Development of an Unsupervised Remote Sensing Methodology to Detect Surface Leakage from Terrestrial CO₂ Storage Sites

by

Rajesh Govindan

A thesis submitted for the degree of Doctor of Philosophy of the University of London and the Diploma of Imperial College

March 2011
Abstract

Carbon Capture and Storage (CCS) is an emerging climate change mitigation technology that can play an important role in preventing the build up of CO₂ in the atmosphere. The potential for CO₂ to migrate from the storage complex towards the surface, eventually resulting in leakage, is the source of significant concern. Therefore, monitoring techniques must be in place to evaluate such risks. Remote sensing has demonstrated success in various environmental monitoring applications. This is attributed to its good areal coverage and continued development in sensor technologies. Since there are no known cases of CO₂ leakages at industrial storage sites, studies are carried out at natural analogue sites in order to develop monitoring methodologies using optical remote sensing. One such study site is located in Latera, a volcanic caldera in west central Italy, with a high geothermal gradient causing production of gases in the subsurface containing CO₂ (>95%) and trace gases like CH₄, H₂S and H₂ (<5%). Another site at the Laacher See lake, located in the East Eifel volcanic district of Renish Massif in Germany, has CO₂ degassing on the surface (upto 99%) at its western shore. Indirect monitoring methodologies developed earlier over these sites using airborne multispectral and hyperspectral datasets are labour intensive in terms of data analysis and interpretation. In order to address the limitations of these approaches, a new unsupervised data processing methodology is proposed to monitor potential CO₂ occurrences on the surface.

This thesis describes the development of an unsupervised methodology based on concepts from geostatistics and probability theory. Prior detection of outliers in the data that correspond to leakages are first determined using a multivariate geostatistical image filtering methodology based on Intrinsic Random Functions of order k (IRF-k) and Independent Component Analysis (ICA). The prior detection is
subsequently updated using a probabilistic information fusion framework, called the Dempster-Shafer (DS) theory of evidence combination, based on the spectral knowledge of indirect effects on the surface environment due to CO$_2$ leakage over time, such as stressed vegetation and mineral alterations, generated by data clustering using the Growing Hierarchical Self-Organising Maps (GHSOM). This enables assigning a confidence measure as posterior detection, thereby identifying potential leakages. Automatic validation of the detection results was carried out with direct field measurements data using a signal theoretic measure for detection performance called the Area Under the Receiver Operating Characteristic curve (AUROC).

The research took a step further by demonstrating that lower resolution spaceborne multispectral data, which are acquired periodically with relatively good spatial and temporal resolutions and a much wider coverage, can also be used for unsupervised detection of leakages. To this end, a study was conducted to systematically degrade the spatial and spectral resolutions of the airborne data of Latera, process the synthetic datasets using the methodology, and compare their respective AUROC measures. Data from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), flying on the Terra satellite as part of NASA’s Earth Observing System (EOS), was also investigated as a case study for Latera.

The validation of the results obtained using all the available multispectral and hyperspectral datasets for Latera and Laacher See study sites demonstrate the potential of the unsupervised methodology as a new technique for rapid detection of surface leakages. However, it is acknowledged that the methodology developed will require more tests and subsequent enhancements in the future for the application to other sites, which differ from Latera and Laacher See, using both the airborne and spaceborne imagery.
DEDICATED TO MY PARENTS
Affirmation

The work submitted in this thesis is my own, and has not been submitted previously for any other degree. The following publications and presentations have resulted from this work:


Acknowledgements

I would like to thank my supervisors Dr Anna Korre and Professor Sevket Durucan for their considerable moral support they have provided me over the past 3 years, and giving me the opportunity to conduct this research. The invaluable discussions I have had with them during this period are intellectually stimulating, and a major factor that has shaped this project to its current form.

I also extend my thanks to Dr. Jonathan Pearce and Mr. Luke Bateson from the British Geological Survey (BGS), Keyworth, UK, and Dr. Michela Vellico from the Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), Trieste, Italy, for providing the multispectral and hyperspectral data that was used to carry out this research.

Thanks to my colleagues and friends from the Mining and Environmental Research Group (MERG) who have accompanied and supported me through these years: Dr. Ji-Quan Shi, Dr. Caglar Sinayuc, Dr. Claire Imrie, Dr. Zhenggang Nie, Dr. Paula Ramirez Muñoz, Giacomo Bacci, Julio Castro Molinare, Maria Lathouri, Sharidah Mohd Amin, James Smith, Meng Wang, Shafiuddin Amer Syed, and Annuwat Wattanawan. Thank you all!

Finally, last, but definitely not the least, I would like to thank my parents and my brother for their support and constant encouragement.
Table of Contents

ABSTRACT  I
AFFIRMATION ........................................................................................................ IV
ACKNOWLEDGEMENTS ............................................................................................ V
TABLE OF CONTENTS ............................................................................................... VI
LIST OF FIGURES ...................................................................................................... X
LIST OF TABLES ........................................................................................................ XV

CHAPTER 1  INTRODUCTION AND OBJECTIVES ................................................ 1
  1.1  INTRODUCTION .............................................................................................. 1
  1.2  RESEARCH OBJECTIVES ............................................................................. 5
  1.3  THESIS STRUCTURE ...................................................................................... 8

CHAPTER 2  METHODS FOR MONITORING CARBON DIOXIDE LEAKAGES ON THE SURFACE ........................................................................................................ 9
  2.1  INTRODUCTION .............................................................................................. 9
  2.2  LONG OPEN PATH INFRA-RED DIODE LASERS ........................................ 10
  2.3  SHORT OPEN PATH INFRA-RED DIODE LASERS ...................................... 13
  2.4  SHORT CLOSED PATH INFRA-RED SYSTEMS .......................................... 15
  2.5  THE EDDY COVARIANCE METHOD ........................................................... 16
  2.6  NEAR SURFACE GAS GEOCHEMISTRY ...................................................... 17

vi
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.7</td>
<td>SOIL GEOCHEMISTRY</td>
<td>20</td>
</tr>
<tr>
<td>2.8</td>
<td>ECO SYSTEM-BASED MONITORING</td>
<td>20</td>
</tr>
<tr>
<td>2.8.1</td>
<td>Monitoring in Latera, Italy</td>
<td>21</td>
</tr>
<tr>
<td>2.8.2</td>
<td>Monitoring at In Salah CO₂ Storage Site</td>
<td>22</td>
</tr>
<tr>
<td>2.8.3</td>
<td>Strengths and Weaknesses</td>
<td>22</td>
</tr>
<tr>
<td>2.9</td>
<td>CONCLUSIONS</td>
<td>23</td>
</tr>
<tr>
<td>3.1</td>
<td>INTRODUCTION</td>
<td>24</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Concepts in Data Acquisition</td>
<td>26</td>
</tr>
<tr>
<td>3.2</td>
<td>DATA PROCESSING</td>
<td>29</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Problems and their Solutions in the Feature Space</td>
<td>29</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Processing the Feature Space for Classification</td>
<td>35</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Processing the Feature Space for Anomaly Detection</td>
<td>38</td>
</tr>
<tr>
<td>3.3</td>
<td>FIELD APPLICATIONS FOR CARBON DIOXIDE LEAKAGE DETECTION</td>
<td>40</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Direct Methods</td>
<td>40</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Indirect Methods</td>
<td>43</td>
</tr>
<tr>
<td>3.4</td>
<td>CONCLUSIONS</td>
<td>53</td>
</tr>
<tr>
<td>4.1</td>
<td>INTRODUCTION</td>
<td>55</td>
</tr>
<tr>
<td>4.2</td>
<td>SPATIO-SPECTRAL DATA PROCESSING WITH GEOSTATISTICAL AND MULTIVARIATE ANALYSIS METHODS</td>
<td>56</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Geostatistical Methods for Structural Representation and Data</td>
<td>58</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Intrinsic Random Functions (IRF)</td>
<td>61</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Multivariate Data Analysis with Independent Component Analysis (ICA)</td>
<td>65</td>
</tr>
<tr>
<td>4.3</td>
<td>DATA CLASSIFICATION USING UNSUPERVISED FUZZY ANALYSIS</td>
<td>69</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Fuzzy c-Means Clustering</td>
<td>71</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Self-Organising Maps (SOM)</td>
<td>73</td>
</tr>
<tr>
<td>4.4</td>
<td>PROBABILISTIC INFORMATION FUSION AND VALIDATION</td>
<td>78</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Information Fusion in Remote Sensing</td>
<td>79</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Dempster-Shafer (DS) Theory of Evidence Combination</td>
<td>81</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Validation using the Receiver Operating Characteristics (ROC) Curve</td>
<td>85</td>
</tr>
<tr>
<td>4.5</td>
<td>CONCLUSIONS</td>
<td>87</td>
</tr>
</tbody>
</table>
# Table of Contents

## CHAPTER 5  FIELD SITES AND DATA PRE-PROCESSING

5.1  **Introduction**  ................................................................. 89

5.2  **The Latera Site in Italy**  .................................................. 90
  5.2.1  **Geology of the Latera Caldera** .................................... 90
  5.2.2  **Surface Lithology of the Latera Caldera Floor** ................. 92
  5.2.3  **Data Available for the Study Site** ................................ 92

5.3  **The Laacher See Study Site in Germany**  .............................. 97
  5.3.1  **Geology of the Caldera** ............................................ 97
  5.3.2  **Data Available for the Laacher See Study Site** ............... 98

5.4  **Data Pre-Processing** ...................................................... 99
  5.4.1  **Noise Reduction** .................................................... 100
  5.4.2  **Dimensionality Reduction** ........................................ 104

5.5  **Conclusions** ............................................................... 110

## CHAPTER 6  UNSUPERVISED LEAKAGE DETECTION USING OPTICAL REMOTE SENSING DATA

6.1  **Introduction** ............................................................... 113

6.2  **IRF and ICA for Prior Detection** ..................................... 114
  6.2.1  **Application of the Prior Detection methodology for the Latera Data** 115
  6.2.2  **Prior Detection using the Laacher See Data** .................... 130

6.3  **GHSOM for Likelihood Map Generation** ................................ 132
  6.3.1  **Likelihood Generation using the Latera Data** ................. 134
  6.3.2  **Likelihood Generation using the Laacher See Data** ........... 140

6.4  **Conclusions** ............................................................... 141

## CHAPTER 7  INFORMATION FUSION FOR POSTERIOR DETECTION

7.1  **Introduction** ............................................................... 143

7.2  **Posterior Detection using DS Theory** ................................ 144
  7.2.1  **Posterior Detection and Validation using Latera Data** ....... 145
  7.2.2  **Posterior Detection and Validation using Laacher See Data** .... 153

7.3  **Conclusions** ............................................................... 154

## CHAPTER 8  DETECTION OF SURFACE LEAKAGES USING LOW SPATIAL RESOLUTION OPTICAL REMOTE SENSING DATA

8.1  **Introduction** ............................................................... 156

8.2  **Variation of Detection Performance with Spatial and Spectral Resolutions** 157
8.2.1 Degradation of the Spatial and Spectral Resolutions of Latera Hyperspectral data ................................................................. 158
8.2.2 CO₂ Leakage Detection using the Synthetic Data Cubes............. 160
8.2.3 Comparison of Relative AUROC Values ................................... 164
8.3 Processing of Latera ASTER Data ............................................ 165
8.4 Conclusions ........................................................................... 167

CHAPTER 9 CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH ................................................................................................................. 169

9.1 Achievements and Conclusions from the Research ....................... 169
9.2 Limitations and Unanswered Questions ........................................ 172
  9.2.1 Early Detection ................................................................. 172
  9.2.2 Change Detection ............................................................. 173
  9.2.3 Leakage Quantification ....................................................... 173
  9.2.4 Masking Factors ............................................................... 173
  9.2.5 Sea Floor Leakages ......................................................... 174
9.3 Recommendations for Future Research ....................................... 175

REFERENCES .................................................................................. 179
List of Figures

Figure 1.1: CO₂ density (and volume) variation with depth (After IPCC (2005)) ........................................... 2
Figure 1.2: A leakage scenario diagram at a hypothetical CO₂ storage site indicating some useful monitoring technologies (After Chadwick et al. (2009)) ........................................... 4
Figure 1.3: The proposed unsupervised methodology for leakage detection using optical remote sensing data ................................................................. 7
Figure 2.1: Air flow (grey arrow) consists of different sizes of eddies (After Burba and Anderson (2007)) ............................................................. 17
Figure 2.2: Study site at Latera indicating the area with highest CO₂ emissions, as a spot lacking all vegetation; towards the outside vegetation changes first to acid-tolerant species and then returns to normal agricultural plants (After Oppermann et al. (2010)) ................................................................. 21
Figure 3.1: The timeline for the Landsat family of satellites starting from the launch of MSS on Landsat 1 and leading up to the Landsat Data Continuity Mission (LDCM), scheduled for December 2012 (After NASA (2010)) ................................................................. 25
Figure 3.3: Pictorial description of a processed data cube from an imaging spectrometer (After CSIRO (2006)) ................................................................. 28
Figure 3.4: Hughes phenomenon of classifier performance, dependent on the number of training samples (m) (After Hughes (1968)) ................................................................. 30
Figure 3.5: Orthogonal projection x' of an arbitrary vector x onto a 2D subspace A in N-dimensional space, and its orthogonal complement x'' ................................................................. 31
Figure 3.6: Application of PCA using a data cloud in 2D space to resolve eigenvectors for subspace determination ................................................................. 32
Figure 3.7: Examples of noise affecting sensor acquisitions: (a) striping, (b) bad lines, and (c) salt and pepper noise (After van der Meer et al. (2009)).

Figure 3.8: (a) Mixtures formed from a heterogeneous surface, (b) Mixtures formed from a homogeneous or intimately mixed surface (After Keshava and Mustard (2002)).

Figure 3.9: Taxonomy for endmember determination algorithms organised based on several criteria such as data interpretation, randomness description and optimisation. Where: DPFT: Dark Point Fixed Transform, FPFT: Fixed Point Free Transform (After Keshava (2003)).

Figure 3.10: A simple representation of statistical testing using GLRT to determine the active hypothesis for a given pixel.

Figure 3.11: (a) Vertical deformation rate derived by stacking of 71 differential interferograms; (b) Areas with absolute value of deformation rate greater than 1 mm/year, at the In Salah CO2 storage site in Algeria (After Onuma et al., 2010).

Figure 3.12: Common surface expressions due to seepage of light and heavy hydrocarbons (After van der Werff (2006)).

Figure 3.13: Map of possible CO2 leakages in the Latera caldera (After Bateson et al. (2008)). The polygon colours correspond to the number of datasets (methods) that showed an anomaly within that polygon.

Figure 4.1: Contour map simulations showing IRF processes with zero mean: (a) k = 0; (b) k = 1; (c) k = 2 (After Chiles and Delfiner (1999)).

Figure 4.2: A 3 x 3 square Laplacian neighbourhood is an ALC for monomials of order k = 1 (note that $\sum_{i=1}^{5} \lambda_i = 0$).

Figure 4.3: A typical trained GHSOM reflecting the structure of data in both lateral and hierarchical directions (After Rauber et al. (2002)).

Figure 4.4: Growing process by inserting row (a) or column (b) of units between e and d (After Rauber et al. (2002)).

Figure 4.5: Example illustration for the types of evidence combinations occurring in information fusion problems: (a) consonant evidences; (b) consistent evidences; (c) arbitrary evidences; (d) disjoint evidences (After Sentz and Ferson (2002)).

Figure 4.6: Confusion matrix for binary classification and the common performance metrics derived from it (After Fawcett (2006)).

Figure 4.7: A basic ROC plot for five different classification/detection algorithms (After Fawcett (2006)).

Figure 5.1: W-E schematic for the geological sections of the Latera caldera (After Annnuziatellis et al. (2008)).

Figure 5.2: (a) The region of Latera in Italy, shown with the flight survey area and Study Area 1, and (b) Study area 1 containing the ground measurement line and CO2 vents, along with local features such as vegetation and road networks.

Figure 5.3: (a) Field gas flux measurements taken at the study area in Latera in 2006 - the boxes indicate known CO2 vents, and (b) DEM data of the study area.

Figure 5.4: (a) The region of Latera in Italy, highlighting features such as the caldera rim and local road networks, and indicating locations of the two study areas; (b) Study Area 2, indicating the ground measurement line and detected regions of high CO2 flux; (c) Study Area 3, indicating all the known CO2 vents in the region.

Figure 5.5: The region of Laacher See lake in Germany (a); indicating the location of the study area on its western shore (b).
Figure 5.6: Image constructed by kriging interpolation of mobile laser data overlaid on an areal photograph (After Jones et al. (2009))

Figure 5.7: Latera Data affected by noise (a) Band 1 of the 2005 hyperspectral data; (b) Band 1 of the 2007 hyperspectral data.

Figure 5.8: Latera data after applying MNF: (a) Band 1 of the 2005 hyperspectral data; (b) Band 1 of the 2007 hyperspectral data.

Figure 5.9: (a) Band 1 of the 2005 multispectral data for Latera affected by noise; (b) Band 1 after applying MNF.

Figure 5.10: (a) Band 1 of the 2007 hyperspectral data for Latera affected by noise; (b) Band 1 after applying MNF.

Figure 5.11: (a) Band 1 of the 2009 hyperspectral data for Latera corrupted by noise; (b) Band 1 after applying MNF.

Figure 5.12: ‘Study Area 1’ data spatial analysis: Cross-variogram functions of eigenvectors (a) for conventional PCA; (b) for proposed approach.

Figure 5.13: ‘Study area 2’ data spatial analysis: Cross-variogram functions of eigenvectors (a) for conventional PCA; (b) for proposed approach.

Figure 5.14: Laacher See data spatial analysis: Cross-variogram functions of eigenvectors (a) for conventional PCA; (b) for proposed approach.

Figure 6.1: Latera ‘Study Area 1’ hyperspectral data (a) 1st eigenvector image; (b) 1st order polynomial trend given by Equation (6.3) for the eigenvector image; (c) Detrended eigenvector image.

Figure 6.2: Effect of increasing neighbourhood size on the frequency response of the ALC-0 filter (a) 3 x 3 filter; (b) 5 x 5 filter; (c) 7 x 7 filter.

Figure 6.3: Latera ‘Study Area 1’ hyperspectral data (a) Detrended eigenvector image; (b) Result obtained after co-kriging with other detrended eigenvectors; (c) Information of the eigenvector that is filtered out.

Figure 6.4: Prior detection result obtained for the Latera ‘Study Area 1’ hyperspectral data after kriging and ICA, superimposed on artificially illuminated elevation data (contains shadows), with black boxes indicating high CO2 flux regions.

Figure 6.5: Latera ‘Study Area 2’ hyperspectral data (a) 1st eigenvector image; (b) 1st order polynomial trend given by equation (6.7) for the eigenvector image; (c) Detrended eigenvector image.

Figure 6.6: Latera ‘Study Area 2’ hyperspectral data (a) Detrended eigenvector image; (b) Result obtained after co-kriging with other detrended eigenvectors; (c) Information of the eigenvector that is filtered out.

Figure 6.7: Prior detection result obtained for the Latera ‘Study Area 2’ hyperspectral data after kriging and ICA, with black boxes indicating high CO2 flux regions.

Figure 6.8: Latera ‘Study Area 2’ multispectral data (a) band 1 image; (b) 1st order polynomial trend given by equation (6.8) for the band 1 image; (c) Detrended band 1 image.

Figure 6.9: Latera ‘Study Area 2’ multispectral data (a) Detrended band 1 image; (b) Result obtained after co-kriging with other detrended bands; (c) Information of the band 1 that is filtered out.

Figure 6.10: Prior detection result obtained for the Latera ‘Study Area 2’ multispectral data after kriging and ICA, with black boxes indicating high CO2 flux regions.

Figure 6.11: Laacher See (a) 1st eigenvector image; (b) 1st order polynomial trend given by equation (6.9) for the eigenvector image; (c) Detrended eigenvector image.
Figure 6.12: Laacher See (a) Detrended eigenvector image; (b) Result obtained after co-kriging with other eigenvectors; (c) Information of the eigenvector that is filtered out.

Figure 6.13: (a) Image constructed by kriging interpolation of mobile laser data overlaid on an areal photograph (After Jones et al. (2009)); (b) Prior detection result obtained for the Laacher See study area after kriging and ICA.

Figure 6.14: SDH representations for different smoothing parameters, after unsupervised learning using a dataset drawn from a Gaussian mixture (After Pampalk et al. (2002)).

Figure 6.15: Hierarchical cluster visualisation of SOMs for Latera data using GHSOM and SDH toolboxes in MATLAB, showing SOM layers 0 (containing 6 units) and 1 (containing 6 SOMs expanded from layer 0).

Figure 6.16: Cluster centres (reflectance signatures) from layer 0 units 1, 2, 4 and 6, for different land cover types in the Latera study area: (a) dense and healthy vegetation; (b) sparse and healthy vegetation and soil mixture; (c) sparse and stressed vegetation and soil (containing mineral alteration) mixture; (d) road network; (e) other unidentified man-made structures.

Figure 6.17: Fuzzy classification map representing the likelihood of CO2 leakage on the surface for Latera ‘Study Area 1’, with black boxes indicating high CO2 flux regions.

Figure 6.18: Fuzzy classification maps representing the likelihood of CO2 leakage on the surface for Latera ‘Study Area 2’, with black boxes indicating high CO2 flux regions (a) using Hyperspectral data; (b) using Multispectral data.

Figure 6.19: Hierarchical cluster visualisation of SOMs for Laacher See data using GHSOM and SDH toolboxes in MATLAB, showing SOM layers 0 (centre) and 1 (surrounding).

Figure 6.20: Cluster centres from Layer 0 units 2 and 8 for (a) stressed vegetation; (b) healthy vegetation in the Laacher See study area.

Figure 6.21: (a) Image constructed by kriging interpolation of mobile laser data overlaid on an areal photograph (After Jones et al. (2009)); (b) Fuzzy classification map representing the likelihood of CO2 leakage on the surface for Laacher See hyperspectral data.

Figure 7.1: Confusion matrix for BPA values range considering an example pair of evidences A and B for DS combination.

Figure 7.2: Posterior detection result obtained for the Latera ‘Study Area 1’ hyperspectral data after applying the DS theory of evidence combination using the prior detection and likelihood maps, superimposed on artificially illuminated elevation data (black patches indicating shadows), with black boxes indicating high CO2 flux regions.

Figure 7.3: Detection result obtained for the Latera ‘Study Area 1’ hyperspectral data after applying the filter proposed by Goovaerts et al. (2005) with the RX statistics, superimposed on artificially illuminated elevation data (black patches indicating shadows), with black boxes indicating high CO2 flux regions.

Figure 7.4: ROC curves comparison between the methodology of Goovaerts et al. (2005) and the proposed methodology for Latera ‘Study Area 1’ hyperspectral data.

Figure 7.5: Detection result obtained for the Latera ‘Study Area 2’ hyperspectral data (a) after applying the DS theory of evidence combination using the prior detection and likelihood maps, with black boxes indicating high CO2 flux regions (b) after applying the methodology of Goovaerts et al. (2005), with black boxes indicating high CO2 flux regions.

Figure 7.6: ROC curves comparison between the methodology of Goovaerts et al. (2005) and the proposed methodology for Latera ‘Study Area 2’ hyperspectral data.
Figure 7.7: Detection result obtained for the Latera ‘Study Area 2’ multispectral data after applying the DS theory of evidence combination using the prior detection and likelihood maps, with black boxes indicating high CO2 flux regions..........................152

Figure 7.8: ROC curves comparison for the methodology using Latera ‘Study Area 2’ hyperspectral and multispectral datasets.................................................................152

Figure 7.9: (a) Image constructed by kriging interpolation of mobile laser data overlaid on an areal photograph (After Jones et al. (2009)); Detection result obtained for the Laacher See study area hyperspectral data (b) after applying the DS theory of evidence combination using the prior detection and likelihood maps; (c) after applying the methodology of Goovaerts et al. (2005)..........................................................153

Figure 8.1: Spatial resolution degradation using Band 1 of the Latera hyperspectral data for (a) 1m (original resolution); (b) 2m; (c) 4m; (d) 8m; (e) 16m....................................................159

Figure 8.2: Detection results using the Latera synthetic datasets at 1m resolution for (a) 63 bands; (b) 32 bands; (c) 16 bands; black boxes indicate high CO2 flux regions.........................161

Figure 8.3: ROC curves comparison for the methodology by using the synthetic datasets with 1m resolutions, containing 63, 32 and 16 bands .................................................162

Figure 8.4: Detection results using the Latera synthetic datasets containing 63 bands for (a) 1m; (b) 2m; (c) 4m; (d) 8m; (e) 16m; black boxes indicate high CO2 flux regions...163

Figure 8.5: ROC curves comparison for the methodology by using the synthetic datasets containing 63 bands at 1m, 2m, 4m, 8m and 16m resolution.................................163

Figure 8.6: The proposed filtering scheme for ASTER data to generate prior detection ..........165

Figure 8.7: (a) Prior detection result for Latera ‘Study area 3’ using ASTER data bands; (b) Likelihood map of CO2 leakages for Latera ‘Study Area 3’.................................166

Figure 8.8: Posterior detection result for Latera ‘Study area 3’ using ASTER data bands; black boxes indicate high CO2 flux regions.................................................167

Figure 9.1: A reconstructed image of the sea floor with gas seepages, showing Pockmarks with different sizes and distribution patterns; the patch seen inside the biggest normal-Pockmark indicates missing data (After Hovland et al. (2010)).................................177
List of Tables

Table 2.1: Wavelengths for CO2 absorption (After Shuler and Tang, 2005) .......................... 11
Table 5.1: The list of Eagle 1K bands used for Latera survey (After Bateson et al., 2008). 95
Table 5.2: The list of ATM bands used for Latera survey (After Pearce et al., 2007) .......... 97
Table 5.3: Summary of ASTER sensor and data characteristics (After Eurimage, 2011). ... 96
Table 5.4: Comparison between the conventional PCA and proposed approach for dimensionality reduction for ‘Study Area 1’ data .............................................. 106
Table 5.5: Comparison between the conventional PCA and proposed approach for dimensionality reduction for ‘Study Area 2’ data .............................................. 108
Table 5.6: Comparison between the conventional PCA and proposed approach for dimensionality reduction for Laacher See data .............................................. 109
Table 6.1: Choosing the candidate to represent prior detection for the Latera ‘Study Area 1’ hyperspectral data (k=4) ................................................................. 121
Table 6.2: Choosing the candidate to represent prior detection for the Latera ‘Study Area 2’ hyperspectral data (k=4) ................................................................. 123
Table 6.3: Choosing the candidate to represent prior detection for the Latera ‘Study Area 2’ multispectral data (k=4) ................................................................. 129
Table 6.4: Choosing the candidate to represent prior detection for the Laacher See study area (k=4) .............................................................................................. 132
Table 8.1: Desired pixel resolutions and their corresponding PSF variances .................... 158
Table 8.2: Comparison of AUROC measures, average values and difference values for different spatial resolutions obtained for Latera multispectral and hyperspectral synthetic datasets ........................................ 163
Table 9.1: Typical values of penetration depths (in metres) for some common low frequencies in different soil types (After Grandjean et al. (1999)). .................... 176
Chapter 1 Introduction and Objectives

1.1 Introduction

Climate change is a problem of global concern that poses many challenges for Earth scientists today. The ever increasing global anthropogenic CO₂ concentrations in the atmosphere emitted by fossil fuel utilisation in human activities, energy production systems, such as electric power generation, and resource extraction and processing industries contribute significantly to this problem. The Intergovernmental Panel on Climate Change (IPCC) reported that, between 1995 and 2001, the average global CO₂ emissions grew at a rate of 1.4% per year, higher than the growth rate of emissions for the previous five years (IPCC, 2005). With such increasing emission rates, the atmosphere will be unable to reach a stabilisation point for greenhouse gas concentrations. Based on many of the models considered by the IPCC, stabilisation occurs at a level not greater than 550 ppmv of CO₂ in the atmosphere. These models indicate that upto 70% reduction in emissions is required by the year 2100 in order to achieve the stabilisation level (IPCC, 2005).

Recent technological research and development has focussed on investigating the idea of Carbon Capture and Storage (CCS), wherein CO₂ is stored in geological formations in the Earth's deep subsurface, such as saline aquifers, oil and gas reservoirs, and coal beds, typically at depths of about 800-1000 m, for long periods
of up to 1000 years. Although there is uncertainty about the total storage capacity of these formations, it is large enough to reduce the rate of emissions. This is attributed to the fact that, owing to high temperature and pressure conditions at such depths, CO₂ is stored in supercritical phase (dense liquid) occupying only about 3.2-11% of the volume it would have occupied otherwise in its gas phase (assuming a geothermal gradient of 25°C/km, surface temperature of 15°C, and hydrostatic pressure (Angus et al., 1976), as illustrated in Figure 1.1). Evidence suggests that, with a probability between 66% and 90%, the net worldwide reservoir capacity available in geological formations with the technical potential to reduce greenhouse gas emissions using CCS is at least 2,000 Giga tonnes CO₂ (GtCO₂). In 2005, the net worldwide anthropogenic CO₂ emissions was reported to be about 24 GtCO₂ per year, which implies that an approximate CO₂ injection time period of 50-100 years may be feasible before the available storage capacity becomes saturated (IPCC, 2005). This data clearly shows that CCS can play an important role in preventing the build up of CO₂ in the atmosphere, and hence mitigate climate change. In the wake of the introduction of CCS, issues from science and engineering to economics are being investigated by researchers to ensure that it provides a safe and sustainable option in the mitigation solutions portfolio.

![Figure 1.1: CO₂ density (and volume) variation with depth (After IPCC (2005)).](image-url)
Even with such large storage capacities and likely time scale, it is important to note that simple physical factors such as the pressure gradient between the surface and reservoir, and the existence of migration pathways, such as faults and abandoned wells, can cause stored CO\textsubscript{2} to migrate towards the surface and eventually back into the atmosphere. In the long term, the interaction between five principal mechanisms in the deep subsurface determine the fate of the stored CO\textsubscript{2} (Metz et al., 2005): immobilisation in structural traps, immobilisation as residual saturation, dissolution into the formation water, geochemical reactions with the rock-forming minerals, and migration out of the reservoir if the caprock seal is fractured. The risk of continuous leakage leading to the loss of a large fraction of stored CO\textsubscript{2} and increasing concentration levels in the vadose zone, equivalent to mole fractions of 0.3 or higher (Farrar et al., 1995, 1999), and \( \geq 1 \% \) by volume in the atmosphere (Smets et al., 2010) can: threaten the health and safety of humans, have adverse effects on flora and fauna, contaminate ground water, and ultimately lead to a regional ecological imbalance. Furthermore, it also represents the risk that CCS may not provide an effective measure to mitigate climate change. Hence, suitable technologies that can reliably monitor potential CO\textsubscript{2} leakage from the reservoir are of great value as they can provide an early warning that warrants necessary mitigation action. Figure 1.2 illustrates a leakage scenario indicating some monitoring methods that are deployed to detect the possibility of CO\textsubscript{2} leakage from a storage reservoir.

Monitoring activities must typically span the entire lifecycle of a storage site that includes: the pre-injection phase, upto 5 years, the injection phase, for a few decades depending on the site and quantity of CO\textsubscript{2} captured, and the post-injection phase, which can last 50 years or more, including closure of the site and transfer of liability to the State. Baseline surveys are generally done at the storage site in the pre-injection phase to provide data for comparison with subsequent monitoring surveys during the injection and post-injection phases. This is particularly important to help understand the changes occurring over time, and relate those changes to any seepage of high concentrations of CO\textsubscript{2} from the subsurface.
Introduction and Objectives

Aifbome monitoring for CO₂ leak
Soil gas survey
Gravimetry
Seismic reflection
Underwater sampling
Permanent soil/gas measurements

Figure 1.2: A leakage scenario diagram at a hypothetical CO₂ storage site indicating some useful monitoring technologies (After Chadwick et al. (2009)).

Monitoring methods for surface leakages are generally classified into two categories, namely: direct techniques, and indirect techniques. The direct methods involve field-based measurement techniques to detect and quantify gas leakage, such as the use of a sampling probe with an Infra Red (IR) gas analyser for local soil sample analyses (Ciotoli et al., 1999), laser measurements for plume concentration (Jones et al., 2009), and measurements of CO₂ flux and concentration using a closed circuit accumulation chamber technique (Hutchinson and Livingston, 1993). Recently, Arts et al. (2009) related these measurements to subsurface fault systems mapped by geophysical techniques, such as 2D reflection and refraction seismics, Multi-channel Analysis of Surface Wave (MASW), Ground Penetrating Radar (GPR), and 2D/3D geo-electrical and electromagnetic properties among others. On the other hand, indirect methods involve using remote sensing techniques to study the effects of leakage on the surface environment, such as vegetation stress (Pickles and Cover, 2004; Bateson et al., 2008; Keith et al., 2009; Lakkaraju et al., 2010; Male et al.,...
Introduction and Objectives

2010; Rouse et al., 2010), mineral alterations patterns (Pickles and Cover, 2004), and surface deformation (Onuma and Okhawa, 2009; Tamburini et al., 2010).

Today, airborne and spaceborne remote sensing platforms, such as the Radio Detection And Ranging (RADAR), Light Detection And Ranging (LIDAR), and multispectral/hyperspectral optical and thermal sensor technologies, are playing an important role in environmental monitoring applications. These technologies are attractive because they provide: good areal coverage considering the large surface area that may be affected by potential CO₂ seepages from a leaking storage complex, synoptic information of the surface in different regions of the electromagnetic spectrum that are complimentary, and good temporal information. Moreover the data are collected in a non-invasive manner, unlike field-based techniques, and are also relatively cheaper to acquire. As CO₂ leakage and related effects on the surface environment exhibit diagnostic features (signature) in the electromagnetic spectrum, there is a strong scientific basis for pursuing research using remote sensing data for CO₂ leakage identification.

For the research described in this thesis, datasets acquired using airborne and spaceborne multispectral/hyperspectral optical remote sensing sensors were investigated. The primary reason for this choice, as opposed to RADAR and LIDAR data, is that information of the surface is acquired in many contiguous wavelengths that allow discriminating CO₂ leakages, directly or indirectly, from other background clutter information. In particular, the spectral signatures containing diagnostic features of the effects on the surface, such as vegetation stress and mineral alterations, are used in order to improve the indirect detection of potential leakages by using unsupervised information processing of the datasets. This gives the leakage locations with a higher confidence that can then be used as independent information for designing field surveys to take direct measurements.

1.2 Research Objectives

The fundamental problem associated with existing indirect techniques that have been developed using optical remote sensing data, based on vegetation stress and mineral alterations, is that they involve manual effort in data collection and interpretation that could be error prone or time consuming. Such techniques may also be referred to as supervised techniques since they rely on a priori knowledge on the locations of
leakage points and their features. Using this knowledge, an extrapolation is generally performed to detect other potential leakages over a larger area. However, surface complexities such as the presence of different land cover types, terrain elevation, varying reflectance and emissivity conditions across a homogenous surface due to the position of the Sun (azimuth and elevation angles change according to the season), among others, can increase false detection rates in current supervised techniques.

It is therefore essential to enhance these methods or develop new ones that can be used to detect locations of potential CO$_2$ leakage in an unsupervised manner using optical remote sensing data. Once such locations are known, the area covered and time spent on monitoring and the collection of field measurements can be reduced significantly. The main objectives of the research described in this thesis were:

(1) to develop an unsupervised methodology using airborne and spaceborne optical remote sensing data for terrestrial ecosystems to detect leakages.

(2) to apply the methodology to different study sites and ensure that it detects leakages consistently.

(3) to assess the performance of the methodology using synthetic datasets generated for different spatial and spectral resolutions.

(4) to develop a general understanding of the types of optical remote sensing data that are mostly applicable for leakage detection.

In order to achieve the stated objectives, an unsupervised methodology based on geostatistical and probabilistic image processing was developed using high resolution airborne hyperspectral and multispectral datasets acquired over analogue study sites located in Latera, Italy, and Laacher See, Germany, where natural seepage of CO$_2$ occurs from the subsurface. The underlying idea behind the methodology is to model the spatial correlations in the images and filter out data redundancies using geostatistical methods to obtain residuals, followed by the estimation of a hidden residual variable based on using Independent Component Analysis (ICA). The hidden variable, as a prior distribution, was fused with fuzzy evidence (likelihood) that conveys the surface effects due to CO$_2$ leakage using the Growing Hierarchical Self-Organising Maps (GHSOM) and the Dempster-Shafer (DS) theory of evidence.
combination to generate a posterior confidence map showing anomalous areas related to leakage. These will be explained in detail in subsequent chapters; Figure 1.3 illustrates a schematic for the proposed unsupervised methodology.

Since the cost and effort involved in collecting data using on-demand airborne surveys are huge, the use of low resolution satellite data was also considered for leakage detection. Such data are readily and cheaply available for environmental research today. They have reasonably good spatial and temporal resolution for detecting changes on the surface environment over a very large area. Hence, the methodology was also applied using spaceborne multispectral data for the Latera site acquired by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor.

![Diagram: Optical remote sensing data](image)

**Figure 1.3:** The proposed unsupervised methodology for leakage detection using optical remote sensing data.
1.3 Thesis Structure

Chapter 2 reviews various field based monitoring techniques. Advantages and disadvantages for each technique are stated in the context of CO₂ leakage monitoring.

Chapter 3 reviews optical remote sensing and its application as both direct and indirect techniques to monitor leakages, along with other remote sensing technologies such as RADAR and Electromagnetic (EM) methods.

A detailed literature review describing all computational techniques that are used to develop the unsupervised methodology for processing optical remote sensing data for leakage detection is presented in Chapter 4.

In Chapter 5, the Latera and Laacher See study sites are described and the airborne and spaceborne data sets acquired over these sites are introduced along with other field data used for validation purposes. The pre-processing of the data is explained here as the starting point for the methodology.

Chapter 6 explains the first stage of the methodology by introducing geostatistical filtering with Intrinsic Random Functions (IRF), multivariate analysis with Independent Component Analysis (ICA), and fuzzy analysis with the Growing Hierarchical Self-Organising Maps (GHSOM) using the airborne data to generate the prior detection and likelihood maps.

Chapter 7 presents the second stage of the methodology using the Dempster-Shafer (DS) theory of evidence combination to generate the posterior detection maps and is followed by validation using the Receiver Operating Characteristics (ROC) curve for airborne data.

In Chapter 8, the performance of the methodology is tested using systematically degraded synthetic datasets generated from airborne hyperspectral data in order to understand the behaviour of the methodology at low spatial and spectral resolution cases. ASTER satellite data for the Latera site are also processed using the methodology to test the potential for leakage detection of low data resolution.

Finally, Chapter 9 presents the main conclusions and accomplishments of the research and makes recommendations for future research.
2.1 Introduction

Monitoring the surface occurrences of greenhouse gases on the surface, such as CO$_2$ and CH$_4$, with surface based instrumentation is a widely used technique. Various surface monitoring methods that are available today were developed by exploiting the knowledge of the physical and chemical nature of these gases, and their interaction with the atmospheric or surface boundary layer. Since some of these methods can allow both the detection and quantification of gases emitted at ground level, they have been investigated widely for CO$_2$ storage monitoring purposes in recent years.

The knowledge of the interaction of CO$_2$ with light, and its absorption features in different regions of the electromagnetic spectrum, has led to the implementation of Infra-Red (IR) diode laser technology with the capability of selectively tuning into CO$_2$ absorption wavelengths to locate the plume rising above the ground surface. Near-surface geochemical methods help in identifying the rising gas fluxes as those originating from either biogenic or geogenic sources. It is also possible to understand the effects of these fluxes, at varying rates, on the surface environment using soil geochemistry and eco-system monitoring. Hence, a detailed review is in order to
describe the various surface monitoring methods and place them in the context of CO₂ storage monitoring activities.

### 2.2 Long Open Path Infra-Red Diode Lasers

Various open path sensing techniques have been developed that measure the path-integrated concentration of a target gas between two points near the ground surface. These methods have been used to detect emission points from various sources, such as landfills, coal mines, and wastewater treatment plants, and to estimate their leakage rates to the atmosphere (Piccot et al., 1996; Hashmonay et al., 2001; EPA, 2006). Recently their potential for monitoring CO₂ geological storage sites has also been explored (Annunziatellis et al., 2007; Humphries et al., 2008; Loh et al., 2009; Trottier et al., 2009).

A number of different open path techniques that were developed for surface monitoring of gas plumes including: Open-Path Fourier Transform Infrared (FTIR) Spectroscopy, Ultra-Violet Differential Optical Absorption Spectroscopy (UV-DOAS), Open-Path Tuneable Diode Laser (TDL), and Absorption Spectroscopy; Differential Absorption LIDAR (DIAL) (EPA, 2006). To date, the TDL and, to a much lesser extent, the FTIR have been applied more in the field of CO₂ monitoring research and development.

TDL, which consist of an integrated transmitter/receiver unit and a reflector, measures the distance-averaged concentration of a specific gas in the near surface atmosphere. A signal emitted from the transmitter propagates through the air to the reflector (positioned a few to hundreds of metres away) and returns back to the instrument where it is focused onto a detector. Because of absorption of the signal by the gas of interest, the decrease in the signal returned to the detector is directly proportional to the total amount of that gas over the entire path length.

The transmitter consists of a tuneable diode laser that emits light at a wavelength that is absorbed by a gas. It typically absorbs at more than one wavelength, each having its own absorption strength. For CO₂ gas, the level of light absorption varies markedly between the 5 main absorbing wavelengths for this gas, listed in Table 2.1. This variability among different wavelengths can be exploited based on the sensitivity and interference requirements of the application, e.g. portable CO₂ Infra-
Red (IR) detectors equipped with an internal short-path cell usually use high absorption bands (e.g. 4.255 \( \mu \text{m} \)) to increase sensitivity. Such highly absorbing bands, however, are not appropriate where large amounts of CO\(_2\) result in significant signal loss, such as in high concentration conditions or over long path lengths, as in the case for long open path systems. Such systems often employ the 1.58 \( \mu \text{m} \) wavelength (Flesch \textit{et al.} 2004; Trottier \textit{et al.} 2009), which has a much lower absorption strength and very little interference, although the band near 2.004 \( \mu \text{m} \) has also been applied (Humphries \textit{et al.} 2008).

**Table 2.1:** Wavelengths for CO\(_2\) absorption (After Shuler and Tang (2005)).

<table>
<thead>
<tr>
<th>Wavelength (( \mu \text{m} ))</th>
<th>Relative absorption strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.432</td>
<td>1</td>
</tr>
<tr>
<td>1.580</td>
<td>3.7</td>
</tr>
<tr>
<td>2.004</td>
<td>243</td>
</tr>
<tr>
<td>2.779</td>
<td>6,800</td>
</tr>
<tr>
<td>4.255</td>
<td>69,000</td>
</tr>
</tbody>
</table>

The choice of the reflector is important in order to balance the amount of light returned to the detector to obtain a good sensitivity and avoid detector saturation. The type and size of the reflector should be chosen based on the optical path length and on the expected amount of monitored gas. For long path lengths, reflectors typically consist of a series of aligned, gold-plated mirrors that maximise light reflection.

Sensitivity of the device will also depend on the gas type being measured, e.g. the relative sensitivity for CH\(_4\) can be 10 times that for CO\(_2\). This will also have an impact on the maximum path length for each gas type measurement. Because the measurements are related to the linear distance of the total path covered by the laser, i.e. twice the distance between the transceiver and reflector, the units are parts per million metres (ppmm). These values can be divided by the total path length to give a ‘path-averaged’ concentration in ppm. Therefore, although longer path lengths ensure that larger areas are monitored, there are problems such as the loss of resolution and greater dilution of a leakage signal thereby reducing the possibility of a plume being detected.

Although originally designed for the monitoring of industrial installations, such as gas pipelines or factory emissions, long open path IR laser systems have also
recently been applied to the monitoring of geological systems. One of the first applications involved study of the release of deep natural gases from volcanic or geothermal systems, and their eventual movement and dispersion to the atmosphere. Belotti et al. (2003) used long open path systems to investigate a gas fumarole located in the ‘Solfatara’ caldera in the Pozzuoli area, near Naples, Italy, in order to map CO2 concentration distribution in the area. Cuccoli et al. (2007) also applied this method at a geothermal degassing point near Siena, and a landfill site near Florence.

Both CO2 and CH4 sensing lasers were recently tested at the Pembina CO2-Enhanced Oil Recovery (EOR) project, located west of Edmonton, Canada (Trottier et al. 2009). Between 50 and 100 tonnes of CO2 have been injected every day since Spring 2005 into the 1,650m deep Cardium formation at this site. Monitoring was conducted during the month of February for three successive years. Although the chosen period minimised variability caused by biological activity, there were logistical problems because of snow cover and freezing temperatures.

A new laser design has also been tested at a controlled CO2 release facility established at a relatively flat, agricultural field at Montana State University (Humphries et al. 2008). In this study a controlled underground release of 300 kg/day was monitored both above and below ground using a distributed feedback diode laser that scans the 2.0027-2.0042μm region. During the CO2 release, the measured average atmospheric CO2 concentration over the injection well was 618 ppm whereas that at the background site was 448 ppm. However, wind conditions during the experiment were not considered.

The long open path technique has been applied at these sites due to its number of advantages with respect to monitoring CO2 leakages. Firstly, the large atmospheric volume that it can cross gives this method the potential to intersect a temporally and spatially variable plume. Secondly, because the lasers can be mounted on automated rotating platforms controlled by aligning software, and most units have an internal reference cell for self-calibration, this approach is well adapted for long-term, unattended monitoring of a site. Combining these advantages means that it may be possible to monitor a large area with a single laser unit and a number of fixed reflectors, i.e. optical pathways. The deployment of multiple open path units has also
been proposed (Loh et al., 2009) although issues related to differential instrument drift must be addressed.

Despite these advantages there are certain limitations that will affect the eventual deployment of this technique. Although it may be physically possible to measure across great distances with the laser, the fact that the method yields a distance-averaged concentration means that with increasing path length there is a corresponding reduction in resolution and sensitivity due to dilution of the plume signature by background atmospheric air. Because the open path technique measures along a line there are also a number of deployment issues that must be considered. For example the pathway between transmitter and reflector could be completely blocked by accumulating snow or growing vegetation, meaning that the optical pathway cannot be placed too close to the ground surface otherwise pathway maintenance would be required. In addition, the signal will also be attenuated by falling snow, rain, fog and dust, all of which may also coat the optical windows of both the transceiver and retro-reflector. Passing cars and people will also block the light path, and safety issues related to potential eye damage of people looking into the laser should be considered. Finally, a single deployment of an open path technique is not capable of locating a leak on its own because of the path averaged concentration measured by the methods. That said, multiple pathways combined with wind data may be able to define its general location if it is within the monitoring network, or indicate a direction if it is outside the monitoring area. More detailed point measurement techniques, e.g. gas flux, would then be needed to precisely locate leakage.

### 2.3 Short Open Path Infra-Red Diode Lasers

The basic principle of the short open path technique is that transmitter and receiver are separated by a fixed distance (typically on the order of 1-2 m), between which there is a free exchange of atmospheric air. Compared to long open path instruments, which are usually fixed installations at ground surface that provide distance-averaged concentrations, short open path equipment is mounted on ground or airborne vehicles for mapping of point concentrations. Due to the fact that the sensor measures concentrations directly in the air, unlike the closed path systems, response times are very rapid, and thus surveys can be conducted at high speeds. Also by reflecting the
sensing beam multiple times across the unit casing, sensitivity can be increased while maintaining a small physical dimension, which may be necessary both for logistical reasons and to provide a ‘point’ measurement.

Short open path lasers were first developed as a tool for locating natural gas leaks along pipelines. Given the great distances involved through often highly vegetated and uninhabited areas, an airborne technique with rapid response times was required to efficiently and cost-effectively monitor this infrastructure. CH₄ sensors were the first units tested and have been in operation for over 10 years. Although the CO₂ unit is less sensitive than that for CH₄, this sensor has recently undergone significant technological advances that have greatly improved its performance.

To date only the TDL systems have been tested in the field of CO₂ monitoring, with research conducted into the use of the ground based CO₂ detecting unit at natural analogue sites for leak locating and monitoring purposes only (Jones et al. 2009; Kruger et al. 2009). In these studies, a Boreal Laser GasFinderAB was mounted on an All Terrain Vehicle (ATV) and surveys were conducted first at the Latera test site in Italy, and then subsequently two times at the Laacher See site, in Germany. The results obtained from the latter study site have been used for the research described in this thesis (see chapter 5, section 5.3).

Surface gas monitoring was also carried out at the In Salah Gas project in 2009 using a Boreal Laser open path laser CO₂ detector, linked to a gasFinder FC analyser mounted on a 4-wheel drive vehicle (Jones et al., 2010). The detector used a wavelength of 2 μm and had a sensitivity of around 5-10 ppm for CO₂. The first traverses were made around the injection well ‘KB-502’, and subsequent measurements were carried out within the area between the ‘KB-502’ and ‘KB-5’ wells. The majority of the measurements were close to a typical atmospheric level of around 380 ppm CO₂. The highest values were observed around KB-502, but were attributed to vehicle exhaust.

In summary, some of the advantages of short open path systems include:

- temperature changes, ground, and cloud conditions do not affect the performance of the TDL.
direct measurement of concentrations in the air results in very rapid response times, thereby permitting fast surveying of large areas;

internal calibration and lack of moving parts should result in good instrument stability and low maintenance costs; and

relatively small size and weight compared to long open path instruments, facilitating the mounting of the instrument on ground or airborne vehicles.

However, disadvantages of short open path systems include:

changes in wind speed and direction will greatly affect the series of absolute and relative point measurements in the atmosphere; and

the influence of background variations in CO₂ concentrations on leak detection capability has not been quantified, i.e. false positives due to natural variability.

2.4 Short Closed Path Infra-Red Systems

Short closed path IR detectors involve the introduction of a gas sample into a ‘closed’ chamber via a pump or by diffusion, and subsequent detection and quantification of a specific gas component by passing a light across the chamber. Thus these methods are similar to long open path and short open path due to the use of optical sources and detectors, but differ due to the use of the measurement chamber. There are two types of IR detectors, namely, non-dispersive IR (NDIR), and dispersive detectors. In a non-dispersive IR detector all the light from the source passes through the sample, after which it is filtered to the desired wavelength. In a dispersive system a grating or prism is used prior to the sample to select a specific wavelength, and thus only this wavelength passes through the sample on its way to the detector. Dispersive IR detectors are typically used in laboratory instruments because they can scan a wide range of wavelengths, however they tend to be more costly and less suitable for portable instruments. Instead NDIR units are the most commonly used short path closed detectors for field applications, due to their robust nature, portability, low cost, stability and selectivity.

Short closed path TDL are also a possibility for CO₂ monitoring applications, as illustrated by work conducted to monitor volcanic gases (Gagliardi, et al. 2002; Richter et al., 2002). However, their large size and greater expense makes widespread deployment more difficult. They do offer the potential for real-time
Methods for Monitoring Carbon Dioxide Leakages on the Surface

Isotopic analysis (δ¹³C) which could provide important data for separating near-surface biological CO₂ from leaking CO₂ (Fessenden et al., 2010). The TDL short closed path instruments are most commonly used along with the eddy covariance technique (Griffis et al., 2004).

An application of NDIR CO₂ sensors, as presented by Lewicki et al. (2010), was to continuously monitor CO₂ concentrations over time during a long-term, shallow CO₂ injection test. The results were used to discern background levels and related diurnal fluctuations, determine leakage breakthrough, and to correlate concentration trends with various atmospheric parameters.

Advantages of the NDIR instruments are their low cost, their flexibility both in terms of concentration range and individual deployment configurations, and the limited interferences from other gases (depending on the wavelength of the chosen detector). In particular, the low cost per NDIR sensor means that a relatively large number could be deployed, while their small size and point measurement nature can allow for deployment also close to the ground (i.e. avoiding line-of-sight or vegetation problems). However, the effect of wind dilution and dispersion on sensitivity remains a disadvantage for NDIR instruments, as with most of the atmospheric methods.

2.5 The Eddy Covariance Method

One of the most useful methods to measure and determine gas fluxes in the atmospheric boundary layer (surface layer or constant flux layer) is the Eddy Covariance Method (ECM). The ECM uses statistics to compute turbulent fluxes of heat, water and gas exchange, e.g. CO₂, CH₄ and trace gases. It averages the integral flux of gases over larger areas and different temporal scales (Baldocchi et al., 2001; Lewicki et al., 2009). Although the method is very complex in terms of hardware design and processing the large amounts of data, it has been proposed as a potential methodology for monitoring CO₂ storage sites (Miles et al., 2005; Oldenburg et al., 2003; Leuning et al., 2008).

The ECM requires a great deal of specific knowledge, in particular regarding the application of mathematical corrections and processing workflows for the specific purpose of measurement. The ECM relies on the assumption that air flow between
the surface and the atmosphere is occurring by turbulent movement. These small confined turbulences are called eddies and the air flow can be imagined as a horizontal flow of numerous rotating three dimensional eddies with different sizes distributed over the measurement height, as illustrated in Figure 2.1.

![Figure 2.1: Air flow (grey arrow) consists of different sizes of eddies (After Burba and Anderson (2007)).](image)

Though more expensive and technically more complex, the eddy covariance method provides a powerful tool that allows for spatial integration and near-continuous, long-term monitoring of the soil-atmosphere flux. This holds for greenhouse gases, such as CO₂ and CH₄ (Smeets et al., 2009), but the application of this method is restricted to onshore sites. Whether the ECM can detect a release of CO₂ from a storage site strictly depends on the ratio between the integral CO₂ flux from the footprint area and the seepage rate from the point sources, e.g. a seepage rate of 0.1 tonnes/day from a release experiment (Lewicki et al., 2009) was not distinguishable from the background CO₂ levels, whereas the release of 0.3 tonnes/day significantly increased the measured flux rates compared to the base line emission of the area. In a recent application at the CO2CRC Otway Project in SE Australia to monitor atmospheric CO₂ and test the possibility of detecting leakage (Etheridge et al., 2010), it was found that under high natural background variability, ECM would be unlikely to be able to detect elevated CO₂ emissions from a leakage point.

2.6 Near Surface Gas Geochemistry

Near surface gas geochemistry for CO₂ detection can be divided into two general areas of study: gas flux measurements at the ground surface, and analysis of soil gas collected from a near-surface region. A short overview of these two techniques is given below.
There are primarily five methods that have been commonly applied to the measurement of gas flux across the ground-atmosphere boundary: alkali absorption, open chamber, vertical profiling, Closed Chamber (CC), and Dynamic Closed Chamber (DCC) (Bekku et al., 1997). The CC and DCC techniques involve the monitoring of gas concentration changes over time within an accumulation chamber placed on the soil surface, with samples collected manually at pre-determined times in the CC method and continuously (typically every second) via an in-line detector in the DCC method. One significant difference between these two approaches is that a number of gas samples must be removed from the chamber using the CC method, whereas the DCC technique is a closed loop system. This means that the CC chamber volume must be large or sample volume small to minimise any pressure drops during sample extraction. The light weight, rapid response (usually 1-2 minutes), and good precision of IR DCC (typically ± 10%) has made it the most widespread technique for measuring CO2 flux rates in various applications, including CO2 storage monitoring. Gas flux measurements with the DCC technique are usually conducted manually via gridded surveys, although automated systems do exist for continual monitoring of a small site, e.g. at well heads.

Soil gas samples are most typically collected using small, lightweight soil probes. The method involves driving a hollow steel tube into the ground, typically to a depth of 0.5-1.0 m, and drawing soil air to the surface for field or laboratory analysis. Alternative sampling methods involve direct push, power hammered, augered or drilled systems. These are slower, less portable, and more costly than the simple probes but may allow sampling in difficult ground, or at greater depths where atmospheric effects are less significant, assuming a deep water table. Samples are collected at surface by drawing gas up from a sampling port on the probe, with the gas either being analysed in the field using portable equipment or stored in pre-evacuated airtight containers for eventual laboratory analysis.

In addition to CO2, various other gas species are also studied for flux and concentration due to their association with the reservoir, e.g. CH4 or H2S, and tracers, such as SF6, that are added to the injected stream (e.g. fluorocarbons), or natural tracer gases, such as He or Rn. Isotopic analyses are also conducted; the most common are those of carbon in CO2 (δ13C to determine origin and δ14C to determine age). The primary motivation for analysing various gas species is to help distinguish
Methods for Monitoring Carbon Dioxide Leakages on the Surface

CO₂ which might be leaking from a geological storage reservoir from that which is produced naturally in the shallow subsurface via biological processes, e.g. plant or microbial respiration, or inorganic processes, e.g. groundwater degassing. Klusman (2003) used both stable and radiogenic carbon isotopic data, combined with CO₂ flux data and other measurements, to estimate the amount of potential CO₂ leakage at surface from the Rangely CO₂-EOR project in Colorado, USA. Numerous near surface gas geochemistry surveys have been conducted at other industrial CO₂ storage or test sites, such as Weyburn (Jones and Beaubien 2006), In Salah (Jones et al., 2006), and Recopol (van Eijndthoven, 2005). However, no evidence of leakage has been observed at these sites.

Baseline surveys were performed at the CO₂CRC Otway Project in SE Australia between March 2007 and March 2010 (Schacht et al., 2010). Monitoring was focused on CO₂, CH₄ and He concentrations. A subset of samples was analysed for δ¹³C and δ¹⁴C in CO₂, and sulphur hexafluoride (SF₆) tracer added to the injected CO₂. The CO₂ concentrations within the soil were found to be moderately high. However, the SF₆ tracer was not detected in any of the samples collected. Analyses of He, δ¹³C and δ¹⁴C indicated that the primary source of the soil CO₂ is the decomposition of organic matter, and that deep subsurface CO₂ was negligible.

As shown by the studies described above, the greatest advantage of these techniques is the fact that they are able to make direct measurements of transfer rates, and thus with judicious selection and design of sampling grids they have the potential to provide the most accurate estimate of CO₂ leakage rates. However, there are two major issues that one must contend with in order to improve the quality of estimates and its timeliness. The first is related to the methods’ ability to locate and clearly delineate the boundaries of any leakage feature. This will depend on the ratio between leakage and total area, between anomalous versus baseline values, and on the capability of the method to correctly identify a leakage anomaly. Continued work on devising improved sampling strategies and better integration with other techniques, e.g. detailed geological site knowledge, and eddy covariance towers, hold great promise for future applications. Secondly, the baseline contribution of shallow biological CO₂ flux must be accurately quantified so that it can be removed from the total flux rate measured for the leakage area. To accomplish this, representative baseline CO₂ flux and soil gas measurements must ideally be
conducted prior to injection using manual spatial sampling and/or automatic temporal sampling.

### 2.7 Soil Geochemistry

Mineralogical studies of the clay-rich soils of the natural CO₂ leakage site at Latera, Italy, have indicated variations in soil geochemistry associated with increased acidity and anoxic conditions (Beaubien et al., 2008; Pettinelli et al., 2008). In particular, the results showed an increase in the concentration of K-feldspar with an associated decrease in albite, and a decline in the occurrence of oxides such as MgO, CaO, Fe₂O₃ and Mn₃O₄ in the region of the gas vent compared with the surrounding soils. This information is used as evidence for leakage in the research described in this thesis (see chapter 6, section 6.3). According to Stephens and Hering (2002; 2004), who reported differences in geochemical composition for volcanic soils with elevated CO₂ levels at Mammoth Mountain, California, the dissolution of minerals in soils is not directly affected by CO₂, but by the reduction in pH due to the increase in organic acids. Therefore, any increase in CO₂ levels in soils would have to cause a decrease in pH before changes in the mineralogical composition could be observed. However, changes in pH could also result from degradation of organic matter.

### 2.8 Ecosystem-based Monitoring

Ecosystem-based monitoring can be employed to qualitatively detect and monitor leakage into near surface systems. This technique can be used for terrestrial and freshwater as well as marine ecosystems. Ideally, ecosystem modelling is combined with a soil gas survey of the area assumed to be affected by leakage from deep reservoirs.

Generally, the monitoring of terrestrial ecosystems combines a detailed analysis of the non-mobile live form, i.e. plants, meiofauna and the microbial populations inhabiting the soil. This includes the identification of the bio-diversity present together with other parameters, including microbial activities and geochemical parameters as indicators for the health status of an ecosystem. The plant and microbial populations determined at a suspected leakage point are then compared with control sites representing the background to be expected without disturbances.
2.8.1 Monitoring in Latera, Italy
To better understand the potential impact of a leakage event, a detailed geochemical and biological study was conducted during two different seasons on a naturally occurring gas vent located within a Mediterranean pasture ecosystem, shown in Figure 2.2. Results from botanical, soil gas, microbiological and gas flux surveys have shown that a significant impact is only observed a few metres either side of the vent, where CO$_2$ flux rates are extremely high. In this 'vent core' there is no vegetation, pH is low, and small changes are observed in mineralogy and bulk chemistry (Beaubien et al., 2008).

![Figure 2.2: Study site at Latera indicating the area with highest CO$_2$ emissions, as a spot lacking all vegetation; towards the outside vegetation changes first to acid-tolerant species and then returns to normal agricultural plants (After Oppermann et al. (2010)).](image)

Microbial activities and populations are regulated in this interval by near-anoxic conditions, and by elevated soil gas CO$_2$ and reduced trace gases (CH$_4$, H$_2$S, and H$_2$). An approximately 20 m wide area surrounding the core forms a transition zone, over which there is a gradual decrease in CO$_2$ concentrations, a rapid decrease in CO$_2$ fluxes, and the absence of reactive gas species. In this transition zone acid-tolerant grasses dominate the near vent core, but these are progressively replaced by clover and a greater plant diversity moving away from the vent centre (see Figure 2.2). Physical parameters, e.g. pH, bulk chemistry and mineralogy, and microbial systems also gradually return to background values across this transition zone. Results indicate that, even at this anomalous high-flux site (2,000-3,000 gm$^{-2}$/day), the effects of the gas vent are spatially limited and that the ecosystem appears to have
adapted to the different conditions through species substitution or adaptation. Further details of this site, including the data collected, are described in a later chapter (see chapter 5, section 5.2).

2.8.2 Monitoring at In Salah CO₂ Storage Site
A botanical and microbiological survey was also undertaken for the In Salah Gas project in 2009 (Jones et al., 2010). The survey area was focused on the three injection wells, with samples taken along transects at a spacing of 100-200 m, and priority was given to vegetated sites. The survey revealed flora typical of a desert environment, with vegetation generally sparse but more abundant in the dry wadi channels. A number of angiosperm plants were found, which gives rise to the potential for CO₂-sensitivity, as observed in similar communities in temperate climates (Beaubien et al., 2008; West et al., 2009). Several of the microbial samples were found be devoid of detectable life. However, suggestions were made that it would be informative to undertake a similar survey after a period of rain to ascertain the microbiological background and change in flora under these conditions.

2.8.3 Strengths and Weaknesses
The monitoring of terrestrial and freshwater ecosystems can provide information to identify leaks from deep subsurface reservoirs into surface systems at a low cost. The sites affected by high CO₂ concentrations are typically visible by bare eye and thus easy to detect, even from a distance or low-flying aircrafts. Effects of CO₂ on biota are mediated through lowering of pH (Magnesen and Wahl, 1993), which may have a toxic effect on the fauna or a hampering effect on the production of calcareous structures. The tolerance will vary from one species to another. If CO₂ that leaks causes effects on the benthic layer community level, e.g. causing changes in tropic dynamics, shift in relative abundance of feeding types, changes in diversity and dominance patterns, such results may be transferable to mesopelagic communities, and may be quite easily detected.

However, to sufficiently describe the health status of an affected ecosystem a detailed and comprehensive assessment of its biodiversity, including plants, meiofauna and microfauna has to be conducted. This is time consuming and requires substantial expertise by the investigators.
2.9 Conclusions

There are various established methods available for performing surface monitoring of CO₂ leakages from storage reservoirs. From example applications, particularly those of open path and close path techniques, it is observed that although plume detection is generally possible there are some problems that have not been addressed. For example, it is not possible to perform monitoring under adverse weather conditions such as rainfall and strong winds that cause dilution of plume concentration. Moreover, for all techniques, including the Eddy Covariance Method (ECM), natural background variability can also pose a major issue for CO₂ plume detection even at high concentrations.

Surveys for the selection of the optimal locations for the placement of monitoring instruments must be done prior to actual monitoring surveys, which introduces logistical problems and increases the effort, time and cost. The choice of suitable locations is based on the absence of natural or artificial obstructions that may lie in the instrument’s line of sight. This, however, may not be an issue for the short open path methods as portability allows the instrument to be mounted on mobile ground-based or airborne platforms.

Studies related to near surface gas geochemistry, soil geochemistry and eco-system monitoring have provided a good insight on various techniques that can reveal locations of leakages based on high CO₂ flux rates and concentration in the soil, geochemical alterations, and impact on the local ecosystem, including plants and animals that thrive in the vicinity. However, the reliability of the analyses depends on good baseline surveys and subsequently monitoring changes occurring on the surface due to CO₂ leakage.

Generally speaking, all surface based methods discussed in this chapter rely on the knowledge of the locations of potential leakages. The lack of such prior knowledge combined with a poor spatial coverage is therefore a major drawback of these methods. Airborne and spaceborne remote sensing technologies provide a relatively much wider coverage in surface information, which has been an important reason for its success in various environmental monitoring applications. The next Chapter therefore provides a review on remote sensing techniques and their application for monitoring CO₂ storage sites.
3.1 Introduction

For over four decades, optical remote sensing technology has been extensively used to study the spectral properties of materials on the Earth’s surface. It began as a result of the response by the United States to the launch of the first Earth orbiting artificial satellite called the Sputnik by the former Soviet Union on 4th October 1957, which instigated the ‘Space Race’ during the Cold War period. The response was immediate and led to the formulation of the National Aeronautics and Space Act in 1958, and thus the inception of the National Aeronautics and Space Agency (NASA) (Landgrebe, 1997). These events caused scientists in the 1960s to think about how views of the Earth from space could be generated effectively using data analysis and interpretation for surface data acquisition.

The modern era of remote sensing of the Earth was pioneered with the development and deployment of the Landsat Multi-Spectral Scanner (MSS) system by NASA in 1972. For the time, good quality data was acquired with: high ground resolution (80m); wide spatial coverage (185 x 185 km²); good repeat coverage (18 days) and multispectral information (4 wavelengths). Since the MSS, several other multispectral sensors in the Landsat family, namely, Thematic Mapper (TM) and
Enhanced Thematic Mapper (ETM+) systems were introduced. A full timeline of the Landsat programme is illustrated in Figure 3.1.

Figure 3.1: The timeline for the Landsat family of satellites starting from the launch of MSS on Landsat 1 and leading up to the Landsat Data Continuity Mission (LDCM), scheduled for December 2012 (After NASA (2010)).

The programme spawned the development of similar space programmes for Earth observation by several countries, including Canada, China, India, Israel, Japan, South Korea and Taiwan, and the establishment of multinational organisations like the European Space Agency (ESA) (Schowengerdt, 2007). Among many others, some important multispectral sensors include: the National Oceanic and Atmospheric Administration’s (NOAA) Advanced Very High Resolution Radiometer (AVHRR); the French Satellite Pour l'Observation de la Terre (SPOT) systems; and NASA’s Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and Moderate Resolution Imaging Spectroradiometer (MODIS). Over the years, sensor development focused on the two key data resolution parameters: the number of spectral bands, and the Ground-projected Sample Interval (GSI), or the spatial resolution, giving birth to a new generation of hyperspectral sensors. Examples include NASA’s Advanced Visible/Infra-Red Imaging Spectrometer (AVIRIS) and HyVista Corporation’s HyMap airborne sensors.

Figure 3.2 shows where many of these sensors stand with respect to their acquisition resolutions. The data acquired by them have found important applications in monitoring the Earth and its atmosphere (Schowengerdt, 2007), including:

- environmental assessment and monitoring (urban growth and hazardous waste);
- global change detection and monitoring (atmospheric ozone depletion, deforestation and global warming);
- agriculture (crop condition, yield prediction and soil erosion);
- non-renewable resource exploration (minerals, oil and natural gas);
- renewable natural resources (wetlands, soils, forests and oceans);
- meteorology (atmospheric dynamics and weather prediction);
- mapping (topography, land use and civil engineering);
- military surveillance and reconnaissance (strategic policy and tactical assessment); and
- news media (illustrations and analysis).


3.1.1 Concepts in Data Acquisition
The Sun is a major source of electromagnetic radiation that interacts with the Earth’s surface. Earth Observation data are a measure of the proportion of radiation that gets reflected and scattered, absorbed and re-emitted (all bodies with temperatures above 0 K emit radiation) by objects and their related processes on the surface. Passive sensors measure pre-dominantly the reflected and scattered solar radiation in the visible, Very Near Infra-Red (VNIR), Very Short Wave Infra-Red (VSWIR), Short Wave Infra-Red (SWIR), and Medium Wave Infra-Red (MWIR) regions of the
spectrum (0.4-5 μm). This depends on various object properties such as their pigmentation, moisture content and cellular structure e.g. in vegetation, the mineral and moisture content of soils, and the suspended particle content of water to name a few. On the other hand, heat capacity and other thermal properties of the surface determine its emitted heat signature measured in the Thermal Infra-Red (TIR) (5-14 μm) (Richards and Jia, 2006). This range covers about 58% of the net exo-atmospheric solar spectral irradiance on the surface. The radiance values measured for different regions of the spectrum (in Wm⁻² μm⁻¹) can be integrated over a specific range of wavelength, called the at-sensor spectral radiance (L), and expressed as the photon flux power per unit area per unit solid angle per wavelength interval. From an engineering perspective, quantification of radiance depends on the sensor’s sensitivity towards (van der Meer et al., 2009): a) radiation reaching the aperture, b) radiation from a certain ground resolution cell, c) radiation from different wavelengths, and d) temporal variability of radiation.

Processing the measured at-sensor radiance values consists of three levels: sensor calibration, atmospheric correction, and solar and topographic correction. In the first level, the raw numbers are calibrated into at-sensor radiances using sensor parameters like gain and offset. Sensor calibration is typically done through a linear transformation that relates the at-sensor radiance with photon-counts, converted to raw Digital Numbers (DN), for each channel. In other words, spectrometers measure the radiance indirectly, such that the digitised signal recorded is directly proportional to the incoming photon energy from the surface, without any physically meaningful unit. The general linear relationship is represented by the form:

\[ L = c_0 + c_1 \times DN \]  

(3.1)

where, \( c_1 \) is the gain parameter, and \( c_0 \) is the offset parameter. The at-sensor radiance values are further processed through the second level atmospheric correction at different wavelength regions to form the surface radiance images. Second level processing is relatively more difficult, when compared to the first level, as it requires knowledge of the view-path atmospheric conditions at the time when the data acquisition is made. Some of the commonly used atmospheric correction models are: ACORN (Atmospheric CORrection Now) (Miller, 2002), ATCOR (ATmospheric
Optical Remote Sensing Methods for Carbon Dioxide Monitoring

CORrection) (Richter and Schlapfer, 2002), ATREM (ATmospheric REMoval) (Gao et al., 1993), and FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) (Adler-Golden et al., 1999; Matthew et al., 2000).

The final level incorporates correction for topographic slope and aspect, atmospheric path length variation due to topographic relief, like mountainous terrain, in order to achieve surface radiance to reflectance calibration, called the data cube, as shown in Figure 3.3.

![Figure 3.3: Pictorial description of a processed data cube from an imaging spectrometer (After CSIRO (2006)).](image)

Geometric correction is additionally carried out in order to reposition the pixels of the data cube with respect to a standard reference grid. The correction is summarised in three stages: distortion polynomial model, coordinate system transformation, and resampling (interpolation). These stages put together are sometimes also referred to as image warping (Wolberg, 1990).

Map projections form the essential spatial framework for remote sensing data. They allow the projection of a 3D spherical system into a 2D Cartesian system for the image formation. Some of the well known map projections are: polar stereographic, Mercator, oblique Mercator, transverse Mercator, and Lambert normal conic. For the
2D system, a polynomial model has to be devised whose coefficients are constrained by Ground Control Points (GCPs), which are reference points with known geographic locations. A large number of GCPs are sometimes required to minimise the error in warping. Once the polynomial model is established, the data is projected into the 2D map system resulting in a distorted grid. The distortion occurs from transformed non-integer position coordinates. This is further corrected by resampling techniques such as: nearest neighbour, bilinear, or cubic interpolation (Schowengerdt, 2007). After the radiometric and geometric correction steps, the data cube is ready for problem specific processing and interpretation.

3.2 Data Processing

The question on how information from the data cube can be processed was answered as early as in the 1960s. The two general schemes of data representation that were thought about are described below (Landgrebe, 1999).

Image Space: This is based on how the Human Visual System (HVS) interprets the images naturally by relating one pixel of the image to another in the geographic sense. However, this spatial relation does not completely represent the information carried by a data cube, i.e., band relationships are not apparent.

Spectral/Feature Space: This focuses on how a pixel is represented as a function of wavelength and is called the spectral signature. It helps in identifying different classes based on their signature, particularly for training algorithms. For computational purposes, each pixel is a vector in an N-dimensional space \( \mathbb{R}^N \), where \( N \) is number of wavelengths sampled in the spectral signature. This allows for a statistical description of background variability and providing a spatial context. Spatial relationships among pixels vary depending on illumination conditions, physiology of plants and soil background, among others. However, until the 1990s, most processing techniques have considered data cubes as unordered listings of spectral measurements with no particular spatial arrangement (Tadjudin and Landgrebe, 1998).

3.2.1 Problems and their Solutions in the Feature Space

An advantage of the feature space representation is its flexibility for multidimensional (multivariate) analysis. On adding more dimensions, the analyst is
able to better visualise and extract essential features from the data cube as it reduces the overlap between them and enhances the potential for good discrimination. However, consider an example, noted by Landgrebe (2002), of a hypothetical hyperspectral data cube where \( N = 100 \) and each pixel is a 10-bit DN, i.e. there are 1,024 distinct levels. This approximates to about \( 1,000^{100} = 10^{300} \) discrete locations in the feature space that the pixels can occupy. Kendall (1961) noted that such high dimensional feature space contains data clusters that are sparsely distributed.

Hughes (1968) experimentally proved that for high dimensional data, the number of training samples required to avoid degradation of classifier performance increases asymptotically, referred to as the Hughes phenomenon. As shown in Figure 3.4, it is therefore possible to achieve the same level of performance with reduced dimensions and lesser number of training samples.

![Figure 3.4: Hughes phenomenon of classifier performance, dependent on the number of training samples (m) (After Hughes (1968)).](image)

This calls for the need to extract information in the data cube by projecting it into a lower dimensional subspace without losing too much information. To reduce the dimensions of a data cube multivariate techniques are generally employed. One of the popular techniques used for this purpose is the Orthogonal Subspace Projection (OSP) (Harsanyi and Chang, 1994), based on the geometric concept of subspaces (Oja, 1983). Consider a \( M \)-dimensional subspace \( A \) in \( \mathbb{R}^N \), where \( M < N \). This
subspace is represented as a ‘$N \times M$’ matrix whose columns are vectors of size $N$, as $A = [a_1 \ a_2 \ a_3 \ ... \ a_M]$. The column vectors $a_1, a_2, a_3, \ldots a_M$, are independent of each other, called basis vectors, and they span the subspace of dimension $M$ (Strang, 2006). For orthogonality, these vectors obey an additional constraint, i.e. for any two column vectors $a_i$ and $a_j$, where $i \neq j$, the dot product (using vector transpose operator $^T$) is given by:

$$a_i^T a_j = 0 \quad (3.2)$$

The representation of a subspace is the foundation for defining subspace projection operators in OSP. Figure 3.5 illustrates the orthogonal projection vector $x'$ of $x$ in a 2D subspace $A$ ($M = 2$), and $x''$ is the orthogonal complement of $x'$.

![Figure 3.5: Orthogonal projection $x'$ of an arbitrary vector $x$ onto a 2D subspace $A$ in $N$-dimensional space, and its orthogonal complement $x''$.](image)

The relationship between the vector $x$ and its projection $x'$ in a subspace $A$ is given by:

$$(AA^T)^{-1} A x' = x \quad (3.3)$$

$$A'' = (AA^T)^{-1} A$$

$$x'' = x - x'$$

where, $A''$ is the Moore-Penrose pseudo-inverse of $A^T$ (Scharf, 1990).
To find the basis vectors $a_i$, $i = 1$ to $M$, the pixels are statistically characterised from the data cloud in $\mathbb{R}^N$ using Principal Component Analysis (PCA), also known as the discrete Karhunen-Loève Transform (KLT) or Hotelling Transform (HT) (Pearson, 1901). PCA works on the principle of linear combination of components of vector $x$ by a suitable weight vector $u$, i.e. the dot product $u^T x$. It finds a solution $u = u_1$ such that it maximises the variances of dot products evaluated on all vectors in the cloud. The variance function is given by:

$$f(u) = E\left\{ (u^T x - E(u^T x))^2 \right\}$$

$$u_1 = \text{arg max} (f(u))$$  \hspace{1cm} (3.4)

The first Principal Component (PC1) is the solution $u_1$ obtained by maximising the variance function. Other PCs are found by using the orthogonality constraint (Equation (3.2)). Figure 3.6 presents a simple illustration of PCA in 2D for two variables Var 1 and Var 2.

![PCA Illustration](image)

**Figure 3.6:** Application of PCA using a data cloud in 2D space to resolve eigenvectors for subspace determination.

Alternatively, Friedman and Tukey (1974) proposed the Projection Pursuit (PP) to determine independent non-Gaussian eigenvectors from the data cloud, while PCA gives Gaussian eigenvectors. For applications, if PCA is preferred over PP, an option is to transform non-Gaussian data through the use of power functions such as the Box-Cox transformation (Box and Cox, 1964).
The presence of noise in the feature space generally impairs the extraction of eigenvectors and hence an attempt to suppress noise is of large importance. The noise originates from sensor electronics, and their characteristics can be defined at the pixel level or by scan lines. Some of the common types of noise that affect image acquisition are: random noise, periodic noise, salt and pepper noise, striped noise, and bad lines, some of which are illustrated in Figure 3.7. In general, analysis of noise in a remote sensing image must be performed before applying geometric correction, which otherwise can result in smearing of noisy information with neighbouring pixels and lines. Both spatial and frequency domain (Fourier based) filters are normally used for noise filtering.

![Figure 3.7: Examples of noise affecting sensor acquisitions: (a) striping, (b) bad lines, and (c) salt and pepper noise (After van der Meer et al. (2009)).](image)

To filter noise in multivariate data cubes, Green et al. (1988) showed that PCA does not necessarily arrange a data cube in the order of image quality due to variable inter-band noise characteristics. They formulated a scheme where eigenvectors are arranged based on their Signal to Noise Ratio (SNR), and hence proposed Maximum Noise Fraction (MNF), later called Noise Adjusted Principal Components (NAPC) (Lee et al., 1990). Unlike PCA, MNF maximises the image quality instead of the data variance. Singular Value Decomposition (SVD) of the noise fraction matrix
generates a set of left eigenvectors (Johnson and Wichern, 2007), ranked in the order of increasing SNR value. Despite the lack of a priori knowledge of the noise type, it is assumed in the spatial domain that the signal value for a pixel is highly correlated with its neighbour while noise exhibits weak correlation. Green et al., (1988) use Minimum/Maximum Autocorrelation Factors (MAF) to estimate the noise fraction.

Figure 3.8: (a) Mixtures formed from a heterogeneous surface; (b) Mixtures formed from a homogeneous or intimately mixed surface (After Keshava and Mustard (2002)).

Another problem encountered in the feature space, related to points in the data cloud, is that of mixed pixel vectors (Keshava and Mustard, 2002). This is particularly true for low resolution acquisitions such as spaceborne data cubes. The radiometric value of a pixel is represented as a contribution of more than one material spectrum called endmembers. Linear models are commonly used to perform spectral unmixing to identify the endmembers and estimate their corresponding abundance fraction. However, such models are ideally suited for spectral mixtures arising from heterogeneous surfaces (Adams et al., 1995; Roberts et al., 1998). For homogeneous or intimately mixed cases, non-linear models have also been proposed (Hapke, 1981; Hapke, 1993; Mustard and Pieters, 1989). Figure 3.8 shows two ways of understanding spectral mixtures for a pixel. The choice between linear and non-linear models is based on the dominating physical characteristics of the surface, which is often difficult to judge (Keshava and Mustard, 2002). However, for most spectral unmixing problems a linear model is sufficient owing to its simplicity.

Endmember determination is a challenging problem in spectral unmixing because the resulting spectra must retain the physical characteristics of the surface constituents, while obeying the mathematical constraints of the process model at the same time.
Several techniques for endmember determination have been proposed. Keshava (2003) surveyed a taxonomical categorisation, as shown in Figure 3.9. The two general categories of techniques in the taxonomy are: the non-statistical, i.e. those that assume that endmembers are deterministic, and the statistical, which allow some degree of uncertainty (fuzziness) or are fully stochastic using probability density functions.

<table>
<thead>
<tr>
<th>Interpretation of data</th>
<th>Non-statistical</th>
<th>Statistical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description of randomness</td>
<td>Non-parametric</td>
<td>Parametric</td>
</tr>
<tr>
<td>Optimization criterion</td>
<td>Non-squared error</td>
<td>Squared error</td>
</tr>
<tr>
<td></td>
<td>Geometric</td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gaussian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-Gaussian</td>
</tr>
<tr>
<td>Outputs</td>
<td>Deterministic endmembers</td>
<td>Stochastic (Gaussian) endmembers</td>
</tr>
<tr>
<td>Inputs</td>
<td>Convex hull of pixel spectra</td>
<td>All pixel spectra and abundances</td>
</tr>
<tr>
<td></td>
<td>All pixel spectra</td>
<td>All pixel spectra</td>
</tr>
<tr>
<td>Noise</td>
<td>No noise in signal model</td>
<td>Additive observation noise</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gaussian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No noise in signal model</td>
</tr>
<tr>
<td>Algorithm</td>
<td>Minimum-volume transform</td>
<td>Penalty-based shrink-wrapping</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Block least squares</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total block least squares</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fuzzy K-means</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-linear least squares</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gaussian class estimation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Independent component analysis</td>
</tr>
</tbody>
</table>

Figure 3.9: Taxonomy for endmember determination algorithms organised based on several criteria such as data interpretation, randomness description and optimisation. Where: DPFT: Dark Point Fixed Transform, FPFT: Fixed Point Free Transform (After Keshava(2003)).

3.2.2 Processing the Feature Space for Classification

Classification is a classic problem in image processing. It is used to generate thematic maps that provide a visualisation of surface features, such as soil, vegetation and surface water, and helps in automatically isolating urban from rural land usage. This idea has long attracted the remote sensing community because it provides the foundation to solving many of the environmental and economic problems. It also serves to compress the large volume of information in the data cube into a single representative map for efficient data transmission (Schowengerdt, 2007). According to a recent survey by Lu and Weng (2007), a wide assortment of classification techniques can be grouped based on several different criteria.
Classification based on the kind of pixel information used:
Per-pixel classification techniques make use of *a priori* knowledge from training pixel signatures for the different features present in the scene. The signatures take into account of the materials that are involved in the feature make-up, but ignore the problem of mixed pixels, particularly for low resolution data cubes. Examples include Maximum Likelihood (ML), minimum distance, Artificial Neural Network (ANN), decision tree, and Support Vector Machines (SVM) classifications.

Sub-pixel classification techniques consider pixel signatures as linear or non-linear combinations of pure material spectra (endmembers) in the scene. Examples include fuzzy-set and Spectral Mixture Analysis (SMA) classifications.

Per-field (Object) classification techniques consider groups of pixels (or parcels), ignoring the potential variation of signatures within the same group. An example is the Geographic Information Systems (GIS) based segmentation.

Classification based on whether training samples are used:
Supervised classification techniques rely on *a priori* knowledge of the scene related to land cover information and surface processes to generate thematic maps. Examples include ML, minimum distance, ANN, and decision tree classifications.

Unsupervised classification techniques form clusters of pixels using inherent statistical properties of the data cube. The analyst then has the responsibility to assign an identity for each cluster. Examples include ISODATA and k-means clustering.

Classification based on whether spatial information is used:
Spectral classification techniques only take into account of the spectral nature of the feature space without considering spatial relationships between pixels. It results in noisy classification if there are spatial variations within the same class. Examples include Spectral Angle Mapper (SAM), ML, and ANN classifications.

Contextual classification techniques use information from neighbouring pixels as well as the pixel in question. Examples include Markov Random Function (MRF), frequency and geostatistical classifications.

Classification based on the degree of classification uncertainty:
Hard classification techniques assign each pixel to a particular class, hence they are not flexible for low resolution data cubes where mixed pixel problem occurs.
Examples include ML, minimum distance, ANN, decision tree, and SVM classifications.

Soft (fuzzy) classification techniques assign each pixel to more than one class, providing potentially more accurate results for low resolution data cubes when compared to hard classification. Examples include fuzzy-set and SMA classifications.

Classification based on the use of inherent data statistics:
Parametric classification techniques model classes using standard distributions such as the Gaussian distributions. However, when the landscape becomes complex, the classification can become noisy. It also does not allow incorporation of non-statistical ancillary variables such as the categorical information given by hard classification that could potentially improve the results. Examples include ML and Linear Discriminant Analysis (LDA).

Non-parametric classification techniques implicitly learns the class distributions from the data, and do not assume standard probability distributions to model the classes. Examples include evidence reasoning, ANN, decision tree, SVM, and expert systems based classifications.

For all criteria, the accuracy of a classification technique depends on both the data resolution and noise content. Low spatial resolution of the data cube generally degrades classification performance (Chen et al., 2004). The size of the surface objects and processes with respect to the spatial resolution of the sensor has a direct impact on the variance of the data cube (Woodcock and Strahler, 1987). According to Strahler et al. (1986), when smaller than the size of the pixel, also called the L-resolution case, objects and their processes are no longer detectable and hence the problem of mixed pixel arises. To overcome problems posed by the resolution factor, and improve accuracy of classification techniques, some of the approaches used are: incorporating ancillary variables, such as Digital Elevation Map (DEM), soil map, housing and population density, road network, temperature, and precipitation (Lu and Weng, 2007), and combining various classification techniques (Benediktsson and Kanellopoulos, 1999; Warrender and Augusteijn, 1999; Steele, 2000; Huang and Lees 2004).
3.2.3 Processing the Feature Space for Anomaly Detection

Anomaly detection is fundamentally different when compared to classification, in that it is essentially a binary Hypothesis Testing (HT) problem – between anomalous and non-anomalous pixels. Anomalous pixels are spectrally different from the background clutter (non-anomalous) and the area occupied by such pixels in the scene is a very small fraction (sparse), hence such pixels are generally characterised as having a low probability of occurrence. The sparseness has important implications for classification, i.e. owing to the insufficient number of pixels in the ‘anomaly class’, classification cannot be supervised. While for the unsupervised case, such as clustering, sparse data would be generally lost in clusters depending on the distance of the anomalous pixels from cluster centres. This calls for a different approach to the detection problem.

In many anomaly detection techniques, the often made assumption is that the spectral signature of the anomaly is unknown. The background clutter is modelled either locally for a neighbourhood, or globally for a very large portion of the image, from which the anomalies significantly deviate. The three common clutter models are (Stein et al., 2002):

**Local Normal model:** This assumes that a neighbourhood of pixels exhibits a Gaussian distribution.

**Global Normal mixture model:** A number of different types of materials can often be found in close proximity (edges and corners). In this case, the local model will fail and a global model is required that represents a mixture of Gaussian distributions representing a multi-component scene. Parameters for mixture models are estimated by using methods such as Stochastic Expectation Maximisation (SEM) (Moon, 1993), Maximum A Posteriori (MAP) classification (Beaven et al., 2000), and Cluster Based Anomaly Detection (CBAD) (Penn, 2002; Duran and Petrou, 2008) based on k-means and fuzzy c-means clustering (Bezdek et al., 1984) or Self Organising Maps (SOM) (Kohonen, 1982).

**Global linear mixture model:** This is a deterministic model that accounts for the problem of mixed pixels for low resolution data cubes. The background clutter is a linear mixture of endmember spectra with pre-determined abundance values.
Deviation from the model is assessed by using matched filter techniques based on subspace projection (Funk et al., 2001; Chang and Chiang, 2002).

The foundation for anomaly detection using any of the above clutter models is the HT. HT is a statistical testing procedure involving two competing hypotheses $H_0$ and $H_1$, also referred to as the null (non-anomalous) and alternative (anomalous) hypothesis respectively. For stochastic models, such as those based on the Gaussian distribution, the HT procedure is called the Generalised Likelihood Ratio Test (GLRT) (Neyman and Pearson, 1928). Figure 3.10 presents a simple representation of HT, where $H_0$ and $H_1$ are represented by their respective hypothetical likelihood functions. The ratio of these functions (GLRT) determines the active hypothesis. This ratio is close to unity at the region of overlap between the two likelihood functions, hence resulting in error detection probabilities (false detections), namely false positive and false negative rates. A GLRT that is currently an industry standard for anomaly detection by per-pixel analysis is the Reed-Xiaoli (RX) anomaly detection statistic (Reed and Yu, 1990) given by:

$$RX(x) = (x - \mu)^T C^{-1} (x - \mu)$$ \hspace{1cm} (3.5)

where, $\mu$ is the mean vector and $C$ is the covariance matrix. This statistic measures the deviation of a pixel from a multidimensional and unimodal Gaussian cluster.

![Figure 3.10: A simple representation of statistical testing using GLRT to determine the active hypothesis for a given pixel.](image)

However, it was also shown that although Gaussian distribution is a simplistic way of representing the data distribution, it fails to fit the tails of the distribution...
Optical Remote Sensing Methods for Carbon Dioxide Monitoring

(Manolakis and Marden, 2002) thereby resulting in an increase of false detections (Manolakis, 2005). Further investigations by Marden and Manolakis (2003), and Schaum (2007) confirmed that tails can be better represented using a general family of Elliptically Contoured (EC) distributions that subsume Gaussian distribution as a special case. It was also found that using non-Gaussian features with the RX statistic improves detection performance (Robila and Varshney 2002; Mei et al., 2008). Non-parametric counterparts for anomaly detection have been recently proposed such as: Support Vector Data Description (SVDD) (Banerjee et al., 2006), Topological Anomaly Detection (TAD) (Basener et al., 2007), and Maximum Entropy Nonparametric Estimation Detection (MENED) (He et al., 2008).

3.3 Field Applications for Carbon Dioxide Leakage Detection

Among many of the applications of remote sensing listed earlier, the problem of CO₂ leakage detection and quantification is particularly recent, and has caught the attention of the research community from the time monitoring activity for CO₂ storage was proposed. In addition to the information acquired by optical remote sensing technologies, such as Light Detection and Ranging (LIDAR), multispectral and hyperspectral data, Electromagnetic (EM) and Interferometric Synthetic Aperture Radar (InSAR) have also been tested.

Investigations using various remote sensing datasets have led to the development of mainly two kinds of approaches for the detection/quantification of CO₂ leakages, namely: direct methods which aim to quantify gas flux/concentration on or above the surface, and indirect methods used to study the effects of leakage on the surface environment without implicit quantification. Although most current methods are indirect, and are largely based on changes in the vegetative cover (Chadwick et al., 2009), two direct methods hold out promise: thermal imaging may potentially detect and quantify leakage if a measurable temperature anomaly is associated with the leakage, and high resolution airborne hyperspectral scanners may also detect CO₂ (and CH₄) directly using absorption features that lie within their wavelength range.

3.3.1 Direct Methods

Light Detection and Ranging

Using the signal reflection (or backscattering) principle, airborne LIDAR sensors have been tested for the direct detection of the CO₂ plume above vegetation surfaces
Optical Remote Sensing Methods for Carbon Dioxide Monitoring

(Benson and Myer, 2002), based on a technique known as the Differential Absorption Lidar (DIAL) (Browell et al., 1998). The suspended particles in the atmosphere, such as aerosols and dust, act as retroreflectors, relative to which the CO₂ plume absorption is measured. This principle is also used in field based instrumentation called the temperature tuneable feedback laser diode, which is capable of measuring the 2.000–2.005 µm region that contains CO₂ absorption lines (Repasky et al., 2006; Humphries et al., 2008).

Hyperspectral Imaging

The presence of CO₂ absorption features in the SWIR region makes it feasible to use hyperspectral sensors for its direct detection and quantification. However, retrieval of CO₂ is generally difficult because most of the absorption bands are relatively narrow, in comparison to the Full Width Half Maximum (FWHM) response of the sensor, and influenced by other atmospheric gases such as water vapour. Nevertheless, Gangopadhyay et al. (2008) attempted to identify the least susceptible wavelengths in the operating region of most present-day hyperspectral sensors with different atmospheric conditions and CO₂ concentrations. The FASCOD model (Clough et al., 1986), which calculates spectral transmittance, radiance or optical depth, was used to simulate atmospheric models accommodating atmosphere profiles, aerosol models, and water and ice cloud models with variable CO₂ concentrations. It was found that the 1.998-2.003 µm and 2.007-2.013 µm ranges are least influenced by water vapour. This information was used to compare two techniques for the application of detecting and measuring anomalous CO₂ flux due to coal fires using spaceborne spectroscopy, namely: the Continuum Interpolated Band Ratio (CIBR) (Green et al., 1989), which is an established method for atmospheric column average water vapour retrieval; and an inversion technique which attempts to obtain a relationship between CO₂ plume-related radiance and its concentration.

Spinetti et al. (2008) used the CIBR method with airborne hyperspectral images to retrieve CO₂ columnar abundance over volcanoes, and subsequently the CO₂ concentration was estimated using a mapping technique based on the assumption that the depth of the CO₂ absorption lines varies with its concentration. However, before estimating the concentration based on absorption characteristics, it may be necessary to first apply corrections accounting for the water vapour in the plume and the background CO₂ level, e.g. the Moderate Spectral Resolution Atmospheric
Transmittance (MODTRAN) algorithm (Berk et al., 1989) was applied by Pickles et al. (2009) on the data acquired by NASA’s multispectral system called the MODIS/ASTER Airborne simulator (MASTER) to remove background atmospheric effects before estimating plume concentration above the surface. In June 2007, a joint test was carried out by the Natural Environment Research Council’s (NERC) Airborne Research and Survey Facility (ARSF) and the British Geological Survey (BGS) to assess the use of hyperspectral data for direct detection of CO2 in the field near Nottingham, UK. A dry run was undertaken with large baths filled with CO2 and lined with tarpaulins of known reflectance. The AISA Hawk was used to obtain the images. The CIBR method was tested and was found to be highly sensitive to noise derived from illumination effects, poorly known sensor calibration and poor weather conditions, which rapidly dispersed the CO2 plume.

Gangopadhyay et al. (2009) found the inversion technique more suitable for estimating CO2 concentration up to a certain level in the atmosphere, and that the results could be further improved by the use of hyperspectral sensors.

AVIRIS data were also employed to identify and map anomalous volcanic CO2 concentrations in the atmospheric column above Mammoth Mountain, California (Martini and Silver, 2002). The 2.060 μm absorption band was selected, and subjected to continuum removal and minimum noise fraction analysis. While there was a degree of success in matching the detected CO2 flux zones with ground observations, complications arose due to confusion between CO2 absorptions and those absorptions due to plant biochemistry.

**Thermal Infra-Red Imaging**

While surface reflectance data acquired by hyperspectral sensors during the daytime is generally found to be useful for the direct detection of CO2 leakages, its applicability under low illumination conditions, and particularly in the night, is very poor. Under such conditions, TIR imaging sensors have the potential to supply useful information, provided they maintain a good sensitivity for emissivity values of leakages against that of the background land cover.

Prata et al. (2004) demonstrated that suitably calibrated cameras operating in the TIR region are capable of detecting the heat signatures of gaseous plumes from volcanoes and industrial sources. However, it is unclear whether these cameras could maintain
their calibration, and hence it was suggested that thermal imagers would be unsuitable to be used to monitor CO₂ leaks on their own. While these results were obtained from ground-based data, airborne versions were also subsequently tested (Prata et al., 2005). Gat et al. (1997) proposed a methodology for the detection and tracking of toxic organic gas plumes using the Airborne Thermal Infrared Imaging Spectrometer (TIRIS), based on a model called the radiative exchange phenomenology, to estimate the difference in temperature of the plume from that of the terrain.

Likewise, Tank et al. (2008) also assessed new remote sensing techniques based on TIR imaging for the detection and quantification of CO₂ degassing from the Earth's surface. The detection of degassing locations from TIR image time series was found to be reliable, particularly under dry, calm and cloudless weather between dusk and dawn. The method quantifies the gas emitted based on the determination of the heat energy required to generate the observed thermal anomaly. While their research employed ground-based remote sensing technology, the authors suggest that they are applicable to airborne and spaceborne measurements, and that they may be useful for monitoring the storage safety of CO₂ storage sites.

Hence, overall, TIR imagery offers the potential to quantify leakage, although a large number of drawbacks would need to be overcome. The most important drawback is that a hyperspectral TIR sensor which can provide information on the main absorption features of CO₂ gas, whose overtones are available in the SWIR spectral region, is currently not available. These overtones correspond to the heat energy absorbed by a gas plume during the day and re-emitted back to the atmosphere during the night. The amount of absorbed/emitted energy could be related to the concentration of the plume, thus providing the scope for an accurate quantification.

3.3.2 Indirect Methods

*Light Detection and Ranging*

LIDAR may also be of great benefit in monitoring vegetation stress related to CO₂ leakages. LIDAR data can be used as a support in the interpretation phase, since the backscatter intensity value helps to identify anomalies related to CO₂ leakages, both for vegetation and bare soil (Bateson et al., 2008).
Electromagnetic techniques

EM techniques involve the use of time-varying source fields to induce secondary electrical and magnetic fields in the subsurface that reveal its conductivity. Their use could potentially detect changes in the resistivity of shallow groundwater associated with the dissolution of CO₂. Although they have yet to be applied to CO₂ storage, airborne EM techniques have been successfully used to detect pollution plumes in shallow aquifers from coal tips, for example.

Airborne EM systems have been used in groundwater exploration studies for three decades, providing information on both the aquifer structure and the water quality (Siemon et al., 2009). Helicopter surveys are more commonly undertaken than those from fixed-wing aircraft, and use a towed rigid boom for frequency domain EM systems and a large transmitter loop with a small receiver for time-domain systems. The depth of penetration depends on the frequency (or time channels) of the signal, and frequency domain EM surveys are appropriate for shallow to medium (1-100 m) depth analyses. Homogeneous or layered half-space models are typically used to invert secondary magnetic field values into resistivities and depths.

Lipinski et al. (2008) used helicopter-borne EM surveys to evaluate the disposal of wastewater into a shallow (2-3 m depth) alluvial aquifer from natural coalbed methane production in the Powder River Basin, Wyoming. The wastewater was characterised as moderately saline and sodic. Beamish and Klinck (2006) used a fixed-wing airborne EM survey to characterise the hydrochemistry of a coal mine plume in Nottinghamshire, UK. The results of the survey revealed extensive zones of elevated subsurface conductivity due to coal mine spoil heap leachate infiltrating the underlying sandstone aquifer.

For airborne EM to be a potential monitoring technique for CO₂ leakage into a shallow aquifer, the ingress must have a measurable impact on the salinity of the formation water. Various studies have been conducted on the effect of CO₂ migration into aquifers, focusing (for example) on the potential release of heavy metals such as Pb and As due to the reduction of pH and subsequent dissolution of minerals (Wang and Jeffe, 2004; Zheng et al., 2009). The dissolution of minerals such as feldspars and carbonates to buffer pH will result in an increase in total dissolved solids (TDS), and hence conductivity. The effects of the CO₂ influx will depend on the nature of
the host rock, and the rate of groundwater flow may also be important (Lu et al., 2009). It is expected that changes in TDS will be more detectable for carbonate aquifers than for sandstones due to the reaction rates involved.

The Frio Brine Pilot experiment in Texas was established to demonstrate the potential for the storage of CO$_2$ in saline aquifers without adverse health, safety or environmental effects (Hovorka et al., 2006). The subsurface CO$_2$ plume was monitored within the injection zone, the overlying formation and the near-surface environment. Monitoring of the groundwater below the surface was found to be ineffective due to noise, sources of which included a variable water table and high natural CO$_2$ flux. It was noted that a long period of pre-injection baseline measurements would be needed to separate any potential leakage signal from the natural background variation. This would be equally valid for airborne EM monitoring.

The use of the airborne EM method may be limited in areas where significant clay contents, or naturally high background levels of TDS are present. Also, the method provides no geochemical discrimination and therefore ground-based sampling techniques must be employed to establish the cause of any enhanced conductivity (Beamish and Klinck, 2006).

**Satellite Interferometry**

Satellite-based Synthetic Aperture Radar interferometry (InSAR) is an established technique which can be used to detect and monitor millimetre-scale vertical land movement due to the subsurface injection of CO$_2$. The technology is based on synthetic aperture radar (SAR) data, whereby phase differences between images obtained at different times over the same region are mapped. The method has been widely applied in various fields, including earthquake deformation, volcanic activity, coastline erosion, landslides, glacier movement and ground subsidence due to extraction of hydrocarbons, minerals and water. Interferometry uses the phase, rather than the amplitude, of reflected microwave radiation. The phase of the reflected wave depends on the distance to the ground, since the total path length consists of a number of whole wavelengths plus a fraction of a wavelength. Ground motion changes the path length and causes a phase difference in the returning wave. Although the total distance to the satellite is unknown, the extra fraction of a
wavelength can be measured extremely accurately. In practice, the phase is also affected by several other factors, which obscure correlations between the pixels in any image. To extract useful information from the phase, some of these effects must be isolated and removed. Topographic interferometry uses two images of the same area taken from slightly different positions and uses the difference in phase between them to determine the topography, based on an image known as an interferogram. The two most commonly used InSAR techniques are: Differential Interferometry (DInSAR) (Berardino et al., 2002), and Permanent Scatterer Interferometry (PSInSAR™) (Ferretti et al., 2000).

In the DInSAR method, two images acquired within a short time interval are used to calculate an interferogram, which is assumed to have no deformation signal and therefore represents the topography. This interferogram is then compared with a third image with a longer time separation to give a residual phase change due to any deformation. The success of the method depends on a number of factors. Primarily, the images must be accurately co-registered to a sub-pixel level to ensure that the same targets are contributing to each pixel. In addition, any difference in viewing angles must not cause the phase to change over the width of a pixel by more than a wavelength. The effects of topography require baselines to be shorter if the relief is steep. Noise may also arise from changes to the ground targets in each pixel caused by vegetation growth, agriculture or snow cover. A further source of error is due to the propagation of the waves through the atmosphere. The velocity of the wave through the atmosphere is affected by the air temperature, pressure and the partial pressure of water vapour, which can change between acquisitions.

These issues can be resolved by using only the phase information of so-called Permanent Scatterers (PS). PSInSAR™ uses targets that retain their characteristics between acquisitions, for example buildings or rocks. By exploiting these scatterers and a sufficiently long time-series of SAR data (typically about 30 images), the atmospheric delay can be modelled and removed from the images and displacement can be measured. As an enhancement to PSInSAR™, the SqueeSAR™ algorithm was introduced in 2009. This exploits the use of Distributed Scatterers (DS), which consist of an extensive area where the back-scattered energy is less strong, but statistically homogeneous. DS may comprise areas of debris, deserts and non-cultivated regions, and the resulting displacement time series are found to be less
noisy. In the absence of PS or DS, engineered reflectors could be installed on ground overlying CO₂ storage reservoirs, such as in Corner Reflector Interferometric SAR (CRInSAR).

DInSAR has been used to detect surface deformation at the In Salah CO₂ storage site in Algeria (Onuma and Ohkawa, 2009; Onuma et al., 2010). A DEM was available with a spatial resolution of 90 x 90 m, which closely matched the resolution of the SAR data acquired from ENVISAT ASAR. SAR images were available with a temporal resolution of 35 days. The results revealed a surface upheaval rate of up to 7 mm yr⁻¹ around each of the three injection wells, and a subsidence rate of 4 mm yr⁻¹ for the gas producing wells. The surface deformation was mapped to show a NW-SE trend, in alignment with the direction of the anticline axis of the field results as shown in Figure 3.11. The time series of the upheaval and subsidence rates were also assessed. The method was concluded to offer a cost-effective monitoring tool in comparison with other conventional geophysical surveys.

Considering its advantages over DInSAR, PSInSAR™ has also been introduced into the field of reservoir monitoring. Tamburini et al. (2010) demonstrated its application for two specific case studies: for an Enhanced Oil Recovery (EOR) field in the Middle East, and the CO₂ storage site at In Salah. The results clearly demonstrate PSInSAR™'s ability to support both surveying and modelling the dynamic behaviour (volumetric change) of a reservoir and associated fault reactivation. Ferretti et al., (2010) also investigated the method for the new generation of X-band satellites, such as the TerraSAR-X which has a repeat cycle of 11 days. Combining data acquired by other sensors such as the COSMO-SkyMed constellation has also been recommended, with a repeat cycle of just 4 days. The use of the method for monitoring CO₂ storage over an urban area has been demonstrated by Kuhn et al. (2009), who observed uplift using this technique over the Ketzin gas storage project in Germany.

Vasco et al. (2008) used the results of a PSInSAR analysis of the In Salah images to estimate the changes in reservoir pressure and permeability. It was indicated that the pressure change resulting from the start of CO₂ injection propagated in a north-west direction along two paths with relatively high permeability possibly due to fractures and faults in the area. While Mathieson et al. (2009) emphasise the need for
corresponding tiltmeters and differential global positioning systems (DGPS) to calibrate the SAR image analyses, they suggest that the agreement between the ground deformation information and the evidence obtained from subsurface monitoring for the movement of CO$_2$ in the reservoir is good enough to warrant further investigation.

SAR data are currently being acquired over the In Salah field by RADARSAT-2, which has a higher spatio-temporal resolution than ENVISAT, taking images every 26 days with a spatial precision of 3 m. If the data can be calibrated, there is the potential to monitor subsurface movement of CO$_2$ (Mathieson et al., 2009). Work has been undertaken at Lawrence Berkeley National Laboratory to assess the use of forward and inverse modelling of the satellite deformation data. Findings reported by Rutqvist et al. (2009), who constructed a coupled reservoir-geomechanical model of the In Salah field using the SAR surface deformation data, have suggested that monitoring surface deformations can be useful for tracking the fluid pressure and for the detection of a permeable leakage paths through the overburden.

While the work undertaken so far at In Salah has proved the potential of SAR techniques in tracking CO$_2$ movement in the subsurface and improving models for
predicting potential leakage paths, it is feasible that the method could also be used to quantify CO₂ leakage if future surface uplift is observed over the injection well region. Xu et al. (2001) demonstrated that it was possible to quantitatively link hydrocarbon production and surface subsidence. Various earlier studies (e.g. Castle et al., 1969; Kumai et al., 1969; Shibasaki et al., 1969; Poland et al., 1975) have indicated a linear relationship between surface subsidence and reservoir fluid fluxes. Strehle (1996) created an empirical predictive tool called the roaming-tank model to correlate surface movement with net injection. This model was employed by Vasco and Ferretti (2005) to establish that it is possible to obtain a model of reservoir volume change that honours both production and injection data as well as surface level observations.

To summarise, with appropriate geomechanical models and calibrated InSAR measurements of surface deformation, it may be possible to use PSInSAR to detect and quantify CO₂ leakage from a subsurface reservoir based on associated surface deformation. However, it is not yet known how much gas would need to have escaped and how long it would take to detect this.

**Multispectral/Hyperspectral Imagery**

One of the foremost applications for leakage detection using optical remote sensing, for gases in general, is that of onshore hydrocarbons seepages. Link (1952) was the first to coin the terms ‘macroseepage’ for heavy hydrocarbon seepages such as tar and oil that are visible to the human eye, and ‘microseepage’ for light hydrocarbon seepages such as methane, ethane, propane, butane and pentane, that are detected by geochemical means. Some of the common surface expressions, shown in Figure 3.12, due to seepage that have been reported are (Schumacher, 1996; Saunders et al., 1999; Yang, 1999; van der Werff et al., 2006):

- local redox zones that result in the formation of calcite, pyrite, uraninite, elemental sulphur and magnetic iron oxides and sulphides;
- edge anomaly formation due to adsorbed or occluded hydrocarbons in soil and ferrous carbonate (delta C);
- ‘red bleaching’ due to reduction of Fe³⁺ (ferric) to Fe²⁺ (ferrous);
- mineralisation of clay; and
- geomorphological and botanical anomalies.
Almeida-Filho (2002) applied a simple band ratioing method using Landsat imagery to detect the bleaching effect of reduced iron in the soil. However, in his PhD research, van der Werff (2006) questioned the ability of simple ratios or the use of hard/soft classification such as the SAM and Minimum Distance to Class means (MDC) to detect smaller and subtle anomalies. Alternatively, a geometric approach was formulated for shape parameterisation at different scales by applying the Hough transform (Hough, 1962). This is based on the hypothesis of halo pattern formation on the surface due to microseepage. Theoretically speaking, some of the surface effects due to hydrocarbon seepage must also hold good for CO₂ surface seepages.

With regards to earlier studies on indirect detection of CO₂ leakages, forming the backbone and inspiration for the work described in this thesis, the potential of optical remote sensing technology have also been identified (Pickles and Cover, 2004; Bateson et al., 2008; Noomen et al., 2008; Keith et al., 2009; Lakkaraju et al., 2010; Male et al., 2010; Rouse et al., 2010). The fundamental advantage of this technology as opposed to the applications of LIDAR and InSAR is that the information is gathered by the sensors at multiple wavelengths simultaneously, commonly ranging from the Visible to LWIR regions of the electromagnetic spectrum. The advantage of
this multibanded data is its inherent ability to discriminate between various land covers and surface phenomenon through their spectral signatures. The most common effects on the surface due to high concentration of CO$_2$ are stressed vegetation and mineral alterations. Some of the investigations that studied these effects involved field-based spectrometric measurements in order to train their data classification procedure or perform statistical analysis for surface detection of leakage, while others used imagery acquired by airborne and spaceborne sensors.

Pickles and Cover (2004) reported some initial analyses using data from the DigitalGlobe’s Quickbird satellite acquired over Rangley enhanced oil recovery field located in Colorado, USA, to study vegetation stress and soil mineralogy, such as kaolinite and montmorillonite (clays), pyrite and haematite. It was found that the SAM technique was suitable for finely detailed mapping of local ecologies or admixtures of specific plant and soil types (meso-ecology mapping), enabling the detection of changes in surface patterns over time.

Bateson et al. (2008) jointly analysed datasets such as airborne multispectral, hyperspectral and LIDAR imagery together with digital orthophotos, for a study site in Latera, Italy, where CO$_2$ seeps naturally occurs at the surface, caused by subsurface gas production and migration related to volcanic processes. Normalised Difference Vegetation Index (NDVI) (Rouse et al., 1974) is used to identify areas of vegetation stress. The LIDAR imagery and digital orthophotos are visually interpreted to identify distinguishing features related to CO$_2$ seepage. This is followed by an integrated statistical analysis using a grouped polygon approach to visually rank regions based on the likelihood of CO$_2$ seepage. Taking into account soil gas flux and concentration measurements by Beaubien et al. (2008) for validation, a success rate of 39% was achieved. The method also extrapolates leakage detection for a large study area based on results obtained for smaller study areas. Figure 3.13 shows an orthophoto for the study area in Latera with the results of detection superimposed.

More recently, Keith et al. (2009) and Rouse et al. (2010) used field-based data collected over the Zero Emission Research Technology (ZERT) site (MSU, 2005) located in Bozeman, Montana, U.S.A., where controlled shallow subsurface CO$_2$ release is made, to study its effects on the local vegetation. One of the methods uses
the Random Forest classifier (Breiman, 2001) to establish a positive correlation between plant health index and distance from the release site (Keith et al., 2009). The other method (Rouse et al., 2010) establishes a regression model between the reflectances in the Red, Green and Near Infra-red (NIR) regions of the spectrum and the NDVI. The regression analysis shows that NDVI alone was best able to statistically separate stressed vegetation regions. Male et al. (2010) applied the SAM technique using hyperspectral plant signatures for the ZERT site. It was found that the minimum amount of soil CO$_2$ concentration that induces vegetation stress within 4 days lies between 4 and 8 % of CO$_2$ by volume. A good correlation was also found between the stressed vegetation classification map and a logarithmic soil CO$_2$ flux map of ZERT site (Lewicki et al., 2010). Further efforts by Lakkaraju et al. (2010) also demonstrated that there exists correlation between vegetation stress and increasing concentration of soil CO$_2$ by using sensitive indices such as: high carotenoid to chlorophyll ratio indicated by: the Structural Independent Pigment Index (SIPI); decrease in Chlorophyll Normalised Vegetation Index (CI NVI); decrease in Pigment Specific Simple Ratio (PSSR), and decrease in Normalised difference First Derivative Index (NFDI).

The University of Nottingham’s ongoing ASGARD (Artificial Soil Gassing And Response Detection) project (http://www.nottingham.ac.uk/geography/asgard) has
also established a soil gassing facility, similar to the ZERT project, which simulates elevated soil CO₂ concentrations that could be caused by a leakage. Remote sensing techniques have been tested for their ability to detect signs of plant stress, using the ratio of the 0.725 μm and 0.702 μm bands. Data were obtained from the IKONOS satellite, which has a high spatio-temporal resolution. The gassed plots were found to be more stressed than the controls, although seasonal differences also have to be taken into consideration.

It is generally seen that all of the above indirect methods developed using optical remote sensing data to some extent rely on prior knowledge of the field under investigation, i.e. knowing the exact location of the leakage, and hence can alternatively be referred to as supervised techniques. Thus the need for the development of a novel unsupervised detection technique that is able to pinpoint the locations of the most probable leakages by solely using the information contained in the data arises. The remaining chapters of this thesis are dedicated to the description of the development and application of such a methodology.

3.4 Conclusions

Optical remote sensing has played a significant role in Earth science applications for over four decades, since the inception of the ‘Space Race’ between the United States and former Soviet Union. One of the earliest concerns of scientists at the time was how large volumes of data generated by the sensors could be processed in order to extract useful information for various science problems. Hence modelling the data cube became an important issue for processing. With the proposition of multidimensional feature space, some of its problems, namely high dimensionality, and possible data corruption due to noise and mixed pixels were tackled using the geometric concept of subspace and multivariate statistics. Subsequent processing approaches from data classification to anomaly detection were also formulated.

Among the large application portfolio of optical remote sensing, what has caught attention of scientists in recent times is that of CO₂ leakage monitoring. Leaking gas from transport pipelines and reservoirs does not only mean economic losses, in terms of the invested energy due to the processes involved, but also hampering the balance of the ecosystem by putting flora and fauna at risk of exposure to high concentration of CO₂. Data processing using various techniques of classification and anomaly
detection can provide information to the data analyst about the surface seepages, providing early warning of potential leakage and averting any risks to the environment and helping deploy mitigation strategies.

Although direct detection techniques have been proposed using optical remote sensing data, it is seen that there are problems related to the ability of the current sensors to capture the main absorption features of CO₂ in the Thermal Infra-Red (TIR) region of the electromagnetic spectrum with high fidelity. This implies that the detection and quantification process for weak seepage signals may be dubious using current methods. On the other hand, state-of-the-art indirect detection techniques that have been proposed are largely dependent on *a priori* field knowledge, i.e. knowing the exact location of some of the leakages and extrapolating their features to detect other leakages for a large area. Some of these techniques also rely on simple measures such as the NDVI that are unable to distinguish between vegetation stress caused by biogenic and petrogenic factors (including CO₂). In other words, they may not effectively use the spatial and spectral information contained in the data. These factors lead to the increase in false detections.

Through the PhD research described in this thesis, a promising solution is proposed which provides potentially valuable information of surface leakages through unsupervised indirect detection using optical remote sensing data. Despite its inability to perform implicit quantification of leakages, it is believed that it provides both reliable and essential information related to the location of seepages, with a measure of confidence, which could help in reducing the burden on field practitioners who set out to take direct measurements with cumbersome instrumentation, thus indirectly leading to seepage quantification.
Chapter 4 Spatial and Spectral Data Analysis Methods

4.1 Introduction

The optical remote sensing data cube acquired over a leaking CO₂ storage or natural analogue site contains a large volume of information. As a result, locating low probability CO₂ leakage anomalies in the feature space is not a trivial problem. The aim of earlier indirect approaches was to look for anomalous surface expressions based on a priori information on the location of at least some known leakage sites and extrapolate the results for a large area in order to detect other potential leakage points. However, the manifestation of a spatially dynamic surface environment that contains both inter and intra spectral variations of land covers increases the chance of missing correct anomalies and labelling the wrong anomalies as CO₂ leakage. This shortcoming of earlier methods was the motivation for the development of a new methodology in this PhD research that involves breaking down the complex problem into manageable components that can be individually studied using spatial and statistical analysis techniques in order to reduce such errors.

The primary objective of the methodology proposed in this thesis is to detect the anomalies in an unsupervised manner without a priori knowledge of leakage targets within the study site. The two key factors investigated to achieve this objective are:
**Redundancy in the data cube:** Data redundancy hinders the detection of weak and low probability anomalies. This is dealt with here by modelling and filtering spatial and spectral redundancies in the data cube, using *geostatistical* and *multivariate analysis methods* to obtain a residual subspace that contains different kinds of surface anomalies including those related to CO₂ leakage.

**Surface expressions:** The task of identifying the likely subset of surface anomalies that are related to CO₂ leakage in the remaining volume (residual subspace) of available information is challenging. This is dealt with here by applying *information fusion* of the theoretical knowledge on surface expressions, obtained through *data classification*, with the filtered subspace to identify this subset.

### 4.2 Spatio-Spectral Data Processing with Geostatistical and Multivariate Analysis Methods

Optical remote sensing provides a synoptic view of the Earth’s surface in terms of both spatial and spectral information. This helps scientists to study and monitor the changes occurring on the surface due to the seasonal cycle, geomorphological changes and human activities (de Jong *et al.*, 2006). Traditionally, an image analyst performs visual interpretation of the scene for object identification, to determine its location with respect to other objects and study their properties, such as colour, pattern, shape, shadow, size, texture and tone (USDA, 1993). With increasing computing power over the past decades, these tasks have become gradually encapsulated by a group of automatic interpretation methods called *spatial analysis* techniques that quantify interesting patterns of the natural world in order to address scientific issues concerning the environment related to its geology, hydrology and ecology, including flora and fauna. To directly or indirectly identify these patterns, spatial analysis methods lay the foundations for image processing tasks such as classification and anomaly detection. Some of these methods are based on:

**Texture measures:** Texture is the spatial signature of an object characterised by the relationship of a pixel with other pixels in the object image (Haralick *et al.*, 1973). The most commonly used statistic to describe this relationship is the covariance between the pixels as a function of distance between them. The covariance function is normally estimated by using the spatial autocorrelation functions within a moving
window (Woodcock and Strahler, 1987), e.g. Moran's I and Geary's C, and gray-
tone spatial dependency matrix, called the co-occurrence matrix (Haralick, 1979).

**Aggregation:** This is a common approach for analysing data and studying pattern
properties from local to regional and global spatial scales (*mesoscale*), sometimes
referred to, by the ecologists, as the Modifiable Area Unit Problem (MAUP). It is an
accepted fact that aggregation of data can affect the consistency of pattern statistics
that may also change the final output. Despite this disadvantage, it has been shown
that certain patterns, that may be interesting, are only revealed at regional or global
spatial scales (Allen *et al.*, 1984). Simple aggregation methods include averaging,
central-pixel resampling and median calculation (Bian and Butler, 1999).

**Fractals and multifractals:** This concept, introduced by Mandelbrot (1983), is
becoming an increasingly useful tool for modelling and understanding natural
phenomena at different scales. It is based on the idea of self-similarity which is
invariant of spatial scale. However, geographical phenomena do not strictly behave
according to this rule (Goodchild and Mark, 1987). Nevertheless, the basic idea of
fractals can be extended to multifractals (or stochastic fractals) where self-similarity
and dimension have dissimilar properties at different scales and are often described
by statistics such as trail lengths, area-perimeter ratios, spatial auto-covariances and
rank-order or frequency distributions (Burrough, 1993). The fractal theory is a useful
alternative in studying image texture properties too (De Cola, 1989; de Jong and
Burrough, 1995).

**Mathematical morphology:** This theory emerges from the study of the geometry of
binary porous media, such as sandstones, comprising of a two phase system - *pores*
in a *matrix* (Soille and Pesaresi, 2002). This idea led Matheron (1975) and Serra
(1982) to consider a set formalism, such as union, intersection, complementation,
and translation, to analyse binary images giving rise to a whole suite of non-linear
grey scale image operators called *morphological operators* for specialised tasks,
such as filtering, segmentation and classification, and texture studies.

**Wavelet analysis:** Meyer (1990) was the first to discuss (in French) the use of this
powerful tool for signal processing, particularly to analyse non-stationary signals
with finite energy such as optical remote sensing data, where simple approaches
using covariance and correlation fails. This approach is closely linked to the original
idea that Mallat (1989) proposed, called Multi Resolution Analysis (MRA), where
images are decomposed into structures and analysed at different scales. A wavelet transform expresses any function as a summation of wavelets (basis functions). Wavelets are generated from a single mother function by dilations and time shifting operations for time series analysis, or spatial domain shifting for image analysis. It has diverse applications in remote sensing such as texture analysis, classification, multi-sensor image registration and fusion, and data compression.

Geostatistics: Spatial representation and quantification of information from the Earth’s surface is also done using the geostatistical theory of regionalised variables (Matheron, 1963). This is based on describing spatial dependence between values of a random variable. The key function that encapsulates this dependence is called the variogram. Although this theory was originally developed for interpolation (estimation) of unknown values at locations using kriging, given some data samples at random locations in a 2D Random Field (RF), it has also lent itself to other applications for remote sensing such as modelling image texture in order to improve classification (Chica-Olmo and Abarca-Hernandez, 2000). These methods were used in the thesis.

The most important reason for the choice of geostatistical tools in this thesis is based on its fundamental ability to determine filter parameters based on the spatial correlation structure in the data. In other words, geostatistics is more data-driven than any of the other types of methods listed above, for which the choice of parameters is merely based on the experience of the data analyst. Recently, geostatistical methods have gained popularity in processing remote sensing data. However, owing to its computational complexity using data cubes, it is normally combined with some form of multivariate analysis. Some of the applications include: image filtering (Wen and Sinding-Larsen, 1997; Oliver et al., 2000), classification (Atkinson and Lewis, 2000), and anomaly detection (Goovaerts et al., 2005). The following sections elaborate on the geostatistical theory of Intrinsic Random Functions (IRF) and a multivariate technique called the Independent Component Analysis (ICA), which have been selected for use in this research.

4.2.1 Geostatistical Methods for Structural Representation and Data Estimation

Geostatistical methods deal with the science of estimation and simulation of spatial variables in a wide spectrum of applications including geology, edaphology, forest
management, environmental studies, oil, gas and water resources assessment, and spatial economy (Chica-Olmo and Abarca-Hernandez, 2000). Some of the earliest studies carried out for remote sensing data also demonstrate its usefulness in information extraction (Curran, 1988). It is based on the concept of a regionalised variable that presents both spatial distribution and variability. The properties of regionalised variables are captured in a spatial auto-correlation function called the variogram. The variogram is based on the intrinsic hypothesis of random and structured spatial variables (Matheron, 1971) and for optical remote sensing data is estimated using the radiometric Digital Number (DN) values of pixels of the data cube. According to the intrinsic hypothesis, finite differences of random variables (or the discrete version of first derivatives) weakly obey the properties of stationarity, namely: the mean value is constant, the variance depends on the spacing (lag distance) between two points, and both the mean and variance parameters are finite and computable. For a random variable \( Z \), this is expressed by the classical definition of the variogram function, given by:

\[
\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(p_i) - Z(p_i + h)]^2
\]  

(4.1)

where, \( \gamma(h) \) is the variogram function, \( Z(p_i) \) and \( Z(p_i + h) \) are pairs of values (realisations) of the random variable at locations \( p_i \) and \( p_i + h \) separated by the lag (displacement) vector \( h \) and \( N(h) \) is the number of pairs available for a given lag. The variogram function serves as a structural analysis tool, and unlike the covariance function it does not require the estimation of the usually unknown mean value of \( Z \).

One of the central problems in geostatistics for incomplete sampled data is the need to estimate the unknown values of data at unsampled locations for a better understanding of the phenomenon under study with respect to its geographic location. Mathematically, this is the problem of interpolation using a set of weights. Conventional interpolation methods approach the problem in a deterministic manner and do not provide estimates of the uncertainty associated with the data estimates. In geostatistics, uncertainty is inherently taken into account by using the structural description of a random variable made by the variogram. This method of
interpolation is called *Kriging*, a term coined in 1963 by G. Matheron in honour of a South African mining engineer named D.G. Krige.

Prior to estimation, the variogram must be suitably fitted by a model function. This could be done by using traditional least squares fitting methods. However, this leads to the possibility that the weights could give negative variance for a linear combination of neighbourhood pixels which is not physically plausible, and hence must be avoided. Instead, standard functions are used to fit to the variogram, namely the spherical, cubic, exponential, Gaussian, and the generalised Cauchy functions (Chiles and Delfmer, 1999). The variance for a linear combination in terms of the variogram function is given by:

\[
E\left\{\left(\sum_{i=1}^{N(h)} \lambda_i Z(p_i)\right)^2\right\} = -\sum_{i=1}^{N(h)} \sum_{j=1}^{N(h)} \lambda_i \lambda_j \gamma(h)
\]

where, \( h = p_i - p_j \), and \( \lambda \) is the kriging weight. The idea that the variance function must be positive is also conditioned on a constraint such that the variogram function is said to be *conditionally positive definite*, hence making the estimation unbiased. This constraint is given by:

\[
\sum_{i=1}^{N(h)} \lambda_i = 1
\]

Kriging thus involves solving a system of constrained linear equations by the method of *Lagrange multipliers* to determine the weights. Convolving the weights with known data points leads to the estimation of data at unknown locations. For real applications, to avoid the bias in estimation due to the presence of inherent data non-stationarity, the Intrinsic Random Function (IRF) of order \( k \) (Matheron, 1973) is used for kriging estimation. The order indicates the degree of the polynomial up to which all Allowable Linear Combinations (ALC) of neighbourhood weights can annihilate non-stationarity in the data. While all the concepts discussed for univariate analysis can be extended for multivariate scenarios, *co-kriging* is simply based on the
Background Theory

structural relationship between dependent variables expressed by a set of variogram functions called the cross-variogram.

4.2.2 Intrinsic Random Functions (IRF)

In practical applications using the regionalised variable assumption, it is not always admissible to consider the Stationary Random Function (SRF) conditions stated earlier, for estimating a unique realisation. This brings in a more robust spatial modelling theory called the IRF for the representation and the estimation of non-stationary phenomena. When passing from SRF to IRF the following changes are considered (Chilès and Delfiner, 1999):

- The SRF fluctuates about a constant value and hence it is 'controlled' by an imaginary elastic force, i.e. the deviations are never too large, nor do they last too long. No such regulation exists in the case of the IRF. Such phenomena can also be described as systems with no memory, e.g. Brownian motion of particles. Hence, IRF is always studied through finite differences or increments of the function values.

- While the structural tool for a SRF case is the covariance function, the variogram function is used in the IRF case owing to a broader pool of mathematical functions that can be used to represent spatial relationships. These functions, unlike the covariance, need only be conditionally positive definite, and can be unbounded.

- The neighbourhood (kernel) variance for a SRF system is represented by an unconstrained linear combination for the centred covariance function. In an IRF system, only special combinations of variogram function have a finite variance, hence must be constrained by the Allowable Linear Combination (ALC) condition. In this manner, IRF helps in getting around the statistical inference problem caused by the bias owing to the spatial relationships in non-stationary random variables.

Early applications of IRF were mainly related to mapping, particularly in automatic contouring procedures for non-stationary phenomena (Matheron, 1973). Some notable examples are ground water modelling and hydrology (Delhomme, 1978; Kitanidis, 1983; Hoeksema and Kitanidis, 1984; Neuman and Jacobson, 1984); Mean Areal Precipitation (MAP) mapping (Chua and Bras, 1982); ocean floor mapping
Background Theory

(David et al., 1986); ecology sampling design for classification (Fortin et al., 1989); soil properties mapping using land attributes from Digital Elevation Maps (DEM) (Odeh et al., 1995); temperature mapping (Hudson and Wackernagel, 1994); and deformation analysis in medical images (Mardia et al., 2006).

Representation of IRF

The simplest and natural form of an IRF, $Z(p)$, which is a limit for most scenarios, is given by the dichotomy (Matheron, 1973):

$$Z(p) = Z_0(p) + P(p)$$

$$E[Z(p)] = 0$$

(4.4)

where, $Z_0(p)$ is the stationary fluctuation, and $P(p)$ is a 2D polynomial of degree $k$ at position vector $p = [x \ y]$, representing data drift. There exists a finite family of continuous monomial functions $f^n$, $n \in \{0, 1, 2, ..., k\}$, in the discrete domain of $Z$, representing $P$, such that:

$$P(p) = \sum_{n=0}^{k} c_n(p) f^n(p) = 0$$

$$f^n(p) = x^{a_1} y^{a_2}$$

$$a_1 + a_2 = n$$

(4.5)

The Brownian motion of particles, that tends to display the characteristic of an IRF of order $k = 0$ process, does not evoke the intuitive idea of stationarity. This situation is further amplified when the order increases, such that IRF theory becomes applicable to larger ranges of non-stationary phenomena. This can be observed from examples of zero mean (detrended) IRF processes with drift patterns, as shown in Figure 4.1.

Generalised Covariance Functions (GCFs) can represent the structural details of an IRF of order $k$, but do so only up to a random polynomial. Nevertheless, in practice, this is seen to be sufficient in encapsulating the spatial anisotropies present in the data (Matheron, 1973).
Background Theory

Figure 4.1: Contour map simulations showing IRF processes with zero mean: (a) k = 0; (b) k = 1; (c) k = 2 (After Chilès and Delfiner (1999)).

**Allowable Linear Combinations of Order k (ALC-k)**

Allowable Linear Combinations (ALC) offer an important constraint applied to an IRF field $Z(p)$ in order to perform statistical inferences under non-stationarity. A neighbourhood containing a set of $m$ points in $Z(p)$ defines an allowable discrete measure $\lambda$ at order $k$, if it locally annihilates a polynomial function $P$ of order less than or equal to $k$ in the neighbourhood, given by:

$$
\lambda = \sum_{i=1}^{m} \lambda_i \delta(p_i)
$$

where, $\delta(p_i)$ is the Dirac delta function defined at location $p_i$ in $Z(p)$. The ALC of order $k$ constraint, abbreviated as ALC-$k$, is given by (equivalent to (4.5)):

$$
P(\lambda) = 0
$$

(4.6)
\[ \sum_{i=1}^{m} \lambda_i f^n(p_i) = 0, \quad n = 0, 1, \ldots, k \quad (4.7) \]

A neighbourhood shown in Figure 4.2 is an example of a set of weights that obeys the ALC-1 constraint (Chilès and Delfiner, 1999). In general, ALC-k constraint applied to an IRF field has the following mathematical properties:

- The discrete measure \( \lambda \) is an element of a class of allowable measures \( \Lambda_k \), and it automatically follows that, up to an arbitrary order of \( K \), \( \Lambda_k \supset \Lambda_{k+1} \supset \Lambda_{k+2} \supset \Lambda_{k+3} \ldots \supset \Lambda_K \). In other words, a set of weights that eliminates a polynomial function of order \( K \) also works well for a system representing the Brownian motion of particles (IRF-0), but the converse is not true.

- The discrete measure \( \lambda \) is independent of drift coefficients of the underlying polynomial function. This in turn leads to the property of translation invariance across the field. Hence, the set of weights are also referred to as generalised increments.

\[
\begin{array}{c|c|c}
0 & 1 & 0 \\
1 & -4 & 1 \\
0 & 1 & 0 \\
\end{array}
\]

**Figure 4.2:** A \( 3 \times 3 \) square Laplacian neighbourhood is an ALC for monomials of order \( k = 1 \) (note that \( \sum_{i=1}^{m} \lambda_i = 0 \)).

**Generalised Covariance Functions (GCF)**

The correlation structure of a stationary random function \( Z(p) \) is generally represented by its covariance function. However, in case of non-stationarity, the correlation structure of the stationary generalised increments of \( Z(p) \) is represented by the GCF \( K(h) \), such that:

\[ K(h) = \sum_{i} \sum_{j} \lambda_i \mu_j K(p_i - p_j) \quad (4.8) \]
where, $\lambda, \mu \in \Lambda_k$, are allowable measures for $Z(\mathbf{p})$. According to the translation invariance property, it is sufficient to say that $\lambda = \mu$. The non-stationary covariance function $R(x, y, h)$ of $Z(\mathbf{p})$ is given by:

$$R(x, y, h) = K(h) + C(x, y)$$

where, $c_{a1}(x)$, $c_{a2}(y)$, and $c_{a1a2}(x, y)$ are the covariance functions in the respective direction and polynomial order. As noted earlier and also observed in Equation (4.8), the GCF is independent of the drift coefficients. Thus, statistical inference is possible by using allowable measures in $\Lambda_k$, such that the part of $R(x, y, h)$ representing drift is eliminated, i.e.:

$$C(x, y) \to 0$$

$$\Rightarrow R(x, y, h) \to K(h)$$

An example of the GCF is the GC-0 function, which is directly related to the variogram function of an IRF-0 process up to an arbitrary constant, given by:

$$K(h) = -\gamma(h) + \text{constant}$$

4.2.3 Multivariate Data Analysis with Independent Component Analysis (ICA)

In multivariate analysis problems related to feature extraction, it is sometimes assumed that the observed data is a linear combination of hidden and information bearing variables, also called latent variables, representing quantitatively and qualitatively independent features that ought to be determined explicitly. One of the earliest theories in practical multivariate problems is that of Blind Source Separation (BSS) (Jutten and Herault, 1991). ‘Blind’ means very little or no prior information is available about the source signals that cause the observed signals. An example of BSS is the cocktail party phenomenon (Bronkhorst, 2000), where it is given that a number of speakers in a room are speaking simultaneously. Although such a situation will result in a mixture of noisy signals, if particular attention is paid, it is possible for the human auditory system to selectively comprehend a speaker’s
conversation from 'background noise'. The literature reveals that while the problem was first addressed in 1983, BSS was given the name Independent Component Analysis (ICA) by Jutten and Herault in 1986 (Comon, 1994). Since then, ICA has become the common terminology for all BSS problems.

ICA is mathematically represented by a generative model where the instantaneous observed variables are linear combinations of some hidden variables, assuming no external noise, given by:

\[
x_i = w_{i1}s_1 + w_{i2}s_2 + \ldots + w_{in}s_n, \quad i = 1, \ldots, n
\]  

(4.12)

where, 'x' is the observed signal, 's' is the hidden signal, and 'w' are the mixing weights. In ICA, both the hidden signal and mixing weights are unknown and must be estimated from the observed signal. According to the generative model and central limit theorem, the hidden signals must generally be more non-Gaussian than the observed signal. The uniqueness of ICA from other multivariate methods, such as PCA, owes to this principle and hence allows it to work in a broad range of multivariate signal processing problems as most practical signals in nature are non-Gaussian. Some measures of non-Gaussianity for ICA, also called objective functions, include (Hyvärinen and Oja, 2000):

**Kurtosis:** This is a classical fourth order statistical measure that characterises the flatness of a data distribution given by:

\[
kurt(x) = \frac{E\{x^4\} - 3\left[E\{x^2\}\right]^2}{\left[E\{x^2\}\right]^2}
\]  

(4.13)

where, \(E\{x\}\) is the mean value. Random variables with negative kurtosis are called sub-Gaussian (platykurtic), while those with positive kurtosis are called super-Gaussian (leptokurtic). The problem of kurtosis, making it less robust for ICA applications, is that its value is very sensitive to the presence of data outliers (Huber, 1985).

**Negentropy:** Also called differential entropy (Cover and Thomas, 1991; Papoulis, 1991), it is an information-theoretic measure based on the entropy of a random variable (Shannon, 1948). Entropy is related to the amount of information conveyed by the variable, i.e., a random and unstructured variable conveys more information
and hence has a large value of entropy. Consider a probability space \( S = (\Omega, F, p) \), where \( \Omega \) is the power set of all possible events also called the sample space, \( F \subset \Omega \) is the event set, and \( p \) is the probability measure on \( F \). The entropy \( H \), and consequently negentropy \( J \), of the sample space is given by:

\[
H(x) = -\int_F p(x) \log(p(x)) \, dx = -E\{\log(p(x))\}
\]

(4.14)

where, \( x \in F \) is a random variable having a non-Gaussian distribution, \( x_{\text{gauss}} \) is a random variable having a Gaussian distribution, and both \( x \) and \( x_{\text{gauss}} \) have the same variance. Among the standard distributions, the Gaussian distribution is considered to have the highest entropy (Hyvärinen et al., 2001), hence serves as a reference for negentropy. The problem sometimes, however, is the difficulty to estimate the probability function of \( x \). Hence, approximations for negentropy are used to overcome this problem (Hyvärinen and Oja, 1998).

**Mutual information:** This is a natural measure, also based on information theory, indicating the statistical independence between random variables. It is based on the Kullback-Leibler (KL) divergence between two probability distributions (Kullback and Leibler, 1951). More specifically, mutual information \( I \) is the entropy measure of KL divergence between the joint probability distribution of data and the product of the marginal distributions of data given by:

\[
I(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} H(x_i) - H(x)
\]

(4.15)

where, vector \( x = [x_1 \, x_2 \, \ldots \, x_n] \). It has been shown that ICA finds directions related to maximum negentropy, hence minimises mutual information between variables.

**Maximum likelihood estimation:** This is based on maximising the entropy of the output of a neural network, also referred to as the Infomax principle (Nadal and Parga, 1994; Bell and Sejnowski, 1995). Pearlmutter and Para (1996), and Cardoso (1997) established that it is equivalent to the Maximum Likelihood Estimation (MLE).

A popular algorithm for ICA is called the FastICA (Hyvärinen, 1999). It performs constrained optimisation of a functional \( f \), given by:
Background Theory

\[
f(G(w)) = E\{G(w^T x)\}
\]

\[
E\{(w^T x)^3\} = \|w\| = 1
\]  

(4.16)

where, \(G\) is a non-quadratic objective function of the form, \(x\) is the input data vector, and \(w\) is the optimised weight vector that denotes the Independent Components (ICs). According to the Kuhn-Tucker conditions (Luenberger, 1969), the constrained optimal of functional \(f\) occurs when:

\[
E\left\{ \frac{d}{dw} f(G(w)) \right\} - \beta w = 0
\]

\[
\beta = E\left\{ w^T \frac{d}{dw} f(G(w)) \right\}
\]  

(4.17)

The FastICA algorithm is derived from these conditions based on approximative Newton iteration in order to optimise \(w\), called the basic fixed point algorithm, given by:

\[
w^* = \frac{d}{dw} f(G(w)) - \frac{1}{2} \left[ \frac{d^2}{dw^2} f(G(w)) \right] w
\]

\[
w^* = \frac{w^*}{\|w^*\|}
\]  

(4.18)

where, \(w^*\) is the normalised vector. Examples of general purpose objective functions for ICA are (Hyvärinen, 1999):

- Kurtosis (Sub-Gaussian): \(G(w) = \frac{1}{4} (w^T x)^4\)
- Kurtosis (Super-Gaussian): \(G(w) = -\exp\left( \frac{-(w^T x)^2}{2} \right)\)
- Skewness: \(G(w) = \log(\cosh(w^T x))\)

(4.19)

Some applications where ICA have had success are: image processing applications such as feature extraction for image compression and noise removal; medical applications such as brain imaging; telecommunications; and analysing hidden factors in financial data. For remote sensing applications in particular, ICA has been
applied for solving issues related to processing the data cube, discussed in chapter 3, such as: high data dimensionality (Lennon et al., 2001); mixed pixel problem (Nascimento and Dias, 2005); classification (Shah et al., 2004; Du et al., 2006); and anomaly detection (Robila and Varshney, 2002).

4.3 Data Classification using Unsupervised Fuzzy Analysis

Classification (or data clustering) is a vital task in information processing for optical remote sensing imagery. It is used to generate thematic maps that show the spatial distribution of identifiable earth surface features, with increasingly widespread applications in monitoring earth surface change. A range of information themes representing the surface features are derived, such as soil types, vegetation and urban/rural land use details, with distinct spectral signatures. However, a number of external factors can cause variability in these signatures, including topography, shadowing, atmospheric variability, sensor calibration changes and class mixing within the Instantaneous Field Of View (IFOV) of the sensor (Schowengerdt, 2007), affecting the accuracy of classification maps. This variability is generally modelled by uncertainty analysis, by means of either statistics or probability theory.

The conventional approach, called ‘hard’ classification, fails to provide good mapping information of the land covers as it is based on the assumption that pixel classes are pure, discrete and mutually exclusive (Foody, 1996), e.g. discriminant analysis and Maximum Likelihood Classification (MLC). However, in recent years, analysts have switched to ‘soft’ or fuzzy classification that takes into account the mixed pixels problem (sub-pixel analysis). It provides a way to deal with vagueness in complex land covers and their imprecise radiometric information. This uncertainty is comparable to the reasoning patterns of the human brain (Zadeh, 1965). In this case, classification is performed by relaxing the crisp membership condition imposed by the conventional approach and allowing a given pixel to have partial membership in all predefined land cover classes of the scene, which potentially gives a more accurate representation of the real world (Woodcock and Gopal, 2000). The three ways in which fuzzy classifications may be obtained are: (1) fuzzy classifiers (Wang et al., 1990; Foody 1992), (2) softening the output of hard classification, such as fuzzy likelihood estimation, and (3) neural networks (Carpenter et al., 1999).
most common example and widely used among these types is the fuzzy $c$-means clustering technique (Bezdek et al., 1984).

An important characteristic of the fuzzy methods is that they are mostly unsupervised in nature. Unsupervised clustering involves the investigation of a given set of samples, without any a priori knowledge or supervised training in order to identify their membership. It has been extensively used in identifying pattern groups in data for applications related to exploratory data analysis, decision making, and machine learning situations, including data mining, document retrieval, image segmentation and image classification (Jain et al., 1999). It has also been used in many disciplines such as biology, psychiatry, psychology, archaeology, geology, geography and marketing (Jain and Dubes, 1988). In remote sensing applications, it groups pixels of the data cube based on their natural relationship using spectral characteristics. Despite these achievements, some analysts believe that unsupervised clustering is an unpromising approach and question on whether or not it is possible to learn anything of value from unlabelled data samples. However, some of the important reasons for the consideration of unsupervised methods include (Duda et al., 2001):

- it can avoid the effort involved in collecting and labelling a large set of samples for learning;

- it is useful to train a classifier using a large pool of unlabelled samples to identify groups and subsequently use minimum supervision to label the groups (semi-supervised);

- it can give insight on the inherent structure of the data distribution and the discovery of various classes – clusters or groups of patterns whose members are more similar to each other than they are to other patterns – is useful for the choice of an appropriate classifier design; and

- it can track changes in the characteristics of class patterns over time.

Alternate terminologies for unsupervised clustering that are referred in the literature are: unsupervised learning (Jain and Dubes, 1988); numerical taxonomy (Sneath and Sokal, 1973); vector quantisation (Oehler and Gray, 1995); and learning by observation (Michalski and Stepp, 1983).
4.3.1 Fuzzy c-Means Clustering

The conventional k-means clustering approach involves the task of dividing a given data set \( Z \) into \( c \) subsets (clusters) that are: non-empty sets, have null intersections with other subsets, and form the original data set \( Z \) by the union with other subsets, represented in set theoretic notations as:

\[
\mathcal{Z} = \bigcup_{i=1}^{c} Z_i = Z
\]

where, \( Z_i \) is the i-th cluster subset containing data points from \( Z \), such that \( Z_i \subset Z \), and \( \emptyset \) is an empty set. In fuzzy c-means clustering, the general condition for the membership of data points in the subsets is relaxed so that the subsets are not mutually exclusive anymore, such that:

\[
Z_i \cap Z_j \neq \emptyset
\]

The fuzzy c-means algorithm is best described by recasting the conditions in equations (4.20) and (4.21) in matrix notations (Bezdek et al., 1984). If \( Z \) contains \( N \) data points, then the cluster partition can be represented by a matrix of size \( c \times N \), \( U = [u_{ij}] \), with the following conditions:

\[
u_{i}[z_j] = u_{ij} = \begin{cases} 
(0,1], & z_j \in Z_i \\
0, & \text{otherwise}
\end{cases} \\[12pt]
\sum_{j=1}^{N} u_{ij} > 0, \; \forall i \\
\sum_{i=1}^{c} u_{ij} = 1, \; \forall j
\]

In order to make the partitions, represented by \( U \), consistent with the data set \( Z \), the generalised least-squared errors functional \( J_m \) needs to be minimised, given by:
\[ J_m(U, v) = \sum_{i=1}^{c} \sum_{j=1}^{N} (u_{ij})^m \| z_j - v_i \|_A^2 \]  

(4.23)

where, \( m \) is the fuzziness exponent (1.5 \( \leq m \) \( \leq 3 \), for most types of data), \( v = [v_i] \) is the vector of cluster centres, and \( \| z_j - v_i \|_A \) is the \( A \)-norm distance on the data set \( Z \).

The \( A \)-norm distance is characterised by a positive definite matrix \( A \), such that:

\[ d_{ij} = \| z_j - v_i \|_A^2 = (z_j - v_i)^T A (z_j - v_i) \]  

(4.24)

Matrix \( A \) controls the shape of the optimal clusters in \( Z \). For the purpose of fuzzy c-means clustering, there is a choice from the three most commonly used \( A \)-norms:

- **Euclidean norm**: The \( A \) matrix is given by the identity matrix \( (I) \), hence \( J_m \) identifies hyperspherical clusters in \( Z \).

- **Diagonal norm**: The \( A \) matrix is given by a diagonal matrix \( (D) \), hence \( J_m \) identifies hyperellipsoidal clusters in \( Z \), with axes proportional to the eigenvalues of \( A \).

- **Mahalanobis norm**: The \( A \) is given by the inverse of covariance matrix \( (C) \), hence \( J_m \) identifies hyperellipsoidal clusters in \( Z \), with axes proportional to the eigenvalues of \( A \).

The optimisation of the \( A \)-norm distance functional \( J_m \) is done using the Picard's iteration (fixed point iteration) on \( U \) and \( v \), given by:

\[ v_i = \frac{\sum_{j=1}^{N} (u_{ij})^m z_j}{\sum_{j=1}^{N} (u_{ij})^m}, \ \forall i \]  

(4.25)

\[ u_{kj} = \left( \frac{\sum_{i=1}^{c} (d_{ij})^{2 (m-1)}}{\sum_{i=1}^{c} (d_{ij})^{m-1}} \right)^{-\frac{1}{m-1}}, \ \forall j, k \]
4.3.2 Self-Organising Maps (SOM)

SOM is a biologically inspired mechanism that closely relates to the manner in which the cells of regions in the brain interact with each other when triggered by a stimulus (signal) from different parts of the body. The spatial interaction between closely spaced brain cells results in signal patterns or classes through an unsupervised learning process. Direct evidence of such localisation can be seen through imaging techniques that show the spatial distribution of 'thought' patterns. Some of the techniques include Positron Emission Tomography (PET), Autoradiography, and Magnetoencephalography (MEG). The images obtained serve as inputs in order to generate conceptual maps of the brain called retinotopic, somatotopic and tonotopic maps (Kohonen, 1982). SOM was originally developed with the intention of providing a viable alternative to traditional neural networks that are primarily feedback (Hopfield, 1982) and feedforward (Rumelhart et al., 1986) networks. SOM is a third type of neural network based on competitive learning. It has been successfully applied in domains such as pattern recognition, robotics, process control, and processing semantic information (Kohonen, 1990).

Static Competitive Learning

Competitive learning is non-linear mapping of n input data vectors $z_i$ (i = 1, ..., n) from an N-dimensional domain $\mathbb{R}^N$ to a 2D map of m N-dimensional output weight vectors $w_j$ (j = 1, ..., m; m<n), such that they represent the topology of the input as faithfully as possible. In classical theory of Vector Quantisation (VQ), this set of output vectors is called the codebook. The change in the state of the codebook (adaptation) is dependent on input such that the codebook vector closest to a given $z_i$, called the Best Matching Unit (BMU) or the 'winner' $w_j$, is always updated. This is represented by (Kohonen, 1990):

$$\min_j \{d(z_i, w_j)\} = d(z_i, w_c)$$

$$\left[\nabla_{w_c} d(z_i, w_j)\right]^T \delta w_j < 0, \ \forall j$$

(4.26)

where, $d$ is a distance measure, $j = C$ is the BMU index, and $\nabla_w$ (del) is the vector gradient operator with respect to the BMU. In analytical terms, this is interpreted as the minimisation of the expectation of the $r^{th}$ power of the distance, or error functional $E$, given by:
\[ E = \int \| z_i - w_j \|^2 p(z) dz, \forall i, j \]
\[
\min_j \left\{ \| z_i - w_j \| \right\} = \| z_i - w_c \| 
\] (4.27)

where, \( r = 2 \) indicates the mean squared error criterion using Euclidean distance. In practice it is hard to attain a closed form solution, hence an approximation scheme is used. This is called the **steepest-descent gradient step optimisation** for functional \( E \), such that the BMU is moved closer to the input vector, given by:

\[
w_c(t+1) = w_c(t) + \alpha(t)(z_i(t) - w_c(t))
w_j(t+1) = w_j(t), \quad j \notin C
\] (4.28)

where, \( t \) is the time (iteration) index, and \( \alpha \) is the adaptation gain coefficient, whose value lies between 0 and 1. It has been pointed out by Gersho (1979) that the adaptation of the codebook vectors with respect to the input signal space described above is such that the point density function in the 2D output map is an approximation of \( p(z) \).

SOM orders the responses to input vectors spatially by considering a time varying neighbourhood \( N_c(t) \) centred about the BMU. The initial neighbourhood size is set large, indicating coarse map resolution during the learning process, and shrinks monotonically with increasing iteration index for higher resolution. A biological lateral interaction can typically be described by a symmetric function such as the bell curve neighbourhood, e.g. a 2D Gaussian function. The neighbourhood is included in the optimisation to modify Equation (4.28), such that:

\[
w_j(t+1) = w_j(t) + \alpha(t) h(t)(z_i(t) - w_j(t)), \quad \forall j \in N_c(t)
w_j(t+1) = w_j(t), \quad \forall j \notin N_c(t)
\]
\[
h(t) = \begin{cases} h_0(t) \exp \left( -\frac{\| r_j - r_c \|^2}{2\sigma(t)^2} \right), & \forall j \in N_c(t) \\ 0, & \forall j \notin N_c(t) \end{cases}
\] (4.29)

where, \( h_0(t) \) is time varying peak value of the Gaussian function, \( \sigma(t) \) guides the time varying changes in neighbourhood size, and \( r_j \) is the positional vector of location \( j \) in
the time varying neighbourhood $N_c(t)$. In order to put more emphasis on the cluster structure and visualisation in the mapping obtained by using SOM, several techniques have been proposed, namely: U-matrix (Ultsch, 1992); Self Organising Field (SOF) (Santini, 1996); adaptive coordinates and cluster connections (Merkl and Rauber, 1997); Generative Topographic Mapping (GTM) (Bishop et al., 1998); LabelSOM (Rauber, 1999); and data histograms (Vesanto, 1999).

**Dynamic Competitive Learning**

There are two problems associated with the SOM that have been identified in literature. The first is that SOM uses fixed network architecture in terms of the number and arrangement of units which must be defined prior to training. This is particularly difficult for a large volume of input data vectors. Thus it is worth considering models that can also learn their architecture during unsupervised training. Examples of such models include incremental grid growing (Blackmore and Miikkulainen, 1993), growing grid (Fritzke, 1995), hypercubical SOM (Bauer and Villmann, 1997), and growing SOM (Alahakoon et al., 2000). Secondly, hierarchical decomposition of the input data is not mirrored in a straight forward manner by the SOM. Some models that have been developed for this purpose merely create a tree format containing a fixed number of units at different levels that are not representative of the input data, such as Hierarchical Feature Map (HFM) (Miikkulainen, 1990), and tree-structured SOM (Koikkalainen and Oja, 1990; Koikkalainen, 1995).

Recently, a more dynamic competitive learning model called the Growing Hierarchical Self Organising Maps (GHSOM) has been proposed by Dittenbach et al. (2000). By overcoming the limitations of fixed size and non-hierarchical architecture in the earlier models, GHSOM fully adapts to the inherent structure of input data with a specified level of detail during unsupervised training, resulting in many layers with each layer containing independently growing SOMs. Starting with the top layer, each map is grown to a size such that it represents a certain granularity (resolution) of the data. Each unit in the layer is subsequently checked for finer granularity so that any unit that corresponds to diverse and inhomogeneous inputs is expanded to form a new SOM at the next level. This strategy is repeated for all units
and levels until the best representation is reached in both lateral and hierarchical
directions. Figure 4.3 illustrates a hypothetical GHSOM learning result.

![Diagram of GHSOM](image)

**Figure 4.3:** A typical trained GHSOM reflecting the structure of data in both lateral and
hierarchical directions (After Rauber *et al.* (2002)).

The general quality of the adaptation of a unit’s (codebook) vector is measured by
the average dispersion of this vector from all the input data vectors that the unit
represents at a given granularity (measure of homogeneity). The two strategies that
are generally used for this measure are: (1) Mean Quantisation Error (MQE), and (2)
absolute value of Quantisation Error (QE). MQE is similar to Equation (4.27), where
\( p(z) \) is a uniform distribution for a set of homogeneous input vectors \( z_i \) for \( r = 1 \),
representing the standard deviation. Although MQE may be intuitive from a
statistical point of view, QE more closely follows the principal characteristic of the
SOM by providing sufficient map space and detail, particularly for more densely
populated regions of the input, called the magnification factor (Rauber *et al*., 2002).

The GHSOM architecture is initialised at layer 0 as the average of all data vectors,
followed by layer 1 as a SOM containing 2 x 2 units of randomly initialised
codebook vectors prior to its training and growth process (see Figure 4.3). A newly
initialised map is trained using the static competitive learning. After a certain number
of iterations, the QEs for all units are checked to find the one with the highest
measure. This unit is referred to as the error unit, denoted by \( e \). Next, the most
dissimilar neighbour \( d \) from the four nearest neighbours of \( e \) is determined using the
Euclidean measure. A new row or column of units is inserted between \( e \) and \( d \),
initialised as the average of their respective neighbours. More formally, the insertion process can be represented by:

\[
e = \arg \max_j \left( \sum_{z_i \in S_j} \| \mathbf{w}_j - z_i \| \right), \quad S_j \neq \emptyset
\]

\[
d = \arg \max_i (\| \mathbf{w}_e - \mathbf{w}_i \|), \quad \mathbf{w}_i \in N_e
\]

where, \( S_j \) is the set of input vectors that is associated with the unit \( \mathbf{w}_j \), and \( N_e \) is a 4-point neighbourhood of \( \mathbf{w}_e \). The row/column insertion process is illustrated in Figure 4.4. Static learning is repeated until the next growing step is encountered.

![Figure 4.4: Growing process by inserting row (a) or column (b) of units between \( e \) and \( d \) (After Rauber et al. (2002)).](image)

As more units are added to the growing SOM their QEs decrease indicating that each of the unit represents a more concise portion of the input space. This continues until the SOM’s global QE falls below a fractional measure of the QE of its parent unit, called the global stopping criterion, given by:

\[
\text{QE}_j = \sum_{z_i \in S_j} \| \mathbf{w}_j - z_i \|, \quad \forall j
\]

\[
\text{QE}_g = \frac{1}{n_g} \sum_{\forall j} \text{QE}_j
\]

\[
\text{QE}_g < \tau_1 \text{QE}_{p(g)}
\]

where, \( \tau_1 \) is a fraction of the QE measured on unit \( p(g) \), the parent of \( g \), and \( S_j \) is a set of input data vectors represented by \( \mathbf{w}_j \). In order to reveal the hierarchical structure of the data, those units of the SOM that represent too diverse input vectors have to be
expanded in the next layer as new SOMs, of size $2 \times 2$ each, until the local stopping criterion is met, given by:

$$Q_E_j < \tau_2 Q_E_{p(j)}$$

(4.32)

where, $\tau_2$ is a fraction of the QE measured on unit $p(j)$, the parent of $j$. At the transition from one layer to the next, the number of input vectors used for training reduces.

It can be seen that the training and growth of the GHSOM is entirely data-driven. The only parameters that are explicitly controlled are $\tau_1$ and $\tau_2$, called the breadth and depth parameters, in order to decide the extent of the lateral and hierarchical growth respectively. The optimal values for these parameters obey the following constraint:

$$0 < \tau_2 << \tau_1 < 1$$

(4.33)

An important characteristic of the GHSOM adaptation is that the training process generally does not lead to a balanced hierarchy, as seen clearly in Figure 4.3, hence it is very closely related to the inherent structure of the input data distribution.

### 4.4 Probabilistic Information Fusion and Validation

Information fusion is the science of the acquisition, processing and synergistic combination of data from disparate sources in order to have a better understanding of a phenomenon for subsequent quality decision making (Hall and Llinas, 2002). A natural example of such a process is that of the human brain evaluating food quality and edibility by integrating the senses, such as the visual appearance, touch and smell. This concept has been used in a variety of fields with the common purpose of: detection of the presence of an object or condition, classification, object tracking, monitoring, and change detection. Its applications are classified under two broad categories (Varshney and Burrus, 1997).

**Military applications:** These include tracking the locations of military entities such as ships, aircrafts and weapons as enemy targets. Air-to-air, surface-to-air, and ocean surveillance are the types of systems used for onshore and offshore monitoring. The volume of data collected is generally very large, and the acquisition systems involve
numerous aircrafts, ships and submarines. Typical sensor technologies used for imaging are RAdio Detection And Ranging (RADAR), SOUnd NAvation and Ranging (SONAR), Infra-Red (IR), and Synthetic Aperture Radar (SAR).

**Civilian applications:** Based on the sensor technologies used in military applications and others, such as millimetre wave radar, TV images, acoustic signals, electromagnetic signals, X-rays, and Nuclear Magnetic Resonance (NMR), these include modern day applications such as air traffic control, robotics, manufacturing, medical image interpretation and diagnosis (Pinz *et al.*, 1995; Bloch, 1996; Solaiman *et al.*, 1999), biometrics (Sanderson and Paliwal, 2002; Ross and Jain, 2003; Zhang and Ji, 2005), and remote sensing.

In remote sensing, information fusion has proved to be a powerful tool for monitoring the earth’s surface and atmosphere. It aims at the integration of complementary multisource earth data in order to enhance the information, which is otherwise less apparent if interpretation were to be carried out with single-source data, thereby increasing the confidence in the final results. Some of the objectives for processing remote sensing data using fusion are: sharpening images (Chavez *et al.*, 1991); improving geometric correction (Strobl *et al.*, 1990); stereophotogrammetry (Bloom *et al.*, 1988); feature enhancement and extraction (Leckie, 1990); classification improvement (Solberg *et al.*, 1994); temporal analysis (Duguay *et al.*, 1999); and substitution of missing information, or defective data, in case of data occlusions due to clouds or shadows (Aschbacher and Lichtenegger, 1990).

### 4.4.1 Information Fusion in Remote Sensing

Fusion in remote sensing applications can be categorised into low, mid and high level techniques (Goshtasby and Nikolov, 2007). *Low level* techniques work at pixel scales, either in the spatial domain, or the transformed spatial domain (Toet, 1990; Nikolov *et al.*, 2001; Li *et al.*, 2002; Goshtasby, 2005). Those that work directly on the spatial domain can sometimes also have the flexibility to focus on specific regions of the image, leaving other regions untouched. The pixels are processed by numerical/statistical methods: for example, addition and multiplication of images for fusing low and high resolution data; difference and ratio of images for change detection studies (Mouat *et al.*, 1993); and Regression/Canonical Variable Substitution (RVS/CVS) for composition of bands (Campbell, 1993). On the other hand, transform domain techniques tend to work on the whole image at once, e.g.
changing between Red-Green-Blue (RGB) and Intensity-Hue-Saturation (IHS) was found useful for geological applications (Haydn et al., 1982; Singh, 1989; Yésou et al., 1993). Additional examples include: Principal Component Analysis (PCA) (Chavez et al., 1991; Yésou et al., 1993; Shettigara, 1992) and Independent Component Analysis (ICA) (Mitianoudis and Stathaki, 2007); multi-resolution and wavelet analysis (Burt and Kolczynski, 1993; Li et al., 1995; Ranchin et al., 1996; Petrović and Xydeas, 2004); and Particle Swarm Optimisation (PSO) (Han and Yao, 2010).

Mid and high level fusion techniques, also referred to as feature-level and symbol-level fusion respectively, segment images into regions and fuse them together based on their properties (Zhang and Blum, 2002; Piella, 2003; Lewis et al., 2007). A wide range of applications in remote sensing use these techniques for multi-sensor image analysis and interpretation. Some of the common examples are (Pohl and van Genderen, 1998): topographic mapping and map updating; land use; agriculture and forestry; flood monitoring; ice and snow monitoring; and geology.

The information fusion framework based on probability theory is useful because it takes into account uncertainty when deciding among many competing hypotheses presented by the sources for a decision making process (Pinz et al., 1996). The Dempster-Shafer (DS) theory (Shafer, 1976) is a generalised probabilistic model where probability values are based on sets of joint hypotheses rather than just mutually exclusive or singleton sets. If the hypothesis space contains $n$ singletons, then a frame of discernment is formed by a power set $\Theta$ containing $2^n$ elements, or possibilities. The null set $\phi$ is one of the elements of the power set that represents a scenario of conflicting evidence between two or more hypotheses in the system. Hence, discounting the null set, the actual sum of all probabilities in the frame of discernment can be less than or equal to unity, unlike classical probability. Like the popular Bayesian inference, the DS theory combines hypotheses (evidences) as and when available to update the system’s prior belief. However, the DS theory offers some advantages over the Bayesian view (Wu, 2003), such as: easy definition of the prior probability; the ability to handle multiple and conditionally dependent events; the ability to account for competing events; and definition of uncertainty range. Hence, the DS theory is sometimes referred to as a generalised Bayesian inference.
4.4.2 Dempster-Shafer (DS) Theory of Evidence Combination

The study of uncertainty and its application has begun to take importance in the scientific and engineering community. This owes in great part to the recent advancement in computational power in order to perform complex mathematical and statistical analyses accounting for the full scope of uncertainty, which traditional probability theory cannot handle. Uncertainty can be classified into two types (Helton, 1997).

**Aleatory uncertainty:** This type of uncertainty is directly attributed to the randomness of the system under study. It is alternatively referred to as stochastic uncertainty, Type A uncertainty, irreducible uncertainty, variability, or objective uncertainty.

**Epistemic uncertainty:** This type of uncertainty is indirectly attributed to a random system, i.e. it is a property of the analyst owing to his/her lack of knowledge about the system. It is alternatively referred to as subjective uncertainty, Type B uncertainty, reducible uncertainty, state of knowledge uncertainty, or ignorance.

Although the frequentist and Bayesian approaches in traditional probability theory are useful in characterising aleatory uncertainty, some analysts have criticised their application to deal with epistemic uncertainty. Any probability analysis requires *a priori* knowledge on the probability of all the events occurring within a system. However, when such knowledge is not available, the use of uniform distribution function is justified under the concept of Lapalce’s *Principle of Insufficient Reason* (Savage, 1972). This is interpreted as all simple events having equal probability. An additional axiom in classical probability is that of the principle of additivity, i.e. the probability of all events sum to unity. These principles are appropriate to model implicit system randomness, however, they are not sufficient when applied to an issue of the analyst’s explicit belief or knowledge about the system.

As a consequence, applied mathematicians seek alternative approaches to represent epistemic uncertainty, in case of a lack of available information to evaluate probability and nonspecific, ambiguous or conflicting information from different sources. Since epistemic uncertainty cannot be characterised by a precise probability value, it is reasonable to assign an interval/set of probability values for a particular event. This has three important implications:
- it is not necessary to have a precise probability measure if it is not feasible or realistic to do so;

- the principle of insufficient reason is not imposed, i.e. it is possible to consider the likelihood of multiple events together without having to consider the likelihoods of exclusive events; and

- the principle of additivity is not imposed, i.e. sum of all the probabilities need not sum to unity, but can be subadditive if there exists an incompatibility between multiple sources of information, or superadditive, if there is a cooperative effect among all the sources.

In the theory of imprecise probability, three major frameworks exist that have approached the problem of representing events by probability intervals, namely: imprecise probabilities (Walley and Fine, 1982); possibility theory (Zadeh, 1978); and Dempster-Shafer (DS) theory (Dempster, 1967; Shafer, 1976). Although application dependent, it is a current research question which framework works best in a general sense. Sentz and Ferson (2002) give four reasons as a motivation for choosing the DS theory among others:

- the relatively high degree of theoretical development in the DS theory when compared to other theories for characterising imprecise uncertainty;

- the direct connection between the DS theory and traditional probability, through set theory, is easier to comprehend;

- the versatility of the DS theory in combining different kinds of evidence from multiple sources; and

- a large number of examples exist of successful engineering applications using the DS theory.

**Evidence Combination Methods**

There are two important aspects while considering the combination of evidence from multiple sources: the type of evidence involved and handling conflicting evidences. For example, consider five different sources of information (sensors) – Sensor 1, Sensor 2, Sensor 3, Sensor 4 and Sensor 5 - giving information about five targets A, B, C, D and E. It is then possible to consider four types of scenarios in the theory of
evidence combination for information fusion (Sentz and Ferson, 2002) as illustrated in Figure 4.5.

Consonant evidences: This type is represented by evidences showing a nested structure of sets, where the target information provided by one set (subset) is encapsulated by a parent set containing information on other targets, and so on, as seen in Figure 4.5 (a).

Consistent evidences: This type is represented by evidences showing dependent structure of sets, where at least one target information is common to all sets, as seen in Figure 4.5 (b).

Arbitrary evidences: This type is represented by evidences showing a dependent structure of sets, where no target information is common to all the sets, as seen in Figure 4.5 (c).

Disjoint evidences: This type is represented by evidences showing mutually exclusive (independent) structure of sets, as seen in Figure 4.5 (d).

It is clear that the level of conflict between the evidential sets based on the target information they provide is maximum for disjoint evidences.

Figure 4.5: Example illustration for the types of evidence combinations occurring in information fusion problems (a) consonant evidences; (b) consistent evidences; (c) arbitrary evidences; (d) disjoint evidences (After Sentz and Ferson (2002)).
The Rule of Evidence Combination

The imprecise probability function on the frame of discernment Θ, called the Probability Mass Function (PMF), m, can be approximated by the classical definition (Chokr and Kreinovich, 1994), bounded by a range between values of two functions [belief, plausibility] called the Basic Probability Assignment (BPA), given by:

\[
\text{Bel}(A) = \sum_{B \subseteq A} m(B) \\
\text{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B)
\]  

(4.34)

where, Bel(A) is the belief function, Pl(A) is the plausibility function, and A and B are evidential sets. Hence, the relation between belief and plausibility functions for set A follows from their definitions, given by:

\[
\text{Pl}(A) = 1 - \text{Bel}^{c}(A)
\]  

(4.35)

The original rule of evidence combination of multiple PMF functions is a generalisation of the Bayes’ rule, called the DS combination (Dempster, 1967). It strongly emphasises the agreement between multiple sources and eliminates conflicts through a normalisation factor. This can also be considered as a strict AND-operation in fuzzy logic terms. For combining a pair of PMF functions of independent evidential sets X and Y in order to generate a new composite hypothesis set A, DS combination rule states that:

\[
m_X \oplus m_Y(A) = \frac{\sum_{X \cap Y = A} m(X)m(Y)}{1 - K} \\
K = \sum_{X \cap Y = \emptyset} m(X)m(Y)
\]  

(4.36)

where, 1 - K is the normalising factor, which has the effect of removing conflicts and attributing any PMF associated with it to the null set. When K = 0, the belief and plausibility values are equal, as in the case of classical probability. The DS combination operator \( \oplus \) essentially obeys the four algebraic properties (Yager, 1986), namely:
\[ \text{Commutativity : } X \oplus Y = Y \oplus X \]
\[ \text{Idempotency : } X \oplus X = X \]
\[ \text{Continuity : } X \oplus Y = X' \oplus Y, \text{ where } X' \approx X \]
\[ \text{Associativity : } (X \oplus Y) \oplus Z = X \oplus (Y \oplus Z) \]  

The DS combination rule has been used extensively for multisource and multiscale classification techniques where data from multiple sensors provide multivariate information on the surface processes (Le Hégarat-Mascle et al., 2002; Le Hégarat-Mascle et al., 2003).

4.4.3 Validation using the Receiver Operating Characteristics (ROC) Curve

In signal detection theory, the ROC curve is a visualisation of performance that depicts the trade-off between true and false labels in classification and detection algorithms (Egan 1975). One of the earliest users of the ROC curve in machine learning was Spackman (1989), who demonstrated its value for comparing different algorithms. Since then, ROC curves have been extensively used in the machine learning community as they give a better judgement when compared to other, simpler performance indicators (Provost and Fawcett, 1997; Provost et al., 1998). ROC curves have also been applied to research in other fields, particularly in medical diagnosis, forensic science, education, manufacturing industry, and military applications (Swets et al., 2000).

Consider an instance of data \( I \) that needs to be mapped onto one of the elements of a decision set \( \{P, N\} \), indicating positive and negative class labels respectively (true class). Detection is a binary classification problem that essentially involves mapping the data instances into such binary sets. In order to distinguish between the actual and predicted class, consider \( \{P', N'\} \) as the predicted classes. The algorithm can choose one of four possible outcomes for \( I \), as shown by the confusion matrix in Figure 4.6. If the instance is positive and it is classified as positive, it is regarded as true positive. On the other hand, if the instance is classified as negative it is regarded as false negative. Likewise, if the instance is negative and it is classified as negative, it is regarded as true negative. On the other hand, if the instance is classified as positive it is regarded as false positive. This confusion matrix forms the basis for defining performance metrics for classification and detection algorithms.
The two common performance metrics derived from the confusion matrix are: True Positive Rate (TPR) that indicates the fraction of correct decisions (TP) on the total number of data instances in class P, and False Positive Rate (FPR) that indicates the fraction of wrong decisions (FP) on the total number of data instances in class N. Other possible performance metrics include: precision, recall, accuracy, and F-measure (Fawcett, 2006).

ROC plots several (TPR, FPR) points to evaluate the overall performance of an algorithm. Figure 4.7 illustrates an ROC space with 5 labels, A through E, denoting different classification/detection algorithms.
When the plot contains points lying on the diagonal line connecting (0, 0) and (1, 1) (point C), it indicates that the algorithm performs random guesses during the mapping process. Points lying above the diagonal line indicate the operating region TPR > FPR in the ROC space (points A, B and D), and hence good classification/detection, and vice versa (point E). Area Under the Receiver Operating Characteristics (AUROC) is a scalar representation, whose value lies between 0 and 1, obtained by integrating the ROC curve. It is also equivalent to the probability that the algorithm will correctly map a data instance into the {P, N} set. An AUROC value > 0.5 indicates acceptable performance (Fawcett, 2006).

4.5 Conclusions

In contrary to the general belief that having a large volume of data from multispectral and hyperspectral sensors can help answer many environmental related issues, detection of low probability anomalies on the earth's surface, such as leakages of CO₂ from a subsurface storage reservoir or a natural analogue site, using the data cube is not a trivial problem. This is owing to the redundancy in data that needs to be carefully analysed and filtered out before features related to leakages could be detected. On the other hand, data redundancy could be avoided by designing new sensors in the future that provide information only in those interesting regions of the reflectance and emissivity spectrum that contain the main absorption lines or other relevant diagnostic features for CO₂ leakage detection and quantification. As for the data that are available from sensors today, a methodology is required that performs systematic decomposition and identification of the available information. It is also necessary that minimum human intervention takes places during this process to reduce subjective errors in the final detection result. This is because the reliability of manual interpretation of the data cube is dependent upon the performance of the analyst’s eye, which generally varies owing to tiredness during the course of sifting through large volumes of information. Moreover, analysis of large volumes of data can become very slow and expensive.

The research conducted and described in this thesis proposes a methodology that makes use of the combination of numerical techniques for data processing, hence requiring less effort on the part of the analyst and avoiding subjectivity. It considers
two important factors: information redundancy in the data cube and fusion of independent information.

The detection problem is viewed as a Bayesian inference problem. In the developed methodology, the geostatistical theory of Intrinsic Random Functions of order k (IRF-k) is used to design the multivariate spatial filter, together with Independent Component Analysis (ICA), for prior detection. The Dempster-Shafer (DS) information fusion framework is then used to update the prior detection to posterior detection by incorporating theoretical knowledge of the spectral signatures of surface expressions caused by CO₂ leakage, through the Growing Hierarchical Self-Organising Maps (GHSOM) data clustering. Finally, the Receiver Operating Characteristic (ROC) curve is used to perform the validation of the results using ground truth data.

The use of IRF-k theory was motivated by its ability to model the non-stationarity in multispectral and hyperspectral datasets. This is considered an essential step because it helps in eliminating spatial redundancy by filtering out complex polynomials representing the long range correlations in the images and retaining residuals containing low probability short range correlations that include anomalies related to CO₂ leakages on the surface. These correlations are further enhanced by transforming the multivariate residuals using ICA which looks for directions that correspond to interesting non-Gaussian features for the representation of prior detection. Data clustering is performed using GHSOM because it naturally represents the inherent density distribution in a non-parametric manner, and learning the spectral signatures for all possible classes that constitute the surface. The signature containing diagnostic features could then be selected to represent the likelihood information for CO₂ leakages on the surface. This information is used to update the prior detection by the application of the DS theory based on imprecise probabilities. Using this theory it is possible to assign a confidence measure as the improved posterior detection thus giving a better estimation of the potential locations of surface CO₂ leakages. Finally, ROC curves allow automatic validation of the posterior detection based on a general detection performance metric called the Area Under the Receiver Operating Characteristics Curve (AUROC) which depends on the True Positive Rate (TPR) and False Positive Rate (FPR) of the detection results.
Chapter 5  Field Sites and Data Pre-Processing

5.1 Introduction

There is currently very limited documentation regarding CO₂ storage project sites that have any signs of leakage on the surface. Results from extensive leakage monitoring at injection sites such as the Sleipner gas field in the North Sea, Norway, and Weyburn oil field in Saskatchewan, Canada, have been positive (Chadwick et al., 2006; Wilson and Monea, 2004). Although this is according to the desired objective of CO₂ storage, data is required that provides the opportunity to develop a new unsupervised indirect detection methodology, which can then be rapidly deployed at industrial sites to monitor and detect if leakage occurs. Such data is generally acquired from analogue sites with natural or controlled surface occurrences of CO₂.

Two natural analogue sites, having continuous gas seepage from the subsurface, were selected for this PhD work. Both sites are located within volcano caldera regions, namely the Latera geothermal field in Italy, and the Laacher See in the East Eifel volcanic district of the Rhenish Massif in Germany. Extensive field work has been conducted previously by other researchers at these sites to determine various subsurface controls of seepage points, such as faults, migration mechanisms of the gas through the faults, locations of seepage points, the corresponding gas flux and concentration measurements, and the effect of seepage on the surface environment,
particularly on the local vegetation and soil geochemistry. In addition, both ground based and airborne remote sensing surveys have also been carried out providing the essential data for the research described in this thesis.

The airborne optical remote sensing datasets include both multispectral and hyperspectral imagery that are pre-processed for radiometric and geometric corrections. Additional processing is, however, required for lowering the subsequent computational effort and increasing the reliability of the unsupervised detection results. Therefore, it is essentially the starting point of the proposed methodology.

5.2 The Latera Site in Italy

This study site is a geothermal field located within the Latera caldera in the west-central part of the Italian peninsula. The caldera is a large, elliptical (10 km x 8 km), NNE-SSW trending structure with several eruptive vents located both within and outside of the caldera. Despite being inactive now for over 0.16 Ma (Vezzoli et al., 1987), the volcanic structure still has an extremely high geothermal gradient. It is believed that the associated thermo-metamorphic reactions in carbonate rocks are the source of large quantities of deep CO₂ (over 90 % by volume) and other trace gases, such as N₂, H₂, CH₄ and H₂S (5 - 10 % by volume), that are released to the atmosphere from various discrete gas vents throughout the valley (Bertrami et al., 1984; Duchi et al., 1992).

5.2.1 Geology of the Latera Caldera

In the area of the Latera caldera, and throughout central Italy, the regional substratum consists of metamorphic phyllites and micaschists onto which a series of ‘tectonic units’ were placed during the initial compressive phase, in the Eocene to Late Miocene period, via thrusts that trend primarily N-S (Annunziatellis et al., 2008). These units mainly comprise of the Mesozoic ‘Tuscan nappe’ and the Cretaceous to Eocene period’s ‘Ligurian flysch’. The Tuscan nappe consists primarily of carbonates and some siliciclastic successions. Due to high secondary permeability in the carbonates, this unit forms a regional aquifer and locally hosts the geothermal reservoir beneath the Latera caldera. The overlying Ligurian flysch is made up of various units, including impermeable shales and siltstones which act as a regional aquitard and which locally form the caprock for the geothermal reservoir. Figure 5.1 illustrates the geological sections of the caldera showing these layers.
At the cessation of Latera volcanic activity, edifice collapse created the caldera via a series of associated normal faults which, along with thrusts, regional normal faults, and extensive fracturing created by the volcanic eruptions themselves, became conduits for upwardly migrating hydrothermal fluids and CO₂. Many of these conduits became sealed via secondary mineral precipitation (Cavarretta et al., 1985). This process has made much of the overlying flysch and volcanic units impervious, except where gas is locally escaping, and helped to partially isolate the underlying geothermal reservoir (see Figure 5.1).

The CO₂ that leaks at the surface is likely the product of metamorphic alteration of the reservoir limestones, driven by the water-dominated system and the high heat flow linked to the original magma intrusion and deeper mantle (Duchi et al., 1992; Chiodini et al., 1995; Minissale, 2004; Chiodini et al., 2007).
5.2.2 Surface Lithology of the Latera Caldera Floor

Surface lithology throughout the caldera floor is dominated by alluvial sediments, while the overlying soils are highly fertile due to formation from the alkali-potassic units of the Latera volcanic complex (Beaubien et al., 2008). The central part of the Latera caldera is characterised by a NE-SW to N-S trending hydrothermally altered area, with silica, alunite, kaolinite and sulphur mines formed in the vicinity of several gas vents and thermal/cold springs (Gianelli and Scandiffio, 1989; Lombardi and Mattias, 1987; Lombardi et al., 1993).

The typical Mediterranean climate, rich volcanic soil, and relatively low population density all contribute to this being an active agricultural area, with grain and corn cultivation and sheep pasturing being the main activities. The limited infrastructure, forested areas, relatively flat topography and extended areas of homogenous crop types are all advantageous for the application of remote sensing techniques (Bateson et al., 2008).

5.2.3 Data Available for the Study Site

The study site itself is located centrally within the caldera. According to earlier studies at this site, crops and natural vegetation vary in growth height, leaf cover and relative stress during different seasons of the year. In areas where the gas vents are located, the vegetation appears continually stressed (Bateson et al., 2008). Three areas of interest were selected for the data analysis in this site. Figure 5.2 shows the location of the study site in the Latera caldera indicating the first area of interest, ‘Study Area 1’.

The optical remote sensing data available for this study area were acquired by the Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), Trieste, Italy, using an AISA Eagle 1K hyperspectral pushbroom scanning system that uses a linear system of Charge-Coupled Device (CCD) detectors mounted onboard an AS350 B2 helicopter. Although the sensor has the capacity to acquire up to 224 bands ranging from visible through NIR wavelengths (0.40 - 0.99 μm), only 63 bands were recorded for the survey (Bateson et al., 2008) with an average Full Width at Half Maximum (FWHM) of about 10 nm. Table 5.1 presents the list of recorded centre wavelengths. The spatial resolution of the data generally depends on the altitude at which the sensor is flown.
For this study area, two separate daytime images were acquired by the same sensor in October 2005 and July 2007 at 2m and 1m spatial resolutions respectively.

In 2006, the University of Rome ‘La Sapienza’ carried out CO₂ flux and concentration measurements in this study area. The flux was measured using a West Systems LI820 accumulation chamber CO₂ flux meter. The chamber CO₂ concentration was monitored for 120 seconds and flux was calculated based on the rate of change, chamber parameters, and environmental conditions (e.g. Lewicki et al., 2005). Soil gas samples were also collected and analysed in a stainless steel
probe for CO₂, O₂, H₂S, and H₂ using infrared and electrochemical analysers (Geotechnical Instruments GA2000 or Draeger Multiwarn). Further, laboratory analysis was conducted for helium on a Varian mass spectrometer, and for light hydrocarbons, CO₂, N₂, and O₂ + Ar on two Fisons 8000-series bench Gas Chromatographs (GC). The results were published by Beaubien et al. (2008) explaining the changes in soil chemistry in the shallow subsurface region.

Major and minor oxides in the soil were analysed using a Philips MagiX-PRO WD-XRFS, equipped with a 60 kV generator and 4 kW rhodium (Super Sharp) end-window X-ray tube, and controlled via a PC running the PANalytical SuperQ XRF application package (Beaubien et al., 2008). All spectral backgrounds and peaks were corrected for instrument drift using two external ratio monitors and inter-element effects were corrected using theoretical alpha coefficients. This study confirmed the presence of useful indicators such as Iron (III) Oxide and Calcium Oxide. These oxides are reduced when the soil pH decreases at the surface due to hydrocarbon or CO₂ seepage (van der Meer et al., 2002).

Based on the flux data, an image was generated indicating the transect line of the measurement, as shown in Figure 5.3 (a). The size of the black ‘+’ markers in this image corresponds to the magnitude of the flux, providing the means for visually validating processing results. Those areas with high flux were highlighted as gas vents. Digital Elevation Model (DEM) data for the study area was also acquired in 2005 by OGS, shown in Figure 5.3 (b), which provides additional terrain information for data processing. It can be seen that the study area is predominantly flat with a homogenous land cover, containing both tall and short vegetation.

The other two areas of interest in the study site are relatively larger, namely ‘Study Area 2’ and ‘Study Area 3’, as illustrated in Figure 5.4. In addition to vegetation, it can be seen that the chosen areas also include features such as road networks and buildings. These features add more heterogeneity to the data.

The optical remote sensing datasets for ‘Study Area 2’ include both airborne multispectral and hyperspectral imagery. The former was acquired by the British Geological Survey (BGS), UK, using a Daedalus 1268 ATM multispectral scanner, mounted on the NERC Airborne Research and Survey Facility (ARSF) aircraft, in May 2005. This sensor records information in 11 bands ranging between the visible
and Thermal Infra-Red (TIR) wavelengths (0.42 - 13 μm) (Bateson et al., 2008) at 2.5m resolution. Error! Reference source not found. The latter was acquired by the OGS in July 2007 using the AISA Eagle 1K sensor at 1m resolution and containing 63 bands, ranging from visible to NIR wavelengths (0.40 - 0.99 μm).

<table>
<thead>
<tr>
<th>Band Number</th>
<th>Wavelength range (μm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.42-0.45</td>
</tr>
<tr>
<td>2</td>
<td>0.45-0.52</td>
</tr>
<tr>
<td>4</td>
<td>0.52-0.60</td>
</tr>
<tr>
<td>4</td>
<td>0.60-0.62</td>
</tr>
<tr>
<td>5</td>
<td>0.63-0.69</td>
</tr>
<tr>
<td>6</td>
<td>0.69-0.75</td>
</tr>
<tr>
<td>7</td>
<td>0.76-0.90</td>
</tr>
<tr>
<td>8</td>
<td>0.91-1.05</td>
</tr>
<tr>
<td>9</td>
<td>1.55-1.75</td>
</tr>
<tr>
<td>10</td>
<td>2.08-2.35</td>
</tr>
<tr>
<td>11</td>
<td>8.5-13.0</td>
</tr>
</tbody>
</table>

Figure 5.3: (a) Field gas flux measurements taken at the study area in Latera in 2006 - the boxes indicate known CO₂ vents, and (b) DEM data of the study area.

Table 5.2: The list of ATM bands used for Latera survey (After Pearce et al. (2007)).

For ‘Study Area 3’, spaceborne imagery is available that was acquired by the multispectral Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) instrument, when the Terra satellite flew over Latera at 10:09:57 UTC on 13th May 2006. The image was pre-processed from raw acquisition with both geometric and radiometric correction, called the level 1-B product, and distributed by
Field Sites and Data Pre-Processing

Figure 5.4: (a) The region of Latera in Italy, highlighting features such as the caldera rim and local road networks, and indicating locations of the two study areas; (b) Study Area 2, indicating the ground measurement line and detected regions of high CO$_2$ flux; (c) Study Area 3, indicating all the known CO$_2$ vents in the region.

the ASTER GDS (Ground Data Systems), Japan. ASTER data are acquired with 15m, 30m and 90m resolutions in the visible and NIR (3 bands), Short Wave Infra-Red (SWIR) (6 bands) and TIR regions (5 bands) of the electromagnetic spectrum respectively (0.52 - 11.65 μm). Additional band information and general sensor details are shown in Table 5.3.
Table 5.3: Summary of ASTER sensor and data characteristics (After Eurimage, 2011).

<table>
<thead>
<tr>
<th>Launch Date</th>
<th>Altitude</th>
<th>Inclination</th>
<th>Swath</th>
<th>Revisit at</th>
<th>Sensors</th>
</tr>
</thead>
<tbody>
<tr>
<td>December, 1999</td>
<td>705 km</td>
<td>98.2°</td>
<td>60 Km</td>
<td>equator</td>
<td>16 days</td>
</tr>
</tbody>
</table>

**VNIR**

<table>
<thead>
<tr>
<th>BAND</th>
<th>Spectral Range</th>
<th>Pixel Size</th>
<th>Quantisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.52-0.60</td>
<td>15</td>
<td>8 bit</td>
</tr>
<tr>
<td>2</td>
<td>0.63-0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3N/3B</td>
<td>0.78-0.86</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SWIR**

<table>
<thead>
<tr>
<th>BAND</th>
<th>Spectral Range</th>
<th>Pixel Size</th>
<th>Quantisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.60-1.70</td>
<td>30 m</td>
<td>8 bit</td>
</tr>
<tr>
<td>5</td>
<td>2.145-2.185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.185-2.225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.235-2.285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2.295-2.365</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2.360-2.430</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TIR**

<table>
<thead>
<tr>
<th>BAND</th>
<th>Spectral Range</th>
<th>Pixel Size</th>
<th>Quantisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>8.125-8.475</td>
<td>90 m</td>
<td>12 bit</td>
</tr>
<tr>
<td>11</td>
<td>8.475-8.825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>8.925-9.275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>10.25-10.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>10.95-11.65</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Raw data for CO₂ flux and concentration are also available from the University of Rome ‘La Sapienza’, collected during the 2006 field survey, for the validation of processing results.

5.3 The Laacher See Study Site in Germany

This study site is a lake in Maria Laach, a nearly circular flooded caldera in the east Eifel volcanic district of the Rhenish Massif in south west Germany. This is the largest maar lake formed, compared to those in the west Eifel region, as a result of the most recent eruption of a huge volcano producing some 5 km³ of highly evolved mafic phonolite magma (Schmincke et al., 1989), about 11,000 years BP (Aeschbach-Hertig et al., 1996). Today, the lake in the Maria Laach caldera has a maximum depth of 52 m and a surface area of 3.31 km², covering 27% of its drainage area. Degassing of up to 99 vol. % of CO₂ occurs from the mantle, as indicated by the Helium and Carbon isotope data collected in the area (Giggenbach et al., 1991). Surface gas emissions have been found from several points in and around the lake (Aeschbach-Hertig et al., 1996).

5.3.1 Geology of the Caldera

The main volcanic activity in the Rhenish Massif occurred during the Tertiary period. Quaternary volcanic activity started about 0.7 million years ago and was essentially confined to the west and east Eifel fields (Lippolt, 1983). The typical
craters called *maars* were formed by phreatomagmatic eruptions, i.e. steam explosions due to the contact of water with a magma chamber near the surface (Lorenz, 1973), and eventually leading to the formation of the lake. The ringwall, rising 90 to 240 m above the basin, is made up by basanitic/tephritic cinder cones and the tephra deposits of the Laacher See eruption.

Numerous mineral springs are located in the Eifel region, most of which are rich in dissolved CO₂ and are commercially exploited (Ulrich, 1958). The internal structure of the Laacher See basin is dominated by a NE-SW striking thrust (Krüger et al., 2009). CO₂ is produced below the extinct volcanic caldera, emerges from degassings of the upper mantle and migrates along faults and fractures to the surface. Griesshaber et al. (1992) found maximum contributions of isotopic Carbon and Helium from the mantle towards Laacher See. Their results for gas analysis also confirmed earlier measurements of Giggenbach et al. (1991) regarding the mantle signature of the CO₂.

Release to the atmosphere typically occurs from gas vents, characterised by a small core of elevated gas flux. Discharge of CO₂ along the eastern shore of Laacher See has been known for many years (Schmidt-Ries, 1955; Bahrig, 1985). Rising gas bubbles can be observed on the water surface at some locations near the Eastern shore. In a recent field survey, Jones et al. (2009) reported the occurrence of CO₂ in the western shores of the lake with concentrations of up to 1,600 ppm against an average background level of 200 ppm.

5.3.2 Data Available for the Laacher See Study Site

The area of interest that was chosen in the study site is a small section of the terrestrial surface along the western shores of the lake, as illustrated in Figure 5.5.

The optical remote sensing data for this study area was acquired by OGS, through surveys conducted in 2009, using the AISA Eagle 1K sensor at 1m resolution and containing 63 bands, ranging from visible through NIR wavelengths (0.40 - 0.99 μm). Figure 5.6 shows the field measurement results related to soil CO₂ concentrations for this area obtained in 2008 (Jones et al., 2009), which are useful for visual validation of the processing results.
Field Sites and Data Pre-Processing

Figure 5.5: The region of Laacher See lake in Germany (a); indicating the location of the study area on its western shore (b).

Figure 5.6: Image constructed by kriging interpolation of mobile laser data overlaid on an areal photograph (After Jones et al. (2009)).

5.4 Data Pre-Processing

Pre-processing, or image enhancement, is an essential step for all kinds of image processing problems. The objective of enhancement is to obtain an image, from the original image, that is useful to solve a particular problem (Gonzalez and Woods, 2002). The multivariate character of optical remote sensing data cube makes it amenable for data enhancement transformations. The new representation enhances
the essential information in such a way that patterns that are not easily discernible in the original data may become relatively more apparent after transformation.

The two pre-processing steps that are considered necessary and sufficient for the retention of valuable information in optical remote sensing data in this thesis are: noise reduction by Maximum Noise Fraction (MNF), and dimensionality reduction by Principal Component Analysis (PCA). Using the MNF transformation, the noisy vectors in the data are first discarded. After the noise is sufficiently minimised, the dimensionality of the data is also reduced (reducing the number of bands) using a band grouping scheme based on PCA. This is discussed here in the context of spatial analysis for leakage detection.

5.4.1 Noise Reduction

Noise affects the data cube in a disproportionate manner, such that the amount of noise variance mixed with useful information is generally unequal from one band to the other. It is possible to apply spatial low pass filters, such as the edge preserving smoothing filter and anisotropic diffusion, to filter noise in each band and lose very little valuable information. However, as an alternative solution, the Maximum Noise Fraction (MNF) or Noise Adjusted Principal Components (NAPC) technique was applied because it can perform noise reduction for multispectral images by estimating the noise covariance matrix given by:

\[
\Sigma_a = E\{[Z(x) - Z(x + \Delta)]^2}\}
\]

(5.1)

where, \(x\) is the pixel location, \(\Delta\) is a very short lag vector, typically equal to unity, and \(Z(x)\) is a multivariate pixel vector.

Based on the noise covariance, the technique calculates the noise fraction matrix and transforms it using PCA or Singular Value Decomposition (SVD). This helps in the estimation of noisy eigenvectors that requires elimination in order to restore the true Digital Numbers (DN).
Latera Data Noise Reduction Results

For 'Study Area 1', both hyperspectral datasets have been affected by noise in some of the bands, particularly in the NIR wavelength region. Figure 5.7 shows examples from the 2005 and 2007 datasets (for Band 1 = 0.402 μm).

![Image of Latera Data affected by noise](image)

Figure 5.7: Latera Data affected by noise (a) Band 1 of the 2005 hyperspectral data; (b) Band 1 of the 2007 hyperspectral data.

It is clearly seen that these daytime images have been affected by different levels of noise even though they were acquired by the same AISA Eagle 1K sensor. This maybe owing to the application of geometric correction before noise removal which tends to smear noise across pixels, particularly for low resolution data as seen in Figure 5.7 (a), which has a lower resolution of 2m against that of Figure 5.7 (b) at 1m.. The condition of sensor electronics at different times corresponding to these acquisitions could also have contributed to different noise levels.

Despite these differences, it is observed that MNF operates very well under such conditions because it uses the 2nd order noise statistics to perform data transformation, and does not assume that all bands have the same level of noise, i.e. their variances can be different. Figure 5.8 illustrates the corresponding results obtained after applying the MNF.

The rule of thumb that was used to identify and discard the noisy eigenvectors in the MNF transformed space is based on the assumption that these vectors generally have skewness values close to zero. This assumption works for most cases except where
the original data is inherently Gaussian or has been explicitly transformed to achieve a Gaussian distribution. This case is however very unlikely in optical remote sensing data.

Figure 5.8: Latera data after applying MNF: (a) Band 1 of the 2005 hyperspectral data; (b) Band 1 of the 2007 hyperspectral data.

Figure 5.9: (a) Band 1 of the 2005 multispectral data for Latera affected by noise; (b) Band 1 after applying MNF.
Field Sites and Data Pre-Processing

For 'Study Area 2', the multispectral data also contains noise in the NIR region (for Band 1 = 0.435 μm). Although the spatial resolution is 2.5m, the noise content appears to be lesser in comparison to the Band 1 image of the 2005 hyperspectral dataset for 'Study Area 1', Figure 5.7 (a), which is generally due to the spectral averaging in multispectral data. Figure 5.9 illustrates the result after applying MNF on the data. Similar results are also obtained using 2007 hyperspectral data in the NIR region (for Band 1 = 0.402 μm), as shown in Figure 5.10. For 'Study Area 3', the spaceborne ASTER data did not contain noise that was relatively significant and hence MNF was not applied.

Laacher See Data Noise Reduction Results
For the study area near the western shore of the Laacher See, the hyperspectral data bands were severely corrupted by noise, particularly for the NIR region. As in the case of Latera data, the noise could have been exaggerated as a result of geometric correction. Geometric transformation into a 2D coordinate system, such as the Universal Transverse Mercator (UTM) coordinates in this case, smears pre-existing noise in the data thereby decreasing its initial quality. Hence it is recommended that
geometric transformation should always be done only after noise in the data has been significantly reduced using techniques such as the MNF transformation. Figure 5.11 illustrates the result obtained for Band 1 (0.402 μm) after applying MNF on the Laacher See data.

![Figure 5.11: (a) Band 1 of the 2009 hyperspectral data for Latera corrupted by noise; (b) Band 1 after applying MNF.](image)

5.4.2 Dimensionality Reduction

PCA is one of the most commonly adopted techniques to reduce dimensionality. However, as the dimensionality increases, PCA might not necessarily be optimal for feature extraction problems, such as classification and anomaly detection (Cheriyadat and Bruce, 2003). In other words, although PCA yields satisfactory outcomes for multispectral data, it copes badly with from large intra-class variations in hyperspectral data (Tsai et al., 2007), hence, a bias is introduced by the Principal Components (PCs) where data variance is majorly accounted only for shorter wavelengths, i.e., visible and NIR regions of the spectrum. In addition to this, PCA overlooks local variances that correspond to interesting anomalous patterns in the global data statistics.

Considering these disadvantages of the PCA, in this research a different approach was developed in this PhD research to achieve dimensionality reduction of the data. Although it is based on PCA, the eigenvectors that are used to represent the dimension reduced space does not necessarily obey a decreasing trend in their eigenvalues. This was performed using a band grouping scheme as follows:
Field Sites and Data Pre-Processing

\[
\text{IF } \lambda_r (\text{PC1}(1:N)) \geq 99% \\
\text{IF } \gamma (\text{PC1}(1:N)) - \gamma (\text{PC1}(1:N-1)) = \delta \\
N = N + 1;
\]

\[
\text{ELSE} \\
N = N + 1; \\
\text{IF } \lambda_r (\text{PC1}(1:N)) < 99% \\
\text{DECLARE \text{Band\_Group}(N-2), Stop;}
\]

\[
\text{ELSE} \\
\text{DECLARE \text{Band\_Group}(N-1), Stop;}
\]

where \(N\) is number of bands, \(\text{PC1}\) the first PC, \(\lambda_r\) the local relative eigenvalue, \(\gamma\): skewness, \(\delta\) a very small difference, \text{Band\_Group} is finalising a band group, and \text{Stop} denotes the end of iteration. The scheme was applied only for the hyperspectral datasets.

The local relative eigenvalue \(\lambda_r\) of a band group containing \(N\) bands is given by:

\[
\lambda_r = \frac{\lambda_{\text{PC1}}}{\sum_{n=1}^{N} \lambda_{\text{PCn}}} \times 100
\]  

(5.2)

The data obtained from this dimensionality reduction scheme has three critical characteristics:

1. \textbf{Band reduction}: The reduction in the number of bands using the proposed approach is comparable to that of conventional PCA approach.

2. \textbf{Data distribution}: The average energy, associated with the anomalous values in the distribution, retained in the data after dimensionality reduction is shown to be higher. This essentially improves the chance for better performance in detection problems.

3. \textbf{Spatial correlation}: While a conventional PCA approach gives components that are also mostly orthogonal in the spatial sense (Goovaerts, 1993), i.e. original spatial correlation in the data are either lost or changed in the components, the proposed approach maintains all short and long range correlations of the data.
**Latera Data Dimensionality Reduction Results**

The proposed band grouping approach is applied for ‘Study Area 1’ and ‘Study Area 2’ hyperspectral datasets only. The eigenvalues obtained are numerically compared to those obtained using conventional PCA. Table 5.4 shows the results obtained for ‘Study Area 1’ data.

<table>
<thead>
<tr>
<th>Eigenvectors</th>
<th>Eigenvalues λ (conventional PCA)</th>
<th>Energy outside the Confidence Interval (for ±2σ)</th>
<th>Eigenvalues λ (proposed approach)</th>
<th>Energy outside the Confidence Interval (for ±2σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>e₁</td>
<td>30878808</td>
<td>1420425</td>
<td>3145931</td>
<td>144713</td>
</tr>
<tr>
<td>e₂</td>
<td>14323239</td>
<td>658869</td>
<td>3040689</td>
<td>139872</td>
</tr>
<tr>
<td>e₃</td>
<td>195355</td>
<td>8986</td>
<td>13418535</td>
<td>617253</td>
</tr>
<tr>
<td>e₄</td>
<td>103457</td>
<td>4759</td>
<td>565329</td>
<td>26005</td>
</tr>
<tr>
<td>e₅</td>
<td>68763</td>
<td>3163</td>
<td>355619</td>
<td>16358</td>
</tr>
<tr>
<td>e₆</td>
<td>14529</td>
<td>668</td>
<td>522542</td>
<td>24037</td>
</tr>
<tr>
<td>e₇</td>
<td>10487</td>
<td>482</td>
<td>23787583</td>
<td>1094229</td>
</tr>
<tr>
<td>e₈</td>
<td>5490</td>
<td>253</td>
<td>412602</td>
<td>18980</td>
</tr>
<tr>
<td>e₉</td>
<td>3111</td>
<td>143</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>e₁₀</td>
<td>2486</td>
<td>114</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>e₁₁</td>
<td>1571</td>
<td>72</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Percentage of Total Data Variance</td>
<td>100%</td>
<td>-</td>
<td>99%</td>
<td>-</td>
</tr>
<tr>
<td>Average Energy</td>
<td>-</td>
<td>190721</td>
<td>-</td>
<td>260181</td>
</tr>
</tbody>
</table>

The following general observations are made:

- Both approaches achieve comparable reduction in dimensionality and nearly equal amounts of information are retained in terms of data variance. It can be seen that the proposed approach represents at least 99% of the total data variance after compression.

- Assuming Gaussian distribution for eigenvectors, the proposed approach has a higher average energy outside the ±2σ Confidence Interval (CI) where anomalous data reside, i.e. within 4.6% of the total data variance which may also
include noise. This implies that the proposed scheme provides a better chance for detection algorithms to work.

The cross-variogram (semi-variance) function for the first two eigenvectors (e₁, e₂) is also generated, as shown in Figure 5.12. It is seen that when the conventional PCA is applied the continuity in spatial correlation is affected, e.g. at about 15m the correlation trend has a sharp turning point, and is not data representative. This introduces difficulty in modelling the variogram function for the purpose of defining the co-kriging filter. Moreover, the nugget value is high which suggests that the components are uncorrelated at very short lag distances, which gives filter parameters (kriging weights) that introduces noisy results. On the other hand, the proposed approach gives components that maintain both short and long range correlations, with a low nugget value.

![Cross-variogram functions](image)

**(a)**

Figure 5.12: ‘Study Area 1’ data spatial analysis: Cross-variogram functions of eigenvectors (a) for conventional PCA; (b) for proposed approach.

Similar analysis was carried out for the ‘Study Area 2’ data, and the results are shown in Table 5.5. The observations remained consistent for a larger area.

Spatial analysis using the cross-variogram for eigenvectors (e₁, e₂), as shown in Figure 5.13 also reveals that, despite a lower nugget value, there exists discontinuity in the spatial correlations, which is not representative of the original correlations before dimensionality reduction was applied.
Table 5.5: Comparison between the conventional PCA and proposed approach for dimensionality reduction for ‘Study Area 2’ data.

<table>
<thead>
<tr>
<th>Eigenvectors</th>
<th>Eigenvectors $\lambda$ (conventional PCA)</th>
<th>Energy outside the Confidence Interval (for $\pm 2\sigma$)</th>
<th>Eigenvectors $\lambda$ (proposed approach)</th>
<th>Energy outside the Confidence Interval (for $\pm 2\sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>12829755</td>
<td>590169</td>
<td>1703964</td>
<td>78382</td>
</tr>
<tr>
<td>$e_2$</td>
<td>5181163</td>
<td>238333</td>
<td>1598001</td>
<td>73508</td>
</tr>
<tr>
<td>$e_3$</td>
<td>278972</td>
<td>12833</td>
<td>247163</td>
<td>11369</td>
</tr>
<tr>
<td>$e_4$</td>
<td>85653</td>
<td>3940</td>
<td>1389260</td>
<td>63906</td>
</tr>
<tr>
<td>$e_5$</td>
<td>30792</td>
<td>1416</td>
<td>189319</td>
<td>8709</td>
</tr>
<tr>
<td>$e_6$</td>
<td>5102</td>
<td>235</td>
<td>168067</td>
<td>7731</td>
</tr>
<tr>
<td>$e_7$</td>
<td>2156</td>
<td>99</td>
<td>200692</td>
<td>9232</td>
</tr>
<tr>
<td>$e_8$</td>
<td>1443</td>
<td>66</td>
<td>10135125</td>
<td>466216</td>
</tr>
<tr>
<td>$e_9$</td>
<td>867</td>
<td>40</td>
<td>2631907</td>
<td>121068</td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>106</td>
<td>5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$e_{11}$</td>
<td>45</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Percentage of Total Data Variance

- 100%
- 99%
- 

Average Energy

- 77013
- 93347

Figure 5.13: ‘Study area 2’ data spatial analysis: Cross-variogram functions of eigenvectors (a) for conventional PCA; (b) for proposed approach.

Laacher See Data Dimensionality Reduction Results

Table 5.6 shows that the average energy values are now close to each other in comparison to the earlier cases – the increase of energy outside the CI is +13% in comparison to the cases in Table 5.4 (+36%) and Table 5.5 (+21%). One reason for this may be because the number of components obtained from both methods is nearly
the same, with only a difference of one eigenvector. However, from all the examples demonstrated here, it is noted that the rate of decrease in the eigenvalues for the conventional PCA is higher. This generally has an impact on its average values, causing it to become lower than those obtained from the proposed approach. This also has an implication that the number of effective directions of finding leakage anomalies is lower for the conventional PCA method.

**Table 5.6:** Comparison between the conventional PCA and proposed approach for dimensionality reduction for Laacher See data.

<table>
<thead>
<tr>
<th>Eigenvectors</th>
<th>Eigenvalues $\lambda$ (using conventional PCA)</th>
<th>Energy outside the Confidence Interval (for $\pm2\sigma$)</th>
<th>Eigenvalues $\lambda$ (using proposed approach)</th>
<th>Energy outside the Confidence Interval (for $\pm2\sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>24059717</td>
<td>0.046 $\lambda$</td>
<td>228800</td>
<td>10525</td>
</tr>
<tr>
<td>$e_2$</td>
<td>5688881</td>
<td>1106747</td>
<td>1337362</td>
<td>61519</td>
</tr>
<tr>
<td>$e_3$</td>
<td>593895</td>
<td>261689</td>
<td>1958169</td>
<td>90076</td>
</tr>
<tr>
<td>$e_4$</td>
<td>164434</td>
<td>27319</td>
<td>2306264</td>
<td>106088</td>
</tr>
<tr>
<td>$e_5$</td>
<td>94295</td>
<td>7564</td>
<td>808011</td>
<td>37169</td>
</tr>
<tr>
<td>$e_6$</td>
<td>29295</td>
<td>4338</td>
<td>19629464</td>
<td>902955</td>
</tr>
<tr>
<td>$e_7$</td>
<td>26928</td>
<td>1348</td>
<td>14161689</td>
<td>191438</td>
</tr>
<tr>
<td>$e_8$</td>
<td>13081</td>
<td>1239</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Percentage of Total Data Variance | 100%                                    | 99%                                                   |
| Average Energy                   | -                                       | 176356                                                | 199967                                          |

**Figure 5.14:** Laacher See data spatial analysis: Cross-variogram functions of eigenvectors (a) for conventional PCA; (b) for proposed approach.
The cross-variogram function of the eigenvectors \((e_1, e_2)\) shows a cyclical variation, in Figure 5.14, after applying conventional PCA. Although this can be modelled using a sinusoidal variogram function, it indicates that it may be necessary to consider a large neighbourhood in order to model such spatial variability, thus increasing the computations required for processing.

5.5 Conclusions

For the purpose of developing a new unsupervised methodology of CO\(_2\) leakage detection, data is required that contains information related to leakage. Since most industrial sites have no leakage reported so far, data has been acquired and used instead from natural analogue sites in the Latera and Laacher See sites in Italy and Germany respectively.

Given the volume of information contained in optical remote sensing data, it is essential to perform pre-processing that can reduce factors such as data corruption by noise and high dimensionality prior to dealing with specialised problems such as detection of leakages. However, care is taken such that information loss is minimal.

Noise is filtered from the available datasets by using the Maximum Noise Fraction (MNF) transformation. The level of noise observed for the data is different in each case even though they were acquired by the same sensor. Additionally, lower spatial resolution and geometric corrections in the data can emphasise noise. In all such cases, however, MNF is seen to work consistently.

Dimensionality reduction is performed only for hyperspectral datasets by using a band grouping approach based on PCA. Primarily, this approach gives similar number of eigenvectors as in the case of conventional PCA. However, these eigenvectors have a higher average energy outside the 95.4% \((\pm 2\sigma)\) confidence interval where the low probability (anomalous) pixels corresponding to leakage are expected to reside. This is also consistent because, almost always, the rate of decreasing eigenvalues in conventional PCA is higher than the proposed approach. The spatial correlation analysis between the eigenvectors reveals that conventional PCA introduces complexities in modelling the variogram, and subsequently the co-
kriging filter parameters. These factors do not affect the proposed approach and this increases the reliability of band grouping for dimensionality reduction.
Chapter 6  
Unsupervised Leakage Detection using Optical Remote Sensing Data

6.1 Introduction

After the pre-processing stage discussed in Chapter 5, the data are ready for analysis using the unsupervised methodology developed in this thesis. The first stage of the methodology aims to process the data in order to locate all surface anomalies, including those that may be related to CO₂ leakage, called the prior detection, and to cluster the data pixels in groups so as to find expressions on the surface that allow the likelihood of leakage to be estimated. This stage is described and implemented in two parts presented in this chapter.

In the first part, geostatistical processing of the images is carried out using Intrinsic Random Functions (IRF) followed by multivariate analysis using Independent Component Analysis (ICA). The ICA is performed in the residual subspace of the data containing anomalies, which are uncharacterised. These are point anomalies with low probability of occurrence, and whose spatial correlation has a short range thereby making them spectrally different from their corresponding neighbourhood. The anomalies generally occur in the tails of a multivariate data distribution. Therefore, the positively skewed non-Gaussian distributions are found in order to enhance the anomalies in the residual subspace and are identified as the prior detection result.

113
The second part involves unsupervised learning and grouping of the data pixels in the prior detection result into natural clusters based on their spectral signature using the Growing Hierarchical Self-Organising Maps (GHSOM). The GHSOM method assists in forming a likelihood map that provides the indirect evidence of CO₂ leakages on the surface.

6.2 IRF and ICA for Prior Detection

A new application of the IRF theory is proposed here to process optical remote sensing data. These data generally exhibit complex non-stationary spatial characteristics that are attributed to the presence of both inter-class (land cover) and intra-class radiometric variability. An analogous process that this variability was compared to is the force equation representing 2D Brownian motion of particles, which is an IRF-0 process. The mean square displacement in particles exhibiting motion of this type is given by the 1ˢᵗ order 1D differential equation (Uhlenbeck and Ornstein, 1930), which can be easily extended to 2D:

\[ m \frac{du}{dt} = -fu + F(x_0, x, t) \]  \hspace{1cm} (6.1)

where, \( m \) is the mass of the particle, \( u \) is the velocity of the particle, \( f \) is the friction coefficient, and \( x_0 \) is the initial particle position at \( t = t_0 \). The influence of the surrounding medium on the particle is divided into two parts: (1) a systematic force (deterministic): \(-fu\) which describes friction is a 1ˢᵗ order monomial, and (2) a fluctuating force (random): \( F(t) \), such that the mean \( E\{F(t)\} = 0 \). The fluctuating force is equivalent in representation to a regionalised variable \( Z(p) \) in geostatistics, as in equation (4.4). The solution for fluctuation force \( F(t) \) is given by:

\[ F(x', x, t) = \left( \frac{1}{4\pi t} \right)^{1/2} e^{-\frac{x^2}{4Dt}} \]  \hspace{1cm} (6.2)

where, \( s = x - x' \)

and \( D \) is a diffusion coefficient of the particles. This solution reveals that the fluctuations have an underlying Gaussian distribution, attributed to the displacement variable \( s \). These fluctuations are made up of the sum of short and long range
correlations, characterised respectively by $Z_0(p)$ and $P(p)$ in equation (4.4). Therefore by applying the IRF theory, the aim is to extract the residual stationary process $Z_0(p)$ containing only the short range correlations for detecting surface anomalies.

6.2.1 Application of the Prior Detection methodology for the Latera Data

The first step was to detrend the pre-processed data. This ensures that the pixel values oscillate in such a way that the global average is essentially near zero. This is done to obey the preliminary constraint stated in equation (4.4). 2D regression was used to model the global behaviour of the data points using a polynomial, which is subsequently removed from the images by subtraction. Although, this trend may represent the global average illumination of the image, which is a multiplicative factor, subtraction is carried out to give a statistical treatment for the detection problem.

For the 'Study Area 1' hyperspectral data, using the eigenvectors obtained after noise and dimensionality reduction discussed in chapter 5, 2D polynomials are regressed individually on each eigenvector by using least-squares fitting based on the Vandermonde matrix, e.g. the polynomial that was used to detrend the first eigenvector is of the 1st order, given by:

$$1.2631 - 0.0118x - 0.0007y \quad (6.3)$$

Figure 6.1 shows the polynomial regressed on the 1st eigenvector and the detrended eigenvector for 'Study Area 1'.

A low order polynomial was chosen because:

- there is a high probability of finding pixel neighbourhoods that exhibit good local homogeneity in high spatial resolution airborne images, so a lower order polynomial is sufficient; and

- in the Fourier domain, low order polynomials correspond to low frequency data components, equivalent to the frictional force affecting the Brownian motion of particles in equation (6.1).
Next, the spatial correlation between different detrended eigenvectors was estimated. Kitanidis (1993) proved that the variogram function model of a detrended data can be used as an approximation for the Generalised Covariance Function (GCF) of an IRF process. Hence, the multivariate spatial correlation structure between the eigenvectors was accounted for by the cross-variogram functions. A suitable model was chosen for the functions through visual curve fitting in the Isatis geostatistical software (Geovariances, 2010). The fitted model is given by a nested function of the form:
Unsupervised Leakage Detection using Optical Remote Sensing Data

\[ \gamma(h) = \begin{cases} c_0, & h = 0 \\ c_0 + c_1 \left[ \frac{3h}{2a_1} - \frac{1}{2} \left( \frac{h}{a_1} \right)^3 \right] + c_2 \left[ \frac{3h}{2a_2} - \frac{1}{2} \left( \frac{h}{a_2} \right)^3 \right], & 0 < h < a_2 \\ c_0 + c_1 + c_2, & h \geq a_2 \end{cases} \]

(6.4)

where, \( c_0 = 121309.501 \) (Nugget), \( a_1 = 25 \text{m} \), \( c_1 = 1180221.41 \) (range and sill of the 1st spherical function), and \( a_2 = 120 \text{m} \), \( c_2 = 1609724.97 \) (range and sill of the 2nd spherical function). The nugget value \( (c_0) \) is the variance of the shortest range correlations in the data, such as noise or point anomalies. The other relatively longer range correlations are represented by spherical functions of 25m and 120m. Using this model, the co-kriging system estimates the weights of the filter to be used to process the data. Two factors were considered that dictate the quality of the co-kriging result using the cross-variogram model.

Cross-validation test: This was used to test the cross-variogram model. In cross-validation, comparison between the true values of the data \( (Z) \) and the estimated values \( (Z^*) \) from co-kriging are quantified by the variance of standardised error, given by:

\[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{Z^*(p_i) - Z(p_i)}{\sigma} \right)^2 \]

(6.5)

where, \( \sigma^2 \) is the kriging variance estimated using the model. This quantity corresponds to the ratio between the experimental and theoretical variance. Hence, for a good variogram model it is desired that the variance of standardised error remains as close as possible to unity, with a very small error value \( (Z^* - Z) \), ensuring unbiased estimation. For the model parameters in equation (6.4), the mean error value is 0.00205 and the standardised error variance is 1.3.

Neighbourhood size: Although it maybe possible to choose an ALC-0 neighbourhood with size up to the range of the cross-variogram function, it is computationally very expensive to use such large neighbourhoods for raster datasets such as optical remote sensing data. Thus, in order to choose an appropriate size the frequency (Fourier) response of the filter was studied, as illustrated in Figure 6.2. It is seen that for a neighbourhood of size 3 x 3, the frequency response obeys the
characteristics of a high pass filter, i.e. a filter that extract high frequency information (only short range correlations) of the image by filtering out low frequency polynomials. As the size increases to $7 \times 7$ and beyond, the filter is seen to over-smooth the short range correlations (low pass filter) thus compromising its ability for potential leakage detection and increasing the cost of computation. However, the $5 \times 5$ neighbourhood (band pass filter) was chosen in this case for co-kriging to ensure that even interesting point anomalies will have a correlation greater than 1 pixel lag in the filter result.

After having tested the model using cross-validation and selection of a suitable neighbourhood, cokriging was performed for all eigenvectors, conditioning on elevation data (see Figure 5.3 (b)) as an external drift variable. Figure 6.3 illustrates only the result obtained for the first eigenvector. It can be seen that most of the information is removed essentially retaining only short range correlations (Figure 6.3 (b)). The portion of image information that is eliminated (Figure 6.3 (c)) is made up of many locally constant values, monomials or polynomials that together describe the longer range correlations.

All the filtered eigenvectors are then used to estimate the new non-Gaussian vectors using the FastICA algorithm. This algorithm is implemented in the ICASSO Matlab toolbox for independent components reliability analysis (Himberg and Hyvärinen, 2003). Skewness was chosen as the ideal objective function because it is assumed that detection of low probability anomalies is primarily done by analysing positively skewed data distributions.

However, in order to speed up ICA computations, an additional step was implemented. This involved performing image downsampling, according to the Nyquist-Shannon sampling theorem (Shannon, 1949), where alternate pixels were removed from each row/column. This is possible because an image that is filtered using a band pass filter, such as the chosen ALC-0 $5 \times 5$ neighbourhood (see Figure 6.2), is a smoothed version of short range correlations. Hence reducing the number of pixels by half, along the $x$ and $y$ dimensions of the image does not affect the useful information for anomaly detection. For an image of size $N \times N$, downsampling was performed to obtain a new size $N/2 \times N/2$. 

118
Figure 6.2: Effect of increasing neighbourhood size on the frequency response of the ALC-0 filter (a) 3 x 3 filter; (b) 5 x 5 filter; (c) 7 x 7 filter.
Figure 6.3: Latera ‘Study Area 1’ hyperspectral data (a) Detrended eigenvector image; (b) Result obtained after co-kriging with other detrended eigenvectors; (c) Information of the eigenvector that is filtered out.

The number of Independent Component (IC) features (potential candidates) obtained from the downsampled eigenvectors is always less than or equal to the number of eigenvectors. There was a pool of 8 potential candidates obtained, only one of which best represents the prior anomaly distribution. In order to choose the most appropriate candidate, a two-stage elimination procedure was employed (see Table 6.1).
It was observed that those candidate distributions with low skewness value represented background clutter information such as shadows, and high skewness values represented ecotones (transition edges between different types of vegetation). Therefore, such ICs were not considered as potential anomaly detection candidates. It was then assumed that in order to represent a prior distribution, one of the remaining candidates must maintain spatial homogeneity in variance, also called *homoscedasticity*. This property was considered essential for selection because it implicitly conveys the fact that all the pixels in the IC are *identically distributed*, i.e. their prior probabilities as anomalous pixels are equally likely. The most appropriate candidate was chosen as the one with the lowest value of the Brown-Forsyth’s (BF) F-statistic (Brown and Forsythe, 1974) indicating its homogeneity in variance, given by:

\[
F = \frac{(N - k) \sum_{i=1}^{k} n (\bar{z}_i - \bar{Z})^2}{(k - 1) \sum_{i=1}^{k} \sum_{j=1}^{n} (z_{ij} - \bar{Z}_i)^2}
\]

where, the data is divided into k pixel groups with n samples each, and \( \bar{z} \) and \( \bar{Z} \) are the mean and median values respectively.

<table>
<thead>
<tr>
<th>Independent Components</th>
<th>IC1</th>
<th>IC2</th>
<th>IC3</th>
<th>IC4</th>
<th>IC5</th>
<th>IC6</th>
<th>IC7</th>
<th>IC8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Skewness</strong></td>
<td>4.79</td>
<td>3.2</td>
<td>0.72</td>
<td>0.55</td>
<td>11.13</td>
<td>4.38</td>
<td>1.78</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>F- Statistic</strong></td>
<td>-</td>
<td>1124.71</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>383.85</td>
<td>599.54</td>
</tr>
</tbody>
</table>

| Stage 1 elimination candidates | Stage 2 elimination candidates | Candidate for prior detection |

Table 6.1: Choosing the candidate to represent prior detection for the Latera ‘Study Area 1’ hyperspectral data (k=4).
The best candidate (IC7) was then upsampled back to the original size from ‘N/2 x N/2’ to ‘N x N' by using inverse-squared distance weighted interpolation, giving the prior detection result shown in Figure 6.4. The ICA projection values were linearly rescaled to give the prior confidence measures for anomalous regions on the surface, with 0 for low confidence and 100 for high confidence. The known high CO₂ flux regions based on field measurements have been marked with black boxes. It can be observed that surface anomalies that may or may not be associated with CO₂ leakage have been assigned with some level of confidence after the geostatistical processing. This is because pure statistical analysis does not inherently associate the anomalies with any specific physical phenomenon such as leakages.

![Image](image.png)

**Figure 6.4:** Prior detection result obtained for the Latera ‘Study Area 1’ hyperspectral data after kriging and ICA, superimposed on artificially illuminated elevation data (contains shadows), with black boxes indicating high CO₂ flux regions.

The prior detection procedure was repeated with the hyperspectral data for ‘Study Area 2’ in the same manner as explained earlier. The polynomial trend for the first eigenvector, shown in Figure 6.5, is given by:

\[
612.5868 - 0.4366x - 2.1727y
\]

(6.7)
Spatial correlation is then modelled using the same nested cross-variogram model structure as in equation (6.4), with the parameters: Nugget: $c_0 = 64763.07$; Spherical function 1 $c_1 = 842098.877$, $a_1 = 15m$; Spherical function 2 $c_2 = 720960.384$, $a_2 = 60m$. This model gave a cross-validation result with mean error of 0.00026, and standardised variance of 1.31. The eigenvectors were filtered based on this model, as shown in Figure 6.6. However, unlike the earlier case no external drift was used as the elevation data or the like was not available for this area.

The residual eigenvectors were subsequently analysed using ICA to obtain 6 ICs as potential candidates for prior detection. Table 6.2 lists the skewness and F-statistic values for all ICs based on which the selection of the best candidate (IC6) was done. The selection procedure based on the skewness and homogeneity of variance values of the ICs was observed to be consistent with the change in size of the image. Figure 6.7 shows the prior detection result obtained after upsampling IC6 through interpolation.

Table 6.2: Choosing the candidate to represent prior detection for the Latera ‘Study Area 2’ hyperspectral data (k=4).

<table>
<thead>
<tr>
<th>Independent Components</th>
<th>IC1</th>
<th>IC2</th>
<th>IC3</th>
<th>IC4</th>
<th>IC5</th>
<th>IC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness</td>
<td>3.11</td>
<td>4.7</td>
<td>8.55</td>
<td>1.27</td>
<td>0.23</td>
<td>3.76</td>
</tr>
<tr>
<td>F- Statistic</td>
<td>891.8</td>
<td>771.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>566.3</td>
</tr>
</tbody>
</table>

The detection result has been scaled linearly to represent prior measures between 0 and 100, for low and high confidence respectively. However, the result now shows that high confidence values are assigned at several places including features that are clearly not leakage areas, such as road networks and possibly a roof-top of a building in the area. This result suggests that it may not possible to accurately determine the desired anomalies for a large area solely through statistical analysis unless additional information is incorporated to improve the result. This issue will be addressed in the following chapter.
Figure 6.5: Latera 'Study Area 2' hyperspectral data (a) 1st eigenvector image; (b) 1st order polynomial trend given by equation (6.7) for the eigenvector image; (c) Detrended eigenvector image.
Figure 6.6: Latera 'Study Area 2' hyperspectral data (a) Detrended eigenvector image; (b) Result obtained after co-kriging with other detrended eigenvectors; (c) Information of the eigenvector that is filtered out.
Figure 6.7: Prior detection result obtained for the Latera ‘Study Area 2’ hyperspectral data after kriging and ICA, with black boxes indicating high CO$_2$ flux regions.

The multispectral data for ‘Study Area 2’ is also analysed here. Since dimensionality reduction is not performed in the earlier chapter for this data, the original set of bands (DN values) were used to perform the prior detection. The 1$^{st}$ order polynomial that is used to detrend the first band (0.435 $\mu$m) of the data, shown in Figure 6.8, is given by:

$$5.8948 - 0.002x - 0.0022y$$

(6.8)

All detrended bands, 11 in number, are used to estimate the spatial correlation model with the same nested cross-variogram model structure as in equation (6.4), with the parameters: Nugget : $c_0 = 212643.521$; Spherical function 1: $c_1 = 772160.509$, $a_1 = 15$m; Spherical function 2: $c_2 = 543875.035$, $a_2 = 60$m. This model gave a cross-validation result with mean error of 0.0282, and standardised variance of 0.85. The bands were filtered based on this model, as shown in Figure 6.9.
Figure 6.8: Latera ‘Study Area 2’ multispectral data (a) band 1 image; (b) 1st order polynomial trend given by equation (6.8) for the band 1 image; (c) Detrended band 1 image.
Figure 6.9: Latera ‘Study Area 2’ multispectral data (a) Detrended band 1 image; (b) Result obtained after co-kriging with other detrended bands; (c) Information of the band 1 that is filtered out.
The residual bands were subsequently analysed using ICA to obtain 11 ICs as potential candidates for prior detection. Table 6.3 lists the skewness and F-statistic values for all ICs based on which the selection of the best candidate (IC8) was made. Unlike earlier cases, however, some of the components have negative skewness values which are not considered because such components do not represent sparse images. Figure 6.10 is the prior detection result obtained after upsampling IC8 through interpolation.

Table 6.3: Choosing the candidate to represent prior detection for the Latera ‘Study Area 2’ multispectral data (k=4).

<table>
<thead>
<tr>
<th>Independent Components</th>
<th>IC1</th>
<th>IC2</th>
<th>IC3</th>
<th>IC4</th>
<th>IC5</th>
<th>IC6</th>
<th>IC7</th>
<th>IC8</th>
<th>IC9</th>
<th>IC10</th>
<th>IC11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness</td>
<td>-1.2</td>
<td>4.5</td>
<td>11.34</td>
<td>1.1</td>
<td>-5.6</td>
<td>2.3</td>
<td>5.1</td>
<td>3.4</td>
<td>8.14</td>
<td>2.6</td>
<td>4.44</td>
</tr>
<tr>
<td>F- Statistic</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>506</td>
<td>-</td>
<td>919</td>
<td>-</td>
<td>-</td>
<td>256</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 6.10: Prior detection result obtained for the Latera ‘Study Area 2’ multispectral data after kriging and ICA, with black boxes indicating high CO₂ flux regions.

Similar to the prior detection result using hyperspectral data, the result using multispectral data also contains surface features that maybe unrelated to leakage,
such as the road networks. However, when visually comparing this result with that obtained using hyperspectral data in Figure 6.7 it was observed that there is an improvement. This maybe due to the fact that in the case of hyperspectral data, the eigenvectors that were used representing different band groups in the original data have large variance values as seen in chapter 5, while the band variance for the multispectral data is relatively lower owing to averaging in the Full Width at Half Maximum (FWHM) region for each band.

6.2.2 Prior Detection using the Laacher See Data

The first order polynomial trend for the first eigenvector using the pre-processed data for the study area in Laacher See is given by:

$$16.3619 - 1.3857x + 0.29394y$$

(6.9)

The result of detrending the eigenvector is illustrated in Figure 6.11.

![Figure 6.11: Laacher See (a) 1st eigenvector image; (b) 1st order polynomial trend given by equation (6.9) for the eigenvector image; (c) Detrended eigenvector image.](image)

Unlike the earlier cases, the detrending result has a small difference when compared to the original eigenvector. The 1st order polynomial that is modelled does not have a large gradient mainly due to the size of the study area and data homogeneity, with values between -124 and +121. This means that almost all the variance in the
eigenvector is retained after detrending. However, this is not the case for all the eigenvectors.

A spatial correlation model for the detrended eigenvectors using the same nested cross-variogram structure as before (See equation (6.4)) is determined, with the parameters: Nugget: $c_0 = 151.63$; Spherical function 1: $c_1 = 16948.73$, $a_1 = 11m$; Spherical function 2: $c_2 = 133798.07$, $a_2 = 70m$. This model gave a cross-validation result with mean error of 0.00244, and standardised variance of 0.91. Figure 6.12 shows the result of filtering the 1$^{st}$ eigenvectors based on this model. Once again no external drift was used as the elevation data or the like was not available for this area.

![Figure 6.12: Laacher See (a) Detrended eigenvector image; (b) Result obtained after co-kriging with other eigenvectors; (c) Information of the eigenvector that is filtered out.](image)

The nugget value is seen to be far lower in comparison to the earlier cases, which is an indication of having lesser anomalous pixels or noise. The residual eigenvectors after co-kriging were subsequently analysed using ICA to obtain 7 ICs as potential candidates for prior detection. Table 6.4 lists the skewness and F-statistic values for all ICs based on which the selection of the best candidate was done. The behaviour for the homogeneity of variance of the prior detection result (IC4) is once again observed to be consistent with earlier cases. It is also noted from all the cases that the skewness of a variable and its F-statistic are independent of each other. Therefore,
the two stage elimination procedure followed here to select the best IC candidate is plausible. Figure 6.13 (b) is the prior detection result after upsampling the candidate using interpolation. It shows higher confidence in the region that is closer to the lake and appears promising when compared to field data map for CO₂ concentration in Figure 6.13 (a).

Table 6.4: Choosing the candidate to represent prior detection for the Laacher See study area (k=4).

<table>
<thead>
<tr>
<th>Independent Components</th>
<th>IC1</th>
<th>IC2</th>
<th>IC3</th>
<th>IC4</th>
<th>IC5</th>
<th>IC6</th>
<th>IC7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness</td>
<td>0.13</td>
<td>12.11</td>
<td>3.45</td>
<td>2.13</td>
<td>10.98</td>
<td>4.38</td>
<td>2.45</td>
</tr>
<tr>
<td>F- Statistic</td>
<td>-</td>
<td>-</td>
<td>435.7</td>
<td>321.3</td>
<td>-</td>
<td>-</td>
<td>919.1</td>
</tr>
</tbody>
</table>

Stage 1 elimination candidates
Stage 2 elimination candidates
Candidate for prior detection

Figure 6.13: (a) Image constructed by kriging interpolation of mobile laser data overlaid on an areal photograph (After Jones et al. (2009)); (b) Prior detection result obtained for the Laacher See study area after kriging and ICA.

6.3 GHSOM for Likelihood Map Generation

Unsupervised learning and clustering using GHSOM was applied to the optical remote sensing datasets that are not pre-processed for dimensionality reduction.
GHSOM assists in automatically creating clusters from the data and classifying the pixels based on these clusters. Each cluster is associated with a spectral signature, called the cluster centre, representing the pixel group. Those clusters that correspond to surface effects caused due to CO2 leakage were used to generate the likelihood distribution. This is possible by identifying relevant diagnostic features in their spectral signatures. However, information about clusters, i.e. their cluster centres, are obtained through visualisation after GHSOM completes the learning process. Visualisation enables the selection of all cluster centres from multiple layers of the map hence identifying the potential clusters for the given data. Clusters are generally considered as areas with high data density. To do this, a SOM-based histogram representation technique for cluster visualisation called Smoothed Data Histograms (SDH) (Pampalk et al., 2002) was used.

The SDH estimates density by interpreting the GHSOM units, obtained after the training process, as bins of a histogram. The width of a bin is a variable defined through the distance between the corresponding unit and its neighbours. The membership degree of an input data vector to a specific bin is determined by a smoothing parameter $s$ and calculated based on the rank of the distances between the data vector and all the available bins, such that membership $M$ to the $s$ closest bins is given by (Pampalk et al., 2002):

$$M(i) = \begin{cases} \frac{s-i}{C_s}, & 0 \leq i \leq s-1 \\ 0, & \text{otherwise} \end{cases}$$

$$C_s = \sum_{i=0}^{s-1} s - i$$

where, $C_s$ is a normalising factor. The influence of the smoothing parameter $s$ on the contour representation of the SOM-SDH is illustrated as an example in Figure 6.14, for a hypothetical dataset drawn from a mixture of five Gaussian distributions. Reducing the smoothing parameter can increase noise ($s = 1$), while increasing the parameter can result in a degenerate density distribution represented by a single composite cluster ($s = 50$).
Figure 6.14: SDH representations for different smoothing parameters, after unsupervised learning using a dataset drawn from a Gaussian mixture (After Pampalk et al. (2002)).

GHSOM and SDH toolboxes in MATLAB (Chan and Pampalk, 2002; Pampalk et al., 2002) were used to replace the conventional Picard’s iteration in the fuzzy c-means technique in order to generate the likelihood distribution of CO₂ leakage using the datasets. This was done because, while using fuzzy c-means, it was necessary to explicitly specify the number of possible classes in the data, which is rather difficult to guess when analysing a heterogeneous study area. On the other hand, with GHSOM, all the natural clusters in the data are automatically discovered. Moreover, owing to the often high correlation between the spectral bands, the data usually lies in a low-dimensional non-linear manifold, which eludes classical unsupervised classification approaches including the fuzzy c-means technique (Villmann and Merényi, 2002).

6.3.1 Likelihood Generation using the Latera Data

Using the hyperspectral data for ‘Study Area 1’, the first step was to normalise the pixel vectors in the data cube by:

\[
x_{\text{norm}} = \left[ \frac{x_1}{\sqrt{x_1 + x_2 + \ldots + x_N}} \quad \frac{x_2}{\sqrt{x_1 + x_2 + \ldots + x_N}} \quad \ldots \quad \frac{x_N}{\sqrt{x_1 + x_2 + \ldots + x_N}} \right]
\]  

(6.11)

where, \( x_{\text{norm}} \) is the normalised vector of a pixel vector \( x \), and \( N = 63 \) bands. The normalisation ensures that all the pixels lie on a unit hypersphere. This is done to ensure that the clustering process is dependent only on the shape of the spectral signature, and not its absolute values. The normalised pixel vectors were then supplied as input to initiate unsupervised learning using the training function available in the GHSOM toolbox. For the proper selection of the breadth (\( \tau_1 \)) and
Unsupervised Leakage Detection using Optical Remote Sensing Data

depth ($\tau_2$) parameters, it was necessary to obey the constraint in equation (4.33). It was observed through experimentation that if $\tau_1$ was $>0.5$, the global criterion is met too soon and lateral growth of the SOM layer is terminated. Likewise, the local criterion is met too soon if $\tau_2$ is $>0.1$, and new SOM layers are not formed. In these cases, the result contains an insufficient number of clusters and hence does not represent the data density completely. The rule of thumb that was followed to avoid this problem was to choose $\tau_2$ an order of magnitude smaller than $\tau_1$. Typical values that were used - $\tau_1 = 0.3$ and $\tau_2 = 0.01$ - seemed to work well for all datasets giving sufficient number of clusters (reducing $\tau_1$ and $\tau_2$ further did not change the number of clusters). After the training process, it was possible to visualise the clusters that formed during the learning process using by SDH toolbox to determine those which represent the data. An example for visualisation of the clusters obtained is illustrated in Figure 6.15.

![Hierarchical cluster visualisation of SOMs for Latera data using GHSOM and SDH toolboxes in MATLAB, showing SOM layers 0 (containing 6 units) and 1 (containing 6 SOMs expanded from layer 0).](image)

**Figure 6.15:** Hierarchical cluster visualisation of SOMs for Latera data using GHSOM and SDH toolboxes in MATLAB, showing SOM layers 0 (containing 6 units) and 1 (containing 6 SOMs expanded from layer 0).
Each SOM has a gridded representation, where a grid block is a unit (or neuron) designated by a codebook vector, e.g. layer 0 SOM is made up of six units, each a vector of length \( N = 63 \). Each unit acts as a parent unit for six SOMs belonging to layer 1. The contour display of cluster indicates density values calculated using equation (6.10). The peak of the contour which corresponds to the mode vector is the cluster centre, therefore, those units containing peaks are chosen from the SOM grids to represent a cluster. Cluster centres were taken from the lowest layer in the hierarchy, i.e. from layer 1 SOMs containing 7 clusters. However, only 5 out of 7 clusters centres, marked with a red ‘X’ mark in Figure 6.15, corresponding to ‘unique’ reflectance signatures as illustrated in Figure 6.16., were used. Using these signatures, fuzzy classification was subsequently performed using the \( A \)-norm distance measure from equation (4.24), to obtain 5 fuzzy classification maps.

![Cluster signatures](image)

Figure 6.16: Cluster centres (reflectance signatures) from layer 0 units 1, 2, 4 and 6, for different land cover types in the Latera study area: (a) dense and healthy vegetation; (b) sparse and healthy vegetation and soil mixture; (c) sparse and stressed vegetation and soil (containing mineral alteration) mixture; (d) road network; (e) other unidentified man-made structures.

One of these maps is a probability map giving the likelihood of \( CO_2 \) leakage on the surface, i.e. given that leakage has occurred the likelihood is the probability distribution for the surface effects, such as vegetation stress or mineral alterations.
To choose the most appropriate map, the corresponding cluster centre (reflectance signature) was checked to see whether it contained specific diagnostic features pertaining to these effects. The spectral information contained in the bands of the hyperspectral data ranges from the visible to the Near Infra-Red (NIR) region of the electromagnetic spectrum. This range holds two potentially important diagnostic features of surface effects caused by CO₂ leakage that were used.

**Shifting of the red edge**: This is an important indicator of early stress in vegetation, proposed by Horler *et al.* (1983), based on the amount of shift in the peak value of the first derivative of reflectance spectra, between 0.55 and 0.75 μm, towards shorter wavelengths. The most important characteristic of this indicator is that it is strongly related to the chlorophyll content, which is in direct agreement with the two-parameter extension of the Beer-Lambert Law, accounting for both reflectance and scattering. Although this is the single most important factor, leaf maturity and Leaf Area Index (LAI) also contribute towards the reflectance properties in the Red region. Nevertheless, it is insensitive to the amount of vegetation cover and distinguishes between stressed vegetation and bare soil, in contrast to band ratioing methods, such as Normalised Difference Vegetation Index (NDVI). Recently, Smith *et al.* (2004) applied Red edge shift for the detection of Natural gas leaks from underground pipelines, and this is used here as a diagnostic feature for CO₂ leakage effects on vegetation.

**Red bleaching**: This occurs when Iron (III) (Ferric) Oxide (e.g. Haematite) in the soils is reduced due to the formation of an acidic environment, such as those created by hydrocarbon seepages (Schumacher, 1996), while CO₂ may also cause this alteration (van der Meer *et al.*, 2002). Ferric iron exhibits an absorption feature near 0.9 μm, and the spectrum falls off sharply towards the visible region, with a maximum at around 0.8 μm. This diagnostic feature was used here to locate bare soil regions that have undergone bleaching induced by CO₂ leakage.

Based on the above knowledge on diagnostic features, it can be seen in Figure 6.16 that 'cluster 3' signature which corresponds to a mixture of sparse and stressed vegetation, and altered soil is the most appropriate to represent the likelihood of CO₂ leakage because:
the high reflectance region between 0.68 and 0.75 μm, when compared to ‘cluster 1’ for dense and healthy vegetation, shows that the Red edge has clearly shifted towards the lower wavelength indicating constantly stressed vegetation; and

around 0.9 μm, a shallow dip in reflectance is observed with a gradual decrease in reflectance towards lower wavelengths, indicating reduction of ferric oxide, when compared to ‘cluster 2’ for sparse and healthy vegetation, and unaltered soils mixture where there is a sharp decrease in reflectance between 0.7 and 0.8 μm.

Thus the fuzzy classification map corresponding to ‘cluster 3’ was chosen to represent the likelihood distribution, shown in Figure 6.17. The patches of high values (close to 1) correspond to pixels whose spectra correspond to sparse and stressed vegetation, and altered soil mixture. This is owing to the fact that even though a dataset of high resolution (1 m) was used, the pixel size is large enough to contain both of these effects encapsulated by the cluster 3 signature. The fuzzy map provides the evidence in the methodology to generate a posterior detection from the prior detection results, which will be illustrated in the following chapter.

Figure 6.17: Fuzzy classification map representing the likelihood of CO₂ leakage on the surface for Latera ‘Study Area 1’, with black boxes indicating high CO₂ flux regions.
The same steps were followed to generate the likelihood distributions for ‘Study Area 2’ using the hyperspectral and multispectral datasets, illustrated in Figure 6.18.

The difference in the results using the hyperspectral and multispectral data is clearly seen owing to: lower resolution of the multispectral data of 2.5m in comparison to the hyperspectral data of 1m, and the presence of the thermal band (8.5-13 μm) in the multispectral data. The spectral resolution of the multispectral data allowed the selection of the map based only on the diagnostic feature related to vegetation stress in the signature, and not mineral alterations.

Figure 6.18: Fuzzy classification maps representing the likelihood of CO₂ leakage on the surface for Latera ‘Study Area 2’, with black boxes indicating high CO₂ flux regions (a) using Hyperspectral data; (b) using multispectral data.
6.3.2 Likelihood Generation using the Laacher See Data

Using GHSOM, it was found that the Laacher See data contained only two unique spectral signatures although 16 clusters were formed during the learning process, as illustrated in Figure 6.19. Many of these clusters are formed from the common contour peak overlapping units 3, 4, 5, 6, 7 and 8 in layer 0. This suggests that a major fraction of the total number of pixels in the image most likely belong to a single feature class, i.e. healthy vegetation in this case, unlike the earlier case in Figure 6.15. The remaining pixels belonging to units 1 and 2 is seen to represent the class of stressed vegetation.

The clusters marked with a red ‘X’ have corresponding codebook vectors with spectral characteristics for healthy and stressed vegetation, as shown in Figure 6.20. Based on these signatures, fuzzy analysis was performed to obtain the likelihood map for potential CO₂ leakage in Figure 6.21. Most of the stressed vegetation areas are concentrated in the region close to the shore of the Laacher See.

Figure 6.19: Hierarchical cluster visualisation of SOMs for Laacher See data using GHSOM and SDH toolboxes in MATLAB, showing SOM layers 0 (centre) and 1 (surrounding).
Figure 6.20: Cluster centres from Layer 0 units 2 and 8 for (a) stressed vegetation; (b) healthy vegetation in the Laacher See study area.

Figure 6.21: (a) Image constructed by kriging interpolation of mobile laser data overlaid on an areal photograph (After Jones et al. (2009)); (b) Fuzzy classification map representing the likelihood of CO₂ leakage on the surface for Laacher See hyperspectral data.

6.4 Conclusions

As the unsupervised detection of weak anomalous pixels such as CO₂ leakages on the surface is a difficult problem, the methodology developed starts by locating all types of surface anomalies without any prior knowledge. The geostatistical theory of Intrinsic Random Functions (IRF) allows modelling the spatial correlations in the pre-processed data and subsequently filtering out the low frequency long range correlations. Assessment of the neighbourhood size for the ALC-0 kriging filter
shows that a small neighbourhood of size 5 x 5 is sufficient since larger neighbourhoods not only increases computations, but it also over-smooths the high frequency features (short range correlations) in the image that are most interesting for the purpose of leakage detection.

Performing Independent Component Analysis (ICA) on the residual subspace in order to iteratively find non-Gaussian components once again increases the cost of computations, particularly for large images. In this regard, image downsampling according to the Nyquist-Shannon sampling theorem was found useful. However, ICA gives a pool of non-Gaussian and sparse images from which only one is a useful result for the prior detection of surface anomalies. Statistical measures such as skewness and the Brown-Forsyth’s F-statistic were found to be consistently useful for the selection, and no specific relation was found between them.

In addition to prior detection, it was also necessary to extract maps from the data that provide suitable evidence for leakages on the surface. Fuzzy analysis using the Growing Hierarchical Self-Organising Maps (GHSOM) and cluster visualisation using Smoothed Data Histograms (SDH) facilitated the generation of a probability map from the data whose dimensions are not reduced. This map represents the likelihood of CO₂ leakages based on diagnostic features of its indirect effects on the surface like vegetation stress and mineral alterations at the Latera and Laacher See study sites. The advantage of this approach when compared to the use of traditional measures, such as the Normalised Difference Vegetation Index (NDVI) for mapping healthy and stressed vegetation, is that it automatically identifies feature classes and utilises their spectral signature to provide a relatively more accurate discrimination between them, with an uncertainty measure. When visually comparing the prior detections and likelihood maps, there appears to be matches in some high confidence regions. Hence combining these maps was considered as the next stage in the methodology.
Chapter 7  Information Fusion for Posterior Detection

7.1 Introduction

The second stage of the methodology aims at improving the prior detection results obtained in Chapter 6 by incorporating the information provided by the fuzzy maps regarding the likelihood of CO$_2$ leakages on the surface. This stage generates a confidence map as the posterior detection of leakages using information fusion ideas borrowed from the Dempster-Shafer (DS) theory of evidence combination.

The fusion is performed by representing each of the maps as Probability Mass Functions (PMF) and applying the DS combination rule in order to estimate the Basic Probability Assignment (BPA) per image pixel, with an uncertainty range [belief, plausibility]. Those pixels with low BPA values and low uncertainty are converted as high confidence measures using the entropy function from information theory, indicating leakage anomalies as posterior detection.

Validation of the confidence measures is also demonstrated using the Receiver Operating Characteristics (ROC) curves with the available field measurement data, thereby allowing a comparison between the proposed methodology and another anomaly detection technique by Goovaerts et al. (2005) which is also based on geostatistical methods. This technique aims to filter unwanted spatial information
from hyperspectral data with applications for the detection of: artificial targets with distinct boundaries and dimensions, and multiple targets of various shapes and sizes in a mine tailings site that has a highly complex landscape. The filter is based on the kriging of local means (Goovaerts, 1997) from the Principal Components (PCs) of the data. The number of PCs that are chosen is generally a small subset of the whole data, depending on their eigenvalues. Surface anomalies are subsequently detected by kernel (local) based correlation methods using Reed-Xiaoli (RX) statistics (Reed and Yu, 1990) and Local Indicator of Spatial Autocorrelation (LISA) statistics, such as the Moran's I (Anselin, 1995).

7.2 Posterior Detection using DS Theory

The DS theory was applied to mainly obtain an improved posterior detection. To achieve this, the PMFs of the prior detection result (ICA values) and the corresponding likelihood map for CO₂ leakages (fuzzy values) on the surface were estimated using a 1D kernel-based density estimation technique (Silverman, 1998; Botev, 2005), given by:

\[
k(x; \theta; \sigma^2) = p(x) \sum_{i=0}^\infty e^{-\frac{1}{2} \sigma^2 \phi_i(x) \phi_i(\theta)}
\]

(7.1)

\[
g(x) = \frac{1}{N} \sum_{i=1}^N k(x; \theta_i; \sigma^2)
\]

where, \( p(x) \) is the empirical density function for an observed \( N \) point neighbourhood, \( \theta \) is the kernel mean, \( \phi \) is an eigenfunction, \( \lambda \) is the corresponding eigenvalue, \( \sigma^2 \) is the kernel variance, and \( g(x) \) is the estimated target density function.

In order to assign a confidence measure for each pixel, the entropy, an information theoretic function, was calculated using the pixel's plausibility value obtained after applying the DS combination rule between the two mass functions. In general, for a set of \( i \) events, the average entropy function is given by (Shannon, 1948):

\[
H = -\sum_{i=1}^N p_i \log_x (p_i)
\]  

(7.2)
where, \( p_i \) is a probability value for an event \( i \), \( x \) is the base of the logarithmic function that determines the unit of entropy – for \( x = 2 \), unit is Bits; for \( x = 10 \), unit is Hartley, and \( x = e \), unit is Nats. In the context of DS theory, Yager (1983) called this function as \textit{dissonance}, where \( p_i \) is the plausibility value, and \( x = 2 \).

7.2.1 Posterior Detection and Validation using Latera Data

Using ‘Study Area 1’ data, the PMFs for the prior detection and likelihood maps in Figure 6.4 and Figure 6.17 respectively (see chapter 6) were first estimated using the kernel density estimation technique according to equation (7.1). Each pixel is now associated with two probability mass values, one estimated from the prior detection, say \( m_1 \), and the other estimated from the likelihood, say \( m_2 \). The DS combination rule was then applied using equation (4.36) to obtain the plausibility value, such that:

\[
\text{plausibility} = \frac{m_1 m_2}{1 - [m_1(1 - m_2) + (1 - m_1) m_2]} \tag{7.3}
\]

The reason for the choice of DS combination rule can be explained by the confusion matrix drawn between the \textit{belief} and \textit{plausibility} functions, shown in Figure 7.1. If \( A \) is the evidential set that corresponds to the probability density of anomalous pixels found in the prior detection, and \( B \) is the evidential set that corresponds to the probability density of anomalous pixels found in the likelihood of \( \text{CO}_2 \) leakages on the surface, the \([\text{belief, plausibility}]\) range of the combination can take three possible cases (the fourth case is not possible because \textit{belief} < \textit{plausibility}).

\[
\begin{array}{c|cc}
\text{Bel}(A, B) & \text{Low} & \text{High} \\
\hline
\text{Low} & [\text{Low, Low}] & [\text{Low, High}] \\
\text{High} & [\text{High, Low}] & \text{Impossible} & [\text{High, High}] \\
\end{array}
\]

\text{Figure 7.1: Confusion matrix for BPA values range considering an example pair of evidences A and B for DS combination.}
Low values are assumed between 0 and 0.5, while high values between 0.51 and 1. If the belief values are considered \( (m_1, m_2) \), it can be seen from the confusion matrix that it can take a low value in \( 2/3^{rd} \) of the cases, where as the plausibility value (equation (7.3)) is low only in \( 1/3^{rd} \) of the cases. This implies that if the entropy function (dissonance) is calculated for plausibility, using equation (7.2), only those pixels that belong to the top left quadrant of the confusion matrix are assigned a high confidence, i.e. pixels that are both anomalies in the prior detection and likelihood maps and having a low uncertainty range between belief and plausibility \( (P(A,B) - Bel(A,B)) \). Thus the posterior detection map is obtained for ‘Study area 1’ using this idea, as shown in Figure 7.2. Dissonance is linearly rescaled to lie between 0 and 100.

![Dissonance Measure](image)

**Figure 7.2:** Posterior detection result obtained for the Latera ‘Study Area 1’ hyperspectral data after applying the DS theory of evidence combination using the prior detection and likelihood maps, superimposed on artificially illuminated elevation data (black patches indicating shadows), with black boxes indicating high CO\(_2\) flux regions.

The result shows an increase in confidence for the regions that correspond to high CO\(_2\) flux regions, marked as black boxes, when compared to the prior detection result in Figure 6.4 (see chapter 6). In order to make a further comparison, a state-of-the-art geostatistical technique for the detection of surface anomalies proposed by
Goovaerts et al. (2005), which is also based on geostatistical filtering, was applied with the RX statistics. This technique was chosen for comparison in this thesis owing to its similarity with the proposed filtering methodology in terms of unsupervised detection. Figure 7.3 shows the result obtained after for 'Study Area 1'.

Figure 7.3: Detection result obtained for the Latera 'Study Area 1' hyperspectral data after applying the filter proposed by Goovaerts et al. (2005) with the RX statistics, superimposed on artificially illuminated elevation data (black patches indicating shadows), with black boxes indicating high CO₂ flux regions.

Visual comparison clearly shows the difference between the detection results of the methodology developed in this thesis and the filtering technique by Goovaerts et al. (2005), which has mostly no detections, or very weak detections. Two possible reasons exist for this observation:

- The use of global spatial statistics through Intrinsic Random Functions of order-k (IRF-k) and Independent Component Analysis (ICA), followed by the fusion between prior detection and likelihood maps generates posterior detection of leakage anomalies independent of their size or shape. It makes the proposed methodology relatively more unsupervised when compared to...
that of Goovaerts et al. (2005), where different kernel sizes need to be used depending on the scale of the anomalies that need to be detected.

- The posterior detection by fusion of spectral information of CO₂ leakages improves the result of pure statistical analysis, which the methodology of Goovaerts et al. (2005) fails to do.

In comparison to the results obtained by earlier conventional (supervised) approaches (e.g. Bateson et al., 2008; see Figure 3.13, chapter 3), the methodology has detected all known vents in addition to other possible high CO₂ flux regions which could not be verified due to insufficient field measurement data. If these are truly regions containing vents, it is may be concluded that the proposed unsupervised methodology has outperformed conventional methods both in terms of its rapidness in data processing and ability to detect the missed vents (false negatives).

ROC curves were used in order to quantify the difference in the detection results in Figure 7.2 and Figure 7.3. Validation was performed using ground measurement data that was taken along a transect line in the study area (see Figure 5.2, chapter 5). The observation line contains 108 flux measurements, and those observations with flux value greater than 200 g/m²/day were considered anomalous for plotting the ROC curves. The ROC plot obtained is shown in Figure 7.4.

![ROC curves comparison](image)

**Figure 7.4:** ROC curves comparison between the methodology of Goovaerts et al. (2005) and the proposed methodology for Latera 'Study Area 1' hyperspectral data.
The Area Under the Receiver Operating Characteristics curve (AUROC) was used as the performance measure. The AUROC values obtained are 0.5955 and 0.8758 respectively for the methodology of Goovaerts et al. (2005) and the proposed methodology. The low AUROC value of the former demonstrates that its performance is close to that of random detection where the chances of a pixel being both anomalous and non-anomalous are equally likely. Hence, such AUROC figures indicate that the detection result is generally poor. On the other hand, a good detection performance is obtained for the proposed methodology.

For ‘Study Area 2’ the same step was repeated for the prior detection and likelihood maps using the hyperspectral data in Figure 6.7 and Figure 6.18 (a) (see chapter 6) respectively, using DS theory. The posterior detection result is shown in Figure 7.5 (a) along with that obtained using the methodology of Goovaerts et al. (2005) for comparison (Figure 7.5 (b)).

The differences in detection between the two results obtained are visually clear. For the proposed methodology, generally low confidence values are assigned to features such as vegetation and road networks. When compared to its prior detection result in Figure 6.7 (see chapter 6), it is seen that there is an improvement by the elimination of such false detections in the posterior. All the known high CO₂ flux regions are also assigned with high confidence.

Validation of the detection results is made using the ROC curves, as shown in Figure 7.6. The AUROC measures obtained for Goovaerts et al. (2005) and the proposed methodologies are now 0.5217 and 0.8238 respectively. These values suggest that the change in the area of analysis does not have a significant effect on the performance of the proposed methodology.
Figure 7.5: Detection result obtained for the Latera ‘Study Area 2’ hyperspectral data (a) after applying the DS theory of evidence combination using the prior detection and likelihood maps, with black boxes indicating high CO₂ flux regions (b) after applying the methodology of Goovaerts et al. (2005), with black boxes indicating high CO₂ flux regions.

In comparison to the results obtained through earlier conventional (supervised) approaches (e.g. Bateson et al., 2008; see Figure 3.13, chapter 3), there are regions of high confidence in Figure 7.5 (a) that could be other potential CO₂ vents, missed by earlier approaches. However, this fact can only be confirmed with additional field measurement data.
Figure 7.6: ROC curves comparison between the methodology of Goovaerts et al. (2005) and the proposed methodology for Latera ‘Study Area 2’ hyperspectral data.

Similar posterior analysis was also carried out for the ‘Study Area 2’ multispectral data using the prior detection and likelihood maps in Figure 6.10 and Figure 6.18 (b). The posterior detection result obtained in Figure 7.7 also shows that all known regions with high CO$_2$ flux are assigned with high confidence measure. When the ROC curve was plotted for validation along with that of hyperspectral detection result, shown in Figure 7.8, it was found that AUROC measures of 0.8107 and 0.8238 were obtained for detections using the multispectral and hyperspectral data respectively.

These AUROC values suggest that reduced spectral resolution does not significantly change the detection performance of the proposed methodology. An important reason for this observation is due to the availability of limited field measurement data (along a transect line) that is used for validation. Another reason for this may also be based on the availability of spectral bands in the data that represent specific regions of the electromagnetic spectrum where diagnostic features related to the effects of CO$_2$ leakage can be found. For instance, the multispectral data contains a thermal band (8.5-13 μm) which is absent in the hyperspectral data. This band shows high emittance values for regions where high CO$_2$ flux is present mainly due to the presence of bare soil/stressed vegetation region around the vents. Given clear sky conditions, these land covers attain a higher temperature during the day when compared to the adjacent healthy vegetation, and hence behaves as a high emittance
thermal anomaly. However, visually, this result shows a possible increase in false detections owing to decreasing spatial and spectral resolution (2.5m and 11 bands) in comparison to the hyperspectral dataset (1m and 63 bands).

![Detection result obtained for the Latera 'Study Area 2' multispectral data after applying the DS theory of evidence combination using the prior detection and likelihood maps, with black boxes indicating high CO₂ flux regions.](image)

**Figure 7.7:** Detection result obtained for the Latera 'Study Area 2' multispectral data after applying the DS theory of evidence combination using the prior detection and likelihood maps, with black boxes indicating high CO₂ flux regions.

![ROC curves comparison for the methodology using Latera 'Study Area 2' hyperspectral and multispectral datasets.](image)

**Figure 7.8:** ROC curves comparison for the methodology using Latera 'Study Area 2' hyperspectral and multispectral datasets.
7.2.2 Posterior Detection and Validation using Laacher See Data
Following the steps as explained earlier for the case of Latera datasets, DS theory was applied per-pixel on the prior detection and likelihood maps obtained using the hyperspectral data for Laacher See study area in Figure 6.13 and Figure 6.21, after estimating their respective PMFs. The posterior detection map obtained by information fusion in Figure 7.9 (b) shows high confidence in the regions close to the lake shore. This result was visually compared with the field result acquired using the mobile laser technique (Figure 7.9 (a)) by Jones et al. (2009), and there appears to be a generally positive spatial correlation. ROC analysis could not be performed for this result as raw field data were not available.

![Image of posterior detection map](image)

**Figure 7.9:** (a) Image constructed by kriging interpolation of mobile laser data overlaid on an areal photograph (After Jones et al. (2009)); Detection result obtained for the Laacher See study area hyperspectral data (b) after applying the DS theory of evidence combination using the prior detection and likelihood maps; (c) after applying the methodology of Goovaerts et al. (2005).
However, additional comparison is made with the result obtained after applying the
detection methodology of Goovaerts et al. (2005) for this study area (Figure 7.9 (c)).
Despite the detections being weak, it appears that their methodology is able to assign
a relatively higher confidence for regions where high CO₂ concentration are found in
comparison to the background, however, boundaries between land covers (ecotones)
are also assigned with the same level of confidence as leakages. Based on this result,
and earlier results, obtained from the Latera dataset, it may be concluded that the use
of higher order PCs in their methodology does not account for the part of the data
variance that is represented by weak surface anomalies, and hence leading to either
no detections or false detections; therefore lower order PCs maybe more useful for
the detection of leakages.

7.3 Conclusions

Combining synergistic information from multiple sources is useful to extract features
in a particular scene under study that are not otherwise evident from individual data.
This fact has been demonstrated by the use of the Dempster-Shafer (DS) theory of
evidence combination for the detection of CO₂ leakages in Latera and Laacher See
study sites. The information fusion stage of the methodology has increased the
confidence values of those pixels in the image that are related to CO₂ leakages from
prior to posterior by incorporating the likelihood of leakage based on its indirect
effects on the surface, and also by reducing confidence in those regions that are not
leakages, hence reducing false detections.

While comparing the results obtained by using the proposed methodology with those
of the geostatistical anomaly detection technique suggested by Goovaerts et al.
(2005), it was found that the latter performed poorly in terms of detecting the known
CO₂ vents in the Latera and Laacher See study sites, where in all the cases it
performs random detection. On the other hand, the proposed methodology has been
able to assign high confidence to known vents, although there may be some false
high confidence values (false positives). The validation was performed by using
limited raw field measurement data with Receiver Operating Characteristics (ROC)
curves, and the results appear to be largely consistent with visual comparison.

On the other hand, in comparison to the results obtained through conventional
(supervised) approaches, e.g. Bateson et al. (2008), the methodology has detected all
the vents in addition to some others which could not be verified due to the lack of comprehensive raw field measurement data. It is possible that some of these high confidence regions are truly high CO₂ flux regions, in which case the proposed methodology has clearly outperformed conventional approaches owing to its rapidness in data processing and leakage detection. However, for lower resolution multispectral data the proposed methodology is seen to be giving higher false positive rates.

Although not assessed in terms of data distributions, the quality of the results obtained through the proposed methodology may depend on the shape of the Probability Mass Functions (PMFs) of the prior detection and likelihood maps. Best results are obtained if both PMFs are purely one sided distributions, such as an F-distribution (with unit degrees of freedom), Chi-square distribution (with unit degree of freedom), Gamma distribution (with α ≤ 1), or a Weibull distribution (with β ≤ 1), ensuring that the false detections are minimised using the DS theory.

With regards to the effect of spectral resolution on detection performance, it is observed that both hyperspectral and multispectral data performed equally well for the transect line measurement data, shown by their respective Area Under the Receiver Operating Characteristic curve (AUROC) measures. However, it is generally believed that lower spectral resolution of a dataset may be compensated by information that it provides on leakage anomalies in regions of the electromagnetic spectrum, such as the Thermal Infra-Red (TIR), that may not be covered by a hyperspectral dataset. This also implies that in the future if hyperspectral/multispectral sensors are designed that provide information in regions where relevant absorption features for CO₂ gas are present, especially in the TIR wavelengths, along with those diagnostic features in the Near Infra-red (NIR) wavelengths for the effects of leakage on the surface, the statistical image analysis presented in this thesis, with further improvements for reducing false detections, could potentially provide a means to both rapidly detect and quantify leakages.

Analysis on the effect of decreasing spatial and spectral resolutions on the performance of the proposed methodology is a subject matter addressed in the following chapter.
8.1 Introduction

The unsupervised CO₂ leakage detection methodology presented in the previous Chapters was primarily developed and applied to high resolution airborne data. However, regular on-demand acquisition of airborne data using helicopters or aircraft to monitor surfaces above CO₂ storage sites may become laborious, time consuming and expensive. It has been pointed out in Chapter 3 that there exist a number of spaceborne optical remote sensing platforms that are currently operational in the Earth’s orbit. The sensors on-board such platforms acquire data with relatively better temporal resolution (with repeat cycles of at least 16 days), moderate spatial resolution (as high as 15m), and having a much wider spatial coverage. Their applicability for monitoring storage sites warrants investigation.

In order to evaluate the performance of spaceborne optical remote sensing data for the detection of potential CO₂ leakages, an equivalent test was carried out using the available high resolution hyperspectral data for the Latera study site by systematically degrading its spatial and spectral resolutions, and generating synthetic datasets. The methodology developed in this research was applied to each synthetic dataset, and the corresponding Area Under the Receiver Operating Characteristics
curve (AUROC) measure estimated. The trend in the AUROC values reveals that low spatial resolution data may well be applicable for leakage detection.

Furthermore, a case study was also performed using the methodology on a real, low resolution multispectral dataset for Latera acquired by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor onboard NASA’s Terra satellite. This dataset contains bands with different spatial resolutions in the visible (15m), Near Infra-Red (NIR) (30m) and Thermal Infra-Red (TIR) (90m) regions of the electromagnetic spectrum. Since these images with different resolutions pose an initial problem for co-processing them using the unsupervised detection methodology developed, a slight modification was made in the procedure using the Nyquist-Shannon sampling theorem.

### 8.2 Variation of Detection Performance with Spatial and Spectral Resolutions

Sensors convert the upwelling radiance from the Earth’s surface and the atmosphere into a digital image that is an approximation of the true spatial radiance distribution. Although several transformations/degradations occur during the image acquisition process, changing the radiometric (spectral), spatial, and geometric properties, it is the spatial factor that is of interest here. In practice, a sensor can never measure a physical signal $i$ (the at-sensor radiance) with infinite precision, and some form of spatial averaging takes place. The measured output radiance signal $o$ is expressed as a linear convolution function, given by Schowengerdt (2007):

$$o(x, y) = \int_{\alpha_{\min}}^{\alpha_{\max}} \int_{\beta_{\min}}^{\beta_{\max}} i(\alpha, \beta) h(x - \alpha, y - \beta) d\alpha d\beta$$  \hspace{1cm} (8.1)

where, $h$ is the sensor response function called the Point Spread Function (PSF), and $(\alpha_{\min}, \alpha_{\max})$ and $(\beta_{\min}, \beta_{\max})$ are the extents of weighting influence of the PSF, along the x and y directions respectively, that defines the spatial resolution of the acquired image. The PSF is a separable function, i.e. it can be represented as a product of two 1D functions for the cross-track (x) and in-track (y) directions of the sensor, given by:
\[ h(x, y) = h(x) h(y) \]  \hspace{1cm} (8.2)

A common separable function model that is used to represent the PSF is the 2D Gaussian function (Kaiser and Schneider, 2008), given by:

\[ h(x, y) = \frac{1}{2\pi ab} e^{-\frac{x^2}{2a^2} - \frac{y^2}{2b^2}} \]  \hspace{1cm} (8.3)

where, \( a^2 \) and \( b^2 \) are the net variances along the x and y directions respectively, contributed by the sensor's optical system, aperture area, motion and electronics. For a commercially available pushbroom CCD-line-scanner, the Leica ADS40 system, the approximate relationship between net variance and the pixel size (\( \delta_{\text{pixel}} \)) (Reulke \textit{et al.}, 2006) is given by:

\[ \sigma \approx 0.75\delta_{\text{pixel}} \]  \hspace{1cm} (8.4)

where, the net variance is assumed to be equal in both directions, i.e. \( \sigma^2 = a^2 = b^2 \).

The Leica ADS40 system forms a part of the AISA Eagle 1K sensor which was used to acquire the hyperspectral data for the Latera study site.

8.2.1 Degradation of the Spatial and Spectral Resolutions of Latera Hyperspectral data

In order to degrade the spatial resolution of the Latera hyperspectral data (from 1m), using the Gaussian PSF function, equation (8.4) was first used to estimate the corresponding variances at different pixel resolutions at which the data was required to be degraded to, as shown in Table 8.1.

\[
\begin{array}{c|c|c}
\delta_{\text{pixel}} & 2m & 4m & 8m & 16m \\
\hline
\sigma^2 & 2.25 & 9 & 36 & 144 \\
\end{array}
\]

Table 8.1: Desired pixel resolutions and their corresponding PSF variances.
These variance values were used to form Gaussian PSFs for different spatial resolutions at 2m, 4m, 8m and 16m, using equation (8.3). By convolution of the PSFs with the high resolution image, as in equation (8.1), spatially degraded images were obtained at the desired resolutions. Figure 8.1 illustrates spatial resolution degradation of the Latera hyperspectral data (for Band 1 = 0.402 μm).

![Spatial resolution degradation using Band 1 of the Latera hyperspectral data for (a) 1m (original resolution); (b) 2m; (c) 4m; (d) 8m; (e) 16m.](image)

Likewise, the spectral resolution was also simultaneously degraded from 63 bands, by equally weighted averaging of the bands, to obtain resolutions of 32 and 16 bands, all ranging from the visible to NIR regions. Equal weighting was assumed because it represents a proportional increase in the Full Width at Half Maximum (FWHM) value of the bands. It serves as a generic case of multispectral data for the purpose of understanding the effects on detection performance by jointly degrading spatial and spectral resolutions. However, it does not represent the spectral averaging that takes place in the ASTER data, where additional effects, such as the spectral
**bleed-through** (or **cross-talk**), are involved. Using the resolution degradation step, a total of 14 synthetic datasets were obtained for different spatial resolutions, at 1 m, 2 m, 4 m, 8 m and 16 m, and containing 63, 32 and 16 bands each.

### 8.2.2 CO₂ Leakage Detection using the Synthetic Data Cubes

The unsupervised detection methodology developed was applied to the synthetic datasets. Figure 8.2 illustrates the posterior detection results obtained for 1m resolution data containing 63, 32 and 16 bands.

By visual comparison, it generally appears that the degradation of spectral resolution does not significantly degrade the detection performance. In fact, validation of the results with transect line measurement data using ROC curves indicates that detection with the synthetic multispectral data (16 bands) showed a slightly better performance in comparison to other datasets, as seen in Figure 8.3. The AUROC values are 0.8238, 0.8288 and 0.9281 for data containing 63 bands, 32 bands and 16 bands data respectively. This is because spectral averaging of the bands has reduced the effects of anomalous spectral behaviour that is not related to CO₂ leakages.

The averaging method has an indirect connection with the spectral range that the dataset covers. If diagnostic features related to CO₂ leakage and its surface effects lie outside this range, it affects the detection result using multispectral data. In this study, the synthetic multispectral data was created by equal weighted averaging of continuous bands in the hyperspectral data (63 bands). Spectral averaging with a Gaussian function, which reduces the FWHM of the data bands, could increase or decrease the detection performance depending on the location of the centre wavelength.

Hence, both the choice of centre band (with a higher weighting) and FWHM are important factors to be considered in multispectral sensor design for detecting leakages. This suggests that future sensors designs should consider assigning a higher weighting to those wavelengths in the NIR and TIR regions that are useful for CO₂ leakage detection and quantification. This design consideration can provide datasets with the potential to reduce false detections.
Figure 8.2: Detection results using the Latera synthetic datasets at 1m resolution for (a) 63 bands; (b) 32 bands; (c) 16 bands; black boxes indicate high CO$_2$ flux regions.
Detection of Surface Leakages using Low Spatial Resolution Optical Remote Sensing

Figure 8.3: ROC curves comparison for the methodology by using the synthetic datasets with 1m resolutions, containing 63, 32 and 16 bands.

Figure 8.4 illustrates the posterior detection results obtained for synthetic datasets containing 63 bands with 1m, 2m, 4m, 8m and 16m resolutions. There is an overall degradation of performance with reduction in the spatial resolution. However, it is seen that the known gas vents marked by the black boxes are generally assigned with high confidence, with an increase in false detections with decreasing resolution. The ROC curves were also plotted for these results, as shown in Figure 8.5.

The AUROC values obtained from these curves were 0.8238, 0.7110, 0.6434, 0.6381 and 0.5880 for 1m, 2m, 4m, 8m and 16m respectively. These values indicate a decreasing trend in detection performance as expected for decreasing spatial resolutions.
Figure 8.4: Detection results using the Latera synthetic datasets containing 63 bands for (a) 1m; (b) 2m; (c) 4m; (d) 8m; (e) 16m; black boxes indicate high CO₂ flux regions.

Figure 8.5: ROC curves comparison for the methodology by using the synthetic datasets containing 63 bands at 1m, 2m, 4m, 8m and 16m resolution.
8.2.3 Comparison of Relative AUROC Values

The AUROC values were similarly estimated for all 15 synthetic datasets in order to understand the general trend in performance of the proposed methodology. Table 8.2 lists the AUROC values obtained for decreasing number of bands at a particular spatial resolution, and for different spatial resolutions.

<table>
<thead>
<tr>
<th>AUROC measure</th>
<th>Spatial resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 m</td>
</tr>
<tr>
<td>63 bands</td>
<td>0.8238</td>
</tr>
<tr>
<td>32 bands</td>
<td>0.8288</td>
</tr>
<tr>
<td>16 bands</td>
<td>0.9231</td>
</tr>
<tr>
<td>Average</td>
<td>0.8586</td>
</tr>
<tr>
<td>Difference between 63 and 16 band cases</td>
<td>0.0993</td>
</tr>
</tbody>
</table>

On inspecting the values, the following observations required explanation.

- The average AUROC values showed an expected decreasing trend as the spatial resolution decreases. This is attributed to the spatial averaging of radiance values by the sensor PSF function.

- At high resolutions, the difference in performance values, between the multispectral (16 bands) and hyperspectral equivalent data (32 and 63 bands), were small. As pointed earlier, this could be attributed to the FWHM of the synthetic multispectral data, which was based on equal spectral weighting. The higher the FWHM, the better the performance that can be achieved with multispectral data. However, this is dependent on the sensor system.

- As the resolution degrades, the difference in performance values becomes more evident. At the lowest spatial resolution of 16 m, where this difference is the highest, it is observed that detection using multispectral data outperforms that of hyperspectral data. This counterintuitive phenomenon can be explained based on the relation between the discriminatory power and Signal to Clutter Ratio (SCR) of the data. Since hyperspectral data has a higher discriminatory power, any given pixel tends to exhibit lower spatial
correlation (compare Figure 8.1 (a) and Figure 8.1 (e)) with its immediate neighbours owing to even smaller differences in their spectral signatures. This leads to unwanted interference (clutter) masking the weaker anomalies (signal), hence reducing the AUROC and SCR for lower resolutions. The converse is true for multispectral data, and hence better performance in detection is observed, hence providing a supporting evidence for the need for new multispectral sensor designs in the future which covers the essential bands in the NIR and TIR regions of the spectrum.

8.3 Processing of Latera ASTER Data

ASTER spaceborne sensor data for the Latera ‘Study Area 3’ (see chapter 5, section 5.2.3) contains different spatial resolution images in different bands, i.e. 15m in the NIR and 30m in the SWIR regions. Therefore, the methodology developed earlier could not be directly applied without having considered an additional step. This step was performed by using the Nyquist-Shannon sampling theorem for signal reconstruction. Since the images were assumed to be bi-partite, i.e. containing low and high frequency components, it is possible to discard alternate pixels in the NIR images (downsampling) after the application of the geostatistical Intrinsic Random Functions (IRF) band pass filter (see chapter 6). The filtered and downsampled NIR bands (to become 30m resolution) were grouped together with those of the filtered SWIR bands to perform the Independent Component Analysis (ICA). This process is illustrated in Figure 8.6.

![Figure 8.6: The proposed filtering scheme for ASTER data to generate prior detection.](image)
Upsampling and interpolation is performed to obtain the prior detection result, as shown in Figure 8.7 (a). The TIR bands (90m) were excluded from the analysis since daytime thermal images posed difficulty in isolating the anomalies in low resolution images. Fuzzy analysis was also performed with the Growing Hierarchical Self-Organising Maps (GHSOM) (see chapter 6) to obtain the fuzzy map for the likelihood of CO₂ leakages, indicating stressed vegetation, as shown in Figure 8.7 (b).

Finally, using Dempster-Shafer (DS) theory of evidence combination (see chapter 7), information fusion was performed to obtain the posterior detection map in Figure 8.8. The black boxes indicating known CO₂ vents in the region have been superimposed on the map. The result generally shows that high confidence areas have a good match with the marked vents. This result demonstrates the there is potential for the unsupervised methodology for detecting leakages using spaceborne data. It is also seen that some of the high confidence areas, particularly in the south and south-east side of the study area where ground measurements were not performed, have also been detected. Although according to the analysis these are related to stressed vegetation, it is an indication of increasing false detection rates. However, owing to the lack of comprehensive ground truthing it is not currently possible to verify these anomalies. Nevertheless, it is believed that they may also represent other potential
leakage points in Latera, which could be validated using ground measurement data in the future.

Figure 8.8: Posterior detection result for Latera 'Study area 3' using ASTER data bands; black boxes indicate high CO₂ flux regions.

8.4 Conclusions

The proposed unsupervised methodology is tested on synthetic datasets that were created by systematic degradation of the high resolution hyperspectral data acquired in Latera. It is generally observed that reducing the spectral resolution does not affect the performance of the methodology for high spatial resolution data. However, this is believed to be an outcome of the equally weighted averaging that was used, hence resulting in synthetic multispectral data bands with a large Full Width Half Maximum (FWHM).

At lower spatial resolutions it is also observed that the performance using multispectral data is better than that of hyperspectral data. Although counterintuitive, it demonstrates that there is potential for the use of satellite imagery for detecting leakages. On the other hand, degrading spatial resolution of the data does affect the performance of the detection process owing to spatial averaging of pixel values.

Results obtained from the analysis using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data for Latera has demonstrated
promise in the application of the unsupervised leakage detection methodology developed in this thesis. However, there are issues related to possible increase in false detections. Minimising such false positives completely remains an important challenge while using low resolution data and may require further improvements in the methodology. Hence this is considered to be a topic for future investigation.

More importantly, analysing the performance of the proposed methodology using synthetic and real ASTER datasets has provided insights related to a potential airborne and spaceborne multispectral CO₂ sensor design for the future. Such a sensor must take into account of all useful wavelengths that correspond to diagnostic features of CO₂ leakages, in the NIR and TIR regions of the spectrum. Hence, the choice of centre wavelengths and their corresponding Full Width at Half Maximum (FWHM), based on the spectral averaging method, are two key factors considered essential for the band formation. The multispectral dataset acquired by such a sensor will also allow for rapid detection of surface leakages with minimum false positives using the methodology proposed in this thesis.
Chapter 9  Conclusions and Recommendations for Further Research

9.1 Achievements and Conclusions from the Research

This PhD research has developed a new unsupervised methodology for detecting CO₂ leakages on terrestrial surfaces using optical remote sensing datasets. It is has been developed based on a combination of computational techniques to perform data processing and thus relieves the task of an image analyst to carry out manual data interpretation that is generally based on a priori field information, as has been done in the case of earlier indirect detection methods.

It was shown that the methodology works well with data acquired for different locations, and is able to detect known leakages with high confidence. It does so by utilising both the spatial and spectral information contained in the data, and not relying on simple measures, such as the Normalised Difference Vegetation Index (NDVI), that make use of information in selected wavelengths and ignore spatial variability. However, it is seen that there is a possible increase in false positive rate with decreasing spatial resolution.

In addition to detecting leakages using high resolution airborne hyperspectral data, the methodology has also been tested with low resolution spaceborne multispectral data that have a much wider coverage and promising results achieved.
Conclusions and Recommendations for Further Research

The map of the results obtained following the application of the methodology developed contains information regarding the potential leakage areas associated with a measure of confidence. Such confidence maps are especially useful as independent information that can save time, cost and effort for field practitioners when designing ground truthing field measurement campaigns using cumbersome instrumentation. For the Latera data acquired by the spaceborne ASTER sensor, the detection result shows that, typically for a confidence measure of 50% or more, the area that needs to be covered for collecting measurement data is only about 0.05175 km², implying a reduction of up to 98% from the total survey area of 2.25 km².

Additionally, this research has provided a number of insights related to processing optical remote sensing data in order to detect potential CO₂ leakages on the surface in an unsupervised manner. The important conclusions that have been drawn from the application of the newly developed methodology on datasets acquired for the Latera, Italy, and the Laacher See, Germany, natural analogue CO₂ leakage sites are considered below.

- Pre-processing of the data cube is essential before unsupervised detection of CO₂ leakages can be carried out using optical remote sensing data. It is seen that noise frequently affects the data and must therefore be filtered out to support better reliability of the leakage detection. The Maximum Noise Fraction (MNF) method works well with noisy data. Additionally, high data dimensionality reduction helps to ease the computations for processing.

- It is shown that the Principal Component Analysis (PCA) method is not very effective in terms of reducing data dimensionality for detecting leakages since the average variance available outside ±2σ Confidence Interval (CI), corresponding to low probability leakage anomalies, is low when compared to an equivalent band grouping approach that is proposed in this thesis. Moreover, PCA complicates the modelling of the spatial correlation between components.

- Filtering the data using the geostatistical Intrinsic Random Functions (IRF) helps to eliminate the low frequency spatial heterogeneities in the images which allows the identification of residuals with short range correlations. The Independent Component Analysis (ICA) enhances the leakage anomalies, along with other possible surface features, which serves as the prior detection.
Conclusions and Recommendations for Further Research

- In order to retain high frequency information in the images the choice of the IRF filter neighbourhood is considered important. Small neighbourhoods (5x5) perform better than large neighbourhoods based on their frequency response in the Fourier domain. This also helps in reducing the processing time for kriging.

- Downsampling the filtered images based on the Nyquist-Shannon sampling criterion ensures that no information is lost and simultaneously provides gains in processing time for the ICA.

- Unsupervised fuzzy analysis of the data using the Growing Hierarchical Self-Organising Maps (GHSOM) helps in automatically identifying the cluster of pixels giving the likelihood probability map for CO2 leakages. This is essential for characterising pixels in the prior detection, without which a leakage confidence cannot be assigned.

- Posterior analysis using the Dempster-Shafer (DS) theory of evidence combination provides a means to reduce false detections by incorporating evidences of surface leakages based on diagnostic features in the spectral signature. Although only one evidence variable was considered, the flexibility of DS theory allows combining many such evidence variables, such as the surface temperature, mineral distribution, and soil pH maps, that can help isolate leakage anomalies with better confidence.

- The Receiver Operating Characteristics (ROC) curve is a validation tool that allows the comparison of results obtained by different types of detection methods, and also results obtained by using the same method for different types of data. Using the ROC, allows the performance of the methodology to be assessed for different spectral and spatial data resolutions, and hence provides insights as to which kind of data may be most useful for the detection of leakages, e.g. it was found that low resolution multispectral data can be used, while low resolution hyperspectral data may be inadequate for leakage detection at the wavelengths studied here.

- The potential of the methodology to detect leakages has been demonstrated for low resolution data acquired by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor onboard NASA’s Terra satellite.
Leakage points have been assigned with high confidence. However, there is an increase in false detections which currently remains a challenge, and points out that improvements in the methodology are necessary.

9.2 Limitations and unanswered questions

Notwithstanding the achievements of the research reported in this thesis, there are a number of limitations and unanswered questions some of which may pave the way for future work in the application of optical remote sensing for monitoring CO₂ storage sites.

9.2.1 Early Detection

The methodology does not provide early detection of leakage as it is solely based on indirect effects. Depending on the vegetation and soil geochemical species present on the surface, the first signs of sub-surface leakage become apparent on the surface at different intervals in time, usually spanning from a few days to several weeks. Hence, the capacity to detect leakages will depend to some extent on the duration of exposure of the surface to high subsurface CO₂ flux and concentration. Under intermittent flux conditions, it is known that vegetation has the capacity to recover from its temporary stressed state, mainly because CO₂ also acts as a fertiliser for plants under low flux and concentration conditions. This also applies to soil mineral alterations which could be reversible under specific conditions of temperature, pH and land use. Lateral movement of the CO₂ plume in the shallow subsurface, called the vadose zone, can also reduce the effective flux of CO₂ at the surface, posing additional detection problems (Oldenburg and Unger, 2003).

Thus it may be necessary to combine surface information from remote sensing platforms with those from other modalities that measure changes in the vadose zone. The CO₂ plume rising from the subsurface, with a higher temperature than that of the surface, could be detected as a thermal anomaly using multispectral and hyperspectral Thermal Infra-Red (TIR) sensors. However, this depends on the time of data acquisition, and the thermal characteristics (heat capacity) of the land cover. Images acquired during the night over land covers with a low heat capacity, combined with direct measurements, has a very good potential for early detection.
9.2.2 Change Detection
Due to the lack of a good time series dataset, the methodology developed does not take into account temporal variation. Studying anomalous patterns within normal patterns over time, such as the presence of constantly stressed vegetation or bare soil within a homogenous healthy vegetation cover, provides an indication for high CO₂ concentration. For taller vegetation, such as forests, it may also be useful to study the changes occurring in the photosynthetic activity of leaves, including gas and water vapour exchange between the canopy and atmosphere, which are normally affected due to the decrease in water and nutrient uptake caused by root asphyxiation (Farrar et al., 1999). These factors contribute to variations in spatial characteristics of an ecosystem, such as ecotones (Holland, 1988) and texture (Coppin et al., 2004), from local to regional and global (mesoscale) scales. In addition to the changes occurring on the surface, accounting for changes in the vadose zone due to varying CO₂ concentrations, such as the electric properties of consolidated soil, such as the dielectric permittivity (Guéguen and Palciauskas, 1994), can potentially provide an early alarm for leakage detection and remediation.

9.2.3 Leakage Quantification
The methodology developed does not provide an estimate of the amount, flux or concentration associated with a leakage anomaly. According to the literature, it is possible to quantify leakage using either the reflective (visible to Short Wave Infra-Red (SWIR)) or the emissive regions (Thermal Infra-Red (TIR)) of the electromagnetic spectrum. CO₂ gas is known to have absorption features in the SWIR at 1.4 μm, 1.6 μm, and 2 μm. However, a reasonable retrieval of CO₂ using typical hyperspectral sensors is difficult because most of the absorption regions are relatively narrow or influenced by other atmospheric gases, such as water vapour. On the other hand, owing to the strong emittance signatures of CO₂, TIR images acquired at night could prove to be relatively more useful for locating heat anomalies on the surface, generated by a high concentration subsurface gas plume, and performing source quantification (Fischer et al., 2009; Coppola et al., 2010).

9.2.4 Masking Factors
The methodology does not work well if the images are contaminated by masking factors, such as shadows and cloud cover, because it is based on processing the data from passive technologies, such as optical remote sensing. Standard anomaly
detection procedures based on radiance equalisation only partially address the problem of shadows (Kanaev and Murray-Krezan, 2010). In addition to the obvious spectral differences with well-illuminated regions in the reflective regime, it is intuitive that weakly illuminated regions under shadows typically suffer from poor Signal to Noise Ratio (SNR). However, as noted earlier, it is still possible to detect heat signatures under such conditions. On the contrary, active remote sensing technologies, such as RA dio Detection And Ranging (RADAR), work equally well under all weather and illumination conditions. More specifically, CO₂ has the tendency to change the relative dielectric permittivity of the vadose zone, which in turn manipulates Ground Penetrating Radar (GPR) responses owing to the differences in signal penetration depths. Investigations by Pettinelli et al. (2008) at one of the gas vents in Latera using the GPR reveal that CO₂ from the subsurface is moisture bearing, usually carried from the ground water layer. Hence the core region of the vent, where the CO₂ concentration is the highest, is usually relatively wetter when compared to the outer regions. Since moisture on the surface introduces complex dielectric permittivity (impedence) (Gueguen and Palciauskas, 1994), signal blanking (or zero penetration) occurs at the vent core. However, this effect is not prominent under generally wet weather conditions.

9.2.5 Sea Floor Leakages

The methodology was primarily developed for processing optical remote sensing datasets for monitoring terrestrial storage sites, and hence unsupervised leakage detection from sub-sea reservoirs has not been addressed. The most commonly used technologies for studying the sea floor are: SO und Navigation And Ranging (SONAR) for the deep floor using water vehicles, and airborne L Ight Detection And Ranging (LIDAR) for the shallow floor using aircrafts. Between the two, acoustic monitoring using SONAR is most commonly reported in the literature for various bathymetric applications, and in particular for identifying gas leakages on the sea floor that result in the formation of nearly circular crater-like expressions, called pockmarks, where fluids escape upwards through the fine-grained sediments (Hovland and Judd, 1988). Gas columns, called ‘chimneys’, below the sea floor are seen as anomalies in echosounder images, such as acoustic blanking, acoustic turbidity, enhanced reflections, bright spots, and columnar disturbances (Schroot and Schüttenhelm, 2003). The mechanics of migration of gas and water through the sea
Conclusions and Recommendations for Further Research

Floor sediments are being carefully studied to understand how early detection of pockmarks can be made possible (Cathles et al., 2010). In terms of the quantification of leakages at the sea floor, underwater photo-optical systems mounted on Remotely Operated Vehicles (ROVs) provide useful imagery of gas bubbles from which it is possible to estimate the flux rates using bubble size and shape distribution and rise velocity (Thomanek et al., 2010).

9.3 Recommendations for Future Research

The main focus of this PhD research was spatial analysis of the surface's radiometric response to locate potential CO₂ leakages on the surface using geostatistics and probability theory. Apart from being unsupervised in nature, an important feature of this methodology is the ability to combine independent information acquired by different modalities. This implies that in the future, there is further scope for improvement in the methodology by addressing the limitations highlighted in section 9.2.

It is perceived that early detection and change detection are complimentary to each other, and this is an important objective that is yet to be achieved in the application of optical remote sensing for the detection of surface leakages. In this regard, the proposed methodology could be extended by incorporating information related to the changes in vadose zone, such as the dielectric permittivity of material, temperature, moisture content, and other potentially useful geophysical and geochemical properties. The DS theory could be used to analyse change information from the shallow subsurface for the early detection of leakages, and provide an alarm even before any noticeable effects are seen on the surface. This would involve jointly analysing optical, thermal and RADAR datasets.

An additional component would be required in the existing methodology which takes care of quantifying the detected leakages. This calls for detailed investigations in radiative transfer analysis to understand the energy/mass exchange occurring at the Atmospheric Boundary Layer (ABL), i.e. between the land cover and atmosphere, using images acquired by multispectral and hyperspectral TIR sensors.

Based on the findings from the analysis of low resolution datasets in this research, a new airborne/spaceborne multispectral sensor could be designed in the future that
provide information at specific wavelengths in the spectrum related to CO₂ leakages and its effects on the surface environment. The choice of centre wavelengths for each band, and their corresponding Full Width at Half Maximum (FWHM) are important design parameters. When data from such sensors are available, it not only reduces the volume of redundant information but could also be processed rapidly using the proposed methodology. This also has an implication for reducing false positives in the detection results.

Furthermore, it is recommended that future work could be extended towards volumetric analysis of data acquired for the vadose zone, the section of soil and rocks between the water table and the surface, which covers at least the top few metres of the Earth. This can facilitate the development of a coupled model between the vadose zone and the surface (or the ABL) that may be necessary to address the limitations of the methodology highlighted earlier. However, imaging 3D information of the subsurface may not be as easy as getting 2D information of the surface with airborne/spaceborne sensor platforms owing to signal attenuation. Grandjean et al. (1999) were amongst the first to discuss systems based on the penetration capabilities of low RADAR frequencies, namely the L-band (1-2 GHz) and the P-band (300-500 MHz) for subsurface Earth observation. Examples of systems that operate at these frequencies are: AIRSAR developed by NASA, and Radar Aéroporté Multi-Spectral d'Etudes des Signatures (RAMSES) developed by the French agency ONERA. Typical signal penetration depths for three general soil types at different GPR frequencies are indicated in Table 9.1.

<table>
<thead>
<tr>
<th>Soil Type</th>
<th>1.5 GHz</th>
<th>1.0 GHz</th>
<th>500 MHz</th>
<th>300 MHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wet sand</td>
<td>3.3</td>
<td>4.9</td>
<td>9.8</td>
<td>16.4</td>
</tr>
<tr>
<td>Dry sand</td>
<td>6.0</td>
<td>9.0</td>
<td>17.9</td>
<td>29.9</td>
</tr>
<tr>
<td>Paleosoils</td>
<td>0.5</td>
<td>0.8</td>
<td>1.6</td>
<td>2.6</td>
</tr>
</tbody>
</table>

In recent years, particularly in the field of hydrogeology, there has been significant application of tomographic techniques using GPR (Vereecken et al., 2008) to model and understand soil hydraulic properties, water and energy fluxes, spatial and
Conclusions and Recommendations for Further Research

temporal dynamics of soil moisture profiles, and flow and transport mechanisms of soil contaminants. Through volumetric signal scattering in the subsurface at multiple polarisations, L- and P-band frequencies have the potential to map heterogeneities, such as lithological variations, water content, sedimentary structures, and oil and gas seepages. Zare et al. (2010) assessed the benefits for cueing optical remote sensing data with GPR data for the purpose of detecting buried targets. Considering these facts, it is hoped that detailed investigations on information fusion between these datasets could provide a solution for early CO₂ leakage detection and quantification.

The study of sea floor is also considered important because it may provide useful insights of the sub-surface regarding fluid-migration pathways, pockmark field evolution, and possible mechanisms for pockmark generation and maintenance. Manifestation of gas seepage is often conveyed by unit-Pockmarks, usually small seabed depressions (<5 m in diameter) which are found in isolation, in groups, and in association with larger Pockmarks, called normal-Pockmarks (Hovland and Judd, 1988). Unit-Pockmarks may occur as solitary features (in isolation), curvilinear strings of varying length and density, and clusters of varying density (Hovland et al., 2010). A typical sea floor image obtained from the Norwegian Trough, near the Troll hydrocarbon field, is illustrated in Figure 9.1.

Figure 9.1: A reconstructed image of the sea floor with gas seepages, showing Pockmarks with different sizes and distribution patterns; the patch seen inside the biggest normal-Pockmark indicates missing data (After Hovland et al. (2010)).

The image contains numerous solitary unit-Pockmarks scattered over an otherwise smooth sea floor, with high density near some of the normal-Pockmarks. In addition, it can be seen that some of these are clearly associated with trawl-scars running
Conclusions and Recommendations for Further Research

across the image at all orientations. The distribution of unit-Pockmarks is guided by diagenetic processes, such as the sub-sea formation of hydrates, e.g. Methane-Derived Authigenic Carbonates (MDAC) first found in North Sea Pockmarks in 1983 (Hovland et al., 1985), that clog normal-Pockmarks thereby causing the seepages to seal and forcing the gas column to find alternative paths of least resistance. Recently, Andrews et al. (2010) proposed an automatic extraction methodology based on the principles of geomorphometry and established algorithms for surface analysis. With this as the motivation, it is hoped that future analysis will contribute towards spatio-temporal processing of SONAR data to establish the relation between time varying sub-sea hydraulic properties owing to CO₂ seepage and the Pockmarked landscape of the sea floor.
References


Annunziatellis, A., Beaubien, S.E., Bigi, S., Ciotoli, G., Coltella, M., and Lombardi, S., 2008, Gas migration along fault systems and through the vadose zone in the Latera


180
References


Botev, Z.I., 2005, Stochastic Methods for Optimization and Machine Learning, B.Sc (Hons.) Thesis, Department of Mathematics, School of Physical Sciences, The University of Queensland, Australia.


References


Duchi, V., Minissale, A., Paolieri, M., Prati, F., and Valori, A., 1992, Chemical relationship between discharging fluids in the Siena-Radicofanì graben and the deep fluids produced by the geothermal fields of Mt Amiata, Torre Alfina and Latera (Central Italy), Geothermics, vol. 21, no. 3, pp. 401-413.


EPA, 2006, OTM10 - Optical remote sensing for emission characterization from non-point sources, United States Environmental Protection Agency.


References


Foody, G.M., 1992, A fuzzy sets approach to the representation of vegetation continua from remotely sensed data: an example from lowland health, Photogrammetric engineering and remote sensing, vol. 58, no. 2, pp. 221-225.


References


References


References


References


Metz, B., Davidson, O., de Coninck, H.C., Loos, M., and Meyer, L.A., 2005, IPCC special report on carbon dioxide capture and storage, prepared by working group III of the
References

intergovernmental panel on climate change, Cambridge UK: Cambridge University Press, Chapter 3.


References


Sanderson, C., and Paliwal, K.K., 2002, Information fusion and person verification using speech and face information, IDIAP Research Report 02-33, Martigny, Switzerland.


Scharf, L.L., 1990, Statistical signal processing: detection, estimation, and time series analysis, Addison-Wesley, Reading.


References


Tadjudin, S., and Landgrebe, D., 1998, Classification of High Dimensional Data with Limited Training Samples, PhD Thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN.


van Eijndthoven, W., 2005, Soil monitoring in the RECOPOL project, NITG 05-181-B1207, Netherlands Institute of Applied Geoscience (TNO).


Zhang, Y., and Ji, Q, 2005, Active and dynamic information fusion for facial expression understanding from image sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 699-714.
