

**Department of Spatial Sciences**

**Developing Land Management Units using Geospatial Technologies:  
an Agricultural Application**

**Georgina Margaret Rose Warren**

**"This thesis is presented as part of the requirements for the  
award of the Degree of Doctor of Philosophy  
of the  
Curtin University of Technology"**

**June 2007**

## **Declaration**

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: .....

Date: .....

## **ABSTRACT**

This research develops a methodology for determining farm scale land management units (LMUs) using soil sampling data, high resolution digital multi-spectral imagery (DMSI) and a digital elevation model (DEM). The LMUs are zones within a paddock suitable for precision agriculture which are managed according to their productive capabilities. Soil sampling and analysis are crucial in depicting landscape characteristics, but costly. Data based on DMSI and DEM is available cheaply and at high resolution.

The design and implementation of a two-stage methodology using a spatially weighted multivariate classification, for delineating LMUs is described. Utilising data on physical and chemical soil properties collected at 250 sampling locations within a 1780ha farm in Western Australia, the methodology initially classifies sampling points into LMUs based on a spatially weighted similarity matrix. The second stage delineates higher resolution LMU boundaries using DMSI and topographic variables derived from a DEM on a 10m grid across the study area. The method groups sample points and pixels with respect to their characteristics and their spatial relationships, thus forming contiguous, homogenous LMUs that can be adopted in precision agricultural applications. The methodology combines readily available and relatively cheap high resolution data sets with soil properties sampled at low resolution. This minimises cost while still forming LMUs at high resolution.

The allocation of pixels to LMUs based on their DMSI and topographic variables has been verified. Yield differences between the LMUs have also been analysed. The results indicate the potential of the approach for precision agriculture and the importance of continued research in this area.

## ACKNOWLEDGEMENTS

There are many people who have assisted in the completion of this thesis and I would like to take this opportunity to acknowledge them. For those I have not mentioned by name, I say thank you, and acknowledge their valuable contribution of advice, support or material assistance.

Firstly I express sincere gratitude to my supervisor Professor Graciela Metternicht for her constant support, constructive comments and editing. Her expertise in remote sensing for agricultural applications provided extensive knowledge and guidance throughout this project. I am indebted to my associate supervisor, Jane Speijers, without her assistance in statistical applications this thesis would not have been possible. Jane was also involved in the painstaking editing process. My associate supervisor Professor Murray McGregor was involved in initial thesis discussions.

Without funding this research would not have been possible, as such I would like to thank the Department of Spatial Sciences at Curtin University of Technology for the initial stipend (Christopher Worth Memorial Scholarship) along with ongoing support from the CPSTOF project funded by a grant from the Australian Academy of Sciences, within the ARC-Linkage program (LP0219752) and its partners; The Department of Agriculture and Food Western Australia and SpecTerra Services. An Australian Postgraduate Award provided the majority of the funding throughout the thesis.

Field data collection has been a large part of this research and I would like to thank those involved for organising The Muresk Institute of Agriculture as a suitable location. Many people have assisted in the collection process they include; Gregor Brockmann who spent several weeks soil sampling, others that assisted were Mandy Jodrell and Tony Warren. For vegetation sampling Tammi Short and Bernhard Klingseisen assisted in the field and laboratory. Several months were spent at the Chemistry Centre WA performing chemical analysis on the soil. I would like to thank Dr Dave Allen and his team; Jeremy Brown, Jenny Harbord, Simon Adams and Katrina Walton for all their help.

Several organisations have shared either their data or resources and I would like to record my appreciation. These include; The Department of Agriculture and Food Western Australia, SpecTerra Services Pty Ltd, Muresk Institute of Agriculture and The Chemistry Centre of WA.

Numerous members of the Department of Agriculture and Food Western Australia have provided valuable assistance in differing ways; Buddy Wheaton, Henry Smolinski, James Fisher, Bill Bowden and Paul Galloway.

I would like to thank several colleagues within the Department of Spatial Sciences for their assistance and valuable lively discussions; Deavi Purnomo, Dr Zahurul Islam, Dr Rachael O'Brien, Todd Robinson, Bernhard Klingseisen, Dr Lesley Arnold and Lori Patterson. In particular Todd Robinson interpolated all the yield data, Bernhard Klingseisen developed LANDFORM while Dr Lesley Arnold and Lori Patterson assisted immensely with editing. I also thank Lisa Warren for her valuable time in editing and Des Thornton for all his tips and tricks in formatting the final manuscript.

Finally I would like to thank my family and close friends for their support during the entire research period. In particular the final 1½ years have been the most challenging in terms of combining being a new mum and research. Many thanks to those who have taken care of Isabella (especially her Gran) so I could work. Without your help, I wonder if this thesis would have ever been finished. To my wonderful husband Tony and gorgeous Isabella, it has all been worth it!

## TABLE OF CONTENTS

ABSTRACT .....	i
ACKNOWLEDGEMENTS .....	ii
TABLE OF CONTENTS .....	iv
LIST OF FIGURES .....	xi
LIST OF TABLES .....	xv
LIST OF ACRONYMS AND OTHER ABBREVIATIONS .....	xviii
CHAPTER 1	
INTRODUCTION .....	1
1. 1    Problem Formulation .....	1
1. 2    Background .....	2
1. 3    Research Objectives .....	3
1.3.1    Aims of the Research .....	3
1.3.2    Expected Outcomes.....	5
1.3.3    Significance and Benefits of the Research.....	5
1.3.4    Research Methodology .....	5
1. 4    Overview of Thesis .....	7
CHAPTER 2	
LITERATURE REVIEW .....	10
2. 1    The Generation of Land Management Units.....	10
2.1.1    Methods of Classification of Spatial Zones .....	12
2.1.1.1    Spatially Constrained Classification .....	20
2. 2    Landscape Attributes Required to Form LMUs.....	22
2.2.1    Soil Properties .....	22
2.2.1.1    Soil Texture.....	25
2.2.1.2    Coarse fragments.....	26
2.2.1.3    Soil Stability.....	27
2.2.1.4    Organic Matter .....	28
2.2.1.5    Cation Exchange Capacity .....	29
2.2.1.6    Soil Acidity .....	31
2.2.1.7    Soil Alkalinity, Sodicity and Salinity .....	32

CHAPTER 2 *continued*

2.2.2	Vegetation .....	33
2.2.3	Topographic Attributes .....	34
2.3	Methods for Recording Landscape Attributes .....	34
2.3.1	Field Sampling Techniques for Soil and Vegetation .....	34
2.3.1.1	Soil .....	34
2.3.1.2	Vegetation .....	39
2.3.2	Remote Sensing of Soil and Vegetation.....	40
2.3.2.1	Reflectance of Vegetation .....	41
2.3.2.2	Remote Sensing of Agricultural Crops .....	42
2.3.2.3	Selected Spectral Indexes for Agricultural Applications .....	46
2.3.2.4	Reflectance of Soil properties .....	48
2.3.3	Digitally Derived Topographic Attributes .....	54
2.3.3.1	Landforms .....	54
2.3.3.2	Compound Topographic Index .....	57
2.3.3.3	Slope.....	59
2.4	Validation Techniques for the LMU classification.....	59
2.4.1	Comparisons with Yield Data .....	60
2.4.2	Spatial Map Comparisons .....	64
2.5	Summary .....	67

## CHAPTER 3

STUDY AREA AND DATA SETS .....	69
3.1 Study Area.....	69
3.1.1 Climate .....	70
3.1.1.1 Rainfall.....	70
3.1.2 Soils.....	73
3.2 Remote Sensing Data .....	77
3.2.1 Calibration of DMSI .....	78
3.2.1.1 Select a Reference Image.....	79
3.2.1.2 Select Invariant Targets.....	80
3.2.1.3 Calculate the Calibration Coefficients .....	83
3.2.1.4 Examine Calibration Curves and Adjust Targets.....	83
3.2.1.5 Calibrate the Image .....	87

CHAPTER 3 *continued*

3.3	Digital Elevation Model.....	88
3.4	Yield Data .....	90
3.5	Field Data.....	93
3.6	Software and Hardware.....	94
3.6.1	Global Positioning System (GPS).....	94
3.6.2	GIS Software.....	94
3.6.3	Statistical Software .....	94
3.7	Summary .....	94

## CHAPTER 4

FIELD DATA COLLECTION AND ANALYSIS.....		96
4.1	Soil Sampling Design.....	96
4.1.1	NDVI and RI Layers .....	98
4.1.2	Mask Layer .....	98
4.1.3	The Variability Map Layer.....	99
4.1.4	Slope Class Map.....	101
4.1.5	Variability Table .....	102
4.1.6	Slope Table .....	104
4.1.7	Sample Matrix.....	105
4.1.8	Sample Zones and Sampling points.....	105
4.2	Field Collection.....	107
4.3	Chemical Analysis of Soil Properties .....	108
4.4	Building the Soil Database.....	111
4.5	Vegetation Sampling.....	111
4.6	Generation of Vegetation Indices.....	115
4.7	Relationship between DMSI and Crop Attributes .....	116
4.8	Factors Affecting the Relationship between DMSI and Crop Attributes	117
4.8.1	Vegetation Indices and LAI .....	118
4.8.2	Wheat .....	121
4.8.3	Lupins.....	122
4.8.4	Canola .....	123
4.9	The Use of DMSI for Characterising Vegetation Variability .....	125
4.10	Relationship between DMSI and Yield.....	125



CHAPTER 4 *continued*

4. 11	Summary .....	126
-------	---------------	-----

## CHAPTER 5

## GENERATION OF TOPOGRAPHIC ATTRIBUTES AND ANCILLARY POINT

DATA.....	128
-----------	-----

5. 1	Landform Classification.....	128
------	------------------------------	-----

5. 2	Implementation of Landform Classification .....	129
------	---	-----

5.2.1	Comparison between Semi-automated and Manually Derived Landforms .....	132
-------	---	-----

5.2.2	Forming Landform Variables.....	133
-------	---------------------------------	-----

5. 3	Compound Topographic Index .....	136
------	----------------------------------	-----

5. 4	Implementation of CTI.....	138
------	----------------------------	-----

5. 5	Generation of a Slope Percentage Map.....	141
------	---	-----

5. 6	Extracting Ancillary Data at Soil Point Locations.....	142
------	--	-----

5.6.1	Extracting Topographic Attributes.....	142
-------	--	-----

5.6.2	Extracting DMSI for Soil Points in Areas without Vegetation Cover	143
-------	---	-----

5.6.2.1	Selecting Locations with Minimum Vegetation Cover.....	144
---------	--	-----

5. 7	Yield Data Sets.....	145
------	----------------------	-----

5.7.1	Extraction of Interpolated Yield Values .....	146
-------	---	-----

5.7.2	Potential Yield Estimation at Sample Points (Yield Data Set 1).....	148
-------	---	-----

5.7.2.1	Multiple Regression Analysis .....	148
---------	------------------------------------	-----

5.7.2.2	Estimation of Potential Yield at Sample Point locations .....	150
---------	---	-----

5.7.3	Actual Yield Data at Soil Points for 2002 and 2003 (Yield Data Set 2)	151
-------	---	-----

5. 8	Extracting Ancillary Data for the Mask Area .....	152
------	---	-----

5.8.1	Extracting Topographic and DMSI Variables on a 10m Grid .....	152
-------	---	-----

5. 9	Summary .....	153
------	---------------	-----

## CHAPTER 6

## SELECTION OF VARIABLES FOR USE IN FORMING LAND MANAGEMENT

UNITS.....	155
------------	-----

6. 1	Selecting and Deriving Stable Soil Properties from Existing Literature .	155
------	--	-----

CHAPTER 6 *continued*

6.2	Selecting Variables Based on their Association with Crop Yield.....	157
6.2.1	Transforming Variables .....	157
6.2.2	Correlation Analysis .....	159
6.2.3	Associations between Yield and Landforms.....	160
6.2.4	Multiple Regression between Yield and Soil Properties.....	161
6.2.5	Discussion of Results .....	163
6.3	Selecting Variables Using Principal Component Analysis.....	165
6.4	Relationships between Selected Soil properties and ancillary data .....	173
6.4.1	Correlation Analysis .....	174
6.4.2	Analysis of Variance with Landforms .....	174
6.5	Summary .....	176

## CHAPTER 7

A FRAMEWORK FOR CREATING LAND MANAGEMENT UNITS .....		177
7.1	The Land Management Unit Classification .....	177
7.2	Part A: Forming LMUs from Soil Point Data.....	180
7.2.1	Transforming Input Variables.....	180
7.2.2	Estimating Geographical Parameters from Variograms .....	180
7.2.3	Determining the Scale and Form of Spatial Variation.....	181
7.2.4	Forming the Similarity Matrix on Soil Variables .....	186
7.2.5	Determining the Geographic Distance between Sample Points.....	187
7.2.6	Forming Matrix of Spatial Weights .....	188
7.2.7	Forming Principal Coordinates from a Similarity Matrix.....	189
7.2.8	Selecting the Appropriate Number of LMUs.....	190
7.2.9	Non-hierarchical Clustering of 250 Soil Points .....	194
7.2.10	Utilising Landform Variables .....	195
7.2.11	Analysis of Variance of Ancillary Data and LMUs.....	195
7.3	Part B: Further Defining LMU Boundaries using High Spatial Resolution Ancillary Data .....	196
7.3.1	Transforming Ancillary Variables .....	196
7.3.2	Similarity between Mask Pixels and LMUs; Option 1 .....	196
7.3.3	Discriminant Analysis between Mask Pixels and LMUs: Option 2	197

CHAPTER 7 *continued*

7.3.4	Forming the Geographically Weighted Dissimilarity Matrix between Soil Based LMUs and Mask Data Set.....	197
7.4	Provisional LMUs .....	198
7.5	The Land Management Unit Map and Unit Descriptions.....	201
7.6	Summary .....	208

## CHAPTER 8

VALIDATION OF LMU CLASSIFICATION .....		210
8.1	Choosing a Validation Procedure.....	210
8.2	Validation of LMU Allocation using Ancillary Data .....	211
8.2.1	The Random Selection .....	213
8.2.2	Part A: Forming LMUs based on Soil Point Properties.....	213
8.2.3	Part B: Assigning Random Points to LMUs based on Ancillary Data .....	213
8.2.4	Part C: Assign Random Points to LMUs based on Soil Properties ..	214
8.3	Sensitivity Analysis of LMU Classification .....	216
8.3.1	Results from the Kappa Map Comparison.....	218
8.4	Relationship between LMUs and Yield.....	222
8.5	Summary .....	224

## CHAPTER 9

CONCLUSIONS AND RECOMMENDATIONS .....		225
9.1	Conclusions.....	225
9.1.1	Determine Stable Soil Properties that have the most influence on Yield Variability in the Agricultural belt of WA.....	225
9.1.2	Employ High Resolution, Readily Available, Cheap Ancillary Data Sets that are related to Soil Properties and/or Landscape Variability .....	226
9.1.3	Develop a Methodology to combine information derived from High Resolution Data Sets and Soil Properties at Point Locations.....	229
9.1.4	Produce a LMU Map with associated Soil Properties at Paddock Scale .....	229
9.1.5	Validate the Methodology for forming LMUs.....	231
9.1.6	Develop an Effective Field Soil Sampling Strategy .....	232

CHAPTER 9 *continued*

9.1.7	Determine the Opportunities and Limitations of using High-Resolution Digital Multi-spectral Imagery as a diagnostic tool for monitoring Crop Growth during the Growing Season.....	232
9.2	Recommendations .....	233
9.2.1	Data Sources .....	233
9.2.2	Opportunities of DMSI .....	234
9.2.3	GIS Programming .....	234
9.2.4	Validations techniques .....	234
	REFERENCES.....	236
	APPENDICES .....	259
A	Daily rainfall 2002 and 2003.....	260
B	Soil Sampling Design.....	263
C	Drainage Classes .....	268
D	Field Texture Estimation.....	270
E	Full Soil Database Schema .....	272
F	Implementation of Landform Classification.....	274
G	Implementation of Compound Topographic Index.....	277
H	Slope Algorithm.....	280
I	Significant Difference between Oats and Oaten Hay Crop Types .....	282
J	Estimated Yield at Sample Points .....	284
K	Soil Variables Transformations.....	287
L	Ancillary Variables Transformations .....	300
M	Correlations between Soil Variables .....	304
N	Regionalized Variable Theory .....	311
O	Experimental Variograms .....	314
P	Box-Plots of Soil Properties and Ancillary Variables .....	319
Q	Results from Kappa Map Comparisons .....	323
	APPENDICES' REFERENCES .....	329

## LIST OF FIGURES

Figure 1.1 Flow diagram of thesis chapters .....	7
Figure 2.1 Spectral Partitioning of solar irradiance by vegetation.....	41
Figure 2.2 Visible and near-infrared reflectance differences in vegetation due to senescence .....	42
Figure 2.3 Soil reflectance spectra for a silt loam soil at varying moisture contents. Percentage of moisture content by weight shown above each curve.....	50
Figure 2.4 The textural triangle defines soil texture on the basis of sand, silt and clay .....	51
Figure 2.5 Example of a profile across terrain divided into morphological types of landform elements.....	56
Figure 3.1 Study Area Location.....	70
Figure 3.2 Rain gauge locations for Muresk Farm .....	71
Figure 3.3 Cumulative rainfall by month for the years 1996 – 2003 .....	72
Figure 3.4 Total and Growing season rainfall for 1996-2003.....	73
Figure 3.5 Muresk farm soils .....	75
Figure 3.6 Geographic distribution of target locations overlayed on reference image (DMSI June 2004).....	81
Figure 3.7 Scatter plots showing the least squares regression line for June 2002 imagery for a) Band 1, b) Band 2, c) Band 3 and d) Band 4.....	84
Figure 3.8 Scatter plots showing the least squares regression line for September 2002 imagery for a) Band 1, b) Band 2, c) Band 3 and d) Band 4 .....	85
Figure 3.9 Scatter plots showing the least squares regression line for September 2003 imagery for a) Band 1, b) Band 2, c) Band 3 and d) Band 4 .....	86
Figure 3.10 DEM systematic error in the form of tiling and a seam running north- south.....	89
Figure 3.11 An example of an interpolated yield surface overlayed with filtered yield points and paddock boundary.....	91
Figure 3.12 Field data collection during June 2004 DMSI flight .....	93
Figure 4.1 Flow chart of sampling design.....	98
Figure 4.2 Histogram of NDVI and RI change % values for mask region .....	100
Figure 4.3 Variability map generated with a 10m spatial resolution .....	101
Figure 4.4 Slope map of study area.....	102

Figure 4.5 Histogram of slope values for the area of interest.....	102
Figure 4.6 Spatial distribution of the soil sample points used for survey of the Muresk farm.....	106
Figure 4.7 Soil sampling Field Data Collection.....	107
Figure 4.8 Soil chemical analysis .....	109
Figure 4.9 Soil database schema .....	111
Figure 4.10 Field data collection and laboratory analysis.....	114
Figure 4.11 Status of crops in year 2002 .....	119
Figure 4.12 Status of crops in year 2003 .....	119
Figure 4.13 Relationship between NDVI and Crops' LAI .....	121
Figure 4.14 Spectral Curves for (a) 2002; (b) 2003 .....	124
Figure 5.1 Flow chart of implementation of Landform .....	129
Figure 5.2 Position of slope elements in a toposequence.....	131
Figure 5.3 Map of landform classes across Muresk farm .....	132
Figure 5.4 Plot of principal coordinate scores .....	135
Figure 5.5 Upslope contributing area ( $A$ ) is the area of land upslope of a length of contour ( $l$ ). Specific catchment area ( $A_s$ ) is $A/l$ .....	136
Figure 5.6 Flow chart of implementation of CTI in GeoMedia Grid.....	139
Figure 5.7 CTI layer across Muresk Farm .....	140
Figure 5.8 Slope percentage map of the Muresk farm .....	141
Figure 5.9 Soil sampling points converted to 10m raster cells overlaid on landforms .....	143
Figure 5.10 Overview of process for extracting DMSI data at soil points with bare soil.....	144
Figure 5.11 Soil sampling points with 3 x 3 neighbourhood window overlaid on a true colour DMSI composite.....	146
Figure 5.12 Soil sampling points and their distance from field recorded GPS yield points overlaid on yield interpolated surface.....	147
Figure 5.13 Plot of residuals against fitted values following Equation (5.5).....	149
Figure 5.14 Plot of residuals against fitted values for the final model .....	150
Figure 5.15 Frequency distribution of the Standardised Yield Estimations for the Soil Points .....	151
Figure 5.16 Frequency distribution of the yield 2002 at soil points .....	151
Figure 5.17 Frequency distribution of the yield 2003 at soil points .....	152

Figure 5.18 Overview of process for extracting DMSI data from 10m cells.....	153
Figure 6.1 The relative amounts of water available and unavailable for plant growth in soils with textures from sand to clay.....	164
Figure 6.2 Biplots based on 250 soil points and the stable soil variables. a) PC1 vs PC2; b) PC1 vs PC3 .....	167
Figure 6.3 Biplots based on 250 soil points and the stable soil variables. a) PC1 vs PC4; b) PC1 vs PC5 .....	168
Figure 6.4 Relationship between pHw and pHca for 10 and 30cm depths.....	172
Figure 7.1 Flowchart of LMU classification.....	179
Figure 7.2 Commonly used variogram models: (a) spherical; (b) exponential; and (c) Gaussian .....	183
Figure 7.3 Experimental and exponential model variograms for soil properties and principal components after outliers have been removed.....	185
Figure 7.4 Plot of a) $g^2S$ against $g$ , b) Criterion $S$ against $g$ and c) $\log S$ against $g$ based on 13 soil variables .....	192
Figure 7.5 Plot of a) $g^2S$ against $g$ , b) Criterion $S$ against $g$ and c) $\log S$ against $g$ based on 19 linearly independent soil variables.....	193
Figure 7.6 Plot of a) $g^2L^*$ against $g$ , b) Criterion $L^*$ against $g$ and c) $\log L^*$ against $g$ based on 13 soil variables .....	193
Figure 7.7 Plot of a) $g^2L^*$ against $g$ , b) Criterion $L^*$ against $g$ and c) $\log L^*$ against $g$ based on 19 linearly independent soil variables.....	193
Figure 7.8 Soil point LMU groups based on 250 soil points for a range of a) 200m, b) 250m and c) 300m .....	194
Figure 7.9 Nearest neighbour similarity assigned to a mask point for each LMU .	198
Figure 7.10 LMU groups based on Option 1: Cluster/Cluster analysis for a range of a) 200m, b) 250m and c) 300m and Option 2; Discriminant analysis for a range of d) 200m, e) 250m and f) 300m .....	200
Figure 7.11 Biplot showing soil points within LMUs and canonical variate loadings for CV1 vs CV2 .....	202
Figure 7.12 Biplot showing soil points within LMUs and canonical variate loadings for CV1 vs CV3 .....	203
Figure 7.13 Biplot showing soil points within LMUs and canonical variate loadings for CV1 vs CV4 .....	203

Figure 7.14 The Land Management Unit map for Muresk Farm.....	206
Figure 8.1 Flowchart of random comparison/validation process.....	212
Figure 8.2 Histogram of the percentage of <i>correct</i> assignment of $S$ (50) random points to LMUs for $c$ (100) cycles .....	215
Figure 8.3 Kappa Map Comparison resultant agreement/disagreement maps for differing parameters a) Cluster/Cluster 200m b) Cluster/Cluster 300m c) Discriminant 200m d) Discriminant 250m e) Discriminant 300m .....	219
Figure 9.1 (a) Muresk soil map versus (b) LMU map generated in this research. .	230



## LIST OF TABLES

Table 2.1 Summary, in chronological order of methods for classifying spatial zones .....	14
Table 2.2 Texture grades based on clay content .....	26
Table 2.3 Determining the soil stability score for soils .....	28
Table 2.4 Ca:Mg ratio for USA soils .....	28
Table 2.5. Percentage of organic matter ratings at surface horizon (A1) .....	29
Table 2.6 Ratings for chemical properties .....	31
Table 2.7 Rating and salinity categories and associated effects .....	33
Table 2.8 Examples of Vegetation Indices .....	44
Table 2.9 Iron oxides and soil colour.....	53
Table 2.10 Morphological type (topographic position) classes .....	56
Table 3.1 Description of Soil Survey hierarchy.....	74
Table 3.2 Description of subsystem, landform and major soil types present in the study area .....	76
Table 3.3 DMSC Flight specifications.....	77
Table 3.4 Dynamic range of pixel values for DMSI.....	79
Table 3.5 Targets selected for calibrations .....	82
Table 3.6 Least-squares regression model for image calibration.....	87
Table 3.7 Raw, Reference and Calibrated image statistics .....	88
Table 3.8 Interpolated yield data available for Muresk Farm.....	92
Table 4.1 Definition of slope classes .....	97
Table 4.2 Change % range for variability class of NDVI and RI .....	100
Table 4.3 Initial and final selection of Pairwise Ratio Matrix values and Weights	104
Table 4.4 Variability proportion determination .....	104
Table 4.5 Number of sample points per slope class.....	105
Table 4.6 Sample Matrix of slope and variability classes.....	105
Table 4.7 Soil sampling field data collection.....	108
Table 4.8 Methods of soil chemical analysis .....	110
Table 4.9 Transect Attributes.....	112
Table 4.10 Methodology to determine crop density .....	113
Table 4.11 Basic descriptive statistics of crop attributes recorded on the transects	115
Table 4.12 Correlation coefficient for crop LAI vs Vegetation Indices .....	117

Table 4.13 Correlation coefficient for canola PAI and LAI vs DMSI bands and LAI vs Visual % Cover.....	117
Table 5.1 Default values used for the generation of primary landforms .....	130
Table 5.2. The landform similarity matrix .....	134
Table 5.3. Percentage variance accounted for by 7 landform classes.....	135
Table 5.4 Principal coordinate scores for landform PC's .....	136
Table 5.5 Number of soil points with NDVI $\leq 0$ .....	145
Table 5.6 A description of the two yield data sets and their attributes .....	146
Table 5.7 List of attributes recorded for soil and vegetation point locations during yield extraction.....	147
Table 6.1 Stable soil properties and variables from literature review.....	156
Table 6.2 Variable transformations and resultant skewness .....	158
Table 6.3 Correlation coefficients from analysis between yield and other variables .....	159
Table 6.4 ANOVA for Yield against Landforms.....	161
Table 6.5 Regression results from analysis between yield and soil properties.....	162
Table 6.6 Percentage of variance explained by each principal component and loadings of variables for each principal component from the 250 soil points .	166
Table 6.7 Correlation of texture and ECEC .....	169
Table 6.8 List of 13 soil variables to be included in LMU model .....	173
Table 6.9 Correlation between 13 selected soil properties and ancillary variables	174
Table 6.10 ANOVA for Soil Properties against Landforms.....	175
Table 7.1 Principal components of correlation matrix of 13 soil properties.....	182
Table 7.2 Exponential variogram models for soil properties and principal components .....	184
Table 7.3 Percentage of variance accounted for by 13 PCO's for effective ranges of 200m, 250m and 300m.....	190
Table 7.4 Summary of principal coordinate analysis over several spatial ranges ..	190
Table 7.5 Significance of differences between LMUs formed at each range for ancillary variables .....	195
Table 7.6 Soil variables used in discriminate analysis.....	201
Table 7.7 Percentage of variance explained and variable loadings of variables for each canonical variate .....	202
Table 7.8 Discriminating tree for LMU groups .....	204

Table 7.9 Description of LMUs present in Figure 7.14 .....	207
Table 8.1 The contingency table in its generic form.....	217
Table 8.2 The Kappa statistics; overall and for individual LMUs, used for comparison with the LMU map based on a Cluster/Cluster method with a range of 250m .....	220
Table 8.3 Summary of the individual LMU group comparisons that have not been mapped well ( $\geq 20$ percent to another LMU) and LMU legend ranking .....	221
Table 8.4 Accumulated analysis of variance for regression model in Equation (8.9) .....	223
Table 8.5 Predicted Yield for each LMU based on the regression model .....	223
Table 9.1 Stable soil properties and variables from literature review.....	226
Table 9.2 Comparison of the cost of some high resolution data sets January 2007	227

## LIST OF ACRONYMS AND OTHER ABBREVIATIONS

ANOVA	Analysis of variance
CaMg	Calcium/Magnesium ratio
CTI	Compound Topographic Index
CV	Canonical Variate
DEM	Digital Elevation Model
dGPS	Differential Global Positioning System
DLI	Department of Land Information
DMSC	Digital Multi-Spectral Camera
DMSI	Digital Multi-Spectral Imagery
DN	Digital Number
ECEC	Effective Cation Exchange Capacity
ESP	Exchangeable Sodium Percentage
GIS	Geographic Information System
GPS	Global Positioning System
GSR	Growing Season Rainfall
LAI	Leaf Area Index
LF_PC	Landform Principal Coordinate
LMU/LMUs	Land Management Unit/Units
MCK	Map Comparison Kit
NDVI	Normalised Difference Vegetation Index
NDVIgreen	Normalised Difference Vegetation Index-Green
OrgC	Organic Carbon
PA	Precision Agriculture
PAI	Petal Area Index
PCO	Principal Coordinates
PPR	Plant Pigment Ratio
PVR	Photosynthetic Vigour Ratio
RI	Redness Index
SAVI	Soil Adjusted Vegetation Index
SSCM	Site Specific Crop Management
VI	Vegetation Index/Indices
WA	West/Western Australia

## CHAPTER 1

### INTRODUCTION

#### 1.1 Problem Formulation

Precision Agriculture has been defined as '*observation, impact assessment and timely strategic response to fine-scale variation in causative components of an agricultural production process*', and thus may cover a range of agricultural enterprises, and can be applied to pre- and post-production aspects of agricultural enterprises (Australian Centre for Precision Agriculture, 2006). Site-specific crop management (SSCM) is one facet of precision agriculture and is defined as '*matching resource application and agronomic practices with soil and crop requirements as they vary in space and time within a field*' (Whelan and McBratney, 2000). A cost effective and practical system of dividing a field into land management units that account for the spatial variability of crop production across spatially separable zones in Australia, needs to be offered for SSCM to be tested, accepted and adopted (Whelan *et al.*, 2002a).

High resolution remote sensing imagery is useful for agricultural spatial planning at the farm scale, particularly when it is incorporated into a Geographic Information System (GIS) containing other secondary data sources such as terrain attributes. It offers farmers, farm advisors and extension officers, a system to identify areas within a paddock which can be managed differently based on their individual characteristics, and can be incorporated into various land use scenarios. Farmers have shown considerable interest in the opportunities presented by remote sensing, especially where its application can lead to improved farm management.

As with all new management tools, the new techniques require further investigation and economic justification before they are widely accepted and adopted by the farming community. This research addresses this requirement by developing a framework for determining land management units across an entire farm.

## 1.2 Background

In Western Australia, current soil and land capability maps provide information suitable for studies at regional or catchment scale. However, such maps lack the level of detail required for sound planning at farm scale. John Blake (2001), the leader of the Western Australian Department of Agriculture and Food's Precision Farming Group, stated that Precision Agriculture has the potential to manage variability in the State's agricultural areas which probably have the most variable soils in the world (e.g. farms can have easily five changes of soil type in one circuit of the paddock) (Blake and Huffer, 2001).

Remote sensing imagery incorporated into a GIS has the potential to be useful in establishing the location and extent of soil variability. This is important as farmers have expressed the need for detailed assessment (at farm scale) of paddock conditions (e.g. soil and land capability maps adequate for decisions at farm scale) to aid in the planning of alternative land uses or land improvements according to the current conditions of the land (Blake and Huffer, 2001).

Dividing a paddock into relatively homogenous management units offers opportunities for the manager to exploit sites according to their specific capability. The delineation of land management units (LMUs), or zones, has been investigated using several techniques. They include a mechanistic simulation model based on detailed soil inventory and climate records (Alphen and Stoorvogel, 1998), multi-year yield estimates derived from Landsat Thematic Mapper imagery (Boydell and McBratney, 1999), multivariate K-means clustering utilising temporal yield data (Cupitt and Whelan, 2001; Dobermann *et al.*, 2003; Whelan and McBratney, 2003), continuous soil electrical conductivity (EC) measures (Nehmdahl and Greve, 2001), morphological and spectral filtering of elevation data and soil EC combined in binary form (Zhang and Taylor, 2001) and the classification of remotely sensed imagery using supervised and unsupervised methods (Yang and Anderson, 1996; Stewart and McBratney, 2001). These techniques have had differing levels of success. Approaches based on intensive sampling or continuous sensors to map crop yield and soil properties tend to be expensive and time consuming.

More cost effective techniques are clearly required in order to classify LMUs with the detail required for applications of precision agriculture. The methods based on cheap high resolution data sets in combination with soil properties recorded at point locations, which are investigated here, have the potential to minimise costs while still providing an output at the required level of detail. Sampling strategies for soil properties can take numerous forms (Burrough and McDonnell, 1998a; McBratney *et al.*, 1999), and a strategy that addresses the variability of soil properties and subsequently makes an optimal placement of sampling points can minimise cost. In this research, various strategies will be reviewed and an appropriate method implemented, which will result in a spatial database of soil properties for the case study area.

An analysis of high resolution data with soil properties and crop attributes will need to be conducted to determine their suitability for forming LMUs. Other high resolution data sets include Digital Multi-Spectral Imagery (DMSI) and terrain variables derived from a Digital Elevation Model (DEM). The uses of these data have shown some promise. For example several studies have highlighted relationships between terrain attributes and soil properties (Moore *et al.*, 1993a; Gessler *et al.*, 1995; Boer *et al.*, 1996; McKenzie and Ryan, 1999) while remotely sensed imagery has been used in numerous vegetation studies (Weigand *et al.*, 1991; Cloutis *et al.*, 1996; Mogensen *et al.*, 1996; Cloutis *et al.*, 1999; Metternicht *et al.*, 2000; Senay *et al.*, 2000; Thenkabail *et al.*, 2000; McNairn *et al.*, 2002) and soil based applications (Bowers and Hanks, 1965; Stoner and Baumgardner, 1981; Karmonova, 1982; Latz *et al.*, 1984; Coleman and Montgomery, 1987; Escadafal, 1989; Henderson *et al.*, 1989; Epema, 1993; Metternicht and Zinck, 1997).

### **1.3 Research Objectives**

#### **1.3.1 Aims of the Research**

The main objective of this research is to *develop a framework for classifying a farm into paddock scale, homogenous land management units (LMUs)* that can be linked to the 'patch zone management' approaches used in Precision Agriculture. There are several principles upon which this framework has been based which should be understood by users before they adopt the methodology:

- i. LMUs are based on stable soil properties;
- ii. The scale of LMUs is finer than the minimum size of current paddocks;
- iii. The cost of producing the LMUs is kept low, so that the adoption by farmers' advisory groups is facilitated, and;
- iv. Data sets which are utilised are available in most agricultural regions of WA.

To satisfy these principles further aims of this research are to;

- a) Determine stable soil properties that have the most influence on yield variability in the agricultural belt of WA;
- b) Employ high resolution, readily available, cheap ancillary data that are related to soil properties and/or landscape variability;
- c) Develop a methodology to combine information derived from high resolution data sets and soil properties at point locations;
- d) Produce a LMU map with associated soil properties at paddock scale, and to;
- e) Validate the methodology for forming LMUs.

Addressing these aims encompasses the following steps;

- Collecting soil information (e.g. physical and chemical soil properties) and crop growth attributes across a selected study area;
- Investigating high resolution ancillary data that could be used in the formation of LMUs;
- Examining relationships between DMSI and crop growth variables to determine their use as diagnostic tools for depicting within paddock crop variability;
- Analysis of soil properties and yield data to determine soil properties that influence yield variability;
- Describing the relationship between LMUs and soil properties, and;
- Examining the success of the LMU classification.

Further derivative aims have evolved which include;

- f) To develop an effective field soil sampling strategy, and;
- g) To determine the opportunities and limitations of using high-resolution digital multi-spectral imagery as a diagnostic tool for monitoring crop growth during the growing season.



### 1.3.2 Expected Outcomes

The outcomes of the research are;

- A framework for rapid and efficient production of farm maps suitable for decision making related to site-specific management within paddocks;
- A soil sampling strategy that reduces the amount and cost of field work and soil laboratory analyses while still generating adequate information for producing LMUs and their associated soil characteristics, and;
- Spatial tools for explaining yield variability at farm and paddock level.

### 1.3.3 Significance and Benefits of the Research

The research will develop a framework for rapid mapping of LMUs that can be used by extension officers, farm advisors and farmers to make appropriate economic and environmental decisions on land uses, while limiting the cost of expensive sampling.

The LMUs will provide spatial zones that can be incorporated into automated land evaluation models and used to conduct land suitability analysis. Outputs that can be generated once the LMUs are defined and characterised include; physical suitability in terms of current use; proposed scenarios for improvements for paddock/sub-paddock level management, and; alternative land use scenarios.

An evaluation of the use of DMSI (flown by SpecTerra Systems) to support the generation of LMUs will provide potential users with the opportunity to assess whether it is an appropriate layer of information to be included in their farm management programs. It is anticipated that the findings will easily be adapted into the so-called “next generation” of satellites (e.g. Resource 21, IKONOS, Quickbird, OrbView-3) with similar spectral, spatial, and radiometric characteristics.

### 1.3.4 Research Methodology

The study comprises the following major components:

- 1) Review of relevant literature to determine techniques appropriate for LMU classification, stable soil properties and their relation to plant growth, suitable soil and vegetation sampling techniques, potential of remote sensing applications for detecting landscape variability, topographic attributes useful in soil/landscape studies and LMU classification validation techniques.

- 2) Process existing data sets, i.e. calibrate DMSI and interpolate yield point data to create a continuous surface.
- 3) Design and implement an optimal soil sampling strategy, and collect and analyse soil samples to form a comprehensive soil database.
- 4) Design and conduct vegetation sampling synchronous with capture of DMSI. Perform statistical analysis between *in situ* crop attributes and DMSI to determine the potential of DMSI in detecting landscape variability.
- 5) Generate topographic attributes, namely landforms (utilising the LANDFORM system), compound topographic index (CTI) and slope percentage.
- 6) Conduct a statistical analysis between yield and soil properties, topographic attributes and DMSI to determine input data for LMU classification. Perform principal component analysis on soil properties to examine appropriate stable soil properties for the classification of LMUs.
- 7) Use a spatially weighted multivariate classification technique to form contiguous, homogenous LMUs.
- 8) Perform a validation process to assess the success of the LMU classification technique.

## 1.4 Overview of Thesis

Figure 1.1 displays a flow diagram indicating the connection of each chapter contained within this thesis. The chapters' components and their relationship to the thesis structure are detailed individually hereafter.

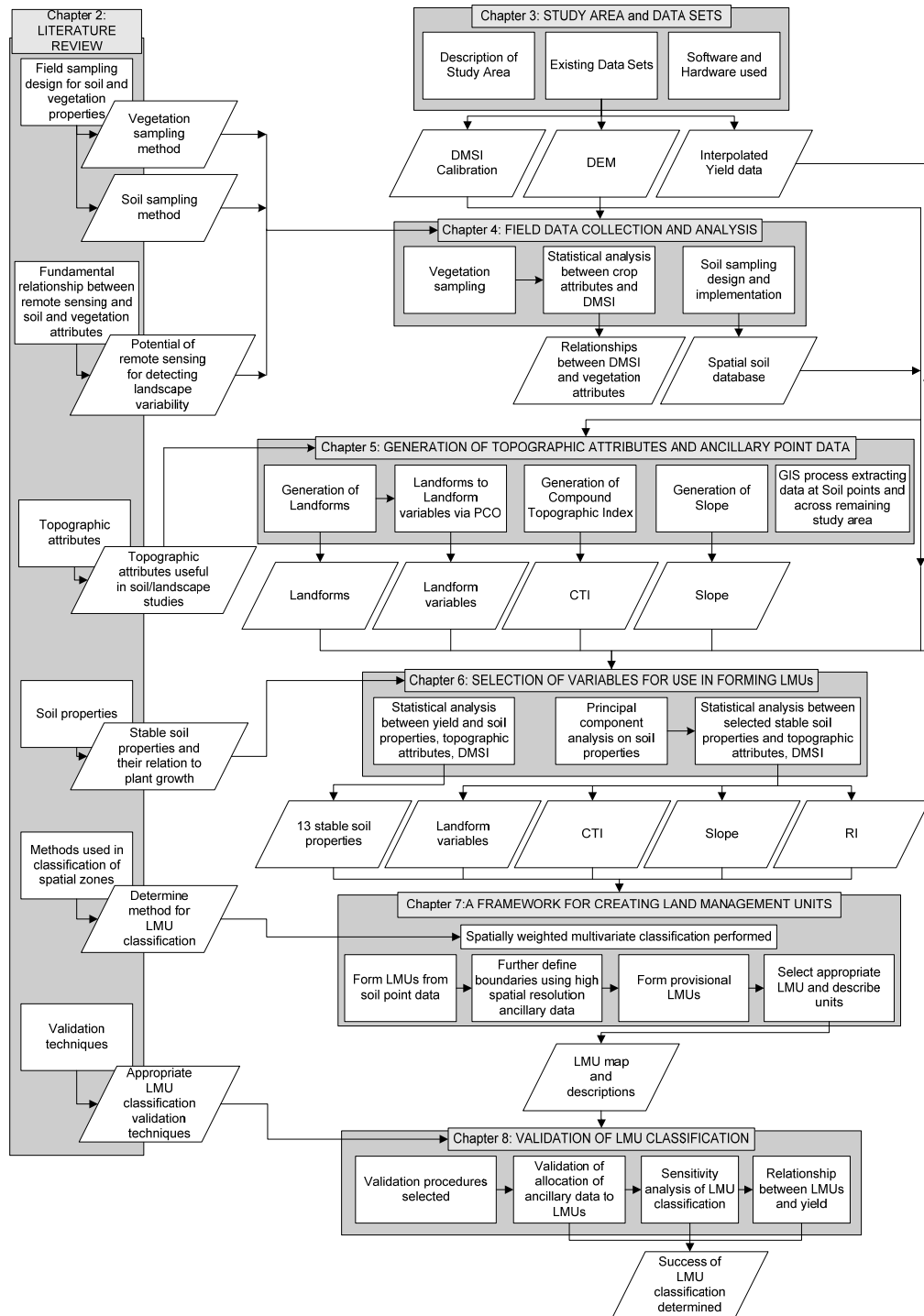


Figure 1.1 Flow diagram of thesis chapters

Chapter 1 describes the research problem, providing relevant background information and setting the research objectives.

Chapter 2 contains a literature review. It defines LMUs with a comprehensive review of methods for classification of the landscape into spatial zones, culminating in the selection of an appropriate methodology for this research. Stable soil properties are identified and described in relation to plant growth. These properties are used in the statistical analysis with yield data (Chapter 6). Methods of field sampling design and implementation for soil and vegetation are described, which will be implemented in the field data collection (Chapter 4). The fundamental relationship between remotely sensed imagery and soil and vegetation attributes are detailed highlighting the potential of high resolution remotely sensed data for characterising landscape variability. Topographic attributes that are related to soil and landscape variables are identified and described; and their generation is illustrated in Chapter 5. Chapter 2 concludes with a description of validation techniques which could be considered for the LMU classification.

The selected study area is described in relation to climate, current soil information and land use in Chapter 3. Existing data sets which include high resolution remote sensing data, digital elevation model, yield data and field data are detailed. Calibration of remote sensing data is included. The software and hardware used in the research are also detailed in Chapter 3.

Chapter 4 includes a detailed description of the soil sampling methodology and implementation. The resulting spatial soil database includes the soil properties to be statistically analysed in relation to yield data (Chapter 6) and selected as inputs into the LMU classification (Chapter 7). The collection of field vegetation attributes is detailed and a thorough analysis of their relation to DMSI is conducted.

The generation of topographic attributes namely; landform, compound topographic index (CTI) and slope derived from the DEM are explained in Chapter 5. These are statistically analysed to determine their appropriateness as inputs to the LMU classification in Chapter 6. Chapter 5 also includes the GIS process devised for extracting the yield, topographic and DMSI data at soil points. This method of

constructing the yield data sets is later used in the statistical analysis (Chapter 6). The procedure for extraction of the topographic and DMSI data for the remainder of the study area is also detailed.

The statistical analysis between yield data and soil properties, topographic attributes and DMSI data is implemented in Chapter 6 with the intention of determining the most influential properties on yield variability. These properties are subsequently used as the input to the LMU classification (Chapter 7). Correlation, multiple regression, analysis of variance and principal component analysis are performed.

The framework for developing LMUs is described and implemented in Chapter 7. A spatially weighted multivariate classification technique was selected as a consequence of reviewing relevant literature (Chapter 2). The classification has been conducted using several parameters and the most appropriate map of LMUs has been selected. The LMUs have been further described based on the soil properties and landscape attributes that would be appropriate to farmers and farm advisors.

A validation of the LMU classification is performed in Chapter 8. Three methods for determining the degree of success in the formation of LMUs have been used. Finally, conclusions and recommendations for future research are provided in Chapter 9.

## CHAPTER 2

### LITERATURE REVIEW

*The purpose of this research, to “develop a framework for classifying LMUs” and the principles upon which the framework has been based, has been presented in the introductory chapter along with an overview of thesis structure. Chapter 2 continues with a relevant literature review.*

*The following literature review covers background information on several topics necessary in this research. Firstly, a review of methods used to classify spatial zones is included with emphasis on studies that would be useful at farm scale, incorporate multivariate data and produce homogenous zones. Compiled from the review of methods for generating LMUs, the most common landscape attributes used as inputs have been identified, namely soil properties, vegetation and topographic attributes. Subsequently, these attributes have been reviewed in relation to their effects on plant growth and/or use in landscape classification applications. Methods for the collection, measuring and/or derivation of the soil properties, vegetation and topographic attributes have been reviewed which range from field sampling, remote sensing through to digitally derived by means of the computation of existing data. Finally an overview of methods for validating the LMU classification approach has been included.*

#### **2.1 The Generation of Land Management Units**

As identified in Chapter 1, the main objective of this research is to develop a framework for classifying farms at paddock scale into homogeneous LMUs that can be linked to site-specific crop management approaches used in precision agriculture. The principals of the framework have been detailed, stipulating that the framework for forming LMUs should be (i) based on stable soil properties; (ii) at within paddock scale; (iii) affordable to farmers, and; (iv) utilise available data. The following section firstly identifies what a LMU is and then goes on to examine methods for their generation, summarising with an approach that will be used in this research.

Dividing a paddock into relatively homogenous LMUs offers opportunities for the manager to exploit sites according to their specific capability. LMUs have been defined by various authors, at various scales, for various purposes. Lloyd (2003) defines a LMU as “*an area of land, with common soils and landforms that should be managed similarly, to maximise production and minimise land degradation*”. Other researchers have not used the actual term LMU, but management zones or units. Doerge (1998) defines a ‘management zone’ as “*a portion of a field that expresses a homogeneous combination of yield-limiting factors for which a single rate of a specific crop input is appropriate.*” And thus, to be successful, the factors included in the delineation method must be based on the true causes that affect the crop yield at that site (Doerge, 1998). Lark *et al.* (2003) define a ‘management zone’ or ‘zone’ as “*a region of a field defined according to some criteria on the assumption that all sites within a zone are expected to be subject to similar constraints on crop performance and, therefore, might be managed in the same way.*” Although not stating an actual definition, several researchers (van Alphen and Stoorvogel, 1998; Boydell and McBratney, 1999; Fleming *et al.*, 1999; Whelan and McBratney, 2000; Cupitt and Whelan, 2001; Stewart and McBratney, 2001; Whelan *et al.*, 2002b; Basnet *et al.*, 2003) have described a management zone as being fairly homogeneous in terms of yield (either being actual yield, potential yield, plant growth, growth conditions, or productivity). For this research, a LMU has been defined as “*an area of land, similar in terms of the physical characteristics and production capabilities, which can be managed uniformly*”.

To begin to identify a method and inputs used to classify LMUs, reviews of predictive methods of determining digital soil property maps (McBratney *et al.*, 2003; Scull *et al.*, 2003) were consulted. These reviews have discussed the origin of work by Jenny (1941), who defined five factors for soil formation, the model taking the mathematical form;

$$S = f(cl, o, r, p, t...) \quad (2.1)$$

where  $S$  is a soil property,  $cl$  represents climate,  $o$  stands for organisms,  $r$  relief,  $p$  parent material and  $t$  time. Numerous surveyors have used this framework as a method for understating the qualitative list of important factors in soil formation (McBratney *et al.*, 2003). The general equation is somewhat difficult to solve, but the model encouraged pedologists to adopt an experimental approach to field studies

(McKenzie *et al.*, 2000). Several researchers have adopted the model, holding factors such as time and organisms constant (McKenzie *et al.*, 2000; McBratney *et al.*, 2003).

In the same vein, a LMU will be a function of influential factors and the LMU model will take the mathematical form;

$$LMU = f(x_1, x_2, x_3, \dots, x_n) \quad (2.2)$$

where *LMU* is a land management unit,  $x_1$  to  $x_n$  are input landscape attributes that can be recorded cheaply and have a dominating influence on within-paddock spatial variability of productivity. Importance not only lies in the  $x_1$  to  $x_n$  factors, but the methodology utilised to combine the data layers to create meaningful, useable LMUs.

The following review collates several methodologies that have been applied for the delineation of fairly homogenous spatial zones within a landscape. These “spatial zones” have been derived for various purposes such as management zones, soil mapping, yield maps, salinity maps and vegetation classification to name a few, where the outputs are a homogenous class for the purpose of the investigation. The importance here lies with the methodology utilised to incorporate several spatial data layers to model the spatial zones.

### 2.1.1 Methods of Classification of Spatial Zones

In a review of approaches for making digital soil maps McBratney *et al.* (2003) discuss several predictive methods, which include, linear models, generalised linear models (GLMs), generalised additive models (GAMs), tree models (classification and regression), neural networks, fuzzy systems, other methods (genetic algorithms, splines), strengthening models: bagging, boosting, expert (knowledge-based) systems, unsupervised classification and geostatistical methods. From the review of approximately 70 studies, McBratney *et al.* (2003) found that GLMs in the form of multiple regression to be the most frequently used followed by co-kriging. While the use of regression trees and neural networks was not common.

O'Brien (2004) included a description and comparison of potential modelling approaches in search of an applicable method for tropical forage selection. The



models included logistic regression, GLMs and GAMs, artificial neural networks, classification and regression trees, environmental envelopes, fuzzy rule-based methods and Bayesian probability models. O'Brien (2004) details these approaches and highlights their strengths and weaknesses. To identify methods for classifying spatial zones further, Scull *et al.*(2003) in a review of predictive soil mapping, classify the modelling methods into four main streams namely; geostatistical methods, statistical methods, decision tree analysis and expert systems.

The classification of the landscape into spatial zones has also been used widely in remote sensing applications. Jensen (1996b) discusses several image classification techniques which are characterised into four main approaches; supervised classification, unsupervised classification, fuzzy classification and methods that incorporate ancillary data in the classification process. The methods that incorporate ancillary data are of particular interest in this research, and they are further subdivided into a) geographical stratification in which ancillary data is used prior to classification to subdivide the image into strata, b) classifier operations, which incorporate ancillary data during the image classification process and c) post classification sorting, which further subdivides an image based on some set of rules after an initial classification has been done. A combination of these methods that incorporate ancillary data are d) layered classification and e) expert systems.

The reviews of McBratney *et al.* (2003), O'Brien (2004), Scull *et al.*(2003) and discussion of Jensen (1996b) have highlighted numerous approaches which could be considered in this research. However, there was not one methodology that could be identified from these reviews as the most promising for this research. Subsequently, studies that classify the landscape into spatial zones have been tabulated (Table 2.1) identifying their model/method used, input data and scale of study, to assist in pinpointing a method to be used to classify LMUs in this case.

Table 2.1 Summary, in chronological order of methods for classifying spatial zones

Spatial Zone		Model	Input Landscape Attributes	Scale Study Area	Authors
Name	Attribute				
Non-hierarchical classes	Soil properties	Spatially constrained cluster analysis	-21 soil properties	1400m x 600m (paddock scale) 100m pixels	(Oliver and Webster, 1989)
Land Cover mapping	Land cover classes in mountain environment	Three-stage Classifier: 1. Quadtree segmentation and homogeneity 2. Minimum distance to means 3. Ancillary data and spectral curves	-SPOT satellite image -Elevation Model: geomorphometric measures of elevation, slope angle, incidence value	National park scale 11 classes	(Franklin and Wilson, 1992)
Geomorphic surfaces	Geomorphic Classes at three levels. -process domain(lv-1) -landform(lv-2) -sublandform(lv-3)	Discriminate functions: linear discriminant analysis.	-SPOT HRV: Spectral , Spatial (Texture analysis of SPOT HRV -Geomorphometric variables (DEM)	-process domain(lv-1) 4 classes -landform(lv-2) 7 classes -sublandform(lv-3) 16 classes.	(McDermid and Franklin, 1994)
Dryland Salinity Risk	Current and future salinity	Decision Tree and rule based classifiers (c4.5)	-Multi-temporal Landsat TM -DEM-derived variables -quantified expert knowledge	Catchment scale 2 classes	(Evans <i>et al.</i> , 1996)
Forest Ecosystem Classes	Vegetation types and soil types	Linear discriminant analysis	-High resolution compact airborne spectrographic Imagery (CASI) -Geomorphometric variables	Landscape scale (1:20,000) classes	(Treitz <i>et al.</i> , 1996) (Treitz and Howarth, 2000)
Management Zones	Crop growth properties. Height, biomass, yield	Unsupervised Classification	-Airborne Videography Green, Red and NIR	2 paddocks (approx 6ha and 11ha) 3 zones	(Yang and Anderson, 1996)

Spatial Zone		Model	Input Landscape Attributes	Scale Study Area	Authors
Name	Attribute				
Management Units	Yield – indicating soil variations	Fuzzy Clustering	-Multi-temporal yield data. 2-4 years	5 farms, 23 paddocks in total. (Mean paddock size 10ha) 2-7 zones	(Stafford <i>et al.</i> , 1998)
Management Units	Growth conditions	Mechanistic simulation model	-Detailed soil inventory (chemical and physical properties) -climatic records	2 paddocks (10ha and 15ha) 13 zones	(van Alphen and Stoorvogel, 1998)
Management Zones	Stable yield estimate zones	“FuzMe” Modified fuzzy k means	-11 years Satellite based yield estimates “Farsite” based on Landsat TM acquired mid season	2 paddocks (each 182ha) 2 zones	(Boydell and McBratney, 1999)
Management Zones	Productivity	Hand drawn vector lines	-Soil colour aerial photography -topography -expert knowledge	1 paddock (71ha) 3 zones	(Fleming <i>et al.</i> , 1999)
Dryland Salinity Risk	Predicted salinity risk (predicted discharge areas) and non-risk areas	Decision Tree Classifier (c5.0)	-Multi-temporal Landsat TM -DEM-derived variables -expert knowledge	Catchment scale 2 classes	(Evans and Caccetta, 2000)
Land Cover Mapping	Land cover classes	Multi-step method. 1. Unsupervised classification. 2. Aggregation to raster polygons 3. Supervised, nonparametric classification 4. Optional postclassification sorting	-Landsat TM image classification -DEM – topographic attributes	1700km <sup>2</sup> 12 classes	(Wheatley <i>et al.</i> , 2000)

Spatial Zone		Model	Input Landscape Attributes	Scale Study Area	Authors
Name	Attribute				
Crop Management Zones	Significant differences in yield; differences in influential soil properties on yield	Multivariate k-mean clustering	-Yield -3 years -Soil EC -Elevation data	1 paddock (100ha) 3 zones	(Cupitt and Whelan, 2001)
Vegetation classes and subsequent bird abundance	Over and understorey vegetation species	Spatially constrained clustering	-9 vegetation variables	45ha 50m pixels	(Hall and Maruca, 2001)
Management Zones	Soil properties	SEC continuous sensor correlated with other soil properties	-Soil electrical conductivity	1 paddock (10ha) 13 zones	(Nehmdahl and Greve, 2001)
Management Zones (Contiguous)	Soil properties	Hard-k-zones algorithm	-Yield – 3 years -Soil OC and K	1 paddock (17ha) 3 – 4 zones	(Shatar and McBratney, 2001)
Management Zones	Yield potential	1. Maximum Likelyhood classification on soil colour. i.e. soil type 2. Unsupervised Classification using k-means clustering	-ETM+ of Landsat 7 30m pixels	2 paddocks (120ha and 150ha) 3 zones	(Stewart and McBratney, 2001)
Tropical forest ecosystems	<i>Selvas</i>	Spatially constrained cluster analysis	-Measures of vegetation activity from multi-temporal AVHRR imagery -water balance variables modelled in GIS -elevation from DEM	Regional scale 8km resolution	(Mora and Iverson, 2002)
Management Zones	Yield performance	User-defined fuzzy set membership function	-Standardised Yield (4 years)	1 paddock (40.5ha) 3 zones	(Basnet <i>et al.</i> , 2003)

Spatial Zone		Model	Input Landscape Attributes	Scale Study Area	Authors
Name	Attribute				
Yield zones	Yield performance	Ward's clustering k-means and fuzzy k-means clustering	-5 years of yield data	1 paddock (62.7ha)	(Dobermann <i>et al.</i> , 2003)
Management zones	Yield response	Combination of standardised yield values	-5 years of yield data	1 paddock (70.8ha)	(Heermann <i>et al.</i> , 2003)
Management Zones	Yield potential	Multivariate K-means clustering	-2 x Yield -EC -Elevation data	1 paddock (75ha) 2 and 3 zones	(Whelan and McBratney, 2003)
Land Cover classes	Managed grassland, woodland, rough grassland	Spatially weighted supervised classification; Method of incorporating spatial weighting into supervised classification	-Simulated remotely sensed imagery	Simulated 64 x 64 pixels	(Atkinson, 2004)
Management classes	Yield for potential management	Fuzzy k-means on 3 paddocks, then multiple discriminant analysis across farm	-Radiometric thorium, total counts, potassium, uranium - Soil EC - elevation, slope and CTI	3000ha farm 7 classes	(Florin <i>et al.</i> , 2005)
Management Zones	Yield response Nitrogen levels	HERMES simulation		2 fields (20ha each)	(Kersebaum <i>et al.</i> , 2005)
Soil property maps	Soil attributes	Linear regression ordinary kriging plus regression simple kriging incorporating ancillary data	-Soil properties: Texture, OM, pH, K -Bare soil aerial colour photograph. BGR wavebands	35m x 20m grid of 86 sample points (6ha)	(López-Granados <i>et al.</i> , 2005)

Spatial Zone		Model	Input Landscape Attributes	Scale Study Area	Authors
Name	Attribute				
Management zones	Soil or landscape conditions Yield limiting factors	Fuzzy cluster analysis	3 approaches: a) relative elevation, organic matter, slope and EC b) 6 years yield data c) relative elevation, organic matter, slope and EC, yield spatial trend, yield temporal stability.	2 fields: 32.8ha: 5 classes 12.5ha: 4 classes	(Miao <i>et al.</i> , 2005)
Agricultural Management Zones	Soil properties and yield	Cluster using filtering and fuzzy k-means	-Elevation, slope, aspect, drainage area	1 field (9.8ha) 5 zones	(Pilesjö <i>et al.</i> , 2005)
Management Units	Soil-Landscape features	Spatially Constrained classification	-Soil map. -relative elevation, slope -soil EC -soil surface reflectance	52ha: 28 units merged to 6 units	(Simbahan and Dobermann, 2006)

Table 2.1 lists a literature review of classification methods for producing spatial zones highlighting their methodology, input factors and scale of study. As a summary of the works cited in Table 2.1, it can be observed that two thirds of the works are applied at paddock scale. The main focus of the review targeted paddock scale studies because the approach in this research will endeavour to provide LMUs that subdivide a paddock, however, in this case it will be implemented for the entire farm. As such, other catchment and regional scale applications have also been included to review methods and input data sets at those scales. The approach by Florin *et al.* (2005) is the only farm scale method, and they make the point that whole farm applications remain elusive. Their approach uses a method that firstly forms zones on three paddocks and then allocates the classes across the entire farm.

Table 2.1 also shows that the four main input landscape attributes are soil properties, yield data, remote sensing data and topographic attributes. Soil properties have been used on more than one quarter of the techniques, almost all of which are paddock scale applications. Yield data have been used in approximately one quarter of the techniques, again all of which are paddock scale applications. Remote sensing data (recording vegetation attributes or soil properties) are used in half of the techniques, with over 50 percent of these being farm, catchment or regional scale applications. While topographic attributes are also used in almost half of the techniques with over 50 percent of these being farm, catchment or regional scale applications. It can also be summarised that all farm, catchment or regional scale applications used remote sensing and topographic attributes, which is not surprising as it is more common that such data are widely available at these scales as compared to more expensive point based data (i.e. soil properties).

Over one third of the methods included in Table 2.1 are either k-means clustering or fuzzy clustering (which is more often fuzzy k-means). Other more popular methods are spatially constrained clustering, decision tree classifiers and discriminant analysis. The k-means and fuzzy clustering methods are all used for paddock scale applications while spatially constrained clustering has been used at paddock and regional scales.

From the studies included in Table 2.1 the spatially constrained classification technique has a number of characteristics which make it the best method for this research. As for a number of methods, it can be used at paddock and regional scale and has the ability to incorporate multivariate data. The pivotal feature that makes it the best method listed is that it incorporates spatial relationships between objects into the classification, which group's objects that are similar in terms of their attribute data. The resulting groups are spatially compact homogenous zones. While other methods can produce homogenous groups, they can be spatially scattered making them impractical for farm management. This problem is usually solved using post classification filters which inevitable introduce greater heterogeneity into the zones. The spatially constrained classification optimises the formation of zones with respect to both attribute homogeneity and spatial compactness, forming zones that can be practically managed by farmers. More details regarding the spatially constrained classification are discussed hereafter.

#### 2.1.1.1 Spatially Constrained Classification

Similar to other classification techniques, spatially constrained multivariate classification (Urban, 2004) allocates samples to clusters based on their similarity with respect to measured variables (i.e. soil or terrain), but also considers their proximity to one another. This forces the clusters to be homogenous in terms of their variables, as well as, geographically contiguous (Urban, 2004). The technique is based on a similarity matrix of size  $n(n+1)/2$  ( $n$ =number of sample points), which becomes very large when classifying high resolution data sets (i.e. >100,000 individuals). Urban (2004) suggests that spatially constrained classification shows some promise, but is not yet computationally feasible for large datasets. In addition he mentions that while the classifications can yield patches that are better resolved spatially, they may be less well defined ecologically (i.e. on their measured variables). The use of spatially constrained classification will depend on the focus of the application. For example Caeiro et al (2003) applied three different classification approaches; (1) spatially constrained clustering followed by indicator kriging; (2) discriminant analysis; (3) a hybrid of 1 and 2, discriminant analysis with indicator kriging, and found method 1 that used the spatially constrained clustering to produce the more realistic pattern that was in better agreement with estuary behaviour (the measured variable).



There have been several differing procedures for incorporating the spatial weighting into the classification which are presented here. Legendre and Legendre (1998) offer a detailed description for spatially constraining a clustering methodology based on the similarity matrix, while Legendre (1987b) (cited in Legendre and Legendre, 1998) suggests weighting the values in an ecological similarity matrix by a function of geographic distance among sampling locations, prior to clustering. McIntire and Fortin (2006) used spatially constrained clustering which was constrained to only those plots that were contiguous to determine the structure of a boundary zone created by fires and mountain pine beetle. Oliver and Webster (1989) used a univariate variogram as the spatial weighting function, while Bourguault *et al.* (1992) applied a multivariate variogram or covariogram. The methods that utilise the variogram are attractive as the spatial weighting incorporated is also modelled on the inherent spatial structure of the variables in that study area. Mora and Iverson (2002) developed a spatially constrained ecological classification based on Oliver and Webster's (1989) rationale. They identify that the spatially constrained technique forms not only ecologically different clusters but also indicates their pattern of distribution, which is more useful for landscape ecologists. Gordon (1996) reviewed constrained classification methods, not restricted to those constrained spatially, and discussed that the most useful developments in this area would be on methods for assessing the results.

Spatially constrained classification has also been examined by Atkinson (2004) who incorporates a spatial weighting into a supervised classification of remotely sensed data. The idea came from Oliver and Webster (1989) and Atkinson (2004) concurs with Urban (2004) stating that the reason why Oliver and Webster's (1989) method is unsuitable in remote sensing applications is that the method requires the use of a distance matrix, which would not be viable using remote sensed images as they generally contain millions of pixels. Atkinson's new approach (which modifies the feature-space distance-based metric with a spatial weighting) shows potential with only a modest increase in computing time; however it was based on a relatively small (64 x 64 pixels) simulated data set.

In summary to the above works reviewed, spatially constrained classification which incorporates the spatial weighting based on the variogram is the most appealing for this research. As mentioned (Mora and Iverson, 2002; Caeiro *et al.*, 2003; Urban, 2004) spatially constrained classification produces units that are better resolved spatially and produces more realistic patterns to natural phenomena than other classification techniques. In this research homogenous units are a priority and as such if the computational feasibility can be overcome, the method is appropriate. In particular, the Oliver and Webster (1989) approach is suitable for several reasons; (i) it can be applied to multivariate data; (ii) it uses both a distance component and variograms to model the spatial structure of the variability of the properties; and (iii) it produces homogeneous units.

## **2.2 Landscape Attributes Required to Form LMUs**

The three main attributes that appear to drive the majority of the LMU classification techniques are soil, vegetation and topography. One of the requirements of this research is that the LMUs are based on stable soil properties, as soil properties provide information on the ability of the ground to support plant growth and are therefore particularly useful for studies that endeavour to classify paddocks into zones of differing crop yield potential. In particular, stable soil properties do not change significantly over a growing season and will not alter through the addition of fertilizers, thus the LMUs resulting from them are assumed to be more “stable” over time. Vegetation attributes are predominantly useful in studies that are classifying landscape into vegetation classes, moreover they provide a surrogate measure of landscape variability or change, based on the relationship between a sites’ properties (i.e. geology, soil and topography) and vegetation type or growth status. And topographic attributes have been shown to be related to soil properties. These three attributes are discussed individually in the following sections.

### **2.2.1 Soil Properties**

Soil supports growing plants by exchanging gases with the atmosphere and supplying water and nutrients. A plant root system is required to carry out these functions and, as such, properties that affect root growth will also need to be considered. If soil is non-limiting to plant growth, that is, it can supply all these requirements without degrading, the potential yield of plant growth will be governed solely by climate. In

south-western Australia most soils (>80 percent) are not in this category due to physical and/or chemical limitations that require management to minimise their effect on crop yields (Moore, 1998c).

For this research soil properties that will be considered appropriate for inputs to the LMU classification are those that are stable over time and affect plant growth and therefore yield potential. Following, is a review of several works to determine what the stable soil properties are, and of those, the ones that govern plant growth.

Moran *et al.* (1997) in a review of the opportunities and limitations for image-based remote sensing in precision crop management, segregates information needs into seasonally stable and seasonally variable conditions. *Seasonally stable conditions are those that are relatively constant through the crop growing season, such as yield based or soil-based management units, and only need to be determined pre-season and simply updated, when and if necessary. While seasonally variable conditions are those that change continually within the season, such as soil moisture, weed or insect infestation, and crop disease, and need to be determined numerous times during the season for proper management.* Moran *et al.* (1997) mention maps of soil fertility and physical attributes are being used in precision crop management approaches to determine the responsive and non responsive parts of the field in relation to seasonally stable management units. They cite Nielsen *et al.* (1995) who identified soil nitrogen, soil organic matter, relative position and slope of the terrain, and Pierce *et al.* (1995) who suggested that soil physical properties or landscape may have a higher priority than soil fertility in explaining yield variations. Moran *et al.* (1997) discuss eight seasonally variable soil and crop condition variables namely; soil moisture content, crop phenology, crop growth, crop evaporation rate, crop nutrient deficiency, crop disease, weed infestation and insect infestation with a discussion on remote sensing techniques for monitoring these.

Atherton *et al.* (1999) tabulate several factors that may be used to explain yield variability including cultural, soil, topography, pest management and climate. Not all of these factors can be managed equally well and the cost of managing them will vary from location to location. The soil factors include, water holding capacity, texture, structure, nutrient availability, organic matter content, cation exchange

capacity, pH, compaction, topsoil depth, depth to restricted layer, water table depth, infiltration rate and hydraulic conductivity.

Crop yield (or the income derived from it) has the greatest bearing on farm management and practices at present (Cupitt and Whelan, 2001). As such the importance of assessing the physical and chemical properties of soils becomes evident for optimising the farming system. The variability of soil attributes and in turn crop production potential, at a given site and time is controlled by several important processes. Geology and pedology that define the soil type and govern the majority of static soil properties are the most influential (Whelan, 1998). Through literature Whelan (1998) makes a segregation of the soil attributes into those that are either static or dynamic in a manner similar to Moran's *et al.* (1997) seasonally stable or variable parameters, identifying that the magnitude of variability is generally lower in those that are static. He mentions texture, horizon colour and cation exchange capacity as static soil properties that are governed by geology and pedological processes, while soil management practices and cropping systems can greatly influence the properties that are more dynamic such as nutrients, water, air and solute regimes. Reiterating Moran's *et al.* (1997) seasonally variable crop condition parameters, Whelan (1998) also makes the connection between within field crop variation being a function of crop insect pest and diseases and weed infestation, which are all important limitations on yield.

Whelan and McBratney (2003) mention that many studies have shown the dominant influences on yield variability (other than climate) are the more static soil physical factors such as soil texture, associated structure and organic matter which indirectly influence cation exchange capacity, nutrient availability and moisture storage capacity of the soil.

Fitzpatrick *et al.* (1999) list six key soil morphological descriptors for assessing the quality of soil in dryland farming regions, namely depth changes in consistence, colour, texture, structure, segregations/coarse fragments (carbonates and ironstones) and abundance of roots in different layers. Cass (1999) intends to propel Fitzpatrick's *et al.* (1999) ideas into *a more quantitative set of physical descriptors of how soils store, supply and regulate the essential physical plant growth requirement,*

mentioning that this involves an interpretation on how static physical properties, dynamic hydrological and mechanical processes influence plant growth requirements. This is difficult to achieve because the measurement of physical properties is demanding on time and sampling procedure, and often few physical properties are routinely measured in commercial laboratories (Cass, 1999). In summary, Cass (1999) adopted the following arbitrary list of properties that might constitute a statement of the physical fertility of soil. Infiltration rate, total plant-available water storage, air-filled porosity at the wettest drained condition, penetration resistance at the wettest drained condition, structural stability to wetting, the balance of salinity to sodicity. Essentially, these properties act as a surrogate measurement of soil structure and resistance to structural breakdown (Cass, 1999).

In summary to the above mentioned soil properties derived from the literature, it appears that the stable soil properties that are important to consider in terms of plant growth are topography, texture (and coarse fragments), structure, organic matter, nutrient availability, pH, salinity / sodicity balance, depth of topsoil and depth to restricted layer. The following sections further discuss the relevant properties in this research.

#### 2.2.1.1 Soil Texture

Soil texture describes the proportions of sand, silt and clay (the particle size distribution). Sands are mineral particles with a size range of 2 to 0.02mm, silt 0.02 to 0.002mm and clay particles are smaller than 0.002mm (Moore, 1998a). Field texture is a measure of the behaviour of a small handful of soil when moistened and kneaded into a ball (bolus) and then pressed out between thumb and forefinger to form a ribbon (McDonald *et al.*, 1990). The behaviour of the soil during bolus formation and the ribbon length determine the field texture grade (Moore, 1998a). There is only an approximate relationship between field texture and particle size distribution analysis performed in the laboratory, as factors other than sand, silt and clay influence field texture (McDonald and Isbell, 1990), such as organic matter, clay mineralogy, amount of sodium and the presence of calcium carbonate (Moore, 1998a). Texture groups from particle size distribution in WA are based largely on the clay content, as the silt content is often small (<15 percent) (Purdie, 1998).

Moore (1998a) provides the following *Texture grades* together with approximate clay content (Table 2.2).

Table 2.2 Texture grades based on clay content (adapted from Moore, 1998a; Purdie, 1998)

Texture Group		Approximate Clay Contents
Sands	Sand	<5%
	Loamy sand	about 5%
	Clayey sand	5-10%
Loams	Sandy loam	10-20%
	Loam	about 25%
	Sandy clay loam	20-30%
	Clay loam	30-35%
	Clay loam, sandy	30-35%
Clays	Light clay	35-40%
	Light medium clay	40-45%
	Medium clay	45-55%
	Medium heavy clay	>50%
	Heavy clay	>50%

Moore (1998a) mentions that texture is useful for predicting soil behaviour in terms of water availability (profile hydrology, available water capacity, waterlogging) and erodibility. It is also significant in relation to other soil properties, including water repellence, nutrient deficiencies, nutrient leaching, subsurface compaction, soil structure decline and pH buffering capacity.

#### 2.2.1.2 Coarse fragments

Fragments within the soil impose limitations on rooting volumes which in turn diminish the soils capacity to supply nutrients and water to the plants, while surface fragments can effect cultivation and harvesting of crops (Hazelton, 1993). However, in dense soils, the spaces between the coarse fragments may form pathways for water drainage and root penetration (Brady and Weil, 1999b) and therefore assist plant growth.

Coarse fragments are described as the size of particles greater than 2mm. Coarse fragments are also referred to as gravel and gravelly soils are often discussed as superior or inferior to soils that are similar but gravel is not present (Moore, 1998a)

depending on where in the profile the gravel is present. Large gravel content reduces the effective soil volume, but it inhibits root growth only when the gravel is cemented (Moore, 1998a). The proportion of gravel in the soil can provide an indirect indication of several factors; (i) the presence of clayey subsoil deeper in profile which can provide water storage, (ii) in the A-horizon the texture can contain slightly more clay in some soils and, thus improve nutrient and water storage slightly, (iii) gravelly soils are generally high in the landscape providing favourable implication for drainage, (iv) another common property is their ability to ‘fix’ phosphorus, (v) and a large proportion in the A-horizon is likely to reduce the susceptibility to traffic compaction (Moore, 1998a; Smolinski, 2005). Gravel over clay soils are also usually better structured and non-sodic (Smolinski, 2005).

#### 2.2.1.3 Soil Stability

The decline of soil structure can lead to hardsetting and crusting and the effects of this decline can reduce crop yield directly or indirectly via reduced infiltration, poor soil workability, delayed seeding, reduced seedling emergence, reduced aeration and reduced trafficability (Needham *et al.*, 1998). Needham *et al.* (1998) describes an ideal soil structure as having a *large proportion of aggregates from 0.5 to 2mm, which are not easily broken down. It has high porosity for water entry and gas exchange, low strength, is stable in wet or mechanically disturbed, and aggregates can reform if subjected to adverse management.*

A soil stability score can be derived from laboratory analysis and used as a general guide to stability. It is based on four soil properties; sodicity, organic matter, calcium:magnesium ratio and electrical conductivity (Needham *et al.*, 1998). Table 2.3 provides an example for a soil with a sandy loam or finer surface texture, (i.e. 10 percent clay or more as according to Table 2.2).





made by these organisms. Plants are the primary source of all soil organic matter and animal matter is secondary (Brady and Weil, 1999c).

Organic matter plays several roles in relation to the physical and chemical properties of soils and therefore needs to be considered here because of the flow on effect to plant growth. Davey (1990) lists several reasons why the structure of soil needs to be stable and well balanced; (1) the bulk density does not limit plant growth, (2) the soil is well-aerated, (3) available water storage is at a maximum, and (4) the soil will withstand cultivation and wind and water erosion. Organic matter influences all of these properties because of its role in the development of soil structure and its action in cementing soil structural units. Nitrogen for plant growth in unfertilised soils comes mostly from organic matter through microbial decomposition (Davey, 1990).

A widely used method in Australia is the measure of organic carbon (Walkley and Black). This is reported as Org C (W/B) and expressed as %C. Table 2.5 can be used to apply a rating to the percentage of organic matter by combustion values.

Table 2.5. Percentage of organic matter ratings at surface horizon (A1) (adopted from Purdie, 1998)

Rating	Percentage of organic carbon. Org C (W/B)	Soil properties
Low	< 1%	Poor nutrient storage, unstable structure
Medium	1 - 2%	
High	> 2%	Good nutrient storage, stable structure

#### 2.2.1.5 Cation Exchange Capacity

By definition the cation exchange capacity (CEC) is a measure of the total capacity of a soil to hold exchangeable cations (Rengasamy and Churchman, 1999), which control the fertility and structural stability and hence the productivity of soils (Davey, 1990). The clay content and minerals contribute to the CEC of soils (Davey, 1990; Purdie, 1998). For most soils the bulk of the cations present in exchangeable format are  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^{+}$  and  $\text{K}^{+}$ . An effective cation exchange capacity (ECEC), which is the sum of the basic cations (Ca, Mg, Na, K) and exchangeable/titratable acidity ( $\text{H}^{+}$ ,  $\text{Al}^{3+}$ ,  $\text{Mn}^{2+}$ ), can be calculated to represent the total amount of exchangeable cations held by a soil (Rengasamy and Churchman, 1999). The acidity occurs in two forms; hydrogen ion acidity ( $\text{H}^{+}$ ) and metal ion acidity ( $\text{Al}^{3+}$ ,  $\text{Mn}^{2+}$ ) (Allen, 2004).

The CEC is usually measured at pH 7 or 8.5 and for neutral/alkaline soils the ECEC is usually very close to measured CEC (Allen, 2004). However, for acidic soils the situation is a little different and ECEC is usually intended to approximate the CEC measured at pH 7 (Allen, 2004).

As most soils contain a proportion of pH-dependant electrical charge, there is often discrepancy between the CEC and ECEC (McArthur, 1991). CEC values determined for soils at a pH value largely different to that which occurs in the field may be misleading (Rengasamy and Churchman, 1999), as the amount of  $H^+$  is partly related to the pH of the soil. The greater the difference between pH 7 and field pH, the greater the  $H^+$  acidity (Allen, 2004). The Chemistry Centre of WA data show that the metal ion acidity ( $Al^{3+}, Mn^{2+}$ ) is usually much larger than  $H^+$  acidity, and the  $H^+$  acidity is usually  $<0.05me\%$  if the  $pH > 4.5$  and thus they are fairly confident that ignoring the  $H^+$  acidity (which can be measured by a tedious method) will provide adequate accuracy (Allen, 2004). The ECEC can be calculated as the sum of  $(Ca+Mg+Na+K+Al+Mn)$  (Allen, 2004). If two soils contain similar clay minerals and a large proportion of permanently charged components, then a larger CEC generally indicates a larger proportion of clay-size particles (Rengasamy and Churchman, 1999). The type of clay mineral has a strong influence on the CEC, for soils with high proportions of permanently charged components and if clay content is known then CEC values can be used to infer the type of clay mineral present (Rengasamy and Churchman, 1999).

Table 2.6 rates soils based on their chemical properties and indicates factors that may need to be considered in managing soils. For CEC values, Purdie (1998) suggests that the ratings seem appropriate for WA soils in an international context where high and low values indicate that the soil has good and poor capacity to store nutrient cations, respectively. For this reason, these ratings are applied in this research to assist in describing the soils ability to store nutrients.

Table 2.6 Ratings for chemical properties (adopted from Purdie, 1998)

Analysis		Ratings		
		Low	Medium	High
CEC measured	(me%)	<5	5-15	>15
exchangeable cations	Ca <sup>2+</sup> (me%)	<5	5-10	>10
	Mg <sup>2+</sup> (me%)	<1	1-5	>5
	K <sup>+</sup> (me%)	<0.5	0.5-1.0	>1.0
	Na <sup>+</sup> (me%)	<0.3	0.3-1.0	>1.0
	Al <sup>3+</sup> (me%)	<0.1	0.1-1.0	>1.0
	Mn <sup>2+</sup> (me%)			

### 2.2.1.6 Soil Acidity

Soil acidification is a process that naturally occurs within soils, however it is accelerated by agriculture. The acidification rate is variable and depends on the soil type, land use, productivity and management of the farming system. The activity of soil acidity is one measure used for soil acidity. It refers to the concentration of hydrogen ions in soil solution of which pH is an indicator (Moore *et al.*, 1998).

The pH can be measured in distilled water (pHw) and calcium chloride (pHca). Using calcium chloride as the dilute salt solution, neutralises the effect of variable soil to solution ratio, and represents the ionic strength of most agricultural soils (Davey, 1990). Although pHw is considered closer to what the plants roots would be exposed to, it is subject to variation due to seasonal change in soil moisture (Slattery *et al.*, 1999). The measured pH in calcium chloride may represent field conditions better than when measured in water (Davey, 1990).

Aluminium toxicity is a major problem associated with acid soils (Moore *et al.*, 1998). It is usually not until soil pHca falls below 4.8 that Al becomes measurable and can become toxic to plants (Slattery *et al.*, 1999). Slattery *et al.* (1999) discuss the results from several studies of the relationship between Al and grain yield or pasture growth highlighting that across a range of soil types exchangeable Al is not a particularly good predictor of plant response, while Al saturation is more generally applicable, but still varies between locations. If the analysis is limited to a particular soil type or group, correlations between exchangeable Al or Al saturation, with yield response are reasonably good.

### 2.2.1.7 Soil Alkalinity, Sodicity and Salinity

Soils that are alkaline have a  $\text{pH}_w > 7.5$  (measured in a 1:5 soil:water suspension) in one or more layers and can have a number of nutrient deficiencies (Scholz and Moore, 1998). High alkalinity leads to sodicity in soils however not all sodic soils are alkaline (Scholz and Moore, 1998). The  $\text{pH}_w$  is a better method for differentiation of the alkalinity than  $\text{pH}_{\text{Ca}}$  (Scholz and Moore, 1998). Possible limitations of alkaline soils can be inferred from  $\text{pH}_w$  and the  $\text{EC}_{\text{SE}}$  (electrical conductivity of a saturation extract) (Scholz and Moore, 1998).

Soils that have large concentrations of exchangeable sodium are referred to as sodic soils (Scholz and Moore, 1998). Sodicity can have direct toxic effects on plants (Scholz and Moore, 1998) and the effect of sodicity on soil stability has been shown in Table 2.3 (i.e. higher sodicity can lead to decreased stability). If the exchangeable sodium percentage (ESP) of soils in Australia is between 6 to 15 they are called sodic, and highly sodic if greater than 15 (Scholz and Moore, 1998).

On the other hand, soil salinity refers to soils that have a large concentration of soluble salts which is usually measured by the electrical conductivity of a 1:5 soil:water suspension ( $\text{EC}_{1:5}$ ). Although  $\text{EC}_{1:5}$  does not reflect salt content directly,  $\text{EC}_{\text{SE}}$  (electrical conductivity of a saturation extract) is more meaningful as it accounts for soil texture i.e. being closer to field water content (Shaw, 1999) and plant response (Moore, 1998b) but is a more tedious method (Shaw, 1999) and thus often not practical. Shaw (1999) describes a conversion method from  $\text{EC}_{1:5}$  to  $\text{EC}_{\text{SE}}$ , which requires the air-dry moisture content and saturation percentage, while Moore (1998b) provides the relationship developed by George and Wren (1985) (as cited in Moore, 1998b p. 153) that includes the saturation point of soil, which can be estimated from a relationship with texture. Moore (1998b) suggests using this when it is not practical to measure  $\text{EC}_{\text{SE}}$ .  $\text{EC}_{\text{SE}}$  values are 5 to 18 times higher than  $\text{EC}_{1:5}$  (Purdie, 1998). McArthur (1991) presents relationships between salinity categories based on  $\text{EC}_{1:5}$  and depth in reference soils, and Purdie (1998) adopted these ratings to provide a pointer to potential problems, not to be misinformed as a specific evaluation, and the associated effects on plant growth which would vary with species and salts present. Campbell and Bowyer (1990) provide an indication of the effect

on crop growth. Table 2.7 is a combination of the information provided by McArthur (1991), Purdie (1998) and Campbell and Bowyer (1990) which will be useful in this research to describe the soils.

Table 2.7 Rating and salinity categories and associated effects (adopted from Campbell and Bowyer, 1990; McArthur, 1991; Purdie, 1998)

<b>EC<sub>1:5</sub> (mS/m)</b>	<b>Salinity Category</b>	<b>Effect on Plant Growth</b>	<b>Effect on Crop</b>	<b>Rating</b>
< 16	Non-saline	Minimal effect on plant growth	Salinity effect negligible	Low
16-31	Slightly saline		Very sensitive crops affected	
32-45				
46-50	Moderately saline	Plant growth is inhibited	Yields of many crops affected	Medium
51-62				
63-90				
91-125	Highly saline		Tolerant crops may grown	High
126-200				
200-250	Extremely saline	Plant growth is severely restricted	Only very tolerant crops grown	
>250				

This section has provided a thorough description of selected stable soil properties. Field soil samples are collected at point locations and thus strategies for soil sampling along with collection via remote sensing approaches are presented in Section 2.3.

### 2.2.2 Vegetation

Vegetation or parent material can provide an indication of the underlying soil and geology and can subsequently be used to determine the substance and variability of the landscape. Traditional soil survey methods (McDonald *et al.*, 1990) use vegetation as an additional description of the landscape and since the onset of remote sensing techniques, vegetation attributes are used in landscape scale classification. Using high resolution remote sensing ( $\leq 10\text{m}$  spatial resolution), vegetation attributes can also be used to help detect variations in growing condition (i.e. variability) at within field scales. The recording of vegetation attributes within a paddock and their detection via remote sensing techniques are further discussed in Section 2.3.

### 2.2.3 Topographic Attributes

Topographic attributes can be divided into primary and secondary (or compound) attributes (Moore *et al.*, 1993a). Examples of primary attributes include elevation, aspect, slope and catchment area, all of which can be directly calculated from regular grid digital elevation models (DEMs) using GIS software or the TAPES-G program (Gallant and Wilson, 2000). Secondary topographic attributes are combinations of primary attributes and can be used to characterise the spatial variability of processes within the landscape (Moore *et al.*, 1993a). Several studies have correlated terrain attributes, derived from a DEM, to soil properties (Moore *et al.*, 1993a; Gessler *et al.*, 1995; Boer *et al.*, 1996; McKenzie and Ryan, 1999). In particular, McKenzie *et al.* (2000 pg 247) refer to several studies (Milne, 1935, Walker *et al.* 1968, Gerrard 1981,1990) that have shown close, although often complex, relationships between soils and landforms. Three topographic attributes namely; landforms, the compound topographic index (CTI) and slope, have been identified as important drivers in soil landscape studies and their derivation (i.e. collection) are further detailed hereafter.

## 2.3 Methods for Recording Landscape Attributes

Soil, vegetation and topographic attributes have been identified and discussed as the attributes required for classifying LMUs. The recording and compilation of these data, which can be achieved through; 1) on ground field sampling techniques, 2) remotely with the use of remote sensing applications, and 3) digitally derived through the computation of existing data, are discussed in that order, hereafter.

### 2.3.1 Field Sampling Techniques for Soil and Vegetation

#### 2.3.1.1 Soil

A soil sampling strategy needs to be developed in order to collect soil property information across the study area. Intensive soil sampling that results in detailed soil maps is costly and time consuming, and therefore, there is a need to reduce the degree of sampling without compromising the accuracy of the information gathered. Accordingly, other approaches for cost effective acquisition of soil property information at farm and paddock scales, which account for the spatial variability of soil properties inherent in agricultural paddocks, need to be explored.

When designing a sampling strategy several factors need to be considered; grouping and stratification, resolution or sampling frequency, intensity or sampling size and pattern of samples in order to determine the layout of points. Grouping, resolution and intensity all relate to variability. Therefore, there is a need to know something about the variability in advance, i.e. prior mapping of some related condition.

At medium scale the variability across a landscape of many soil properties is related primarily to differences in particular soil-forming factors, such as topography (drainage) or parent material (Brady and Weil, 1999a) which can therefore be useful for inferring the soil variability in advance. Several studies have correlated terrain attributes, derived from a DEM, to soil properties (Moore *et al.*, 1993a; Gessler *et al.*, 1995; Boer *et al.*, 1996; McKenzie and Ryan, 1999). Terrain analysis has the potential to improve soil surveys in three ways by (1) generating high-resolution environmental information of direct use in land evaluation (slope, net radiation, etc); (2) creating explicit environmental stratifications for survey design; and (3) providing quantitative spatial predictions of individual soil properties (McKenzie *et al.*, 2000).

In a preliminary analysis of the relationships between topography and soil attributes, Moore *et al.* (1993a) found that slope was the most highly correlated terrain attribute, and suggested that the relationship between terrain and soil attributes could be applied as a guide to sampling strategies. Furthermore, McKenzie *et al.* (2000) states that terrain analysis can be used to generate stratified random sampling schemes in a way that previously has not been possible in conventional soil survey. This indicates that topographic attributes can be useful in guiding the soil sampling strategy.

The high cost of collecting soil attribute data at many locations across the landscape has created the need for inferring air and water properties of soils using economical surrogates derived from soil morphological properties. This is demonstrated in studies by Gessler *et al.* (1995) and Moore *et al.* (1993a) which are based on deriving relationships between terrain values and soil properties. The most common surrogates used are soil texture, organic matter, structure, and bulk density (Moore *et al.*, 1993a). The rationale is that in many landscapes, catenary soil developments occur in response to the way the water moves through the landscape. Therefore, it

may be hypothesised that the spatial distribution of topographic attributes that characterise water flow paths also captures the spatial variability of soil attributes at meso-scale.

Moore *et al.* (1993a) suggest that the DEM data at an appropriate scale could be used to calculate terrain attributes that would enhance soil surveys as a source of soil attribute data. Their results indicate that slope and wetness index (compound topographic index) are the terrain attributes most highly correlated with soil attributes. The terrain attributes could also be used to segment the landscape into essentially stationary process zones, where attribute prediction may be done in a more statistically robust fashion. This indicates that DEM derivatives can potentially be used to enhance sampling designs for soil properties due the relationship between terrain values and soil properties, and therefore they will be investigated to be incorporated in the soil sampling strategy in this research.

Optimal soil sampling should characterise the variability within the sampling area, and at the same time maximise efficiency of soil sample points. The variance quad tree method by McBratney *et al.* (1999) demonstrates this concept. They use crop yield variability to direct sampling for soil attributes that are considered to influence crop yield, and suggest that the technique would also be suitable for designing a soil or crop sampling scheme based on elevation data and aerial or satellite imagery.

An existing layer of digital multi-spectral imagery could also provide some prior knowledge of variability. This is indicated by Lamb (2000) who states that beyond canopy cover closure, spatial variations in the biomass or vigour of the crop canopy itself may indirectly indicate variations in underlying soil structure, and hence, be useful for delineating soil zones. Lamb (2000) also states that regardless of the seasonal effect, the identification and exact location of soil zone boundaries, either directly or indirectly from qualitative multispectral imagery, is a useful basis for planning more detailed soil sampling and analysis.

Drysdale *et al.* (2002) based their soil sampling design on these concepts. This strategy increased the intensity of sampling points in areas of heterogeneous soil properties while sampling less intensively in areas of the paddock that appeared



homogeneous. A combination of slope data, derived from a DEM and change in NDVI (from high resolution airborne remote sensing), is used to provide a surrogate layer of the soil properties variation. This optimised the location of the sampling points by detecting within paddock soil variability prior to field sampling. Detecting the spatial variation in advance addresses the first three factors to be considered in sampling designs. The pattern of soil sampling points is considered in the following section.

#### 2.3.1.1.1 Soil Sampling Pattern

The location of the sample points can be critical for subsequent analysis. Ideally for mapping, samples should be located evenly over the area however, a completely regular sampling network can be biased if it coincides in frequency with a regular pattern in the landscape (Burrough and McDonnell, 1998a). Therefore, Burrough and McDonnell (1998a) note that statisticians suggest that some kind of random sampling is preferred for computing unbiased means and variances. However, completely random locations of sample points can lead to an uneven distribution of points leaving unwanted holes. Therefore, a good compromise between random and regular sampling is stratified random sampling (Burrough and McDonnell, 1998a). This method provides an adequate "even" coverage with a sufficient range of inter-sample distances. Dicks and Lo (1990) (cited in Jensen, 1996a, p249) also indicate that some combination of random and stratified sampling provides the best balance between statistical validity and practical application. This suggests that a stratified random sampling pattern should be used in this research to overcome the inherent problems of other patterns.

#### 2.3.1.1.2 Size of Support

The support is the name used in geostatistics for the area or volume of the physical sample on which the measurement is made. Therefore, while considering factors for field soil sampling, the size of support also needs to be determined. The soil properties sampled in this research will be analysed in conjunction with high-resolution remote sensing imagery in order to determine if relationships exist between the two. If the support sizes of both sets of observations are not matched it may be difficult to combine the data sets for modelling and spatial analysis (Burrough and McDonnell, 1998a).

When data collected on a given support are used to predict values of the same attributes at unsampled locations then the predictions refer to the locations that also have that support. Procedures such as bulking or spatial averaging are used to relate the observations to larger areas or volumes. Bulk sampling consists of several small samples taken within a defined area around a spatially located sample point, which are mixed together before analysis to homogenise the area. This is useful when data collected by different methods need to be combined such as soil and information collected by remote sensors. The remote sensors collect data, which are recorded as single grid values (pixel values), and the numbers recorded are area-weighted averages of the radiation received, so the pixel area (resolution) defines the size of support (Burrough and McDonnell, 1998a).

In this research the size of support will need to be considered for the resolution of the remotely sensed imagery. Sampling units smaller in area than a pixel will lead to unrepresentative observations of larger areas. Sample units equal to the area of one pixel are not recommended, because of the problems of accurate ground location in terms of an external coordinate reference system. Justice and Townshend (1981) offer a formula (Equation (2.3)) to calculate the minimum dimension of a sample area based on the pixel resolution and accuracy of its location. In general the minimum dimension of the sample area,  $A$ , chosen may be estimated as follows:

$$A = P(1 + 2L) \quad (2.3)$$

Where  $P$  = the pixel dimensions, and  $L$  = the accuracy of the location in terms of the number of pixels (Justice and Townshend, 1981). The accuracy in this scenario also relates to the differential global positioning system used to determine the spatial location.

In summary to the literature related to the design and implementation of a soil sampling strategy, several aspects have been considered. The optimisation of the strategy will place sample points with respect to the inherent variability in soil properties in the paddocks, in doing so, place points more intensively in areas of heterogeneous properties and less often in areas of similar soils. Without prior knowledge of the current soil variability, data sets that could provide a surrogate measure are terrain attributes, in particular slope, and the surface reflectance via remote sensing techniques depicting variability in plant/crop growth and/or soil.

### 2.3.1.2 Vegetation

Vegetation attributes can be recorded by several methods and intensities depending on the level of information the collector wants to gather. In principle, the sample intensity should be sufficient to characterise accurately the population under study (Cihlar *et al.*, 1987), the population in this case being agricultural crops. There has been little research in this area, and Cihlar *et al.* (1987) provide a comprehensive instruction manual for field collection of information on crops to be used in remote sensing studies.

Before sampling, the method of vegetation attribute collection is considered along with sample locations and intensity. The vegetation sampling will need to detect variability in crop growth. Transect sampling can achieve this by setting the length of transect and sampling point intervals *in situ* appropriate to the magnitude of crop variation. The advantage of transect sampling as apposed to other methods (i.e. grid/random) is its simplicity to be assembled in the field and spatially located in remotely sensed imagery. Statistical analysis between remote sensing data and vegetation attributes will be conducted on a minimum of 30 sampling points per crop type. This will simplify statistical calculations based on the assumption of normality.

Cihlar *et al.* (1987) describe procedures for characterising agricultural crops in remote sensing studies. Plant height, defined as the distance between the canopy of interest and soil surface, is recommended to be recorded by placing a measuring or jousting stick vertically near selected plants and measuring the total maximum height (to the top of canopy) a minimum of five times. Plant cover is defined as the proportion of the soil surface masked by the plant material. The authors suggest using a nadir-viewing photograph taken at 2m or higher and overlaying the photograph with a grid to estimate the percentage of cover. Visual estimates can also be useful, although they tend to be highly subjective and inconsistent (Cihlar *et al.*, 1987). Plant density is defined as the number of plant stems per unit area (Cihlar *et al.*, 1987). For row crops, it is recommended to count the number of stems along the row, and convert it into density using the average row width, while for randomly distributed plants, counting the number of stems in a predefined area is advised.

The leaf area index (LAI) is defined as the total area of green leaves (one side only) from plants within an area, divided by the ground surface area within which these plants grow, in units of percent (Cihlar *et al.*, 1987). Cihlar *et al.* (1987) suggest the most efficient approach is to use a leaf area meter which measures the area of leaves that pass through it. In addition, it is recommended that a minimum of five randomly selected plants be included in one sample and then the LAI calculated using the plant density value.

The basis of recording field vegetation attributes in this research will be to determine whether relationships exist between the crop growth properties and high resolution remote sensing data. And furthermore to establish if the remote sensing layer can be used as an input to the LMU classification, based on its ability to depict landscape variability. The methods of field vegetation collection detailed by Cihlar *et al.* (1987) appear appropriate for this research as they are focused on remote sensing applications.

### 2.3.2 Remote Sensing of Soil and Vegetation

Remote sensing offers a non-invasive method for detecting information about the features on the earth surface. Reflective optical radiation is defined as propagating electromagnetic energy with characteristic wavelengths between the 0.4 and 3 $\mu$ m. When optical radiation interacts with a surface, a portion of the radiation is either absorbed or transmitted; the remainder is said to be reflected. The ratio of the reflected radiation to the total radiation falling upon the surface is defined as reflectance (Baumgardner *et al.*, 1985). A soil sampling strategy will need to be implemented in this research and the use of vegetation indices (i.e. NDVI) derived from remote sensing of vegetation have been shown to be useful in soil sampling strategies (Drysdale *et al.*, 2002). The reflectance of the surface (either soil or vegetation) is a useful method of determining the characteristics of the landscape and as such, could be a layer to be included in a LMU classification. Accordingly the principles behind remote sensing of vegetation and soil and examples of the techniques used are discussed hereafter.

### 2.3.2.1 Reflectance of Vegetation

Vegetation is unique in its three segment partitioning (reflected, transmitted or absorbed) of optical radiation (Figure 2.1). In the visible part of the spectrum (400 - 700nm), reflectance is low, transmittance is almost zero, and absorptance is high. The plant pigmentation is the fundamental control of the energy-matter interactions with vegetation (Lusch, 1989).

Within the visible part of the spectrum, absorptance of the bluish (400-500nm) and reddish (600-700nm) wavelengths is pronounced by the chlorophyll compounds, which are dominant plant pigments. The absorption of solar energy is required by vegetation to support photosynthesis. The transmittance by vegetation is very low in the visible wavelengths as shown in Figure 2.1. Energy that has not been absorbed will be reflected, thus chlorophyll-bearing vegetation appears green as a result of its minor peak in the 500-600nm wavelength.

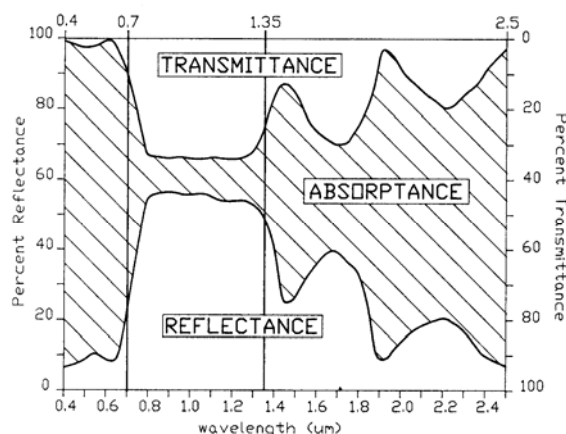


Figure 2.1 Spectral Partitioning of solar irradiance by vegetation (adapted from Gausman, 1985)(cited in Lusch, 1989)

Other plant pigments, the carotenes and xanthophylls, produce yellow or orange reflectance. These compounds have a single, broad absorptance band centred at about 450nm. While anthocyanins, absorb the bluish and greenish wavelengths, producing their red reflectance. The features of these three compounds are usually masked by chlorophyll absorptance. However, during senescence or a stress period, the chlorophyll production usually declines shifting the spectral absorptance and allowing these compounds to display their features. The relative abundance of these various pigments change during the plant senescence process, which produces shifts

in the spectral absorptance and reflectance. Figure 2.2 displays the temporally dynamic nature of visual foliar reflectance (Lusch, 1989).

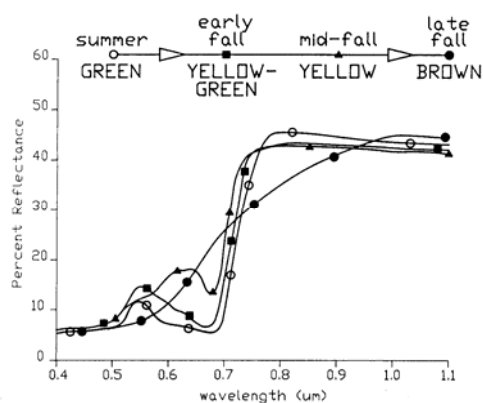


Figure 2.2 Visible and near-infrared reflectance differences in vegetation due to senescence (adopted from Lusch, 1989)

Within the near-infrared part of the spectrum, a relationship exists between leaf structure and maturity. Since young immature plants contain less chlorophyll and fewer air voids than older leaves, they reflect more visible light and less infrared radiation. In addition, as the number of leaf layers increases (increasing leaf area), infrared reflectance increases, especially in the near-infrared portion (Lusch, 1989).

### 2.3.2.2 Remote Sensing of Agricultural Crops

Based on the understanding of how vegetation reflectance differs through the electromagnetic spectrum the use of remote sensing applications for distinguishing between agricultural crop types and internal crop characteristics has been extensively researched during the past decade (Weigand *et al.*, 1991; Cloutis *et al.*, 1996; Mogensen *et al.*, 1996; Cloutis *et al.*, 1999; Metternicht *et al.*, 2000; Senay *et al.*, 2000; Thenkabail *et al.*, 2000; McNairn *et al.*, 2002). The trends being developed between specific crop types, maturity, nutrient levels and their reflectance values in spectral bands and relationship to vegetation indices (VI) are becoming well known and useful when limited ground truth data are available (Senay *et al.*, 2000). An explanation of VI followed by examples of their use in agricultural application provides an indication that remote sensing of vegetation attributes could depict landscape variability and be a useful method of collecting vegetation data and an appropriate layer within the LMU classification technique.

#### 2.3.2.2.1 Vegetation Indices

There is a wide range of image transformation/enhancement techniques available, such as image reduction and magnification, contrast enhancements, band ratioing, spatial filtering and special filtering, which includes principal components analysis and vegetation indices (VI). A VI is a mathematical combination of several bands of remote sensing data that utilises the significant differences in reflectance of vegetation in the blue, green, red and near-infrared wavebands (as explained in the previous section). The index is typically a sum, difference, ratio or other linear combination that reduces multi-band observations to a single numerical index (Weigand *et al.*, 1991). VI are a simple tool for exploring or quickly evaluating the state of vegetation over large areas. They enhance the spectral contrast of vegetation and minimise the influence of other factors (e.g. topography, illumination). Their usefulness depends on the empirical relations that may be found between them and variables of interest, such as vegetation stress and biomass.

Jackson and Huete (1991) classify VI into two groups; ratios and linear combinations and a third group can be added called orthogonal transformations. Ratio VI or slope-based indices (Thiam and Eastman, 2003), represent simple arithmetic combinations that focus on the contrast of the reflectance in bands (Table 2.8). Linear VI or distance based indices (Thiam and Eastman, 2003), are designed to eliminate the effect of background soil brightness and detect the features only of vegetation (Table 2.8). Orthogonal transformations, undertake a transformation of the available bands to form a new set of uncorrelated bands in which a green vegetation index can be defined (Thiam and Eastman, 2003). Thiam and Eastman (2003) state that the link between orthogonal transformation techniques is that they all express green vegetation through the development of their second component. Some examples of VIs and their authors are listed in Table 2.8.

Table 2.8 Examples of Vegetation Indices

Name	Acronym	Group	Author
Ratio	Ratio	Ratio	(Rouse <i>et al.</i> , 1974)
Normalised Difference Vegetation Index	NDVI	Ratio	(Rouse <i>et al.</i> , 1973)
Transformed Vegetation Index	TVI	Ratio	(Deering <i>et al.</i> , 1975)
Normalised Difference Vegetation Index-Green	NDVIgreen	Ratio	(Gitelson and Merzlyak, 1997)
Photosynthetic Vigour Ratio	PVR	Ratio	(SpecTerra Services, 1999)
Plant Pigment Ratio	PPR	Ratio	(SpecTerra Services, 1999)
Perpendicular Vegetation Index	PVI	Linear	(Richardson and Weigand, 1977)
Perpendicular Vegetation Index 3	PVI3	Linear	(Qi <i>et al.</i> , 1994)
Soil Adjusted Vegetation Index	SAVI	Linear	(Huete, 1988)
Weighted Difference Vegetation Index	WDVI	Linear	(Richardson and Weigand, 1977)
Atmospherically Resistant Vegetation Index	ARVI	Linear	(Kaufman and Tanré, 1992)
Soil Adjusted and Atmospheric Resistant Vegetation Index	SARVI	Linear	(Kaufman and Tanré, 1992)
Principal Components Analysis	PCA	Orthogonal	(Singh and Harrison, 1985)
Green Vegetation Index of the Tasseled Cap	GVI	Orthogonal	(Kauth and Thomas, 1976)

The underlying premise of using remote sensing to monitor crop condition is that important crop parameters related to growth and yield are manifested in the multi-spectral reflectance of crop canopies (Bauer, 1985). The Leaf Area Index (LAI), representing the ratio of leaf surface area to ground area, is the fundamental canopy parameter in two basic physiological processes: photosynthesis and evapotranspiration, which are most dependant on solar radiation (Bauer, 1985). Most models of crop growth and yield require an estimate of green LAI, and therefore the strong relationship of infrared reflectance to LAI of crop canopies is the basic mechanism for linking multispectral remote sensing data to crop growth and condition (Bauer, 1985; Clevers, 1997).



In remote sensing a common approach for measuring or monitoring crop growth is the correlation of vegetation indices or ratios with crop variables, such as percentage of vegetation cover and LAI (Moran *et al.*, 1997). Moran *et al.* (1997) suggest that measurements of crop properties at sample sites combined with multi-spectral imagery could produce accurate, timely maps of crop characteristics for defining precision management units. Some examples of works that have shown relationships between remotely sensed data and crop attributes are detailed hereafter.

Senay *et al.* (2000) used high resolution multi-spectral data to identify corn and soybean crops at various growth stages. They used multi-spectral sensor (MSS) data, with 12 spectral bands whose wavelength range included the visible, near-infrared and mid-infrared and was obtained at 1m resolution on four occasions representing different growth stages. Spectral analysis of the individual bands and three vegetation indices were performed. The correlation analysis between the MSS and ground reference data highlighted that generally the near-infrared bands were more highly correlated than were the visible or the mid-infrared band. On the individual acquisition dates the VI performed better than the red and green bands, but similar to the near-infrared band. However, the VI correlated better with plant height and plant nitrogen than any of the individual bands. The strongest correlations were between the NDVI and plant height and nitrogen content in the leaves.

NDVI has shown good correlation with plant growth variables (i.e. height, LAI, biomass and yield); in particular, it is highly related to yield and could be used to estimate yield-based within field management zones (Yang and Anderson, 1996). However, Corner *et al.* (1998) used NDVI from Landsat TM and high resolution multi-spectral video to estimate yield at different scales. Reasonable correlations were achieved at a regional scale (paddock scale) using Landsat TM. However, high resolution multi-spectral video, at a local scale (sub-paddock scale), resulted in poor correlations. The regional scale achievement was attributed to a considerable degree of spatial and temporal averaging, such as varying sowing dates and localised weather events. At a local scale other factors that cause variation in crop production predominate, and therefore high resolution NDVI measurements may be best used as a diagnostic tool to determine these factors, rather than as a yield predictor.

Previous investigations distinguishing internal canola crop variations using passive remote sensors (Cloutis *et al.*, 1996; Mogensen *et al.*, 1996; Cloutis *et al.*, 1999) provide support to the hypothesis that variations in canola growth can be depicted by remote sensed imagery. For instance, Mogensen *et al.* (1996) investigated the use of a spectral reflectance index for determining early water stress on canola grown under controlled field conditions (lysimeter tanks). The reflectance index (RI) being defined as the ratio of incoming and reflected infrared radiation in the range of 740 to 820nm, to the incoming and reflected photosynthetically active radiation between 400 and 700nm. They simulated a drought type effect within the trial and used a relative reflectance index (RRI) defined as the ratio of the reflectance index of the droughted crops to the fully irrigated reference crops to analyse the effect of water stress on RI. The authors concluded that the RRI was an index sensitive to water stress, seeming most appropriate in the vegetative stage of growth, as changes in spectral response of crop surfaces due to senescence or changes in architecture due to leaf wilting of the crop may change the RI values.

The above works have highlighted the use of VIs for agricultural applications, in particular depicting variations in crop growth and in some cases suggesting their use for determining management units. Five VIs namely, NDVI, NDVIgreen, SAVI, PVR and PPR (Table 2.8) will be used to analyse relationships with crop growth variables in this research. The aim is to determine if they are able to depict landscape variability and in turn be useful inputs in the LMU classification. Thus, the selected VI are explained in detail hereafter.

### 2.3.2.3 Selected Spectral Indexes for Agricultural Applications

One of the most successful vegetation indices based on band ratioing was developed by Rouse *et al.* (1973). They computed what is known as the normalised difference vegetation index, referred to as the NDVI, whereby a new image is created by transforming the pixels according to the equation,

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (2.4)$$

The NIR and RED are reflectance values in those bands. Highly vegetated land produces a NDVI close to unity while in non-vegetated areas NDVI is close to zero or it assumes negative values (Eastman, 1999; Lamb, 2000).

The normalised difference vegetation index-green (NDVI-green) is a VI developed for the remote estimation of chlorophyll content in higher plant leaves. Chlorophyll content in the higher plant leaves changes throughout different stages of plant development. The vegetation being exposed to various stresses affects the content of the pigments. Thus, a measure of chlorophyll content can aid as a guide to detection of physiological states and stresses in plants (Gitelson and Merzlyak, 1997). Gitelson and Merzlyak (1997) found that the use of the reflectance in the green channels increases the sensitivity to the chlorophyll content with a wide range of chlorophyll variation. Thus the use of the green channel increases the sensitivity of the NDVI to chlorophyll content by about five-fold. The equation is as follows,

$$NDVI_{green} = \frac{(NIR - GREEN)}{(NIR + GREEN)} \quad (2.5)$$

As the chlorophyll content increases, the absorption in the green band increases (i.e. small digital number). Thus, large chlorophyll content lead to a larger NDVI-green value.

The soil-adjusted vegetation index (SAVI) is a transformation technique to minimise soil brightness influences from the spectral vegetation indices involving red and near-infrared wavelengths (Huete, 1988). A constant soil adjustment factor,  $L$ , is incorporated into the denominator of the NDVI equation.  $L$  varies according to the reflectance characteristics of the soil (eg. colour and brightness) (Eastman, 1999). The equation takes the form,

$$SAVI = \frac{(NIR - RED)}{(NIR + RED + L)} \times (1 + L) \quad (2.6)$$

The  $L$  factor chosen depends on the density of the vegetation cover being analysed. Huete (1988) suggests that for very sparse vegetation, use a  $L$  factor of 1, for intermediate 0.5 or larger vegetation density, 0.25.

The photosynthetic vigour ratio (PVR) uses the green band, as a reference band, and the strong chlorophyll absorption red band. The ratio is calculated as:

$$PVR = \frac{GREEN}{RED} \quad (2.7)$$

This ratio is large for leaves with strong chlorophyll absorption (photosynthetically very active) and small for weakly active vegetation with smaller chlorophyll absorption (SpecTerra Services, 1999).

The plant pigment ratio (PPR) is a combination of the green band, as a reference band and the blue band, related to pigment absorption (SpecTerra Services, 1999).

The PPR is as follows,

$$PPR = \frac{GREEN}{BLUE} \quad (2.8)$$

The green band is intentionally made the numerator so that strongly pigmented foliage, absorbing more energy in the blue band, will have a large PPR value while the weakly pigmented foliage will have a small PPR (Metternicht, 2003).

The five VIs described above are statistically analysed in this research in relation with crop growth attributes collected in the field.

#### 2.3.2.4 Reflectance of Soil properties

Optical remote sensing measurements record the radiation emitted and reflected from the soil surface, as there is very little penetration of electromagnetic energy through the soil body. Soil reflectance derives from the inherent spectral behaviour of the heterogeneous combination of the biogeochemical (mineral and organic) constituents, geometrical-optical scattering (particle size, aspect, roughness), and moisture conditions of the surface (Baumgardner *et al.*, 1985; Irons *et al.*, 1989; Ben-Dor *et al.*, 1999).

In an endeavour to effectively interpret soil spectral reflectance for land management studies, several researchers have reported good correlations between soil reflectance and soil properties such as organic matter, soil moisture, particle size distribution, iron oxide content, colour, soil mineralogy, salts, and parent material (Bowers and Hanks, 1965; Stoner and Baumgardner, 1981; Karmonova, 1982; Latz *et al.*, 1984; Coleman and Montgomery, 1987; Escadafal, 1989; Henderson *et al.*, 1989; Epema, 1993; Metternicht and Zinck, 1997). Soil spectral reflectance signatures result from the presence or absence, as well as the position and shape of specific absorption features of its constituents. Soils are mixtures of many inorganic and organic constituents so it is not straightforward to evaluate the composition of soils from their spectral signatures (Ben-Dor *et al.*, 1999).

Early studies to quantify soil reflectance, and determine difference in signatures between soil reflectance spectra were conducted by Condit (1970), Stoner and Baumgardner (1981) and Huete and Escadafal (1991). Under laboratory conditions, Condit (1970) identified three main types of spectral soil curves in the range of 0.32 to 1 $\mu\text{m}$ , though no attempts were made to relate spectral shape to soil properties quantitatively. The five distinct soil reflectance spectral curves of Stoner and Baumgardner (1981) are based on curve shape, the presence or absence of absorption bands, the predominance of soil organic matter and iron oxide composition in the range of 0.52 to 2.32 $\mu\text{m}$ . Whereas the four soil spectral curves identified by Huete and Escadafal (1991) using spectral decomposition and mixture modelling techniques in the range of 0.4 to 0.9 $\mu\text{m}$ , represent soil brightness, red iron oxides, organic carbon and reduced iron oxide (goethite) contents (Metternicht *et al.*, 2002).

#### 2.3.2.4.1 Moisture

A common observation with most soils is that they appear darker when wet than dry. This is due to the decreased reflectance of incident radiation in the visible region of the spectrum (Baumgardner *et al.*, 1985). Strong water absorption bands at 1.45 and 1.95 $\mu\text{m}$  affect the shape of the soil reflectance curve (Baumgardner *et al.*, 1985). Increasing soil moisture content generally decreases soil reflectance across the entire shortwave spectrum (Irons *et al.*, 1989). Bowers and Hanks (1965) (cited in Baumgardner *et al.*, 1985, p.16) found a lowering in reflectance for Newtonia silt loam at six increasing soil moisture contents over the wavelength range of 0.5 to 2.5 $\mu\text{m}$  and is shown in Figure 2.3.

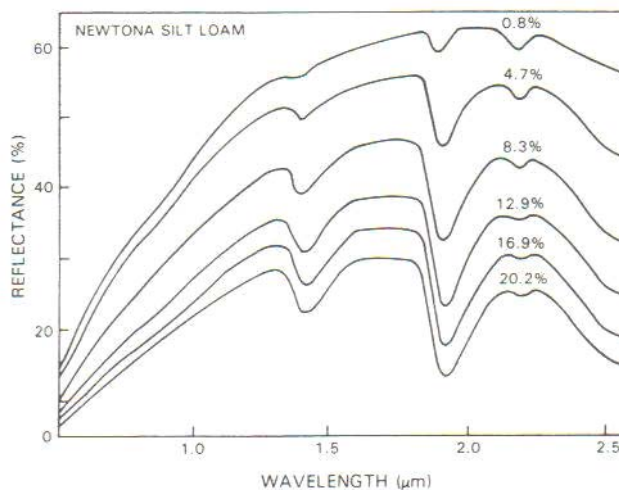


Figure 2.3 Soil reflectance spectra for a silt loam soil at varying moisture contents. Percentage of moisture content by weight shown above each curve (Bowers and Hanks, 1965) (shown in Irons *et al.*, 1989, p.90)

#### 2.3.2.4.2 Organic Matter

The soil organic matter content and the composition of organic constituents have a strong influence on soil reflectance. In general, as organic matter content increases, soil reflectance decreases throughout the 0.4 to 2.5 $\mu\text{m}$  wavelength range (Hoffer and Johannsen, 1969) (cited in Baumgardner *et al.*, 1985; Lusch, 1989). Baumgardner *et al.* (1970) (cited in Baumgardner *et al.*, 1985) found that organic matter content plays a dominant role in bestowing spectral properties to soils when the organic matter content exceeds 2 percent. In addition, as the organic matter content diminishes below 2 percent, it becomes less effective in masking the effects on reflectance of other soil constituents. The spectra of soils with organic matter contents greater than 5 percent often have a concave shape between 0.5 and 1.3 $\mu\text{m}$  as compared to the convex shape of spectra for soils with lower organic-matter content (Stoner and Baumgardner, 1981).

Obukhov and Orlov (1964) (cited in Baumgardner *et al.*, 1985) found that organic constituents, including humic and fulvic acid and non-specific compounds including decomposing plant residues, can influence soil reflectance to differing degrees. However, the contribution of each constituent has been difficult to quantify (Baumgardner *et al.*, 1985). Henderson *et al.* (1992) investigated the high dimensional reflectance in the 400 to 2500nm wavelength range with soils ranging in organic carbon content from 0.99 to 1.72 percent. Statistical analysis of the

reflectance values indicated that the visible, near-infrared and middle-infrared bands provided information about the organic carbon content, but not organic matter composition.

#### 2.3.2.4.3 Texture

The size and arrangement of the soil particles in relation to the soil air and water also influence the soil reflectance. Soil texture refers to the size distribution of the soil mineral particles. The relative proportions of sand, silt, and clay-sized particles determine textural classes as shown in Figure 2.4. The soil texture has a strong influence on the reflectance; in particular, sandy soils tend to be brighter than clayey soils (Irons *et al.*, 1989).

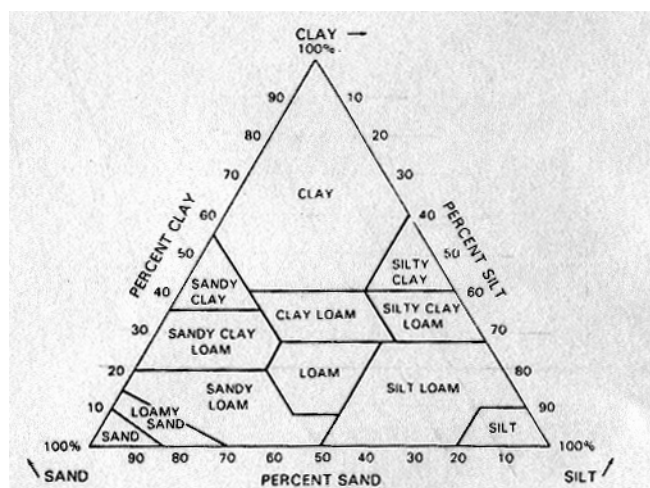


Figure 2.4 The textural triangle defines soil texture on the basis of sand, silt and clay (adopted from Irons *et al.*, 1989)

However, Lusch (1989) mentions that if all other factors are equal, the finer particle sizes will exhibit greater soil reflectance. Bowers and Hanks (1965) (cited in Baumgardner *et al.*, 1985) found a rapid exponential increase in reflectance at all wavelengths between the 0.4 and 1.0 $\mu\text{m}$  with decreasing particle size.

#### 2.3.2.4.4 Colour

Spatial and temporal variations in soil surface colour provide important clues to land degradation processes such as salinization, erosion, and drainage status of a soil (Latz *et al.*, 1984; Cihlar *et al.*, 1987; Mougnot *et al.*, 1993; Metternicht, 1996; Thompson and Bell, 1996). The visible soil reflectance, or colour, is one of the most useful co-variant soil properties used to identify and describe a particular soil

material. Soil scientists use soil colour as a first indirect measure of soil type and its physical, chemical and hydraulic properties in the field (Mattikalli, 1997). Furthermore, soil colour provides a simple surrogate measure to estimate many other difficult-to-measure soil constituents, and is widely used in soil mapping as an indicator of the presence of organic matter, iron oxides, and carbonates as well as moisture content (Huete, 2002).

For a few soils, colour is directly derived from the parent material. However, more common organic matter, which consists of dark compounds, tend to mask the colours of iron oxides, and manganese oxides darken the soil (Moore, 1998a). Soils with little organic matter often display the iron oxide colours (Brady and Weil, 1999a), such as the red of haematite and yellowish colour of goethite (Baumgardner *et al.*, 1985; Escadafal, 1994; Brady and Weil, 1999a) and in general colour is related to the type(s) of oxides present rather than the amount (Bigham *et al.*, 1978), and in turn, the type of iron oxides present can be inferred from the colour (Moore, 1998a). Iron compounds, such as organo-iron, amorphous, weakly crystallized and non-silicate irons, have dissimilar effects on the visible spectral reflectance of soils. Moreover, if all the non-silicate iron is removed then soil reflectance becomes practically identical, as it is controlled by the achromatic grey colour of clay minerals (Karmonova, 1982).

The iron oxides goethite, haematite, maghemite, lepidocrocite and ferrihydrite have a distinctive colour range from yellow to red in aerobic soils. Table 2.9 provides that specific colour range in regards to hue, value and chroma (Moore, 1998a).



Table 2.9 Iron oxides and soil colour (adapted from Moore, 1998a)

Iron Oxide	Munsell Soil Colour Code	Colour
Goethite	10YR → 7.5YR	Brownish yellow
Haematite	5YR → 5R	Blood red
Lepidocrocite	7.5YR with value >6	Orange
Ferrihydrite	5YR → 7.5YR with values >6	Brown to dark reddish brown
Other iron compounds		
Jarosite	5Y (mottles)	Yellow mottles
Maghemite	2.5YR → 5YR	<i>Can be identified with a magnet</i>

Waterlogged soils may lack pigmentation of iron oxides in the anaerobic zones, as the oxides have been removed through bacterial reduction and leaching. For soils that have insignificant organic matter, the soil is white, grey or green (Moore, 1998a). The drainage status of clayey subsoils can be inferred via their colour (Moore, 1998a). Moore (1998a) mentions that generally permeability decreases as subsoil colour changes from red → brown → yellow → dark (black to very dark brown) → grey (light greys to bluish and greenish greys).

Fitzpatrick *et al.* (1999) offer some indication of good and poor soil conditions for most forms of plant growth. They mention that dark brown colours near the surface and bright yellowish and reddish colours in the sub-soils are good indicators as the darker surface soils are often associated with higher organic matter, well aerated and above average nutrient levels. While yellow and reddish subsoils, indicate iron and therefore suggest good drainage. Poor soil conditions can be identified by mottles (varying colours) and rust coloured specks, as both indicate water logging. Very pale grey or white colours may indicate leaching.

Many of the absorption features in soil reflectance spectra are due to the presence of iron in some form (Irons *et al.*, 1989). The redness index, RI (Equation (2.9)) (Escadafal and Huete, 1991) has been found to increase the sensitivity of vegetation indices (VI), such as the NDVI and SAVI, by estimating the proportion of bare soil (soil noise) in a VI with the RI in sparsely vegetated areas. Based on the findings of Escadafal (1989)(cited in Escadafal and Huete, 1991) that the saturation of soil colour (or "redness") was correlated with the red band/green band ratios in the case

of arid soils, the normalised difference between the two bands has been used to express the "redness" of bare soil surfaces.

$$RI = \frac{RED - GREEN}{RED + GREEN} \quad (2.9)$$

Stewart and McBratney (2001) had encouraging results when they identified potential managements zones from remote imagery of bare soil. They found statistical and agro-economic evidence that suggests red soil (assumed Chromosols) and the dark soil (assumed Vertosols) have very different cotton yield potentials. They thought that this difference is most likely linked with the differing soil textures found in Chromosols and Vertosols and their associated water holding capacity.

The relationship between remote sensing and soil properties may provide a useful measure to detect variations in soil properties across the landscape and provide a layer for inclusion in the LMU classification. In particular the RI, which is simple to generate, would be a useful surrogate to identify soil variation in areas of sparse vegetation.

### 2.3.3 Digitally Derived Topographic Attributes

As mentioned, topographic attributes have been recognised as important inputs into many land classification techniques (Table 2.1) and in particular landforms, the compound topographic index (CTI) and slope have been identified as important drivers in soil landscape studies, and thus their derivations are described individually hereafter.

#### 2.3.3.1 Landforms

There are several techniques for the development of landform units and these differ in terms of categorical structure. Zinck (1988) defines the geopedological approach to landform classification. This approach is based on a strong integration of geomorphology and pedology and uses geomorphology as a tool to improve and speed up the soil survey. Ventura and Irvin (2000) provide a sample of different methods for division of the landscape into identifiable sections (Ruhe and Walker, 1968; Troeh, 1964; Hugget, 1975; Conacher and Dalrymple, 1977; Pennock *et al.*, 1987; Speight, 1974,1990) (cited in Ventura and Irvin, 2000), which are either based on soil-forming processes, landform and landscape elements and/or a combination of these concepts.

Several studies have correlated terrain attributes, derived from a digital elevation model (DEM), to soil properties (Moore *et al.*, 1993a; Gessler *et al.*, 1995; Boer *et al.*, 1996; McKenzie and Ryan, 1999). Ventura and Irvin (2000) propose methods for delineating landform elements based on terrain attributes, by partitioning the landscape into geomorphologic units using specific terrain attributes that relate to soil properties and assigning them to landform elements. They used continuous classification (fuzzy logic) and unsupervised (ISODATA-crisp boundaries) classification techniques to classify each pixel in a DEM (10m resolution) according to its membership in a landform class based on six primary and secondary topographic attributes. Fuzzy classification permits soil class membership values to be treated as continua in geographic space, while crisp models have delineated mapping boundaries in geographic space (Burrough *et al.*, 1997). Both methods offer products that can aid those people interested in soil-landscape processes. However, although fuzzy models provide additional information at each point, they do not offer easily visualized and directly usable output products (i.e. maps) (Irvin *et al.*, 1995).

A key Australian classification of landforms was developed by Speight (1974; 1990). Speight (1974) proposed a two-level descriptive procedure for a systematic, parametric description of landforms into *landform patterns* and *landform elements*. The landform is viewed as a mosaic of tiles whereby the larger tiles, *landform patterns* are generally on the order of 300m radius. The smaller tiles, which are mosaics within landform patterns, are *landform elements* that are commonly of the order of 20m radius (Speight, 1990). Speight (1990) defined about 40 types of landform patterns including for example flood plain, dunefield and hills and more than 70 types of landform elements such as cliff, footslope and valley flat. Relief and stream occurrence describe landform patterns while landform elements may be described by five attributes namely slope, morphological type (topographic position), dimensions, mode of geomorphological activity and geomorphological agent. Speight (1990) distinguished ten types of topographic positions in which landform elements fall into, as listed in Table 2.10.

Table 2.10 Morphological type (topographic position) classes by Speight (1990)

Name	Definitions of Speight (1990)
<i>Crest</i>	Area high in the landscape, having positive plan and/or profile curvature
<i>Depression</i> ( <i>open, closed</i> )	Area low in the landscape, having negative plan and/or profile curvature, <i>closed</i> : local elevation minimum; <i>open</i> : extends at same or lower elevation
<i>Flat</i>	Areas having a slope < 3%
<i>Slope</i>	planar element with an average slope > 1%, sub classified by relative position:
<i>Simple slope</i>	adjacent below a crest or flat and adjacent above a flat or depression
<i>Upper slope</i>	adjacent below a crest or flat but not adjacent above a flat or depression
<i>Mid-slope</i>	not adjacent below a crest or flat and not adjacent above a flat or depression
<i>Lower slope</i>	not adjacent below a crest or flat but adjacent above a flat or depression
<i>Hillock</i>	compound element where short slope elements meet at a narrow crest < 40m
<i>Ridge</i>	compound element where short slope elements meet at a narrow crest > 40m

A full description of each morphological type can be found in Speight (1990). Figure 2.5 provides an example of a profile across the terrain divided into morphological types of landform elements as classified by Speight (1990).

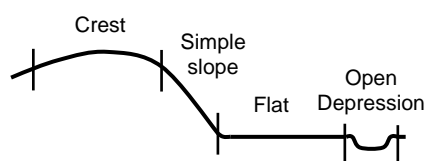


Figure 2.5. Example of a profile across terrain divided into morphological types of landform elements (adapted from Speight, 1990)

Speight's (1990) description of landforms is a key component that contributes towards the systematic recording of field observations in Australian soil and land surveys, and as such, many existing survey records consist of Speight's (1990) landform descriptions. Coops *et al.* (1998) produced a set of techniques that allow topographic position to be predicted from 25m DEMs in which the classes are equivalent to Speight's (1990) morphological types that are used by field botanists, ecologists and other natural resource scientists and managers.

The strong relationship reviewed between landform elements and soil properties highlights landform as a potentially fundamental topographic attribute to be incorporated in the LMU classification in this research, with particular emphasis on landform elements classified in an Australian context (i.e. Speight (1974; 1990)).

### 2.3.3.2 Compound Topographic Index

The compound topographic index (CTI) is a quantification of the topographic position of a site in the landscape. Originally formulated for hydrological forecasting by Bevan and Kirkby (1979), the CTI is also referred to as the topographic wetness index (TWI) (McKenzie *et al.*, 2000; Wilson and Gallant, 2000) or wetness index ( $w$ ) (Moore *et al.*, 1993a). The CTI is intended to represent the topographic control on soil wetness (Wilson and Gallant, 2000), and furthermore, is a guide to water and sediment movement in the landscape (McKenzie and Ryan, 1999; McKenzie *et al.*, 2000). Moore *et al.* (1993a) mentions that CTI is one of the potential compound indices for predicting the spatial distribution of soil properties and in soil specific crop management applications.

The CTI has been used for sampling designs (Gessler *et al.*, 1995; McKenzie and Ryan, 1999), classification techniques (Irvin *et al.*, 1995; Corner, 1999; Herron *et al.*, 2004) and correlation analysis (Moore *et al.*, 1993a) within soil studies, which are discussed hereafter.

Gessler *et al.* (1995) used the CTI to guide soil sampling by postulating that the spatial dependence structure of the CTI related, in a general way, to the spatial dependence structure of the soil attribute of interest. Soil attributes show varying degrees of spatial dependence, and this reduces the efficiency of random sampling. Spacing sample sites using information about the spatial dependence structure increases the information content of samples. Accordingly, Gessler *et al.* (1995) designed the sampling plan following the hypothesis that CTI was a strong controlling variable, and their field data supported this.

McKenzie and Ryan (1999) used a classification of the CTI in conjunction with climate and geology parameters for a digital stratification of the landscape whereby each stratum represents a discrete environment. Density functions of the CTI and climate data were used to calculate quantiles on an equal area basis (McKenzie *et al.*, 2000). Potential sample sites were then randomly selected within each stratum. The authors draw the conclusion that statistical sampling plans are explicit, consistent and repeatable in contrast to conventional survey methods. In regards to the predictive

relationship between soil properties and environmental variables, when using a regression tree approach for the prediction of soil profile depth (m) they found that CTI appeared to be one of the explanatory variables.

Corner (1999) demonstrates the use of CTI in several examples of an expert (knowledge based) system, namely *Expector*, to soil mapping. The CTI forms an evidence layer within a system of Bayesian inference to assign a varying probability of a soil property occurring with a given study area. Evidence is weighted according to the uncertainty associated with it (Cook *et al.*, 1996). Irvin *et al.* (1995) used the CTI as one of six topographic attributes to classify each pixel of a 10m resolution DEM to a landform class which were determined by the natural clustering of data in attribute space. Furthermore, Herron *et al.* (2004) used the CTI within the salinity benefit index, where by the salinity benefit index is a measure of the relative change in stream salinity from current salinity levels, caused by changes in land use/management within an area of interest. Quickflow, which is the surface runoff, rapid subsurface flow and interflow component of stream flow, is a function of rainfall, soil, topography and land use. The CTI forms the topography function and is used for weighting quickflow zones, and is assumed that a cell's contribution to quickflow was directly proportional to its CTI value.

Moore *et al.* (1993a) refer to their own works (Moore *et al.*, 1988; 1993b), which have used the CTI to characterise the spatial distribution of zones of surface saturation and soil water content in landscapes. Moore *et al.* (1993a) found that slope and CTI were the terrain attributes most moderately correlated to soil attributes. CTI individually accounted for about one quarter the variability of several soil attributes, such as silt percentage ( $r=0.61$ ), organic matter content ( $r=0.57$ ), A-horizon depth ( $r=0.55$ ) and phosphorus ( $r=0.53$ ) in the soil surface.

These examples highlight that incorporating the CTI in the classification of LMUs will include topographic information that is related to soil properties and the way water moves through the landscape.

### 2.3.3.3 Slope

Slope is a key primary topographic attribute, it is utilised in the computation of many attributes that describe the topographic position in the landscape, with landform and CTI previously discussed to name a few. *Slope is the means by which gravity induces flow of water and other materials, so it is of great significance in hydrology and geomorphology. It affects the velocity of both surface and subsurface flow and hence soil water content, erosion potential, soil formation and many other important processes* (Gallant and Wilson, 2000). In a preliminary analysis of the relationships between topography and soil attributes, Moore *et al.* (1993a) found that slope was the most moderately correlated terrain attribute to soil attributes ( $R = 0.45-0.64$ ), and suggested that the relationship between terrain and soil attributes could be applied as a guide to sampling strategies. Furthermore, McKenzie *et al.* (2000) state that terrain analysis can be used to generate stratified random sampling schemes in a way that previously had not been possible in conventional soil survey. These examples highlight the importance of slope in soil and landscape classifications and demonstrate that it would add important topographic information to the proposed LMU classification.

## 2.4 Validation Techniques for the LMU classification

Evaluation techniques are required for ensuring that foreseen models for the creation of LMUs are, *appropriate for their intended purpose*. In a review on sensitivity analysis in multicriteria spatial decision-making, Qureshi *et al.* (1999) (cited in Delgado and Sendra, 2004) states that a model evaluation could be divided into three components: a) Verification, b) Validation and c) Sensitivity analysis, which are further individually described. *Validation* is defined by Qureshi *et al.* (1999) (cited in Delgado and Sendra, 2004 p. 1174) as “*Modelers have to ensure that the structure of the model is correctly built from a conceptual and operational point of view (if it is appropriate for its intended purpose), according to a specific methodology.*” As such, validation techniques for the foreseen model for the creation of LMUs are reviewed.

Determining the success of a classification technique is difficult, as more often than not there is little reference information that can be used. Crop yield (or the income derived from it) has the greatest bearing on farm management and practices at

present. Management zones should display significant differences in yield (or production) for variable rate application of crop inputs to be worthwhile. Ensuring the differences displayed in crop yield are genuine, let alone significant, is difficult (Cupitt and Whelan, 2001). However, one method of determining the success of the LMU classification would be in comparison with yield data collected across the same study area. Several works have outlined methods for deriving management zones based on yield inputs (as shown in Table 2.1).

Methods of utilising yield data for validation purposes are discussed hereafter in view of either using the LMUs as a spatial base and assessing whether significant differences (in yield) are present between them, or forming a yield based spatial zone map for subsequent map comparison.

#### 2.4.1 Comparisons with Yield Data

Cupitt and Whelan (2001) use local block kriging with local variograms within VESPER (Whelan *et al.*, 2001) to predict five attributes namely; yield data for 1996, 1997 and 1998, soil EC and elevation data onto a single 5m grid. From the five interpolated attributes multivariate k-means clustering was then used to delineate three potential management zones. The kriging process provides an estimate of the mean prediction variance ( $\sigma^2 krig$ ) and Cupitt and Whelan (2001) show that the confidence interval (95% C.I.) surrounding the mean yield estimate within a field ( $\mu$ ) can be calculated according to Equation (2.10).

$$95\%C.I. = \mu \pm \left( \sqrt{\sigma^2 krig} \times 1.96 \right) \quad (2.10)$$

Cupitt and Whelan (2001) suggest that the absolute difference between the mean zone yields ( $|\bar{Y}_{zone1} - \bar{Y}_{zone2}|$ ) should then follow Equation (2.11) (Moore and McCabe, 2003) for the potential management zones to be considered representative regions of significantly different yield ( $p < 0.05$ ).

$$\left| \bar{Y}_{zone1} - \bar{Y}_{zone2} \right| \geq \sqrt{2 \times \sigma^2 krig} \times 1.96 \quad (2.11)$$

Whelan and McBratney (2003) also utilise Equations (2.10) and (2.11) to examine differences in yield between potential management zones. This suggests that the



LMU classification zones could be compared in terms of significant difference in yields using the above mentioned methods.

Whelan and McBratney (2000) present a methodology for assessing the temporal variability of a paddock. The variability across time in crop yield at within-field scales can be estimated based on the yield data by Equation (2.12).

$$\sigma_{T,i}^2 = \frac{\sum_{j=1}^n (Y_{i,j} - \bar{Y}_i)^2}{n-1} \quad (2.12)$$

where;  $\sigma_{T,i}^2$  = temporal variance at point  $i$ ,  $Y_{i,j}$  = yield value at point  $i$  in each year  $j$  and  $\bar{Y}_i$  = mean yield value at point  $i$  for all years. Fixed points within the field must be used for comparison, whereby the data are obtained from spatial prediction onto a single grid. The estimate should provide an indicator of seasonal influences on crop yield (Whelan and McBratney, 2000). In relation to the LMU classification, a mean yield value could be calculated for each LMU that could be considered,  $Y_{i,j}$ . In this way the LMUs could be assessed in terms of their temporal stability. Unfortunately, problems would arise when using different crop types.

Yield map standardisation is essential when combining multiple year and crop type yield data. Stafford *et al.* (1998) and Basnet *et al.* (2003) have assessed this issue in different ways. Basnet *et al.* (2003) delineate management zones using multiple crop yield data. They use four years of consecutive yield that had been interpolated on a 10m grid with VESPER (Minasny *et al.*, 2002). The output yield data were converted to deciles (i.e. divide each data set into ten equal parts) and then rescaled (using linear interpolation between adjacent points) to values between 0 and 1. The four scaled yield layers were then combined spatially using arithmetic operators, such as addition and division. Average yield values were calculated on a cell by cell basis, and thus a map of average cell values created (0-100 percent yield). This was then reclassified into high, medium and low yield zones based on even divisions (i.e. 0 to 0.33 = low). Unfortunately, Basnet's *et al.* (2003) output map of three management zones does not produce contiguous zones. Stafford *et al.* (1998) used a fuzzy classification on the yield data after first standardising the yield for each season to zero mean and unit variance (i.e. each value is given its z-score value).

Cluster analysis was carried out using the FCM algorithm of Bezdeck *et al.* (1984)(cited in Lark and Stafford, 1997). Basnet *et al.* (2003) used deciles to get the control points and make some form of standardisation prior to fuzzy input; while, Stafford *et al.*(1998) use zero mean and unit variance (i.e. z-score) for standardisation prior input to the fuzzy function.

Several researchers (Boydell and McBratney, 1999; Aspinal, 2000; Whelan and McBratney, 2000; Kelly *et al.*, 2002; Basnet *et al.*, 2003) have investigated the use of yield data to examine the spatial variability of yield over time.

The development of stable yield zone estimates was the aim of Boydell and McBratney's (1999) research. Derived from multi-year yield estimates from mid-season Landsat TM imagery over 11 consecutive years, they wanted to discover the number of consecutive years of yield estimates required to give similar 'stable' estimates of yield zones. Boydell and McBratney (1999) used modified fuzzy k-means for predictive classification (FuzME (Minasny and McBratney, 2002)) to cluster the data and determine the most suitable number of groups. Their fields showed a strong degree of temporal stability and the general conclusions drawn were that stable yield zone patterns may emerge from multi-year yield estimates. More specifically, they state that five years of data (+/- 2 years) seem to give reasonable stable estimates of yield zones.

Most previous works have hypothesised that spatial trends in yield data would become more stable over time. However, Blackmore *et al.* (2003) utilising six years of yield data from four paddocks found that historical yield map trends cannot be used to extrapolate yield patterns in the future. Nonetheless, he suggests that spatial and temporal trend maps can help create homogenous management zones.

Detailing Blackmore's *et al.* (2003) approach, the yield data collected in 1995-2000 inclusive was standardised on a 20m grid using a 20m search radius (they found that larger grids sizes tended to smooth the data too much, while smaller grids became too reliant on low number of data points). Blackmore *et al.* (2003) firstly produced two maps (spatial trend and temporal stability), which were later combined to form a spatial and temporal trend map.

a) Spatial Trend Map: The spatial trend map is designed to show this trend by calculating the arithmetic temporal yield at the same grid point over a number of years, which can subsequently be divided into tonne/ha classes. During this process Blackmore *et al.*(2003) found a large difference between yields from year to year. They documented this as the *Temporal stability: inter-year offset*; which was calculated as the difference between the arithmetic mean yield values between two years in the same field. Histograms of yield were computed and an offset judged as the difference between the yearly curves.

b) Temporal stability - temporal variance map: The temporal variance at a point has been identified by Blackmore *et al.* (2003) when one part of the field yields relatively high in one year and relatively low in another when compared to the mean. The variance from the mean (point yield minus the field mean) over time (1995-2000); is calculated for each year and then divided by the number of years (6 in this case) as follows;

$$\sigma_i^2 = \frac{\sum_{t=95}^{t=00} (Y_{t,i} - \bar{Y}_t)^2}{6} \quad (2.13)$$

where:  $\sigma_i^2$  is the temporal variance at grid point  $i$ ,  $t$  is the time in years between 1995 and 2000,  $Y$  is the yield in years  $t$  at point  $i$ , and  $\bar{Y}_t$  is the mean of the yield for the whole fields in years  $t$ . This variance can be converted to standard deviation and maybe more useful as it is in tonnes/ha. The temporal variance will be small if an area of the field were to always yield close to the mean. This could be considered Stable in Time (SIT), as it would have a small temporal variance. Another area could sometimes yield large and sometimes small relative to the mean, this would be temporally unstable and it would give a large value of temporal variance (Blackmore *et al.*, 2003). The temporal variance map can be classified into areas that are SIT and areas considered UNSIT (unstable in time) by setting a particular threshold in the temporal standard deviation data. However as yet, Blackmore *et al.*(2003) have no conclusive method defined to set this level. Therefore, they suggest looking at the sensitivity of the areas within the map that are deemed UNSIT at several levels.

c) Spatial and temporal trend map: The spatial trend map and temporal variance map are brought together to form a single overview of the field and classified into four homogenous classes:

1. high yielding area: above the grand mean (for all years) for the field:
2. low yielding area: below the grand mean (for all years) for the field:
3. stable area – low inter-year spatial variance (arbitrary threshold)
4. unstable area – high inter-year spatial variance (arbitrary threshold)

As such, four possible combinations are possible: i) High and stable (HS); ii) High and unstable (HU); iii) Low and stable (LS) and iv) Low and unstable (LU) (Blackmore *et al.*, 2003).

Blackmore's *et al.*(2003) work offers an appealing approach as it provides a spatial layer based on yield data for comparison purposes that takes in consideration the fact that yield patterns do not necessarily become more stable over time. It is envisaged that a spatial and temporal trend map of this kind could be analysed against the map of LMUs using map comparison techniques to determine the appropriateness of LMUs.

#### 2.4.2 Spatial Map Comparisons

The need to compare two maps spatially has been derived mostly from remote sensing applications where producers and users of the land-use classification want to know its accuracy. Remote sensing studies use the *error matrix* to quantitatively assess the accuracy of the land use classification (Jensen, 1996b) which Foody (2002) state as the “core of accuracy assessment”. The error matrix is the summary of the relationship between two sets of information, which in remote sensing application are typically: (1) a remote sensing derived classification map and (2) reference information map (Jensen, 1996b). A pixel-by-pixel comparison is performed and reported in the error matrix, and several statistics can be calculated such as, the *overall accuracy*, *producer's accuracy* and *user's accuracy*. The overall accuracy is calculated by dividing the total correct pixels by the total number of pixels in the error matrix (Jensen, 1996b). The producer's accuracy is calculated by dividing the total number of correct pixels in that category by the total number of pixels in that category, while the user's accuracy is calculated by dividing the total number of correct pixels in that category by the total number of pixels that were actually classified in that category (Jensen, 1996b). These statistics provide the producer of the map a measure of how well an area is classified and the user of the

map a probability that a pixel classified on the map actually represents that category on the ground (Jensen, 1996b).

The Kappa statistic is also a measure of agreement or accuracy, expressed as a single number, which can be calculated from the error matrix. It is a discrete multivariate technique therefore appropriate for use on remote sensing data because these data are discrete, not continuous and also binomially or multinomially distributed not normally distributed (Jensen, 1996b). The Kappa is the fraction of agreement, which is corrected for the fraction of agreement statistically expected from the random relocation of all pixels in the maps, using the observed frequency distribution for the classification map and reference map (Hagen, 2002). Fitzgerald and Lees (1994) conclude that the Kappa statistic is superior to the overall accuracy percentage, as it is able to test the null hypothesis that there is no agreement between the two maps.

Foody (2002) mentions that there has been a call for standard measures and reporting to be used in accuracy assessment however, this has not been achieved as there is a variety of different needs and interpretations that exist and in reality it is probably impossible to specify a single all-purpose measure. However, the Kappa statistics has been recommended by analysts to be adopted as one of the standard measures (e.g. Smits *et al.* 1999)(cited in Foody, 2002 pg.189). The Kappa statistic for comparing maps has been used extensively in remote sensing applications, some examples include Fitzgerald and Lees (1994), Michelson *et al.* (2000) and South *et al.* (2004).

South *et al.* (2004) refer to Montserud and Leamans (1992) who evaluated Kappa statistics and classification methodologies and proposed that a Kappa value of 0.75 or greater indicates a very good to excellent classification performance. While Foody (2002) cite Thomlinson *et al.* (1999) who set a target of an overall accuracy of 85 percent with no class less than 70 percent accurate.

Fuzzy set theory (Zadeh, 1965) has also been used by several authors to assess the accuracy of maps or map comparisons (Metternicht, 1999; Power *et al.*, 2001; Metternicht *et al.*, 2005) and it was introduced to address some of the map comparison issues (Hagen, 2002). Issues include; allowing for some level of

positional tolerance, finding the spatial distribution of error and to differentiate the error magnitude (i.e. some errors are more significant than others) (Hagen, 2002). Fuzzy sets deal with inexact concepts in a definable way (Burrough and McDonnell, 1998b). Some classes for various reasons cannot, and do not, have sharply defined boundaries and fuzziness can characterise this imprecision (Burrough and McDonnell, 1998b). A pixel is given a grade of membership expressed in terms of a scale varying continuously between 0 and 1. The grade corresponds to the degree to which that pixel belongs to its class. There are different kinds of fuzzy membership functions and their selection depends on the user's requirement in defining the partial membership.

For the application of accuracy assessment or the comparison of maps Hagen (2003) deals with two sources of fuzziness, fuzziness of location and fuzziness of category. Fuzziness of location allows for vagueness of a category's spatial position, i.e. the category may be present somewhere in the proximity of that location. Fuzziness of category allows some categories in the map to be more similar to each other than others. In a similar manner to the Kappa statistics, the  $K_{Fuzzy}$  can also be calculated which provides an overall value of similarity between two maps which results in values between 1 (identical maps) and 0 (total disagreement).  $K_{Fuzzy}$  corrects the percentage of agreement for the expected percentage of agreement and differs from Kappa in its calculation of the expected similarity (Hagen, 2003). For further details of the  $K_{Fuzzy}$  derivation, readers are referred to Hagen (2003).

The need to assess the accuracy of a map or simply compare two maps derived from differing applications has been addressed by the Research Institute of Knowledge Systems in The Netherlands in their development of the Map Comparison Kit (MCK)(RIKS, 2005a). They have bundled together several spatial map comparison techniques, including Kappa and  $K_{Fuzzy}$  outlined above, into user friendly software which handles a series of maps. The Kappa algorithm included in the MCK is dissected into further statistics; Kappa location (measuring the similarity of location) and Kappa histogram (measuring the similarity of quantity).

## 2.5 Summary

The following provides a summary combining the subsections included in the literature review and their link to the following chapters.

Several classification techniques have been reviewed concluding that for this research the LMU classification will be an adaptation of the spatially constrained classification of Oliver and Webster (1989) which in this case has the ability to incorporate large datasets (i.e. high resolution remote sensing). The two stage methodology to classify LMUs is described in Chapter 7.

Soil properties, vegetation and topographic attributes have been identified as important inputs to the LMU classification. Soil properties that appear to be stable in time and influence plant growth have been identified, namely texture, structure, organic matter, nutrient availability, pH, salinity/sodicity balance, depth of topsoil and depth to restricted layer. The relevant properties have been further detailed, highlighting the relationships between the properties, which support scientists in understanding their effect on plant growth. The field collection and chemical analysis of soil properties is detailed in Chapter 4. The soil properties are analysed statistically in Chapter 6, which will assist in determining the appropriate soil properties to be used in the LMU classification.

Based on the evidence of soil sampling strategies, the design will be based on a DEM and remote sensing data to provide an indirect (surrogate) measure of soil variability providing the opportunity to optimise the strategy. A stratified random sampling pattern that uses bulk sampling of surface soil, respecting the size of support will be used in this research. The soil sampling design is presented and implemented in Chapter 4 along with methods for sampling vegetation attributes that are used in remote sensing studies.

The potential of remote sensing data to detect landscape variation based on the relationships that exist between remote sensing systems and soil and vegetation properties has been highlighted. It appears that the RI would be a useful index on bare soil, while the vegetation indices (in particular the NDVI) would be useful for cropped areas. Transect sampling through three crop types, common to the Western

Australian wheatbelt, will be conducted depicting variation in crop growth. An analysis between crop variability and high resolution remote sensing data is discussed in Chapter 4.

Three topographic attributes namely; landforms, compound topographic index (CTI) and slope have been identified as important drivers in soil forming processes and their uses in past studies has been discussed. Their generation utilising GIS software is detailed in Chapter 5 and their relationship with soil properties in this instance are examined in Chapter 6.

Based on the discussion of validation techniques appropriate for the LMU classification, it appears that the use of spatial and temporal yield maps would be the most appropriate method which is mentioned in Chapter 8. The sensitivity of the LMU classification to changing parameters (another form of validation) can be assessed with algorithms provided in the MCK (RIKS, 2005a) (Chapter 8). Specifically, the Kappa, with Kappa location and Kappa histogram provided in the MCK, will be useful in this research to assess the sensitivity of the LMU classification to differing parameters (i.e. distance applied in the spatial constraint). It is preferred over the  $K_{Fuzzy}$  in this case, for its simplicity.

The following chapter includes a description of the study area and existing data sets. The software and hardware used throughout this research are briefly noted.



## CHAPTER 3

### STUDY AREA AND DATA SETS

*Chapter 2 provided appropriate background information and reviewed several approaches for classifying landscapes into spatial zones. It was concluded that a spatial weighted multivariate classification worth a case study. This chapter introduces the case study region and describes how the various data for the experimentation were derived, from the primary data sources; terrain, remote sensing and yield data along with the software used.*

#### 3.1 Study Area

The study site selected is the Muresk Institute of Agriculture Farm, part of Curtin University of Technology. It lies between latitudes 31° 41' and 31° 46'S and longitudes 116° 39' and 116° 45E, and is approximately 100km north-east of Perth city in the mid southern region of the Shire of Northam in the south-west of Western Australia (WA) (Figure 3.1). It covers approximately 1,780ha of which approximately 1,250ha is arable (Muresk Institute of Agriculture, 2005).

Typical to its region in WA, the farm is dominated by dryland agriculture and regions of natural vegetation. The main crops are wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), oaten hay (*Avena sativa*), canola (*Brassica napus*), and lupins (*Lupinus angustifolius*), and pastures for animal grazing. Animal production includes Merino sheep for prime lamb and wool production and beef cattle. The WA Pig Skills Centre Pty Ltd is also on the property.

Certain advantages arise from selecting Muresk farm; yield data have been collected in the past, a complete database of paddock history exists, the farm has a variety of crop types and has topographic variability. Being part of Curtin University creates cohesion between the campuses and departments, and fosters future research relationships in GIS and agriculture. The farm is close to Perth (approximately 1 ½ hours drive) and can provide accommodation.

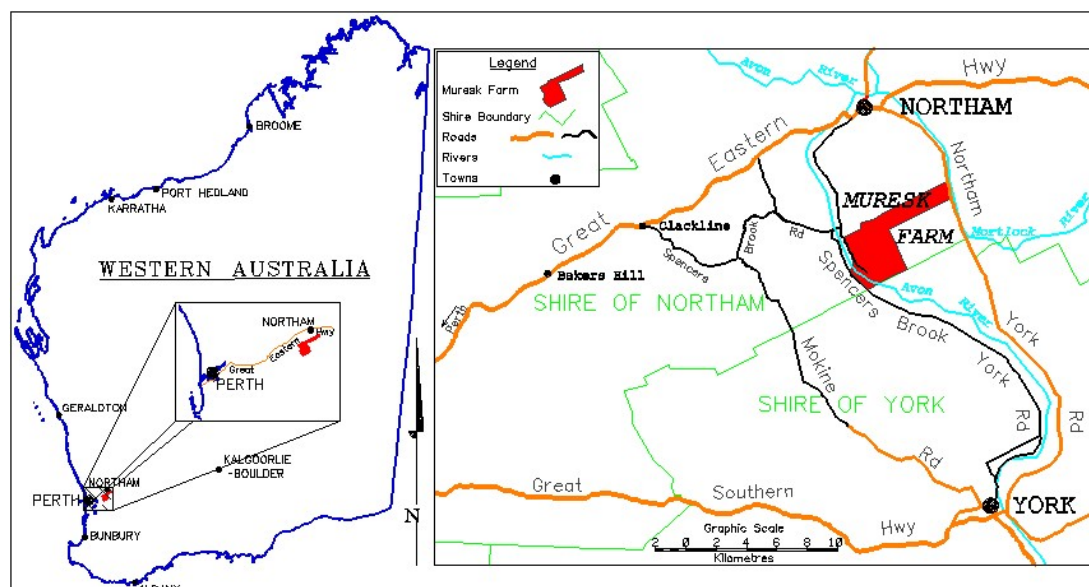


Figure 3.1 Study Area Location

### 3.1.1 Climate

The climate is Mediterranean with hot, dry summers and mild, wet winters (Lantzke and Fulton, 1993). Rainfall received for Northam (approximately 10km away) is on average, 440mm per annum. Average maximum temperatures range from 16.7°C in July to 34°C in January, while average minimum temperatures range from 5.6°C in August and 17.4°C in February (Weaving, 1999).

#### 3.1.1.1 Rainfall

The amount of rainfall falling within a year is an uncontrollable limiting factor for dry land agriculture. For Muresk the rainfall data are recorded at one point which is allocated to the entire study area (synonymous with metrological data in general); so spatially varying climate data is not available for analysis. As a consequence, it is not possible to include rainfall as a variable in the LMU classification. However, it needs to be considered as a factor when analysing variables that are greatly dependent on water availability, such as crop growth and subsequent yield, which have been recorded over several years.

Historical records of monthly rainfall have been recorded at Muresk farm from 1926 to 2001. The rain gauge is within the campus area of Muresk farm as indicated in Figure 3.2. However, no rainfall has been recorded at the Muresk gauge since 2001 and therefore daily rainfall data for 2002 and 2003 and monthly data in 2004 have

been obtained from neighbouring “Paradym” farm (Dymond, 2004) (Figure 3.2). Daily rainfall values for 2002 and 2003 are listed in Appendix A. There would be small differences between the monthly rainfall from Muresk farm and Paradym farm.

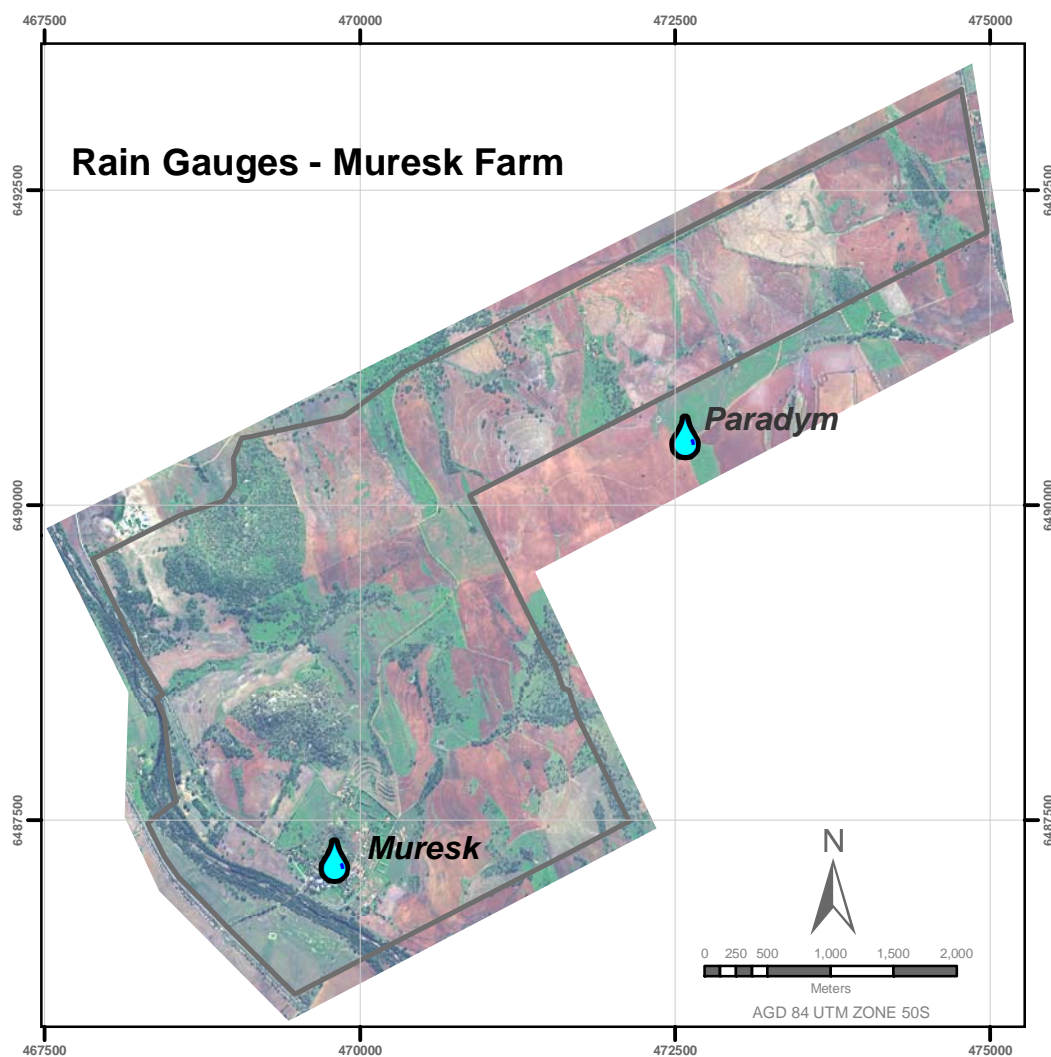


Figure 3.2 Rain gauge locations for Muresk Farm

The average monthly cumulative rainfall has declined at Muresk as indicated by the historical 76 year average (445mm), 30 year average (413mm), 15 year average (390mm) and 5 year average (352mm). On this basis the years included in the study 1996-2004 (for which yield data is also available) have been compared to the 15-year average rather than a longer term average for a sensible assessment.

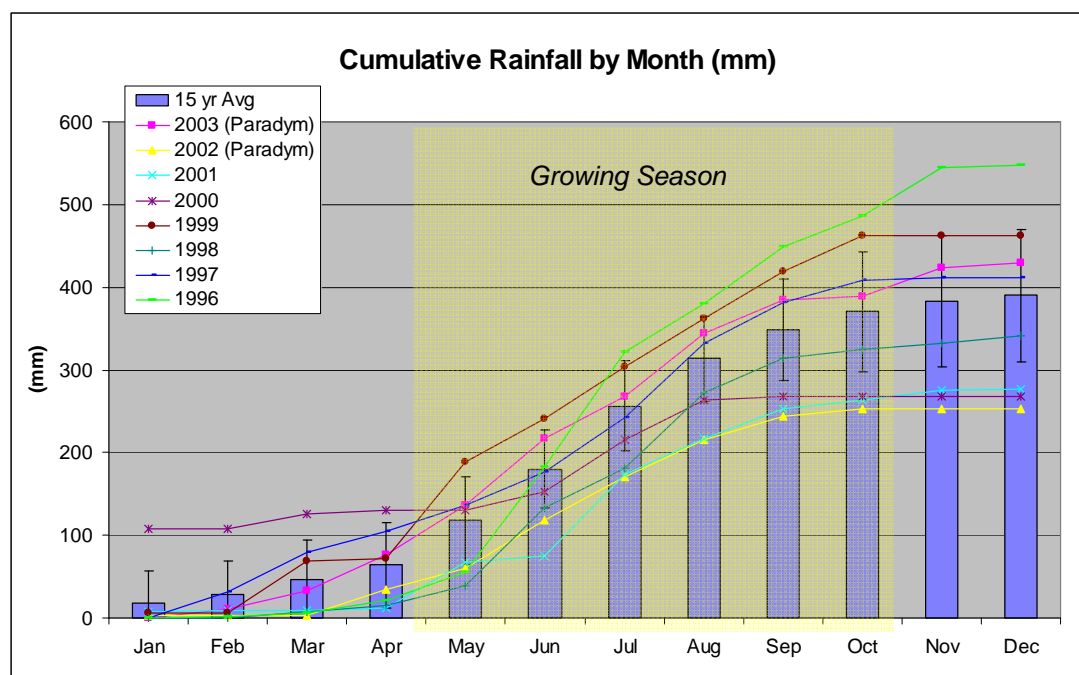


Figure 3.3 Cumulative rainfall by month for the years 1996 – 2003 as well as the 15 year average (1987-2001) with  $\pm$  standard deviation bars

The first year for which yield records are available, 1996, had well above average rainfall while the years 2000 and 2001 had well below average rainfall culminating with 2002, which had the lowest cumulative total rainfall since 1995. The remaining years, 1997, 1998, 1999 and 2003 had close to average total rainfall. What is also of interest for cropping programs is the growing season rainfall (GSR), as the amount of rain that falls during this time has the greatest bearing on crop growth. The GSR in the South West of WA is denoted by the period May – October (inclusive). This is highlighted as the yellow area and by the steeper gradient on the cumulative curves in Figure 3.3. The year 2000 in particular was a dry year for cropping. It had below average total rainfall (269mm), of which only 51 percent (138mm) fell within the growing season, while the year 2002 had an even lower total rainfall (253mm) but an above average percent (86 percent) falling in the growing season (Figure 3.4). The effect of the drier than average conditions in 2002 on yield was enhanced by the two preceding drier than average years, i.e. 2000 and 2001 (Figure 3.3 and Figure 3.4) leading to very low levels of soil moisture.

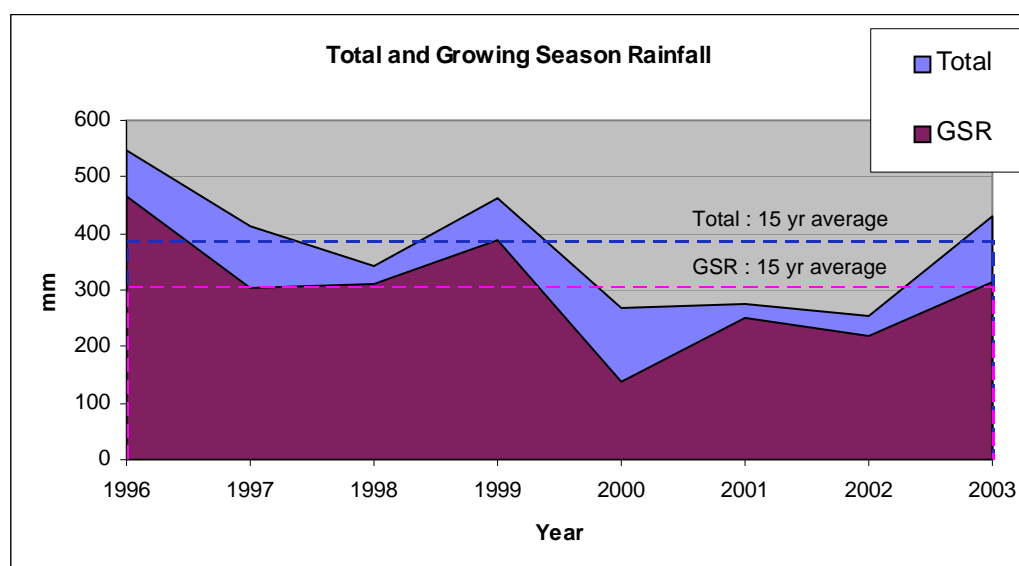


Figure 3.4 Total and Growing season rainfall for 1996-2003 with the 15 year average total and growing season rainfall shown as a dashed line

In summary,

- 1996 had above average rainfall in relation to recent years.
- 1997, 1998, 1999 and 2003 were equivalent to an average year (15 year average).
- 2002 was below average rainfall even in relation to recent years.
- The years 2000 and 2001 were also below average which would contribute to even drier conditions in 2002, through limited water reserves in the soil. The drier than average conditions in 2002 coupled with the below average rainfall in the preceding years could lead to anomalies in analysis with yield data for that year.

The amount of rainfall does not limit using the study area for this research, but it is recommended to consider these factors during statistical analysis. In particular, rainfall is important for understanding the variation in crop growth, which is discussed in Chapters 4 and 6.

### 3.1.2 Soils

Soil data are necessary for the study area to gain an appreciation for the landscape and highlight the detail that is currently available. The soil information was gathered from the Northam Region Land Resources Survey (Lantzke and Fulton, 1993), which is a regional survey intended for use at a scale of 1:100,000 and is currently the most

detailed survey available. The survey follows standard procedures, which divide the landscape using a six level hierarchy, namely: regions, province, zones, system, sub-systems and phases (Schoknecht and Tille, 2002). The features of the units within this hierarchy are summarised in Table 3.1.

Table 3.1 Description of Soil Survey hierarchy

NAME	DESCRIPTION	USEFUL SCALE(S)
Region	Broad subdivisions of the Australian continent.	Approx 1:20,000,000
Province	Provides a broad overview of the whole state.	Approx 1:5,000,000
Zone	Areas defined on geomorphological or geological criteria, suitable for regional perspectives.	Approx 1:1,000,000
System	Area with recurring patterns of landforms, soils and vegetation.	Approx 1: 250,000
Subsystem	An area of characteristic landforms features containing definite suites of soils.	1:150,000 - 1:100,000
Phase	Division of subsystems based on land use interpretation requirements.	1:100,000 – 1:25,000

The case study property, Muresk lies in the Northern Zone of Rejuvenated Drainage, which is characterised by dissection of the landscape forming steeper, narrower valleys which contain rivers and creeklines that flow every winter. Small remnants of sand plain occur, often bordered by a scarp or breakaway. In areas where the lateritic profile has been completely removed, such as in the Avon Valley, extensive areas of rocky, red and greyish soils have developed from fresh rock. The valley floors contain alluvial clays, loams and sands (Lantzke and Fulton, 1993).

Two soil-landscape systems are present on the Muresk property, namely the Avon Flats and Jelcobine. The Avon Flats system follows the Avon River, which cuts through the southern corner of the property and falls just outside the boundary running from the south to the northwest. The system is described as alluvial flats with brown loamy earth, grey non-cracking clay and brown deep sand vegetated by york gum (*Eucalyptus loxophleba*), salmon gum (*Eucalyptus salmonphloia*), flooded gum (*Eucalyptus rudis*) and river sheoak woodland (*Allocasuarina obesa*) (Lantzke and Fulton, 1993). The Jelcobine system covers the remainder of the property extending from the adjoining Avon Flats system at the south of the property to the northern corner. There is also an area of the Jelcobine system in the southern most corner of the property. The system is described as isolated steep low hills with

undulating low granite hills and isolated lateritic remnants with gravels, and grey shallow to deep sandy duplexes vegetated by york gum, jam (*Acacia acuminata*) wandoo (*Eucalyptus wandoo*), salmon gum and sheoak woodland (*Allocasuarina huegeliana*) predominate (Lantzke and Fulton, 1993).

Existing soil-landscape mapping (Lantzke and Fulton, 1993) has comprehensive detail of mapping units down to the subsystem level, with several phases mapped only where appropriate, and possible. The subsystem units are unique in terms of the proportions of soils they contain within each system. However, they often comprise of similar landforms and similar broad suites of soils. Table 3.2 describes each subsystem, its landform, major soil types and vegetation found within the study area. Figure 3.5 highlights the location of the soil subsystems described.

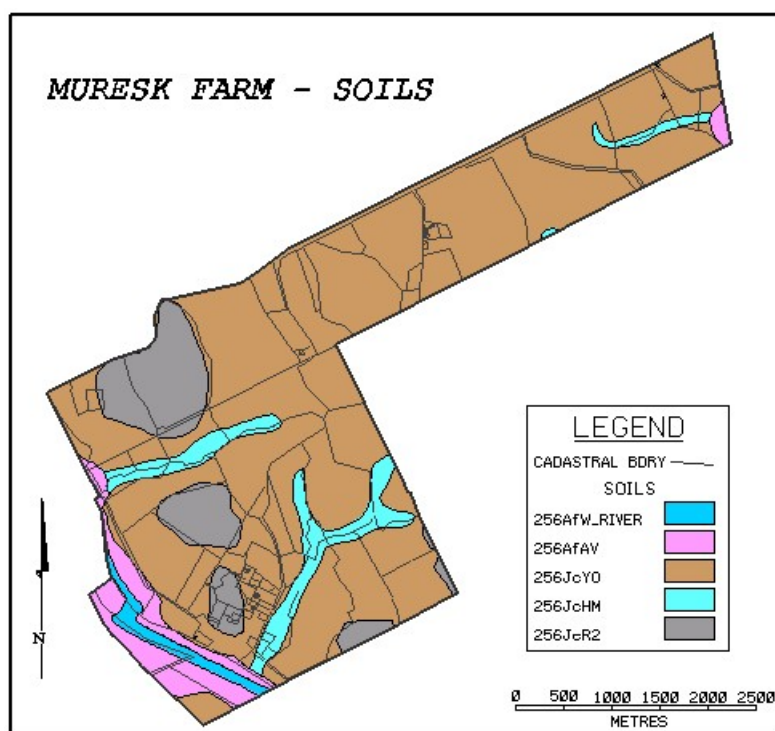


Figure 3.5 Muresk farm soils

Table 3.2 Description of subsystem, landform and major soil types present in the study area (Lantzke and Fulton, 1993)

Map Label	Subsystem Name	Landform	Major soils	Dominant Vegetation
256AfW_RIVER	Avon flat wet, river phase	River	Wet soils, water	NIL
256AfAV	Avon Flats	Alluvial floodplains which may be up to 2km wide in upstream areas but much narrower (<200m) downstream from Northam. The gradients of these values are about 1 in 250. The unit includes small areas of dunes that occur immediately adjacent to the river channel. Soil salinity is rare, though the river is quite salty.	Red brown alluvial loam (40%). Grey alluvial clay (30%). Orange alluvial loamy sand (20%).	Woodland containing Salmon gum, York gum, Flooded gum and Sheok
256JcYO	Jelcobine York	Irregular, often hilly country where streams or rivers have dissected the lateritic profile to expose bedrock. This unit occurs on the mid and lower slopes but can occur higher up in the landscape adjacent to rock outcrop. Slopes are generally in the order of 3 to 12%.	Rocky red brown loamy sand/sandy loam (65%). Brownish grey granitic loamy sand (15%). Red brown doleritic clay loam (15%). Rock outcrop (3%).	York gum and Jam woodland. Salmon gum grows on some of the heavier soil types while White gum and Sheoak generally grow on gritty, lighter soils.
256JcHM	Hamersley (Jc)	First and second order streams occur as midslope drainage lines. This unit possesses a 'V' shaped morphology but can have thin, alluvial terrace up to 20m wide. Rock outcrop is common. Some areas of this unit are saline. Gullies may form in the drainage line. Slopes of 1 to 6% occur along the water course.	Waterlogged grayish loamy sand/sandy loam (60%). Rocky red brown loamy sand/sandy loam (15%). Brownish grey granitic loamy sand (15%). Rock outcrop (10%).	York gum, Jam and Flooded gum in wetter areas.
256JcR2	Steep Rocky Hills 2 (Jc)	Steep sloping hills, which contain large areas of rock outcrop, generally occurring on the mid and upper slopes. Slope range from 5% to greater than 30%.	Rock outcrop (30%). Rocky red brown loamy sand/sandy loam (30%). Brownish grey granitic loamy sand (20%). Red brown doleritic clay loam (15%).	Jam, Sheok and York gum.



### 3.2 Remote Sensing Data

The remote sensing data used throughout this research are SpecTerra Services Digital Multi-Spectral Imagery (DMSI). The DMSI is an airborne high resolution remote sensing system captured using SpecTerra Services Digital Multi-Spectral Camera (DMSC) MkII. The DMSC MkII system comprises four 12-bit digital CCD cameras recording 1024 lines of 1024 pixels per line. Four interchangeable narrow band-pass interference filters were used to generate imagery in the blue, green, red and near-infrared bands. Band-pass filters down to 10nm in width, within the range 400-900nm are easily interchanged for specific applications, and the four spectral bands were nominally set as shown in Table 3.3. The system records the spectral characteristics of vegetation, soil and water from a single-engine light aircraft, such as a Cessna 182, flying between 380m and 3050m from the target producing images with a spatial resolution between 0.25m – 2m and radiometric resolution of 12 bits. Imagery was captured of Muresk Farm on four occasions during 2002 through to 2004 at a flying height of 3048m above sea level. This equates to an average cell resolution of 1.7m, resampled to 2m, the DMSC specifications are outlined in Table 3.3.

Table 3.3 DMSC Flight specifications

Band	Channel Name	Central Wavelength (nm)	Band Pass (nm)
1	Blue	450	20
2	Green	550	20
3	Red	675	20
4	Near-infrared	780	20

Image-to-image geo-referencing of the individual frames to historical ortho-rectified aerial photography was performed using a first-order polynomial warping and bicubic convolution resampling methodology. The mosaicking was performed using a technique based on cut-line feathering over 3 pixels (PCI Geomatica, 2003). The radiometric correction was carried out using an in-house developed software (SpecTerra Services, 2003b), based on inversion of the bidirectional reflectance model proposed by Roujean *et al.* (1992). Current corrections achieve a reduction of frame brightness from typically 20 percent of the dynamic range across individual frames, to less than 3 percent (SpecTerra Services, 2003a).

Atmospheric corrections were not performed on the data set, as the airborne data were collected on a clear dry day, close to noontime, when the solar zenith angle changes slowly with time. Although other researchers (Vermote *et al.*, 1997; Karnieli *et al.*, 2001) have applied atmospheric corrections on a remote sensing data set of characteristics similar to DMSI (e.g. in terms of remote sensing platform, spatial and spectral resolution), their studies show that the effects of the atmosphere on this type of airborne data are minimal; and given that clear sky conditions prevailed during data acquisition for this study, it is concluded that the atmospheric effects would also be minimal. However, image to image calibration has been performed where applicable and is discussed in the following sections.

### 3.2.1 Calibration of DMSI

The spectral data extracted from the four images, and in particular the June captured data, is used to analyse relationships between soil properties and DMSI. These comparisons require the digital numbers from each scene to be calibrated to common reference values, in a similar manner to those utilising images from different dates for change detection analysis (Furby and Campbell, 2001).

The calibration technique used in this study is similar to that of Furby and Campbell (2001), and it has been adopted because it is a relatively simple method using regression techniques. Furby and Campbell (2001) developed and implemented a ‘like-values’ calibration procedure that uses robust regression and large numbers of targets which are likely to be invariant over time. Their method converts the raw digital numbers of an image(s) so that they are consistent with the digital numbers of a reference image. This technique is an extension of the scene-to-scene correction methodology proposed by Casselles and Garcia (1989), which is designed to be robust against targets that are not truly invariant (Furby and Campbell, 2001). The calibration is based on the assumption that the relationship between the digital numbers for any image and the digital numbers of a chosen reference image for the same scene are linear, following;

$$R_i = g_i DN_i + o_i \quad (3.1)$$

as derived by Casselles and Garcia (1989) where  $R$  is the digital number of the target in the reference image,  $DN$  is the digital number of the target in the raw image,  $g$  and  $o$  are the gain and offset derived from the linear relationship, used to calibrate the

raw image data values, and  $i$  signifies the image band (Furby and Campbell, 2001). The gain and offset account for atmospheric, sensor and on-ground processing differences (Furby and Campbell, 2001).

The ‘like-value’ calibration procedure consists of five steps, namely: (a) selecting a reference image; (b) selecting invariant targets; (c) calculating the calibration coefficients to calibrate each image to the reference image; (d) examining the calibration curves and refine the target selection if necessary; and (e) using the estimated coefficients to calibrate the image. These steps are discussed in more detail in Sections 3.2.1.1 to 3.2.1.5.

### 3.2.1.1 Select a Reference Image

The June flights are determined to be the most suitable for soil studies in this research because they have been captured when limited vegetation cover was present on the ground. It is important that the reference image is (i) cloud free, (ii) the atmosphere is fairly clear, (iii) the data storage are within the format range for all bands, (iv) the time of year is appropriate for the application and (v) the image has the best possible dynamic range of pixel values. This concurs with the principles of selection advocated by Furby and Campbell (2001). As all images are free of cloud and haze and acquired from the same image source (DMSI), the choice of reference comes down to the data storage format, dynamic range of pixel values and time of image flight. Table 3.4 provides the dynamic range of the pixel values for each band for each flight.

Table 3.4 Dynamic range of pixel values for DMSI

<b>Flight Date</b>	<b>Flight No.</b>	<b>Band 1 Blue</b>	<b>Band 2 Green</b>	<b>Band 3 Red</b>	<b>Band 4 NIR</b>
28 <sup>th</sup> June 2002	1	0-4222	0-4541	0-3993	0-2214
17 <sup>th</sup> Sept 2002	2	0-4422	0-4672	0-3902	0-2319
3 <sup>rd</sup> Sept 2003	3	0-4128	0-4128	0-4118	0-3660
14 <sup>th</sup> June 2004	4	0-2004	0-3272	0-2825	0-1215

As the imagery is 12-bit radiometric resolution, a maximum range of 4096 is possible (0-4095). However, in Table 3.4 it can be seen that some of the bands exceed this range. This is due to offsets being applied during bidirectional reflectance correction (SpecTerra Services, 2003b) and whiteouts (where the reflectance value exceeds the maximum recorded digital number). Therefore, any

value greater than 4095 can be truncated to 4095. From Table 3.4, Flight 4, captured during June 2004, is used as the reference image as the data range are all within 0-4095 for all bands (minimising the uncertainty in bright targets) and it is an appropriate time of year for image capture. However, this comes at the expense of compressing the digital range, as the Flight 4 image has the minimum range of pixels in contrary to Furby and Campbell's (2001) suggestion of using a reference image where the range of digital numbers is maximum preventing the range being compressed further.

### 3.2.1.2 Select Invariant Targets

Invariant targets are features that have constant reflectance over time, and they should cover a range of bright, midrange and dark data values. There are various automatic and manual methods for target selection (Hall *et al.*, 1991; Furby and Campbell, 2001). The manual technique can be labour intensive and problematic, particularly if the images utilised have a significant time lapse (several years) as many targets may have changed due to land use change (Schott *et al.*, 1988; Hall *et al.*, 1991). Automatic techniques also have limitations, such as the method used by Hall *et al.* (1991), which finds only dark and bright features (Furby and Campbell, 2001). In this study, the manual target selection technique has been used, as there is a good knowledge of the study area's geographic features and land use from each date of image capture, giving rise to appropriate target selection for dark, mid and bright ranges. Once the labour-intensive selection of targets is completed, repeat calibrations can be generated easily. The number and size of bright and dark targets should be balanced in order to estimate the calibration coefficients successfully using regression. Furby and Campbell (2001) suggest selecting 10-20 targets that contain four to nine pixels with good geographic distribution across the image. Some suggested dark targets in the Western Australian wheatbelt include lakes, water in dams and reservoirs. Potential mid-range targets are rock outcrops, airfield runway crossings, quarries, gravel scrapes and open mines, bright targets could be selected from roaded catchments, beach sand and bare ground. However, vegetated targets should be avoided as they are inclined to show seasonal trends (Furby and Campbell, 2001). The presence of midrange targets allows for the confirmation of a linear relationship between the images (Furby and Campbell, 2001).

The targets selected consisted of man-made and natural features and each target contained a minimum configuration of two-by-two pixels. Narrow targets were avoided where possible in order to eliminate mixed pixel effects. During the selection process the reference image (Flight 4) was used as a base for *heads-up digitizing* the targets. Other images (Flight 1,2 and 3) were overlaid and pixels analysed to assist in the selection of targets that were consistent across each image. Table 3.5 lists the features that were selected for each target and its predefined range, while Figure 3.6 displays the geographic location of the selected targets used in the following regression analysis.



Figure 3.6 Geographic distribution of target locations overlaid on reference image (DMSI June 2004). Where B, represents bright targets; M, the midrange targets; and D, represents dark targets

Table 3.5 Targets selected for calibrations. Only those highlighted in grey were used in the final estimation of calibration coefficients

Range	Target No.	GIS Layer No:	Feature	No. of Pixels	
Dark Targets	D1	1	1	River water	8
	D2	2	2	Dam water	20
	D3	3	3	Dam water	24
	D4	4	4	Dam Water (limited amount of water not consistent across all images)	9
	D5	5	5	River water	16
	D6	6	6	Water in dam	6
	D7	7	7	Water in Dam	16
	D8	8	8	Water in Dam	12
	D9	9	9	Shadow (are not good targets)	4
	D10	10	10	Water in Dam (Sept 03 looks empty of water in image)	6
Mid Targets	M1	11	1	Bitumen roadway	7
	M2	12	2	Gravel track	9
	M3	13	3	Rocky outcrop	12
	M4	14	4	Bitumen Road crossway	12
	M5	15	5	Gravel track	9
	M6	16	6	Gravel track	8
	M7	17	7	Gravel track	6
	M8	18	8	Bitumen road	12
Bright	B1	19	1	Dam bank/bare ground (whiteout)	6
	B2	20	2	Rock outcrop	20
	B3	21	3	Bare sand	9
	B4	22	4	Sand/roaded catchment	11
	B5	23	5	Roaded catchment	9
	B6	24	6	Building roof (whiteout)	8
	B7	25	7	Building roof (whiteout)	32
	B8	26	8	Building roof (whiteout)	9
Bright New	B2_1	27	1	Sand in river (small patch and inconsistent across images)	4
	B2_2	28	2	Sand in river	6
	B2_3	29	3	Sand near tree (Not a good target small and inconsistent)	4
	B2_4	30	4	Cement slab	9
	B2_5	31	5	Sand contour bank	6

The pixel values for each band for each flight of imagery were extracted using the zonal statistics module of Spatial Analyst in ARCGIS. The median value for each target was used for regression analysis. Furby and Campbell (2001) would typically use the values from each pixel in each target and ensure that the number of pixels, rather than the number of targets, is balanced if the sizes are uneven. However, Furby (2004) suggests use of the median for each target as the most robustly representative measure.

### 3.2.1.3 Calculate the Calibration Coefficients

The calibration coefficients were calculated by the least-squares approach following Equation (3.1) where  $R$  is the digital number of the target in the reference image,  $DN$  is the digital number of the target in the raw image. The least-squares approach was used because it is simpler to apply than other methods (unweighted least-squares regression and robust regression) if a good set of targets are used. Furby and Campbell (2001) demonstrate the calculation of coefficients using two approaches: (i) unweighted least-squares regression and (ii) robust regression based on S-estimation followed by weighted least-squares regression. The robust regression procedure is proposed and utilised by Furby and Campbell (2001) because it is assumed that up to 50 percent of selected targets are not invariant, and those targets that change between image and reference image are automatically omitted from the calculation. Robust regression approaches have been shown to be less affected by outliers caused by clouds and over emphasis on dark and bright targets (Furby and Campbell, 2001). The unweighted least-squares approach is demonstrated by Furby and Campbell (2001) for comparison purposes and to highlight the strong influence of atypical values from targets that have changed between image dates on the calibration line. The authors note that if a good set of targets is used then there is very little difference between the S-estimation and weighted least-squares calibration lines. While, it can be visualised in the scatter plots provided by Furby and Campbell (2001) that the same would apply between robust regression and unweighted least-squares with appropriate target selection. As stated in Section 3. 2, the images in this study are cloud free, and the manual selection of targets has ensured that the proportion, type and variant nature of targets is carefully controlled and recorded (Table 3.5). Although considered, the robust regression approach was not utilised for the above mentioned reasons and the least-squares approach is demonstrated hereafter. The median digital number (pixels values) were imported into GenStat® (Laws Agricultural Trust, 2003a) for least-squares linear regression.

### 3.2.1.4 Examine Calibration Curves and Adjust Targets

Following a least squares approach, initial regression equations were formed. It was clear from the residual plots and a re-assessment of the targets that several targets were not appropriate due to their variant nature across images and pixel whiteouts (Table 3.5). They were therefore removed from the data set and new targets selected.

Table 3.5 lists all targets and those used in the following analysis are highlighted. Figure 3.7 to Figure 3.9 provide the scatter plots showing the least squares regression calibration line.

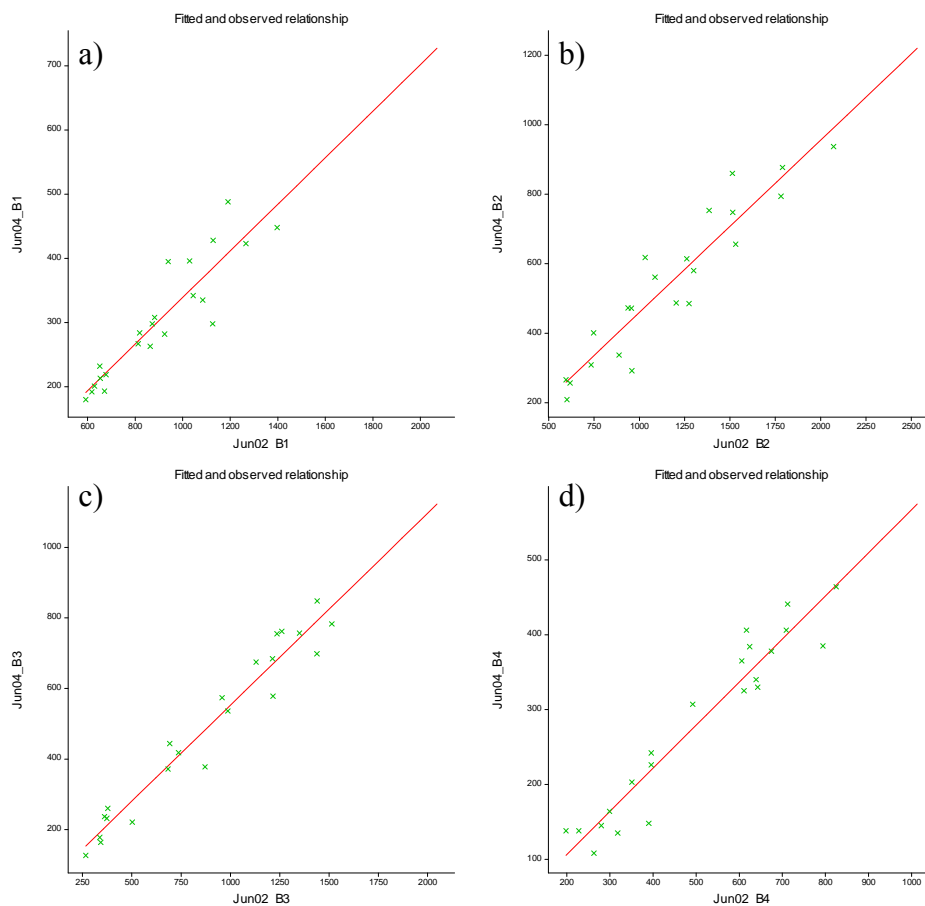


Figure 3.7 Scatter plots showing the least squares regression line for June 2002 imagery for a) Band 1, b) Band 2, c) Band 3 and d) Band 4. (y axis = reference image, x axis = raw image)



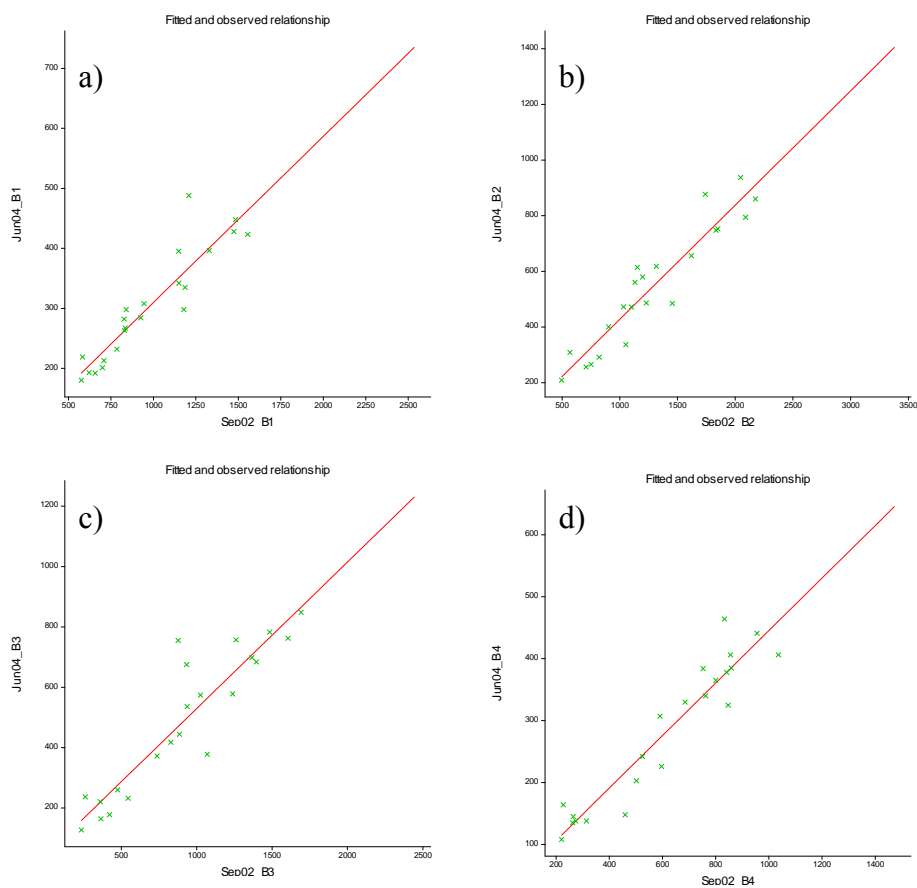


Figure 3.8 Scatter plots showing the least squares regression line for September 2002 imagery for a) Band 1, b) Band 2, c) Band 3 and d) Band 4 (y axis = reference image, x axis = raw image)

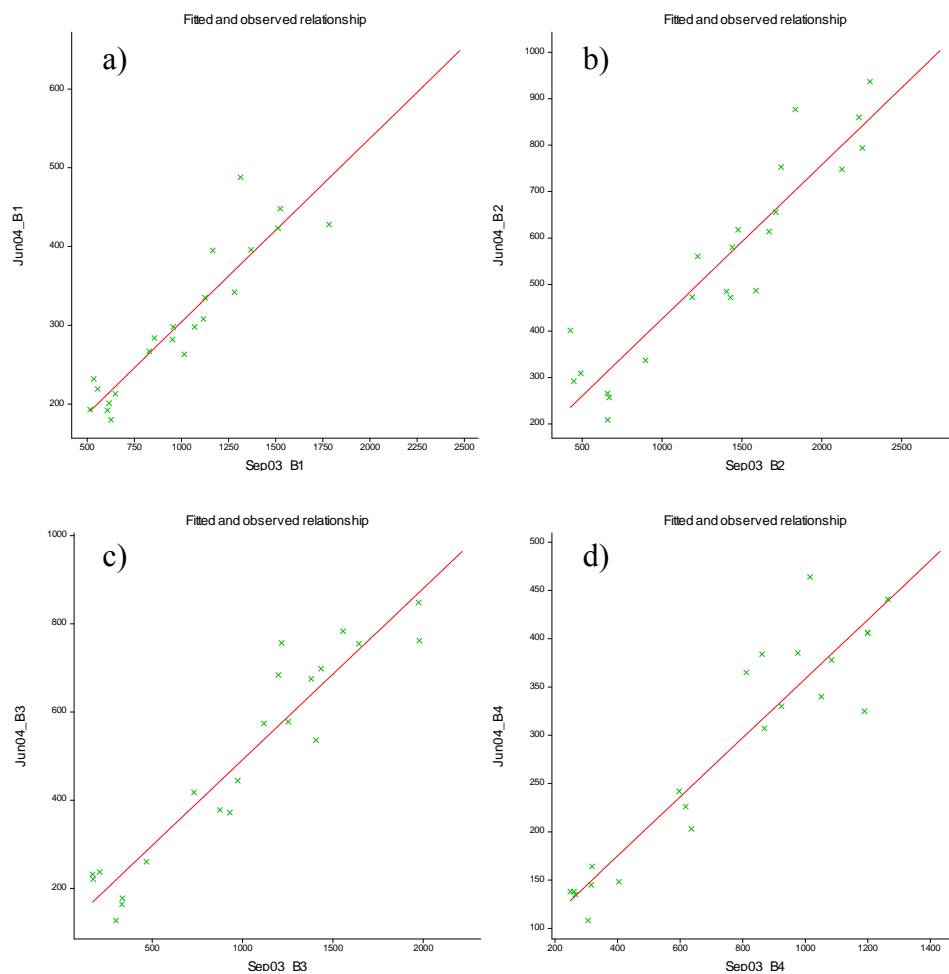


Figure 3.9 Scatter plots showing the least squares regression line for September 2003 imagery for a) Band 1, b) Band 2, c) Band 3 and d) Band 4 (y axis = reference image, x axis = raw image)

Table 3.6 provides the least-squares regression lines, percentages of the variance in the reference bands from June 2004 that each model accounted for, the residual standard errors, and significance.

Table 3.6 Least-squares regression model for image calibration

<b>Band</b>	<b>Model:</b> $R_i = gain_i DN_i + offset_i$	<b>Adjusted R<sup>2</sup></b>	<b>Standard Error</b>	<b>P</b>
June 02 – B1	$R_{B1} = 0.3629_{B1} DN_{B1} - 23.8_{B1}$	82.5	38.6	<0.001
June 02 – B2	$R_{B2} = 0.4960_{B2} DN_{B2} - 36.4_{B2}$	87.1	78.6	<0.001
June 02 – B3	$R_{B3} = 0.5439_{B3} DN_{B3} + 8.7_{B3}$	94.5	55.5	<0.001
June 02 – B4	$R_{B4} = 0.5760_{B4} DN_{B4} - 8.9_{B4}$	91.6	33.9	<0.001
Sept 02 – B1	$R_{B1} = 0.2771_{B1} DN_{B1} + 32.5_{B1}$	84.0	36.9	<0.001
Sept 02 – B2	$R_{B2} = 0.4104_{B2} DN_{B2} + 17.3_{B2}$	88.1	75.5	<0.001
Sept 02 – B3	$R_{B3} = 0.4851_{B3} DN_{B3} + 44.4_{B3}$	82.9	98.2	<0.001
Sept 02 – B4	$R_{B4} = 0.4242_{B4} DN_{B4} + 21.2_{B4}$	89.5	38.0	<0.001
Sept 03 – B1	$R_{B1} = 0.2333_{B1} DN_{B1} + 70.8_{B1}$	84.3	36.6	<0.001
Sept 03 – B2	$R_{B2} = 0.3318_{B2} DN_{B2} + 94.1_{B2}$	85.4	83.6	<0.001
Sept 03 – B3	$R_{B3} = 0.3884_{B3} DN_{B3} + 103.1_{B3}$	89.1	78.4	<0.001
Sept 03 – B4	$R_{B4} = 0.3061_{B4} DN_{B4} + 52.4_{B4}$	87.3	41.8	<0.001

### 3.2.1.5 Calibrate the Image

The raw band images are calibrated to the reference image by applying Equation (3.1) in accordance with the appropriate gain and offset. The gains and offsets applied to each band, derived from the regression line equation, are based on the shift required to make the mean and standard deviation of the raw pixel set match the reference pixel set (Yuan *et al.*, 1999).

Table 3.7 displays the statistics and highlights that the mean and standard deviation of the reference and calibrated image are more closely matched. The mean and standard deviation of the pixel values have been calculated and extracted for each band within the reference image, raw image and calibrated image. For consistency, a mask layer was used to ensure the same spatial extent was analysed for each band of imagery.

Table 3.7 Raw, Reference and Calibrated image statistics

Image		Band	Mean	Standard Deviation
June 2002	Raw June 02	B1	773.93	81.13
	Reference June 04		247.98	29.15
	Calibrated June 02		257.06	29.44
	Raw June 02	B2	981.23	152.46
	Reference June 04		420.75	72.20
	Calibrated June 02		450.29	75.62
	Raw June 02	B3	612.06	215.74
	Reference June 04		344.18	93.75
	Calibrated June 02		341.60	117.34
	Raw June 02	B4	631.26	163.64
	Reference June 04		294.66	60.59
	Calibrated June 02		354.70	94.26
Sept 2002	Raw Sept 02	B1	754.42	101.34
	Reference June 04		247.98	29.15
	Calibrated Sept 02		241.55	28.08
	Raw Sept 02	B2	1017.28	184.57
	Reference June 04		420.75	72.20
	Calibrated Sept 02		434.79	75.75
	Raw Sept 02	B3	540.78	198.54
	Reference June 04		344.18	93.75
	Calibrated Sept 02		306.73	96.31
	Raw Sept 02	B4	870.13	165.73
	Reference June 04		294.66	60.59
	Calibrated Sept 02		390.31	70.30
Sept 2003	Raw Sept 03	B1	702.60	105.12
	Reference June 04		247.98	29.15
	Calibrated Sept 03		234.72	25.52
	Raw Sept 03	B2	1069.84	190.08
	Reference June 04		420.75	72.20
	Calibrated Sept 03		449.07	63.07
	Raw Sept 03	B3	453.60	210.42
	Reference June 04		344.18	93.75
	Calibrated Sept 03		279.28	81.73
	Raw Sept 03	B4	1187.45	270.62
	Reference June 04		294.66	60.59
	Calibrated Sept 03		415.88	82.84

### 3.3 Digital Elevation Model

The quality of digital elevation model (DEM) has a strong influence on the reliability of topographic attributes derived from them, and the effect is amplified for secondary topographic attributes. A DEM covering the study area was provided by the Department of Agriculture and Food-Western Australia. The Department of Land Information (DLI) is the custodian of the original DEM data, which were created as

part of the National Dryland Salinity Program under the guidance of the Land Monitor Group. Heights have been derived with a vertical resolution of 0.01m on a 10m grid from stereo aerial photography flown at 1:40,000 scale, using soft copy automatic terrain extraction (image correlation) techniques (Allen and Beetson, 1999). The correlation technique has an expected vertical accuracy of +/-1.5m at the 90 percent confidence level. Due to the automated processes used, isolated large errors may be found in this data due to poor correlation in paddocks, large rock outcrops, sharply rising breakaway country, sand dunes and heavy tree covered areas (Department of Land Information, 2000). Although the DEM was produced within the accuracy specifications, errors are apparent in the Muresk study area and can be visualized in a shaded relief as shown in Figure 3.10a. This undesirable patchwork effect is illustrated in Figure 3.10a, which highlights systematic errors that need to be minimized prior the derivation of topographic attributes.

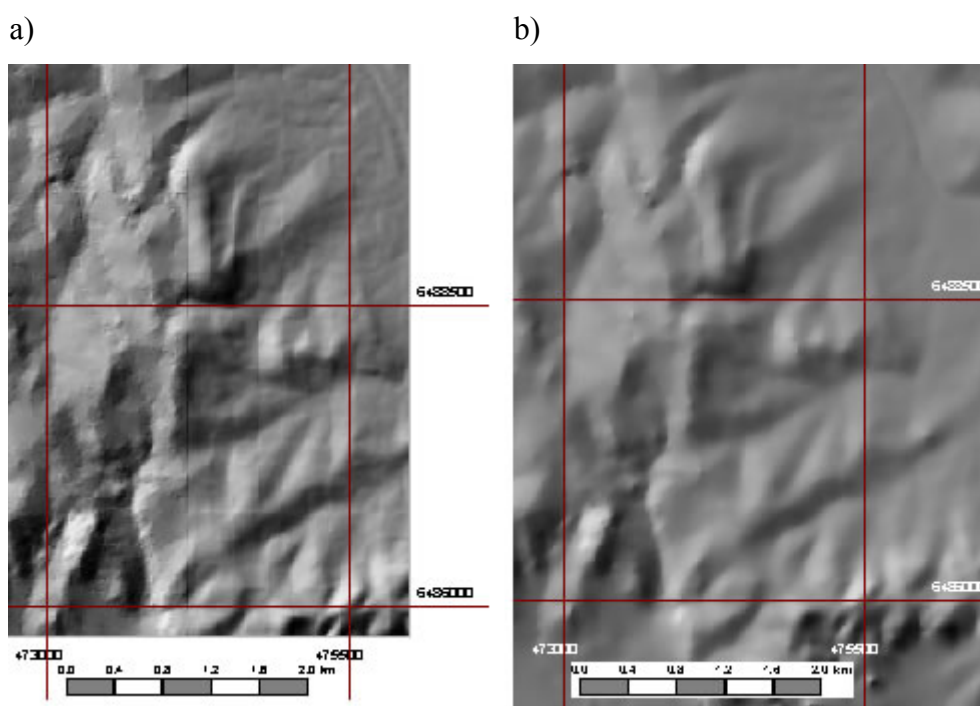


Figure 3.10 DEM systematic error in the form of tiling and a seam running north-south. a) Shaded relief of Landmonitor DEM, b) Shaded relief of smoothed Landmonitor DEM after algorithm applied by Cacetta (2000)

This research endeavours to utilise any existing data sets available and as such the output can be applied across spatially separable areas of WA for which these data sets cover. Klingseisen (2004a) tested several methods for removing errors, taking

into account the findings of López (2002) and in particular investigated the methods of Hannah (1981) and Felicísimo (1994). Klingseisen (2004a) found that both methods were limited to the reduction of non systematic errors and, when applied to the Muresk study area, the prominent terrain edges were over-smoothed while the visible patchwork remained largely unaffected.

The Muresk DEM was smoothed to reduce the discontinuities using Caccetta's (2000) adaptive iterative filter. Caccetta (2000) developed a method to reduce the systematic errors in the Land Monitor DEM. The algorithm regulates the level of smoothing via a height change criteria whereby more smoothing is applied in flat areas and less in non flat regions or the level of smoothing is reduced on areas of strong curvature (either profile, tangential or plan curvature) (Caccetta, 2000). The portion of the smoothed DEM of Muresk is displayed in Figure 3.10b highlighting the improvement post application of Caccetta's (2000) algorithm.

### **3.4 Yield Data**

Yield data were recorded using the GreenStar™ yield monitor, by John Deere Limited, which is coupled with an OmniSTAR differential global positioning real time system, mounted on a harvester. The yield monitor is equipped with a mass-flow sensor, which measures the slightest movement, or impact of grain flow from the top of the clean-grain elevator. This measurement, along with the clean-grain elevator speed, is used to calculate harvested mass (Deere and Company, 2002).

Point yield data were collected at 1-second intervals along the path of the harvester with a swath width of 7.493m. The data were downloaded into JDOffice (version 1.3) software (John Deere Ag Management Solutions, 2003) then exported as an XYZ format (Easting, Northing, kg/ha) for analysis. The raw yield data contain errors in yield estimates from a combination of sources which have been reviewed and mitigated as described in Robinson (2004) and Robinson and Metternicht (2005). As a consequence, the filtered yield datasets were interpolated onto a 4m grid using block kriging and local variograms (Robinson, 2004) within the VESPER (version 1.5.8) software (Minasny *et al.*, 2002). This produces interpolated yield surfaces for the available paddocks with an associated variance layer.

Table 3.8 provides a list of the number of years of yield data available for each paddock and the crop type associated with that years data. The yield monitor was not used during 2001 and therefore no yield data were collected during that year. Table 3.8 highlights that for some paddocks there is complete coverage of 7 years of data while for other paddocks yield data have never been collected. This suggests that these paddocks might not be arable or are used for other farming purposes such as grazing.

Post-filtering the field recorded point data is used for interpolating onto a 4m grid (in this case). Figure 3.11 displays an interpolated yield surface with the filtered field recorded point yield data overlaid. It can be visualised in Figure 3.11 that for some areas of the paddock there is a higher density of point values that will contribute to the interpolated value.

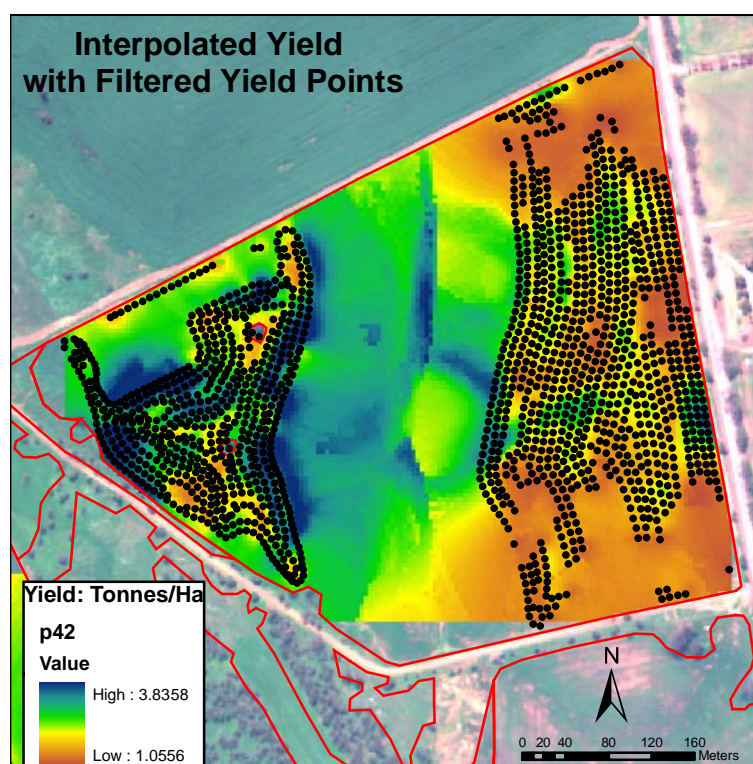


Figure 3.11. An example of an interpolated yield surface overlaid with filtered yield points and paddock boundary

Table 3.8 Interpolated yield data available for Muresk Farm

Paddock No.	1996	1997	1998	1999	2000	2002	2003	Total Yield Layers
1	Wheat	Lupins	Wheat	Canola	Wheat	Wheat	Wheat	7
2 and 3	Lupins		Faba Beans		Wheat	Wheat	Wheat	5
4								0
5	Oaten Hay	Lupins		Canola	Oats/Wheat	Wheat	Wheat	6
6							Wheat	1
7	Wheat	Oats/Oaten Hay	Barley	Lupins	Wheat	Canola	Wheat	7
8								0
9				Barley				1
10						Oaten Hay	Wheat	2
11			Lupins				Wheat	2
12	Faba Beans	Wheat	Lupins		Barley	Wheat		5
13 and 14	Faba Beans	Barley	Lupins		Barley	Wheat		5
15								0
16								0
17		Lupins			Oats		Barley	3
18	Wheat	Barley	Lupins		Lupins	Barley		5
19								0
20			Oats					1
21			Wheat		Canola	Barley		3
22	Barley	Lupins	Wheat		Wheat	Barley	Canola	6
23						Wheat		1
24	Lupins	Wheat		Canola			Wheat	4
25	Barley	Lupins				Canola	Wheat	4
26	Wheat	Lupins	Wheat	Canola		Lupins	Wheat	6
27						Oats	Oats	2
28	Wheat	Wheat				Canola	Wheat	4
29			Wheat		Canola	Barley		3
30	Wheat/Barley	Wheat/Barley	Lupins	Oats	Canola		Barley	6
31								0
32	Wheat	Faba Beans			Barley	Wheat	Wheat	5
33	Lupins		Lupins		Lupins	Wheat		4
34			Wheat		Wheat	Oats	Oats	4
35			Wheat		Wheat	Barley	Lupins	4
36		Wheat	Wheat		Wheat	Lupins	Wheat	5
37		Faba Beans/ Lupins			Lupins	Wheat	Canola	4
38	Wheat				Wheat		Wheat	3
39	Oaten Hay	Barley	Lupins		Wheat	Wheat	Lupins	6
40					Wheat	Wheat		2
41								0
42	Faba Beans	Oats	Barley		Wheat	Wheat	Wheat	6
43			Faba Beans		Wheat	Wheat	Wheat	4
Total Paddocks	18	18	20	7	23	25	24	136

Shaded cells have available yield data however no soil sampling points fall in the region for that year.

Any area outside the perimeter of yield points will in fact extrapolate values rather than interpolate and thus these areas must be discarded from analysis. The paddock



boundary is displayed in red in Figure 3.11. VESPER provides a variance layer associated with the kriged interpolated layer, which can be used to guide the user as to the degree of confidence one can have with the interpolated value. Although yield data have their limitations, the best possible care has been taken during the interpolation process and subsequent extraction of the yield data to avoid any undue error that may be associated with the data. These yield data are the only ones available for the study area and as such make a valuable contribution to the research.

### 3.5 Field Data

Field data were collected on several occasions at Muresk farm to build up a database of vegetation and soil attributes. The two main streams of field data sets are the soil sampling and analysis and vegetation sampling and analysis, which are both described in Chapter 4. A field visit was also made synchronous with DMSI capture during June 2002 and June 2004 to provide an indication of the vegetation cover present. Attributes recorded include the spatial location, vegetation type, vegetation density, digital photo and any other relevant comments. Approximately 30 locations have data recorded for each year. Figure 3.12 provides visual display of the 1m<sup>2</sup> quadrants in which vegetation data were recorded along with associated attributes.

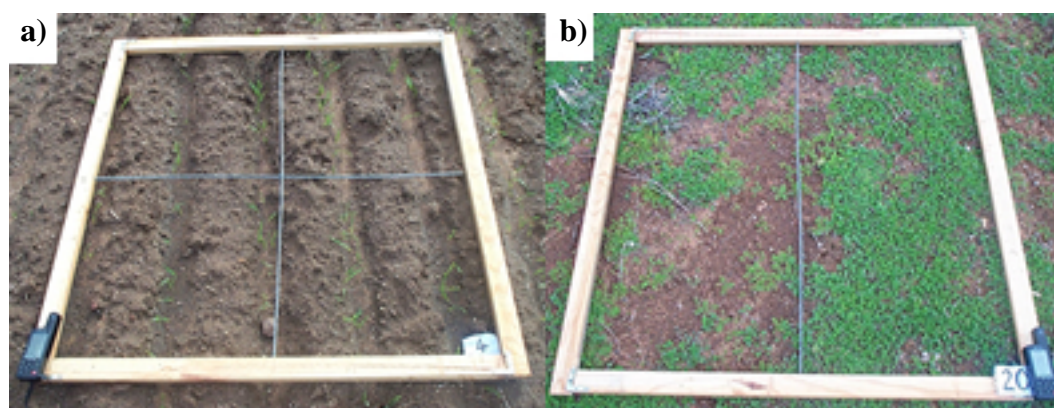


Figure 3.12 Field data collection during June 2004 DMSI flight. a) sample point in cropped paddock with <2 percent vegetation cover. b) sample point in paddock with pasture and 55 percent vegetation cover

### **3.6 Software and Hardware**

#### **3.6.1 Global Positioning System (GPS)**

The Department of Agriculture and Food-Western Australia in South Perth provided the Global Positioning System (GPS) used to locate field sample locations throughout the research. The Trimble GeoExplorer II is a differential GPS system (dGPS) and consists of a hand-held receiver that is taken into the field to collect point, line or area positions. The data are then taken back to the office to be processed for differential correction. Differential correction is a technique that uses an extra GPS receiver (base station) with a known position to increase the accuracy of a GPS position. The dGPS can provide accuracies of sub-metre to 5m Circular Error Probable, meaning 50 percent of the collected points are within a 5m-radius circle on a horizontal plane (Trimble Navigation Limited, 1994). The dGPS configuration was set to record data in the AMG coordinate system and WGS84 datum. During the post-processing, the data are transferred to AGD84 datum to match other spatial data sets used throughout this research.

#### **3.6.2 GIS Software**

Several suites of GIS software have been used throughout this research namely, Intergraph GeoMedia Professional (Intergraph Corporation, 2003) with GeoMedia Grid (Keigan Systems Inc., 2003a), MF Works (Keigan Systems Inc., 1999) and ARCGIS (ESRI Inc, 2004). They have been used on a variety of different levels and where appropriate highlighted throughout the thesis.

#### **3.6.3 Statistical Software**

SPSS for Windows (SPSS Inc., 2003) and Genstat® (Laws Agricultural Trust, 2003a) are the two statistical packages used during this research. Genstat® is the dominant package and was used as the basis for the LMU classification described in Chapter 7.

### **3.7 Summary**

In this chapter the selected study area, Muresk, has been described in terms of; climate, soil types and land use. DMSI data for Muresk has been captured on 4 occasions and provided at a spatial resolution of 2m, along with a DEM of 10m and

yield interpolated surfaces (4m) where available. The GPS, GIS and statistical software used throughout this research are mentioned.

The following Chapter 4 uses the described software based on the Muresk case study region, and explores the potential of DMSI data for detecting crop growth variability. Chapter 4 describes the soil sampling strategy design and the collection and chemical analysis of soil properties identified in Chapter 2 as relevant for the formation of LMUs.

## CHAPTER 4

### FIELD DATA COLLECTION AND ANALYSIS

*This chapter presents the strategy adopted for the soil sampling design and its implementation. Explanations of the field data collected and chemical analysis performed are also included. This phase of the research culminates in a soil spatial database to be utilised for exploring relationships between soil properties, yield, topographic attributes and Digital Multi-Spectral Imagery (DMSI), which are explained in Chapter 6.*

*The chapter also analyses the potential of DMSI for detecting the variability of crop leaf area index (LAI) within several paddocks sown to wheat, lupins and canola. The aim is to examine whether DMSI can be used as a layer for detecting variability in crop growth and should thus be incorporated into the LMU classification.*

*Transect sampling within wheat, lupin and canola crops was conducted synchronous with the capture of DMSI for two consecutive years, 2002 and 2003. Statistical analysis between DMSI spectral values and crop LAI has been performed and this chapter highlights phenological factors affecting the relationship between LAI and spectral values.*

#### **4.1 Soil Sampling Design**

When determining the appropriate number of soil samples, consideration must be given to the practicability and economic feasibility of the collection and analysis process. The density of soil observations in Australia is recommended to be a minimum density of 0.25 ground observations per  $\text{cm}^2$  of published map, as suggested by FAO in 1979 (Gunn, 1988). This equates to between 25 to 100 soil observations per  $\text{km}^2$  for published maps scales from 1:10,000 to 1:5,000 respectively (Gunn, 1988). However, when integrating remotely sensed data into the sampling design, Metternicht *et al.* (2002) showed that a reduction of 60 to 80 percent of field observations can be achieved when compared to traditional survey methods. In this research, 250 samples have been determined to be adequate in the study area, which covers approximately 1780ha ( $17.8\text{km}^2$ ). This equates to

approximately 14 soil observations per km<sup>2</sup> which, after taking into account a reduction in field observations of approximately 60 percent due to the integration of high resolution remote sensing and terrain information, is equivalent to 35 observations per km<sup>2</sup>.

The farm scale sampling methodology is a refinement of the paddock scale sampling strategy approach of Drysdale *et al.* (2002), which used slope in a combination with NDVI to design an optimal soil sampling strategy. In this current research slope is used to subdivide the landscape into slope strata (classes) based on Australian soil survey standards (McDonald *et al.*, 1990) as described in Table 4.1. Slope on Muresk farm ranges from 0 to 28 percent (i.e. Class 1-4).

Table 4.1 Definition of slope classes (adopted from McDonald *et al.*, 1990)

Class	% Slope	Name
1	<1%	Level
2	1%-3%	Very gently inclined
3	3%-10%	Gently Inclined
4	10%-32%	Moderately Inclined
5	32%-56%	Steep

The DMSI captured across the Muresk farm provided some indication of the variability through spatial variations reflected in the imagery. The NDVI (Rouse *et al.*, 1973) was used as a surrogate measure to aid in determining variations in soil conditions in areas of high vegetation. Similarly, the Redness Index (RI) (Escadafal and Huete, 1991) was used in areas of minimal vegetation cover due to its relationship with the redness of soils, as this typified variations in soil conditions (see Section 2.3.2.4.4). Figure 4.1 is a flow chart that provides a graphical representation of the soil sampling design which was performed using commands available in MFWorks and GeoMedia GIS software. The three classes of variability generated from the NDVI and RI were combined with the four slope classes present on Muresk farm to define 12 sampling zones. The number of sample points selected from each zone was based on the area of each zone and its spatial variability. Appendix B provides a step by step description of the commands used while an overview is explained hereafter.

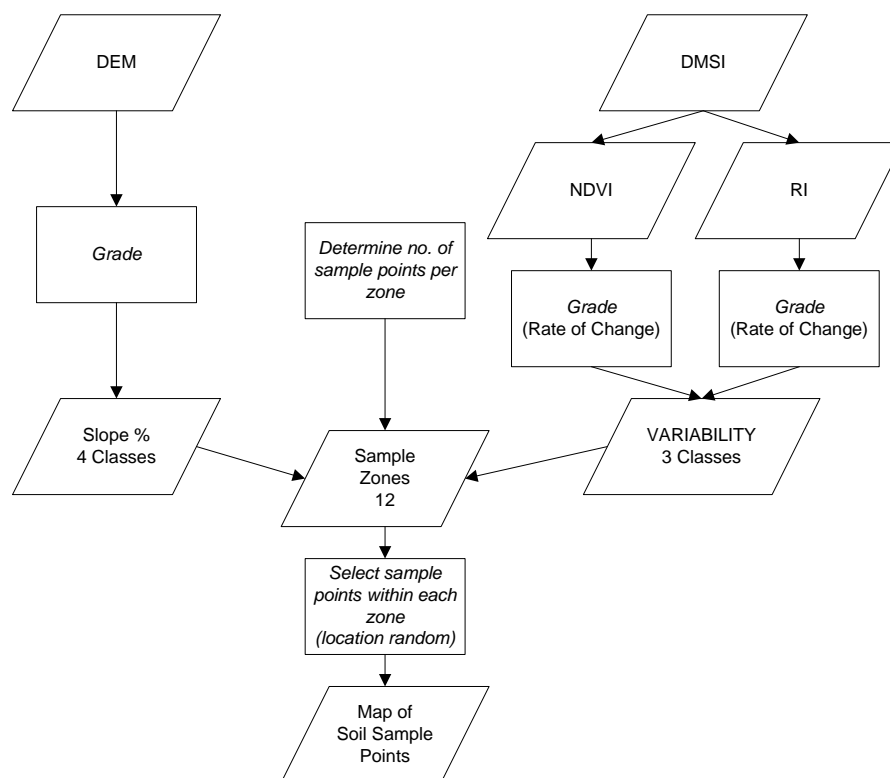


Figure 4.1 Flow chart of sampling design

#### 4.1.1 NDVI and RI Layers

The NDVI and RI layers were generated from the DMSI imagery (flown 28<sup>th</sup> June 2002) for the entire farm following Rouse *et al.* (1973) Equation (2.4) and Escadafal and Huete (1991) Equation (2.9). Rates of change in NDVI and RI layers were computed. This step detected areas of no change (flat) and areas where rapid changes occur over a short distance (steep). These layers are hypothesised to be surrogates for the degree of changes in soil conditions.

#### 4.1.2 Mask Layer

A mask layer of the study area was required to delimit the soil sampling strategy to arable regions of the farm. This was formed by *heads-up digitizing* on a true colour image of the study area. Rocky outcrops, buildings, dams, roads and remnant vegetation were masked and individual polygons created for areas that have similar visual appearance, i.e. bare soil, cropped paddocks, gullies and drainage lines.

The farm had varying levels of vegetation cover (from bare soil to well established crops or fallow) during image capture (28<sup>th</sup> June 2002). As such, it was required to determine which index (NDVI or RI) would be the most appropriate surrogate for determining the degree of variability in soil conditions. If the land presents a moderate to dense vegetation cover, then the NDVI would be the most appropriate option, whereas areas of bare soil would be best represented with the RI. This was achieved using the NDVI values for each region. For each unique region the area (number of cells) with a  $NDVI \leq 0$  and the area (number of cells) with a  $NDVI > 0.05$  was determined. Thus, if the majority of NDVI values for an area are equal or less than 0, the RI values will be used for that particular region. Conversely, if the area presents a majority of NDVI values greater than 0.05, then the NDVI was the information layer adopted for determining variability.

#### 4.1.3 The Variability Map Layer

The variability map layer is required to provide a spatial representation of the variability across the farm based on the change in NDVI and RI. This map layer could subsequently be divided into categories of variability and the soil sampling points placed accordingly.

A histogram of the rate of change, as expressed by the percentage of slope for NDVI and RI values in their associated regions, is displayed in Figure 4.2. This provides an indication of the area (as indicated by the number of cells) in each variability category.

The NDVI and RI output images were grouped into three classes of variability. The class boundaries selected for representing low, medium and high variability in each index are displayed in Table 4.2 along with the area (as indicated by the number of cells) of each class. As these categories represent rate of change it is reasonable to use the same boundaries for each index.

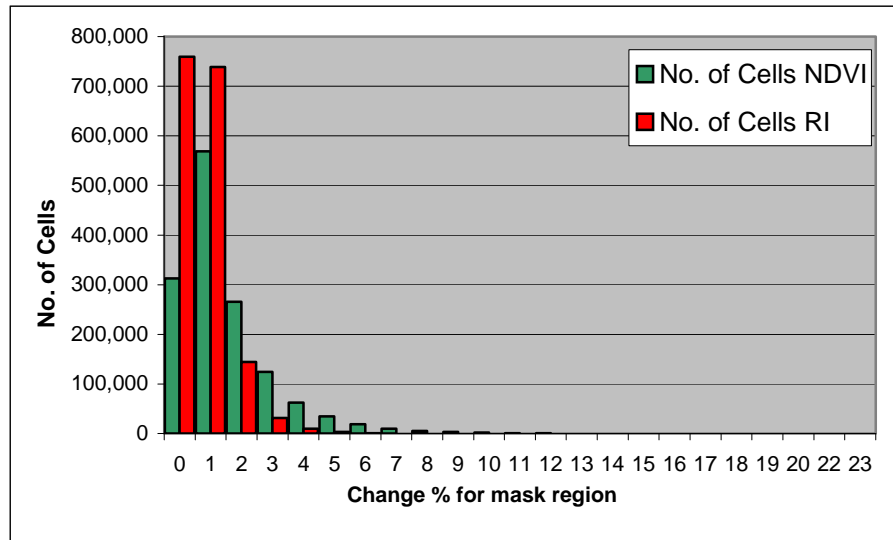


Figure 4.2 Histogram of NDVI and RI change % values for mask region

Table 4.2 Change % range for variability class of NDVI and RI

<b>NDVI</b>			
<b>Class</b>	<b>Change % Range</b>	<b>No. of Cells</b>	<b>NDVI Variability Class</b>
1	0 to <1	312946	Low
2	1 to <2	568287	Medium
3	2 to <24	527796	High
<b>RI</b>			
<b>Class</b>	<b>Change % Range</b>	<b>No. of Cells</b>	<b>RI Variability Class</b>
1	0 to <1	759485	Low
2	1 to <2	738547	Medium
3	2 to <13	189567	High

The final variability map (Figure 4.3) was required to provide one layer that expressed the 3 classes of variability for the entire study area. This layer was then combined with the slope classes as described in the following section in order to produce one map layer consisting of all combinations of variability regions and slope strata.



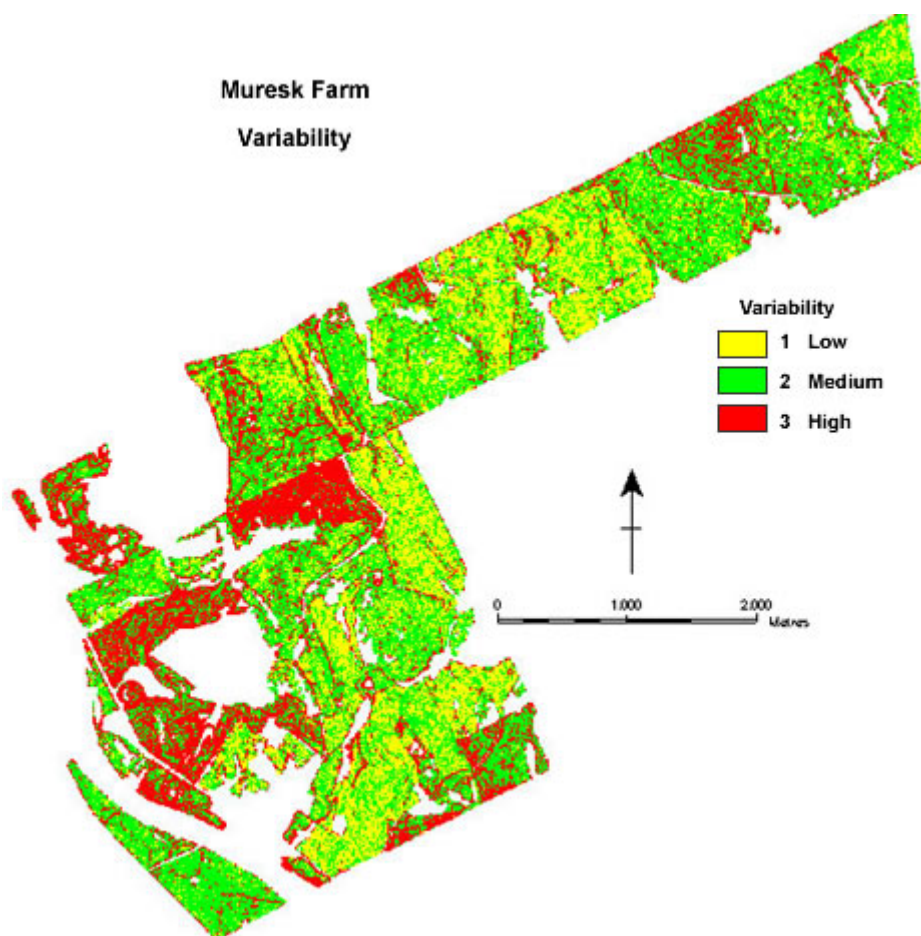


Figure 4.3 Variability map generated with a 10m spatial resolution

#### 4.1.4 Slope Class Map

The slope class map is an integral part of the sampling strategy because it provides the spatial slope strata to which an equal sampling density of soil points will be allocated. The slope layer was generated from the DEM for the entire farm. The output slope map is expressed in percentages (Figure 4.4) and demonstrates that the majority of the arable area of the farm is gently inclined (3 to 10% slope). A histogram of the slope values within the study area was generated (Figure 4.5). The slope map was then *re-classed* into 4 classes according to the Australian Soil and Land Survey classification standards (Table 4.1) and the output map layer was generated providing the spatial regions in which the soil sampling points could be targeted.

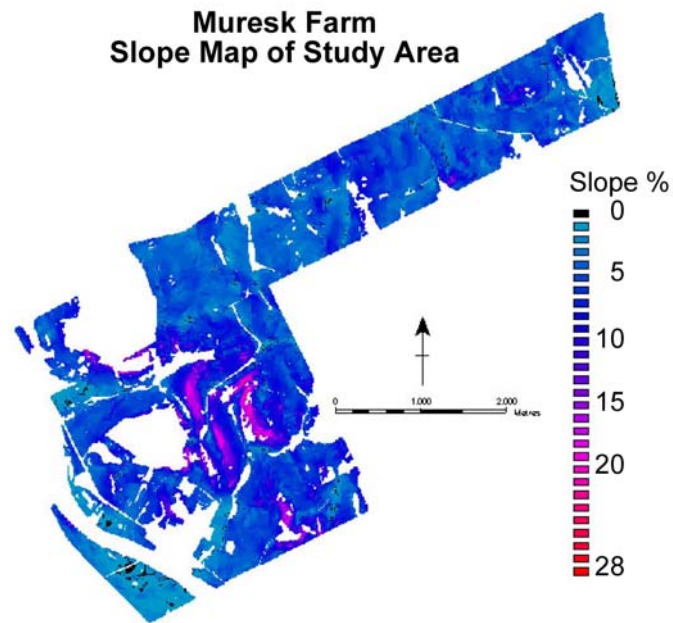


Figure 4.4 Slope map of study area

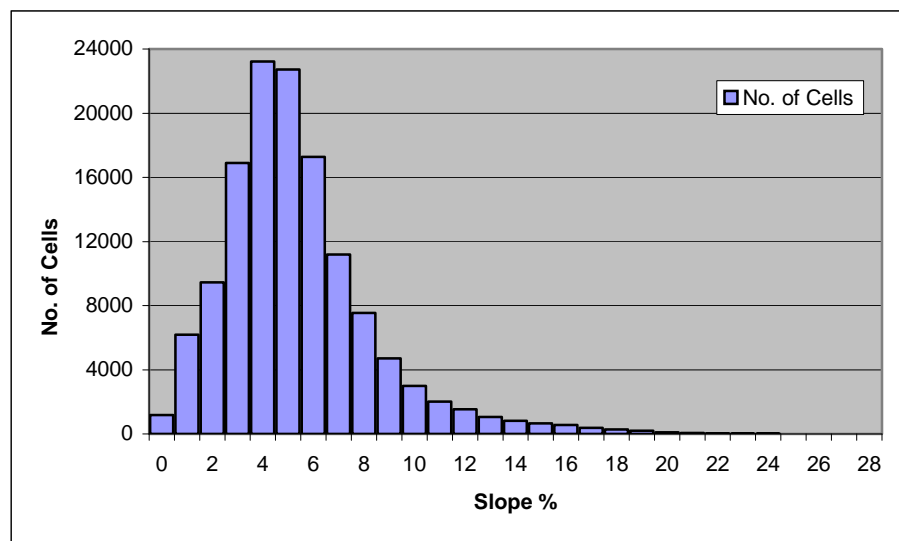


Figure 4.5 Histogram of slope values for the area of interest

#### 4.1.5 Variability Table

The sampling strategy was designed with the intention of sampling more intensively in areas of high variability (as represented by the rate of change of the NDVI and RI layers). In addition the density of sample points would be distributed equally within each slope strata, with respect to the area that each of these strata comprised within the total mask region. Since it is hypothesised that areas of the same slope class

would contain similar types of soils, this would ensure that the representation of soil types within the sample is proportional to their area.

Sampling theory indicates the sample size for each stratum in a sample should be proportional to the product of the size of the stratum and the standard deviation of the stratum (Cochran, 1959 pg. 74). A variability proportion was computed to enable the appropriate number of sampling points to be allocated to each variability class. It takes into consideration the percentage of area of each variability class and a weight which is intended to reflect the standard deviation of the variability class.

The weight given to each of the three variability classes has a strong influence on the number of points assigned to each class and consequently their spatial position in the study area. The pairwise comparison method developed by Saaty in 1980 (Malczewski, 1999) was implemented to achieve this. The method involves ratios of the standard deviation in each class to create a pairwise comparison matrix ( $R_{ij}$ ). It takes the pairwise comparison as an input and outputs weights,  $w_1$ ,  $w_2$  and  $w_3$  (Malczewski, 1999) such that  $\sum w_i = 1$  and  $(w_j/w_i)$  is a good approximation of  $(R_{ij})$ . The three major steps in the procedure, (i) generation of the pairwise comparison matrix, (ii) computation of the criterion weights, and (iii) the estimation of the consistency ratio, were performed following Malczewski (1999). The pairwise comparison matrix is developed based on an assigned value to rate the relative standard deviation between two classes of variability. For example, one class may be considered 10 times more variable which approximates to 3 standard deviations. Likewise for a class considered to be 50 times more variable than another, 7 would be the assigned ratio.

The pairwise comparison approach to determine appropriate weights for each variability class was implemented using IDRISI GIS software by developing a map layer for each class and allocating a value for each layer relative to other layers based on the relative standard deviation. Table 4.3 displays the input value within the ratio matrix and the weight generated by the pairwise comparison approach within IDRISI for each class.

The consistency index (CI) can be calculated by comparing the pairwise comparison matrix to the new ratio matrix based on the computed weights. The consistency ratio (CR), the ratio of CI to a random inconsistency index (tabulated values for a range of class numbers 1-15, (Malczewski, 1999)), can be used to determine if pairwise comparisons are consistent with the weights. If the  $CR \geq 0.1$  the values of the ratio are indicative of a ratio matrix which is not consistent, and a revision of the original values is required until a  $CR < 0.1$  is achieved (Eastman, 1999; Malczewski, 1999). The initial selection of pairwise comparisons (Table 4.3) gave a CR of 0.1 and thus required re-evaluation. The CR for the final values assigned was 0.01 (Table 4.3).

Table 4.3 Initial and final selection of Pairwise Ratio Matrix values and Weights

Initial Values					
	Low	Medium	High	CR	
Low	1			0.10	
Medium	5	1			
High	9	5	1		
Final Values					
	Low	Medium	High	CR	Weight
Low	1			0.01	0.09
Medium	3	1			0.24
High	7	3	1		0.67

The variability proportion was then calculated, based on the weight for each variability class (Table 4.3) and the percentage of the area that each class comprised within the mask region. By multiplying the percentage of the area of each class by its weight, a sample ratio was computed. The resulting variability proportion was then the percentage of the sample ratio for each class in relation to the total sample ratio (Table 4.4).

Table 4.4 Variability proportion determination

Class	Variability	Percentage of mask region	Weight	Sample Ratio	Variability Proportion
1	Low	26.47	0.09	2.38	0.08
2	Medium	47.54	0.24	11.41	0.36
3	High	25.99	0.67	17.41	0.56
Total		100.00	1.00	31.20	1.00

#### 4.1.6 Slope Table

The Slope Table was calculated to identify the correct proportion of the 250 sample points to be allocated to each slope class, based on the area of each slope class within

the mask region. An equal sample density of points were placed within each slope strata, and Table 4.5 displays the number of sample points calculated for each slope class based on the classes' areas.

Table 4.5 Number of sample points per slope class

Slope Class	Percentage of mask region	No. of Points
1	0.89	2
2	11.93	30
3	79.00	198
4	8.17	20
Total	100.00	250

#### 4.1.7 Sample Matrix

The sample matrix (Table 4.6) was generated to determine the number of sample points to be randomly placed within each sample zone (slope strata by variability class). The number of sample points to be placed in each sample zone is determined by multiplying the variability proportion from Table 4.4 by the number of sampling points allocated to each slope strata.

Table 4.6 Sample Matrix of slope and variability classes

		Variability 1 (low)	Variability 2 (medium)	Variability 3 (high)
		0.08	0.36	0.56
Slope 1	2	0	1	1
Slope 2	30	2	11	17
Slope 3	198	16	71	111
Slope 4	20	2	7	11
Total	250	20	90	140

#### 4.1.8 Sample Zones and Sampling points

The 12 sampling zones were formed to generate the spatial regions on the farm within which sampling points would be selected according to the sampling matrix (Table 4.6). The sample zones map layer was created using a cross tabulation between the slope classes (Slope% Class 1-4) and variability classes (Variability 1-3); thus generating an output map layer with 12 unique zones derived from all possible combinations of the slope and variability classes.

Sample points were randomly selected from each sample zone. The raster output map layer was exported as a text file containing the spatial location (easting and northing) of the 250 sample points (XYZ file). The sample points were then overlaid with the true colour composite of Muresk farm (Figure 4.6). This provided a field map for use during fieldwork to aid the approximate spatial location of each point. The list of coordinates associated with each point in the text file was used to accurately locate the points with a hand held differential global positioning system (dGPS).

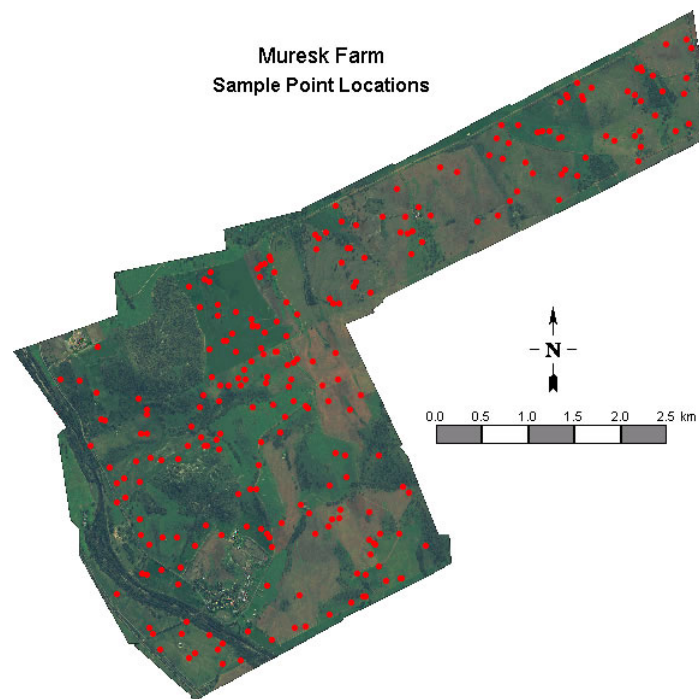


Figure 4.6 Spatial distribution of the soil sample points used for survey of the Muresk farm

## 4.2 Field Collection

The collection of soil samples and field description was conducted between 22<sup>nd</sup> July 2002 and 15<sup>th</sup> August 2002. Point locations were spatially located in the field using a hand held dGPS receiver. The positions generated during the sampling design were used as waypoints. Point locations were recorded with the dGPS. A site description was made at each of the 250 soil samples along with a record of the soil field texture and colour at 10cm intervals to a maximum depth of 80cm (where achievable). The field data collection process is shown in Figure 4.7, it includes; (a) auguring the soil pit, (b) digging the soil pit, (c) recording the dGPS point and (d) an example of a soil pit profile. A list of attributes collected and classes and units recorded is shown in Table 4.7. Soil samples were collected at the topsoil and 30cm depth for laboratory analysis. The topsoil samples were bulk sampled within a 6m diameter from the sampling point following Justice and Townshend (1981) suggestions for the minimum size of unit that should be sampled (Section 2.3.1.1.2 Equation (2.3)) based on a 2m pixel dimension and dGPS advertised accuracy of 5m, while the subsurface was sampled only from that particular location.



Figure 4.7 Soil sampling Field Data Collection. a) Auguring soil pit to maximum depth of 80cm, b) Digging soil pit, c) Recording the dGPS position of soil sampling point, d) Soil sample point 213, soil pit

Table 4.7 Soil sampling field data collection

Attribute	Measuring technique/scale
Point Number	1 – 250
Easting and Northing	dGPS – AGD84 UTMz50s
Date Sampled	22/07/02
Slope	Visual estimation measured in percentage of slope
Aspect	Visual estimation measured in degrees with compass
Drainage Class	E: Excessively drained S: Somewhat excessively drained W: Well drained M: Moderately well drained I: Somewhat poorly (imperfectly) drained P: Poorly drained V: Very poorly drained (Appendix C contains a full description)
Vegetation Cover	Visual estimation in percentage
Vegetation Type and remarks	i.e. crop or pasture, height of crop, land use
Comments	Any further remarks relevant to site location
A-Horizon	Depth in centimetres
Texture*	Field texture estimation. (Appendix D contains a full description)
Soil Colour*	Munsell Colour Chart Code: (Munsell Color Company, 1998 revised edition)
Comments*	Further remarks relevant to identification of changes in soil properties (i.e. rock fragments, compaction, nodules, mottles etc.)

\* recorded at 10cm depth intervals  $\leq$  80cm

### 4.3 Chemical Analysis of Soil Properties

The soil samples collected from the field were transported to the Chemistry Centre (WA) for laboratory analysis. The samples were oven dried and sieved through a 2mm sieve in preparation for analysis. Examples of the soil chemical analysis process are shown in Figure 4.8 namely; (a) a sieved soil sample; (b) weighing soil samples for analysis; (c) particle sizing analysis; (d) organic carbon analysis and (e) soil samples prepared for analysis.



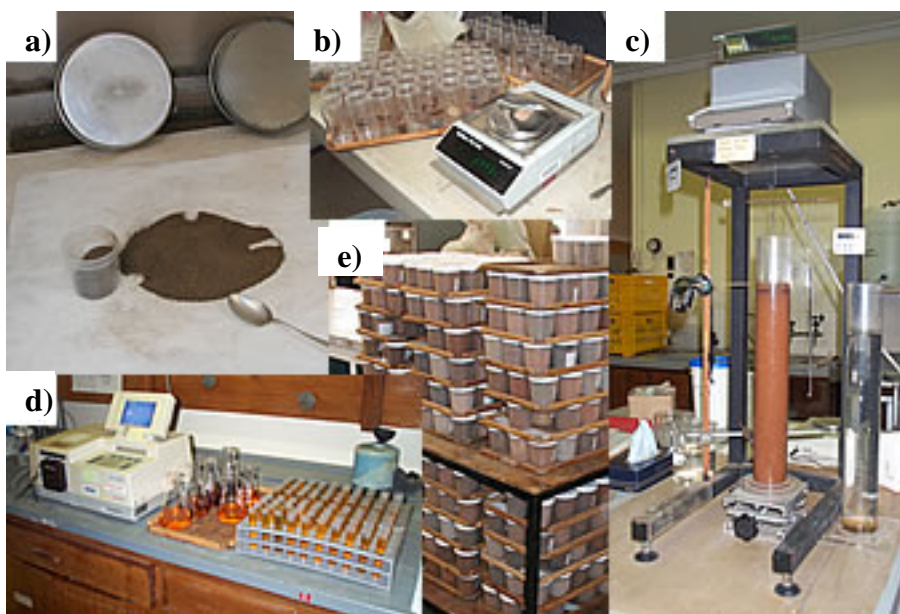


Figure 4.8 Soil chemical analysis; a) sieved soil sample; b) weighing soil samples for analysis; c) particle sizing analysis; d) organic carbon analysis; e) soil samples prepared for analysis

The analysis and procedures followed were performed according to the methodology advocated by Allen (2004). These procedures are detailed in Table 4.8.

Table 4.8 Methods of soil chemical analysis

Analyses	Measurement Method
Stones (%)	Weight of fractions > 2mm
pH(H <sub>2</sub> O)	pH was determined using a glass electrode on a 1:5 extract of soil and deionised water, following equilibration by shaking (end-over-end) for 1 hour. Values corrected to 25°C. Method 4A1 (Rayment and Higginson, 1992).
EC (1:5) (mS/m)	Electrical conductivity was measured by a conductivity meter in the 1:5 extract of soil and deionised water used for measurement of pH (H <sub>2</sub> O). Values corrected to 25°C. Method 3A1 (Rayment and Higginson, 1992).
pH (CaCl <sub>2</sub> )	pH was determined using a glass electrode on a 1:5 extract of soil and 0.01 M CaCl <sub>2</sub> . Method 4B1 (Rayment and Higginson, 1992).
Sand (%)	Particle Sizing was determined using 50g samples of soil (not pre-treated to remove organic matter or CaCO <sub>3</sub> ) dispersed with a solution of "Calgon" (water containing 1g sodium hexametaphosphate and 0.2g NaOH) by overnight rolling at 10rpm. The slurry was then transferred to sedimentation cylinders and made up to 1250mL. Silt (0.002 – 0.020mm) and clay (<0.002mm) concentrations were calculated using the density of the suspension measured by a plummet suspended below an electronic balance at a depth and time determined by Stoke's Law. Method adapted from Loveday (1974).
Silt (%)	
Clay (%)	
OrgC (W/B) %	Organic Carbon was determined on soil ground to less than 0.15mm using Metson's colorimetric method (Metson, 1956) a modification of the Walkley and Black method (Walkley, 1947). The procedure is based on oxidation of soil organic matter by dichromate in the presence of sulphuric acid. The heat for the reaction is supplied by dilution of the sulphuric acid with the aqueous dichromate.
Fe (AmOx) mg/kg	Amorphous forms of hydrous iron and aluminium oxides were extracted from soil using acidic ammonium oxalate solution (pH 3.0). Concentrations of iron and aluminium in the extract solutions were measured by inductively coupled plasma - atomic emission spectrometry (ICP-AES). Method 13A1 (Rayment and Higginson, 1992).
Al (AmOx) mg/kg	
Ca (exch) me%	Exchangeable Cations were determined by extraction using one of the following procedures: a) Neutral Soils: With 1M ammonium chloride (NH <sub>4</sub> Cl) pH 7.0 – used for neutral soils (pH (H <sub>2</sub> O) between 6.5 and 8). The concentrations of cations (Ca, Mg, Na and K) were measured by inductively coupled plasma – atomic emission spectrometry (ICP-AES). Soluble salts were removed from the soils with EC(1:5)>20mS/m by washing with glycol-ethanol. Method 15A1, 15A2 (Rayment and Higginson, 1992).  b) Acidic Soils: With unbuffered 0.1M Barium Chloride (BaCl <sub>2</sub> ) pH <6.5 - used for acidic soils only. The concentrations of cations (Ca, Mg, Na, K, Al and Mn) were measured by inductively coupled plasma – atomic emission spectrometry (ICP-AES). Soluble salts were removed from the soils with EC(1:5)>20mS/m by washing with glycol-ethanol (Tucker, 1985).  c) Calcareous Soils: With 1M ammonium chloride (NH <sub>4</sub> Cl) pH 8.5 – used for calcareous soils. The concentrations of cations (Ca, Mg, Na and K) were measured by flame atomic absorption spectrophotometry (ASS). Modification of Method 15C1 (Rayment and Higginson, 1992).
Mg (exch) me%	
Na (exch) me%	
K (exch) me%	
Al (exch) me%	
Mn (exch) me%	

#### 4.4 Building the Soil Database

The dGPS soil coordinates along with the field data and chemical analysis results were entered into spreadsheets and subsequently transferred to GIS software in order to build a spatial soil database of the study area. Recorded information collected during the field sampling was available for the surface and at 10cm intervals to a maximum depth of 80cm, while chemical properties were associated with depths of 10 and 30cm.

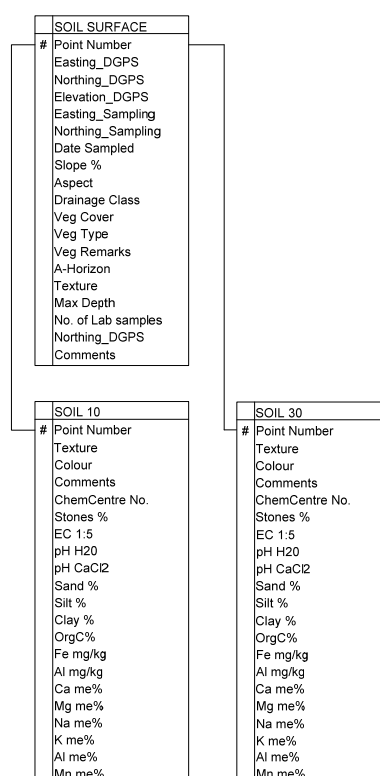


Figure 4.9 Soil database schema

The statistical analysis between soil properties, yield, remote sensing data and terrain attributes (described in the following chapters) will be based on soil information, largely available from the 10cm and 30cm depths, and as such has been shown in Figure 4.9 while Appendix E shows the full soil database schema.

#### 4.5 Vegetation Sampling

Transect sampling was conducted through three crop types namely wheat (*Triticum aestivum*), lupins (*Lupinus angustifolius*), and canola (*Brassica napus*) for two consecutive years, 2002 and 2003 (i.e. one transect per crop type (paddock) per year). Crop attributes recorded from the three crop types are used to explore the

relationships that exist with DMSI bands and vegetation indices. Table 4.9 provides a detailed description of each transect's attributes.

Table 4.9 Transect Attributes

Transect Label	Growing Season	Crop Type	Seeding Date	Seeding Rate (kg/ha)	Field Sampling Date	Laboratory analysis Date	DMSI flight date	Crop age at time of sampling (weeks / days)
Wh02	2002	Wheat	18/05/02	80	25/09/02	25/09/02	17/09/02	18w / 4d
Lu02	2002	Lupins	01/06/02	80	23/09/02	24-25/09/02		16w / 2d
Ca02	2002	Canola	07/05/02	4	18-20/09/02	19-22/09/02		19w / 2d
Wh03	2003	Wheat	21/05/03	70	3/09/03	4/09/03	3/09/03	15w
Lu03	2003	Lupins	23/04/03	100	27/08/03	27-28/08/03		18w
Ca03	2003	Canola	29/04/03	3	25/08/03	26-27/08/03		16w / 6d

Transects were placed through variable crop growth conditions. This was crucial to the field sampling approach in order to collect variable crop condition and to examine if DMSI could depict this variability. Transect locations were determined by a reconnaissance of the farm paddocks in consultation with the farm manager to determine areas of variability. Once these regions were recognised, transects of 310m in length were constructed through selected paddocks. Crop attributes were recorded at 10m intervals within a 1m<sup>2</sup> quadrant equating to 30 sample sites for each transect. Thirty sample sites simplify statistical calculations based on the assumptions of normality. The spatial location of each transect was recorded with a dGPS.

Field data were collected synchronous with the capture of DMSI. At each sample location crop density, height, LAI and weed infestation was recorded. The crop density was determined using a combination of the row crop and randomly distributed plants methods (TopCrop Australia, 1999a) (Table 4.10). Cihlar *et al.* (1987) provide detailed descriptions for collecting crop attributes to be used in remote sensing studies and as such have been adopted here. Crop height was calculated from the mean of five randomly selected plants. At each sample location, five plants were randomly removed and taken to the laboratory to measure the leaf area. All green leaves were stripped from the stems and weighed. For large leaf samples (larger than 30 grams) a sub sample was taken. The leaf area was then determined for the sub sample, and subsequently calculated for the five plants by multiplying by the sub sample proportion. In the year 2002, the average leaf area of five plants within the 1m<sup>2</sup> quadrant was then determined using the LI-COR LI-3000A portable area meter coupled with the LI-3050A transparent belt conveyer

accessory. In the year 2003 it was calculated using the LI-COR LI-3100 area meter (Figure 4.10 (d)). This was then converted to a LAI utilising the density calculated at each corresponding field location. For the canola transects, the flower petal area was also determined for each sample and a petal area index (PAI) calculated in a similar manner to LAI.

One factor influencing seed yield can be the number of flowers present in a crop (Yates and Steven, 1987). The canola plants were flowering and bright yellow when sampled in both years. The PAI was determined to evaluate the impact of flowers on the digital numbers (DN) of DMSI bands. At the time of field data collection, it was unknown what relationship the area of petals (which would be related to the number of flowers) would have on the final yield of canola. Figure 4.10 displays a mosaic of photos captured during field and laboratory data collection.

Table 4.10 Methodology to determine crop density

Crop Cover	Visual density	Count Method	Density $m^{-2}$
Row Crops	N/A	$P_m$ = No. of stems per 50cm row x 2 rows.	$\frac{P_m \times 100}{\text{Seeder row spacing (cm)}}$
Randomly distributed plants	High	$P_{25}$ = No. of stems in 25cm <sup>2</sup>	$P_{25} \times 16$
	Medium	$P_{50}$ = No. of stems in 50cm <sup>2</sup>	$P_{50} \times 4$
	Low	$P_1$ = No. of stems in 1m <sup>2</sup>	$P_1$

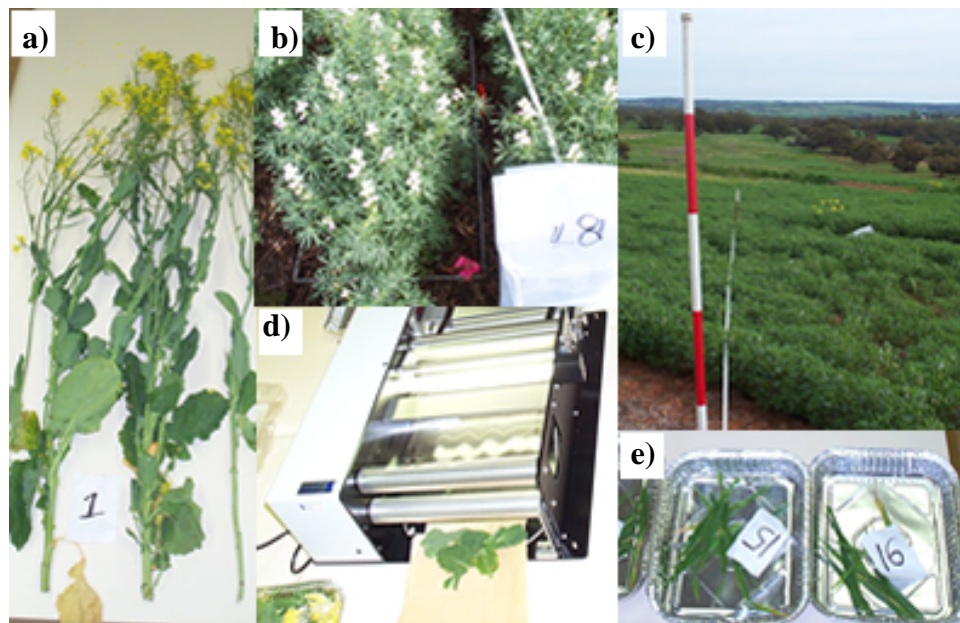


Figure 4.10 Field data collection and laboratory analysis a) canola sample, b) lupin field quadrant, c) lupin transect, d) leaf area machine, e) wheat leaf sample

The results from the field and laboratory data collection are summarised in Table 4.11 to highlight the variability in crop attributes which are used to assist in explaining the relationships with DMSI discussed in Section 4.7. There are fewer than 30 samples for some of the transects because of contour banks, rocky outcrops, water logging or mis-seeding. The software SPSS (SPSS Inc., 2003) was used for the statistical analysis. The data sets were tested for normality in order to determine if parametric or non-parametric approaches would be required in the correlation analysis. When data sets are less than 50 samples the Shapiro-Wilk statistic (Shapiro and Wilk, 1965) is calculated to test for normality and when significance is greater than 0.05, normality is assumed (Coakes and Steed, 2001). Table 4.11 displays the results from the Shapiro-Wilk statistic and whether parametric or non-parametric statistical approaches will be applied in the following Section.

Table 4.11 Basic descriptive statistics of crop attributes recorded on the transects

Crop Attribute	Statistic	Wh02	Lu02	Ca02	Wh03	Lu03	Ca03	Ca03 PAI sites (n=20)
Height (cm)	Min	63	29	108	36	33	56	
	Max	92	73	170	57	83	137	
	Mean	81	53	149	45	64	97	
	Std. Dev.	6.78	10.67	15.02	5.96	15.08	23.55	
	CV%	8.37	20.13	10.08	13.24	23.56	24.27	
Density + (m <sup>-2</sup> )	Min	8	11	4	44	3	4	
	Max	120	72	88	359	84	148	
	Mean	64.7	42.7	26.41	187.3	46.06	52.1	
	Std. Dev.	32.89	17.81	23.23	76.19	20.79	28.65	
	CV%	50.83	41.71	87.96	40.68	45.14	54.99	
Visual % of green vegetation ground cover	Min	5	3	Not Recorded	3	10	Not Recorded	
	Max	75	65		40	100		
	Mean	41.07	23.79		19.59	89.72		
	Std. Dev.	18.28	15.50		11.51	24.40		
	CV%	44.51	65.15		58.75	27.19		
	Shapiro-Wilk Sig.	0.323	0.018		0.045	0.000		
	Distribution	N	NP		NP	NP		
LAI	Min	0.02	0.04	0.28	0.15	0.38	0.08	0.08
	Max	1.34	3.59	5.40	1.88	6.22	7.23	7.23
	Mean	0.39	1.25	1.25	0.84	3.07	1.53	1.98
	Std. Dev.	0.30	0.74	1.03	0.43	1.45	1.83	2.06
	CV%	76.92	59.2	82.4	51.19	47.23	119.6	104.04
	Shapiro-Wilk Sig.	0.027	0.104	0.000	0.171	0.995	0.000	0.002
	Distribution*	NP	N	NP	N	N	NP	NP
PAI	Min	Not Recorded	Not Recorded	0	Not Recorded	Not Recorded	0	
	Max			0.27			0.06	
	Mean			0.082			0.021	
	Std. Dev.			0.074			0.015	
	CV%			90.24			71.42	
	Shapiro-Wilk Sig.			0.008			0.011	
	Distribution*			NP			NP	
		Wh02	Lu02	Ca02	Wh03	Lu03	Ca03	PAI sites

+ The magnitude of the wheat densities should not be compared between 2002 and 2003 due to a different counting technique, however this does not affect the LAI.

\* N:normal, NP:non-parametric

CV%: Coefficient of Variation

#### 4.6 Generation of Vegetation Indices

Using the radiometrically corrected DMSI imagery flown in September 2002 and 2003, five vegetation indices (VI), namely the normalised difference vegetation index (NDVI), normalised difference vegetation index green (NDVIgreen), soil adjusted vegetation index (SAVI), plant pigment ratio (PPR) and photosynthetic vigour ratio (PVR) were generated. New images for each index were created by transforming the pixels according to the Equations (2.4) to (2.8) described in Chapter 2.

The PPR and PVR were proposed by SpecTerra Services (2003b) in ratio form and as such named in that way, but have been manipulated into an index in this research for uniformity in the analysis and comparison of vegetation indices. This concurs with Metternicht (2003), who also utilised the PPR and PVR in index form.

The  $L$  factor chosen for SAVI (Equation (2.6)) depends on the density of the vegetation cover being analysed. The SAVI has been calculated using three  $L$  values, 1 for areas of very low vegetation density and 0.5 and 0.25 for intermediate and high vegetation density respectively, following Huete (1988).

#### **4.7 Relationship between DMSI and Crop Attributes**

Digital numbers were extracted for each band in a 3 x 3 window (6m x 6m) around the sample points to accommodate spatial location errors that may occur. The mean digital number for the 9 cells (36m<sup>2</sup>) at each point within each band and vegetation indices were used for the statistical analysis. When correlating two variables, Spearman's rho correlation coefficient was applied if at least one of the variables was non-parametric while Pearson's correlation coefficient was used otherwise. Table 4.12 and Table 4.13 provide the correlation coefficients obtained from the analysis between all crop LAI, canola PAI with the DMSI bands and vegetation indices. These will be used to examine the relationships between DMSI individual bands, VI and crop attributes. The SAVI and NDVI correlation coefficients will be equal as the adjustments made for the vegetation cover ( $L$  factor) is the same for all points along the transect, as such only NDVI results have been presented.



Table 4.12 Correlation coefficient for crop LAI vs Vegetation Indices

Transect Label ( <i>n</i> )	Vegetation Index or DMSI band	Correlation Coefficient LAI vs DMSI
Wh02 (28)	NDVI	0.130
	NDVIgreen	0.244
	PPR	-0.471**
Lu02 (29)	NDVI	0.502**
	NDVIgreen	0.524**
	PPR	-0.195
Ca02 (29)	NDVI	0.192
	NDVIgreen	0.198
	PPR	0.136
Wh03 (29)	NDVI	0.760**
	NDVIgreen	0.778**
	PPR	0.386*
Lu03 (18)	NDVI	0.279
	NDVIgreen	0.476*
	PPR	-0.644**
Ca03 (29)	NDVI	0.428*
	NDVIgreen	0.334*
	PPR	0.502**

(*n*) Number of samples. \* Correlation is significant at the 0.05 level.

\*\* Correlation is significant at the 0.01 level.

Table 4.13 Correlation coefficient for canola PAI and LAI vs DMSI bands and LAI vs Visual % Cover

Transect Label ( <i>n</i> )	DMSI Band	Correlation Coefficient LAI vs DMSI	Correlation Coefficient PAI vs DMSI	Correlation Coefficient LAI vs Visual% Cover
Ca02 (29)	B1	-0.334*	-0.306	
	B2	0.003	0.625**	
	B3	-0.086	0.641**	
	B4	0.320	0.399*	
Ca03 (20)	B1	-0.499*	-0.388*	
	B2	0.437*	0.260	
	B3	0.387*	0.172	
	B4	0.443*	0.222	
Wh02 (28)				0.110
Lu02 (29)				0.506**
Wh03 (20)				0.815**
Lu03 (20)				0.330

(*n*) Number of samples.

\* Correlation is significant at the 0.05 level.

\*\* Correlation is significant at the 0.01 level.

#### 4.8 Factors Affecting the Relationship between DMSI and Crop Attributes

Analyses of the variability within each transect will be explored in order to determine if each transect has been placed through variable conditions and the relationship between crop attributes and DMSI will be used to examine the potential of DMSI for detecting variability in crop growth attributes. There are many factors that play a role in crop germination and growth. The number of plants established depends on factors such as soil moisture, surface crusting, seedling vigour, sowing depth, fertiliser level, disease and insect attack (Martin and Gill, 1993). However,

evaluating these effects is beyond the scope of this research. Different crop species and/or varieties will also react differently to these factors, and so the crop types cannot be compared without some standardisation. The high variability of crop attributes provided in the summary table (Table 4.11) is taken as evidence that the sampling strategy depicted the variability that can be present in cropped paddocks due to factors outlined above.

#### 4.8.1 Vegetation Indices and LAI

An examination of the significant correlations between the vegetation indices and three crop types (Table 4.12) revealed little consistency over the years. The only significant correlations that were evident in the two years were between NDVIgreen and Lupins, and PPR and Wheat (which was negative in 2002 and positive in 2003). The number of days elapsed between flights and LAI measurements differ, and the crops were not at the same growth stage during both flights, which could have had an influence on the results. There is no evidence that having sampled within eight days of the DMSI flight had an effect on the correlation as there is no trend indicating that as the number of days between sampling and flight increased the correlations decreased. Therefore it appears that the growth stage of the crop is the dominating cause of the lower than expected correlation coefficients. Figure 4.11 and Figure 4.12 show the status of the crops when sampled, however variability occurred in each transect.

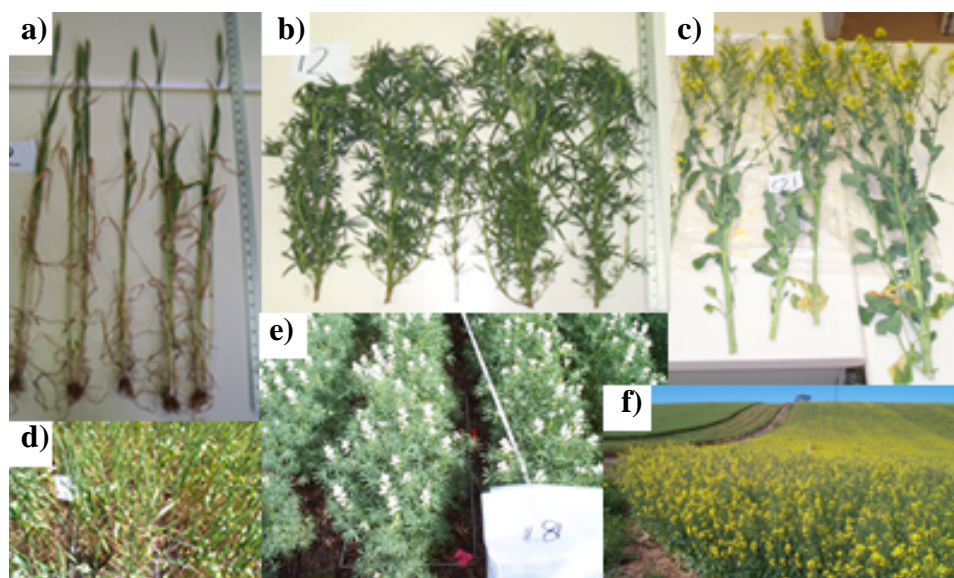


Figure 4.11 Status of crops in year 2002: a) Wheat sample, b) Lupin sample, c) Canola sample, d) Wheat in field, e) Lupins in field, f) Canola in field

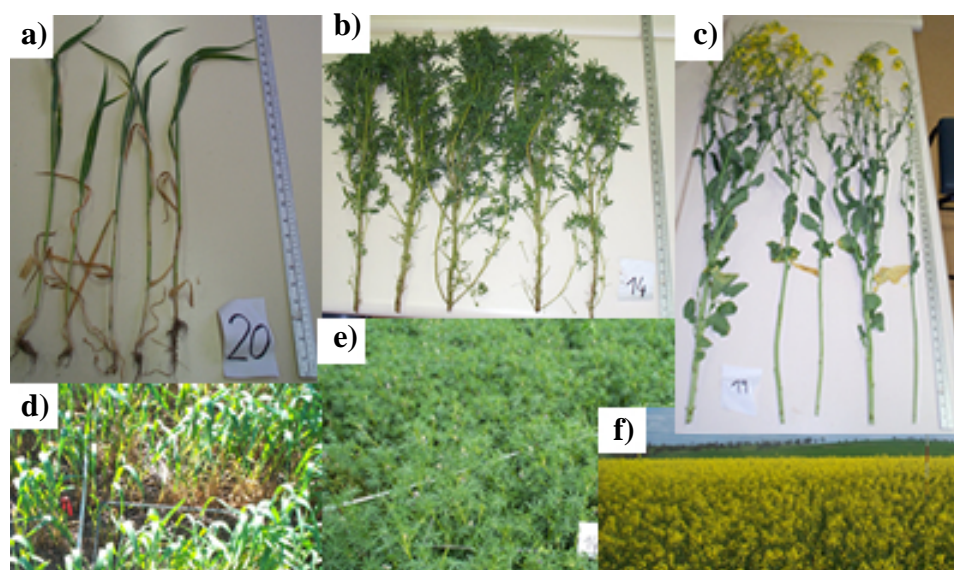


Figure 4.12 Status of crops in year 2003: a) Wheat sample, b) Lupin sample, c) Canola sample, d) Wheat in field, e) Lupins in field, f) Canola in field

Based on a number of investigations (Hatfield *et al.*, 1985; Cloutis *et al.*, 1999; Senay *et al.*, 2000; Xiao *et al.*, 2002; Warren and Metternicht, 2005) which produced strong relationships between the LAI of crops and spectral response (e.g.  $r > 0.70$ ) it was hypothesised that the strength of (as indicated by the correlation coefficient)

correlations between LAI and vegetation indices in this research would be greater. This has led to an examination into the cause of the weaker than expected results.

Bauer (1985) mentions, there is some concern as to whether a single multi-spectral measurement (i.e. one flight, as in this research) contains sufficient information to reliably estimate LAI of crop canopies. Previous studies by Sellers (1985) analysed the effects of saturation for large LAIs, and leaf angle for small LAIs. Sellers (1985) displayed the saturation effect of a large LAI for growing wheat canopy data with an asymptotic curve, which shows the vegetation index<sup>1</sup> responses diminishing for LAI values above 2.0. Sellers (1985) mentions that a near maximum absorption value of photosynthetically active radiation is achieved at a LAI the range of 2.0 to 3.0. He (1985) also points out that for extreme solar elevations (towards 90° above the horizon) and small values of LAI, leaf angle can cause a wide spread in the vegetation index response. For vertical leaves the effect is extreme because of a zero value for optical thickness relative to the direct beam of overhead sun and a maximum optical thickness at low solar angles. Horizontal leaves, show no change in vegetation index with solar angle, given that the optical thickness is always the same with the direction of incident radiation (Sellers, 1985). These factors might be contributing to the weaker than expected correlation in this research.

---

<sup>1</sup>  $(a_{cNIR} - a_{cVIS}) / (a_{cNIR} + a_{cVIS})$ ; where  $a_{cNIR}$  and  $a_{cVIS}$  are the hemispherically integrated albedo for the near-infrared radiation (0.7 -3.0 $\mu$ m) and visible radiation (0.4-0.7 $\mu$ m) respectively.

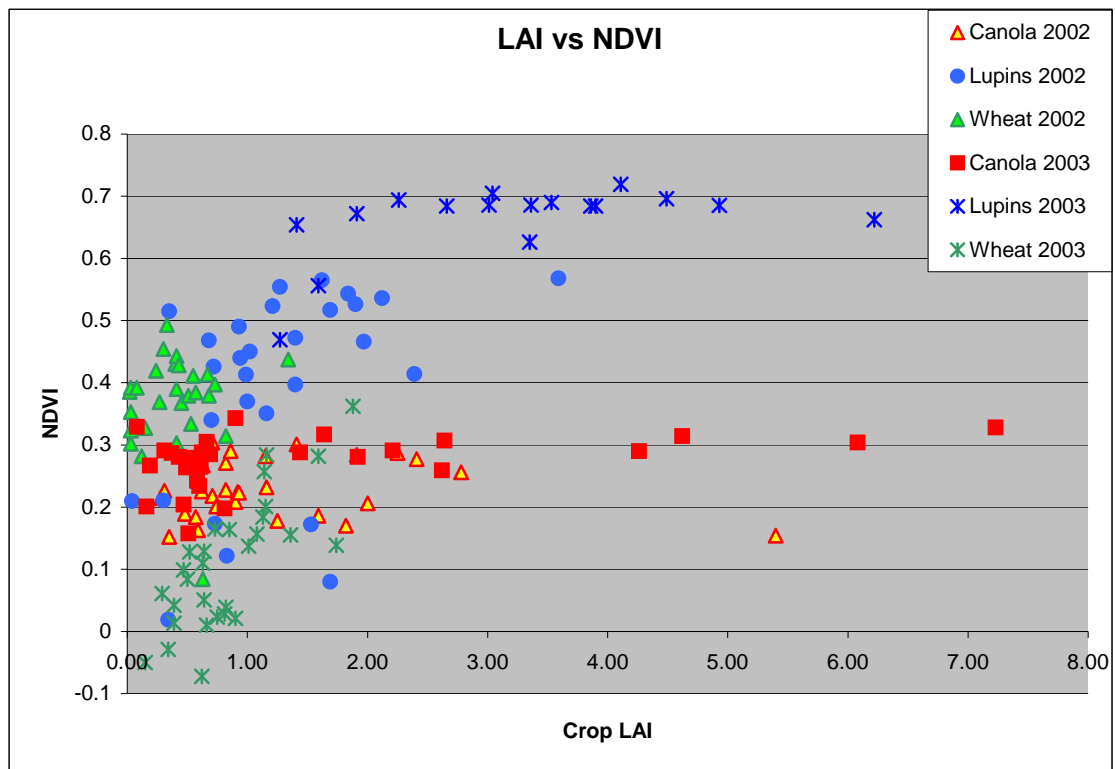


Figure 4.13 Relationship between NDVI and Crops' LAI

Figure 4.13 displays the NDVI and LAI for each transect. As the DMSI flights have not been spectrally calibrated to each other, the comparison between years should be based on the response to LAI values not the absolute values of NDVI. Figure 4.13 shows the diminishing differences in response of NDVI as LAI increases, especially between 2.0 and 3.0, and beyond. This concurs with Sellers' (1985) findings and can assist in explaining the correlation results for those crops with LAI values exceeding 3.0. This is particularly evident for canola (Ca02, Ca03) and lupins (Lu03) as the LAI of lupins collected in year 2003 recorded a mean value of 3.07 with values as high as 6.22 (Table 4.11). Aside from the saturation effect due to full canopy cover, the relationship between LAI and NDVI can also be affected by factors such as, leaf angle and crop flowering stages. These factors exert different effects on the crops analysed and, as a consequence, the data for each crop are analysed separately in the following sections.

#### 4.8.2 Wheat

There is a moderately linear relationship between the NDVI and LAI for Wh03 (0.76), though that is not the case for wheat corresponding to year 2002 (0.13) (Table 4.12) as shown in Figure 4.13. The majority of the points for Wh02 produce a

random scattering around a NDVI of 0.4 irrespective of the LAI, while there is a wider range of LAI values for 2003.

Figure 4.11 (a and d) and Figure 4.12 (a and d) display the wheat at the time of sampling and show the different wheat growth stages. The wheat in year 2002 (Wh02) is at flowering stage, whereas in 2003 (Wh03) it is early to full boot (TopCrop Australia, 1999b). Thus leaf orientation in Wh02 is vertical, while a portion of the leaves present a horizontal direction on the Wh03 transect. The solar angle position would be similar for both flights as the two were at similar times of the year and close to noon (when the solar zenith angle changes slowly with time). As such, the vertical effect of zero value optical thickness for overhead sun (Sellers, 1985) may be affecting the Wh02 results, producing weak correlations between LAI and the VI.

The weeds within the Wh02 transect might also be contributing to the weak correlations between NDVI and LAI for that year. NDVI accounts for all green vegetation present in the sample area (i.e. crop and weeds) whereas the LAI corresponds to the crop fraction only. The visual percentage of green vegetation ground cover (crop and weeds) was recorded in the field (Table 4.11) and shows that 2002 had a larger mean (41.1 percent) and maximum (75 percent) compared to Wh03, 19.6 percent and 40 percent respectively. Correlation was calculated between the visual percentage cover and LAI (Table 4.13). The results show that the visual percentage cover is strongly correlated to LAI for Wh03 (0.82) but not for Wh02 (0.11). This shows the larger density of vegetation cover recorded for Wh02 (crop and weeds) caused an increase in the NDVI which is not correlated to LAI. It could also be argued that the wider range of LAI values has led to the better relationship between LAI and NDVI for 2003 however, crop growth factors are also important.

#### 4.8.3 Lupins

There is a significant relationship between the NDVI and LAI for Lu02 ( $r = 0.50$ ) though this is not the case for lupins corresponding to year 2003 ( $r = 0.28$ ) (Table 4.12) as shown in Figure 4.13. The LAI for Lu03 scatters around an NDVI of 0.66 shows the diminishing differences in response of NDVI at LAIs beyond 2.0, as reported by Sellers (1985). The mean LAI for Lu02 and Lu03 was 1.25 and 3.07

respectively, while the mean visual percentage cover for Lu02 and Lu03 was approximately 24 percent and 90 percent, respectively. These values also show that the canopy saturation effect (Sellers, 1985) occurs for dense levels of vegetation cover evident during the year 2003 (Lu03). This is shown in Figure 4.12 (e) where 100 percent canopy cover is shown for 2003, while in Figure 4.11 (e) partial canopy cover and soil background is visible. The layers of leaves present in a lupin plant can be seen in Figure 4.11 (b) and Figure 4.12 (b), therefore at a certain value of LAI the leaves would have reached full canopy cover, while the actual LAI recorded would continue to increase.

At the time of sampling the lupins were 16 weeks and 18 weeks of age for Lu02 and Lu03 respectively, and for year 2002 (Lu02) the plants ranged from flowering to mid podding, while the lupins for year 2003 (Lu03) was in flower abortion to early podding stage (Nelson and Delane, 1991). There were more flowers in the lupin crop when surveyed during 2002 than 2003 (Figure 4.11 (e) and Figure 4.12 (e)). These differing growth stages of crops between years would in turn result in differing reflective responses in DMSI.

#### 4.8.4 Canola

Several aspects may be affecting the results of the canola correlations; namely the reflectance of the flowers, leaf position and wilting and canopy saturation. The results of the correlation between the PAI and DMSI bands (Table 4.13) show consistency over the two years in terms of band 1 (blue) having a negative correlation and bands 2 (green), 3 (red), 4 (NIR) showing positive correlations. As yellow results from the additive reflectance of green and red, and the absence of blue, these results are logical as the reflectance of the brilliant yellow colour of the canola petals is being depicted in the DMSI bands.

The significant correlations between the PAI and DMSI bands (Table 4.13) (e.g. 0.388 to 0.641) may provide a reason for the weak correlations between DMSI and LAI, as the petals are affecting the leaves' reflectance. For the 2002 canola transect, the PAI correlation coefficients were larger than that of the LAI for Bands 2, 3 and 4 indicating that the presence of flowers is affecting the results (Table 4.13). For the 2003 transect, the PAI correlation coefficients were lower than the LAI for all bands.

The mean PAI for 2003 was 0.021 while in 2002 it was 0.082. Therefore, it is thought that when the PAI reaches a certain threshold it hinders the reflectance of leaves.

Due to the nature of the canola plant this means the largest photosynthetically active area of the canopy (leaves) is shaded by the flowers and their supporting structure (Yates and Steven, 1987). Figure 4.14 displays the mean spectral curves for each crop transect in the year 2002 and 2003 and shows that canola has similar or lower digital numbers than other crops in the blue band, though higher digital numbers in the green and red bands. This also follows the principle of colour formation (e.g.  $G+R = Y$ ) indicating that the yellow flowers affect the reflectance captured by the DMSI.

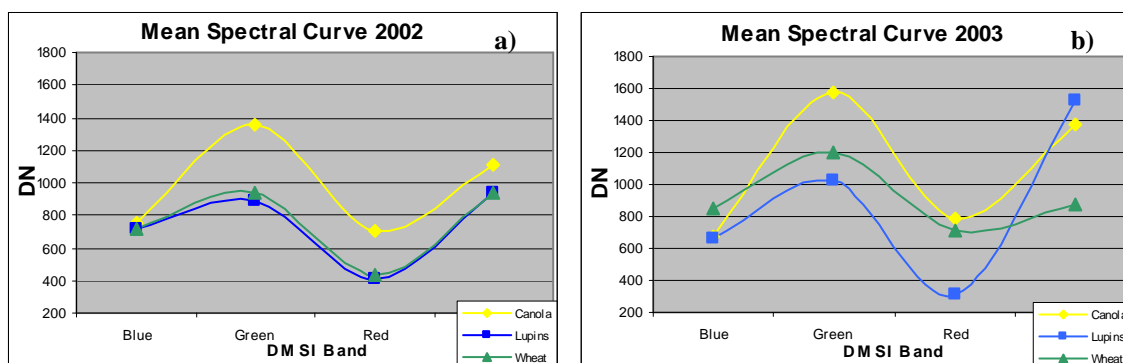


Figure 4.14 Spectral Curves for (a) 2002; (b) 2003

These results accord with those of Yates and Steven (1987) who found that the flowering canopy of canola had a reduction in absorption between the 500 and 700nm and a slight increase in absorptivity in the 400-500nm portion of the spectrum. Yates and Steven (1987) conclude that the spectral absorptivity of the flowering canopy does not give a true indication of the radiation absorbed by its leaves, because the flowers themselves absorb a substantial fraction of radiation they intercept.

Warren and Metternicht (2005) correlated the NDVI computed from DMSI and the LAI of canola, and found a significant correlation of  $r = 0.893$ . The reasons for such a strong correlation have depended on the stage of development of the canola plant. Their plants were 9.5 weeks old and the growth stage of the plant ranged from



seedling to vegetative (cabbage) whereas in this research the plants were 19 and 17 weeks of age for year 2002 and 2003 respectively, and at flowering. During the seeding to vegetative stage of crop growth the broad-leaved canola plant masks the soil considerably. While during flowering (post stem extension) the plant can grow up to 170cm tall and the leaves are distributed along the stem and thus as the LAI of canola is increasing, the percentage of projected foliage cover by the leaves may not be increasing at a similar rate.

#### **4.9 The Use of DMSI for Characterising Vegetation Variability**

The results indicate that DMSI portrays variability in crop growth and as such may be an appropriate remote sensing tool for further analysis inline with site-specific crop management applications. The analysis showed that the cause of the weaker correlation for 2002 and 2003 was that the LAI measurements for that year are not directly correlated to the percentage of ground cover due to leaf orientation (vertical/horizontal), leaf layering, leaf wilting, weed infestation and the impact of flowers. This suggests that a stronger correlation between DMSI and LAI would be during an earlier growth stage, prior to flowering and canopy saturation.

However, it is also important to consider at what stage of crop development attributes such as the LAI or other growth variables (i.e. biomass), should be recorded if the research involves predicting final yield. Future research needs to determine the balance between the spectral and phenological aspects of field crops in order to produce sensible models of crop yield where modelling parameters are extracted from remotely sensed data.

Further examination of the relationship between the reflectance of flowering canola and final yield would be of interest. The petal reflectance may prove to be a more suitable variable to include in crop growth models than the more commonly recorded, LAI. In this way, the capture of remotely sensed data during flowering would be more appropriate.

#### **4.10 Relationship between DMSI and Yield**

Halloran (2004) investigated whether there exists a relationship between DMSI and yield data from the Muresk farm. Based on her findings it appears that individual

bands and vegetation indices representing vegetation variability or both would not aid LMU classification.

Halloran (2004) reviews several studies investigating the statistical relationships between remotely sensed data and yield for inclusion in crop growth models, where yield predictions seem accurate. The review suggested that there would also be a relationship for Muresk with DMSI.

A detailed methodology and discussion of results shows the limiting factors is provided by Halloran (2004) and summarised here. Her results indicate that there is a weak to moderate relationship between DMSI and yield data. The potential factors affecting the results have been attributed to the following.

(1) The scale of the study. Corner *et al.* (1998) found that at a paddock scale reasonable correlations were achieved, whereas at a sub-paddock level, as in Halloran's (2004) study, there were weak relationships only.

(2) The drier than average conditions in 2002 hindered crop growth along with the weed and/or insect infestations and occurrences of frost, which were not closely monitored.

The strongest correlating crops with DMSI were Lupins and Canola which still produced correlation coefficients' below 0.4. Vegetation indices tended to correlate stronger than individual bands, although the two varied amongst crop types. Halloran (2004) found that imagery acquired closer to harvest produced the strongest results. This concurs with Pinter *et al.* (2003) who mentions in general the reliability of remotely sensed imagery to estimate yields decreases as the time before harvest increases, this is because there is more opportunity for factors such as drought, insect infestation and disease to impact yield.

#### **4. 11 Summary**

Chapter 4 describes the soil sampling design, which incorporates the use of slope and DMSI data to gather an understanding of the variability of landscape attributes present across the study area. The design is intended to sample more intensively in areas of heterogeneous soil properties and less often in homogeneous zones to optimise the approach. The field data collection and laboratory analysis of the soil properties are described, culminating in a database of 250 soil sampling points and

their associated properties. These will be used for statistical analysis and classification of LMUs which has been described in Chapter 7.

This chapter also provides discussion on factors that may affect relationships between DMSI and LAI for different crop types, and it is concluded that DMSI has the potential to monitor crop growth conditions. However, without appropriate field scouting, it is difficult to assume that changes in reflectance are purely associated with variability in crop growth. As such vegetation indices derived from the DMSI imagery will not be included in the LMU classification (Chapter 7) as a layer that depicts vegetation variability. However the RI will be analysed in Chapter 6 to assess its relationship in areas of bare soil.

## CHAPTER 5

### GENERATION OF TOPOGRAPHIC ATTRIBUTES AND ANCILLARY POINT DATA

*This chapter describes the topographic attributes that were generated for input into the LMU classification. The topographic position as described by landform, compound topographic index (CTI) and slope have been identified as important factors driving soil forming processes in the landscape. A unique piece of software called LANDFORM was developed by Klingseisen (2004b; Klingseisen et al., 2004), in conjunction with this research. LANDFORM is a customised GIS application for semi-automated landform classification based on the definitions by Speight (1990). The CTI originally formulated by Bevan and Kirkby (1979) has been implemented to provide an indication of the relative average soil wetness, as represented by the spatial location of a point within the catchment. Slope has been described as the single most important element of surface form since all surfaces are composed of slopes; gravitational force, controlled by slope angle, drives geomorphic processes (Evans, 1972).*

*This chapter also explains the process followed for extracting all variables at soil points, and the remaining study area for further analysis discussed in Chapter 6 in preparation for the LMU classification described in Chapter 7. It outlines the yield data available and the output consisting of two sets of yield data; (i) an estimated potential yield value at soil point locations encompassing yield data from 1996-2003 (inclusive); and (ii) actual yield at the soil point location during 2002 and 2003 harvest. The regression model used for the estimation of potential yield values at sample point locations is shown and the two yield data sets derived are utilised for subsequent analysis in Chapter 6.*

#### **5.1 Landform Classification**

Several Landform classification techniques have been reviewed in Chapter 2 highlighting Speight (1974; 1990) as a key Australian classification. Coops' *et al.* (1998) methodology has provided key background information in relation to algorithm development and threshold values, and as such it has been utilised in conjunction with Speight's (1990) descriptions of morphological types for the

development of LANDFORM (Klingseisen, 2004b; Klingseisen *et al.*, 2004), whereby, GeoMedia GIS technology has been customised to implement the software. Users can test default threshold values to generate landforms and then adjust them to suit a specific landscape type and DEM resolution. The resulting output landform classification can be used as a layer of input parameters in the classification of LMUs.

## 5.2 Implementation of Landform Classification

Landform Classification has been performed using *LANDFORM* (version 1.0) (Klingseisen, 2004b; Klingseisen *et al.*, 2004) and details of the algorithms can be found in Klingseisen (2004a) and Klingseisen *et al.* (2004). Figure 5.1 provides an overview of the LANDFORM classification process.

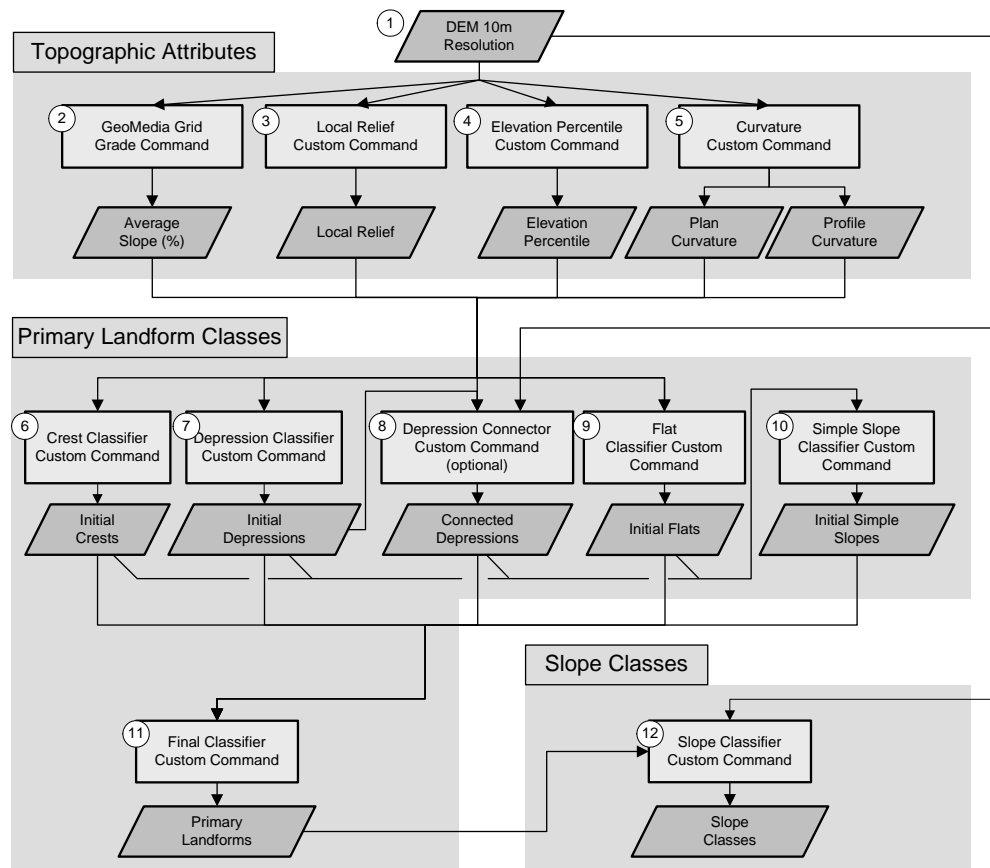


Figure 5.1 Flow chart of implementation of Landform (adapted from Klingseisen, 2004a)

Appendix F provides a step by step description of the commands and threshold values used for the Muresk study area while an overview is explained hereafter.

Based on the Landmonitor DEM, (smoothed using the techniques proposed by Caccetta (2000) as explained in Section 3. 3), five topographic attribute layers were generated namely; slope, local relief, elevation percentile, plan and profile curvature. These were subsequently used as input layers for deriving the primary landform classes. The primary landforms were initially formed as individual layers and then combined to create a single primary landforms layer, which consists of four classes namely; crests, simple slopes, depressions and flats. The default values specified in each case are outlined in Table 5.1 and explained in more detail hereafter.

Table 5.1 Default values used for the generation of primary landforms

<b>Primary Landform</b>	<b>Topographic Attribute values</b>
Crest	Elevation Percentile >6.5 and Local Relief >7.5 Plan Curvature >0.0 or Profile Curvature >0.0
Simple Slopes	Remaining areas not already classified as Crests, Depressions and Flats
Depressions	Elevation Percentile <0.4 or Plan Curvature <-0.50
Connect Depressions	Elevation Percentile <0.48; Acceptable elevation difference 0.00.
Flats	Slope % < 3; minimum radius 20m.

The crest default values specified in Table 5.1 ensure that crests follow Speight's (1990) definition, in that they are typically areas that stand above most other points in the adjacent terrain and have smoothly convex plan or profile curvature (or both). The local relief value ensures that a crest is a significant elevation above the local terrain (Klingseisen, 2004a).

Depressions, as defined by Speight (1990), are areas that stand below most other points in the local terrain and concave upwards. These requirements are fulfilled by the negative plan curvature and low elevation percentile thresholds specified in Table 5.1. In rather flat areas of the terrain, distinct depressions are not present and, as such, the depressions identified solely on these parameters do not form a connected depression network (Klingseisen, 2004a). As a consequence, Klingseisen (2004b) developed the depression connector function to form logical connections between the initial classified depressions (step 8 in Figure 5.1). The function begins with cells classified as depressions and investigates the neighbouring cells that lie in a downflow direction. If these cells are below a specified elevation percentile (Table 5.1), they are subsequently classified as depressions (Klingseisen, 2004a).

Furthermore, setting the acceptable elevation difference to zero ensures that the subsequent cells in the downflow direction are restricted to equal or lower elevation (Klingseisen, 2004b).

Speight (1990) defines flats as areas with a slope gradient of less than 3 percent. However, this means that small areas and thin strips could be formed into flats. Coops *et al.* (1998) suggest introducing an additional parameter to overcome this anomaly. Flats have to be of a minimum width and, although not explicitly defined by Speight (1990), ensures the minimum dimension of a 20m radius for landform elements at map scales of 1:10,000. This concurs with Speight's (1990) recommendation.

Landform elements are arranged (in sequence) down a slope line called a *toposequence* and the position in the toposequence is used to define the unique morphological type of slope elements that can occur between a crest and depression (Speight, 1990) (Table 2.10). Using Boolean algebra, simple slopes are classified as areas not previously classified as crests, flat or depressions in step 10 of Figure 5.1. The final classifier (step 11 in Figure 5.1) combines the four primary landforms into one layer prioritised as; crests > depressions > flats > simple slopes.

Step 12 in the generation of landforms is called the slope classifier which subdivides the simple slopes into upper, mid and lower slope classes. The definition of these classes is presented in Table 2.10 and Figure 5.2 shows their position in a toposequence.

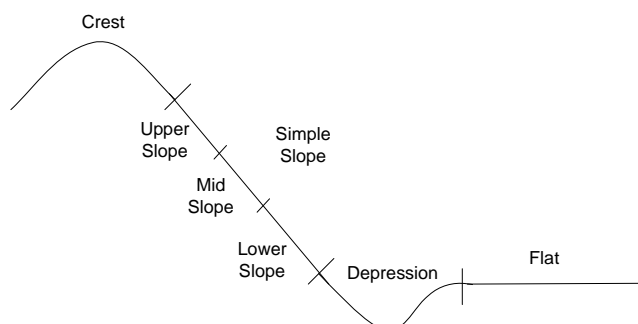


Figure 5.2 Position of slope elements in a toposequence

The slope classifier function generates slope profiles in areas of simple slopes and then scans the profiles for breaks in slope gradient. Cells within simple slopes zones are then subdivided into upper, mid and lower slope classes, depending on the number of breakpoints along a profile and the relative position of a cell (Klingseisen, 2004b). The resultant slope layer is smoothed, with an optional noise reduction filter (e.g. median filter), and combined with the primary landform classes to derive the final landform class map (Figure 5.3).

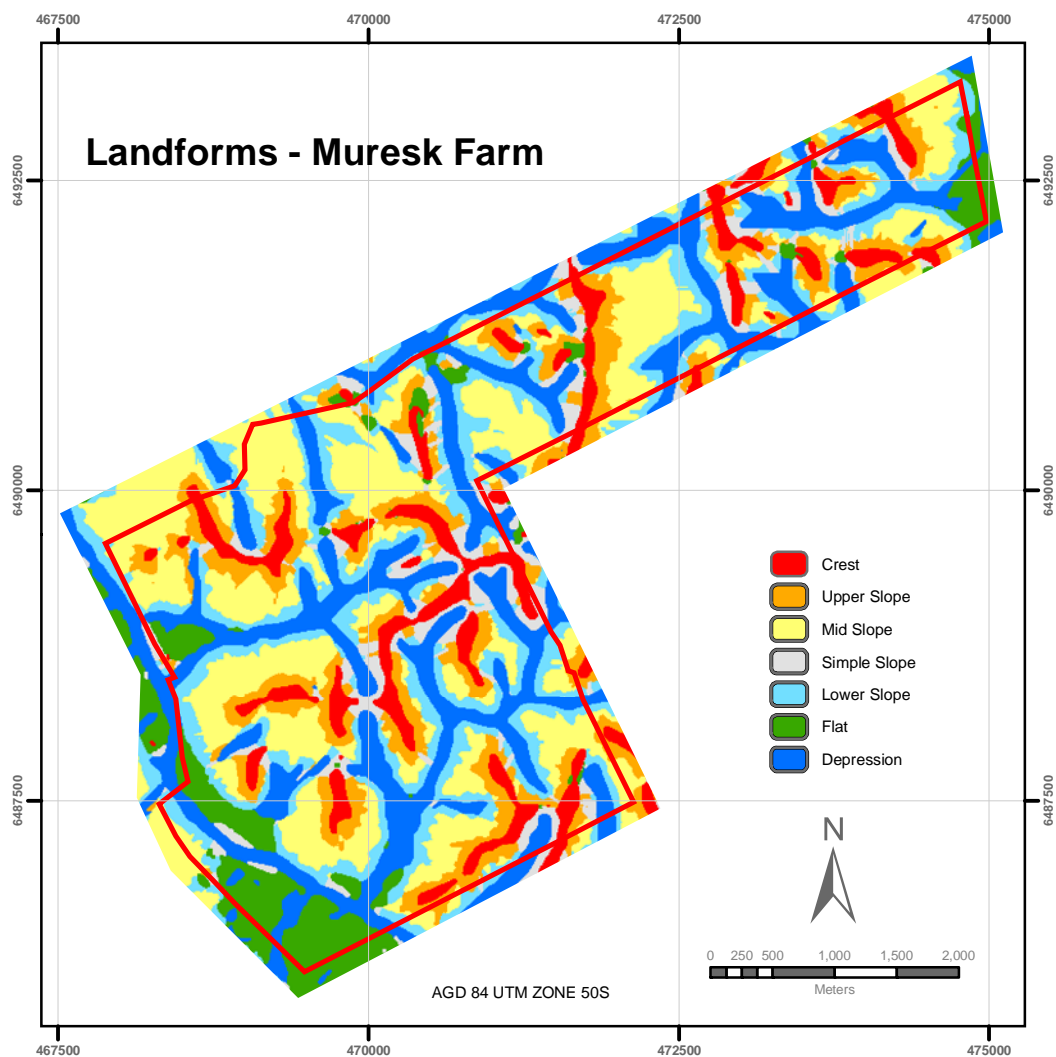


Figure 5.3 Map of landform classes across Muresk farm

### 5.2.1 Comparison between Semi-automated and Manually Derived Landforms

In order to determine the effectiveness of the LANDFORM software, Metternicht *et al.* (2005) compare the similarity between the map generated by LANDFORM and the visual photo interpretation conducted by a soil expert over the same area, using a fuzzy set algorithm. This validation compared the results of LANDFORM against a



landform map produced by 'traditional' photo interpretation methods following guidelines established by Zinck (1988) to determine geomorphic units. Using the Map Comparison Kit (MCK) software the fuzzy set approach proposed by Hagen (2003) was used to compare the similarity of both outputs.

The fuzzy map comparison yielded an average similarity of 0.61. The areas with the highest data matching discrepancies tend to correspond with flat and simple slope landform elements. Metternicht *et al.* (2005) attributed this error to the difficulties in expert classification of subtle changes in slope percentage.

Metternicht *et al.* (2005) found that the most significant differences between outputs of the methodologies can be observed visually, and concluded that while the expert recognises landforms as larger homogeneous areas, LANDFORM generates smaller landform elements. The differences were attributed to the expert having difficulty estimating slope gradation and the inability to determine breaks in slope objectively, when differences are subtle. This resulted in the identification of larger simple slope areas, instead of the subdivision of upper, mid or lower slope, and areas, with slope less than 3 percent, were often not identified as flats by the expert. However, experts are superior at identifying natural drainage systems with connected depressions. These could not be detected by LANDFORM due to roads or other anthropogenic buildings causing barriers in the DEM.

Metternicht *et al.* (2005) conclude that the advantages of LANDFORM, and other automated processes, is the standardisation, increased objectivity and repeatability of the landform classification procedure. This would be particularly evident if used in case studies over large areas (e.g. catchments, regions), as along with the standardisation of the landforms, there is considerable reduction in time required to map them compared to using manual photo interpretation and cartographic processes.

### 5.2.2 Forming Landform Variables

Landforms were required to be converted into landform variables so that they could be used in the LMU classification (detailed in Chapter 7). The conversion has been achieved using principal coordinate analysis.

The principal coordinate analysis is based on a similarity matrix, constructed according to the following rationale. Upper slope (US), mid slope (MS) and lower slope (LS) are all subsets of simple slope (SS) if a breakpoint(s) (significant change in slope) is found in the slope profile of the SS area. An area will remain classified as SS if no breakpoints are found along the profile. In the case of 1 breakpoint the SS area is divided into upper and lower slope. However, when 2 or more breakpoints are detected, SS is divided into US, LS and one or more MS subsets. Flats (F) can be located high in the terrain adjacent crests (C); however they are more likely to be close to depression (D) along the valley floors. A flat is classified as an area with <3 percent slope and a minimum width of 40m. The spatial adjacency of the landform elements present on the Muresk study area was also taken into consideration and the following landform similarity matrix formed.

Table 5.2. The landform similarity matrix

	<b>C</b>	<b>US</b>	<b>MS</b>	<b>LS</b>	<b>SS</b>	<b>F</b>	<b>D</b>
<b>Crest</b>	1						
<b>Upper slope</b>	0.2	1					
<b>Mid slope</b>	0	0.4	1				
<b>Lower slope</b>	0	0.5	0.5	1			
<b>Simple slope</b>	0.2	0.5	0.4	0.5	1		
<b>Flat</b>	0.2	0.2	0.1	0.2	0.2	1	
<b>Depression</b>	0	0	0.2	0.2	0.2	0.3	1

The matrix indicates, for instance, that simple slopes are more similar to upper, mid and lower slopes than they are to crests, flats and depressions. Based on the information above, the best possible allocation has been made; nevertheless a high degree of subjectivity will always be apparent in allocating such values. It is recommended that the landform similarity matrix should be based on a particular study area. If the landform is not being used as the topographic model, then the rationale behind forming the topographic landscape units will need to be reconsidered to create the similarity matrix.

Using the landform similarity matrix (Table 5.2), principal coordinate analysis was performed specifying 6 dimensions. The percentage of variance accounted for by each principal coordinate (PCO1-6) is displayed in Table 5.3, and Figure 5.4 shows the relative positions of each landform when plotted using the PCO1-6 scores.

Table 5.3. Percentage variance accounted for by 7 landform classes

PCO	1	2	3	4	5	6
Percentage of Variance	29.96	23.29	14.90	12.93	10.08	8.84
Cumulative percentage	29.96	53.25	68.15	81.08	91.16	100

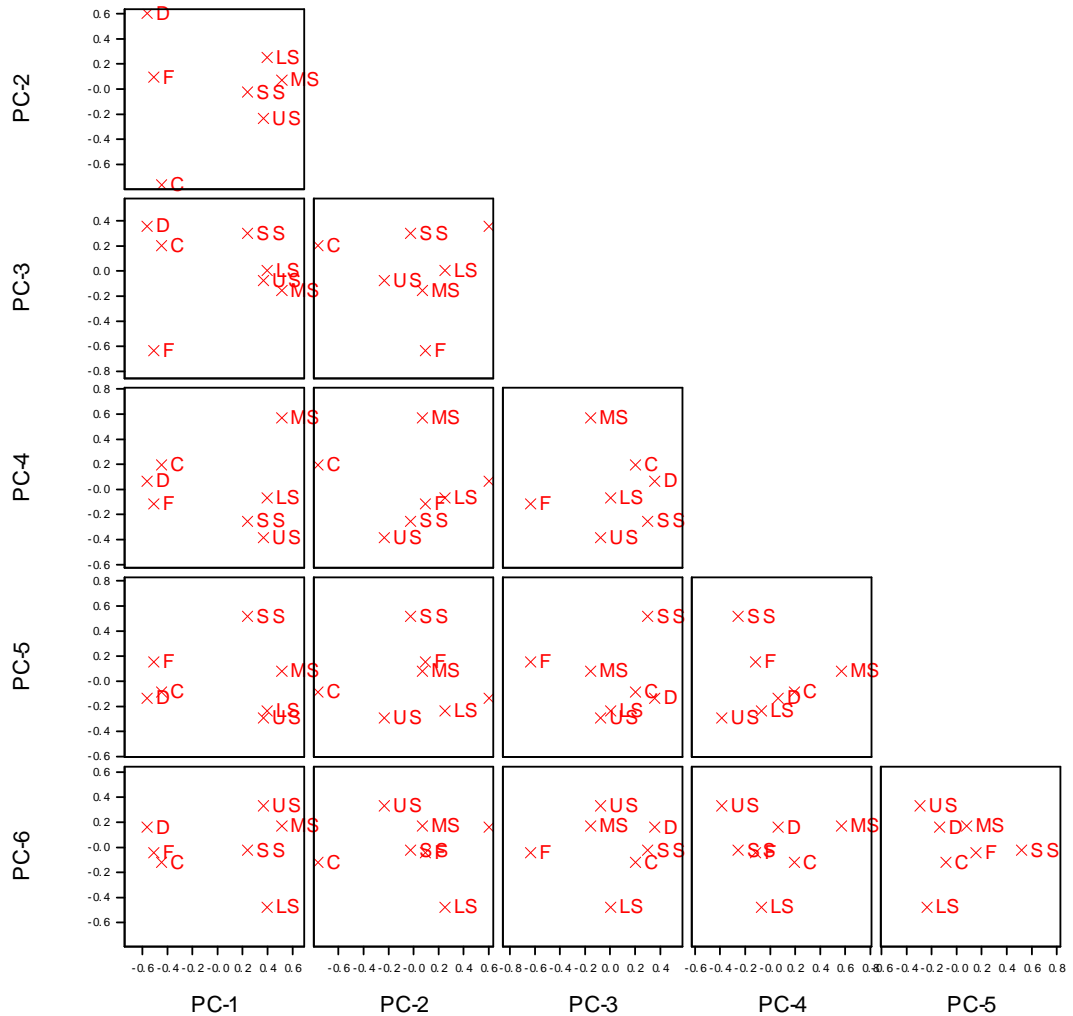


Figure 5.4 Plot of principal coordinate scores

The first two principal coordinates account for over 50 percent of the variance and together separate slopes, crests, flats and depressions. Different slopes are separated by PCO3-6. The principal coordinate scores (Table 5.4) were subsequently used to form six landform variables, which could be treated as continuous data types in the LMU classification.

Table 5.4 Principal coordinate scores for landform PC's

	PCO1	PCO2	PCO3	PCO4	PCO5	PCO6
<b>Crest</b>	-0.446	-0.763	0.203	0.193	-0.086	-0.121
<b>Upper slope</b>	0.366	-0.233	-0.075	-0.386	-0.293	0.332
<b>Mid slope</b>	0.514	0.072	-0.155	0.570	0.081	0.171
<b>Lower slope</b>	0.398	0.252	0.005	-0.069	-0.237	-0.477
<b>Simple slope</b>	0.239	-0.024	0.300	-0.255	0.518	-0.022
<b>Flat</b>	-0.510	0.094	-0.634	-0.117	0.154	-0.044
<b>Depression</b>	-0.561	0.603	0.356	0.064	-0.136	0.161

### 5.3 Compound Topographic Index

The use of the compound topographic index (CTI) Equation (5.1) (Gessler *et al.*, 1995; Irvin *et al.*, 1995) in soil landscape studies has been thoroughly reviewed in Chapter 2, and identified as a secondary topographic attribute that represents the topographic control on soil wetness. The CTI is computed as follows;

$$CTI = \ln (A_s / \tan \beta) \quad (5.1)$$

where  $A_s$  is the specific catchment area (area  $m^2$  per unit width orthogonal to the flow of direction), and  $\beta$  is the slope angle (in degrees) (Gessler *et al.*, 1995; Wilson and Gallant, 2000). The CTI can also be calculated using the TAPES-G program (Irvin *et al.*, 1995). The equation incorporates seven key assumptions and limitations which are mentioned in Wilson and Gallant (2000 pg. 108). The specific catchment area  $A_s$  is the ratio of the contributing area to the contour length,  $A/l$ . The upslope contributing area ( $A$ ) is the area above a certain length of contour that contributes flow across the contour (Figure 5.5) (Gallant and Wilson, 2000). For a raster data type, the contour length ( $l$ ) is approximately the size of a grid cell, and in the simplest case, the contributing area is determined by the number of cells contributing flow to a grid cell (Gallant and Wilson, 2000).

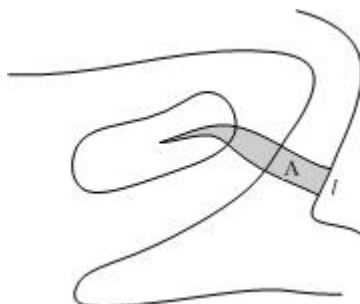


Figure 5.5 Upslope contributing area ( $A$ ) is the area of land upslope of a length of contour ( $l$ ). Specific catchment area ( $A_s$ ) is  $A/l$  (reproduced from Gallant and Wilson, 2000)

The calculation of upslope contributing area ( $A$ ) and resultant specific catchment area ( $A_s$ ), depends on the algorithm for the calculation of flow direction. Gallant and Wilson (2000) provide four different approaches for calculating contributing areas: D8, Rho8, FD8/FRho8, and DEMON. Only the D8 (deterministic eight node) algorithm has been used in this instance. However, other algorithms have been analysed by Gallant and Wilson (2000) and they mention that the D8 algorithm is frequently used primarily because of its simplicity as has been done here. Depressions and flat areas must be dealt with prior to the application of these algorithms.

The D8 algorithm guarantees that the slope calculated at a cell corresponds to the slope in the primary flow direction (Gallant and Wilson, 2000). Flow is only possible to one of eight nearest neighbours (based on primary flow direction). However, D8 is subject to limitations such as an inability to model flow divergence. Nonetheless it is adequate for delineating catchment boundaries (Gallant and Wilson, 2000). The number of cells that drain into each cell is determined post calculation of flow direction for each cell. The upslope contributing area ( $A$ ) is then determined by multiplying the number of contributing cells with the cell resolution. The flow width,  $w$  (in terms of cells i.e.  $l$  in Figure 5.5) is equal to the cell width for flow in all cardinal directions and the cell diagonal for the flow in diagonal directions:

$$w = \begin{cases} h & \text{cardinal directions} \\ h\sqrt{2} & \text{diagonal directions} \end{cases} \quad \text{where } h = \text{cell width}$$

Specific catchment area ( $A_s$ ) is calculated as the ratio of the contributing area to the flow width,  $w$  (Equation (5.2)).

$$A_s = A / w \quad (5.2)$$

Therefore, for a cell resolution of 10m, in cardinal directions  $w = 10$  and for diagonal directions  $w = 14.142$ .

The D8 algorithm calculates slope using the steepest downhill slope to one of the eight nearest neighbours according to the following equation;

Z <sub>7</sub>	Z <sub>8</sub>	Z <sub>1</sub>
Z <sub>6</sub>	Z <sub>9</sub>	Z <sub>2</sub>
Z <sub>5</sub>	Z <sub>4</sub>	Z <sub>3</sub>

$$S = \max_{i=1,8} \frac{z_9 - z_i}{h\Phi(i)} \quad (5.3)$$

where;  $\Phi(i) = 1$  for cardinal (north, south, east, and west) neighbours ( $i = 2, 4, 6$  and  $8$ );  $\Phi(i) = \sqrt{2}$  for diagonal neighbours (in order to account for the extra distance to those cells) (Gallant and Wilson, 2000); and  $h$  is the cell width. For the computation of CTI, if the slope ( $S$ ) is in units of percent the *tan* is not taken, as slope (in a mathematical sense) = rise/run =  $\tan \beta$ . However, slope percent is often  $S \times 100$  and therefore, the slope computed from the algorithm may need to be divided by 100. It is envisaged that the CTI will provide information about topographic position and relative average soil wetness for the LMU classification (Chapter 7), as the contributing area represents how much water arrives at a specific location while the slope gives an indication on how long it will linger.

#### 5.4 Implementation of CTI

The CTI was generated with GeoMedia Grid using the mathematical concepts described above (Section 5.3), and the series of steps flowcharted in Figure 5.6. A more detailed description of the commands used and values set, are described in Appendix G along with a detailed description of the algorithm for the slope function (Appendix H).

Like the landform generation, the CTI is generated from the Landmonitor DEM, which was smoothed using the algorithm proposed by Caccetta (2000). Firstly, a “depressionless” DEM is formed filling pits, ponds and depressions, which is useful for hydrological analysis commands. The downhill path is then computed to produce a grid layer of downhill path directions, where the value of each cell represents the direction of steepest slope (Keigan Systems Inc., 2003b). When based on a “depressionless” DEM, flow is possible in only one of the eight directions. This is synonymous with the D8 algorithm discussed in Section 5.3. The upslope contributing area ( $m^2$ ) is calculated by firstly generating a downhill accumulation layer that assigns each cell a value representing the number of cells on a direct uphill path from the selected cell (Keigan Systems Inc., 2003b). The cell value is multiplied by the cell area, which in this case is  $100m^2$  (i.e. 10m cell resolution).

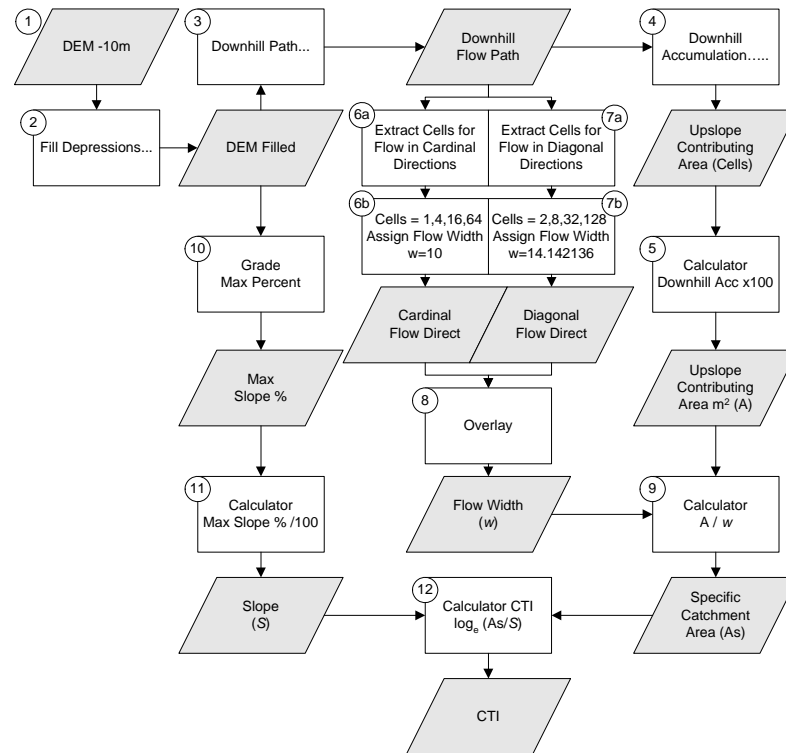


Figure 5.6 Flow chart of implementation of CTI in GeoMedia Grid

The specific catchment area ( $A_s$ ) layer is generated by dividing the catchment area by flow width ( $w$ ). Flow width is determined using a series of steps (6-8 as shown in Figure 5.6), which assigns a value of  $w=10$  and  $w\approx 14.14$  for cells with flow in a cardinal and diagonal direction, respectively. The maximum slope in percent is generated and subsequently divided by 100 to compute  $S$ . This is because the slope command in GeoMedia Grid follows Equation (5.3) of Gallant and Wilson (2000) as shown in steps 10 and 11 of Figure 5.6. Finally CTI is calculated as follows:

$$\log_e(A_s/S) \quad (5.4)$$

where  $A_s$  is the Specific Catchment Area layer and  $S$  the slope layer.

In order to preclude the generation of very large values along major streams and void values due to a denominator of zero for slope, the parameters of McKenzie and Ryan (1999) were adopted for the generation of CTI, by editing cells with a value of 0 in the slope layer to 0.0001 and assigning a value of 100,000 to values  $\geq 100,000$  in the specific catchment area layer. McKenzie and Ryan (1999) used the CTI as one of the parameters for a digital stratification of the landscape in which they limited the contributing area (i.e. upslope contributing area in  $m^2$ ) to a maximum value of

<100,000m<sup>2</sup>. They also recoded zero slope cells with a small value (0.01 percent) to avoid a denominator of zero, and slopes above 300 percent were set to that value.

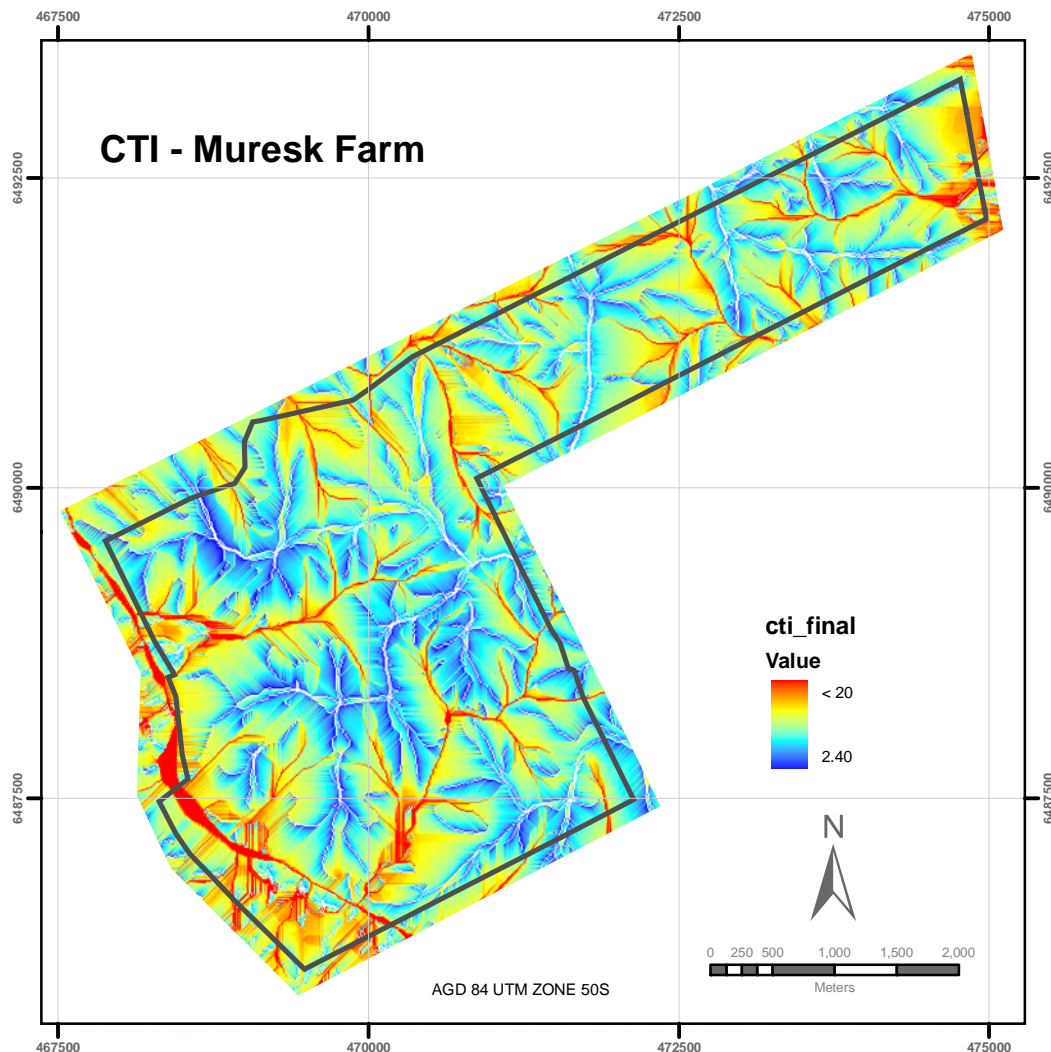


Figure 5.7 CTI layer across Muresk Farm

The CTI in Figure 5.7 is somewhat stripy. This is indicative of CTI layers created using D8 algorithm approaches. The image could be enhanced visually by applying a smoothing filter or more simply by re-classing the values into integer ranges following Herron *et al.* (2004) and Moore *et al.* (1993a). Herron *et al.* (2004) computed the CTI and reclassified the values into integer classes with all CTI values greater than 19 being classed as 19, as these areas coincide with stream networks, while Moore *et al.* (1993a) provides a 3D representation of CTI for the study area in which integer classes have been selected. However, this would be for only visual enhancement and therefore has not been performed. Instead, CTI was grouped into 0.05 range classes, and these values were extracted for further analysis. Some cells



(5 percent of the layer) within the final CTI layer have a void value. This was derived from a value of zero present within the Specific Catchment Area layer, i.e. cells, which are high in the landscape with no other cells contributing flow to that cell, as is the case along crests.

### 5.5 Generation of a Slope Percentage Map

Slope percentage has been identified as one of the most highly correlated terrain attributes to soil attributes (Moore *et al.*, 1993a), and identified as the single most important element of surface form (Evans, 1972). Therefore it will subsequently be analysed for its use in the LMU classification in Chapter 7. The slope in percent was generated across the study area using GeoMedia Grid and the algorithm is described in Appendix H. The slope layer was imported into ARCGIS for subsequent extraction of data. Figure 5.8 displays the slope layer generated across the Muresk farm. It shows that the majority of the Muresk farm consists of gently inclined slopes with a mean slope of 5 percent.

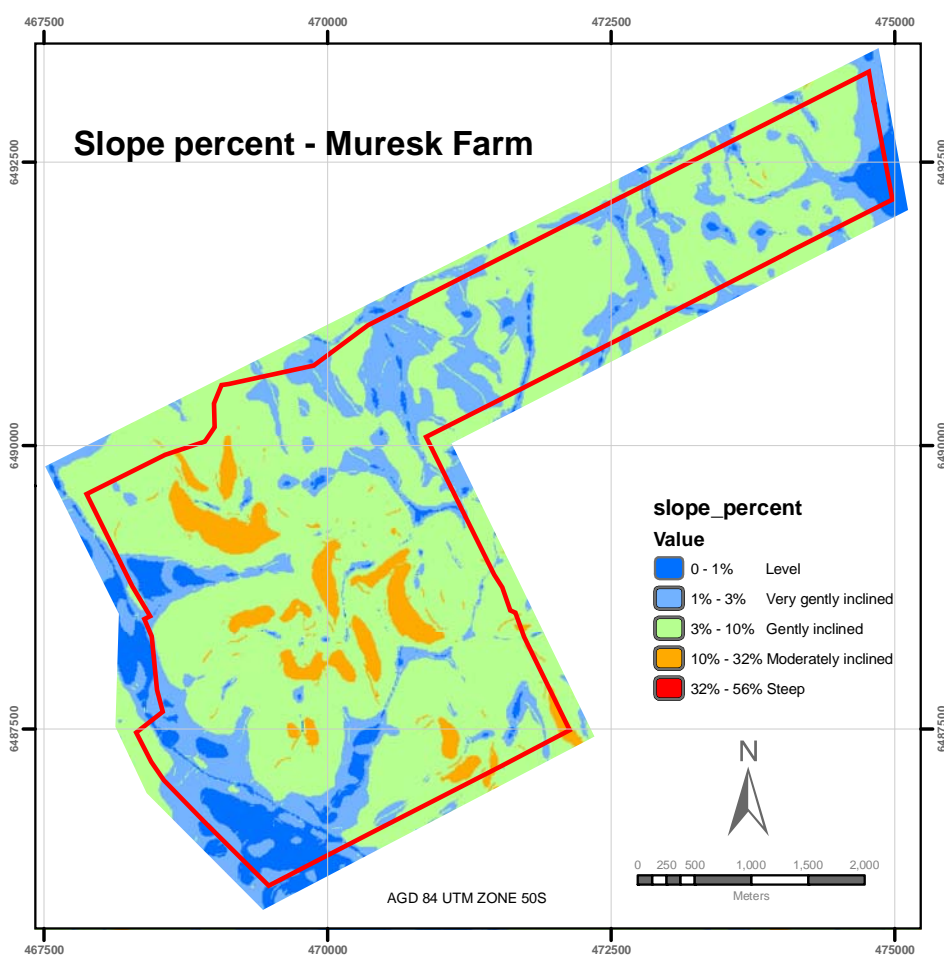


Figure 5.8 Slope percentage map of the Muresk farm based on McDonald *et al.* (1990) slope classes

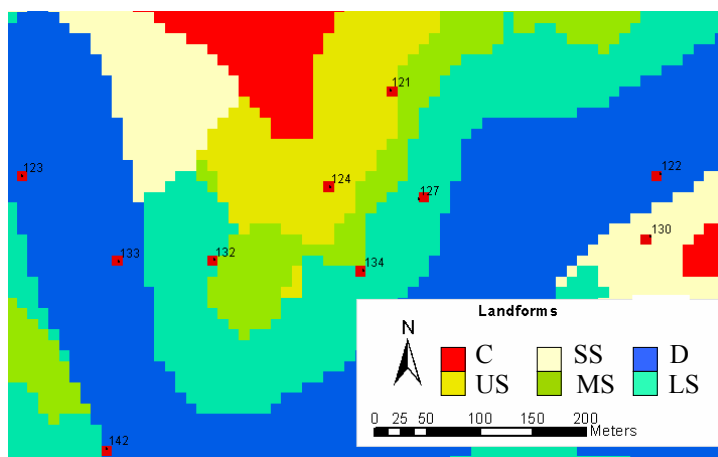
## **5.6 Extracting Ancillary Data at Soil Point Locations**

For analysis between variables, all variables need to be extracted at the 250 soil point locations surveyed in the field, so they can be used as input data for the LMU classification. However, yield data is not available for all 250 soil points and therefore it will not be used as an input to the LMU classification. Nevertheless, yield data will be used for statistical analysis against soil properties to provide an indication of soil variables in order to explain the variability in yield (Chapter 6).

The extraction of yield data at points have been defined by two yield data sets explained in Section 5.7. The soil physical and chemical properties, at each of the soil points, has been stored against a unique identifier with associated geographical coordinates in the soil database. The unique identifier has been used for the following extraction of ancillary variables.

### **5.6.1 Extracting Topographic Attributes**

Landform, CTI and slope were extracted at each soil point location as they will be used as inputs in the LMU classification described in Chapter 7. Firstly, the landform, CTI and slope layers were exported from GeoMedia Grid software and imported into ArcGIS as raw binary files. For each soil point the topographic value for the 10m cell was extracted. For the CTI layer, some soil points did not have any data, as they were associated with a specific catchment area of zero and therefore a void CTI value was assigned instead. Figure 5.9 displays a section of the study area where the soil sampling points are overlaid on the landform map.



Crest (C); Upper slope (US); Mid slope (MS); Simple slope (SS); Lower slope (LS); Depression (D)

Figure 5.9 Soil sampling points converted to 10m raster cells overlaid on landforms

### 5.6.2 Extracting DMSI for Soil Points in Areas without Vegetation Cover

DMSI data based on each soil point location will be used as an input in the LMU classification described in Chapter 7. The Redness Index (RI) (Escadafal and Huete, 1991)(Equation (2.9)) was generated for the entire study area and thus it was appropriate to determine DMSI locations with minimum vegetation cover referred to in this case as *bare soil*. Two flights (Table 3.3: Flights 1 and 4) of DMSI data were acquired during June, in order to capture the farm with minimum vegetation cover.

Seeding offers an optimal time when most of the soil is bare. In liaison with the farm manager, the appropriate time for image capture was determined to be when the majority of the crops were seeded, but prior to extensive leaf growth. The window of capture time is varied by the winter break, crop types and cloud cover. During 2002 the imagery was captured on 28<sup>th</sup> June (Table 3.3: Flight 1) while in 2004 the imagery was captured on 15<sup>th</sup> June (Flight 4). The other two flights of imagery, acquired during September, also provide some locations where vegetation cover is minimal and as a consequence, will be analysed.

Figure 5.10 provides an overview of the steps used to extract DMSI data at each soil point ensuring that the location has minimum vegetation cover. The steps are discussed hereafter.

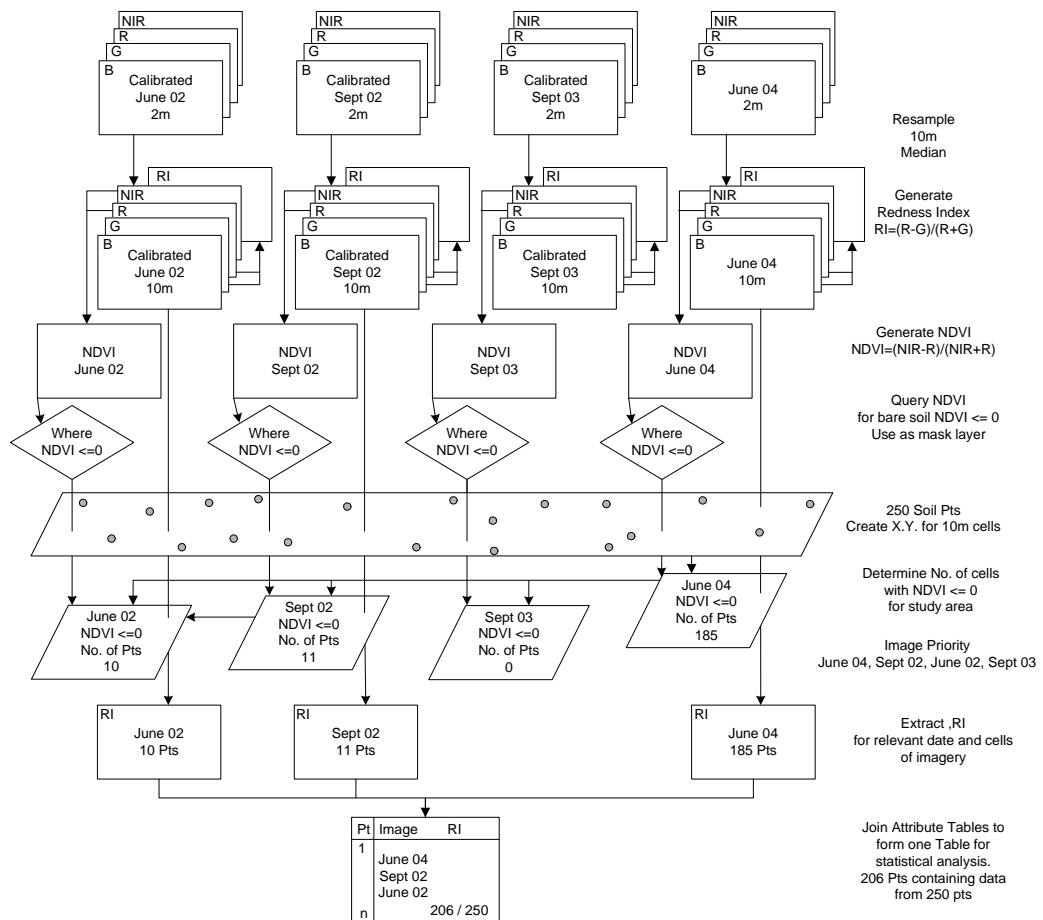


Figure 5.10 Overview of process for extracting DMSI data at soil points with bare soil

### 5.6.2.1 Selecting Locations with Minimum Vegetation Cover

The DMSI data has a spatial resolution of 2m. However, when trying to create a similarity matrix for the entire property, a 2m grid will form over 3 million points for the study area. This scenario requires a powerful computer to process the data. Therefore, DMSI data were resampled to a 10m grid. This divided the total points by 25 using the median value. There are several options that could be considered including the mean, mode and minimum.

In this study the median was used. The mean or median are the most common choice when the objective is neither to highlight nor smooth features in the imagery. The mean and median value provides a good measure of the central value of the data; however the mean can be influenced by extreme values and as such if the distribution of the data is unknown using the median would be more sound.

Not all soil points are located in cropped areas and nor will they have minimum vegetation cover in all years. During the June flights (1 and 4) image capture, field data were collected (Section 3. 5) in order to provide an indication of the amount of vegetation present at the time of flight. A visual density was not recorded in the field, but was estimated from the photos captured. Several paddocks were visited on that day, photos taken and crop and vegetation attributes recorded. The NDVI (Equation (2.4)) has been calculated for Flights 1 and 4 and compared against field samples to determine a NDVI threshold value that indicates minimum vegetation cover. This provides a quantitative method of determining whether a point location can be considered for analysis between spectral data and soil properties. By comparing the field vegetation density measures with the NDVI for June 2002 and 2004, it appears that a NDVI equal to 0 is the appropriate threshold, as any value higher than this may contain vegetation cover at levels greater than 10 percent.

A NDVI layer was generated for each flight and the value of NDVI extracted for each soil point. Table 5.5 provides an indication of the number of sampling points from each flight with  $NDVI \leq 0$ . This can subsequently be used for extracting DMSI values and then used in the statistical analysis with soil properties.

Table 5.5 Number of soil points with  $NDVI \leq 0$

	<b>June 2002</b>	<b>Sept 2002</b>	<b>Sept 2003</b>	<b>June 2004</b>	<b>Total Points</b>
Unique points	10	11	0	185	206

From Table 5.5 all soil points available for the June 2004 image will be used with DMSI from that associated image. Any new unique points for Sept 2002 followed by June 2002 will be used with DMSI extracted from there associated image. The order of imagery is selected based on the descending order of number of unique points available with  $NDVI \leq 0$ .

## 5.7 Yield Data Sets

A discussion and presentation of the available yield data are in Section 3. 4 however, the following table (Table 5.6) summarised the number of points where yield data are

available, the range of crop types and years for each of the yield data sets. The method for extracting the yield data is explained below.

Table 5.6 A description of the two yield data sets and their attributes

Yield Data Set	Description	No. of points with yield data	Range of Crop Types	Range of Years
1	Estimated Potential Yield at Soil Points	184	7	7
2	Actual Yield at Soil Points	125	6	2

### 5.7.1 Extraction of Interpolated Yield Values

Using ArcGIS spatial analyst, a 3 x 3 (12m x 12m) neighbourhood window was generated around each soil point. Figure 5.11 provides a visual display of some of the soil sampling points and their 3 x 3 window overlaid on a true colour DMSI composite (Blue: Band 1, Green: Band 2, Red: Band 3).



Figure 5.11 Soil sampling points with 3 x 3 neighbourhood window overlaid on a true colour DMSI composite

Using the interpolated surface as a base, the mean yield value (kg/ha) for the 3 x 3 window (i.e. nine, 4m<sup>2</sup> raster cells) was extracted and recorded along with the associated crop type and year. In the same manner, the mean variance value from the variance surface created during the interpolation was also extracted. During the extraction process it was recognised that some of the soil sample points lay at some distance (or on the perimeter) from the field recorded GPS yield points that were used for the interpolation (e.g. point's 157 and 158 Figure 5.12). As such, the

distance from these points or perimeter location was noted, as it was thought that this could provide useful information during the regression analysis. If some points were giving large standardised residuals or leverage in the model, it may be due to the confidence of the interpolation. A spreadsheet containing each soil point with the attributes discussed above, as listed in Table 5.7 was formed.

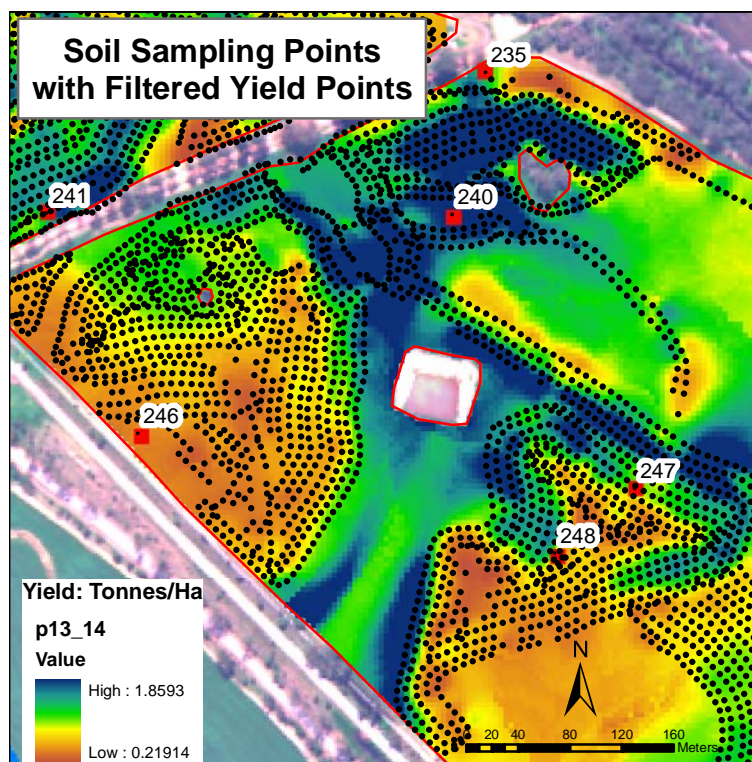


Figure 5.12 Soil sampling points and their distance from field recorded GPS yield points overlaid on yield interpolated surface

Table 5.7 List of attributes recorded for soil and vegetation point locations during yield extraction

Attribute
Paddock Number
Crop Type
Year
Sample Point Number
Mean Yield (kg/ha)
St. Deviation
Distance from GPS yield points (m)

### 5.7.2 Potential Yield Estimation at Sample Points (Yield Data Set 1)

Due to concerns of herbicide resistance, plants or weeds that host disease, fertiliser history, nutrient balance, stubble residue (TopCrop Australia, 1999a) and economical factors, paddocks undergo a rotation of crop types. Consequently, each paddock and thus the soil sampling point location within them, will have experienced a range of crop types between the years 1996 – 2003 inclusive (excluding 2001, as yield data was not recorded). Yield data for paddocks from which GPS yield data was recorded were interpolated to produce continuous surfaces of estimates of yield at each soil sample point (Figure 5.12). Some soil sampling points were not located in arable areas of the farm, and thus not all 250 points have interpolated yield values.

A standardised value of yield is required at each soil sampling point location, as an indicator of the yield potential across all years and crop types. Such a value is calculated by adjusting the interpolated yields at each sampling point according to the effects of crop type and year on measured yield values. A multiple regression model incorporating crop type and year effects (i.e. weather impacts), as well as, sample point effects, was used to estimate a potential yield at each point for use in subsequent analyses. The dependent variable in the regression was the interpolated yield available for each soil sampling point and year.

#### 5.7.2.1 Multiple Regression Analysis

The data were imported into GenStat® (Laws Agricultural Trust, 2003a) for analysis and model building. Crop Type, Year and Sample Pt, are all set as factors<sup>2</sup> while Mean Yield (kg/ha) is a variate<sup>2</sup>. In total there were 691 records of yield for 184 soil sample points with yield data available. The regression model fitted was as follows;

$$Y_{ijk} = \mu + S_i + C_j + Yr_k + CYr_{jk} \quad (5.5)$$

where  $Y_{ijk}$  is the average interpolated yield for sampling point  $i$  in year  $k$  with crop type  $j$ ;  $\mu$  is intercept;  $S_i$  is the effect of sample point  $i$ ;  $C_j$  is the effect of crop  $j$ ;  $Yr_k$  is the effect of year  $k$ ; and  $CYr_{jk}$  is the effect of the interaction between crop  $j$  and year  $k$ .

---

<sup>2</sup> Structures that store lists of numbers in GenStat® are known as variates whereas the categorical structures are called factors Laws Agricultural Trust (2003b) *Introduction, GenStat for Windows 7th Edition*, VSN International, Oxford, UK. 336 pp.



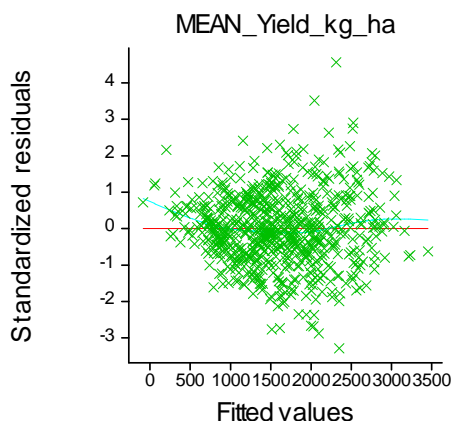


Figure 5.13 Plot of residuals against fitted values following Equation (5.5)

Statistical assumptions underlying regression analysis were examined using plots of residuals. The plot of residuals against fitted values (Figure 5.13) indicates that the variance of the residuals increases with mean yield. This is of some concern. However, it was eliminated by transforming the yield data using a natural log transformation. Subsequent models incorporated the log Yield data, and the second model tested took the form;

$$\log_e(Y_{ijk}) = \mu + S_i + C_j + Yr_k + CYr_{jk} \quad (5.6)$$

Three sample points (i.e. 167, 169 and 176) were measured only once in 1998 for oats sown, and were the only points where oats were measured in 1998. Therefore, a good estimate of adjusted yield at these sample points is not possible unless a lack of significant difference between oats and another crop type is shown. The most logical comparison is between oats and oaten hay, and this was performed and found to be highly significant ( $P < 0.001$ ; see Appendix I) indicating that sample points 167, 169 and 176 should be removed from the data set.

The regression analysis highlighted three records with large standardised residuals and 21 records with high leverage. The later were all sampling points that had been measured only once and thus were left in the data set. However, the three records with large standardised residuals were removed from the following model, as those sampling points had yield data associated on four or more occasions.

The final regression model accounted for 54.2 percent of the variance in log Yield with a standard error of 0.204 and was highly significant ( $P < 0.001$ ). The residual plot shown in Figure 5.14 indicates an adequate model and therefore, this model was used to estimate yield values for each sampling point in the model.

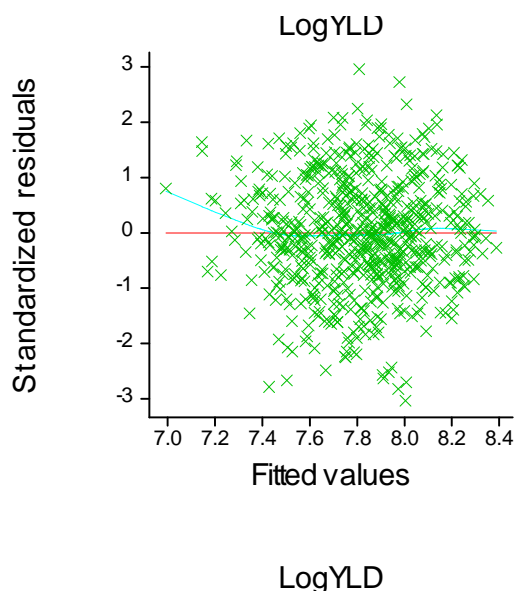


Figure 5.14 Plot of residuals against fitted values for the final model

#### 5.7.2.2 Estimation of Potential Yield at Sample Point locations

Based on the above model, standardised log yield values, representing an average across crop types and years, were estimated for each of the 181 sample points. Appendix J lists the log Yield, transformed Yield (kg/ha) and standard error associated with their estimations. This will be utilised as the potential estimated value at available soil points in subsequent analysis.

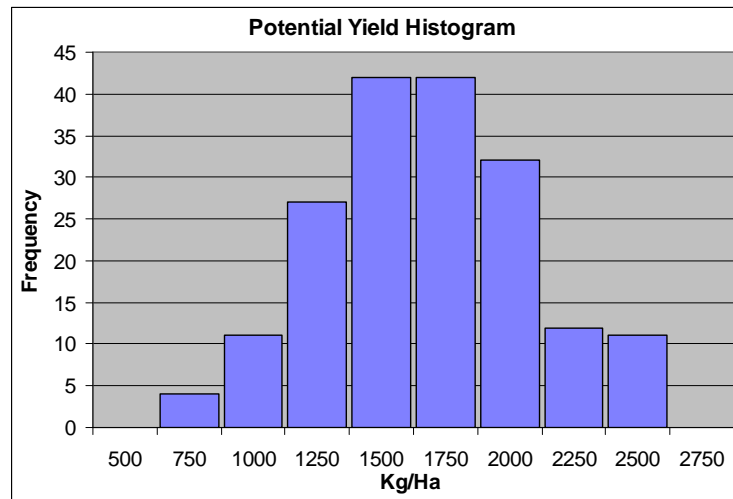


Figure 5.15. Frequency distribution of the Standardised Yield Estimations for the Soil Points

### 5.7.3 Actual Yield Data at Soil Points for 2002 and 2003 (Yield Data Set 2)

Yield data set 2 contains the actual yield data recorded at the soil points that are available for the years 2002 and 2003. The values were extracted as described in Section 5.7.1 and their frequency is displayed in Figure 5.16 and Figure 5.17.

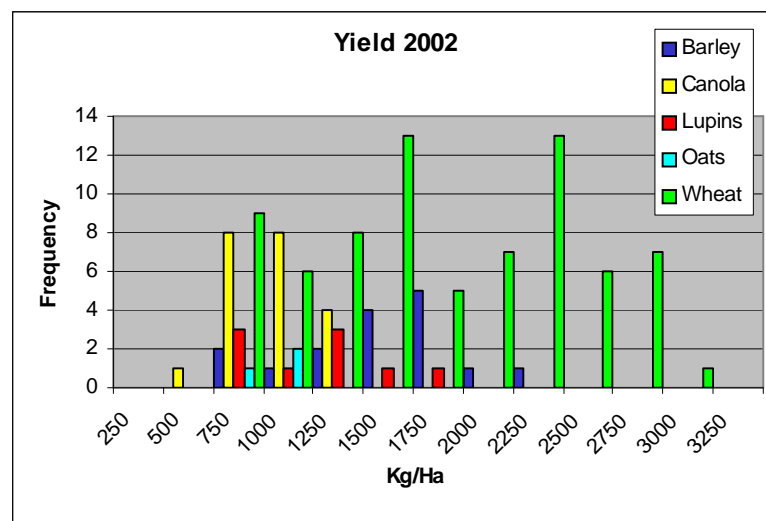


Figure 5.16. Frequency distribution of the yield 2002 at soil points

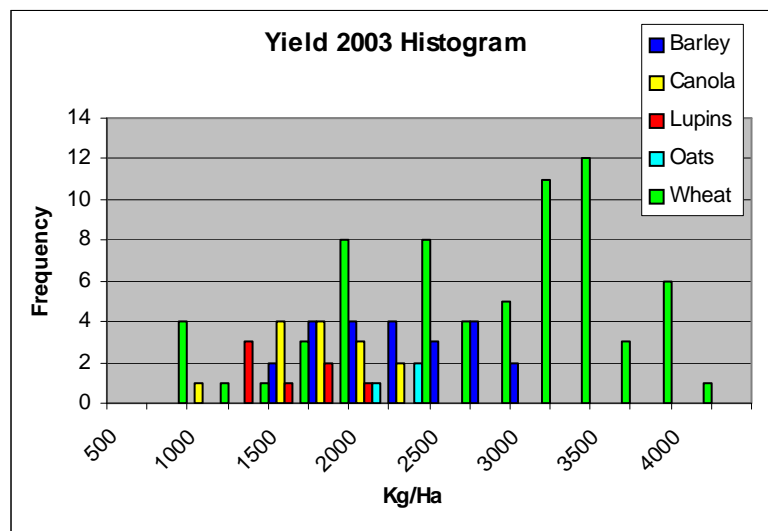


Figure 5.17. Frequency distribution of the yield 2003 at soil points

## 5.8 Extracting Ancillary Data for the Mask Area

DMSI, Landform, CTI and Slope are the ancillary data available for the entire study area. These data layers were extracted and stored in a spreadsheet to be utilised in Part B of the LMU classification described in Chapter 7.

The entire study area includes areas of remnant vegetation, rocky outcrops, creeklines, tracks, and dams. A layer was formed by *heads-up digitizing* to exclude these type of features on a true colour base image and the remaining area to be included in the classification has been named the mask area. The mask area file was converted to a raster layer of 10m pixel resolution and subsequently, each 10m pixel allocated an ID, and its easting and northing assigned based on the centre of each 10m pixel. This formed a file of 118864 points which equates to approximately 1190ha of arable land (approximately 67 percent of total area of Muresk farm) to be included in the classification.

### 5.8.1 Extracting Topographic and DMSI Variables on a 10m Grid

Using the mask points located at the centre of each 10m pixel, the landforms, CTI, slope% and the Redness Index (RI) were extracted for the 118864 pixels in the mask region of the study area. Of these points, 242 are soil points and, as such, only 118622 will be included as the mask area for the LMU classification and the remaining 242 as soil points.

The following flow diagram (Figure 5.18) displays the methodology for extracting DMSI across the entire study area on a 10m grid. Ensuring pixels used with the RI are associated with bare soil, in a similar manner as described for the soil points in Section 5.6.2.

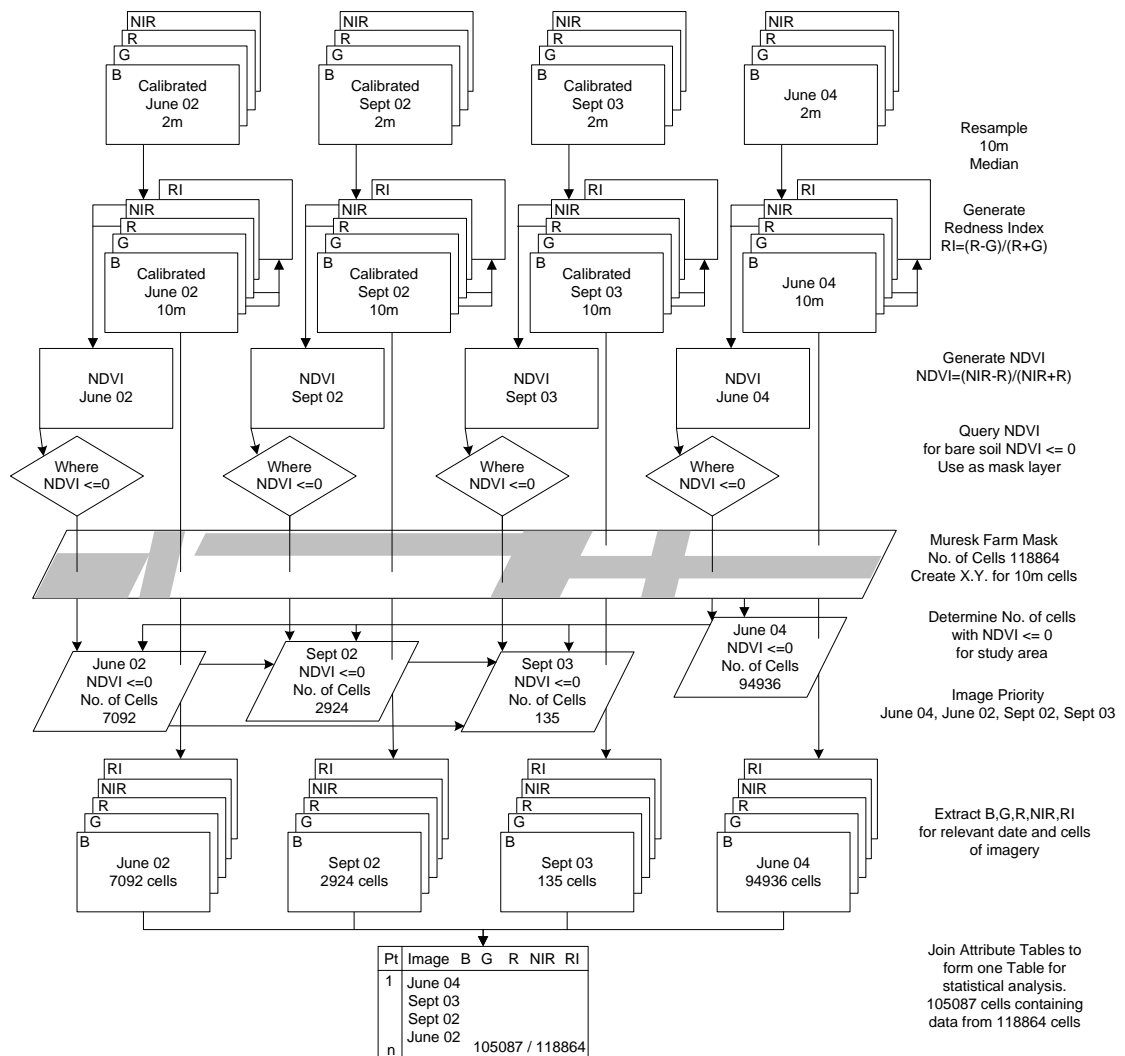


Figure 5.18 Overview of process for extracting DMSI data from 10m cells

### 5.9 Summary

This chapter has presented the generation of landforms, CTI and slope layers. All three layers were generated using GeoMedia software and will be analysed for their appropriateness as inputs to the LMU classification in the following chapters. Landform variables have been formed from the landforms using principal coordinate analysis. This chapter has also described the extraction of variables to be used for statistical analysis and classification of LMUs. The process of estimating a potential yield value at soil points to be used in subsequent analysis has been shown. The

following chapter (6) uses these extracted data layers in statistical analysis to determine the appropriate input attributes in the LMU classification (Chapter 7).

## CHAPTER 6

### SELECTION OF VARIABLES FOR USE IN FORMING LAND MANAGEMENT UNITS

*The purpose of this chapter is to determine what information should be used to form land management units (LMUs) and to examine the methods available for making this decision. In this study a very large number of soil properties were measured at two depths at each soil sampling point. It is expected that only a subset of these properties will be required to form LMUs and that data collection will be restricted to these variables if the methodology is successfully applied on-farm afterwards.*

*However, since the subset of soil properties selected for forming LMUs at Muresk may not be the best set in diverse environments, it is also important to examine ways of selecting an appropriate subset. In addition to the soil properties at soil sampling points, other data sets available include yield measurements over seven years and a number of ancillary variables (landform, CTI, slope and RI) at very high resolution which cover the remaining study area. The yield data will be used in statistical analysis with the soil properties and ancillary variables to determine the landscape attributes most influencing yield variability.*

#### **6.1 Selecting and Deriving Stable Soil Properties from Existing Literature**

Based on the literature review undertaken in Section 2.2.1, stable soil properties that are important for plant growth are; position in the landscape (topography), texture (and coarse fragments), structure (stability), organic matter, nutrient availability, pH, salinity / sodicity balance, depth of topsoil and depth to restricted layer. The soil properties available for this project that could be considered to represent these aspects of soil are; landform, CTI, slope, soil particle size distribution (percentage of sand, silt and clay) and percentage of stones, organic carbon, ECEC, pH, EC, ESP, depth of A-horizon and depth to restricted layer (i.e. compaction or impeding stones) (Table 6.1).

Table 6.1 Stable soil properties and variables from literature review

Property	Variable
Topography	Landform, CTI, Slope
Texture	Percentage of sand, silt and clay
Coarse fragments	Percentage of stones
Organic matter	Percentage of Organic Carbon
Capacity of soil to store nutrients cations	Effective cation exchange capacity (ECEC)
Acidity/Alkalinity	pH
Salinity	EC
Sodicity	Exchangeable sodium percentage (ESP)
Soil stability	Ca:Mg ratio
Depth of topsoil	Depth of A-Horizon
Depth of restricted layer	Depth of compaction or impeding stones

Some derivative soil properties needed to be formed from the soil properties that were determined in the chemical analysis of soils (Section 4. 3). The effective cation exchange capacity (ECEC) in this research has been calculated as the sum of the Ca, Mg, Na, K, Al and Mn (Allen, 2004) as shown in Equation (6.1) (see Section 2.2.1.5 for further information). The exchangeable sodium percentage (ESP) is determined as the percentage of sodium relative to ECEC (Allen, 2004) shown in Equation (6.2), while the Calcium to Magnesium ratio was determined as the proportion of Ca (me%) divided by the proportion of Mg (me%) (Allen, 2004) shown in Equation (6.3).

$$ECEC = (Ca + Mg + Na + K + Al + Mn) \quad (6.1)$$

$$ESP = \frac{Na}{ECEC} * 100 \quad (6.2)$$

$$Ca : Mg = \frac{Ca}{Mg} \quad (6.3)$$

The ancillary variables, landform, CTI and slope percentage account for the topographic position and are considered an indication of the potential water availability in the terrain. The redness index (RI) derived from DMSI provides an indication of soil colour as an indirect indicator of soil type as mentioned in Section 2.3.2.4.4.



## **6.2 Selecting Variables Based on their Association with Crop Yield**

One rationale for selecting variables for use in forming LMUs is that they can be viewed as drivers influencing the productivity of a paddock. The following statistical analysis was performed to determine the variables that explain the variability in yield. The potential yield, actual yield in the year 2002 and actual yield in the year 2003 data sets (Section 5. 7) will be used to this end.

### 6.2.1 Transforming Variables

Regression analysis is generally most successful when the explanatory (independent) variables have reasonably symmetric distributions. Accordingly, the histogram and the skewness statistic were used to examine visually and statistically the distribution of each variable. Following this a decision was made whether to transform the data to form more symmetrical distributions.

Variables with an absolute skewness coefficient greater than one were transformed. Subsequently box-plots were used to detect outliers, and extreme outliers were examined and removed if appropriate. Table 6.2 lists the variables examined and the transformation applied (if required) while Appendix K and L provide the summary statistics, histograms and box-plots for each of the variables. The resultant normally distributed variables will be used in the statistical analysis presented in the following sections.

Table 6.2 Variable transformations and resultant skewness

Variable	Transformation	Resultant No. of Sample Points	Resultant skewness	Variable Name
A-Horizon depth (cm)	$\text{Log}_e(\text{A-Horizon}+1)$	246	-0.814	Log_A_HORIZ
Restricted Depth (cm)		229	0.800	RESTRICT_DPTH_CM
10_Stones_ %		250	0.762	S10_STONES_ %
10_EC(1:5)	$\text{Log}_e(10\_EC)$	250	0.729	Log_10_EC
10_pHw	$\text{Log}_e(10\_pHw+1)$	250	0.737	Log_10pHw
10_pHca	$\text{Log}_e(10\_pHca+1)$	250	0.880	Log_10_pHca
10_Sand	$\text{Log}_e(100-10\_Sand)$	250	0.646	Log_10_SAND%
10_Silt	$\text{Log}_e(10\_Silt)$	250	0.368	Log_10_SILT
10_Clay	$\text{Log}_e(10\_Clay)$	249	0.714	Log_10_CLAY%
10_OrgC		250	0.944	S10_OrgC_ %
10_ECEC	$\text{Log}_e(10\_ECEC)$	250	0.480	Log_10_ECEC
10_ESP	$\text{Log}_e(10\_ESP+1)$	250	0.931	Log_10_ESP
10_Ca_Mg		250	0.188	S10_Ca_Mg
30_Stones_ %	$\text{Log}_e(30\_Stones+1)$	231	-0.908	Log_30_STONES
30_EC(1:5)	$\text{Log}_e(30\_EC+1)$	230	0.830	Log_30_EC
30_pHw		231	0.418	S30_pH_H2O
30_pHca		231	0.494	S30_pH_CaCl2
30_Sand	$\text{Log}_e(100-30\_Sand)$	231	0.348	Log_30_SAND%
30_Silt	$\text{Log}_e(30\_Silt)$	231	0.078	Log_30_Silt
30_Clay	$\text{Log}_e(30\_Clay)$	231	0.253	Log_30_CLAY
30_OrgC	$\text{Log}_e(30\_OrgC+1)$	230	0.930	Log_30_OrgC
30_ECEC	$\text{Log}_e(30\_ECEC+1)$	231	0.814	Log_30_ECEC
30_ESP	$\text{Log}_e(30\_ESP+1)$	230	0.935	Log_30_ESP
30_Ca_Mg		231	0.860	S30_Ca_Mg
CTI	$\text{Log}_e(\text{CTI})$	237	-0.085	LogCTI
Slope%_DEM	$\text{Log}_e(\text{Slope}\%DEM+10)$	250	0.167	LogSlope%DEM
RI_10m		206	-0.092	RI_10m
PredYield		181	0.1	PredYield
Yield 2002		125	0.5	Yield 2002
Yield 2003		114	0	Yield 2003

## 6.2.2 Correlation Analysis

Initially, all pairwise correlations were calculated between soil properties, topographic variables, the redness index (RI) and yield (Table 6.3) to examine possible variables related to yield.

Table 6.3 Correlation coefficients from analysis between yield and other variables

Variable	Potential Yield (n=132)		Actual Yield 2002 (n=98)		Actual Yield 2003 (n=85)	
	Correlation	Probability	Correlation	Probability	Correlation	Probability
Log A HORIZ	0.052	0.551	0.118	0.245	-0.020	0.855
Restrict Depth	-0.121	0.167	0.106	0.301	-0.224	*
S10 STONES %	0.167	0.055	0.078	0.445	0.237	*
Log 10 EC	0.002	0.979	-0.110	0.282	0.030	0.788
Log 10pHw	-0.045	0.608	-0.265	**	0.125	0.255
Log 10 pHca	-0.069	0.432	-0.277	**	0.105	0.337
Log 10 SAND% <sup>a</sup>	0.022	0.805	-0.298	**	0.323	**
Log 10 SILT	0.053	0.546	-0.329	**	0.279	**
Log 10 CLAY%	-0.006	0.942	-0.239	*	0.333	**
S10 OrgC %	-0.022	0.806	-0.328	**	0.169	0.122
Log 10 ECEC	-0.021	0.815	-0.354	**	0.238	*
Log 10 ESP	0.180	*	0.319	**	0.154	0.159
S10 Ca Mg	0.083	0.344	-0.045	0.659	-0.099	0.368
Log 30 STONES	0.148	0.091	-0.003	0.979	0.157	0.152
Log 30 EC	-0.085	0.332	-0.135	0.186	0.136	0.214
S30 pH H2O	-0.200	*	-0.072	0.482	0.090	0.415
S30 pH CaCl2	-0.171	0.052	-0.060	0.559	0.123	0.263
Log 30 SAND% <sup>a</sup>	-0.046	0.603	-0.243	*	0.184	0.092
Log 30 Silt	0.005	0.956	-0.355	**	0.309	**
Log 30 CLAY	-0.027	0.760	-0.188	0.064	0.145	0.185
Log 30 OrgC	-0.027	0.759	-0.232	*	0.183	0.094
Log 30 ECEC	-0.045	0.605	-0.201	*	0.147	0.178
Log 30 ESP	-0.207	*	-0.028	0.788	0.009	0.937
S30 Ca Mg	0.121	0.167	0.050	0.628	-0.033	0.768
LogCTI	-0.009	0.915	-0.052	0.609	0.173	0.114
LogSlope%DEM	0.148	0.091	0.148	0.146	0.066	0.549
RI 10m	0.117	0.181	0.063	0.538	0.031	0.776

\*\* : Significant correlations  $\leq 0.01$

\* : Significant correlations  $\leq 0.05$

a: An inverse transformation was performed on Sand and subsequently correlations that would normally be expected to be negative are positive and vice versa.

n: number of records

Given the large number of samples it is surprising that there are minimal significant correlations. Only three soil properties have a significant correlation with potential yield, although they are all weak ( $r \leq 0.207$ ). ESP at 10cm depth has a significant positive correlation with potential yield, whereas ESP at 30cm has a significant

negative correlation. pHw at 30cm depth also has a significant negative correlation with potential yield.

Correlations with actual yield in 2002 and 2003 show more promising results. In 2002 at the 10cm depth, properties pH (water and Ca), sand, silt, clay, organic carbon and ECEC are all significantly negatively correlated with actual yield, while ESP has a significant positive correlation. At a depth of 30cm, sand, silt, organic carbon and ECEC have significant negative correlations. In 2003 actual yield has a significant negative correlation with restricted depth and significant positive correlations with stones, sand, silt, clay and ECEC at 10cm depth. At the 30cm depth a significant positive correlation with silt is evident.

In summary, there are significant correlations between actual yield and soil texture, organic matter and capacity of the soil to store nutrient cations at both 10cm and 30cm depths, and between actual yield and restricted depth, coarse fragments, pH and ESP at the 10cm depth, however they are all weak ( $r \leq 0.355$ ). The number of significant correlations between soil properties and yield are not only fewer than expected, but the nature of the correlations varies between 2002 and 2003. Perhaps this can be attributed to the large difference in the total rainfall between those years (177mm) (Figure 3.4). It is notable that the topographic variables or RI are not significantly correlated with any of the yield data sets.

### 6.2.3 Associations between Yield and Landforms

A simple analysis of variance was performed to examine the effect of landform on potential yield. The statistical assumptions underlying analysis of variance were examined using plots of residuals. The effect of landform was subdivided into the difference between primary landforms (Crests, Primary slopes, Flats and Depressions) and the difference between slope classes (Upper slope, Mid slope, Simple slope and Lower slope). This subdivision of landform effects is appropriate because the slope classes (US, MS, SS, LS) are a further subdivision of existing simple slopes, which have been labelled primary slopes in this case (Section 5. 2).

There was no effect of primary landforms ( $P=0.514$ ) or slope classes ( $P=0.635$ ) in 2003 but there were significant differences in 2002 ( $P=0.045$  and  $P=0.054$ ,

respectively) when rainfall was less. It appears that crop growth was limited in some landforms in 2002 due to low rainfall, as average yields for primary slopes and crests in 2002 are higher than average yields for flats and depressions (Table 6.4). The potential yield, which is an average over several years (maximum 7 years), showed a slight effect of primary landforms ( $P=0.067$ ) with crests and primary slopes having higher average yields than flats.

Table 6.4 ANOVA for Yield against Landforms

Variable		LANDFORMS								Significance of Primary Landforms	Significance of Slope Classes
		C	US	MS	SS	LS	PS	F	D		
Potential Yield kg/ha	Mean	1661	1553	1581	1657	1541	1574	1307	1501	0.067	0.784
	N	15	23	46	17	37	123	14	29		
	S.Error	102	83	59	96	65	36	106	73		
Yield 2002 kg/ha	Mean	1441	1198	1667	1391	1418	1468	1063	1111	0.045	0.054
	N	12	19	35	9	22	85	10	18		
	S.Error	178	139	102	201	129	67	195	145		
Yield 2003 kg/ha	Mean	2128	2284	2113	2483	2158	2190	2693	2399	0.514	0.635
	N	10	13	35	8	27	83	3	18		
	S.Error	244	216	131	275	150	85	446	182		

Crest (C); Upper slope (US); Mid slope (MS); Simple slope (SS); Lower slope (LS)  
 Primary slope (PS) = (US+MS+SS+LS); Flat (F); Depression (D).

#### 6.2.4 Multiple Regression between Yield and Soil Properties

Many of the soil properties recorded are correlated and the pairwise correlations examined above, while identifying individual variables related to yield, do not identify a group of variables for use in forming LMUs. Therefore, further analysis through multiple regression was performed in an endeavour to choose a group of soil properties that are influencing final yield.

Backward elimination multiple regression was performed using yield as a response variate and the soil properties as explanatory variables. In addition, a factor characterising crop type (wheat, barley, canola, lupins, oats and oaten hay) was included as an explanatory factor. The pH<sub>w</sub> at 10cm and 30cm depths was not included in the models, as it is highly correlated with pH<sub>ca</sub>. The percentage of silt at 10cm and 30cm depth was also not included as silt percent, sand percent and clay percent sum to 100 percent. In backward elimination the full model, including all explanatory variables, is fitted initially. Terms are then progressively dropped according to the value of the variance ratio corresponding to the change in the regression mean square when a particular term is dropped. Terms giving the smallest

variance ratio are dropped first. The process continues until the dropping of any term gives a variance ratio greater than a specified value; 2 in this case. Statistical assumptions underlying regression analysis were examined using plots of residuals. The significance, percentage of variance accounted for and residual standard deviation are shown for the full model and final model in each case, as well as, the significance of the terms remaining in the final model (Table 6.5).

Table 6.5 Regression results from analysis between yield and soil properties

Variable	Potential Yield Significance of terms remaining in Model	Actual Yield 2002 Significance of terms remaining in Model	Actual Yield 2003 Significance of terms remaining in Model	
Log A HORIZ			0.055	
Restrict Depth				
S10 STONES %	0.035	0.019	0.007	
Log 10 EC		0.001		
Log 10 pHca			0.006	
Log 10 SAND%				
Log 10 CLAY%				
S10 OrgC %			0.007	
Log 10 ECEC			<0.001	
Log 10 ESP	<0.001	<0.001		
S10 Ca Mg	0.060		0.085	
Log 30 STONES				
Log 30 EC				
S30 pH CaCl2			0.006	
Log 30 SAND%		0.103		
Log 30 CLAY		0.160	0.002	
Log 30 OrgC				
Log 30 ECEC			0.001	
Log 30 ESP	<0.001	0.008		
S30 Ca Mg	0.128		0.078	
Crop Type ( <i>Factor</i> )	Not Included	<0.001	<0.001	
Full Model	Sig	0.012	<0.001	<0.001
	Variance %	12.2%	40.1%	52.6%
	Residual SD	373	485	539
Final Model	Sig	<0.001	<0.001	<0.001
	Variance %	16.3%	47.2%	55.9%
	Residual SD	364	456	520

Although only a low percentage of variance in yield has been accounted for, especially for potential yield <20 percent, these regressions indicate groups of soil properties that are associated with yield. Soil properties that had a significant effect on potential yield or actual yield in 2002 or 2003 include: stones, EC, pHca, OrgC%, ECEC and ESP at 10cm depth, and pHca, clay, ECEC and ESP at 30cm depth.

### 6.2.5 Discussion of Results

It is not within the scope of this thesis to analyse the strengths and weaknesses of various soil measurements with respect to land management. Rather, given a comprehensive set of measurements, the intention is to examine methods for choosing a subset of those variables and choose a subset appropriate for this particular study area.

The small percentage of variance in yield, explained by the multiple regressions incorporating a wide range of soil properties and crop types, highlights that other variables not considered have a significant influence on final yield. In dryland agriculture, rainfall is an overriding, uncontrollable limiting factor on productivity. However, multiple regressions for potential yield, which is effectively adjusted for rainfall since year effects have been removed (see Section 5.7.2), explain considerably less variance than those for yield in 2002 and 2003. Factors, such as weed and insect infestation, frost, fertilizer application and management decisions, are among those not considered in this analysis which has a substantial effect on yield.

In Western Australia, texture groups derived from particle size distribution are based largely on clay content (Purdie, 1998). Kramer and Boyer (1995) mention that, in general, larger clay content increases the storage capacity of soils for water and minerals (cation exchange capacity) but decreases aeration which is essential for good root growth and functioning. In turn, the proportion of clay in the soil is crucial in determining the suitability of a soil for plant growth (Kramer and Boyer, 1995).

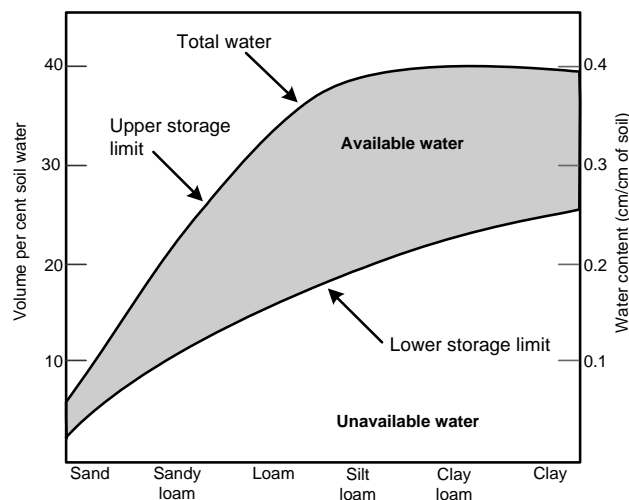


Figure 6.1 The relative amounts of water available and unavailable for plant growth in soils with textures from sand to clay (redrawn from Kramer 1983, as shown in Kramer and Boyer, 1995)

If factors such as the interception of plant canopy, evaporation, transpiration, deep drainage, surface runoff and lateral water movement are ignored, and instead the relationship between available water and texture (Figure 6.1) are viewed; it can be seen that as the percentage of clay increases, the available water to the plant increases (with the exclusion of heavy clays) leading to positive correlation between yield and percentage of clay in years when rainfall is not limiting. However, as the percentage of clay increases, the lower storage limit also increases. In years of low rainfall the lower storage limit might not be reached in soils with large clay content, and thus, water would not be available to the plant leading to a negative correlation between yield and percentage of clay.

Not only was rainfall below average in 2002 but the two years prior were in fact limiting in terms of water availability to plant growth (Figure 3.3 and Figure 3.4) There was a lack of water penetrating the soil over several years and subsequently, it would take a considerable amount of rain to fill the lower storage limit of soils. Rainfall in 2003 was average for the study area. This helps to explain why correlations between clay content and yield have been weaker than expected and also alternate, from negative in 2002 to positive in 2003.



Other soil properties show a similar pattern of correlations to yield in 2002 and 2003, to the percentage of clay, because of their correlation with percentage of clay. Log<sub>10</sub>\_SAND%, Log<sub>10</sub>\_Silt and Log<sub>10</sub>\_ECEC are highly correlated with Log<sub>10</sub>\_CLAY% (0.958, 0.797 and 0.793, respectively) while Log<sub>10</sub>\_pHw, Log<sub>10</sub>\_pHca and S10\_OrgC% have lower correlations with Log<sub>10</sub>\_CLAY% (0.287, 0.306 and 0.422, respectively) (Appendix M).

In summary, the results from the statistical analysis between yield and potential predictors for LMUs does not provide thorough evidence to determine which variables should be used in the classification. As such, principal components analysis was undertaken and is explained hereafter.

### **6.3 Selecting Variables Using Principal Component Analysis**

The next step in determining important variables to include in defining LMUs was to perform principal components analysis. This analysis enabled the identification of variables that explain variance between sampling points and to examine relationships between variables in the multivariate data set. In a principal components analysis, the first principal component (PC1) is the linear combination of the variables and accounts for the largest percentage of variance between sampling points. The second principal component (PC2) is independent of PC1 and is the linear combination of the variables and accounts for the next largest percentage of variance between the sampling points and so on.

The biplot is a graph of PC1 versus PC2 (or other pairs of components) and shows both the sampling points and the original variables as vectors. It is the two dimensional graph that shows the separation between sampling points best and is most useful when the first two principal components account for a large proportion of variation between sampling points. The biplot can be used to examine the relationships between variables, as well as the relative position of the sampling points. In this case, biplots have been used for exploratory data analysis to indicate the level of redundancy in the data set (i.e. variables that are highly correlated) and assist with the selection of appropriate variables for subsequent analysis. The principal components were calculated and the biplots were drawn using GenStat®.

Table 6.6 Percentage of variance explained by each principal component and loadings of variables for each principal component from the 250 soil points

	PC1	PC2	PC3	PC4	PC5
<b>Percentage</b>	37.95	11.82	8.86	5.87	5.58
<b>Cumulative</b>	37.95	49.77	58.63	64.50	70.08
<b>Variable</b>	<b>Loadings</b>				
Log 10 SAND% <sup>a</sup>	-0.300	-0.057	-0.093	0.099	-0.003
Log 30 SAND% <sup>a</sup>	-0.282	0.012	-0.123	0.206	0.003
Log 10 SILT	-0.277	-0.091	-0.100	0.056	-0.030
Log 30 Silt	-0.233	-0.045	-0.104	0.079	-0.139
Log 10 CLAY%	-0.289	-0.022	-0.086	0.130	0.024
Log 30 CLAY	-0.267	0.008	-0.116	0.211	0.042
Log 10 ECEC	-0.292	-0.112	0.172	-0.054	0.090
Log 30 ECEC	-0.296	0.001	-0.086	0.204	-0.069
S10 OrgC %	-0.195	-0.061	0.055	-0.191	0.324
Log 30 OrgC	-0.226	-0.241	-0.061	0.130	0.058
S10 Ca Mg	0.147	-0.317	0.097	-0.049	0.000
S30 Ca Mg	0.165	-0.350	0.029	-0.084	0.031
Log 10 ESP	0.006	0.378	-0.166	-0.219	0.142
Log 30 ESP	-0.035	0.480	-0.012	-0.238	0.058
Log 10 EC	-0.183	-0.099	0.121	-0.322	0.460
Log 30 EC	-0.192	0.089	0.061	-0.256	0.171
Log 10pHw	-0.147	-0.075	0.532	-0.091	-0.148
S30 pH H2O	-0.205	0.319	0.084	-0.060	-0.212
Log 10 pHca	-0.155	-0.103	0.540	-0.116	-0.004
S30 pH CaCl2	-0.201	0.266	0.109	-0.111	-0.196
S10 STONES %	-0.112	-0.199	-0.328	-0.291	0.042
Log 30 STONES	-0.035	-0.157	-0.276	-0.542	-0.262
Log A HORIZ	0.056	0.044	-0.118	0.135	0.608
Restrict Depth	0.134	0.203	0.206	0.247	0.214

a: An inverse transformation was performed on Sand and subsequently loadings that would normally be expected to be positive are negative and vice versa.

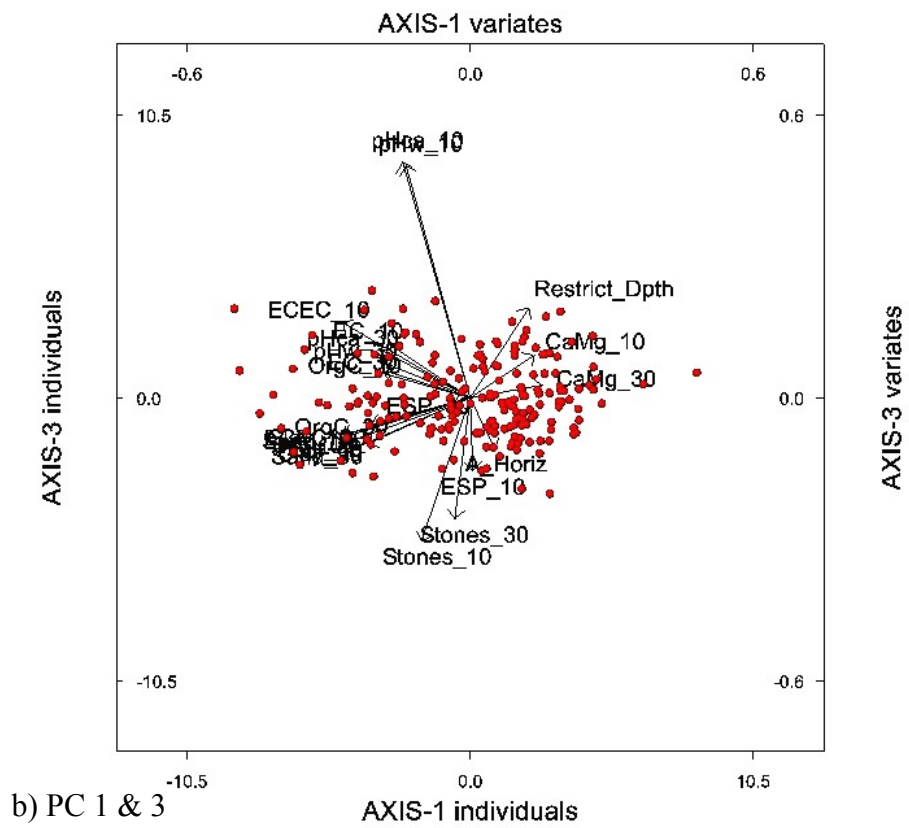
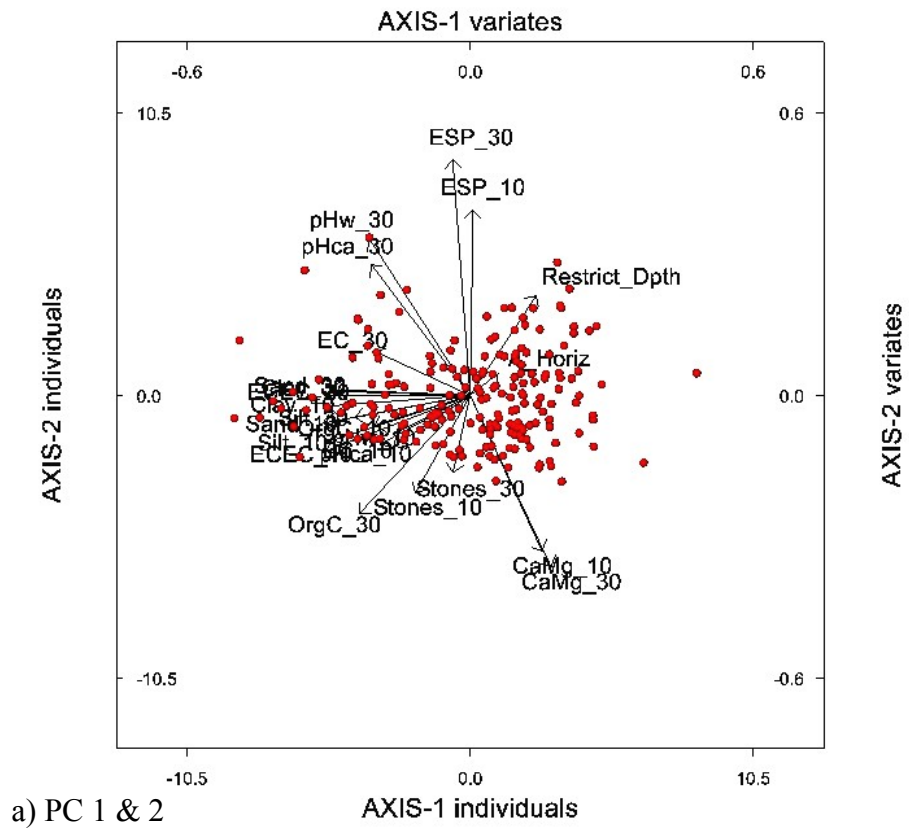


Figure 6.2 Biplots based on 250 soil points and the stable soil variables.  
a) PC1 vs PC2; b) PC1 vs PC3

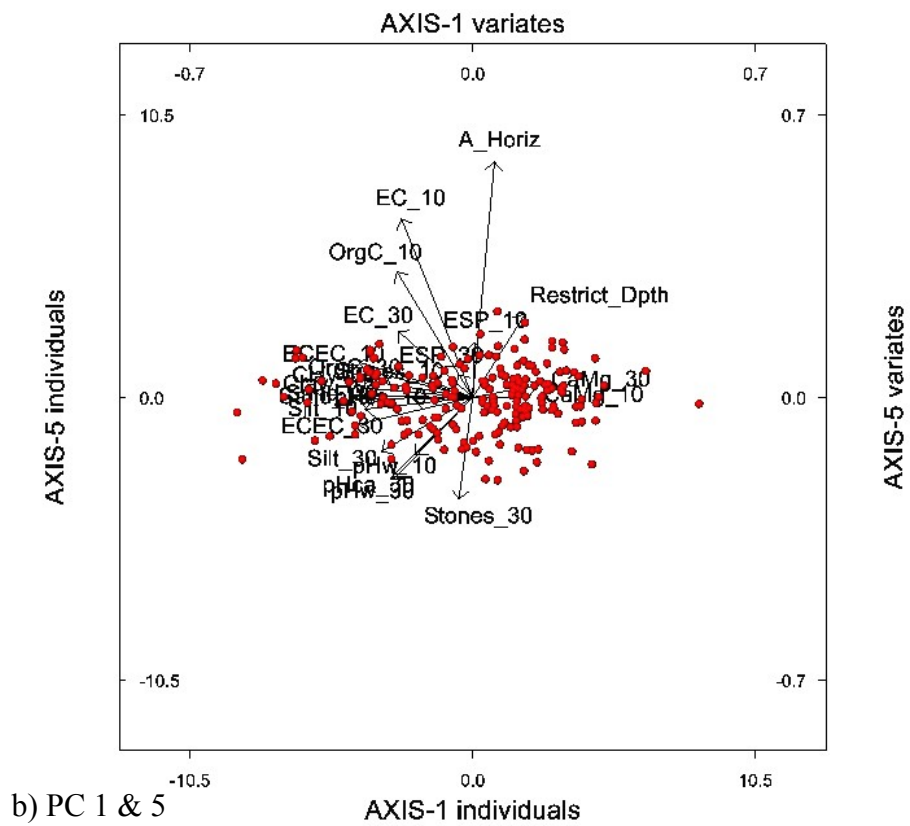
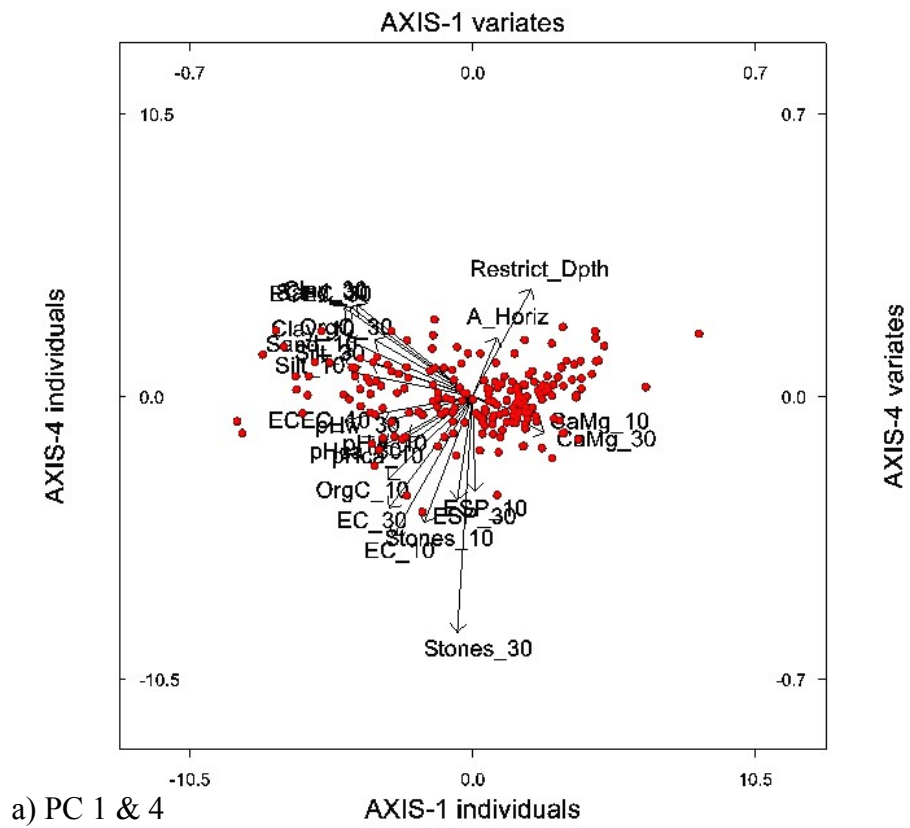


Figure 6.3 Biplots based on 250 soil points and the stable soil variables.  
a) PC1 vs PC4; b) PC1 vs PC5

The first five principal components, based on the correlation between all measurements from 10cm and 30cm depths at all 250 soil points (Appendix M), account for 70.08 percent of the total variance between soil sampling points (Table 6.6). PC1 (Table 6.6 and Figure 6.2) is associated with texture (Sand%, Silt%, Clay%) and ECEC for the 10cm and 30cm depths. ESP at 10cm and 30cm depths, dominates PC2 along with the Ca:Mg ratio (10cm and 30cm) and to a lesser extent, pH<sub>w</sub> and pH<sub>ca</sub> at 30cm. PC3 is loaded with pH<sub>w</sub> and pH<sub>ca</sub> at 10cm and the percentage of stones at 10cm and 30cm. The percentage of stones at 10cm and 30cm along with restricted depth and EC at 10cm and 30cm, contributed to PC4; while PC5 is loaded with A-horizon, EC and OrgC at 10cm.

Table 6.7 Correlation of texture and ECEC

<b>Variables<sup>a</sup></b>	<b>log 10_Sand</b>	<b>log 10_Silt</b>	<b>log 10_Clay</b>	<b>log 10_ECEC</b>	<b>log 30_Sand</b>	<b>log 30_Silt</b>	<b>log 30_Clay</b>	<b>log 30 ECEC</b>
<b>Log10_Sand<sup>b</sup></b>	1							
<b>Log10_Silt</b>	0.935	1						
<b>Log10_Clay</b>	0.958	0.797	1					
<b>Log10_ECEC</b>	0.833	0.782	0.793	1				
<b>Log30_Sand<sup>b</sup></b>	0.721	0.636	0.725	0.603	1			
<b>Log30_Silt</b>	0.681	0.717	0.577	0.580	0.613	1		
<b>Log30_Clay</b>	0.675	0.569	0.705	0.566	0.979	0.471	1	
<b>Log30_ECEC</b>	0.779	0.701	0.771	0.709	0.888	0.629	0.852	1

a: Shaded variables have been removed from further analysis

b: An inverse transformation was performed on Sand and subsequently correlations that would normally be expected to be negative are positive and vice versa.

The biplots and principal component loadings indicate that ECEC and texture variables are correlated with one another. As explained in the literature review Section 2.2.1.1, soil texture is described by the percentage of sand, silt and clay sized particles in each soil sample. These variables are correlated with one another (Table 6.7) because they sum to 100 percent. Sand at 10cm is very highly correlated with silt, clay and ECEC at 10cm and ECEC and sand at 30cm. The sand at 30cm is also very highly correlated with clay and ECEC at 30cm. Some of these variables can be eliminated without removing any information about the differences between sampling points.

As mentioned, texture groups from particle size distribution in Western Australia are based largely on clay content, so for subsequent analysis of results it would be useful to retain the clay content. Therefore, the sand variable was removed for both depths.

After excluding sand at 10cm and 30cm, many of the remaining stronger correlations in Table 6.7 involve ECEC at 10cm or 30cm. This indicates that at least one of these variables may be eliminated. ECEC at 30cm was retained while ECEC at 10cm was removed. This is because; as ECEC relates to the capacity of the soil to store nutrients it was thought this was more important at root depth (i.e. 30cm). The percentage of silt particles at 10cm and 30cm depths are also highly correlated, as such only silt at 10cm was retained. Of the variables retained, ECEC at 30cm provides an indication of the soils capacity to store nutrients, whilst clay content at 10cm and 30cm and silt content at 10cm, provide a texture component at both depths.

Other variables that are correlated (i.e. correlation coefficient greater than  $r = 0.500$ ) are; OrgC at 10cm with EC ( $r=0.603$ ) and silt ( $r=0.553$ ) at 10cm; EC at 30cm with sand ( $r=0.507$ ), clay ( $r=0.511$ ) and ECEC ( $r=0.510$ ) at 30cm; pHca at 30cm with ECEC at 30cm ( $r=0.539$ ); OrgC at 30cm with silt ( $r=0.545$ ) and clay ( $r=0.594$ ) at 10cm and clay ( $r=0.594$ ) and ECEC ( $r=0.680$ ) at 30cm; ESP at 30cm with ESP at 10cm ( $r=0.525$ ) and Ca:Mg at 30cm ( $r=-0.522$ ).

As indicated previously the OrgC at 30cm is correlated with silt and with clay at 10cm and clay and ECEC at 30cm. It also has moderate correlation with OrgC ( $r=0.437$ ), EC ( $r=0.390$ ), stones ( $r=0.354$ ) and pHca ( $r=0.301$ ) at 10cm and restricted depth ( $r=-0.363$ ). OrgC at 30cm does not dominate any of the PC's while OrgC at 10cm is dominant in PC5. Expectedly, the amount of OrgC is much higher at a depth of 10cm than at 30cm (median of 1.12 vs 0.26, respectively) as it consists of plants and animals at various stages of decomposition.

Significant correlations exist between actual yield in 2002 and OrgC at 10cm ( $r=-0.328$ ) and OrgC at 30cm ( $r=-0.232$ ) (Table 6.3). OrgC at 10cm has a slightly stronger correlation with actual yield 2002, it is dominant in PC5 and is not correlated with as many other soil properties as OrgC at 30cm. As a consequence, OrgC at 30cm was removed.

The Ca:Mg ratio at both 10cm and 30cm are dominant in PC2 (Table 6.6). As Ca:Mg at 10cm and 30cm have a mixture of moderate correlation with other variables, only Ca:Mg at 30cm was retained.

ESP at 30cm is moderately correlated with ESP at 10cm ( $r=0.525$ ) and Ca:Mg at 30cm ( $r=-0.522$ ). ESP at 10cm and 30cm are dominant in PC2 with ESP at 30cm being the most dominant. Although both weak, ESP at 30cm is a less weak correlation with potential yield ( $r=-0.207$ ) than ESP at 10cm ( $r=0.180$ ). In view of the summary statistics of the ESP levels against indicators of sodic soils (Section 2.2.1.7), 98.5 percent of the samples at 10cm are non sodic with the remaining 1.5 percent being sodic, while ESP at 30cm has 85 percent of samples non sodic, 12 percent sodic and 3 percent highly sodic. Thus, ESP at 30cm provides a better indication of levels that may affect yield and it was therefore retained, while ESP at 10cm was eliminated.

The evaluation of the summary statistics for EC at 10cm and 30cm and the relationship of these values to effects on plant growth show that, 97 percent of samples from 10cm and 30cm are non saline, and the remaining 3 percent only slightly saline (i.e. minimal effect on plant growth or only very sensitive plants affected). There is only one outlier in the 30cm depth which is highly saline.

This indicates that soil salinity is currently not a land degradation issue at Muresk. As the levels of EC are not affecting plant growth, both variables could be excluded from the analysis. However, EC at 30cm was retained because its high loading in PC4 indicates that it can assist in explaining the variability in the data set. Furthermore, subsurface salinity is a good indicator of the possibility of surface salinity occurring in the future. Neither EC at 10cm or 30cm are significantly correlated with potential yield. Therefore, EC at 10cm was eliminated.

The biplots and correlation matrix indicate that pHw and pHca are very highly correlated at both 10cm and 30cm (0.960 and 0.956, respectively). pHca represents field conditions better than when pH is measured in water (Section 2.2.1.6) and consequently, pHw was removed.

Examining the correlations of pHca and other soil properties, pHca at 10cm has significant correlations with clay at 10cm ( $r=0.306$ ), EC at 30cm ( $r=0.335$ ), while pHca at 30cm is correlated with silt at 10cm ( $r=0.424$ ), clay at 10cm ( $r=0.431$ ), OrgC at 10cm ( $r=0.308$ ), EC at 30cm ( $r=0.451$ ), clay at 30cm ( $r=0.352$ ), ECEC at 30cm ( $r=0.539$ ), ESP at 30cm ( $r=0.322$ ) and Ca:Mg at 30cm ( $r=-0.389$ ). pHca at 30cm is dominant in PC2 and less so in PC1, while pHca at 10cm is dominant in PC3.

In relation to the amount of pHca that can cause restriction to plant growth, soils with a pHca < 4.8 are acidic and may have associated aluminium toxicity (Section 2.2.1.6) while soils that are alkaline have a pHw >7.5, and is equivalent to pHca of approximately > 6.5 (Figure 6.4) which can lead to sodicity (Section 2.2.1.7). Only 2 percent of the soils at 10cm, and 7 percent of the soils at 30cm are alkaline, while 34 percent at 10cm and 9 percent at 30cm are acidic. Given the dominance of pHca at both 10cm and 30cm depths in the principal components and the percentage of samples that indicate acidity, it would be useful to include both variables in the LMU classification.

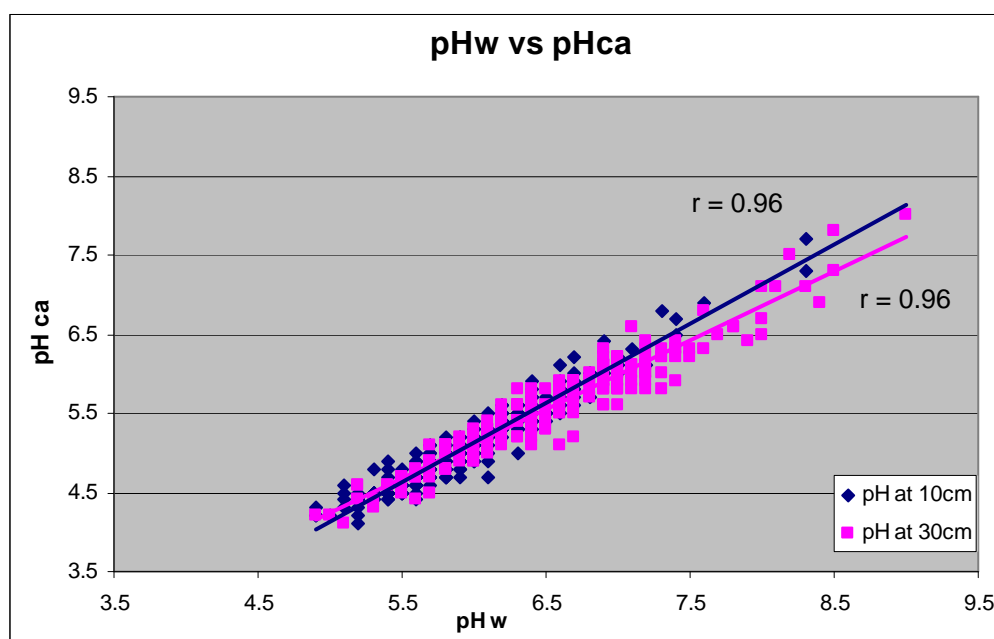


Figure 6.4 Relationship between pHw and pHca for 10 and 30cm depths. Linear best fit regression line.

The percentage of stones at 10cm and 30cm have high loadings in PC3 and PC4, with stones at 10cm being slightly greater for PC3 and stones at 30cm dominating PC4. There is a significant positive correlation between the two stone variables



( $r=0.432$ ) and between stones at 10cm and restricted depth ( $r=-0.385$ ). This indicates that shallower soils have a higher percentage of stones. In terms of correlations with potential yield, the stone variables were very close to significant with a correlation coefficient of 0.167 for stones at 10cm and 0.148 with stones at 30cm (Table 6.3). As discussed in Section 2.2.1.2, gravely soils or soils with a high percentage of stones, can be superior as they can have improved drainage and clayey subsoils, providing water storage. Stones at 10cm have a stronger correlation with potential yield and significant correlation with actual yield in 2003 (Table 6.3). This provides useful information about sample points. However, stones at 30cm are eliminated from further analysis because they have weaker correlations to yield.

Thirteen soil variables have been chosen to define LMUs on the basis of their contribution to the principal components, and their correlations with other soil variables, yield and properties that affect plant growth (Table 6.8).

Table 6.8 List of 13 soil variables to be included in LMU model

Variable
Log A Horizon
Restricted Depth (cm)
S10 STONES %
Log 10 pHca
Log 10 SILT
Log 10 CLAY%
S10 OrgC
Log 30 EC
S30 pHca
Log 30 CLAY
Log 30 ECEC
Log 30 ESP
S30 Ca Mg

#### 6.4 Relationships between Selected Soil properties and ancillary data

It is only sensible to define high resolution boundaries for LMUs using the ancillary data, if the ancillary variables are associated with stable soil properties. Therefore, correlation analysis was used to examine associations between continuous ancillary variables and soil properties, while analysis of variance was used to examine the association between categorical ancillary data and soil properties. This analysis will identify if the ancillary variables are appropriate inputs to the LMU classification.

### 6.4.1 Correlation Analysis

Correlation analysis was performed between the 13 selected soil properties and the continuous ancillary data (Table 6.9). Correlations between RI and soil properties measured at 30cm depth were not performed as DMSI properties relate to surface reflectance only.

Table 6.9 Correlation between 13 selected soil properties and ancillary variables

Variable	CTI <i>n= 193</i>		Slope% <i>n=205</i>		RI <i>n=169</i>	
	Correlation	Probability	Correlation	Probability	Correlation	Probability
Log A_HORIZ	-0.023	0.752	-0.193	**	0.263	**
Restricted Depth (cm)	0.207	**	-0.006	0.928	-0.071	0.362
S10 STONES %	-0.175	*	0.086	0.220	0.244	**
Log_10_pHca	-0.032	0.658	-0.023	0.742	-0.273	**
Log_10_SILT	0.031	0.664	-0.092	0.188	-0.054	0.484
Log_10_CLAY%	0.139	0.054	-0.178	*	-0.132	0.086
S10_OrgC %	0.098	0.173	-0.064	0.360	-0.156	*
Log_30_EC	0.035	0.626	-0.183	**	-0.028	0.720
S30_pH_CaCl2	0.102	0.159	-0.058	0.410		
Log_30_CLAY	-0.017	0.813	-0.052	0.458		
Log_30_ECEC	-0.058	0.425	-0.027	0.698		
Log_30_ESP	0.352	**	-0.299	**		
S30_Ca_Mg	-0.311	**	0.199	**		

\*\* : Significant correlations  $\leq 0.01$

\* : Significant correlations  $\leq 0.05$

All ancillary variables are correlated with some soil properties but not all soil properties are associated with an ancillary variable. It is notable that of the texture variables and the associated ECEC, only clay at 10cm is correlated with an ancillary variable (Log\_10\_CLAY% vs Slope%:  $r=-0.178$ ;  $P<0.05$ ).

### 6.4.2 Analysis of Variance with Landforms

An analysis of variance was carried out to determine whether there were differences between landforms for each of the selected properties. The effect of landform was subdivided into the difference between primary landforms (Crests, Primary slopes, Flats and Depressions) and the difference between slope classes (Upper slope, Mid slope, Simple slope and Lower slope).

Table 6.10 ANOVA for Soil Properties against Landforms

Variable	LANDFORMS									Sig. of Primary Landform	Sig. of Slope Classes
	C	US	MS	SS	LS	PS	F	D			
Log A-Horizon	Mean <sup>L</sup>	1.970	2.089	2.101	1.913	2.050	2.059	2.409	2.012	0.386	0.852
	Mean	6.2	7.1	7.2	5.8	6.7	6.8	10.1	6.5		
	N	28	31	60	21	50	162	16	40		
	S.Error	0.162	0.155	0.111	0.188	0.122	0.067	0.215	0.136		
Restricted Depth (cm)	Mean	25.54	22.66	34.33	28.57	32.72	30.56	43.33	37.71	0.002	0.017
	N	28	32	52	21	46	151	15	35		
	S.Error	3.27	3.01	2.36	3.72	2.51	1.41	4.47	2.93		
10 STONES	Mean	7.893	6.676	6.500	9.714	8.700	7.612	4.765	5.625	<0.001	<0.001
	N	28	34	60	21	50	165	17	40		
	S.Error	0.727	0.640	0.482	0.814	0.528	0.299	0.933	0.608		
Log 10_pHca	Mean <sup>L</sup>	1.841	1.785	1.795	1.815	1.786	1.793	1.810	1.799	0.062	0.586
	Mean	5.30	4.96	5.02	5.14	4.97	5.01	5.11	5.04		
	N	28	34	60	21	50	165	17	40		
	S.Error	0.017	0.015	0.011	0.019	0.013	0.007	0.021	0.014		
Log 10_SILT	Mean <sup>L</sup>	2.073	1.882	1.753	2.121	2.030	1.910	1.855	2.018	0.097	<0.001
	Mean	7.95	6.57	5.77	8.34	7.61	6.75	6.39	7.52		
	N	28	34	60	21	50	165	17	40		
	S.Error	0.077	0.068	0.051	0.086	0.056	0.032	0.099	0.065		
Log 10 CLAY	Mean <sup>L</sup>	1.825	1.701	1.679	2.015	1.959	1.810	1.749	1.830	0.946	0.002
	Mean	6.20	5.48	5.36	7.50	7.09	6.11	5.75	6.23		
	N	28	34	60	21	49	164	17	40		
	S.Error	0.092	0.082	0.062	0.104	0.068	0.038	0.118	0.077		
10_OrgC	Mean	1.210	1.243	1.195	1.158	1.216	1.206	1.171	1.207	0.992	0.913
	N	28	34	60	21	50	165	17	40		
	S.Error	0.085	0.078	0.058	0.099	0.064	0.035	0.109	0.071		
Log 30_EC	Mean <sup>L</sup>	1.853	1.559	1.477	1.733	1.644	1.577	1.569	1.625	0.056	0.117
	Mean	5.4	3.8	3.4	4.7	4.2	3.8	3.8	4.1		
	N	25	29	57	21	44	151	17	37		
	S.Error	0.094	0.087	0.062	0.102	0.070	0.038	0.114	0.077		
30 pH CaCl2	Mean	5.716	5.403	5.423	5.819	5.641	5.538	5.418	5.784	0.068	0.032
	N	25	29	57	21	44	151	17	38		
	S.Error	0.125	0.115	0.082	0.135	0.093	0.051	0.152	0.101		
Log 30 CLAY	Mean <sup>L</sup>	2.554	2.382	2.281	2.673	2.601	2.448	2.128	2.196	0.137	0.133
	Mean	12.86	10.83	9.79	14.48	13.48	11.57	8.40	8.99		
	N	25	29	57	21	44	151	17	38		
	S.Error	0.165	0.152	0.109	0.179	0.124	0.067	0.200	0.134		
Log 30_ECEC	Mean <sup>L</sup>	2.148	1.753	1.631	2.155	1.942	1.818	1.455	1.778	0.020	0.017
	Mean	7.568	4.772	4.109	7.628	5.973	5.160	3.284	4.918		
	N	25	29	57	21	44	151	17	38		
	S.Error	0.144	0.131	0.094	0.154	0.107	0.059	0.174	0.117		
Log 30_ESP	Mean <sup>L</sup>	1.081	1.174	1.368	1.109	1.513	1.337	1.622	1.501	0.005	0.014
	Mean	1.948	2.235	2.927	2.031	3.540	2.808	4.063	3.486		
	N	25	29	57	21	44	151	17	37		
	S.Error	0.112	0.102	0.073	0.120	0.083	0.046	0.136	0.092		
30 Ca_Mg	Mean	4.440	5.168	3.908	4.559	3.182	4.029	3.231	0.326	0.103	0.003
	N	25	29	57	21	44	151	17	38		
	S.Error	0.472	0.428	0.305	0.503	0.347	0.192	0.573	0.383		
		<b>C</b>	<b>US</b>	<b>MS</b>	<b>SS</b>	<b>LS</b>	<b>PS</b>	<b>F</b>	<b>D</b>		

Crest (C); Upper slope (US); Mid slope (MS); Simple slope (SS); Lower slope (LS)

Primary slope (PS) = (US+MS+SS+LS); Flat (F); Depression (D).

<sup>L</sup>= Log Mean. Sig.=Significance.

Nine out of the 13 selected soil properties show significant differences between either primary landforms, slope classes or both. Although they were significant in the general the percentage of variance accounted for was <10 percent. In particular,

silt and clay at 10cm and ECEC at 30cm show significant differences between landforms. Upper slopes and mid slopes have lower silt and clay at 10cm and ECEC at 30cm, than lower slopes and simple slopes. This indicates that landforms contain information about soil properties and will be a useful ancillary data set for input into the LMU classification.

## **6.5 Summary**

From the statistical analysis performed in this Chapter, 13 soil properties (from a possible set of 24) have been selected as inputs into the LMU classification. This was achieved by analysing their correlation with yield data, principal components analysis and the properties effects on plant growth. A particular soil property was eliminated if it could be shown that it was highly correlated with, and/or its effect on plant growth was already accounted for by other soil properties retained. This method for selecting soil properties is a practical approach which can be adopted by others. However, the resultant selected soil properties in this research are specific to this study area. Other properties that will be used to further define the LMU boundaries for the remainder of the study area are landform, CTI, slope percentage and RI, as the correlation and analysis of variance confirms that it is valid to use these ancillary data to define higher resolution boundaries. The LMU classification using the selected inputs is applied in the following Chapter 7.

## CHAPTER 7

### A FRAMEWORK FOR CREATING LAND MANAGEMENT UNITS

*The framework for creating LMUs is described and implemented in Chapter 7. The processes of selecting the most appropriate variables for input to the LMU classification and classification technique have been described in the preceding Chapters. The 13 stable soil properties collected at the 250 soil points, topographic attributes (landform, CTI and slope) and DMSI (RI) are the selected variables for deriving LMUs across Muresk farm. The process uses a two stage methodology based on Oliver and Webster's (1989) spatially weighted multivariate classification. The methodology initially classifies soil sampling points into LMUs based on a geographically weighted similarity matrix. The second stage delineates higher resolution LMU boundaries by using the geographic location, topographic attributes and DMSI on a 10m grid across the remaining study area, and assigning each pixel to an appropriate LMU. The method groups sample points and pixels with respect to their variables and their spatial relationship on the ground, thus forming contiguous, homogenous LMUs. A LMU map with associated soil description is provided, highlighting the use of discriminant analysis based on a selection of soil properties for identifying the soil properties that differentiate the LMUs and ranking them in terms of their similarity.*

#### **7.1 The Land Management Unit Classification**

The methodology used to form LMUs is an extension of Oliver and Webster's (1989) spatially weighted multivariate classification from the family of spatially constrained classification techniques. These techniques group samples based on the values of multiple variables, as well as, their spatial position. This forces the groups to be homogenous in terms of their variables' values and geographically contiguous (Urban, 2004). The techniques are based on a similarity matrix with dimension  $N^2$  ( $N$ =number of sample points). When classifying high resolution data sets, such as DMSI or DEM data at a resolution of 10m, for a study area of only 2000ha this produces a matrix the size of  $200,000 \times 200,000$ . Urban (2004) mentions that this has led to the development of specialised supervised and unsupervised classification methods for these data sets, and suggests that spatially constrained classification

shows some promise in this regard, but is not yet computationally feasible for large datasets.

The LMU classification consists of two distinct sections: Part A, which uses Oliver and Webster's (1989) spatially weighted cluster analysis to subdivide the 250 soil points into LMUs based on the stable soil properties; and Part B, which uses an additional spatially weighted multivariate analysis to further define the LMU boundaries using high resolution ancillary data. As such, the two stage methodology is an evolutionary coupling of spatially constrained classifications designed to accommodate large datasets.

In both sections the usual multivariate methods have been modified by adjusting dissimilarities between points according to the distance between the points and the spatial structure of the input variables. At several stages the classification offers the opportunity of adjustments appropriate to the study area, including the choice of input variables and their spatial structure. The LMU classification is described in the following sections according to the flowchart presented in Figure 7.1.

When dealing with large data sets i.e. high resolution remote sensing data (>100,000 individuals), the amounts of computer time and memory required to form LMUs are prohibitive for current personal computers. The process described in this thesis minimises the resources required by compartmentalising sub-processes wherever possible and taking advantage of commands available in GenStat®. Similar commands will be found in other statistical packages.

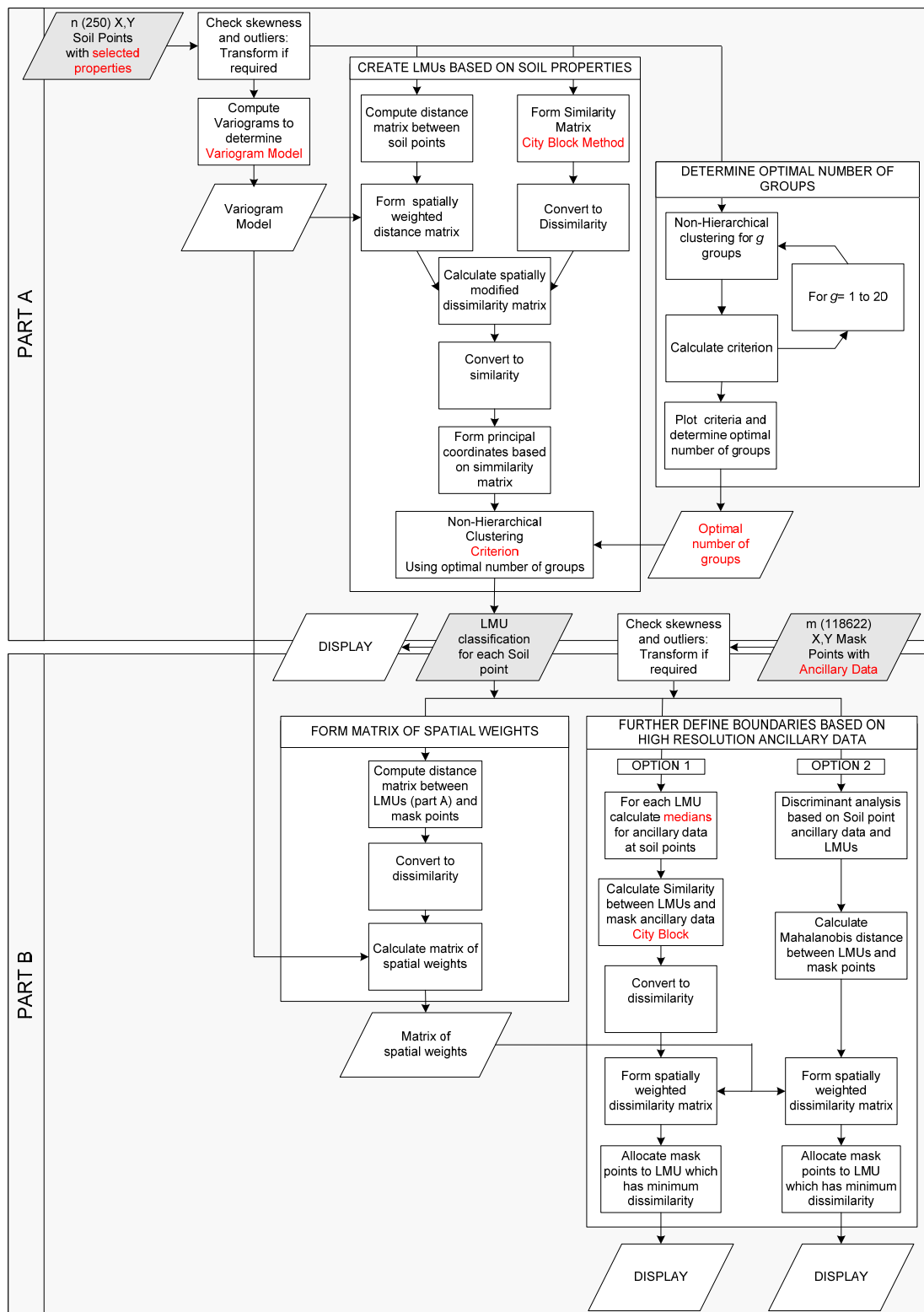


Figure 7.1 Flowchart of LMU classification

## 7.2 Part A: Forming LMUs from Soil Point Data

### 7.2.1 Transforming Input Variables

Variables with skewed distributions and extreme outliers can influence the resulting cluster analysis. For example, outlying values at a soil point can cause that point to be in a cluster by itself, and similarly extreme values in a skewed distribution can result in clusters containing few points with extreme values. This outcome is generally not useful or practical. When skewness is apparent, transformations can assist to make the variable approximately symmetric. Outliers can be removed, but their removal must be treated with care, as they may be of interest.

The selection of input soil variables has been discussed in Chapter 6. The stable soil properties selected as the input variables have been described via summary statistics including histograms and box-plots (Appendix K) and transformed if required as discussed in Chapter 6. The resulting transformed data has been used in the LMU classification.

### 7.2.2 Estimating Geographical Parameters from Variograms

A spatially weighted classification is performed by modifying the dissimilarities between soil points according to the geographic distance between the sampling points and the spatial variation of the input variables (Oliver and Webster, 1989). Following Oliver and Webster (1989) this can be achieved by multiplying the dissimilarity matrix of variables by a function of geographical separation;

$$e_{ij}^* = e_{ij} \cdot f(z_i - z_j) \quad (7.1)$$

where  $e_{ij}$  is the dissimilarity and  $e_{ij}^*$  is the modified dissimilarity between soil points  $i$  and  $j$  and  $z_i$  and  $z_j$  are the locations of  $i$  and  $j$  in one, two or three dimensions. Subsequently, whichever scale and form best describes the spatial variation for the particular investigation can be inserted into Equation (7.1). The geographic distance between the sampling points can be calculated from the easting and northing locations, while the scale and form of spatial variation can be examined through geostatistical analysis according to the variogram parameters of the individual variables.



### 7.2.3 Determining the Scale and Form of Spatial Variation

Variograms are used to describe the spatial variation of a variable. They show the range over which values of a variable are correlated and the variance between values separated by distance  $h$  and direction. Their derivation from the Regionalized Variable Theory is detailed in Appendix N.

Bourgault *et al.* (1992) suggest the multivariate variogram or its counterpart, the multivariate covariogram, as spatial weighting functions rather than the univariate variogram. The multivariate variogram is used to increase dissimilarities between distant samples while the multivariate covariogram is used to decrease similarities between distant samples. In addition Bourgault *et al.* (1992) propose that the metric that has been used to calculate dissimilarities between sampling points should also be used for the multivariate variogram or covariogram. Bourgault *et al.* (1992) make the observation that the main difficulty with Oliver and Webster's (1989) approach is the choice of the variogram model which is a univariate spatial function used to represent the spatial structure of a multivariate measure (dissimilarities).

However, from an initial investigation of the experimental variograms for each of the selected properties the form and range appear to be consistent, indicating that an appropriate variogram model representing the multivariate data set may be derived without resorting to a multivariate variogram.

Experimental variograms were plotted for all soil variables to be used in the model for the entire study area and in several directions. Variograms of the first few principal components were also plotted to take into account the multivariate character of the dissimilarities as did Oliver and Webster (1989). The study area was also split into two regions which, from knowledge of it, appeared to have different underlying geology, and the range and spatial structure were examined. Inspection of the experimental variograms indicated that the differing directions and regions have no effect. Voltz and Webster (1990) suggested that you need over 300 samples to detect anisotropy properly and, as such, it is not really feasible to explore directional effects from this dataset.

Subsequently, experimental variograms were based on all of the soil points and in all directions with a maximum distance of 800m, and lag interval of 20m. The experimental variograms for the 13 selected soil properties, and the first 6 principal components (which accounted for approximately 80 percent of the variance as shown in Table 7.1) from those variables, were plotted. In some cases there were outliers, which had undue influence on the model fitting and parameters. The soil points contributing to the outliers were identified in a variogram cloud (Webster and Oliver, 2001a) and appropriate points were removed. The resultant experimental variograms are displayed in Appendix O.

Table 7.1 Principal components of correlation matrix of 13 soil properties

<b>Principal Component</b>	1	2	3	4	5	6
<b>Percentage of Variance</b>	37.35	13.46	8.75	7.45	6.71	6.21
<b>Cumulative percentage</b>	37.35	50.80	59.56	67.01	73.72	79.93

The three most common variogram models namely, spherical (Equation (7.2)), exponential (Equation (7.3)), and Gaussian (Equation (7.4)) were fitted (Figure 7.2) to the experimental variograms.

$$\begin{aligned} \gamma(h) &= c_0 + c_1 \left[ \frac{3h}{2a} - 1/2(h/a)^3 \right] && \text{for } 0 < h < a \\ &= c_0 + c_1 && \text{for } h \geq a \\ &= 0 && \text{for } h = 0 \end{aligned} \quad (7.2)$$

$$\begin{aligned} \gamma(h) &= c_0 + c_1 [1 - \exp(-h/r)]; \quad a \approx 3r && \text{for } h > 0 \\ &= 0 && \text{for } h = 0 \end{aligned} \quad (7.3)$$

$$\begin{aligned} \gamma(h) &= c_0 + c_1 [1 - \exp(-h/r)^2]; \quad a^2 \approx 3r^2 && \text{for } h > 0 \\ &= 0 && \text{for } h = 0 \end{aligned} \quad (7.4)$$

where  $a$  is the range,  $h$  is the distance between points or lag,  $c_0$  is the nugget variance,  $c_0 + c_1$  equals the sill variance. The exponential and Gaussian models approach their sills asymptotically and therefore do not have a finite range. Parameter  $r$  controls the rate at which the variogram approaches the asymptote. In these models the effective range is defined as  $a \approx 3r$  and  $a \approx \sqrt{3}r$  respectively, and is the distance at which the variance is approximately 95 percent of the sill variance.

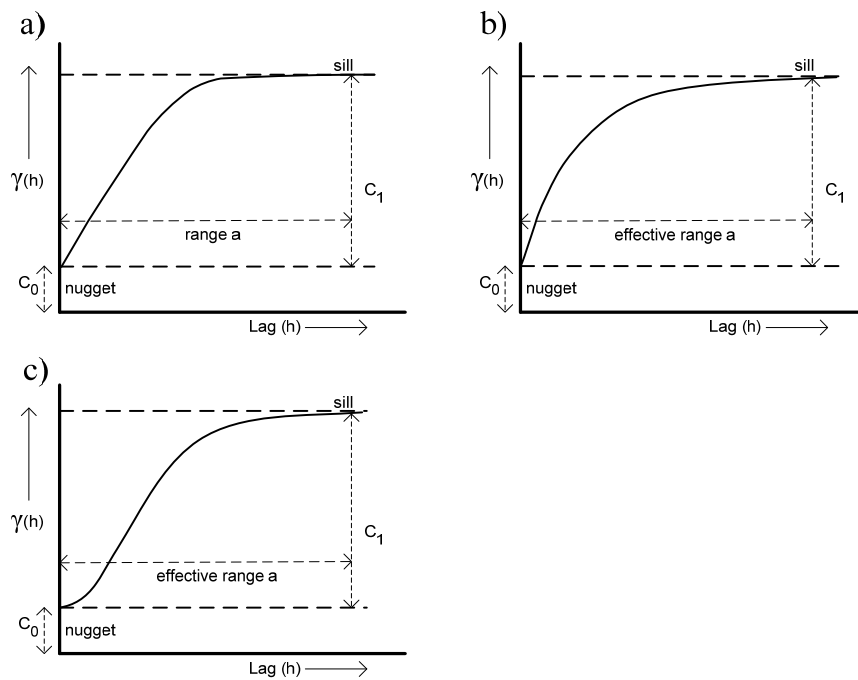


Figure 7.2. Commonly used variogram models: (a) spherical; (b) exponential; and (c) Gaussian (adapted from Burrough and McDonnell, 1998c; Webster and Oliver, 2001c)

The best fit model was selected in the least squares sense. For seven variables, none of the variogram models accounted for a significant percentage of variation in experimental variogram. The exponential model provided the best fit for eight of the 12 fitted variables, the Gaussian model for only three variables and the spherical model for only one of the variables. Subsequently, the exponential model was used for all 12 variables for which a variogram could be fitted (Figure 7.3) and the range and the nugget variance as a percentage of the sill variance were determined (Table 7.2).

Table 7.2 Exponential variogram models for soil properties and principal components

Soil Property	% Nugget variance	Effective Range (m)	r <sup>2</sup>
Restricted Depth (cm)	4.86	160	75.26
10 Stones%	32.57	297	81.17
Log (10-pHca)	6.96	253	90.68
Log (10-Silt)	10.53	409	93.59
Log (10-Clay)	4.36	268	89.44
Log (30-EC)	36.22	339	74.06
30-pHca	0.00	114	83.65
Log (30-Clay)	0.00	190	86.39
Log (30-ECEC)	0.00	137	88.18
Log (30-ESP)	0.00	341	89.25
30-Ca:Mg	0.00	207	80.19
PCP-3	16.16	129	51.13

The mean and median effective range equate to approximately 240m and 230m respectively, with a standard deviation of approximately 95m and a standard error of 8m. The nugget variance is zero for five of the 12 variograms and ranges up to 36 percent. Exponential models with a nugget variance of 10 percent and effective ranges of 200m, 250m and 300m were used as a function of geographical separation,  $f(z_i - z_j)$ , to weight spatially the classification of soil points into LMUs.

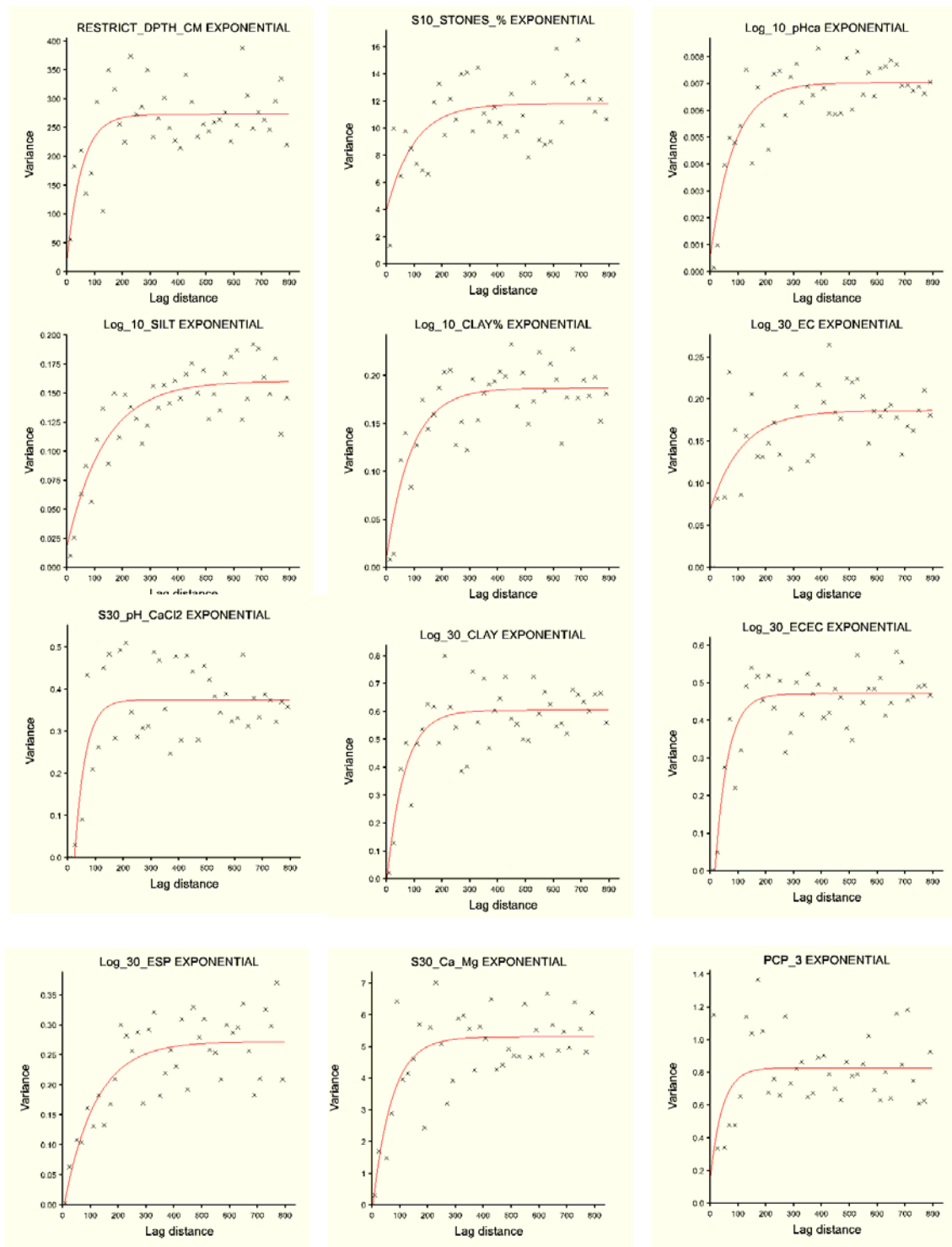


Figure 7.3 Experimental and exponential model variograms for soil properties and principal components after outliers have been removed

#### 7.2.4 Forming the Similarity Matrix on Soil Variables

For every pair of individuals  $i$  and  $j$  there will be a distance separating them in character space, often known as the *Pythagorean distance*, *Euclidean distance*, or *taxonomic distance* and a corresponding similarity,  $S_{ij}$ , referred to as the *similarity coefficient* or *similarity index* (Webster and Oliver, 1990c). The *similarity index* is based on the complement of the dissimilarity, which is scaled so it lies in the range 0 (identity) to 1 (maximum dissimilarity). Using the soil variables, two sample points (individuals) will have a similarity of unity when both have identical values for all variables and a similarity of zero when the values for both samples differ maximally for all variables. If there are  $N$  sample points, the similarity matrix will be of the size  $N \times N$ . Dissimilarity is related to similarity as shown in Equation (7.5), which is equivalent to the definition used by Oliver and Webster (1989) and this definition has been used hereafter.

$$e_{ij} = \sqrt{1 - S_{ij}} \quad (7.5)$$

where  $S_{ij}$  measures the similarity and  $e_{ij}$  measures the dissimilarity between the soil points  $i$  and  $j$ .

The coefficient of similarity,  $S_{ij}$ , presented by Oliver and Webster (1989), following the Gower (1971) generalization, takes the form;

$$S_{ij} = \frac{\sum_{k=1}^p (1 - |x_{ik} - x_{jk}| / r_k) v_{ijk}}{\sum_{k=1}^p v_{ijk}} \quad (7.6)$$

where  $x_{ik}$  and  $x_{jk}$  are observed values of the  $k$ th soil property (variable) and  $r_k$  is the empirical range of the variable sampled (or the range of that variable known to exist) and  $v_{ijk}$  is the weight assigned to the  $k$ th property in the comparison of individuals  $i$  and  $j$ . The division by  $r_k$  standardises quantitative variables.

In summary, the measure of similarity is formed by multiplying each contribution by the corresponding weight, summing all these values, and then dividing by the sum of the weights. The similarity shown in Equation (7.6) is called the *city block* method, when  $v_{ijk} = 1$ . It is based on the sum of the absolute difference for each soil variable divided by the number of soil variables ( $p$ ) and is usually known as the *mean*

*character distance* (Webster and Oliver, 1990c). The similarity measure was converted to dissimilarity using Equation (7.5).

### 7.2.5 Determining the Geographic Distance between Sample Points

The distance between the sampling points can be calculated using *Pythagoras theorem* (synonymous with Euclidean distance);

$$D_{ij} = \sum_{k=1}^2 (y_{ik} - y_{jk})^2 \quad (7.7)$$

where  $y_{i1}$  and  $y_{i2}$  are the easting and northing values at soil point  $i$ . As mentioned earlier, when dealing with large data sets distance matrices become extremely large and computationally demanding. Utilisation of low level programming languages, as used in the Genstat® command to form a similarity matrix will improve the efficiency of the calculation. Accordingly, a Euclidean measure of similarity based on easting and northing was used as it gives a similarity between two sampling points ( $i$  and  $j$ ) following the form;

$$P_{ij} = 1 - \frac{1}{2} \sum_{k=1}^2 \left[ \frac{(y_{ik} - y_{jk})}{R_k} \right]^2 \quad (7.8)$$

where  $R_k$  is the range for each variable. Ensuring that the range for both eastings and northings is equal ( $R$ ) then;

$$P_{ij} = 1 - \frac{1}{2R} \sum_{k=1}^2 (y_{ik} - y_{jk})^2 \quad (7.9)$$

or,

$$P_{ij} = 1 - \frac{1}{2R} D_{ij}^2 \quad (7.10)$$

where,  $D_{ij}$  is the geographic distance between two sample points  $i$  and  $j$ . Thus, we can calculate  $D_{ij}$  from,

$$D_{ij} = R\sqrt{2(1 - P_{ij})} \quad (7.11)$$

In order to minimise computer resources it is useful to calculate dissimilarities,  $d_{ij}$ , from the distances as follows;

$$d_{ij} = \frac{D_{ij}}{R\sqrt{2}} = \sqrt{(1 - P_{ij})} \quad (7.12)$$

As such, the dissimilarity matrix proportional to distance ( $d_{ij}$ ) for  $N$  points can be efficiently calculated by adding an extra point, with appropriate easting and northing, to equalise the range of the easting and northing data. Then the similarity can be calculated using the Euclidean measure, which forms  $P_{ij}$   $((N+1) * (N+1))$  matrix. This is transformed to the  $d_{ij}$  matrix following Equation (7.12) and the extra point is

deleted leaving a matrix of dissimilarity proportional to distances between each of the  $N$  points.

### 7.2.6 Forming Matrix of Spatial Weights

The spatially weighted dissimilarity matrix ( $e_{ij}^*$ ) is formed through the product of the dissimilarity matrix of soil variables ( $e_{ij}$ ) and a function based on the variogram (Equation (7.1)). When the function is evaluated at all soil points the values can be stored in a matrix of spatial weights ( $w_{ij}$ ). Subsequently, the spatially weighted dissimilarity matrix is formed as follows;

$$e_{ij}^* = e_{ij} \cdot w_{ij} \quad (7.13)$$

Using a spherical model, Equation (7.2) as in Oliver and Webster (1989) the weights take values;

$$w_{ij} = \frac{c_0}{c_0 + c_1} + \frac{c_1}{c_0 + c_1} \left[ \frac{3}{2} \frac{D_{ij}}{a} - \frac{1}{2} \left( \frac{D_{ij}}{a} \right)^3 \right] \quad \text{for } 0 < D_{ij} < a \quad (7.14)$$

$$w_{ij} = 1 \quad \text{for } D_{ij} \geq a$$

Note that division by the sill variance ( $c_0 + c_1$ ) ensures that  $w_{ij}$  will lie between 0 and 1.

Equation (7.14) can be re-written as;

$$w_{ij} = c_0^* + (1 - c_0^*) \left[ \frac{3}{2} \frac{d_{ij}}{a^*} - \frac{1}{2} \left( \frac{d_{ij}}{a^*} \right)^3 \right] \quad \text{for } 0 < d_{ij} < a^* \quad (7.15)$$

$$w_{ij} = 1 \quad \text{for } d_{ij} \geq a^*$$

where;  $c_0^* = \frac{c_0}{c_0 + c_1}$  is called the adjusted nugget variance

and  $a^* = \frac{a}{R\sqrt{2}}$  is called the adjusted range

If an exponential model is used;

$$w_{ij} = c_0^* + (1 - c_0^*) [1 - \exp(-3d_{ij} / a^*)] \quad \text{for } 0 < d_{ij} < a^* \quad (7.16)$$

$$w_{ij} = 1 \quad \text{for } d_{ij} \geq a^*$$

Alternatively, if a Gaussian model is used;

$$w_{ij} = c_0^* + (1 - c_0^*) [1 - \exp(-\sqrt{3}d_{ij} / a^*)^2] \quad \text{for } 0 < d_{ij} < a^* \quad (7.17)$$

$$w_{ij} = 1 \quad \text{for } d_{ij} \geq a^*$$



A similarity matrix for use in a cluster analysis is calculated from the spatially weighted dissimilarity matrix as follows;

$$S_{ij}^* = 1 - e_{ij}^{*2} \quad (7.18)$$

### 7.2.7 Forming Principal Coordinates from a Similarity Matrix

The similarity matrix  $S_{ij}^*$  can be used for a wide variety of clustering techniques and is the starting point for most agglomerative hierarchical techniques (Webster and Oliver, 1990a). However, it has been suggested that non-hierarchical cluster analysis is more appropriate for variables that are not hierarchically structured, such as soil properties (Oliver and Webster, 1989). Mora and Iverson (2002) follow Oliver and Webster's (1989) rationale. Unfortunately, the similarity matrix cannot be used directly in non-hierarchical methods, as they use information in variables. Oliver and Webster (1989) propose that the similarity matrix can be converted to variables through principal coordinate analysis.

Principal coordinate analysis takes a similarity matrix, which describes relationships between a set of objects, and forms variables which give an ordination of the objects in multidimensional space. When the similarity matrix is a correlation matrix, the principal coordinate analysis is the same as principal component analysis.

Using a similarity matrix, which has been spatially modified using a variogram with a selected range (200m, 250m and 300m), a principal coordinate analysis was performed specifying 250 dimensions. The percentage of variance accounted for by the first 13 principal coordinates (PCO's) for each of the three ranges are listed in Table 7.3.

Table 7.3 Percentage of variance accounted for by 13 PCO's for effective ranges of 200m, 250m and 300m

PCO	Range = 200m		Range = 250m		Range = 300m	
	Percentage of Variance	Cumulative percentage	Percentage of Variance	Cumulative percentage	Percentage of Variance	Cumulative percentage
1	21.55	21.55	21.53	21.53	21.61	21.61
2	9.02	30.57	9.03	30.56	9.14	30.75
3	7.54	38.11	7.58	38.14	7.59	38.33
4	5.96	44.07	5.96	44.10	6.02	44.35
5	4.57	48.64	4.70	48.80	4.72	49.07
6	3.91	52.56	3.99	52.79	4.01	53.09
7	3.65	56.20	3.75	56.54	3.78	56.87
8	3.21	59.41	3.35	59.89	3.40	60.26
9	2.77	62.18	2.81	62.70	2.91	63.17
10	2.50	64.68	2.64	65.34	2.70	65.88
11	2.12	66.80	2.29	67.63	2.37	68.25
12	1.98	68.78	2.17	69.80	2.27	70.52
13	1.89	70.67	2.03	71.83	2.15	72.66

The first 13 principal coordinates, which account for more than 70 percent of the variance between sampling points (Table 7.4), were used for non-hierarchical clustering described hereafter. There is no guarantee that a modified similarity matrix will generate positive roots, but in this case there are very few negative roots and they only occur well after most of the variance between sampling points has been accounted for. This concurs with Webster's (2006) research, who concludes that under these circumstances when the negative roots are all small in absolute value they can be ignored as they have no important effect on the final result.

Table 7.4 Summary of principal coordinate analysis over several spatial ranges

	Spatial Range		
	200m	250m	300m
<b>Sum of positive roots</b>	65.41	67.57	69.37
<b>Sum of negative roots</b>	-15.39	-17.68	-19.60
<b>Sum of roots</b>	50.02	49.90	49.76
<b>First negative root</b>	143	134	127
<b>No. of PCO</b>	13	13	13
<b>Percentage of variance accounted for by 1<sup>st</sup> 13 PCO's</b>	70.67	71.83	72.66

### 7.2.8 Selecting the Appropriate Number of LMUs

Non-hierarchical cluster analysis forms a pre-determined number of clusters (LMUs) and therefore, some knowledge of an appropriate number is required. It is always difficult to make a decision on the correct number of clusters and the choice is often based on the desires or needs of the users. Users may consider practicability, scale

and size of clusters, and the within cluster variance of soil properties that one is prepared to accept.

After preliminary testing of the LMU classification it appeared that eight or more clusters would lead to divisions of the landscape at a finer scale than paddock boundaries, while any number greater than 15 would create too many management options. When the number of well-defined clusters is unknown *a priori*, Oliver and Webster (1989) suggest computing Wilks' criterion for each subdivision to determine the optimal number. Wilks' criterion  $L$ , proposed by Friedman and Rubin (1967), is given by the ratio of the determinants

$$L = \frac{|W|}{|T|} \quad (7.19)$$

where  $W$  is the within clusters sums of the squares and products matrix, for the original variables and  $T$  is the total sums of the squares and products matrix (Webster and Oliver, 1990b). Following Marriott (1971), Oliver and Webster (1989) plotted  $g^2L$  against  $g$ , the number of groups (from 2 to  $m$ ), and the appropriate number of clusters was identified at the value of  $g$ , at which  $g^2L$  has a sharp decline below the general trend.

The appropriate number of clusters in this case has been determined following Oliver and Webster's (1989) approach. However, a *modified Wilks' ( $L^*$ )* and *sums ( $S$ )* criterions' were used. The modification of Wilks' criterion is a ratio of the determinants of means squares rather than sums of squares, and is as follows;

$$L^* = \frac{|W|}{|T|} * \frac{(n-1)}{(n-g)} \quad (7.20)$$

The *sums* criterion is calculated as;

$$S = \text{Trace}(W) \quad (7.21)$$

The modified Wilks' criterion  $L^*$ , is synonymous with GenStat ® *within* criterion that minimises the variance within clusters, whereas the Wilks' criterion minimises the sum of squares within clusters. The *sums* criterion maximises the total between-group sums of squares (Digby, 2003). The *sums* criterion has been selected as the

most appropriate for this methodology as explained in the following Section 7.2.9 while the modified Wilks' criterion  $L^*$  (*within*) should also provide an indication of the appropriate number of groups when plotted as described above.

Using the 13 selected soil variables non-hierarchical classification was performed with the *sums* and *within* criterion selected. For non-hierarchical cluster analysis no 'missing values' can be present, and as such 46 soil points were excluded. Non-hierarchical cluster analysis was used to classify the remaining 204 soil points into 2 to 21 groups (Figure 7.4). The procedure was repeated using the largest subset of linearly independent soil variables. It was not possible to include all 22 soil variables because of the linear dependency between sand, silt and clay percentages at 10cm and 30cm depths. Thus sand at 10cm and sand and silt at 30cm depths were not included resulting in 19 soil variables (Figure 7.5).

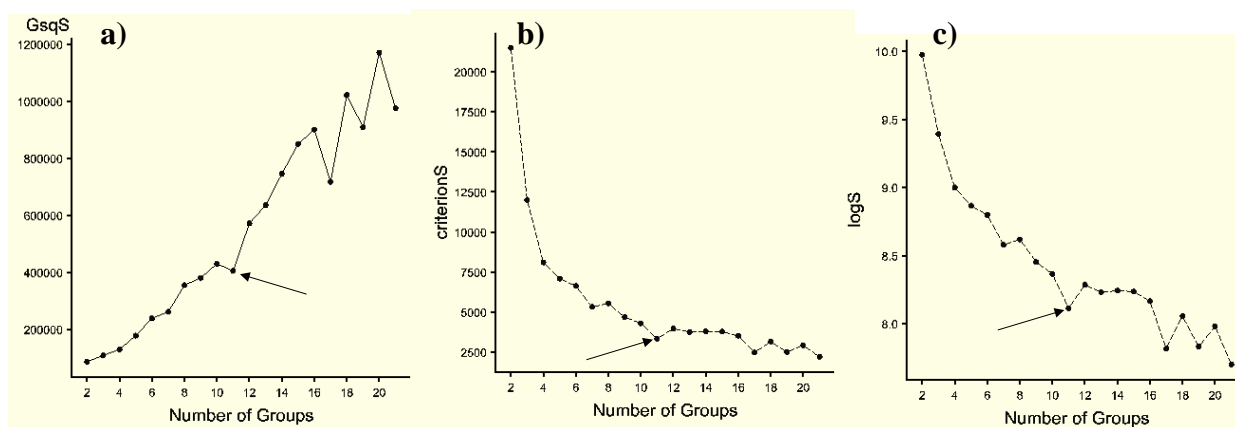


Figure 7.4 Plot of a)  $g^2S$  against  $g$ , b) Criterion  $S$  against  $g$  and c)  $\log S$  against  $g$  based on 13 soil variables

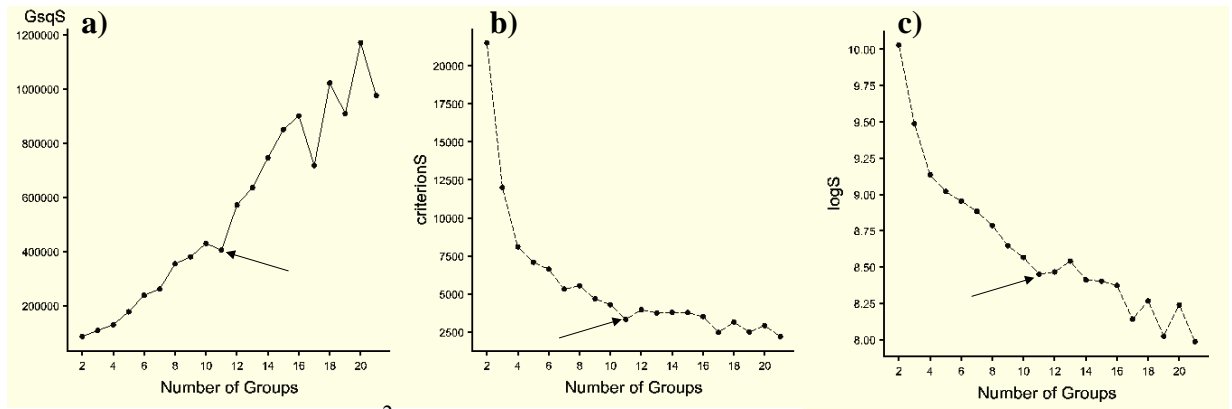


Figure 7.5 Plot of a)  $g^2S$  against  $g$ , b) Criterion  $S$  against  $g$  and c)  $\log S$  against  $g$  based on 19 linearly independent soil variables

Using the *within* criterion, criterion  $L^*$  was calculated as specified in Equation (7.20) and  $g^2 L^*$ ,  $L^*$  and  $\log L^*$  were plotted against  $g$  based on 13 selected soil variables (Figure 7.6) and 19 soil variables (Figure 7.7).

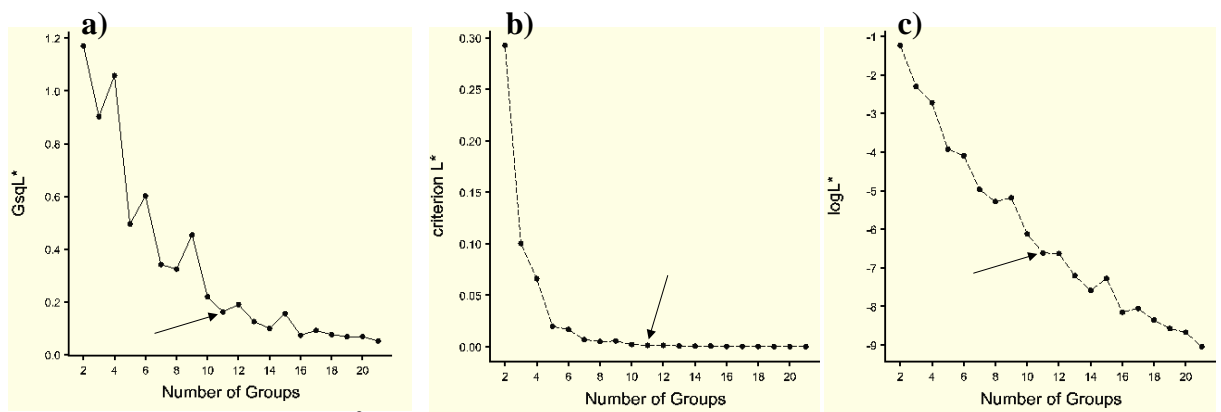


Figure 7.6 Plot of a)  $g^2 L^*$  against  $g$ , b) Criterion  $L^*$  against  $g$  and c)  $\log L^*$  against  $g$  based on 13 soil variables

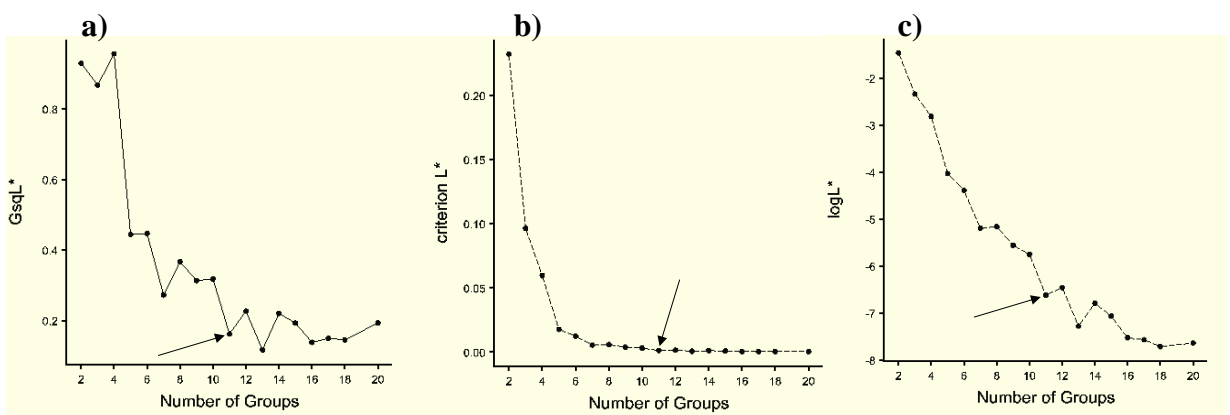


Figure 7.7 Plot of a)  $g^2 L^*$  against  $g$ , b) Criterion  $L^*$  against  $g$  and c)  $\log L^*$  against  $g$  based on 19 linearly independent soil variables

Figure 7.4 through to Figure 7.7 give an indication, however weak, that at  $g = 11$  criterion  $S$  falls below the general trend and criterion  $L^*$  does not provide any reason for this choice to be changed. As such, 11 LMUs will be specified for the clustering described in the following section.

### 7.2.9 Non-hierarchical Clustering of 250 Soil Points

Non-hierarchical cluster analysis subdivides a group of objects on which a number of measurements have been made into a specified number of clusters such that a specified criterion is optimised (Webster and Oliver, 1990b). In GenStat® both the modified *Wilks'* criterion and the *sums* criterion are available.

Utilising the 13 PCOs and a selection of ranges as indicated in Table 7.4, the non-hierarchical clustering was performed specifying 11 LMUs and the *sums* criterion, which maximises the between-group sum of squares (Payne, 2003). The *sums* criterion has been selected as it seems logical that the formation of a new LMU is only necessary if it is different to its neighbouring LMU. Farmers will make the choice to manage an area differently only if there is a substantial difference in the landscape properties and the potential for an economic gain. The resulting LMUs using spatial ranges between 200m to 300m are provided in Figure 7.8.

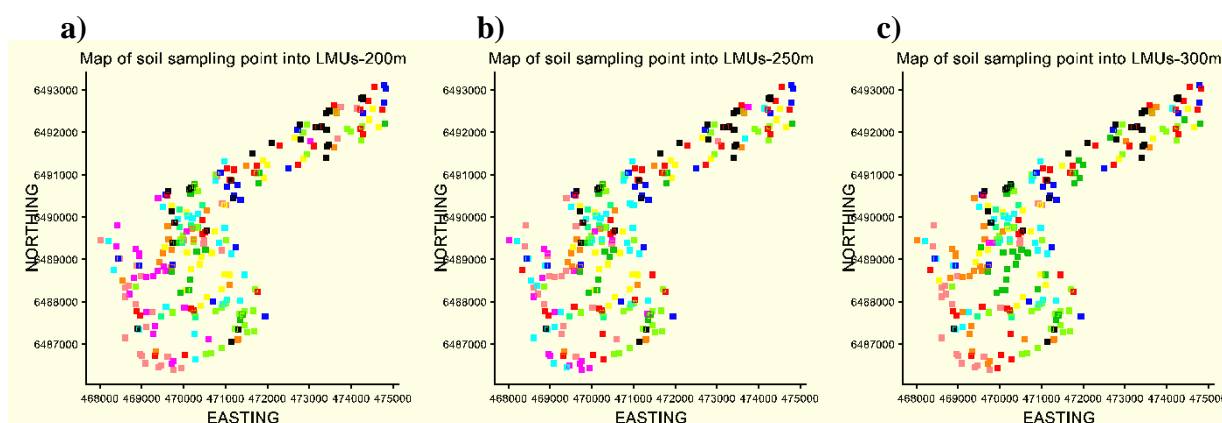


Figure 7.8 Soil point LMU groups based on 250 soil points for a range of a) 200m, b) 250m and c) 300m

### 7.2.10 Utilising Landform Variables

While Gowers' (1971), equation (Equation (7.6)) and subsequently GenStat® has the ability to accommodate both quantitative and some qualitative variables into the one equation for similarity, an appropriate measure of similarity is not available for landforms. The methodology to create landforms is described in Section 5.2 and the resultant location of landform types within the landscape is displayed in Figure 5.3. Based on the methodology used to identify landform types, it is inappropriate to apply a similarity measure that was used for continuous or binary data types, as landform outputs have a quantitative and qualitative nature. As such, the landform classes were converted into variables using principal coordinate analysis and new landform variables namely, LF\_PC1 to LF\_PC6 are scores from the principal coordinates described in Section 5.2.2. This process has not been displayed in the flow diagram as it is unique to this particular data set.

### 7.2.11 Analysis of Variance of Ancillary Data and LMUs

Part B of the model utilises ancillary variables, i.e. landform (PCO1-6), CTI, slope% and RI, for the remainder of the study area (called mask data) to map the LMU boundaries at higher resolution. For Part B to be viable a relationship must exist between the ancillary data and the LMUs just formed. A simple analysis of variance (ANOVA) was performed for each of the ancillary variables at each soil point with respect to the LMU groups to which the soil points have been allocated (Table 7.5).

Table 7.5. Significance of differences between LMUs formed at each range for ancillary variables

Variable	No. soil points	Significance		
		Effective Range = 200m	Effective Range = 250m	Effective Range = 300m
LF-PCO1	250	0.047	0.567	0.924
LF-PCO2	250	<0.001	<0.001	<0.001
LF-PCO3	250	0.002	0.010	0.011
LF-PCO4	250	0.004	0.022	0.033
LF-PCO5	250	0.520	0.516	0.614
LF-PCO6	250	0.069	0.092	0.104
Log CTI	237	<0.001	<0.001	<0.001
Log Slope%	250	<0.001	<0.001	<0.001
RI	206	<0.001	<0.001	<0.001

The results confirm that LMUs are associated with changes in most of the landform variables, CTI, slope% and RI, and that it is appropriate to use them for identifying the LMU boundaries at higher resolution. Although LF-PCO1 and LF-PCO5 do not

show a significant difference between LMUs they are included in Part B of the model. This is because they are available at all mask points and may provide information about LMUs in association with other variables.

### 7.3 Part B: Further Defining LMU Boundaries using High Spatial Resolution Ancillary Data

Part B of the LMU classification delineates higher resolution boundaries based on ancillary data collected on a 10m resolution grid across the entire study area. The method assigns each mask pixel to one of the existing LMUs formed in Part A. This assignment is based on the pixels similarity in terms of landform types, CTI, slope% and RI. Initially a dissimilarity matrix ( $f_{kl}$ ) is formed between the mask points and LMUs from: (1) a similarity between each mask point and the median values of landform, CTI, slope% and RI for each LMU; or (2) a discriminant analysis based on the landform types, CTI, slope% and RI using the Mahalanobis squared distance between each mask point and each LMU.

The dissimilarity matrix is then modified using a spatially weighted matrix ( $w_{kl}$ ) based on the distances between the mask points and LMUs. At the final step each mask point is assigned to a LMU based on the minimum dissimilarity. The modification of the dissimilarity matrix in this case will take the form;

$$f_{kl}^* = f_{kl} \cdot w_{kl} \quad (7.22)$$

where  $f_{kl}$  and  $w_{kl}$  are the dissimilarity and spatial weighting between mask points  $k$  and LMU  $l$ .

#### 7.3.1 Transforming Ancillary Variables

Following the same methodology as for the stable soil properties, the ancillary variables, CTI, slope%, RI and LF-PCO1-6 (landform variables) were inspected for symmetry and transformed if appropriate. Appendix L contains the summary statistics, histograms and box-plots. The CTI and slope% required logarithmic transformation.

#### 7.3.2 Similarity between Mask Pixels and LMUs; Option 1

Using the soil points that form each LMU, a median value was calculated for the six landform variables, CTI, slope% and RI. The similarity between each mask point



and the median for each LMU was then calculated following the *city block* method as shown in Equation (7.6). The similarity measure was then converted to a dissimilarity following Equation (7.5) to form  $f_{kl}$ .

### 7.3.3 Discriminant Analysis between Mask Pixels and LMUs: Option 2

Option 2 uses discriminant analysis to form a dissimilarity between each mask point and each LMU. A discriminant analysis (canonical variate analysis) forms linear combinations of the landform variables, CTI, slope% and RI (canonical variates) measured at each soil point that best discriminate between the LMUs defined in Part A. Using measurements of the same variables for each mask point, canonical variate scores and Mahalanobis squared distances to LMU means are then calculated for all mask points. Unfortunately, any point with a missing value in any of the variates is excluded from the analysis. Missing values are apparent in this data set for some variables. In order to overcome this problem, discriminant analysis was performed several times to calculate Mahalanobis distances based on as many variables as are available at each mask point. Initially all variables were included in calculating Mahalanobis distances. The variables with missing values were subsequently removed and Mahalanobis distances were then calculated for mask points, which previously could not be included. This forms a matrix of Mahalanobis distances ( $f_{kl}$ ) which can be transformed to dissimilarities between each mask point and LMU by dividing by a suitable constant.

### 7.3.4 Forming the Geographically Weighted Dissimilarity Matrix between Soil Based LMUs and Mask Data Set

In a similar manner to that described in Section 7.2.5, a matrix of  $N \times M$  (soil points x mask points) similarities was formed using easting and northing data and the programming language of Genstat®. Firstly, the maximum range of the eastings and northings was calculated, and the easting and northing range set to be equal by adding an extra point at the appropriate spatial location. Similarity between the soil points and mask points using an Euclidean measure was calculated following Equation (7.8). The LMU to which each soil point had been assigned, was associated with the similarities and the nearest neighbour similarity (i.e. maximum similarity; Figure 7.9) between each mask point and each LMU was stored in an  $M \times P$  matrix (mask points x LMUs).

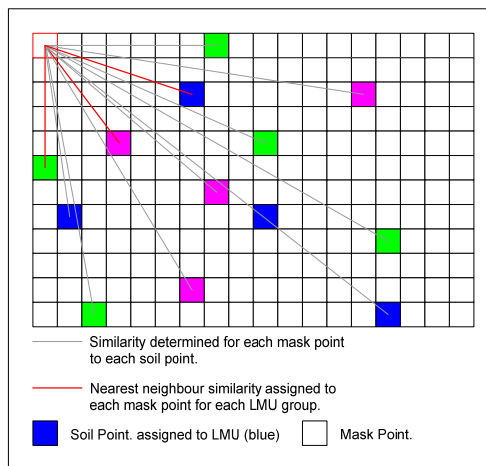


Figure 7.9 Nearest neighbour similarity assigned to a mask point for each LMU

The matrix of similarities between the mask points and LMUs was converted to dissimilarities ( $d_{kl}$ ) as in Equation (7.5). The spatially weighted matrix ( $w_{kl}$ ) between soil based LMUs and mask data is formed by evaluating the appropriate function, as in Equation (7.15) to (7.17) using the dissimilarities ( $d_{kl}$ ) and appropriate range and nugget values.

The dissimilarity matrix, or the Mahalanobis distance matrix, is then modified spatially by multiplying by the spatially weighted matrix  $w_{kl}$ , following Equation (7.22) to form,  $f_{kl}^*$ , a spatially weighted dissimilarity or distance matrix ( $M \times P$ ) between each mask point and each LMU. Each mask point is then assigned to the LMU to which it has minimum dissimilarity or Mahalanobis distance.

#### 7.4 Provisional LMUs

Using a selection of ranges and the two options for forming the LMUs, the provisional maps of LMUs are provided in Figure 7.10. When comparing the Option 1 (Cluster / Cluster) with Option 2 (Discriminant) outputs, the maps are visually similar. In view of adopting farm management options, preference would be given to the Cluster / Cluster output due to the more homogeneous unit outputs. An effect called ‘bullseyeing’, which refers to small circular LMUs, which occur when the distance between the sample points is greater than the range used for weighting, is visible in Figure 7.10 (a) and (d) and to a lesser extent Figure 7.10 (b) and (e). This problem is overcome when the range equals 300. However, as the mean range was

in fact 240m, it seems that an over smoothing would be applied in the latter case, losing some of the higher level details. The 250m range with the Cluster/Cluster method has thus been selected as the appropriate classification to further examine a practical way of describing the units. However, the final decision will always be up to the user's requirements i.e. level of detail (size of units), homogeneity, etc.

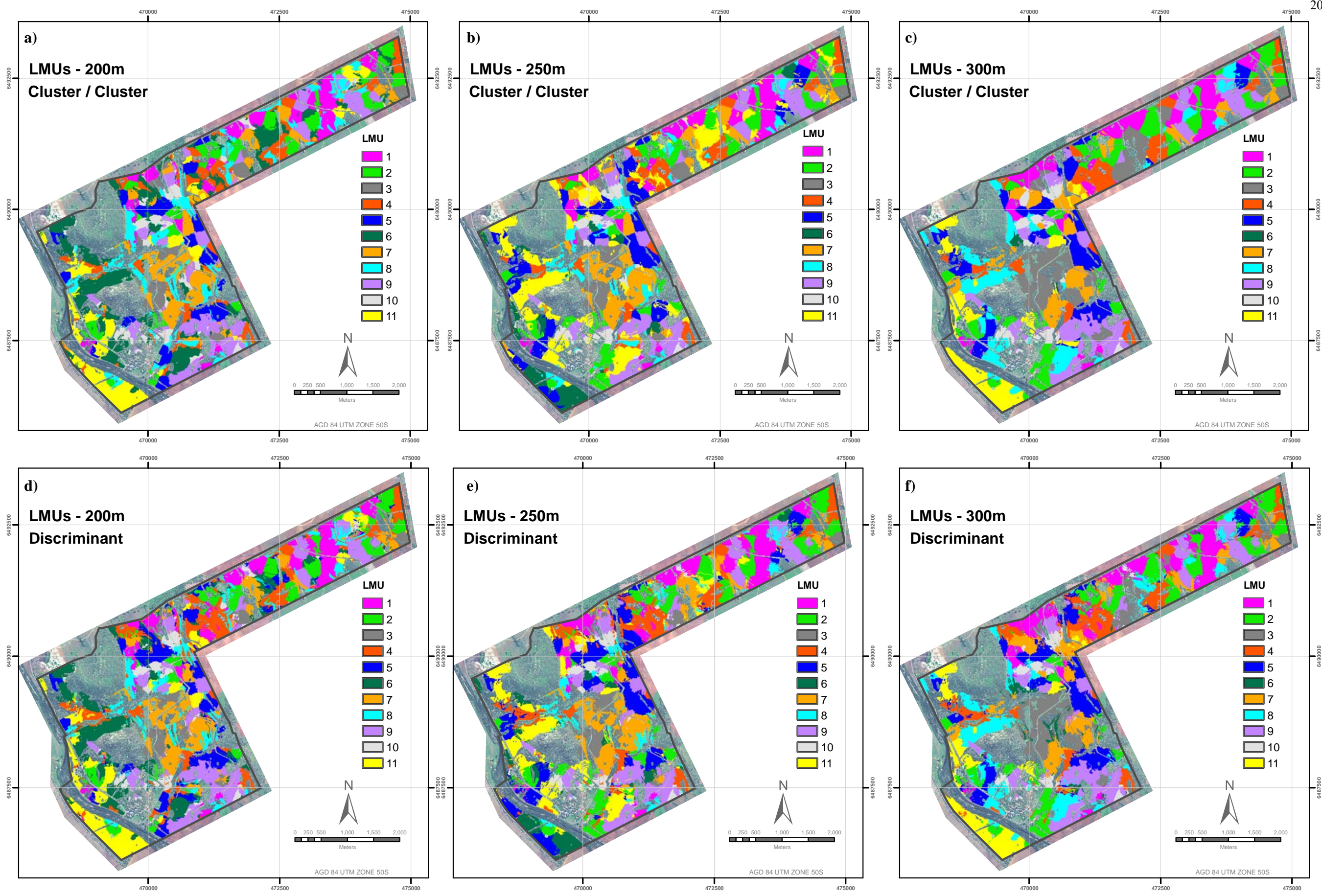


Figure 7.10 LMU groups based on Option 1: Cluster/Cluster analysis for a range of a) 200m, b) 250m and c) 300m and Option 2; Discriminant analysis for a range of d) 200m, e) 250m and f) 300m

## 7.5 The Land Management Unit Map and Unit Descriptions

Discriminant analysis based on a selection of soil properties (Table 7.6) has been used to identify the soil properties that differentiate the LMUs. These soil properties were chosen as they are commonly used in soil survey classification.

Table 7.6 Soil variables used in discriminate analysis

Variable
Log_A_Horizon
Restricted Depth
Log_10_Clay
S10_OrgC
Log_30_Clay
Log_30_Stones
S30_pHca
Log_30_EC
Log_30_ECEC
Log_Slope%

Some of the sampling points had missing values for some of the variables. In this case the mean of the variable for that LMU was used as a substitute. The ten variables were then standardised and discriminant analysis performed. In a discriminant analysis, linear combinations of the original variables (canonical variates) are found that maximise the ratio of between-group to within-group variation, thereby giving functions of the original variables that can be used to discriminate between the groups (LMUs). The first canonical variate (CV1) maximises the discrimination between groups. The second canonical variate (CV2) maximises discrimination between groups that have not been differentiated by CV1. The maximum number of canonical variates equals either the number of original variables or one less than the number of groups, whichever is smaller, and will be 10 in this case.

Canonical variate means were calculated for the soil sampling points in each LMU to indicate the principal differences between LMUs. Biplots were used to graph values of CV1 versus CV2 (or other canonical variates) for each sampling point, as well as, their associated LMUs and the contribution of the soil variables to each canonical variate (Figure 7.11 to Figure 7.13). The position of the LMUs, with respect to the variables on the biplot, has been used as a tool to identify important differences

between LMUs. The discriminant analysis was performed in GenStat® and Excel (Microsoft Corporation, 2003) was used to draw the biplots. The first four canonical variates accounted for over 95% of the variance (Table 7.7).

Table 7.7 Percentage of variance explained and variable loadings of variables for each canonical variate

	CV1	CV2	CV3	CV4
<b>Percentage</b>	56.88	21.12	11.42	5.81
<b>Cumulative</b>	56.88	78.00	89.42	95.23
<b>Variable</b>	<b>Loadings</b>			
Log_A_Horizon	0.258	1.704	0.443	0.145
Restricted Depth	0.470	0.170	-1.297	0.051
Log_10_Clay	-0.520	0.100	0.159	0.555
S10_OrgC	-0.277	-0.011	0.021	0.404
Log_30_Clay	-0.739	0.386	-0.546	-0.753
Log_30_Stones	0.004	0.159	0.469	-0.555
S30_pHca	-0.155	0.374	-0.185	-1.079
Log_30_EC	-0.004	0.033	-0.287	-0.350
Log_30_ECEC	-1.039	-0.045	-0.092	1.102
Log_Slope%	-0.074	-0.280	0.014	0.242

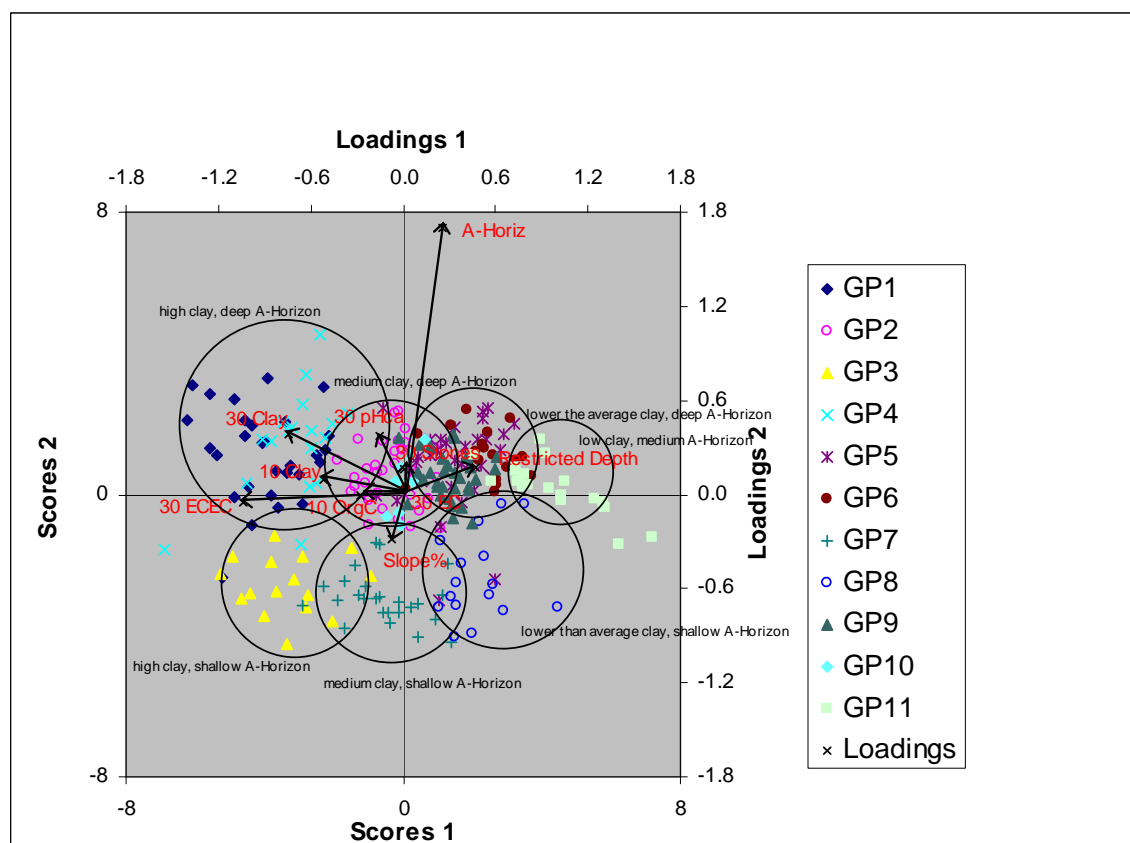


Figure 7.11 Biplot showing soil points within LMUs and canonical variate loadings for CV1 vs CV2



CV1 is associated with ECEC at 30cm depth, clay at 10cm and 30cm depth, and soil restricted depth. The A-Horizon depth dominates CV2, while CV3 is dominated by restricted depth and, to a lesser extent, clay at 30cm depth and stones at 30cm depth. CV4 is loaded with ECEC and pHca at 30cm depth and, to a lesser extent, clay at 30cm depth. Categories have been identified and represented by circles drawn on the biplots to help describe the main discriminating features for each LMU.

It can be seen in Figure 7.11 that while groups 1,3 and 4 all have high clay content, group 3 can be separated from groups 1 and 4 because it has a shallower A-horizon. Figure 7.12 indicates that groups 1 and 4 can be differentiated using restricted depth. Table 7.8 provides a classification tree which shows the characteristics that discriminate between the 11 LMUs, however other characteristics will be associated with these variables (i.e. the dominant variables for each CV). Since CV1 gives the best discrimination between groups, the median value of CV1 for each LMU has been used to form a new order of LMUs. This ranks them in terms of their similarity (New LMU: alphabetical A to K) (Table 7.8). Outputs from non-hierarchical cluster techniques do not form such order.

Table 7.8 Discriminating tree for LMU groups

Clay content	A horizon Depth	Restricted depth	pH at 30cm	ECEC at 30cm	LMU	Median CV1	New LMU
High	Deep	Shallow			1	-3.94	A
		Deep			4	-2.85	C
	Shallow				3	-3.52	B
Medium	Deep		Higher than average.	Lower than average.	10	-0.08	F
			Lower than average	Higher than average.	2	-0.23	E
	Shallow				7	-0.67	D
Lower than average	Deep	Shallow			9	1.42	G
		Medium			5	1.59	H
		Deep			6	2.53	J
	Shallow				8	2.10	I
Low	Medium				11	4.13	K

The final LMU map is provided in Figure 7.14 with a description for each unit, which is based on summaries of values for soil variables of sampling points located within each LMU. The descriptors are provided with reference to the associated



rating tables provided in the soil properties literature review Section 2.2.1. Boxplots for several soil properties and ancillary variables are displayed in Appendix P. A further description of the major characteristic of dominant landscape is provided in Table 7.9.

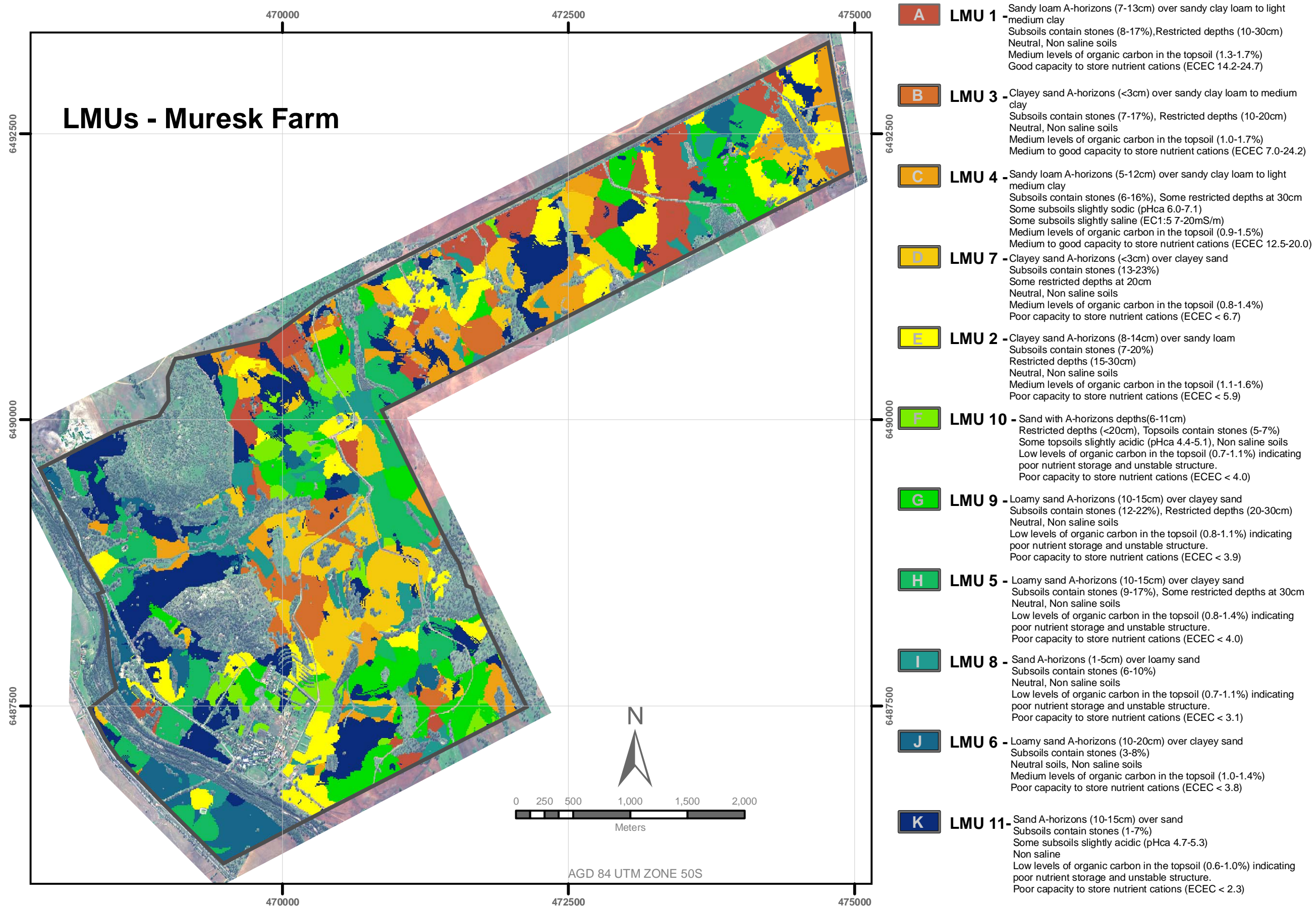


Figure 7.14 The Land Management Unit map for Muresk Farm

Table 7.9 Description of LMUs present in Figure 7.14

LMU	Major characteristics of dominant soil	Major characteristic of dominant Landscape
1 : A	Sandy loam A-horizons (7-13cm) over sandy clay loam to light medium clay Subsoils contain stones (8-17%) Restricted depths (10-30cm) Neutral, Non saline soils Medium levels of organic carbon in the topsoil (1.3-1.7%) Good capacity to store nutrient cations (ECEC 14.2-24.7)	Dominated by somewhat poorly drained soils Predominately mid and lower slopes Gently inclined slopes
3 : B	Clayey sand A-horizons (<3cm) over sandy clay loam to medium clay Subsoils contain stones (7-17%) Restricted depths (10-20cm) Neutral, Non saline soils Medium levels of organic carbon in the topsoil (1.0-1.7%) Medium to good capacity to store nutrient cations (ECEC 7.0-24.2)	Dominated by well drained and moderately well drained soils, Predominately crests Gently inclined slopes
4 : C	Sandy loam A-horizons (5-12cm) over sandy clay loam to light medium clay Subsoils contain stones (6-16%) Some restricted depths at 30cm Some subsoils slightly sodic (pH <sub>ca</sub> 6.0-7.1) Some subsoils slightly saline (EC <sub>1:5</sub> 7-20mS/m) Medium levels of organic carbon in the topsoil (0.9-1.5%) Medium to good capacity to store nutrient cations (ECEC 12.5-20.0)	Dominated by somewhat poorly drained soils Predominately lower slopes and depressions Gently inclined slopes
7 : D	Clayey sand A-horizons (<3cm) over clayey sand Subsoils contain stones (13-23%) Some restricted depths at 20cm Neutral, Non saline soils Medium levels of organic carbon in the topsoil (0.8-1.4%) Poor capacity to store nutrient cations (ECEC < 6.7)	Dominated by well drained soils Predominately lower and mid slopes Gently inclined slopes
2 : E	Clayey sand A-horizons (8-14cm) over sandy loam Subsoils contain stones (7-20%) Restricted depths (15-30cm) Neutral, Non saline soils Medium levels of organic carbon in the topsoil (1.1-1.6%) Poor capacity to store nutrient cations (ECEC < 5.9)	Dominated by well drained soils Predominately slopes Gently inclined slopes
10 : F	Sand with A-horizons depths(6-11cm) Restricted depths (<20cm) Topsoils contain stones (5-7%) Some topsoils slightly acidic (pH <sub>ca</sub> 4.4-5.1) Non saline soils Low levels of organic carbon in the topsoil (0.7-1.1%) indicating poor nutrient storage and unstable structure Poor capacity to store nutrient cations (ECEC < 4.0)	Dominated by well drained soils Predominately upper slopes Gently inclined slopes
9 : G	Loamy sand A-horizons (10-15cm) over clayey sand Subsoils contain stones (12-22%) Restricted depths (20-30cm) Neutral, Non saline soils Low levels of organic carbon in the topsoil (0.8-1.1%) indicating poor nutrient storage and unstable structure Poor capacity to store nutrient cations (ECEC < 3.9)	Dominated by well drained soils Predominately crests and upper slopes Gently inclined slopes

LMU	Major characteristics of dominant soil	Major characteristic of dominant Landscape
5 : H	Loamy sand A-horizons (10-15cm) over clayey sand Subsoils contain stones (9-17%) Some restricted depths at 30cm Neutral, Non saline soils Low levels of organic carbon in the topsoil (0.8-1.4%) indicating poor nutrient storage and unstable structure Poor capacity to store nutrient cations (ECEC < 4.0)	Dominated by well drained soils Predominately mid and lowers slopes to depressions Gently inclined slopes
8 : I	Sand A-horizons (1-5cm) over loamy sand Subsoils contain stones (6-10%) Neutral, Non saline soils Low levels of organic carbon in the topsoil (0.7-1.1%) indicating poor nutrient storage and unstable structure Poor capacity to store nutrient cations (ECEC < 3.1)	Dominated by well drained soils Predominately depressions Gently inclined slopes
6 : J	Loamy sand A-horizons (10-20cm) over clayey sand Subsoils contain stones (3-8%) Neutral soils, Non saline soils Medium levels of organic carbon in the topsoil (1.0-1.4%) Poor capacity to store nutrient cations (ECEC < 3.8)	Dominated by well drained soils Predominately flats and mid slopes Very gently to gently inclined slopes
11 : K	Sand A-horizons (10-15cm) over sand Subsoils contain stones (1-7%) Some subsoils slightly acidic (pH <sub>ca</sub> 4.7-5.3) Non saline Low levels of organic carbon in the topsoil (0.6-1.0%) indicating poor nutrient storage and unstable structure Poor capacity to store nutrient cations (ECEC < 2.3)	Dominated by well drained and somewhat excessively drained soils Predominately mid slopes Gently inclined slopes

## 7.6 Summary

The LMU classification framework has been described in detail concluding with a map of LMUs and their associated soil and dominant landform descriptions. The two stage methodology described uses the application of a spatially weighted multivariate classification to high resolution data sets (i.e. DEM and DMSI) demonstrating the technique is computationally feasible for large data sets. The usefulness of multivariate methods for discriminating between LMUs and colouring map units in a way that highlights progressive change, has been demonstrated.

The spatially weighted multivariate classification procedure has formed clusters that are similar in terms of their multivariate data and spatially contiguous. The classification incorporates the scale and form of the spatial variation of the soil variables by spatially weighting the dissimilarities between locations via a geostatistical function. The process may be adjusted at several steps to suit any other

study area location, scale and application. The success of the LMU classification is analysed in the following chapter.

## CHAPTER 8

### VALIDATION OF LMU CLASSIFICATION

*This chapter describes the analytical procedures that have been carried out to examine the success of the LMU classification explained in Chapter 7. Success is defined as how well the classification procedure has performed. The procedures include the use of simulation to verify the allocation of points to LMUs using ancillary data, the use of Kappa map comparisons to compare LMU groups under varying parameters, and an analysis of variance of yield with respect to each LMU.*

#### **8.1 Choosing a Validation Procedure**

In essence, the hypothesis underlying the classification procedure is that homogeneous LMUs based on soil properties can be managed to maximise profits by being able to match inputs to potential yield more accurately than is possible with present paddocks. Ultimately, success is measured by confirming that this is true. However, the validation of higher profits is beyond the scope or aims of this thesis, which aims to *develop a framework for classifying a farm into paddock scale LMUs* using soil properties sampled at low resolution coupled with less expensive high resolution ancillary data sets.

One aspect of this underlying hypothesis, that the LMUs formed are associated with potential yield has been examined. The association between LMUs and actual yield over recent seasons has been examined on the basis that actual yield should be correlated with potential yield. While an association would support the hypothesis that LMUs are correlated with potential yield, a lack of association would not indicate LMUs are not correlated with potential yield, as it could result that actual yield is not correlated with potential yield.

The success of the classification procedure itself has been examined by experimentation. A simulation has been performed using ancillary data for allocating points to LMUs with similar soil properties. The sensitivity of the process to the choice of method and distance parameters has been examined by comparing LMU maps resulting from different choices using a Kappa statistic. No further

validations have been carried out to address other aspects of the classification procedure itself as there are several variables and parameters, which will inevitably alter the classification output if their values are changed (i.e. the choice of number of groups).

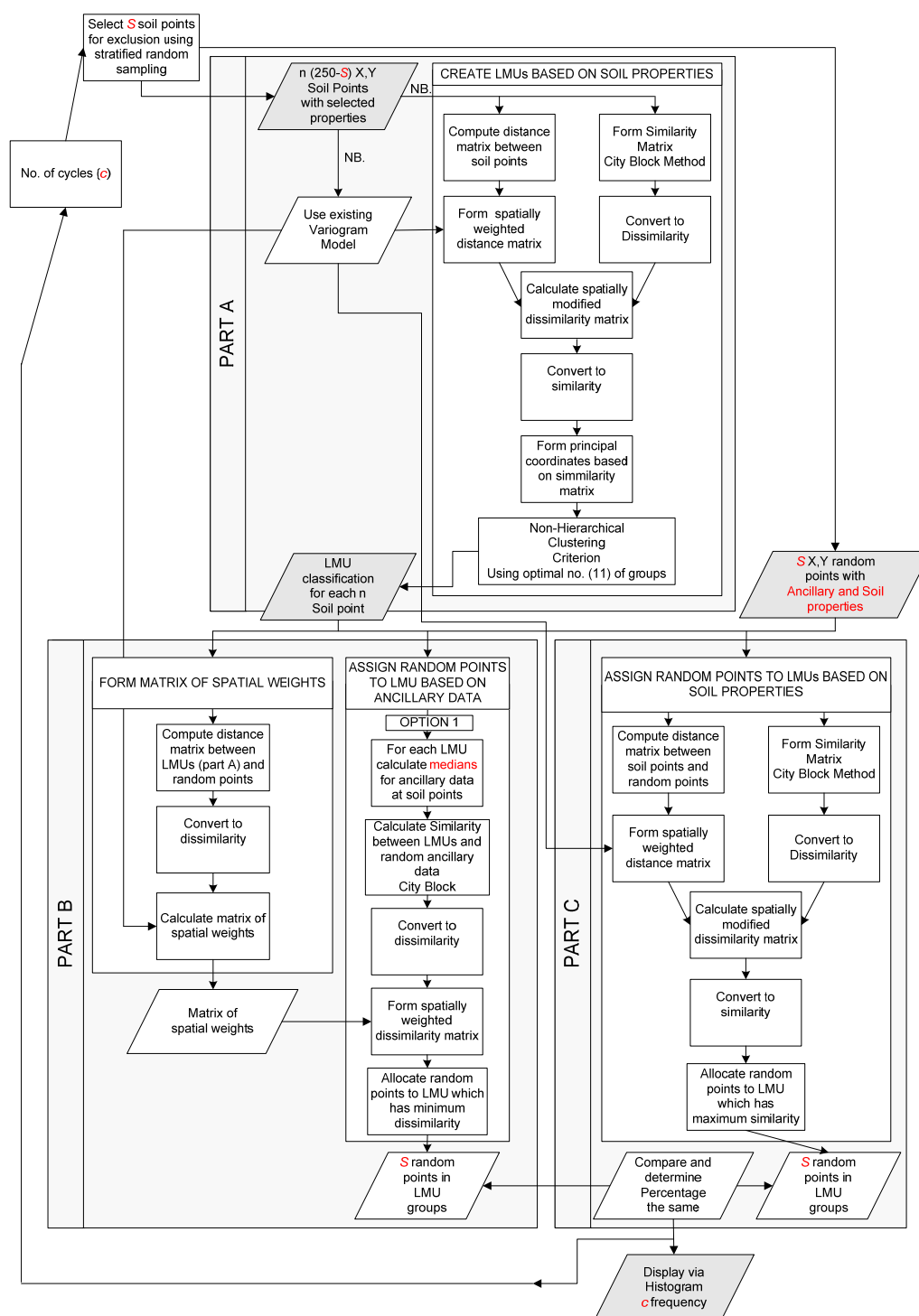
## 8.2 Validation of LMU Allocation using Ancillary Data

One way to test that the LMU classification technique has correctly represented the landscape is to take further field soil samples within the LMUs and analyse their soil properties. Then determine how similar the extra field samples are to the LMU to which they have been assigned based on the 13 selected stable soil properties identified to be included in the formation of LMUs (Table 6.8).

In order to achieve this without further field sampling, the existing 250 points have been randomly subdivided into two subsets. Only points in the first subset have been used for the LMU classification using stable soil properties. Points in the second subset (verification subset) have been allocated to the LMUs they would normally be assigned, if they only had ancillary data available (landform, CTI, slope, RI). In addition, the LMU to which each point is most similar based on the 13 selected soil properties has been identified as the *correct* LMU. An analysis has been carried out to evaluate how often the LMU based only on ancillary data corresponds to the correct LMU. This is discussed in the following section.

The validation of the process of LMU allocation using ancillary data consists of three parts after  $S$  soil points are randomly excluded from the available 250 soil points (Figure 8.1). Part A uses a spatially weighted cluster analysis to subdivide the remaining soil points into LMUs based on the stable soil properties. Part B uses an additional spatially weighted multivariate analysis to assign the  $S$  excluded soil points to the LMUs using the high resolution ancillary data (landform, CTI, slope, RI). Part C assigns the  $S$  excluded soil points to LMUs using a spatially weighted multivariate analysis based on the 13 stable soil properties and compares the LMUs selected for each soil point in parts B and C. The random selection of points for exclusion and Parts A to C are repeated  $c$  times and a summary of the comparisons are displayed as a histogram.

The program used for Parts A and B is the same used to classify the LMUs (Chapter 7) with an identical set of parameters. The random comparison validation process is explained in the following sections according to the flowchart in Figure 8.1.



NB. Transformed (if required) variables used as in Chapter 7 (Figure 7.1)

Figure 8.1 Flowchart of random comparison/validation process



### 8.2.1 The Random Selection

Using the 250 soil points,  $S$  points are randomly excluded for the validation process. In order to ensure that the points are relatively evenly spatially distributed across the farm, the 250 soil points are assigned to 10 strata based on their easting and northing coordinates. The  $S$  points are then selected with an equal number from each stratum. As such,  $S$  has a minimum value of 10 and must be a multiple of 10.

The question of how many points should be used for validation purposes is open for debate. McBratney *et al.* (2003) suggest one quarter to one third of the total sample. However, this is not based on empirical evidence. In this study fifty random points (i.e. 20 percent) is determined as the minimum number required to examine how well the ancillary data process is working. In addition, the remaining 200 soil points are sufficient to generate a useful LMU classification.

### 8.2.2 Part A: Forming LMUs based on Soil Point Properties

Following the process identical to that used for Part A of the LMU classification (Section 7. 2), the remaining ( $n-S$ ) soil points are allocated to LMUs using a spatially weighted multivariate classification based on the 13 selected soil properties. Several parameters, which were trialled and selected in the initial LMU classification, have been used. These include the exponential variogram model with a 250m range and nugget variance of 10 percent, along with 11 LMU groups specified for the non-hierarchical cluster analysis on the 13 PCO's derived from the 13 soil properties. At the end of Part A all ( $n-S$ ) soil points are assigned to one of 11 LMUs.

### 8.2.3 Part B: Assigning Random Points to LMUs based on Ancillary Data

Part B follows a methodology identical to that in the original LMU classification (Section 7. 3), but uses  $S$  randomly excluded points as mask pixels in this instance. Option 1 (Cluster/Cluster) with a range of 250m has been chosen as the most appropriate output map in Chapter 7. This results in the  $S$  random points being assigned to one of the 11 existing LMUs based on the similarity in ancillary data with an exponential spatial weighting.

#### 8.2.4 Part C: Assign Random Points to LMUs based on Soil Properties

In Part C the  $S$  random points are assigned to one of the 11 existing LMUs using a spatially weighted similarity based on the 13 stable soil properties and methodology similar to Part B. Firstly, the similarity between the  $S$  random points and  $(n-S)$  soil points is determined based on the 13 stable soil properties and converted to dissimilarity.

The geographic distance between the  $S$  random points and  $(n-S)$  soil points is calculated using an Euclidean measure of similarity based on the easting and northing. This ensures the range for both easting and northing are equal. A spatially weighted distance matrix is generated using an exponential variogram model with a range of 250m and nugget variance of 10 percent.

Using the spatially weighted distance matrix, a spatially weighted dissimilarity matrix between the  $S$  random points and  $(n-S)$  soil points is computed and converted to similarity. The final step allocates each of the  $S$  random points to the LMU group with which it has maximum similarity. This is regarded as the *correct* LMU to which the point should have been allocated and is compared to the LMU chosen in Part B using only ancillary data.

The above process is run in a loop  $c$  times and a frequency histogram showing the distribution of percent correct for  $c$  cycles is displayed. With  $c = 100$  and  $S = 50$  the results are displayed in the histogram in Figure 8.2.

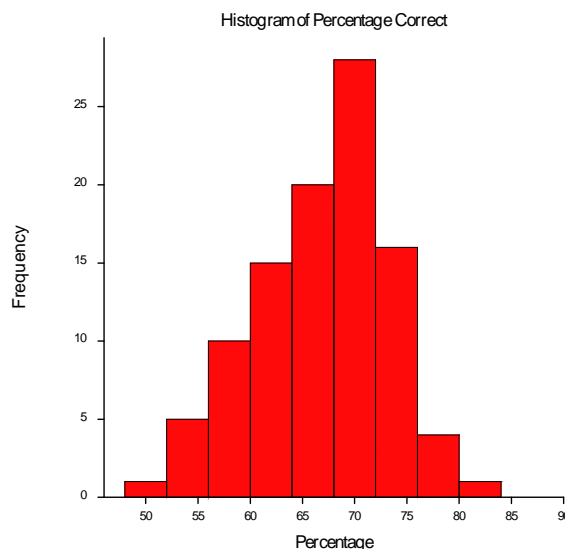


Figure 8.2 Histogram of the percentage of *correct* assignment of  $S$  (50) random points to LMUs for  $c$  (100) cycles

In the absence of soil properties, the ancillary data (landform, CTI, slope, RI) has the potential to assign points to the appropriate LMU. The results indicate a mode of 70 percent correct, which ranges from a minimum of 52 percent to a maximum of 82 percent. As such, 84 percent of the time, greater than 60 percent of the points are correctly allocated. This confirms that it is possible to define high resolution boundaries using ancillary data with reasonable accuracy.

An important point to note is that the allocation of points to LMUs as either correct or incorrect is rather harsh, as some of the LMUs have similar soil properties. While the histogram above indicates that greater than 15 percent of the time the number of points incorrectly allocated will be approximately 40 percent, many of these misallocations are to LMUs with properties that are similar to the *correct* LMU. A fuzzy comparison approach will be briefly discussed in Section 9.2.4 and is recommended here in order to include partial associations to LMUs with similar soil properties.

A natural comparison would be to assess the LMU group output when some points have been removed, with, the original LMU classification selected in Section 7.5. However, as soon as one or more soil points are removed the parameters associated with each original LMU change and therefore, points may be assigned to different LMUs. It is quite possible, for instance, that even though the LMUs are similar, they

may be re-ordered. Under these circumstances it is difficult to measure the correctness of the classification with regard to the original.

### 8.3 Sensitivity Analysis of LMU Classification

The impact of the different distance parameters and choice of method on the spatial variation in LMUs can be observed using the resulting LMU maps. These LMU maps, resulting from different parameters, were compared utilising the tools of the Map Comparison Kit (MCK) © (RIKS, 2005a). It is of interest to know spatially areas of agreement/disagreement that occur when comparing classification parameters.

MCK offers several comparison options including the Kappa statistic, which has been used here. The Kappa statistic which results from a pixel-by-pixel comparison between Map A and Map B, is dissected into two further statistics (Kappa Location and Kappa Histogram) that focus on different aspects of differences between the maps. Kappa Location (*KLoc*) measures similarity of location and Kappa Histogram (*KHisto*), measures similarity of quantity. *Location* refers to the spatial distribution of LMUs in Maps A and B and *histogram* refers to the frequency distribution of LMUs in the maps (RIKS, 2005b). The dissection of the kappa statistics, which the MCK utilises, results from Pontius' clarification (Pontius, 2000) of Klocation and Kquantity. However, Hagen (2002) defines Khisto as a replacement for Kquantity, which avoids many of the practical problems of Kquantity by scaling it using the Expected Fraction of Agreement,  $P(E)$  (Equation (8.2)).

The calculation of Kappa is based on the contingency table (confusion or error matrix) (Hagen, 2002), which details how pixels in each category in Map A are categorised in Map B (Table 8.1). Each row contains the fractions of the pixels in the map, which have been identified in a particular category in Map A and each category in Map B. For example, a value of 0.75 for cell  $P_{21}$  indicates that 75 percent of the mapped area is category 2 in Map A and category 1 in Map B. The last column gives the total fraction of pixels in each category in Map A. Since all categories together make up the whole map, the total sum of these fractions equals 1. Similarly, the last row gives the fraction of pixels in each category for Map B and must add up to 1.

		Map B categories				Total
		1	2	...	c	
Map A categories	1	$p_{11}$	$p_{12}$	...	$p_{1c}$	$p_{1T}$
	2	$p_{21}$	$p_{22}$	...	$p_{2c}$	$p_{2T}$
	⋮	⋮	⋮	...	⋮	⋮
	c	$p_{c1}$	$p_{c2}$	...	$p_{cc}$	$p_{cT}$
Total		$p_{T1}$	$p_{T2}$	...	$p_{Tc}$	1

Table 8.1. The contingency table in its generic form (redrawn from Hagen, 2002)

The following three statistics presented in Hagen (2002) can be calculated from the contingency table and are used here to further explain the calculation of Kappa and the dissection of Kappa into Klocation and Khisto.

$$P(A) = \sum_{i=1}^c p_{ii} \quad (8.1)$$

$$P(E) = \sum_{i=1}^c p_{iT} * p_{Ti} \quad (8.2)$$

$$P(\max) = \sum_{i=1}^c \min(p_{iT}, p_{Ti}) \quad (8.3)$$

$P(A)$  stands for the Fraction of Agreement (Equation (8.1)),  $P(E)$  is the Expected Fraction of Agreement subject to the observed distributions of categories in Maps A and B (Equation (8.2)) and  $P(\max)$  is the Maximum Fraction of Agreement subject to the observed distributions of categories (Equation (8.3))(Hagen, 2002).

Hagen (2002) defines the Kappa statistic  $K$  as follows;

$$K = \frac{P(A) - P(E)}{1 - P(E)} \quad (8.4)$$

Therefore, the Kappa statistic is the fraction of agreement  $P(A)$ , which is corrected for the fraction of agreement statistically expected from a random allocation of all cells in the map using the observed frequency distributions for Map A and Map B (Hagen, 2002). Klocation is calculated according to Equation (8.5) and Khisto according to Equation (8.6).

$$Klocation = \frac{P(A) - P(E)}{P(max) - P(E)} \quad (8.5)$$

$$Khisto = \frac{P(max) - P(E)}{1 - P(E)} \quad (8.6)$$

As such, K is the product of Klocation and Khisto (Equation (8.7)). A more detailed description can be found in Hagen (2002) which is based on, and further described in, Pontius (2000). Hagen (2002) indicates the relative merits of the Kappa statistic and points out that the Kappa statistic by itself offers insufficient information. This is because the value of 0.7 in one case may be considered very high, while in another case can indicate a poor result and, as such, suggest using a reference level of similarity via a reference map.

$$K = Klocation * Khisto \quad (8.7)$$

Using the MCK, the Kappa statistic separated into Klocation and Khisto has been calculated for all LMUs combined and per LMU.

### 8.3.1 Results from the Kappa Map Comparison

The Cluster/Cluster 250m map has been selected as the final LMU map in Chapter 7 and, as such, has been set as Map A and compared to all other LMU maps. In this case the Cluster/Cluster 250m map (Map A) can also be considered the reference map and the results for other maps are interpreted relative to each other. The resultant agreement/disagreements maps are shown in Figure 8.3 and a summary of the Kappa statistics and P(A) listed in Table 8.2, while the matrix for each map comparison are provided in Appendix Q.

The unequal areas of the agreement/disagreement maps (Figure 8.3) are scattered throughout the study area with the majority in all maps occurring in the south west of the farm where the sandier soils are known to be. Visually, the Discriminant 250m (Figure 8.3 d) is the most similar to the Cluster/Cluster 250m map.

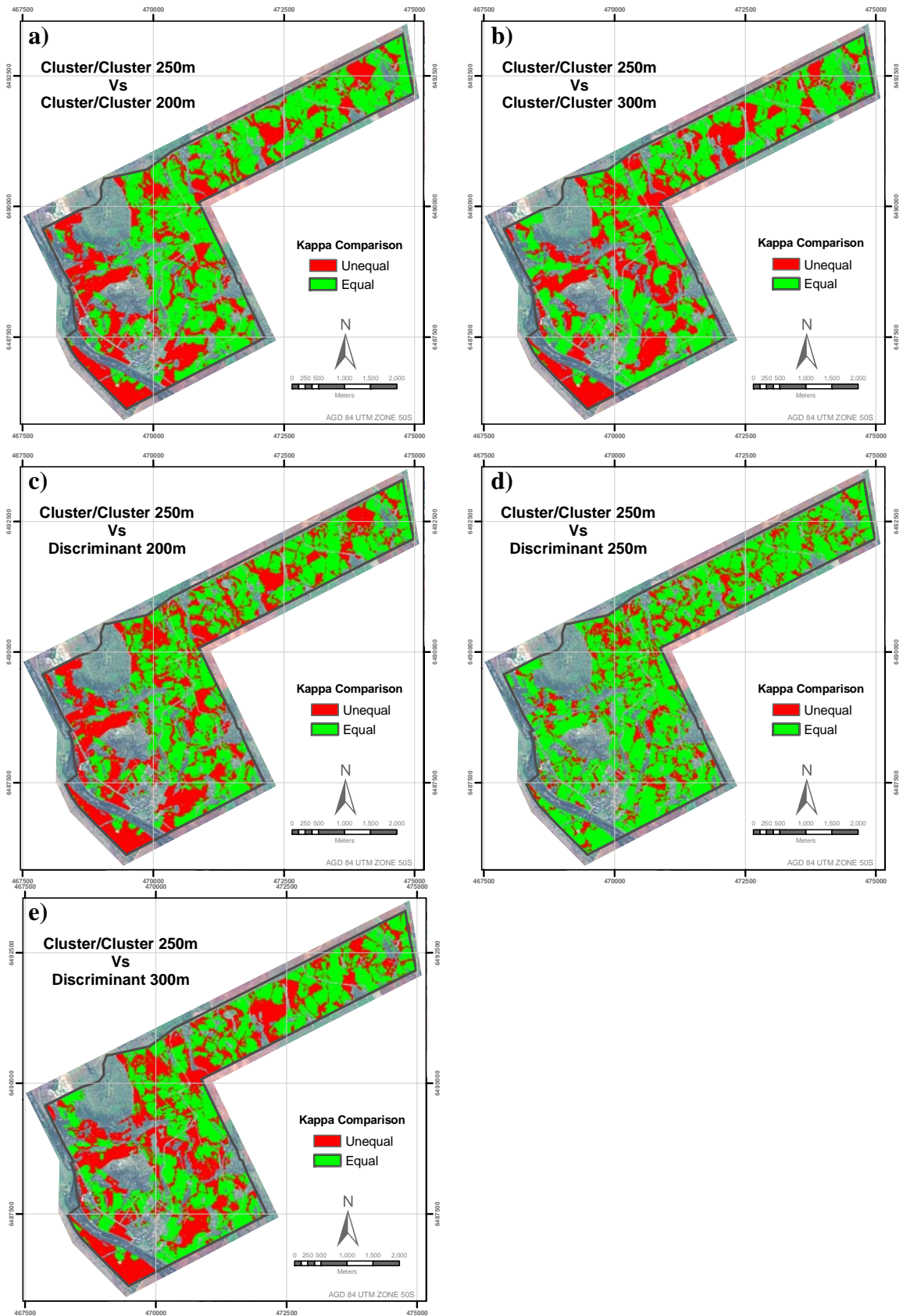


Figure 8.3 Kappa Map Comparison resultant agreement/disagreement maps for differing parameters a) Cluster/Cluster 200m b) Cluster/Cluster 300m c) Discriminant 200m d) Discriminant 250m e) Discriminant 300m

Table 8.2. The Kappa statistics; overall and for individual LMUs, used for comparison with the LMU map based on a Cluster/Cluster method with a range of 250m

P(A)	Cluster/Cluster 200m			Cluster/Cluster 300m		
	0.595			0.597		
	Kappa	KLoc	KHisto	Kappa	KLoc	KHisto
Overall	0.55	0.69	0.81	0.55	0.68	0.82
LMU 1	0.87	0.92	0.95	0.83	0.92	0.90
LMU 2	0.76	0.92	0.83	0.81	0.88	0.91
LMU 3	0.82	0.94	0.87	0.49	0.84	0.59
LMU 4	0.82	0.87	0.94	0.79	0.81	0.98
LMU 5	0.43	0.62	0.71	0.71	0.76	0.94
LMU 6	-0.01	-0.02	0.45	-0.01	-0.06	0.13
LMU 7	0.82	0.83	0.99	0.33	0.36	0.92
LMU 8	0.60	0.81	0.74	0.08	0.11	0.73
LMU 9	0.84	0.85	0.98	0.83	0.91	0.92
LMU 10	0.76	0.86	0.88	0.82	0.83	1.00
LMU 11	-0.08	-0.12	0.69	0.13	0.22	0.61

P(A)	Discriminant 200m			Discriminant 250m			Discriminant 300m		
	0.543			0.762			0.539		
	Kappa	KLoc	KHisto	Kappa	KLoc	KHisto	Kappa	KLoc	KHisto
Overall	0.49	0.58	0.85	0.74	0.81	0.91	0.49	0.60	0.82
LMU 1	0.79	0.80	0.99	0.80	0.86	0.93	0.74	0.82	0.90
LMU 2	0.66	0.75	0.88	0.77	0.82	0.94	0.72	0.77	0.93
LMU 3	0.65	0.69	0.95	0.75	0.76	0.98	0.47	0.71	0.66
LMU 4	0.65	0.82	0.78	0.70	0.86	0.81	0.67	0.79	0.85
LMU 5	0.34	0.39	0.86	0.76	0.82	0.94	0.62	0.67	0.93
LMU 6	0.01	0.02	0.59	0.76	0.76	1.00	-0.02	-0.04	0.52
LMU 7	0.72	0.73	0.98	0.76	0.76	0.99	0.29	0.31	0.92
LMU 8	0.56	0.62	0.89	0.57	0.67	0.85	0.06	0.09	0.71
LMU 9	0.76	0.80	0.95	0.79	0.82	0.96	0.75	0.81	0.93
LMU 10	0.69	0.72	0.96	0.76	0.76	0.99	0.72	0.80	0.90
LMU 11	-0.07	-0.10	0.67	0.65	0.87	0.74	0.14	0.24	0.57

#. KLoc refers to Klocation and KHisto refers to Khistogram.

In terms of the overall Kappa values, the discriminant method with a range of 250m provided the best results with a Kappa value of 0.74, which comprises Klocation of 0.81 and Khistogram 0.91. This highlights that the range parameter has had a greater influence on the distribution of LMUs over the methodology (i.e. Cluster/Cluster or Discriminant). For each comparison the Klocation value is at least 0.10 lower than the Khistogram. This indicates that the spatial location of the LMUs is not as reliable when compared to the frequency distribution of categories. The Cluster/Cluster 200m and Cluster/Cluster 300m are the next most similar with an overall Kappa of 0.55 in each case. Discriminant 200m and Discriminant 300m have an overall Kappa result of 0.49 in each case. This also suggests the strength of the range parameter over method in the comparisons.



Looking at the results for each LMU group comparison, the Klocation results are often lower than the Khistogram. In each case, this brings down the overall Kappa value. The results for each LMU group highlight that LMU 6 and LMU 11, and less often LMU 8, have very low Kappa values with the Klocation being the lowest in each case. Table 8.3 displays a summary of how some of the individual LMUs are mapped based on the results in Appendix Q.

Table 8.3 Summary of the individual LMU group comparisons that have not been mapped well ( $\geq 20$  percent to another LMU) and LMU legend ranking

LMU		Cluster/Cluster 250m						Total of Other LMUs	
		3	5	6	7	8	11		
Cluster/Cluster 200m	5		38%				37%	25%	A LMU 1
	6			15%			72%	13%	B LMU 3
	11			87%			4%	9%	C LMU 4
Cluster/Cluster 300m	6			0		28%	55%	17%	D LMU 7
	7	53%			37%			10%	E LMU 2
	8				30%	20%		50%	F LMU 10
Discriminant 200m						38%	17%	45%	G LMU 9
	5		36%				36%	28%	H LMU 5
	6			13%			70%	17%	I LMU 8
Discriminant 250m	11			62%			5%	33%	J LMU 6
	Mapped well								K LMU 11
	6				<1%		22%	57%	20%
Discriminant 300m	7	44%			33%			23%	
	8				30%	18%		52%	
	11					44%	17%	39%	

LMUs 6 and 11 are the least well mapped followed by LMUs 5,7 and 8. Depending on the method and variogram range selected, pixels in LMU 11 often move to LMUs 5 and 6, and pixels in LMU 6 move to LMU 11 with corresponding changes in the definition of map units. LMU 6 has been mapped as LMU 11 72 percent in the Cluster/Cluster 200m map and 55 percent in the Cluster/Cluster 300m map along with 70 percent in the Discriminant 200m map and 57 percent in the Discriminant 300m map (Table 8.3). LMU 5, 8, 6 and 11 are the sandier soils and have been ranked similarly as shown in the final LMU map (Figure 7.14) and reproduced legend above (Table 8.3). It appears that when using the different classification techniques and ranges, the boundaries (both spatial and in terms of their soil properties) between LMUs with similar soil properties are likely to change.

#### 8.4 Relationship between LMUs and Yield

The LMU to which each soil point has been assigned was added to the Yield Data Set 1 (detailed in Section 5.7.2). A multiple regression model incorporating crop type, year effect and LMU was then used to determine whether yield varied significantly between LMUs and to estimate a yield for each LMU. The regression model fitted was as follows;

$$Y_{ijkl} = \mu + C_j + Yr_k + CYr_{jk} + LMU_l \quad (8.8)$$

where  $Y_{ijkl}$  is the average interpolated yield for sampling point  $i$  in year  $k$  with crop type  $j$  and LMU  $l$ ;  $\mu$  is the intercept;  $C_j$  is the effect of crop  $j$ ;  $Yr_k$  is the effect of year  $k$ ;  $CYr_{jk}$  is the effect of the interaction between crop  $j$  and year  $k$ ;  $LMU_l$  is the effect of LMU  $l$ .

Statistical assumptions of normality, independence and constant variance underlying regression analysis were examined using plots of residuals. The plot of residual against fitted values indicated that the variance of the residuals increases with mean yield. While this is of some concern, it was overcome by transforming the yield data using a natural log transformation. Subsequent models incorporated the log Yield data and took the form;

$$\log_e(Y_{ijkl}) = \mu + C_j + Yr_k + CYr_{jk} + LMU_l \quad (8.9)$$

Results of the model fitting process have been summarised in an analysis of variance table with the effect of each term in the model reported after removing the effects of the terms above it in the table (Table 8.4). There was a highly significant difference in yield for LMUs after adjusting for effects of crop type and year indicating there will be significant differences in yield between some LMUs. The variance between LMUs is over 3 times the unexplained variance within LMUs as indicated by the variance ratio (3.07). The model accounted for 43.9 percent of the variance in yield and the residual standard error of observations is 0.228.

Table 8.4 Accumulated analysis of variance for regression model in Equation (8.9).  
The model accounted for 43.9 percent of the variance in yield and the residual standard error of observations is 0.228.

Factor	Degrees of freedom	Sum of squares	Mean square	Variance Ratio	F. probability significance
Crop Type	6	16.04066	2.67344	51.58	<.001
Year	6	8.23235	1.37206	26.47	<.001
CropYear	22	4.37249	0.19875	3.83	<.001
LMU	10	1.59152	0.15915	3.07	<.001
Residual	646	33.48137	0.05183		
Total	690	63.71839	0.09235		

From the regression model, yield was predicted for each LMU (Table 8.5). These values are comparable because differential effects of year and crop type have been eliminated.

Table 8.5 Predicted Yield for each LMU based on the regression model

Yield kg/ha	LMU	Texture group from final LMU classification	Approximate Clay Content %
1093.10	8	Sand A-horizons (1-5cm) over loamy sand	<5% over about 5%
1296.63	11	Sand A-horizons (10-15cm) over sand	<5% over <5%
1391.32	4	Sandy loam A-horizons (5-12cm) over sandy clay loam to light medium clay	10-20% over 20-45%
1466.36	5	Loamy sand A-horizons (10-15cm) over clayey sand	About 5% over 5-10%
1511.66	6	Loamy sand A-horizons (10-20cm) over clayey sand	About 5% over 5-10%
1549.62	2	Clayey sand A-horizons (8-14cm) over sandy loam	5-10% over 10-20%
1550.13	1	Sandy loam A-horizons (7-13cm) over sandy clay loam to light medium clay	10-20% over 20-45%
1560.10	10	Sand with A-horizon depth (6-11cm)	<5%
1588.41	3	Clayey sand A-horizons (<3cm) over sandy clay loam to medium clay	5-10% over 20-55%
1591.26	7	Clayey sand A-horizons (<3cm) over clayey sand	5-10% over 5-10%
1636.48	9	Loamy sand A-horizons (10-15cm) over clayey sand	About 5% over 5-10%

Looking at the soil properties that are associated with each LMU from the final soil map (Table 8.5), it is found, in general, that the very low clay content top soil and sub soils are associated with the lower yield values. The difference in yield between

LMUs has a maximum of just less than 600kg. At this level a farmer would consider alternative management in view of increased productivity if a cost benefit can be shown. However, it would not be worth while managing LMUs differently with only a 1kg/ha change (i.e. LMU 1 and 2) and in this case a farm manager may choose to manage these units in the same way. However, providing the results in such detail gives the farmer a decision management tool.

### **8.5 Summary**

Analyses of the success of the LMU classification in this chapter have resulted in three conclusions:

- a) the method of allocating points/pixels to a LMU based on ancillary data has resulted in the correct allocation of 60 percent of points over 80 percent of the time;
- b) the LMU classification is not very sensitive to the change in variogram range or alternative techniques, except for a small number of individual LMUs (i.e. 6 and 11) which are less well discriminated in terms of their soil properties; and
- c) there is a relationship between the LMUs formed from the classification procedure and yield, giving farmers an opportunity to look at different management practices based on the LMU classification.

This LMU classification incorporating soil information from a small number of sample points and high resolution ancillary data has been successful. The implications of these results are discussed further in the following Chapter 9.

## CHAPTER 9

### CONCLUSIONS AND RECOMMENDATIONS

*The main objective of this research has been to develop a framework for classifying a farm into homogeneous LMUs that can be used in precision agriculture applications. The framework has been described in Chapter 7 and the success of the classification technique analysed in Chapter 8.*

*The principles underlying the LMU framework have been addressed in Chapters 4, 6 and 7. The main conclusions based on the findings of this research and recommendations for further research are presented below.*

#### **9.1 Conclusions**

This research has provided a framework for creating LMUs. The process starts with the compilation of data sets and finishes with the production of a final map of soil properties relevant to site-specific crop management. The strength of the framework lies in the integrity of each of the steps (aims) which make up the process. These aims are reiterated and dealt with individually below.

##### 9.1.1 Determine Stable Soil Properties that have the most influence on Yield Variability in the Agricultural belt of WA

A set of 11 stable soil properties (Table 9.1) were established from the 24 soil properties measured by examining the relevant literature and using statistical methods to select those having the greatest effects on yield.

No conclusive subset of soil properties could be identified from a regression analysis between yield and soil properties. At best, only just over 50 percent of the variance in yield in 2003 was accounted for using the full set of soil properties.

Table 9.1 Stable soil properties and variables from literature review

Property	Variable
Topography	landform, CTI, slope
Texture	Percentage of sand, silt and clay
Coarse fragments	Percentage of stones
Organic matter	Percentage of Organic Carbon
Capacity of soil to store nutrients cations	Effective cation exchange capacity (ECEC)
Acidity/Alkalinity	pH
Salinity	EC
Sodicity	Exchangeable sodium percentage (ESP)
Soil stability	Ca:Mg ratio
Depth of topsoil	Depth of A-Horizon
Depth of restricted layer	Depth of compaction or impeding stones

Principal components analysis was used to reduce the initial set of 24 soil properties to a subset of 13. It aided in the identification of relationships between soil properties and the soil properties that explain variance between sampling points. This approach would be useful in diverse environments with large multivariate data sets or scenarios that do not have the luxury of historical yield data.

#### 9.1.2 Employ High Resolution, Readily Available, Cheap Ancillary Data Sets that are related to Soil Properties and/or Landscape Variability


The research focuses on the appropriateness of the ancillary data which, on the basis of utility, have been constrained to meet a number of criteria. These criteria do not limit the framework to this particular set of data. The criteria are:

- i. *High Spatial Resolution* - the spatial resolution of differing remote sensing systems range from sub-metre (0.5m) to 1000m, with high resolution referring to the sub-metre end of the scale. Jensen (1996a) offers a useful rule that the spatial resolution of a system should be less than half the size of the feature measured in its smallest dimension. Since the intention of this research was to delineate LMUs at within paddock scales (with an average paddock size of greater than 35ha), a spatial resolution less than or equal to 10m is considered appropriate.
- ii. *Readily available to farmers* - the data sets can be easily purchased from a government organisation or private company, and are not exclusive to any particular study region or limited study area. This criterion limits the process to data sets that are not historical in nature, i.e. there is no requirement that a farmer has collected data prior to the project.

- iii. *Low cost* - it is difficult to stipulate what constitutes low cost ancillary data. Low cost is a relative statement and depends on the funds available for a project. What may be considered high to one user, might not to another. However, this research undertakes a comparison of the commercial cost of the data sets used, against other high resolution data sets widely available. This is illustrated in Table 9.2, which shows that the ancillary data used in this research are relatively low cost.

Table 9.2 Comparison of the cost of some high resolution data sets January 2007

<b>Ancillary Data</b>	<b>Products</b>	<b>Resolution</b>	<b>Source</b>	<b>Total Cost based on 3000ha</b>	<b>Cost/ha \$AUD</b>
DEM	Landform, CTI, Slope	0.01m vertical +/-1.5m accuracy 10m spatial	Department of Agriculture-WA	\$300	\$0.10
DEM (with EM38 inclusive)	DEM	0.02m vertical 2m spatial	Precision Agronomics	\$33,000	\$11.00
DEM and EM38	DEM	0.02m vertical 20m swath transects	Australian Centre of Precision Agriculture	\$22,500	\$7.50
	EC soil variability	20m swath transect			
EM38	EC soil variability	35m swath transect interpolated onto 5m spatial	Precision Agronomics	\$19,500	\$6.50
DMSI	B, G, R, NIR 450-780nm RI, NDVI	1.5m spatial	SpecTerra Services	\$6,900	\$2.30
HyMap	0.45-2.5um	3-10m spatial	HyVista Corporation	\$5,000- \$10,000	\$1.60 - \$3.35
Quickbird	B&W B,G,R,NIR	0.65m 2.5m	Sinclair Knight Merz	\$3000	\$1.00
Radiometrics	Potassium (K), Thorium (Th) Uranium (U) Total Counts	25m spatial	Fugro Airborne Surveys	\$4950 (excluding mobilization cost)	\$1.65

 data sets used on this research highlighted in grey

Other elevation data sources are approximately 10 times more expensive than the DEM used in this case, although it is acknowledged that they have higher spatial and vertical resolution. The DMSI used in this research is similar in cost to other

high resolution multi-spectral data i.e. Hymap. The Quickbird imagery appears to be another option that future projects may investigate as it is half the cost of the DMSI used in this case. However, Quickbird imagery is captured from satellite; therefore the data must be captured during the satellite pass and additional charges may apply. EM38 data, which is used to detect EC soil variability, is almost 3 times more expensive than DMSI, while radiometrics, which have been used to detect site-specific management zones (Florin *et al.*, 2005), are comparable in cost but have a coarser spatial resolution (25m). From Table 9.2 it can be seen that the elevation data used in this research is low cost in comparison to other sources, while the DMSI is similar to other high resolution multi-spectral data sets.

- iv. *Related to soil properties and/or landscape variability* – vegetation indices (VI) were generated from the DMSI and analysed in relation to *in situ* crop attributes to test the hypotheses that VIs could be used as high resolution inputs to the LMU classification. It was concluded that although the VIs captured the spatial variation in crop growth, they were not suitable as inputs to the LMU classification in this case. The results of this analysis have been discussed in Chapter 4 and further concluded below (Section 9.1.7).

A correlation analysis between the ancillary variables CTI, slope, RI and the 13 selected soil properties was performed in Chapter 6, demonstrating that all these ancillary variables were correlated with some of the soil properties. An analysis of variance of the selected soil properties with respects to landforms showed significant differences between either primary landforms, slope classes or both, for nine of the 13 selected soil properties (Chapter 6) indicating that the landforms contain information about soil properties.

The justification of the use of the ancillary data (landforms, CTI, slope, RI) has also been demonstrated in Chapter 8. A validation process was performed in which pixels with known soil properties were allocated to a LMU based on their ancillary data only. Sixty percent of the pixels were allocated correctly over 80 percent of the time.



### 9.1.3 Develop a Methodology to combine information derived from High Resolution Data Sets and Soil Properties at Point Locations

The framework for classifying LMUs is a spatially weighted multivariate classification procedure that combines soil properties sampled at low resolution and high resolution ancillary data. The two stage methodology described in Chapter 7 forms clusters that are spatially contiguous and similar in terms of their multivariate data and is computationally feasible for large data sets. The classification incorporates the scale and form of the spatial variation of the soil variables by spatially weighting the dissimilarities between locations via a geostatistical function. The process may be adjusted at several steps to suit other study areas, scale, input variables and applications.

The success of the LMU classification was addressed in Chapter 8 by showing that: a) the output LMUs show significant differences ( $P < 0.001$ ) in actual yield, b) the LMU classification is not very sensitive to changes in parameters (e.g. variogram range), and c) the ancillary data satisfactorily defined high resolution LMU boundaries associated with stable soil properties.

It has been clearly shown that, by utilising the two stage approach, a spatially weighted multivariate classification technique may now be applied to high resolution data sets. The expense of data acquisition can be kept low by supplementing high cost soil data with lower cost ancillary data which is related to the stable soil properties influencing plant growth and appropriate for defining high resolution boundaries.

### 9.1.4 Produce a LMU Map with associated Soil Properties at Paddock Scale

A LMU map, with a description of the stable physical and chemical soil properties (e.g. texture, pH) associated with each LMU, has been produced (Figure 7.14) for the Muresk farm. Several LMUs make up each paddock, and the output can be given to farmers for use in site specific management.

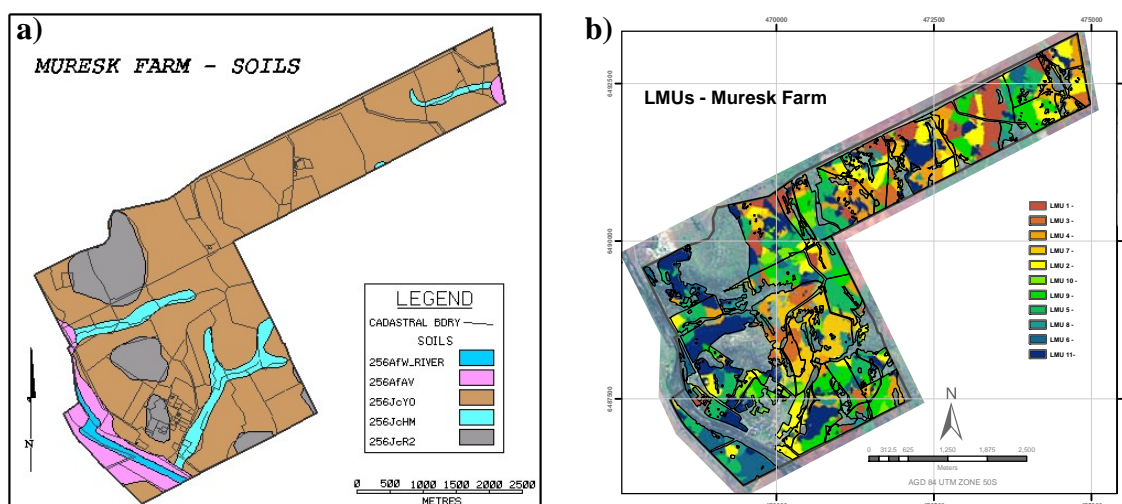


Figure 9.1 (a) Muresk soil map versus (b) LMU map generated in this research. Legends of map units are provided in Chapter 3 and 7 respectively

Figure 9.1 a) is the current Muresk soils map, while Figure 9.1 b) shows the map of LMU produced in this research. The comparison illustrates the greater detail of the LMU map. The current Muresk soils map has only five classes, one of which is the Avon flat wet, river (256AfW\_RIVER) and another non arable, Steep Rocky Hills (256JcR2). In essence there are only three soil classes with the Jelcobine York unit (256JcYo) dominating the arable farming area. This renders the Muresk soils map unsuitable for farm management applications although it is useful at regional scales.

The new LMU map has subdivided the arable area of the paddocks into 11 LMUs enhancing the detail of information on the soil properties, and providing a product that could be used for site specific precision agriculture. It is envisaged that other land use scenarios and optimisation approaches to farming could also be based on the new LMU map.

Multivariate methods for discriminating between LMUs and colouring map units (doubled-ended colour plan (Dent, 1999)) according to progressive changes in soil properties has overcome the issue of a non-hierarchical output (Chapter 7). The non-hierarchical cluster analysis used for the LMU classification has been suggested as most appropriate when classifying objects that are not hierarchically structured, such as soil (Oliver and Webster, 1989). The corollary of this is an output that is not graduated. The use of discriminant analysis to identify the soil properties that

differentiate the LMUs, then using the median value of the first canonical variate for each LMU to rank them in the map legend, has overcome this problem and provides a process that could be adopted by other users in a similar predicament.

#### 9.1.5 Validate the Methodology for forming LMUs

The methodology for forming LMUs has been validated in Chapter 8. Three themes have been addressed; a) points will be allocated to the correct LMU using only ancillary data, b) the LMU classification is insensitive to changes in parameters and c) there is a relationship between LMUs and yield.

- a) Subsets of 50 randomly chosen soil points from the total sample of 250 soil points have been allocated to LMUs as if they had only ancillary data (landform, CTI, slope, RI). Using the known values for the 13 selected stable soil properties they have been compared to other soil points in the LMUs to which they have been assigned. The process highlighted that in the absence of soil properties the ancillary data can be used to assign points to the appropriate LMU with greater than 60 percent of the points being allocated correctly 84 percent of the time. This has confirmed that it is possible to define high resolution boundaries using ancillary data with reasonable accuracy.
- b) LMU maps were produced using differing parameters (variogram range and classification methodology) and the resulting maps were compared using the Kappa statistic within the Map Comparison Kit © (RIKS, 2005a). The results indicated that while the variogram range had a greater influence on the distribution of LMUs than the classification methodology, the LMU classification is not very sensitive to changes in either of these parameters. A small number of LMUs which were less well discriminated in terms of their soil properties appeared to be the major contributor to spatial differences between the maps.
- c) An analysis of variance of potential yield with respect to each LMU showed a highly significant difference in yield between LMUs ( $P < 0.001$ ). In general, it was found that the top soils and sub soils with very low clay content were

associated with the lower yields. A difference in potential yield of approximately 500kg/ha was shown between soils with high and low clay contents.

The three validation procedures performed confirm that the LMU classification has been successful.

#### 9.1.6 Develop an Effective Field Soil Sampling Strategy

The soil sampling strategy developed in this research used a GIS based approach for assessing soil variability. The strategy utilises the RI and NDVI spectral indices (derived from DMSI) in combination with slope classes (derived from DEM) as a surrogate layer to identify soil variability based on their relationship with soil properties. The location of soil sampling points is intensified in areas of heterogeneous soil types, and less concentrated in homogeneous areas.

The approach is effective, more strategic (i.e. based on variability rather than random or grid based placement of soil observations) and uses low cost procedures which combine basic image processing and spatial modelling techniques. The strategy is easy to implement in any GIS and offers an alternative to the more sophisticated and expensive field surveys that account for spatial variability of soil properties within agricultural fields.

#### 9.1.7 Determine the Opportunities and Limitations of using High-Resolution Digital Multi-spectral Imagery as a diagnostic tool for monitoring Crop Growth during the Growing Season

An analysis between the four individual bands and VIs derived from DMSI with crop growth attributes across two years was performed in Chapter 4. The analysis has demonstrated the opportunity for high resolution DMSI to detect seasonal crop growth variability at within paddock scale, and as such may be an appropriate remote sensing tool for further use with site-specific crop management applications. Field scouting still needs to be undertaken to provide an indication of the cause and effect of relationships.

Three commonly grown crops in the WA wheatbelt, Wheat, Lupins and Canola, were correlated with LAI measurements. An initial view of the linear correlation coefficients showed somewhat poorer results than expected, particularly in terms of consistency across the two years. As the NDVI is the most common VI used, a more in depth analysis was carried out between the NDVI and LAI for the three crop types.

A plot of the relationship between NDVI and all crop types LAI (Figure 4.13) showed the diminishing change in NDVI at values of LAI above 2.0. This concurred with Sellers' (1985) findings and helped explain the weaker than expected linear correlations. There was a significant correlation between the NDVI and Wheat LAI in the year 2003 (0.76;  $P \leq 0.01$ ), but a poor correlation in the year 2002 (0.13) which could be attributed to the effect of a weed infestation and crop growth stage. The correlations between NDVI and Lupin LAI was significant in 2002 (0.50;  $P \leq 0.01$ ) and low in 2003 (0.28). The low correlations for Lupins case were attributed again to crop growth stage, leaf layering and saturation effect that occurs at a LAI of 2.0 and 3.0 and beyond. For the canola crop, canopy saturation, the reflectance of flowers, leaf position and wilting, which are all associated with the crop growth stage, have been identified as the causes of lower correlations (0.19 and 0.42;  $P \leq 0.05$  for the year 2002 and 2003 respectively). An analysis between the canola petal area index (PAI) and individual DMSI bands indicated that when the PAI reaches a certain threshold, it hinders the reflectance of the leaves.

## 9.2 Recommendations

The research has thoroughly investigated the proposed objectives and has provided a framework for determining LMUs that can be used in precision agriculture applications. During the implementation of this research some limitations have been identified and the following recommendations are proposed for future work.

### 9.2.1 Data Sources

When producing primary and secondary topographic attributes from a DEM it is recommended to ensure that the DEM is "*hydrologically sound*" (Ryan, 2003), in particular for hydrological applications, and the use of such products as ANUDEM (Hutchinson, 1989) to form these. When deriving the CTI and landforms a

*depressionless* DEM was formed by filling pits, ponds and depressions which was considered to be “*hydrologically sound*”, thus transforming a readily available data set. Future works may address this issue in more detail, comparing outputs from ANUDEM with those used within this study.

### 9.2.2 Opportunities of DMSI

The study presented in Chapter 4 has identified possibilities for further use of DMSI.

Aspects that should be considered are:

- Examine the relationship between canola petal (flower) reflectance and final yield. If a strong relationship exists, then DMSI data captured during the flowering stage of canola growth may indicate that DMSI could be used for within season models of final canola yield;
- Examine whether the relationship between LAI and DMSI to can be improved using non-linear regressions;
- Compare high resolution DMSI and broader scale remote sensing systems (i.e. 30m spatial resolution) for depicting large scale variations, taking into account the cost differences.

### 9.2.3 GIS Programming

The successful implementation of the spatially weighted multivariate classification has been programmed in Genstat ® using GIS data sets derived in ArcGIS and Intergraph GeoMedia. There is potential to program the LMU classification in GIS or remote sensing software. This would provide software users with the ability to perform spatially weighted multivariate classification more simply by avoiding the tireless task of extracting and importing data between software. The LMU classification could also be implemented in the statistical package R which is free to all users.

### 9.2.4 Validations techniques

Several validation processes have been implemented to analyse the success of the LMU classification technique presented in Chapter 8, further options have been identified and are suggested hereafter.

While the ancillary data are doing a good job of assigning pixels to LMUs in the absence of soil properties (Section 8. 2), assessing the result as either correct or incorrect (e.g. a Boolean type assessment) is not informative, as some of the LMUs have similar soil properties. It is therefore recommended that a fuzzy method (e.g. based on the principles of fuzzy logic proposed by Zadeh (1965)) would be more appropriate where pixels would be given a probability of being in a number of LMUs including the correct LMU. The Map Comparison Kit (RIKS, 2005a) offers a fuzzy map comparison set which should be investigated.

The LMUs show significant differences in actual yield after adjusting for crop type and year effects (Section 8. 4). It would be interesting to assess the merits of spatial and temporal trend maps that Blackmore (2003) provides (Section 2.4.1), and analyse them in relation to the LMUs to determine whether the LMUs are spatially similar to areas that can be considered stable in time, based on yield data.

To test the hypothesis that the LMUs are an improvement on current paddock boundaries, the LMUs would need to be adopted and tested on-farm over several years (due to the nature of dry land agriculture and climatic effects) and analysed against existing paddock records. Such research would require a large amount of resources and the willingness of a farmer to offer a study area. Promisingly, in discussion with a local farm consultant group (Farmanco Management Consultants) a proposal is being put together to assess the commercial costs and quantitative benefits of the LMU framework developed in this research.

## REFERENCES

- Allen, A. and Beetson, B. (1999) The Land Monitor Project: A Multi-agency project of the Western Australian Salinity Action Plan supported by the Natural Heritage Trust, *Proceedings of WALIS Forum 1999*, Perth, WA, Australia.
- Allen, D. (2004) *Personnel Communication*, Principal Chemist, Natural Resources Chemistry Laboratory, Chemistry Centre (WA), 125 Hay Street, East Perth, WA 6004.
- Alphen, B.J. and Stoorvogel, J.J. (1998) A Methodology to Define Management Units in Support of an Integrated, Model-Based Approach to Precision Agriculture, *Proceedings of 4th International Conference on Precision Agriculture*, St. Paul, Minnesota, USA, 19-22 July, pp. 1267-1278.
- Aspinal, D. (2000) *Infosheet, Precision Farming, Ministry of Agriculture and Food, Ontario*, [online]:  
<http://www.gov.on.ca/OMAFRA/english/environment/precision>, accessed (2002).
- Atherton, B.C., Morgan, M.T., Shearer, S.A., Stombaugh, T.S. and Ward, A.D. (1999) Site-Specific Farming: A Perspective on Information Needs, Benefits and Limitations, *Journal of Soil and Water Conservation*, Vol. 54, No. 2, pp. 455-461.
- Atkinson, P.M. (2004) Spatially weighted supervised classification for remote sensing, *International Journal of Applied Earth Observation and Geoinformation*, Vol. 5, No. 4, pp. 277-291.
- Australian Centre for Precision Agriculture (2006) *Precision Agriculture*, [online]:  
<http://www.usyd.edu.au/su/agric/acpa/pag.htm>, accessed (17/11/06).
- Basnet, B., Apan, A., Kelly, R., Jensen, T., Strong, W. and Butler, D. (2003) Delineation of Management Zones using Multiple Crop Yield Data, *Proceedings of 16th Triennial Congress of the International Soil Tillage Research Organisation (ISTRO)*, University Of Queensland, Brisbane, 13-18 July.
- Bauer, M.E. (1985) Spectral Inputs to Crop Identification and Condition Assessment, *Proceedings of the IEEE*, Vol. 73, pp. 1071-1085.
- Baumgardner, M.F., Silva, L.F., Biehl, L.L. and Stoner, E.R. (1985) Reflectance Properties of Soils, *Advances in Agronomy*, Vol. 38, pp. 1-44.
- Ben-Dor, E., Irons, J. and Epema, G. (1999) Soil reflectance, in: *Remote Sensing for the Earth Sciences*, Rencz, A. M. (ed.), J Wiley & Sons, New York, USA, pp. 111-188.



- Bevan, K.J. and Kirkby, M.J. (1979) A physically based, variable contributing area model of basin hydrology, *Hydrological Science Bulletin*, Vol. 24, No. 1, pp. 43-69.
- Bigham, J.M., Golden, D.C., Buol, S.W., Weed, S.B. and Bowen, L.H. (1978) Iron Oxide Mineralogy of Well-drained Ultisols and Oxisols: II. Influence on Color, Surface Area, and Phosphate Retention, *Soil Science Society of America journal*, Vol. 42, pp. 825-830.
- Blackmore, B.S., Godwin, R. and Fountas, S. (2003) The analysis of spatial and temporal trends in yield map data over six years, *Biosystems Engineering*, Vol. 84, No. 4, pp. 455-466.
- Blake, J. and Huffer, A. (2001) Managing spatial variability in our agricultural system, *Study report incorporating the 'Precision Farming Forum' and outcomes from the Industry workshop, July 2001*, Technology Park: Bentley, WA.pp.
- Boer, M., Del Barrio, G. and Puigdefábregas, J. (1996) Mapping soil depth classes in dry Mediterranean areas using terrain attributes derived from a digital elevation model, *Geoderma*, Vol. 72, pp. 99-118.
- Bourgault, G., Marcotte, D. and Legendre, P. (1992) The Multivariate (Co)Variogram as a Spatial Weighting Function in Classification Methods, *Mathematical Geology*, Vol. 24, No. 5, pp. 463-478.
- Bowers, S.A. and Hanks, R.J. (1965) Reflection of radiant energy from soils, *Soil Science*, Vol. 100, pp. 130-138.
- Boydell, B. and McBratney, A.B. (1999) Identifying potential within-field management zones from cotton yield estimates, in: *Precision Agriculture '99, Proceedings of the 2nd European Conference on Precision Agriculture*, Stafford, J. V. (ed.), Sheffield Academic Press, Sheffield, UK, pp. 331-341.
- Brady, N.C. and Weil, R.R. (1999a) *The Nature and Properties of Soils*, (12th edn), Prentice Hall, Upper Saddle River, New Jersey, USA. 881 pp.
- Brady, N.C. and Weil, R.R. (1999b) Soil Architecture and Physical Properties, in: *The Nature and Properties of Soils*, (12th edn), Prentice Hall, Upper Saddle River, New Jersey, USA, pp. 117-170.
- Brady, N.C. and Weil, R.R. (1999c) Soil Organic Matter, in: *The Nature and Properties of Soils*, (12th edn), Prentice Hall, Upper Saddle River, New Jersey, USA, pp. 446-490.
- Bruce, R.C. (1999) Calcium, in: *Soil Analysis: an Interpretation Manual*, Peverill, K. I., Sparrow, L. A. & Reuter, D. J. (eds.), CSIRO Publishing, Collingwood, Vic, Australia, pp. 247-254.

- Burrough, P.A. (1996) *Principles of geographical information systems for land resources assessment*, Clarendon Press, Oxford, UK. 194 pp.
- Burrough, P.A. and McDonnell, R.A. (1998a) Creating Continuous Surfaces from Point Data, in: *Principals of Geographic Information Systems*, Burrough, P. A., Goodchild, M. F., McDonnell, R. A., Switzer, P. & Worboys, M. (eds.), Oxford University Press, Oxford, pp. 98-131.
- Burrough, P.A. and McDonnell, R.A. (1998b) Fuzzy Sets and Fuzzy Geographical Objects, in: *Principals of Geographic Information Systems*, Burrough, P. A., Goodchild, M. F., McDonnell, R. A., Switzer, P. & Worboys, M. (eds.), Oxford University Press, Oxford, pp. 265-297.
- Burrough, P.A. and McDonnell, R.A. (1998c) Optimal Interpolation Using Geostatistics, in: *Principals of Geographic Information Systems*, Burrough, P. A., Goodchild, M. F., McDonnell, R. A., Switzer, P. & Worboys, M. (eds.), Oxford University Press, Oxford, pp. 132-161.
- Burrough, P.A., van Gaans, P.F.M. and Hootsmans, R. (1997) Continuous classification in soil survey: spatial correlation, confusion and boundaries, *Geoderma*, Vol. 77, pp. 115-135.
- Caccetta, P.A. (2000) Technical note-A simple approach for reducing discontinuities in digital elevation models (dems), *Report Number 2000/231*, CSIRO Mathematical and Information Sciences, Wembley, Perth, Western Australia, 6 pp.
- Caeiro, S., Goovaerts, P., Painho, P. and Costa, H. (2003) Delineation of Estuarine Management Areas Using Multivariate Geostatistics: The Case of Sado Estuary, *Environmental Science and Technology*, Vol. 37, No. 18, pp. 4052-4059.
- Campbell, K.O. and Bowyer, J.W. (1990) *The Scientific Basis of Modern Agriculture*, Sydney University Press, Sydney, Australia. 479 pp.
- Cass, A. (1999) Interpretation of some Soil Physical Indicators for Assessing Soil Physical Fertility, in: *Soil Analysis: an Interpretation Manual*, Peverill, K. I., Sparrow, L. A. & Reuter, D. J. (eds.), CSIRO Publishing, Collingwood, Vic, Australia, pp. 95-102.
- Casseltes, V. and Garcia, M.J.L. (1989) An alternative simple approach to estimate atmospheric correction in multitemporal studies, *International Journal of Remote Sensing*, Vol. 10, pp. 1127-1134.
- Cihlar, J., Dobson, M.C., Schmugge, T., Hoogeboom, P., Janse, A.R.P., Baret, F., Guyot, G., Le Toan, T. and Pampaloni, P. (1987) Procedures for the description of agricultural crops and soils in optical microwave remote sensing studies, *International Journal of Remote Sensing*, Vol. 8, No. 3, pp. 427-439.

- Clevers, J.G.P.W. (1997) A simplified Approach for Yield Prediction of Sugar Beet Based on Optical Remote Sensing Data, *Remote sensing of Environment*, Vol. 61, pp. 221-228.
- Cloutis, E.A., Connery, D.R. and Dover, F.J. (1999) Agricultural crop monitoring using airborne multi-spectral imagery and C-band synthetic aperture radar, *International Journal of Remote Sensing*, Vol. 20, No. 4, pp. 767-787.
- Cloutis, E.A., Connery, D.R., Major, D.J. and Dover, F.J. (1996) Airborne multi-spectral monitoring of agricultural crop status: effect of time of year, crop type and crop condition parameter, *International Journal of Remote Sensing*, Vol. 17, No. 13, pp. 2579-2601.
- Coakes, S.J. and Steed, L.G. (2001) *SPSS Analysis without Anguish: version 10.0 for Windows*, John Wiley & Sons, Milton, Qld, Australia. 266 pp.
- Cochran, W.G. (1959) *Sampling Techniques*, John Wiley & Sons Inc, New York, USA. 330 pp.
- Coleman, T. and Montgomery, O. (1987) Soil moisture, organic matter and iron content effect on the spectral characteristics of selected vertisols and alfisols in Alabama, *Photogrammetric Engineering & Remote Sensing*, Vol. 53, pp. 1659-1663.
- Condit, H.R. (1970) The spectral reflectance of American soils, *Photogrammetric Engineering & Remote Sensing*, Vol. 36, pp. 955-966.
- Cook, S.E., Corner, R.J., Grealish, G., Gessler, P.E. and Chartres, C.J. (1996) A Rule-based System to Map Soil Properties, *Soil Science Society of America Journal*, Vol. 60, pp. 1893-1900.
- Coops, N.C., Gallant, J.C., Loughhead, A.N., Mackey, B.J., Ryan, P.J., Mullen, I.C. and Austin, M.P. (1998) Developing and testing procedures to predict topographic position from Digital Elevation Models (DEM) for species mapping (Phase I), Report to Environment Australia, CSIRO Forestry and Forest Products: Client Report No: 271, Canberra, 56 pp.
- Corner, R.J. (1999) Knowledge representation in Geographic Information Systems, *Ph.D Thesis*, School of Spatial Sciences, Curtin University of Technology, Perth, WA, 220 pp.
- Corner, R.J., Cook, S.E., Wheaton, G.A. and Caccetta, P.A. (1998) The effects of scale on the utility of remotely sensed grain yield estimates, *Proceedings of 9th Australasian Remote Sensing Conference*, Sydney, Australia.
- Cupitt, J. and Whelan, B.M. (2001) Determining potential within-field crop management zones, *Proceedings of the 3rd European Conference on Precision Agriculture*, Montpellier, France, 18-20 June, pp. 7-12.

- Davey, B.G. (1990) The Chemical Properties of Soils, in: *The Scientific Basis of Modern Agriculture*, Campbell, K. O. & Bowyer, J. W. (eds.), Sydney University Press, South Melbourne, Australia, pp. 54-78.
- Deere and Company (2002) *Ag Management Solutions: Yield Mapping*, [online]: [http://www.deere.com/en\\_US/ag/servicesupport/ams/yieldmonitor-mapsub2.html/](http://www.deere.com/en_US/ag/servicesupport/ams/yieldmonitor-mapsub2.html/), accessed (15/11/02).
- Deering, D.W., Rouse, J.W., Haas, R.H. and Schell, J.A. (1975) Measuring "Forage Production" of Grazing Units From Landsat MSS Data, *Proceedings of the 10th International Symposium of Remote Sensing of Environment, II*, pp. 1169-1178.
- Delgado, M.G. and Sendra, J.B. (2004) Sensitivity Analysis in Multicriteria Spatial Decision-Making: A Review, *Human and Ecological Risk Assessment*, Vol. 10, No. 6, pp. 1173-1187.
- Dent, B.D. (1999) Principles for Color Thematic Maps, in: *Cartography: Thematic Map Design*, (5th edn), Kane, K. T. (ed.), WCB/McGraw-Hill, Boston, USA, pp. 288-309.
- Department of Land Information (2000) Land Monitor DEM Metadata, 1 Midland Square, Morrison Road (cnr Gt Northern Hwy), Midland, Western Australia, 6936, 1 pp.
- Digby, P.N.G. (2003) Multivariate and cluster analysis: Discriminant procedure, in: *The Guide to Genstat Release 7.1, Part 2: Statistics*, Payne, R. W. (ed.), VSN International, Wilkinson House, Jordan Hill Road, Oxford, UK, pp. 648-651.
- Dobermann, A., Ping, J.L., Simbahan, G.C. and Adamchuck, V.I. (2003) Processing of yield data for delineating yield zones, in: *Precision Agriculture '03, Proceedings of the 4th European Conference on Precision Agriculture*, Stafford, J. & Werner, A. (eds.), Wageningen Academic Press, Berlin, Germany, pp. 177-185.
- Doerge, T. (1998) Defining Management Zones for Precision Farming, *Crop Insights*, Vol. 8, No. 21, p. 6.
- Drysdale, G., Delfos, J. and Metternicht, G.M. (2002) Low cost remote sensing approach for designing effective sampling strategies of soil properties for site-specific crop management, *Proceedings of 29th International Symposium on Remote Sensing of Environment Conference*, Buenos Aires, Argentina, April 8-12.
- Dymond, J. (2004) *Personnel Communication*, Farm Manager, "Paradym" farm, Rogers Rd, Northam, 6401.
- Eastman, J.R. (1999) *Idrisi 32 Guide to GIS and Image Processing (Volume 2)*, Clark Labs, Worcester, MA, USA. 170 pp.

- Epema, G. (1993) Mapping surface characteristics and their dynamics in a desert area in southern Tunisia with Landsat Thematic Mapper. *Thesis*, Agricultural University of Wageningen, Wageningen, The Netherlands, pp.
- Escadafal, R. (1989) *Caracterisation de la surface des sols arides par observation de terrain et par teledetection*, Paris, France: Editions de l'ORSTOM, pp.
- Escadafal, R. (1994) Soil Spectral Properties and their Relationship with Environmental Parameters - Examples from Arid Regions, in: *Imaging Spectrometry - a Tool for Environmental Observations*, Hill, J. & Mégier, J. (eds.), Kluwer Academic, Dordrecht, The Netherlands, pp. 71-87.
- Escadafal, R. and Huete, A. (1991) Improvement in remote sensing of low vegetation cover in arid regions by correcting vegetation indices for soil "noise", *C.R. Acad. Sci. Paris, t. 312, Series II*, pp. 1385-1391.
- ESRI Inc (2004) ARCGIS (Version 9.0), <http://www.esri.com/>, (program).
- Evans, F.H. and Caccetta, P.A. (2000) Salinity risk prediction using Landsat TM and DEM-derived data, *Proceedings of Tenth Australasian Remote Sensing Conference*, Adelaide, Australia, 21-23 August.
- Evans, F.H., Caccetta, P.A. and Ferdowsian, R. (1996) Integrating remotely sensed data with other spatial data sets to predict areas at risk from salinity, *Proceedings of 8th Australasian Remote Sensing Conference*, Canberra, Australia, 25-29 March.
- Evans, I.S. (1972) General geomorphometry, derivations of altitude and descriptive statistics, in: *Spatial Analysis in Geomorphology*, Chorley, R. J. (ed.), Methuen & Co. Ltd., London, UK, pp. 17-90.
- FAO (1990) *Guidelines for Soil Description*, (3rd Revised edn), Food and Agriculture Organization of the United Nations, Rome, Italy. 70 pp.
- Felicisimo, A.M. (1994) Parametric statistical method of error detection in digital elevation models, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 49, No. 4, pp. 29-33.
- Fitzgerald, R.W. and Lees, B.G. (1994) Assessing the classification accuracy of multisource remote sensing data, *Remote Sensing of Environment*, Vol. 47, No. 3, pp. 362-368.
- Fitzpatrick, R.W., McKenzie, N. and Maschmedt, D.J. (1999) Soil Morphological Indicators and their Importance to Soil Fertility, in: *Soil Analysis: an Interpretation Manual*, Peverill, K. I., Sparrow, L. A. & Reuter, D. J. (eds.), CSIRO Publishing, Collingwood, Vic, Australia, pp. 55-69.

- Fleming, K.L., Westfall, D.G. and Heermann, D.F. (1999) Farmer Developed Management Zone Maps for Variable Rate Fertilizer Application, in: *Precision Agriculture '99, Proceedings of the 2nd European Conference on Precision Agriculture*, Stafford, J. V. (ed.), Sheffield Academic Press, Sheffield, UK, pp. 917-926.
- Florin, M.J., McBratney, A.B. and Whelan, B.M. (2005) Extending Site-Specific Crop Management from individual fields to an entire farm, in: *Precision Agriculture '05, Proceedings of the 5th European Conference on Precision Agriculture*, Stafford, J. (ed.), Wageningen Academic Press, Uppsala, Sweden, pp. 857-863.
- Foody, G.M. (2002) Status of land cover classification accuracy assessment, *Remote Sensing of Environment*, Vol. 80, No. 1, pp. 185-201.
- Franklin, S.E. and Wilson, B.A. (1992) A Three-Stage Classifier for Remote Sensing of Mountain Environments, *Photogrammetric Engineering & Remote Sensing*, Vol. 58, No. 4, pp. 449-454.
- Friedman, H.P. and Rubin, J. (1967) On some invariant criteria for grouping data, *Journal of the American Statistical Association*, Vol. 62, pp. 1159-1178.
- Furby, S. (2004) *Personnel Communication*, CSIRO, Mathematical and Information Sciences, Leeuwin Centre for Earth Sensing Technologies, 65 Brockway Road, Floreat Park, Western Australia, 6014.
- Furby, S. and Campbell, N.A. (2001) Calibrating images from different dates to 'like-value' digital counts, *Remote Sensing of Environment*, Vol. 77, No. 2, pp. 186-196.
- Gallant, J.C. and Wilson, J.P. (2000) Primary Topographic Attributes, in: *Terrain Analysis: Principals and Applications*, Wilson, P. W. & Gallant, J. C. (eds.), Wiley, New York, USA, pp. 51-85.
- Gessler, P.E., Moore, I.D., McKenzie, N.J. and Ryan, P.J. (1995) Soil-landscape modelling and the spatial prediction of soil attributes, *International Journal of Geographic Information Systems*, Vol. 9, No. 4, pp. 421-432.
- Gitelson, A.A. and Merzlyak, M.N. (1997) Remote estimation of chlorophyll content in higher plant leaves, *International Journal of Remote Sensing*, Vol. 18, No. 12, pp. 2691-2697.
- Gordon, A.D. (1996) A survey of constrained classification, *Computational Statistics and Data Analysis*, Vol. 21, No. 1, pp. 17-29.
- Gower, J.C. (1971) A General coefficient of similarity and some of its properties, *Biometrics*, Vol. 27, No. 4, pp. 857-871.

- Gunn, R.H. (1988) Developing Land Survey Specifications, in: *Australian Soil and Land Survey Handbook: Guidelines for Conducting Surveys*, Gunn, R. H., Beattie, J. A., Reid, R. E. & van de Graaff, R. H. M. (eds.), Inkata Press, Melbourne, Australia, pp. 73-89.
- Hagen, A. (2002) Multi-method assessment of map similarity, *Proceedings of 5th AGILE Conference on Geographic Information Science*, Palma, Balearic Islands, Spain, 25-27 April, pp. 171-182.
- Hagen, A. (2003) Fuzzy set approach to assessing simialrity of catergorical maps, *International Journal of Geographic Information Science*, Vol. 17, No. 3, pp. 235-249.
- Hall, F.G., Strebel, D.E., Nickeson, J.E. and Goetz, S.J. (1991) Radiometric rectification: Toward a common radiometric response among multidade, multisensor images, *Remote Sensing of Environment*, Vol. 35, No. 1, pp. 11-27.
- Hall, K.R. and Maruca, S.L. (2001) Mapping a forest mosaic - A comparison of vegetation and bird distributions using geographic boundary analysis, *Plant Ecology*, Vol. 156, No. 1, pp. 105-120.
- Halloran, K. (2004) Analysing the Relationship between High Resolution Digital Multispectral Imagery and Yield Data, *Honours Thesis*, Department of Spatial Sciences, Curtin University, Perth, Western Australia, 148 pp.
- Hannah, M.J. (1981) Error Detection and Correction in Digital Terrain Models, *Photogrammetric Engineering and Remote Sensing*, Vol. 47, No. 1, pp. 63-69.
- Hatfield, J.L., Kanemasu, E.T., Asrar, G., Jackson, R.D., Pinter, P.J., Reginato, R.J. and Idso, S.B. (1985) Lead-area estimates from spectral measurements over various planting dates of wheat, *International Journal of Remote Sensing*, Vol. 6, No. 1, pp. 167-175.
- Hazelton, P.A. (1993) The Formation of Soil, its Profile Description and Attributes, in: *Soil Technology-Applied Soil Science*, Hazelton, P. A. & Koppi, A. J. (eds.), Australian Society of Soil Science Inc. (NSW Branch) and Department of Agricultural Chemistry and Soil Science, University of Sydney, Sydney, Australia, pp. 1-10.
- Hazelton, P.A. and Murphy, B.W. (1992) *What do all the numbers mean? A guide for the interpretation of soil test results*, Department of Conservation and Land Management (incorporating the Soil Conservation Service of NSW), Sydney, N.S.W.pp.

- Heermann, D.F., Diker, K., Buchleiter, G.W. and Brodahl, M.K. (2003) The value of additional data to locate potential management zones in commercial corn fields under center pivot irrigation, in: *Precision Agriculture '03, Proceedings of the 4th European Conference on Precision Agriculture*, Stafford, J. & Werner, A. (eds.), Wageningen Academic Press, Berlin, Germany, pp. 279-284.
- Henderson, T., Szilagy, A., Baumgardner, M., Chen, C. and Landgrebe, D. (1989) Spectral Band Selection for Classification of Soil Organic Matter Content, *Soil Science Society of America Journal*, Vol. 53, pp. 1778-1784.
- Henderson, T.L., Baumgardner, M.F., Franzmeier, D.P., Stott, D.E. and Coster, D.C. (1992) High Dimensional Reflectance Analysis of Soil Organic Matter, *Soil Science Society of America Journal*, Vol. 56, pp. 865-872.
- Herron, N., Peterson, P. and Black, D. (2004) The Salinity Benefit Index, Centre of Natural Resources, N.S.W. Department of Infrastructure, Planning and Natural Resources, Parramatta, 23 pp.
- Huete, A. (2002) *Soil Optical Properties and Monitoring Soil Processes*, [online]: [http://tbrs.arizona.edu/education/553/Soils\\_lecture.pdf](http://tbrs.arizona.edu/education/553/Soils_lecture.pdf), accessed (14/11/02).
- Huete, A.R. (1988) A Soil-Adjusted Vegetation Index (SAVI), *Remote Sensing of Environment*, Vol. 25, pp. 295-309.
- Huete, A.R. and Escadafal, R. (1991) Assessment of Biophysical Soil Properties Through Spectral Decomposition Techniques, *Remote Sensing of Environment*, Vol. 35, pp. 149-159.
- Hutchinson, M.F. (1989) A new method for gridding elevation and stream line data with automatic removal of pits, *Journal of Hydrology*, Vol. 106, pp. 211-232.
- Intergraph Corporation (2003) GeoMedia Professional (Version 05.01.13.41), <http://www.intergraph.com>, (program).
- Irons, J.R., Weismiller, R.A. and Petersen, G.W. (1989) Soil Reflectance, in: *Theory and Applications of Optical Remote Sensing*, Asrar, G. (ed.), Wiley, New York, USA, pp. 67-106.
- Irvin, B.J., Ventura, S.J. and Slater, B.K. (1995) Landform Classification for Soil-Landscape Studies, *Proceedings of the 1995 ESRI International User Conference*, Palm Springs, California, USA, 22-26 May.
- Jackson, R.D. and Huete, A.R. (1991) Interpreting vegetation indices, *Preventive Veterinary Medicine*, Vol. 11, pp. 185-200.
- Jenny, H. (1941) *Factors of Soil Formation: A system of Quantitative Pedology*, McGraw-Hill, New York, USA, pp.



- Jensen, J.R. (1996a) *Introductory digital image processing: a remote sensing perspective*, (2nd edn), Prentice-Hall, Upper Saddle River, N.J., USA. 316 pp.
- Jensen, J.R. (1996b) Thematic Information Extraction: Image Classification, in: *Introductory digital image processing: a remote sensing perspective*, (2nd edn), Prentice-Hall, Upper Saddle River, N.J., USA, pp. 197-256.
- John Deere Ag Management Solutions (2003) JDOffice (1.3), (program).
- Justice, C.O. and Townshend, J.R.G. (1981) Integrating ground data with remote sensing, in: *Terrain Analysis and Remote Sensing*, Townshend, J. R. G. (ed.), Allen and Unwin, London, p. 232.
- Karmonova, L.A. (1982) Effect of Various Iron Compounds on the Spectral Reflectance and Color of Soils, *Soviet Soil Science*, pp. 53-60, Re print from 1981 *Pochvovedeniya*, 9, 57-64.
- Karnieli, A., Kaufman, Y.J., Remer, L. and Wald, A. (2001) AFRI-aerosol free vegetation index, *Remote Sensing of Environment*, Vol. 77, pp. 10-21.
- Kaufman, Y.J. and Tanré, D. (1992) Atmospherically Resistant Vegetation Index (ARVI) for EOS-MODIS, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 30, No. 2, pp. 261-270.
- Kauth, R.J. and Thomas, G.S. (1976) The Tassled Cap - A Graphic Description of the Spectral Temporal Development of Agricultural Crops As Seen By Landsat, *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, Purdue University, West Lafayette, Indiana, pp. 41-51.
- Keigan Systems Inc. (1999) MFworks (Version 2.6), 633 Colborne St. Suite 250, London, Ontario, Canada, (program).
- Keigan Systems Inc. (2003a) GeoMedia Grid (Version 05.01.01.28), 633 Colborne St. Suite 250, London, Ontario, Canada, (program).
- Keigan Systems Inc. (2003b) 'GeoMedia Grid Help Manual', 633 Colborne St. Suite 250, London, Ontario, N6B 2V3, Canada.
- Kelly, R.M., Strong, W.M., Jensen, T.A., Butler, D., Town, B. and Adams, M.L. (2002) Temporal Changes in Grain Yield and Grain Protein in Northern Australia, *Proceedings of the 6th International Conference on Precision Agriculture*, Madison, Wisconsin, p. 9.
- Kersebaum, K.C., Reuter, H.I., Lorenz, K. and Wendroth, O. (2005) Long term simulation of soil/crop interactions to estimate management zones and consequences of site specific nitrogen management considering water protection, in: *Precision Agriculture '05, Proceedings of the 5th European Conference on Precision Agriculture*, Stafford, J. (ed.), Wageningen Academic Press, Uppsala, Sweden, pp. 795-802.

- Klingseisen, B. (2004a) GIS based generation of topographic attributes for landform classification, *Diploma Thesis*, School of Geoinformation, University of Applied Sciences, Villach, Austria, 122 pp.
- Klingseisen, B. (2004b) LANDFORM (Version 1.0), CPSTOF project, Department of Spatial Sciences, Curtin University of Technology, GPO Box U1987, Perth, WA, 6845, (<http://www.cage.curtin.edu.au/~graciela/projects/cpstof/>), (program).
- Klingseisen, B., Warren, G. and Metternicht, G. (2004) LANDFORM: GIS based generation of topographic attributes for landform classification in Australia, in: *Applied Geoinformatics 2004, Proceedings of 16th AGIT Symposium Salzburg*, Strobl, J., Blaschke, T. & Griesebner, G. (eds.), Wichmann, Heidelberg, pp. 344-353.
- Kramer, P.J. and Boyer, J.S. (1995) Soil and Water, in: *Water Relations of Plants and Soils*, Academic Press, London, UK, pp. 84-114.
- Lamb, D.W. (2000) The use of qualitative airborne multispectral imagery for managing agricultural crops - a case study in south-eastern Australia, *Australian Journal of Experimental Agriculture*, Vol. 40, pp. 725-738.
- Lantzke, N. and Fulton, I. (1993) Land resources of the Northam Region, *Land Resources Series No. 11*, Department of Agriculture - Western Australia, South Perth, Australia, 157 pp.
- Lark, R.M. and Stafford, J.V. (1997) Classification as a first step of interpretation in the interpretation of temporal and spatial variation of crop yield, *Annals of Applied Biology*, Vol. 130, No. 1, pp. 111-121.
- Lark, R.M., Wheeler, H.C., Bradley, R.I., Mayr, T.R. and Dampney, P.M.R. (2003) Developing a cost-effective procedure for investigating within-field variation of soil conditions, *Project Report No. 296*, The Home-Grown Cereals Authority (HCGA), UK, 163 pp.
- Latz, K., Weismiller, R., van Soyoc, G. and Baumgardner, M. (1984) Characteristic variation in spectral reflectance of selected eroded Alfisols, *Soil Science Society of America Journal*, Vol. 48, pp. 1130-1134.
- Laws Agricultural Trust (2003a) GenStat (Windows 7th Edition), VSN International Trust, Hemel Hempstead, UK, (program).
- Laws Agricultural Trust (2003b) *Introduction, GenStat for Windows 7th Edition*, VSN International, Oxford, UK. 336 pp.
- Legendre, P. and Legendre, L. (1998) Spatial analysis, in: *Numerical ecology*, (2nd edn), Elsevier Science, Amsterdam, pp. 707-785.
- Lloyd, B. (2003) Agricultural Definition for Small Landholders, *Bulletin 4570*, Department of Agriculture - Western Australia, 26 pp.

- López-Granados, F., Jurado-Expósito, M., Peña-Barragán, J.M. and García-Torres, L. (2005) Using geostatistical and remote sensing approaches for mapping soil properties, *European Journal of Agronomy*, Vol. 23, No. 3, pp. 279-289.
- López, C. (2002) An experiment on the elevation accuracy improvement of photogrammetrically derived DEM, *International Journal of Geographical Information Science*, Vol. 16, No. 4, pp. 361-375.
- Loveday, J. (1974) *Methods for Analysis of Irrigated Soils*, Technical Communication No. 54 of the Commonwealth Bureau of Soils, Commonwealth Agricultural Bureau. 208 pp.
- Lusch, D.P. (1989) Fundamental Considerations for Teaching the Spectral reflectance Characteristics of Vegetation, Soil and Water, in: *Current trends in remote sensing education*, Nellis, M. D., Lougeay, R. & Lulla, K. (eds.), Geocarto International Centre, Hong Kong, pp. 5-27.
- Malczewski, J. (1999) Criterion Weighting, in: *GIS and Multicriteria Decision Analysis*, John Wiley & Sons, New York, USA, pp. 177-195.
- Marriott, F.H.C. (1971) Practical problems in a method of cluster analysis, *Biometrics*, Vol. 27, pp. 501-514.
- Martin, R.J. and Gill, G.S. (1993) Weed competition and crop yield, in: *Management of agricultural weeds in Western Australia*, Dodd, J., Martin, R. J. & Malcolm Howes, K. (eds.), Department of Agriculture - Western Australia, Bulletin 4243, Perth, Australia, pp. 81-84.
- Mattikalli, N.M. (1997) Soil Color Modeling for the Visible and Near-Infrared Bands of Landsat Sensors Using Laboratory Spectral Measurements, *Remote Sensing of Environment*, Vol. 59, pp. 14-28.
- McArthur, W.M. (1991) *Reference soils of south-western Australia*, Australian Society of Soil Science Incorporated (W.A. Branch), Perth, Western Australia. 265 pp.
- McBratney, A.B., Mendonça-Santos, M.L. and Minasny, B. (2003) On digital soil mapping, *Geoderma*, Vol. 117, pp. 3-52.
- McBratney, A.B., Whelan, B.M., Walvoort, D.J.J. and Minasny, B. (1999) A purposive sampling scheme for precision agriculture, in: *Precision Agriculture '99, Proceedings of the 2nd European Conference on Precision Agriculture*, Stafford, J. V. (ed.), Sheffield Academic Press, Sheffield, UK, pp. 101-110.
- McDermid, G.J. and Franklin, S.E. (1994) Spectral, Spatial, and Geomorphometric Variables for the Remote Sensing of Slope Processes, *Remote Sensing of Environment*, Vol. 49, pp. 57-71.

- McDonald, R.C. and Isbell, R.F. (1990) Soil Profile, in: *Australian Soil and Land Survey-Field Handbook*, (2nd edn), McDonald, R. C., Isbell, R. F., Speight, J. R., Walker, J. & Hopkins, M. S. (eds.), Inkata Press, Melbourne, Australia, pp. 103-152.
- McDonald, R.C., Isbell, R.F., Speight, J.R., Walker, J. and Hopkins, M.S. (1990) *Australian Soil and Land Survey-Field Handbook*,(2nd edn), Inkata Press, Melbourne, Australia. 198 pp.
- McIntire, E.J.B. and Fortin, M.J. (2006) Structure and function of wildfire and mountain pine beetle forest boundaries, *Ecography*, Vol. 29, No. 3, pp. 309-318.
- McKenzie, N.J., Gessler, P.E., Ryan, P.J. and O'Connell, D.A. (2000) The Role of Terrain Analysis in Soil Mapping, in: *Terrain Analysis: Principals and Applications*, Wilson, J. P. & Gallant, J. C. (eds.), Wiley, New York, USA, pp. 245-265.
- McKenzie, N.J. and Ryan, P.J. (1999) Spatial prediction of soil properties using environmental correlation, *Geoderma*, Vol. 89, pp. 67-94.
- McNairn, H., Ellis, J., van der Sanden, J.J., Hirose, T. and Brown, R.J. (2002) Providing crop information using RADARSAT-1 and satellite optical imagery, *International Journal of Remote Sensing*, Vol. 23, No. 5, pp. 851-870.
- Metson, A.J. (1956) *Methods of Analysis for Soil Survey Samples*, New Zealand Soil Bureau Bulletin 12, 208 pp.
- Metternicht, G. (1999) Change detection assessment using fuzzy sets and remotely sensed data: an application of topographic map revision, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, pp. 221-233.
- Metternicht, G. (2003) Vegetation indices derived from high-resolution airborne videography for precision crop mangement, *International Journal of Remote Sensing*, Vol. 24, No. 14, pp. 2855-2877.
- Metternicht, G., Honey, F., Beeston, G. and Gonzalez, S. (2000) Airborne Videography For Rapid Assessment Of Vegetation Conditions In Agricultural Landscapes, *Proceedings of 10th Australasian Remote Sensing Photogrammetry Conference*, Adelaide, Australia, 21-25 August, unpaginated CD-ROM.
- Metternicht, G., Klingseisen, B. and Paulus, G. (2005) A Semi-Automated Approach for GIS Based Generation of Topographic Attributes for Landform Classification, *Proceedings of 22nd International Cartographic Conference*, A Coruña, Spain, 9-6 July, p. 10.

- Metternicht, G., Newby, T., van der Berg, H., Paterson, G. and Booyens, B. (2002) Feasibility of using Aster data for Rapid Farm Scale Soil Mapping in South Africa, *Proceedings of the 11th Australasian Remote Sensing and Photogrammetry Conference*, Brisbane, Australia, 2-6 September, pp. 454-470.
- Metternicht, G.I. and Zinck, J.A. (1997) Spatial discrimination of salt- and sodium-affected soil surfaces, *International Journal of Geographic Information Systems*, Vol. 18, pp. 2571-2586.
- Miao, Y., Mulla, D.J. and Robert, P.C. (2005) Combining soil-landscape and spatial-temporal variability of yield information to delineate site-specific management zones, in: *Precision Agriculture '05, Proceedings of the 5th European Conference on Precision Agriculture*, Stafford, J. (ed.), Wageningen Academic Press, Uppsala, Sweden, pp. 811-818.
- Michelson, D.B., Liljeberg, B.M. and Pilesjö, P. (2000) Comparison of Algorithms for Classifying Swedish Landcover Using Landsat TM and ERS-1 SAR Data, *Remote Sensing of Environment*, Vol. 71, No. 1, pp. 1-15.
- Microsoft Corporation (2003) Excel, (program).
- Minasny, B. and McBratney, A.B. (2002) FuzME (3.0), Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney, NSW 2006, <http://www.usyd.edu.au/su/agric/acpa>, (program).
- Minasny, B., McBratney, A.B. and Whelan, B.M. (2002) VESPER (1.6), Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney, NSW 2006, <http://www.usyd.edu.au/su/agric/acpa>, (program).
- Mogensen, V.O., Jensen, C.R., Mortensen, G., Thage, J.H., Koribidis, J. and Ahmed, A. (1996) Spectral reflectance index as an indicator of drought of field grown oilseed rape (*Brassica napus* L.), *European Journal of Agronomy*, Vol. 5, pp. 125-135.
- Moore, D.S. and McCabe, G.P. (2003) Comparing the Means, in: *Introduction to the Practise of Statistics*, (4th edn), Freeman and Company, New York, USA, pp. 765-780.
- Moore, G. (1998a) Distinctive Morphological Features and their Agricultural Significance, in: *Soilguide: A handbook for understanding and managing agricultural soils*, Moore, G. (ed.), Agriculture Western Australia Bulletin No. 4343, Department of Agriculture, Perth, Australia, pp. 43-50.
- Moore, G. (1998b) Soil Salinity, in: *Soilguide: A handbook for understanding and managing agricultural soils*, Moore, G. (ed.), Agriculture Western Australia Bulletin No. 4343, Department of Agriculture, Perth, Australia, pp. 146-158.

- Moore, G. (1998c) *Soilguide: A handbook for understanding and managing agricultural soils*, Agriculture Western Australia Bulletin No. 4343, Department of Agriculture, Perth, Australia. 381 pp.
- Moore, G., Dolling, P., Porter, B. and Leonard, L. (1998) Soil Acidity, in: *Soilguide: A handbook for understanding and managing agricultural soils*, Moore, G. (ed.), Agriculture Western Australia Bulletin No. 4343, Department of Agriculture, Perth, Australia, pp. 127-140.
- Moore, I.D., Burch, G.J. and Mackenzie, D.H. (1988) Topographic Effects on the Distribution of Surface Soil Water and the Location of Ephemeral Gullies, *Transactions of the American Society of Agricultural Engineers*, Vol. 31, No. 4, pp. 1098-1107.
- Moore, I.D., Gessler, P.E., Nielsen, G.A. and Peterson, G.A. (1993a) Terrain analysis for soil specific crop management, in: *Proceedings of first workshop, soil specific crop management*, Robert, P. C., Rust, R.C. and Larson, W.E. (ed.), American Society of Agronomy : Crop Science Society of America : Soil Science Society of America, Madison, USA, pp. 27-55.
- Moore, I.D., Turner, A.K., Wilson, J.P., Jenson, S.K. and Band, L.E. (1993b) GIS and Land-Surface-Subsurface Process Modeling, in: *Environmental Modeling with GIS*, Goodchild, M. F., Parks, B. O. & Steyaert, L. T. (eds.), Oxford University Press, New York, USA, pp. 198-230.
- Mora, F. and Iverson, L. (2002) A spatially constrained ecological classification: rationale, methodology and implementation, *Plant Ecology*, Vol. 158, No. 2, pp. 153-169.
- Moran, M.S., Inoue, Y. and Barnes, E.M. (1997) Opportunities and Limitations for Image-Based Remote Sensing in Precision Crop Management, *Remote Sensing of Environment*, Vol. 61, pp. 319-346.
- Munsell Color Company (1998 revised edition) 'Munsell Soil Color Charts', Munsell Color, GretagMacbeth, New Windsor, NY, U.S.A.
- Muresk Institute of Agriculture (2005) *The Muresk Farm*, [online]: <http://muresk.curtin.edu.au/campus/northam/farm.html>, accessed (17/05/05).
- Needham, P., Moore, G. and Scholz, G. (1998) Soil Structure Decline, in: *Soilguide: A handbook for understanding and managing agricultural soils*, Moore, G. (ed.), Agriculture Western Australia Bulletin No. 4343, Department of Agriculture, Perth, Australia, pp. 64-79.
- Nehmdahl, H. and Greve, M.H. (2001) Using Soil Electrical Conductivity Measurements for Delineating Management Zones on Highly Variable Soils in Denmark, *Proceedings of the 3rd European Conference on Precision Agriculture*, Montpellier, France, 18-20 June, pp. 461-466.

- Nelson, P. and Delane, R. (1991) *Producing Lupins in Western Australia, Bulletin 4179*, (Revised edn), Department of Agriculture - Western Australia, Perth. 94 pp.
- Nielsen, D.R., Wendroth, O. and Parlange, M.B. (1995) Opportunities for Examining On-Farm Soil Variability, *Proceedings of Site-Specific Management for Agricultural Systems: second international conference*, Minneapolis, MN, USA, March 27-30 1994, pp. 95-132.
- O'Brien, R.A. (2004) Spatial decision support for selecting tropical crops and forages in uncertain environments, *PhD Thesis*, Department of Spatial Sciences, Curtin University of Technology, Perth, Western Australia, 278 pp.
- Oliver, M.A. and Webster, R. (1989) A Geostatistical Basis for Spatial Weighting in Multivariate Classification, *Mathematical Geology*, Vol. 21, No. 1, pp. 15-35.
- Payne, R.W. (2003) Multivariate and cluster analysis, in: *The Guide to Genstat Release 7.1, Part 2: Statistics*, Payne, R. W. (ed.), VSN International, Wilkinson House, Jordan Hill Road, Oxford, UK, pp. 609-754.
- PCI Geomatica (2003) *OrthoEngine User Guide, Geomatica Version 9*, PCI Geomatics, Richmond Hill, Ontario, Canada. 158 pp.
- Pierce, F.J., Warncke, D.D. and Everett, M.W. (1995) Field and nutrient variability in glacial soils of Michigan, *Proceedings of Site-Specific Management for Agricultural Systems: second international conference*, Minneapolis, MN, USA, March 27-30 1994, pp. 133-151.
- Pilesjö, P., Thylén, L. and Persson, A. (2005) Topographical data for delineation of agricultural management zones, in: *Precision Agriculture '05, Proceedings of the 5th European Conference on Precision Agriculture*, Stafford, J. (ed.), Wageningen Academic Press, Uppsala, Sweden, pp. 819-826.
- Pinter, P.J., Hatfield, J.L., Schepers, J.S., Barnes, E.M., Moran, M.S., Daughtry, C.S.T. and Upchurch, D.R. (2003) Remote Sensing for Crop Management, *Photogrammetric Engineering & Remote Sensing*, Vol. 69, No. 6, pp. 647-664.
- Pontius, R.G. (2000) Quantification Error Versus Location Error in Comparison of Categorical Maps, *Photogrammetric Engineering & Remote Sensing*, Vol. 66, No. 8, pp. 1011-1016.
- Power, C., Simms, A. and White, R. (2001) Hierarchical fuzzy pattern matching for the regional comparison of land use maps, *International Journal of Geographic Information Science*, Vol. 15, No. 1, pp. 77-100.
- Purdie, B. (1998) Understanding and Interpreting Soil Chemical and Physical Data, in: *Soilguide: A handbook for understanding and managing agricultural soils*, Moore, G. (ed.), Agriculture Western Australia Bulletin No. 4343, Department of Agriculture, Perth, Australia, pp. 313-332.

- Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H. and Sorooshian, S. (1994) A Modified Soil Adjusted Vegetation Index, *Remote Sensing of Environment*, Vol. 48, No. 2, pp. 119-126.
- Rayment, G.E. and Higginson (1992) *Australian Laboratory Handbook of Soil and Water Chemical Methods*, Inkata Press, Melbourne, Australia. 330 pp.
- Rengasamy, P. and Churchman, G.J. (1999) Cation Exchange Capacity, Exchangeable Cations and Sodicity, in: *Soil Analysis: an Interpretation Manual*, Peeverill, K. I., Sparrow, L. A. & Reuter, D. J. (eds.), CSIRO Publishing, Collingwood, Vic, Australia, pp. 147-157.
- Richardson, A.J. and Weigand, C.L. (1977) Distinguishing Vegetation From Soil Background Information, *Photogrammetric Engineering and Remote Sensing*, Vol. 43, No. 12, pp. 1541-1552.
- RIKS (2005a) Map Comparison Kit (2.0), Research Institute for Knowledge Systems, Maastricht, The Netherlands, <http://www.riks.nl/mck> [25/02/06], (program).
- RIKS (2005b) *Map Comparison Kit 2: User Manual*, Research Institute for Knowledge Systems, Maastricht, The Netherlands, <http://www.riks.nl/mck> [25/02/06]. 60 pp.
- Robinson, T.P. (2004) *Interpolation of Yield Maps for 2002 Cropping Season: CPSTOF Report*, [online]: <http://www.cage.curtin.edu.au/~graciela/projects/cpstof/publications.html>, accessed (30/10/04).
- Robinson, T.P. and Metternicht, G. (2005) Comparing the performance of techniques to improve the quality of yield maps, *Agricultural Systems*, Vol. 85, No. 1, pp. 19-41.
- Roujean, J.L., Leroy, M. and Deschamps, P.Y. (1992) A Bidirectional Reflectance Model of the Earth's Surface for the Correction of Remote Sensing Data, *Journal of Geophysical Research*, Vol. 97, No. D18, pp. 20455-20468.
- Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W. (1973) Monitoring vegetation systems in the Great Plains with ERTS, *Third ERTS Symposium*, Vol. 1, pp. 309-317.
- Rouse, J.W.J., Haas, R., Deering, D.W., Schell, J.A. and Harlan, J.C. (1974) Monitoring the Vernal Advancement and Retrogradation (green wave effect) of natural vegetation, NASA/GSFC Type III Final Report, Greenbelt, MD, 317 pp.
- Ryan, P.J. (2003) *Personnel Communication*, Senior Research Scientist, CSIRO Division of Forestry and Forest Products, PO Box E4008, Kingston, ACT, 2604, Australia.



- Schoknecht, N. and Tille, P. (2002) *Soil-landscape mapping in south-western Australia: an overview*, [online]: [http://agspsrv38.agric.wa.gov.au/servlet/page?\\_pageid=449&\\_dad=portal30&\\_schema=PORTAL30](http://agspsrv38.agric.wa.gov.au/servlet/page?_pageid=449&_dad=portal30&_schema=PORTAL30), accessed (05/11/03).
- Scholz, G. and Moore, G. (1998) Soil Alkalinity and Soil Sodicity, in: *Soilguide: A handbook for understanding and managing agricultural soils*, Moore, G. (ed.), Agriculture Western Australia Bulletin No. 4343, Department of Agriculture, Perth, Australia, pp. 141-145.
- Schott, J.R., Salvaggio, C. and Volchok, W.J. (1988) Radiometric scene normalization using pseudoinvariant features, *Remote Sensing of Environment*, Vol. 26, No. 1, pp. 1-14.
- Scull, P., Franklin, J., Chadwick, O.A. and McArthur, D. (2003) Predictive soil mapping: a review, *Progress in Physical Geography*, Vol. 27, No. 2, pp. 171-197.
- Sellers, P.J. (1985) Canopy reflectance, photosynthesis and transpiration, *International Journal of Remote Sensing*, Vol. 6, No. 8, pp. 1335-1372.
- Senay, G.B., Lyon, J.G., Ward, A.D. and Nokes, S.E. (2000) Using High Spatial Resolution Multispectral Data to Classify Corn and Soybean Crops, *Photogrammetric Engineering & Remote Sensing*, Vol. 66, No. 3, pp. 319-327.
- Shapiro, S.S. and Wilk, M.B. (1965) An Analysis of Variance Test for Normality (Complete Samples), *Biometrika*, Vol. 52, No. 3/4, pp. 591-611.
- Shatar, T.M. and McBratney, A.B. (2001) Subdividing a Field into Contiguous Management Zones using a K-Zone Algorithm., *Proceedings of the 3rd European Conference on Precision Agriculture*, Montpellier, France, 18-20 June, pp. 115-120.
- Shaw, R.J. (1999) Soil Salinity - Electrical Conductivity and Chloride, in: *Soil Analysis: an Interpretation Manual*, Peeverill, K. I., Sparrow, L. A. & Reuter, D. J. (eds.), CSIRO Publishing, Collingwood, Vic, Australia, pp. 129-145.
- Simbahan, G.C. and Dobermann, A. (2006) An algorithm for spatially constrained classification of categorical and continuous soil properties, *Geoderma*, Vol. 136, No. 3-4, pp. 504-523.
- Singh, A. and Harrison, A. (1985) Standardised Principal Components, *International Journal of Remote Sensing*, Vol. 6, No. 6, pp. 883-896.
- Slattery, W.J., Conyers, M.K. and Aitken, R.L. (1999) Soil pH, Aluminium, Manganese and Lime Requirements, in: *Soil Analysis: an Interpretation Manual*, Peeverill, K. I., Sparrow, L. A. & Reuter, D. J. (eds.), CSIRO Publishing, Collingwood, Vic, Australia, pp. 103-128.

- Smolinski, H. (2005) *Personnel Communication*, Research Officer, Client and Resource Information Group, Land Management Services Division of Policy and Business Services, Department of Agriculture - Western Australia, Locked Bag 4, Bentley Delivery Centre, WA, 6983.
- South, S., Qi, J. and Lusch, D.P. (2004) Optimal classification methods for mapping agricultural tillage practices, *Remote Sensing of Environment*, Vol. 91, No. 1, pp. 90-97.
- SpecTerra Services (1999) *Presentation and Analysis of Data*, [online]: [http://www.specterra.com.au/dmsv\\_data\\_frame.html](http://www.specterra.com.au/dmsv_data_frame.html), accessed (29 Oct 2002).
- SpecTerra Services (2003a) DMSI Technical specifications and costs, SpecTerra Services Pty Ltd, Leederville, Western Australia, 2 pp.
- SpecTerra Services (2003b) *SpecTerra Services airborne remote sensing*, [online]: <http://www.specterra.com.au/>, accessed (08/07/04).
- Speight, J.G. (1974) A parametric approach to landform regions, in: *Progress in Geomorphology*, Brown, E. H. & Waters, R. S. (eds.), Alden Press, London, pp. 213-230.
- Speight, J.G. (1990) Landform, in: *Australian Soil and Land Survey-Field Handbook*, (2nd edn), McDonald, R. C., Isbell, R. F., Speight, J. R., Walker, J. & Hopkins, M. S. (eds.), Inkata Press, Melbourne, Australia, pp. 9-57.
- SPSS Inc. (2003) SPSS (12.0.1 for Windows), Chicago, (program).
- Stafford, J.V., Lark, R.M. and Bolam, H.C. (1998) Using Yield Maps to Regionalize Fields into Potential Management Units, *Proceedings of the 4th International Conference on Precision Agriculture*, St. Paul, Minnesota, USA, 19-22 July, pp. 225-237.
- Stewart, C.M. and McBratney, A.B. (2001) Using Bare Soil Imagery to Determine Management Zones for the Variable-Rate Application of inputs for Cotton, *Proceedings of the 3rd European Conference on Precision Agriculture*, Montpellier, France, 18-20 June, pp. 319-324.
- Stoner, E.R. and Baumgardner, M.F. (1981) Characteristic Variations in Reflectance of Surface Soils, *Soil Science Society of America Journal*, Vol. 45, pp. 1161-1165.
- Thenkabail, P.S., Smith, R.B. and De Pauw, E. (2000) Hyperspectral Vegetation Indices and Their Relationships with Agricultural Crop Characteristics, *Remote Sensing of Environment*, Vol. 71, pp. 158-182.
- Thiam, A. and Eastman, J.R. (2003) Vegetation Indices, in: *IDRISI Kilimanjaro: Guide to GIS and Image Processing*, Eastman, J. R. (ed.), Clark Labs, Clark University, Worcester, MA, USA.

- TopCrop Australia (1999a) *Crop Monitoring Guide*, (4th edn), TopCrop Publications, Grains Research and Development Corporation and Primary Industries and Resources, SA. 120 pp.
- TopCrop Australia (1999b) 'Growth Stages of Cereals', Primary Industries and Resources, SA.
- Treitz, P.M. and Howarth, P.J. (2000) Integrating Spectral, Spatial, and Terrain Variables for Forest Ecosystem Classification, *Photogrammetric Engineering and Remote Sensing*, Vol. 66, No. 3, pp. 305-317.
- Treitz, P.M., Howarth, P.J., Shepherd, P.R. and Miller, J.R. (1996) Integrating Remote Sensing Data and Terrain Variables to Discriminate Forest Ecosystems in Northwestern Ontario, *Proceedings of the 26th International Symposium on Remote Sensing of Environment*, Vancouver, BC, Canada, 25-29 March, pp. 515-518.
- Trimble Navigation Limited (1994) *GeoExplorer Operation Manual, Part Number 21281-00 Rev A*, Surveying and Mapping Division, Sunnyvale, CA, USA. 232 pp.
- Tucker, B.M. (1985) Laboratory Procedures for Soluble Salts and Exchangeable Cations in Soils, *Division of Soils Technical Paper No. 47*, CSIRO Australia, 36 pp.
- Urban, D.L. (2004) *Multivariate Analysis: Nonhierarchical Agglomeration, Course Notes, Multivariate Methods for Environmental Applications, Nicholas School of the Environment and Earth Sciences at Duke University, Durham, NC, USA*, [online]: [www.env.duke.edu/landscape/classes/env358/mv\\_pooling.pdf](http://www.env.duke.edu/landscape/classes/env358/mv_pooling.pdf), accessed (21/01/2005).
- van Alphen, B.J. and Stoorvogel, J.J. (1998) A Methodology to Define Management Units in Support of an Integrated, Model-Based Approach to Precision Agriculture, *Proceedings of the 4th International Conference on Precision Agriculture*, St. Paul, Minnesota, USA, 19-22 July, pp. 1267-1278.
- Ventura, S.J. and Irvin, B.J. (2000) Automated Landform Classification Methods for Soil-Landscape Studies, in: *Terrain Analysis: Principals and Applications*, Wilson, P. W. & Gallant, J. C. (eds.), Wiley, New York, USA, pp. 267-294.
- Vermote, E.F., Tanré, D., Deuze, J.L., Herman, M. and Morcrette, J.J. (1997) Second simulation of the satellite signal in the solar spectrum: an overview, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 35, pp. 675-686.
- Voltz, M. and Webster, R. (1990) A comparison of kriging, cubic splines and classification for predicting soil properties from sample information, *Journal of Soil Science*, Vol. 41, pp. 473-490.

- Walkley, A. (1947) A critical examination of a rapid method for determining organic carbon in soils - effect of variations in digestion conditions and of inorganic soil constituents, *Soil Science*, Vol. 63, pp. 251-264.
- Warren, G. and Metternicht, G. (2005) Agricultural Applications of High-Resolution Digital Multispectral Imagery: Evaluating Within-Field Spatial Variability of Canola (*Brassica napus*) in Western Australia, *Photogrammetric Engineering & Remote Sensing*, Vol. 71, No. 5, pp. 595-602.
- Weaving, S. (1999) *Native Vegetation Handbook for the Shire of Northam*, Department of Agriculture - Western Australia, South Perth, Australia. 62 pp.
- Webster, R. (2006) *Personnel Communication*.
- Webster, R. and Oliver, M.A. (1990a) Numerical Classification: Hierarchical Systems, in: *Statistical Methods in Soil and Land Resource Survey*, Webster, R. & Oliver, M. A. (eds.), Oxford University Press, New York, USA, pp. 168-190.
- Webster, R. and Oliver, M.A. (1990b) Numerical Classification: Non-Hierarchical Methods, in: *Statistical Methods in Soil and Land Resource Survey*, Webster, R. & Oliver, M. A. (eds.), Oxford University Press, New York, USA, pp. 191-212.
- Webster, R. and Oliver, M.A. (1990c) Relationships between individuals: Similarity, in: *Statistical Methods in Soil and Land Resource Survey*, Webster, R. & Oliver, M. A. (eds.), Oxford University Press, New York, USA, pp. 112-125.
- Webster, R. and Oliver, M.A. (2001a) Estimating the variogram, in: *Geostatistics for Environmental Scientists*, John Wiley and Sons, Chichester, England, pp. 65-103.
- Webster, R. and Oliver, M.A. (2001b) *Geostatistics for Environmental Scientists*, John Wiley and Sons, Chichester, England. 271 pp.
- Webster, R. and Oliver, M.A. (2001c) Modelling the variogram, in: *Geostatistics for Environmental Scientists*, John Wiley and Sons, Chichester, England, pp. 105-134.
- Weigand, C.L., Richardson, A.J., Escobar, D.E. and Gerbermann, A.H. (1991) Vegetation Indices in Crop Assessments, *Remote Sensing of Environment*, Vol. 35, pp. 105-119.
- Wheatley, J.M., Wilson, J.P., Redmond, R.L., Ma, Z. and DiBenedetto, J. (2000) Automated Land Cover Mapping Using Landsat Thematic Mapper Images and Topographic Attributes, in: *Terrain Analysis: Principles and Applications*, Wilson, J. P. & Gallant, J. C. (eds.), Wiley, New York, USA, pp. 355-389.

- Whelan, B.M. (1998) Reconciling Continuous Soil Variation and Crop Yield - a study of some implications of within field variability for site-specific management, *Unpublished PhD Thesis*, University of Sydney, Sydney, Australia, 356 pp.
- Whelan, B.M., Boydell, B.C. and McBratney, A.B. (2002a) PA:Agriculture through the looking glass, *Proceedings of the 11th Australasian Remote Sensing and Photogrammetry Conference*, Brisbane, Australia, 2-6 September.
- Whelan, B.M., Cupitt, J. and McBratney, A.B. (2002b) Practical definition and interpretation of potential management zones in Australian dryland cropping, *Proceedings of the 6th International Conference on Precision Agriculture*, Madison, Wisconsin, p. 15.
- Whelan, B.M. and McBratney, A. (2003) Definition and interpretation of potential management zones in Australia, *Proceedings of 11th Australian Agronomy Conference*, Geelong, Australia, 2-6 February, CDROM ISBN 0-9750313-0-9.
- Whelan, B.M. and McBratney, A.B. (2000) The 'Null Hypothesis' of Precision Agriculture Management, *Precision Agriculture*, No. 2, pp. 265-279.
- Whelan, B.M., McBratney, A.B. and Minasny, B. (2001) Vesper - Spatial Prediction Software for Precision Agriculture, *Proceedings of 3rd European Conference on Precision Agriculture*, Montpellier, France, 18-20 June, pp. 139-144.
- Wilson, J.P. and Gallant, J.C. (2000) Secondary Topographic Attributes, in: *Terrain Analysis: Principals and Applications*, Wilson, J. P. & Gallant, J. C. (eds.), Wiley, New York, USA, pp. 87-131.
- Xiao, X., He, L., Salas, W., Li, C., Moore III, B., Zhao, R., Froelking, S. and Bolos, S. (2002) Quantitative relationships between field-measured leaf area index and vegetation index derived from vegetation images for paddy rice field, *International Journal of Remote Sensing*, Vol. 23, No. 18, pp. 3595-3604.
- Yang, C. and Anderson, G.L. (1996) Determining within-field management zones for grain sorghum using aerial videography, *Proceedings of the 26th International Symposium on Remote Sensing of Environment*, Vancouver, BC, Canada, 25-29 March, pp. 606-611.
- Yates, D.J. and Steven, M.D. (1987) Reflexion and absorption of solar radiation by flowering canopies of oil-seed rape (*Brassica napus* L.), *Journal of Agricultural Science, Cambridge*, Vol. 109, pp. 495-502.
- Yuan, D., Elvidge, C.D. and Lunetta, R.S. (1999) Survey of Multispectral Methods for Land Cover Change Analysis, in: *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*, Lunetta, R. S. & Elvidge, C. D. (eds.), Taylor and Francis Ltd, London, UK, pp. 21-39.
- Zadeh, L.A. (1965) Fuzzy Sets, *Information and Control*, Vol. 8, pp. 338-353.

- Zhang, N. and Taylor, R.K. (2001) Applications of a Field-Level Geographic Information System (FIS) in Precision Agriculture, *Applied Engineering in Agriculture*, Vol. 17, No. 6, pp. 885-892.
- Zinck, J.A. (1988) Physiography and Soils, *Soil Survey Courses*, ITC, International Institute for Aerospace Survey and Earth Science, Enschede, The Netherlands, 156 pp.

**APPENDICES**

**APPENDIX A**

**DAILY RAINFALL 2002 AND 2003**



## RAINFALL - PARADYM FARM 2002

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1							3.5		4			
2					4		4.5		7.5			
3						1			3			
4						10.5		5				
5						4	2.5	7				
6					5.5	2.5				9		
7					2	2	3					
8				1.5	4							
9				2	3			7				
10								4				
11					7		6.5					
12							13					
13						3.5	3					
14						5	1.5					
15	3								11			
16				15.5				1				
17				6								
18				6.5								
19				0.5		8	4.5					
20						0.5						
21												
22						8						
23									1.5			
24												
25							7.5	2	1			
26						4	1					
27							1	6				
28						8.5						
29												
30								12.5				
31							1	0.5				
	<b>3</b>	<b>0</b>	<b>0</b>	<b>32</b>	<b>25.5</b>	<b>57.5</b>	<b>52.5</b>	<b>45</b>	<b>28</b>	<b>9</b>	<b>0</b>	<b>0</b>

**RAINFALL - PARADYM FARM 2003**

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1						0.5						
2						2.5						
3						0.5		18.5				
4								3				
5							10			0.5		
6							0.5			0.5		
7								1				3.5
8				14.5								
9				20	15		2		5			
10		5		5			9.5		3			
11							8	16	2.5			
12					4.5			4.5	3			
13					3.5				3		5	
14							2.5					
15					1		3.5					
16					22			2				
17		5			3						28.5	
18												3
19								2.5				
20							1.5	2.5	2.5			
21			2	2	7		1					
22				1	2	5.5		23.5	7			
23						10.5			6			
24												
25						2.5		1.5				
26						13						
27				1		12.5	1			4		
28						8						
29						23		0.5				
30			3.5			3	9		8.5			
31			17		2		2.5					
	<b>0</b>	<b>10</b>	<b>22.5</b>	<b>43.5</b>	<b>60</b>	<b>81.5</b>	<b>51</b>	<b>75.5</b>	<b>40.5</b>	<b>5</b>	<b>33.5</b>	<b>6.5</b>

**APPENDIX B**

**SOIL SAMPLING DESIGN**



- 1) Generate NDVI and RI layers from DMSI across the entire study area following Equations (2.4) and (2.9) respectively with a cell resolution of 2m, Figure A.1 (1).
- 2) Determine rates of change in NDVI and RI layers using the grade module within MFworks and output as percentage slope. NDVI\_Change% and RI\_Change% were created in this way, Figure A.1. (2). The gradient is calculated as the average slope (rise over run) multiplied by 100. Flat areas (no change) are assigned a 0 percent, whereas a 45° slope (e.g. steep, rapid changes) takes a value of 100 percent (Keigan Systems Inc., 1999).
- 3) Generate a true colour image of the study area by combining the Blue, Green and Red bands of DMSI.
- 4) Create a mask layer by heads-up digitizing within GeoMedia software. Rocky outcrops, buildings, dams, roads and remnant vegetation masked out. Individual polygons created with similar visual appearance, i.e. bare soil, cropped paddocks, gullies and drainage lines. The vector file was exported from GeoMedia and imported into MFworks as a raster layer with 2m-cell size. Each polygon is associated with a unique number, Figure A.1 (3), so they can be identified individually.
- 5) Generate a 10m mask, (Figure A.1 (4)). Created using the *re-space* module and the *average of values* option, in which the resulting cell values are the average of the cell values in the corresponding area in the input map layer (2m\_Mask) (Keigan Systems Inc., 1999).
- 6) Create the NDVI\_2mMask (Figure A.1 (5)) by combining the 2m\_Mask and the NDVI layer. The *combine* module within MFworks generates unique values for all existing combinations of two or more maps layers (Keigan Systems Inc., 1999), and thus could be used to determine the distribution of NDVI values within a unique region of the mask layer.
- 7) Export and analyse the NDVI\_2mMask layer. For each unique region of the mask, the area (number of cells) with a  $NDVI \leq 0$  and area (number of cells) with a  $NDVI > 0.05$  was determined.
- 8) Reclass the NDVI\_2mMask layer into regions predominantly vegetated (majority of area  $NDVI > 0.05$ ) labelled NDVI or bare soil (majority of area  $NDVI \leq 0$ ) labelled RI. Resulting in the output layer 2mMask\_NDVI\_RI, shown in Figure A.1 (6).

- 9) Generate output layers for the NDVI (NDVI\_Change%\_NDVI) and the RI (RI\_Change%\_RI) by combining the mask layer (2mMask\_NDVI\_RI) with each of the 'Change percent layers' (NDVI\_Change% and RI\_Change%) as shown in Figure A.1 (6), using the *combine* module.
- 10) Generate histograms of the rate of change as expressed by the percentage of slope for the NDVI and RI values and their associated regions.
- 11) Determine the number of variability classes and class boundaries for each index (NDVI and RI) by assessing the histograms. Three classes of variability (high, medium and low) applied with class boundaries located at the natural breaks highlighted in the histograms for each index.
- 12) *Reclass* the NDVI and RI layers into 3 classes of variability (Figure A.1 (8)).
- 13) Create final variability map (3 classes) by combining the NDVI and RI variability layers. The *cover* operation, used to assemble map layers into a final product, was applied (Keigan Systems Inc., 1999). The 2m spatial resolution layers were resampled to a 10m-cell resolution image using the *average of values* option with the *re-space* module of MFWorks. This operation smoothed the layers, and produced the final variability layer shown in Figure 4.4, (Variability 1-3 at 10m, Figure A.1 (9)).
- 14) Generate the slope layer from the DEM for the entire farm using the *grade* module with the *average slope* option within MFWorks (Slope%, Figure A.1 (10)).
- 15) Combine the slope map with the 10m mask layer (Figure A.1 (4)) to generate an output slope layer that coincides with the area of interest on the variability map.
- 16) Generate a histogram of the slope values for the study area.
- 17) *Reclass* the slope map into 4 classes according to the Australian Soil and Land Survey classification standards (Table 4.1). This forms the output map layer (Slope% Class 1-4, Figure A.1 (11)) providing the spatial regions in which the soil sampling points could be targeted.
- 18) Compute the variability proportion to enable the appropriate number of sampling points to be allocated to each stratum. This takes into consideration the percentage of area of each variability class, and a weight is assigned to each class in terms of high, medium and low variability in order to increase the intensity of sample points in areas of high variability.

- 19) Determine the appropriate weight for each variability class based on pairwise comparison method developed by Saaty in 1980 (Malczewski, 1999).
- 20) Determine the correct proportion of the 250 sample points to be assigned to each slope class based on the area of each slope class within the mask region.
- 21) Generate the sample matrix to determine the number of sample points to be randomly located with each sample zone (Figure A.1 (12)). This is determined by multiplying the variability proportion by the number of sampling points allocated to each slope class.
- 22) Generate the sample zones map layer (Figure A.1 (12)) using the *cross* module, which performs map layer cross tabulation cell by cell (Keigan Systems Inc., 1999) between the slope classes map and variability class map; thus generating an output map layer with 12 unique zones derived from all possible combinations of the slope and variability classes.
- 23) Randomly select the appropriate number of sample points from each sample zone producing a new raster output layer 'Sample Points' (Figure A.1 (13)).
- 24) Export 'Sample Points' as a XYZ file-a text file (Sample Pts.xls) (Figure A.1 (14)) containing the spatial location (easting and northing) of the 250 sample points.
- 25) Overlay the sample points onto the true colour composite generated using the DMSI blue, green and red bands.

**APPENDIX C**

**DRAINAGE CLASSES**



Drainage Classes defined in the Manual for Guidelines for Soil Descriptions (FAO, 1990).

Code	Name	Description
E	Excessively drained	Water is removed from the soil very rapidly. The soils are commonly very coarse textured or rocky, shallow or on steep slopes.
S	Somewhat excessively drained	The water is removed from soil rapidly. The soils are commonly sandy and very pervious.
W	Well drained	Water is removed from the soil readily but not rapidly. The soils commonly retain optimal amounts of moisture, but wetness does not inhibit growth of roots for significant periods.
M	Moderately well drained	Water is removed from the soil somewhat slowly during periods of the year. The soils are wet for short periods within the rooting depth. They commonly have an almost impervious layer, or periodically receive heavy rainfall.
I	Somewhat poorly (imperfectly) drained	Water is removed slowly so that the soil is wet at a shallow depth for significant periods. The soils commonly have an almost impervious layer, a high water table, additions of water by seepage, or very frequent rainfall.
P	Poorly drained	Water is removed so slowly that the soils are commonly wet at a shallow depth for considerable periods. The soils commonly have a shallow water table which is usually the result of an almost impervious layer, seepage, or frequent rainfall.
V	Very poorly drained	Water is removed so slowly that the soils are wet at shallow depths for long periods. The soils have a very shallow water table and are commonly in level or depressed sites or have very high rainfall almost every day.

## **APPENDIX D**

### **FIELD TEXTURE ESTIMATION**

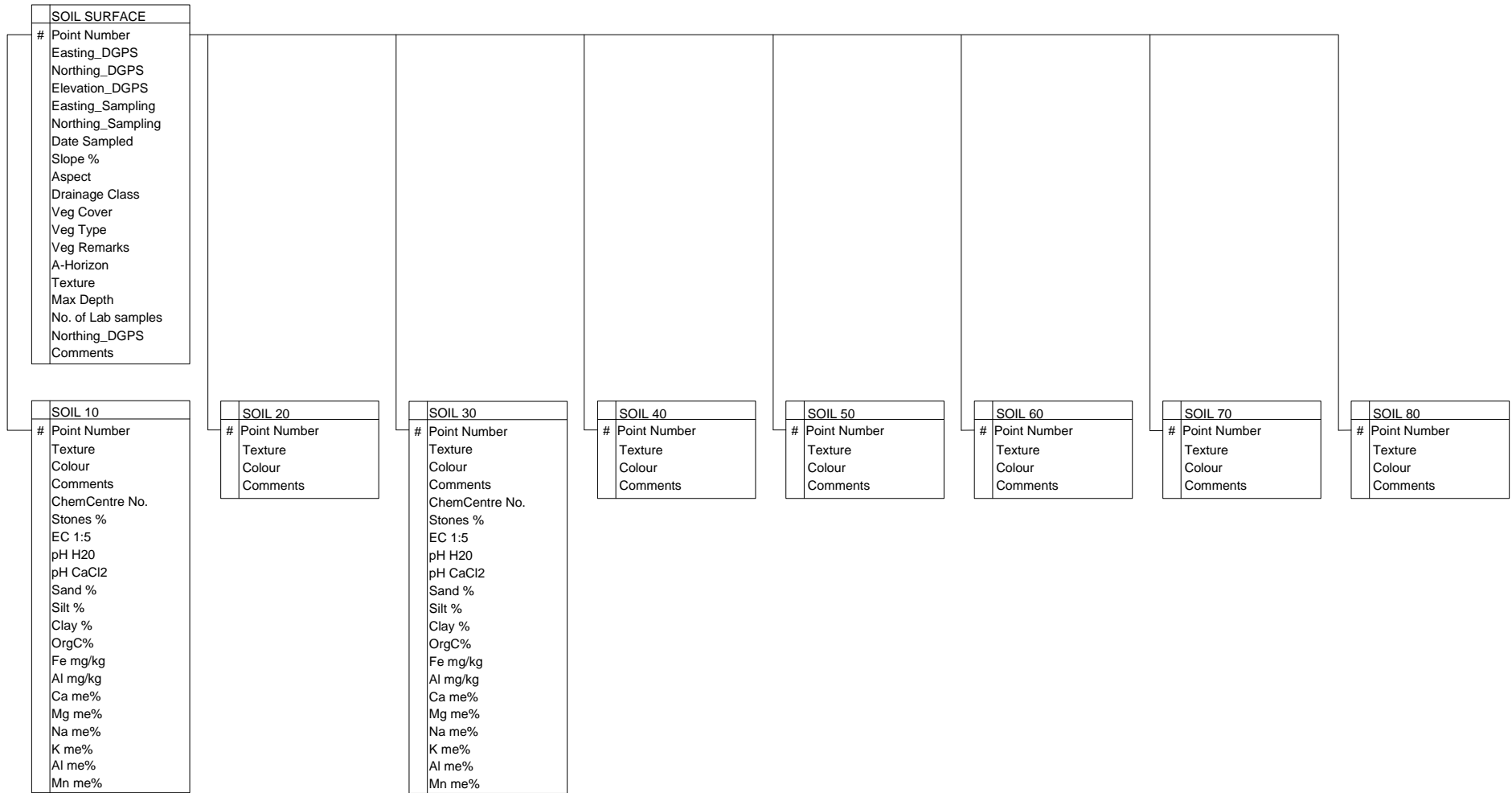
*Field texture is the measure of behaviour of a small handful of soil when moistened and kneaded into a ball (bolus) and then pressed out between thumb and forefinger to form a ribbon. The behaviour of the soil during bolus and formation and the ribbon length determine the field texture grade Table A.1 (Moore, 1998a).*

Table A.1 Field texture grades and approximate percentage of clay content when determined from laboratory particle size analysis (reproduced from McDonald and Isbell, 1990).

Code	Texture Grade	Behavior of moist bolus	Approximate clay content
S	Sand	Coherence nil to very slight, cannot be moulded; sand grains of medium size; single sand grains adhere to fingers.	< 5%
LS	Loamy sand	Slight coherence; sand grains of medium size; can be sheared between thumb and forefinger to give minimal ribbon of about 5mm.	About 5%
CS	Clayey sand	Slight coherence; sand grains of medium size; sticky when wet; many sand grains stick to fingers; will form minimal ribbon of 5-15mm; discolours fingers with clay stain.	5%-10%
SL	Sandy loam	Bolus coherent but very sandy to touch; will form ribbon of 15-25mm; dominant sand grains are of medium size and are readily visible.	10%-20%
L	Loam	Bolus coherent and rather spongy; smooth feel when manipulated but with no obvious sandiness or 'silkeness'; may be somewhat greasy to the touch if much organic matter present; will form ribbon of about 25mm.	About 25%
ZL	Silty loam	Coherent bolus; very smooth to often silky when manipulated; will form ribbon of about 25mm.	About 25% and with silt 25% or more
SCL	Sandy clay loam	Strongly coherent bolus, sandy to touch; medium size sand grains visible in finer matrix; will form ribbon of 25-40mm.	20%-30%
CL	Clay loam	Coherent plastic bolus, smooth to manipulate; will form ribbon of 40-50mm.	30%-35%
CLS	Clay loam, sandy	Coherent plastic bolus; medium size sand grains visible in finer matrix; will form ribbon of 40-50mm.	30%-35%
ZCL	Silty clay loam	Coherent smooth bolus, plastic and often silky to touch; will form ribbon of 40-50mm.	30%-35% and with silt 25% or more
LC	Light clay	Plastic bolus; smooth to touch; slight resistance to shearing between thumb and fore-finger; will form ribbon of 50-75mm.	35%-40%
LMC	Light medium clay	Plastic bolus; smooth to touch; slight to moderate resistance to ribboning shear; will form ribbon of about 75mm.	40%-45%
MC	Medium clay	Smooth plastic bolus; handles like plasticine and can be moulded into rods without fracture; has moderate resistance to ribboning shear; will form ribbon of 75mm or more.	45%-55%
MHC	Medium heavy clay	Smooth plastic bolus; handles like plasticine; can be moulded into rods without fracture; has moderate to firm resistance to ribboning shear; will form ribbon of 75mm or more.	50% or more
HC	Heavy clay	Smooth plastic bolus; handles like stiff plasticine; can be moulded into rods without fracture; has firm resistance to ribboning shear; will form ribbon of 75mm or more.	50% or more

**APPENDIX E**

**FULL SOIL DATABASE SCHEMA**



**APPENDIX F**

**IMPLEMENTATION OF LANDFORM CLASSIFICATION**

1. Import DEM: The LandMonitor DEM (smoothed by Caccetta (2000)) is imported into GeoMedia Grid

Generate Topographic Attributes: The input layer used in the following commands for steps 2-5 was the Smoothed DEM at 10m resolution.

2. Compute Slope Layer: The *Grade* command was used to calculate *average slope in percent*. The output layer was saved as “Grade\_percent”.
3. Compute Local Relief: The *Local Relief...* command was used with the default window radius of 150m. The output layer was saved as “Local\_Relief”.
4. Compute Elevation Percentile: The *Elevation Percentile...* command was used with a default window radius of 150m. The output layer was saved as “Percentile”.
5. Compute Plan and Profile Curvature: The *Curvature Zevenbergen...* command was used to generate plan and profile curvature. The output layers were saved as “PlanCurve” and “ProfileCurve” respectively.

Generate Primary Landform Classes: The input layers for the following commands in steps 6-11, are those derived from steps 2-5 above.

6. Classify Crest: The *Classify Crests...* command was used with the *predefined mask* view. Input layers and there default values were as follows; (Percentile > 0.65 and Local\_Relief > 7.5) and (PlanCurve > 0.0 or ProfileCurve > 0.0). The output layer was saved as “Crests”.
7. Classify Depression: The *Classify Depressions...* command was used with the *predefined mask* view. Input layers and there default values were as follows; Percentile < 0.40 or PlanCurve < -0.50. The output layer was saved as “Depressions”.
8. Connect Depressions: The initial depressions were not connected and therefore the Depression Connector was implemented. The *Connect Depressions...* command was used with the following input layers; DEM, Depressions and Percentile < 0.48. The acceptable elevation difference default of 0.00 was set, which restricts the following cells in downflow direction to equal or lower elevation. The output layer was saved as “ConDepressions”.
9. Classify Flats: The *Classify Flats...* command was used with input layer Grade\_percent. The default threshold for grade percent < 3 percent was

specified and a minimum radius of flat areas of 20m. The output layer was saved as “Flats”.

10. Classify Simple Slope: Simple slopes are the remaining areas not already classified as crest, depressions or flats. The *Classify Simple Slopes...* command with input layers Crests, ConDepressions and Flats was used. The output layer was saved as “SimpleSlopes”.
11. Generate Primary Landform layer: The *Final Classification...* command was used to combine the individual landform layers into one layer and remove noise from the resultant layer. Input layers were Crests, SimpleSlopes, ConDepressions and Flats. The optional noise reduction filter was applied with a 5x5 window and Median filter. The output layer was saved as “PrimaryLandforms”.

#### Generate Slope Classes and Final Landform Classification layer

12. Classify slope classes and Final Landforms: The *Slope Classification...* command was used to break up simple slopes into upper, mid and lower slope, remove noise and combine the resultant layer with the Primary Landforms to produce the Final Landforms layer. Input layers were the DEM and PrimaryLandform layers. The slope difference threshold was set at 0.1 and optional noise reduction filter applied with a 5x5 window and Majority filter. The optional overlay with final results layer was also specified and the output layer saved as “FinalLandforms”.



## **APPENDIX G**

### **IMPLEMENTATION OF COMPOUND TOPOGRAPHIC INDEX**

1. Import DEM: The LandMonitor DEM (smoothed by Caccetta (2000)) is imported into GeoMedia Grid
2. Fill Depression: The *Fill Depressions* command creates a “depressionless” DEM by filling pits, ponds and depressions and is useful in hydrological analysis commands such as *Downhill Path* (Keigan Systems Inc., 2003b).
3. DownHill Path: The *Downhill Path* command produces a grid layer of downhill path directions, in which the value of each cell represents the direction of the steepest slope (Keigan Systems Inc., 2003b). When this command is run on a “depressionless” DEM, flow is possible in one of only eight directions and each cell will be assigned one of nine values (0 = No Flow) as follows;

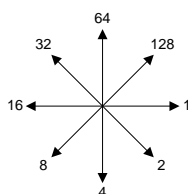


Figure A.2. Flow direction for Downhill path command of GeoMedia Grid.

This simulates the D8 algorithm discussed in Section 5. 3, which downhill path values 1, 4, 16 and 64 are in cardinal flow direction while values 2, 8, 32 and 128 are assigned to diagonal flow direction.

4. Compute Upslope Contributing Area in cells: The *Downhill Accumulation* generates a flow accumulation grid layer from a flow direction grid layer generated from the *Downhill Path* command. The algorithm assigns each cell a value that represents the number of cells on a direct uphill path from that cell (Keigan Systems Inc., 2003b).
5. Compute Upslope Contributing Area in m<sup>2</sup> (A): Using the *Calculator*, upslope contributing area in square metres is determined by multiplying the cell value by the cell area, which in this case is 100m<sup>2</sup> (i.e. 10m cell resolution).
6. a). Extract cells for flow in cardinal directions: Using the *Calculator* cells that represent flow in cardinal directions are extracted from the Downhill Flow Path layer. i.e. Cell value = 1,4,16 and 64 and Cardinal Flow Direction layer created.

- b). Cardinal flow directions are assigned flow width: Assign a value of 10 to cells in Cardinal Flow Direction layer.
7. a). Extract cells for flow in diagonal directions: Using the *Calculator* cells that represent flow in diagonal directions are extracted from the Downhill Flow Path layer. i.e. Cell value = 2,8,32 and 128 and Diagonal Flow Direction layer created.
- b). Diagonal flow directions are assigned flow width: Assign a value of 14.142136 to cells in Diagonal Flow Direction layer.
8. Create Flow Width layer ( $w$ ): Using the *Overlay* command the Cardinal Flow Direction and Diagonal Flow Direction layers are joined to create Flow Width layer, whereby cells that have a cardinal and diagonal flow direction have a value of 10 and 14.142136 respectively.
9. Calculate Specific Catchment Area ( $A_s$ ): Using the *Calculator* command the Upslope contributing area layer is divided by the Flow width layer. Ie ( $A_s = A / w$ ).
10. Generate maximum slope in percent: Use the *Grade* command and specify a maximum slope with an output in percent. These parameters within GeoMedia Grid follow Equation (5.3) of Gallant and Wilson (2000) which is then multiplied by 100 for percent slope. The maximum slope in percent layer is created.
11. Calculate Slope ( $S$ ): Using the *Calculator* command, divide the maximum slope in percent layer by 100 to compute  $S$ .
12. Generate CTI: Using the *Calculator* command the CTI is created as follows:

$$\log_e(A_s/S) \quad (\text{A.1})$$

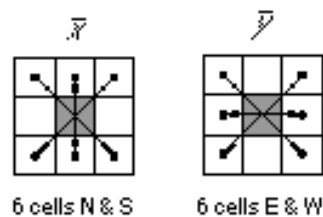
where  $A_s$  is the Specific Catchment Area layer and  $S$  the slope layer.

**APPENDIX H****SLOPE ALGORITHM**

Slope in GeoMedia Grid is calculated using the *Grade* operation.

The user can select Average slope or Maximum slope results, expressed as percents or degrees.

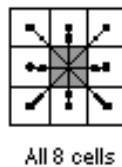
When the Average slope option is selected, Grade determines gradient by calculating the average slopes of the lines that run from north to south and diagonally through the centre cell. It then calculates average slopes for the lines that run from east to west cells and diagonally through the cell centre.



The two averages are squared and added together. The square root of this result is multiplied by 100, yielding the average grade (intersection of the north-south plane with the east-west plane) of the cell:

$$Grade = \sqrt{x^2 + y^2} * 100 \quad (A.2)$$

If the Maximum slope option is selected, the result value is the maximum of the eight slopes:



All 8 cells

(reproduced from Keigan Systems Inc., 2003b)

**APPENDIX I**

**SIGNIFICANT DIFFERENCE BETWEEN OATS AND OATEN HAY CROP  
TYPES**

Determining whether there is a significant difference between the crop type oats and oaten hay was performed by creating a new factor called Crop Type Recode, in which oats and oaten hay have the same value and then Crop Type ( $C_j$ ) is fitted after Crop Type Recode ( $Cr_j$ ) (i.e. adding the difference between oats and oaten hay) and the accumulated variance analysed. The regression model fitted was as follow;

$$\log_e(Y_{ijk}) = \mu + S_i + Y_k + Cr_j + C_j + CrY_{jk} + CY_{jk} \quad (\text{A.3})$$

where  $Y_{ijk}$  is the average interpolated yield for sampling point  $i$  in year  $k$  with crop type  $j$ ;  $\mu$  is intercept;  $S_i$  is the effect of sample point  $i$ ;  $Cr_j$  is the effect of crop  $j$  recoded;  $C_j$  is the effect of crop  $j$ ;  $Y_k$  is the effect of year  $k$ ;  $CrY_{jk}$  is the effect of the interaction between crop  $j$  recoded and year  $k$ ; and  $CY_{jk}$  is the effect of the interaction between crop  $j$  and year  $k$ .

The regression model accounted for 52.2 percent of the variance in log Yield with a standard error of 0.210 and was highly significant ( $P < 0.001$ ). The accumulated analysis of variance is shown in Table A.2.

Table A.2 Accumulated analysis of variance.

Change	d.f.	Sum of squares	Mean Square	Variance Ratio F	Significance
+ Sample Pt	183	20.27636	0.11080	2.51	<0.001
+ Year	6	8.26105	1.37684	31.16	<0.001
+ CropTRecode	5	9.78625	1.95725	44.30	<0.001
+ CropType	1	0.55424	0.55424	12.54	<0.001
+ Year.CropTRecode	20	3.86786	0.19339	4.38	<0.001
+ Year.CropType	1	0.02903	0.02903	0.66	0.418
Residual	474	20.94360	0.04418		
Total	690	63.71839	0.09235		

As shown in Table A.2 the difference between CropTRecode ( $Cr_j$ ) and Crop Type ( $C_j$ ) is highly significant ( $P < 0.001$ ) and as such the difference between oats and oaten hay is highly significant indicating that sample points 167, 169 and 176 should be removed from the data set.

**APPENDIX J**

**ESTIMATED YIELD AT SAMPLE POINTS**



Sample Point NO.	Estimated LogYld	Std Error	Estim_Yld (kg/ha)
1	8.0140	0.0885	2023.0489
2	7.9782	0.0973	1916.6664
8	7.7169	0.1239	1245.8941
9	7.8208	0.1191	1491.9024
10	7.8269	0.0942	1507.1067
12	7.6303	0.0861	1059.7261
13	8.0158	0.1236	2028.3078
14	7.5642	0.0948	927.9671
15	7.9685	0.1191	1888.5328
16	8.0388	0.1191	2098.8000
17	7.7904	0.1191	1417.3162
19	7.6765	0.1454	1156.9928
20	7.8102	0.1191	1465.6160
23	7.8590	0.1191	1589.0271
24	7.7908	0.1053	1418.1661
25	7.9269	0.1454	1770.7969
26	7.9057	0.1459	1712.5779
28	7.8809	0.1191	1646.3007
29	7.9713	0.1191	1896.7292
30	8.1194	0.1454	2359.0518
31	8.0252	0.1454	2057.0305
32	8.0056	0.1191	1997.7870
33	8.0019	0.1191	1986.6721
34	7.7391	0.1053	1296.4251
35	8.1300	0.1454	2394.8246
36	7.8590	0.1053	1588.8170
37	7.7519	0.1454	1326.0750
40	7.7779	0.1053	1387.2960
41	7.9778	0.2055	1915.5298
42	7.8049	0.1053	1452.5924
43	7.8257	0.1053	1504.1659
44	7.7245	0.1224	1263.1670
45	8.0975	0.0938	2286.3633
46	7.6700	0.1113	1143.0124
47	7.8967	0.1290	1688.2681
49	7.7793	0.1250	1390.6507
50	7.7772	0.1113	1385.6401
51	7.7723	0.1290	1373.9357
53	8.0538	0.1271	2145.5772
55	8.0403	0.1054	2103.5497
57	7.7480	0.1250	1316.8223
59	8.0367	0.1054	2092.3263
60	7.9564	0.1271	1853.6589

Sample Point NO.	Estimated LogYld	Std Error	Estim_Yld (kg/ha)
63	7.6591	0.1075	1119.9370
65	7.6517	0.1250	1104.2647
66	7.4911	0.1051	792.0514
67	7.5989	0.1051	995.9302
69	7.8536	0.1051	1574.9482
70	7.6441	0.1051	1088.2034
71	7.8003	0.1051	1441.4496
74	7.5725	0.1051	944.0700
76	7.5635	0.1202	926.6537
80	8.0878	0.1203	2254.4896
81	7.7401	0.1044	1298.7771
84	7.6196	0.1051	1037.8459
85	7.8574	0.1215	1584.8587
86	7.5598	0.1051	919.4750
87	7.7602	0.0940	1345.3764
88	7.6973	0.1219	1202.3060
89	7.6824	0.1051	1169.9253
91	7.8177	0.1202	1484.1095
92	7.6036	0.2069	1005.4935
93	8.1396	0.2075	2427.5651
94	7.7970	0.0940	1433.2760
95	7.8337	0.0940	1524.3169
99	7.8228	0.0946	1496.8832
100	7.8913	0.1050	1673.8546
101	7.8086	0.1050	1461.7999
102	7.8419	0.1050	1544.9289
103	7.8382	0.0997	1535.7575
104	7.8706	0.0940	1619.1802
105	7.5824	0.0997	963.2949
106	7.9909	0.0946	1954.0729
107	7.8627	0.0946	1598.4781
108	7.7491	0.0780	1319.4683
109	7.8737	0.1214	1627.2443
110	7.8024	0.0946	1446.5370
111	7.4652	0.1207	746.1670
112	7.4680	0.1207	751.1632
113	7.9295	0.0946	1778.1579
114	7.8893	0.1237	1668.5610
116	7.6912	0.1237	1189.0902
117	7.7631	0.1237	1352.2620
118	7.8821	0.1203	1649.3010
119	7.6295	0.1207	1058.0499
120	7.8470	0.1042	1558.1106

Sample Point NO.	Estimated LogYld	Std Error	Estim_Yld (kg/ha)
121	7.7490	0.1042	1319.2392
122	7.8135	0.1042	1473.6766
123	7.5473	0.1494	895.6291
124	7.7688	0.1042	1365.5637
126	7.6440	0.1042	1088.0937
127	7.8924	0.1203	1676.8651
128	7.9946	0.0841	1964.9947
129	7.9181	0.0780	1746.5799
130	7.8800	0.1042	1643.8689
131	7.7767	0.2143	1384.2871
132	7.8486	0.2075	1562.1261
133	7.8555	0.1042	1579.9643
134	7.9420	0.1042	1812.9918
135	7.8938	0.1042	1680.7121
136	7.4264	0.0780	679.8249
137	8.0039	0.1528	1992.5414
139	7.9998	0.0780	1980.3159
141	7.9259	0.1528	1767.9830
148	7.8634	0.1528	1600.2223
149	7.8364	0.0780	1531.1124
150	7.9482	0.0949	1830.3986
152	7.5147	0.0780	834.7726
153	7.6149	0.0780	1028.1474
155	8.1548	0.2065	2479.8992
157	7.8930	0.1071	1678.4049
158	7.9116	0.1071	1728.7455
159	7.7448	0.0848	1309.4949
161	7.6929	0.1475	1192.7632
162	7.6183	0.2107	1035.1088
163	7.6058	0.0848	1009.8102
164	7.6821	0.0848	1169.0990
165	7.7187	0.0848	1250.1240
170	7.7085	0.0859	1227.1298
171	7.9439	0.0848	1818.2840
172	7.6521	0.0931	1105.1345
173	7.9348	0.0859	1792.9184
177	7.7936	0.0859	1425.1406
178	7.9806	0.1039	1923.6517
179	7.8950	0.1032	1683.7924
181	7.7864	0.0940	1407.7280
184	7.6390	0.1214	1077.6274
186	7.7187	0.0859	1250.1323
187	8.0759	0.2057	2216.0202
189	8.1211	0.2055	2364.7765
190	7.7970	0.0945	1433.2351
191	8.0938	0.2193	2273.9907
192	7.9620	0.2055	1869.6751
193	8.1417	0.2055	2434.6354

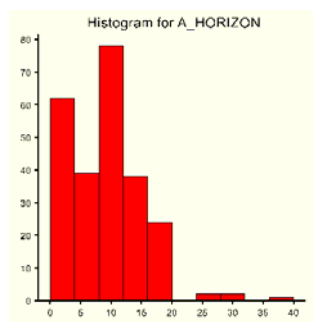
Sample Point NO.	Estimated LogYld	Std Error	Estim_Yld (kg/ha)
194	7.7019	0.1202	1212.5613
196	7.9136	0.2054	1734.1349
197	7.7744	0.2055	1378.9475
198	7.8330	0.1230	1522.5909
199	7.8759	0.1230	1632.9924
200	7.9143	0.1230	1736.1148
201	8.0448	0.1038	2117.5974
202	7.9285	0.0945	1775.3148
203	8.0000	0.0857	1981.0887
205	7.9280	0.0927	1773.7619
206	7.7615	0.0953	1348.3486
207	7.9612	0.0953	1867.3858
208	7.8987	0.1230	1693.7820
209	7.8525	0.0945	1572.1886
212	7.9840	0.0927	1933.5431
213	8.0377	0.1038	2095.5380
215	8.0253	0.1230	2057.4401
217	7.9883	0.2101	1946.1821
218	7.8578	0.2057	1585.9155
219	7.7924	0.2057	1422.1124
221	7.8876	0.2055	1664.0244
222	8.1207	0.1473	2363.4603
223	8.0095	0.2057	2009.3175
224	7.7633	0.1204	1352.6053
225	7.7699	0.1038	1368.2678
226	7.9695	0.1038	1891.5396
227	7.4502	0.1455	720.1545
229	7.7563	0.0947	1336.1428
230	7.9938	0.0857	1962.4441
231	7.5633	0.2066	926.1769
232	7.7544	0.1196	1331.8542
233	7.9553	0.1039	1850.6738
234	7.9750	0.1039	1907.4035
236	7.9595	0.0857	1862.5122
237	7.9761	0.0857	1910.5721
238	7.8267	0.0947	1506.6606
239	8.0980	0.2047	2287.9845
240	7.8620	0.0959	1596.7745
241	7.8770	0.1067	1636.0423
242	7.4407	0.0959	703.9177
244	7.7039	0.0959	1216.8832
245	7.7004	0.0959	1209.2794
246	7.8368	0.0959	1532.1602
247	7.6966	0.0959	1200.9558
248	7.7509	0.0959	1323.7519
249	7.7799	0.1203	1392.0521
250	7.6421	0.0959	1084.0712

**APPENDIX K**

**SOIL VARIABLES TRANSFORMATIONS**

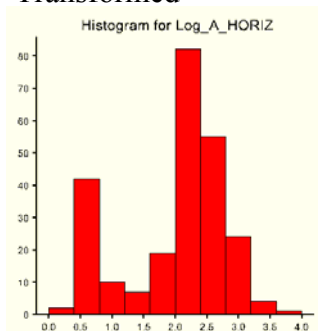
## Stable Soil Properties: Data Summary.

### A Horizon: Untransformed

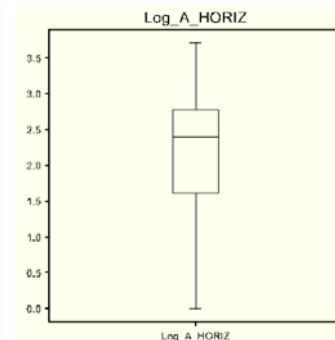


<b><i>n</i></b>	246
<b>Mean</b>	9.45
<b>Median</b>	10
<b>SD</b>	6.59
<b>Skewness</b>	0.69

### Transformed



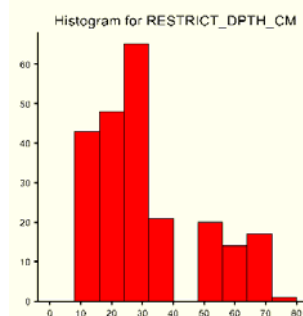
<b><i>n</i></b>	246
<b>Mean</b>	2.064
<b>Median</b>	2.398
<b>SD</b>	0.858
<b>Skewness</b>	-0.814



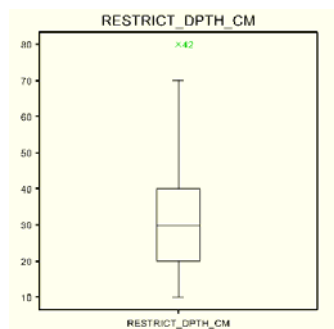
<b>No. of Mild Outliers</b>	0
<b>No. of Extreme Outliers</b>	0

Log A-Horizon =  $\log_e(\text{A-Horizon}+1)$

### RESTRICTED DEPTH

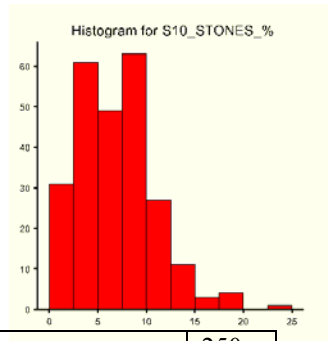


<b><i>n</i></b>	229
<b>Mean</b>	31.88
<b>Median</b>	30.00
<b>SD</b>	17.78
<b>Skewness</b>	0.80

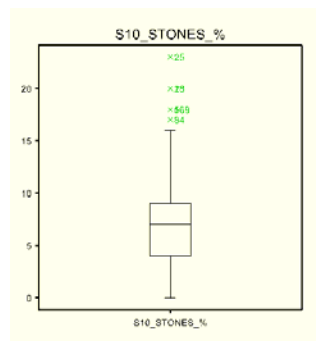


<b>No. of Mild Outliers</b>	1
<b>No. of Extreme Outliers</b>	0

### 10cm STONES

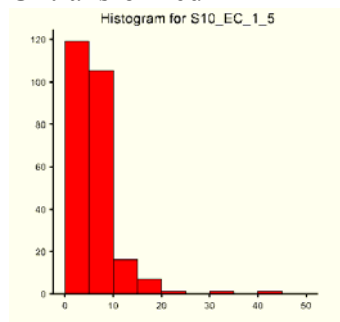


<b><i>n</i></b>	250
<b>Mean</b>	7.13
<b>Median</b>	7
<b>SD</b>	3.95
<b>Skewness</b>	0.76



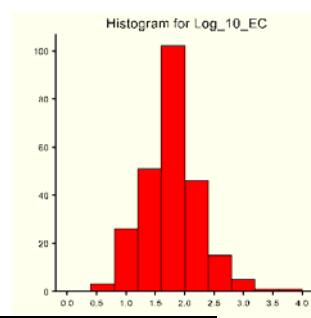
<b>No. of Mild Outliers</b>	
<b>No. of Extreme Outliers</b>	0

### 10cm EC(1:5) Untransformed

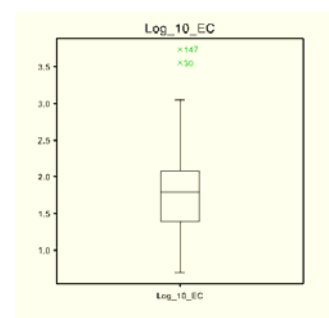


<b><i>n</i></b>	250
<b>Mean</b>	6.69
<b>Median</b>	6
<b>SD</b>	4.37
<b>Skewness</b>	3.85

### Transformed



<b><i>n</i></b>	250
<b>Mean</b>	1.769
<b>Median</b>	1.792
<b>SD</b>	0.481
<b>Skewness</b>	0.729

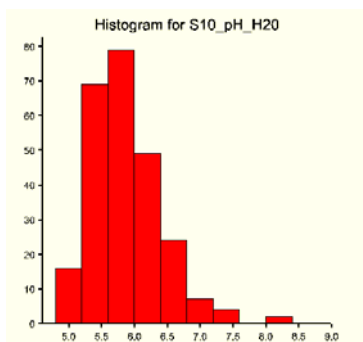


<b>No. of Mild Outliers</b>	2
<b>No. of Extreme Outliers</b>	0

$\text{Log EC} = \log_e(\text{EC})$

### 10cm pH (H<sub>2</sub>O)

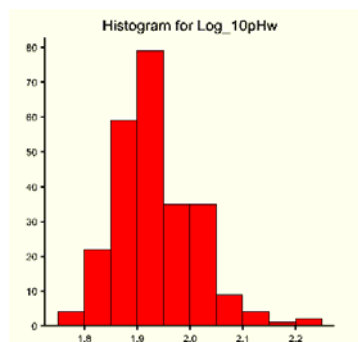
Untransformed



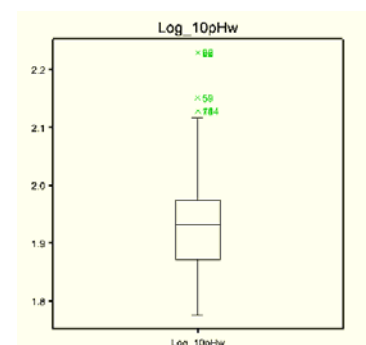
<b><i>n</i></b>	250
<b>Mean</b>	5.936
<b>Median</b>	5.900
<b>SD</b>	0.544
<b>Skewness</b>	1.051

$$\text{Log } 10_{\text{pHw}} = \log_e(10_{\text{pHw}}+1)$$

### Transformed



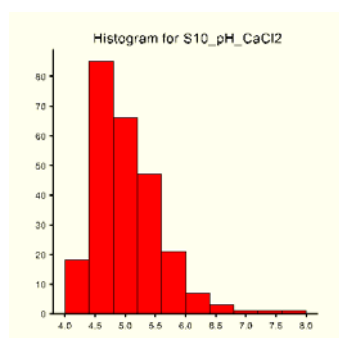
<b><i>n</i></b>	250
<b>Mean</b>	1.934
<b>Median</b>	1.932
<b>SD</b>	0.076
<b>Skewness</b>	0.737



<b>No. of Mild Outliers</b>	
<b>No. of Extreme Outliers</b>	0

### 10cm pH (CaCl<sub>2</sub>)

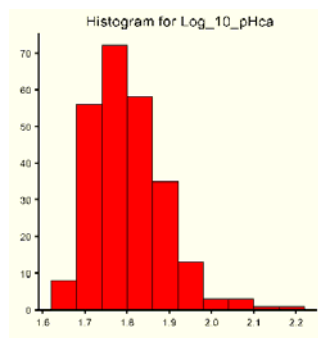
Untransformed



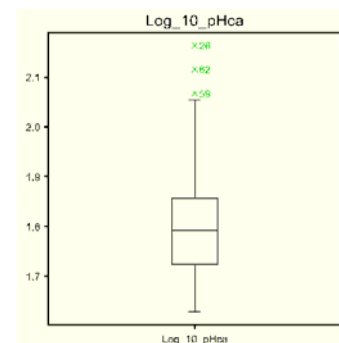
<b><i>n</i></b>	250
<b>Mean</b>	5.075
<b>Median</b>	5.0
<b>SD</b>	0.562
<b>Skewness</b>	1.221

$$\text{Log } 10_{\text{pHca}} = \log_e(10_{\text{pHca}}+1)$$

### Transformed

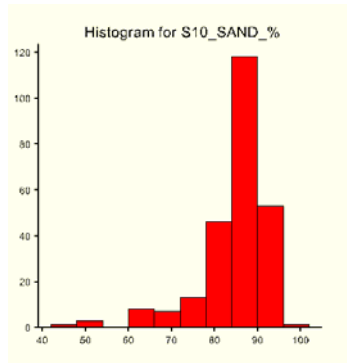


<b><i>n</i></b>	250
<b>Mean</b>	1.8
<b>Median</b>	1.792
<b>SD</b>	0.089
<b>Skewness</b>	0.880



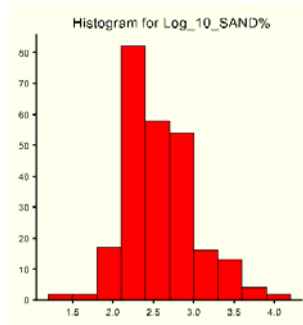
<b>No. of Mild Outliers</b>	3
<b>No. of Extreme Outliers</b>	0

**10 SAND %**  
Untransformed

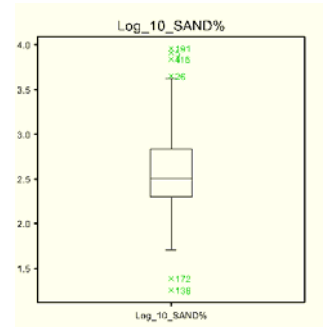


<b><i>n</i></b>	250
<b>Mean</b>	85.23
<b>Median</b>	87.75
<b>SD</b>	7.87
<b>Skewness</b>	-2.15

Transformed



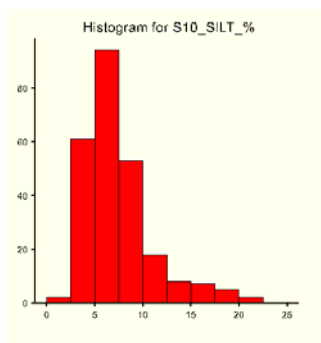
<b><i>n</i></b>	250
<b>Mean</b>	2.588
<b>Median</b>	2.505
<b>SD</b>	0.437
<b>Skewness</b>	0.646



<b>No. of Mild Outliers</b>	7
<b>No. of Extreme Outliers</b>	0

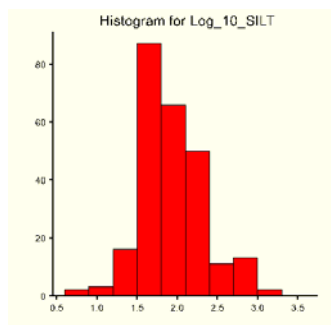
$\text{Log}_{10\_SAND\%} = \text{Log}_e(100 - 10\_SAND)$

**10 Silt %**  
Untransformed

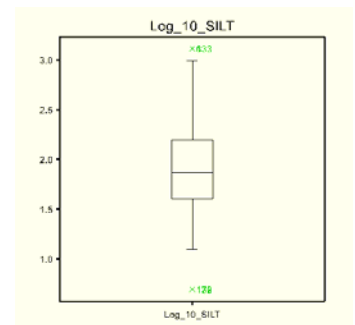


<b><i>n</i></b>	250
<b>Mean</b>	7.62
<b>Median</b>	6.50
<b>SD</b>	3.56
<b>Skewness</b>	1.70

Transformed



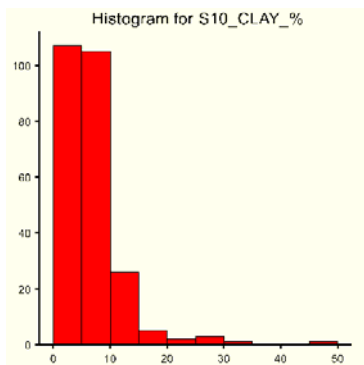
<b><i>n</i></b>	250
<b>Mean</b>	1.942
<b>Median</b>	1.872
<b>SD</b>	0.412
<b>Skewness</b>	0.368



<b>No. of Mild Outliers</b>	433
<b>No. of Extreme Outliers</b>	128

$\text{Log}_{10\_Silt} = \text{log}_e(10\_Silt)$

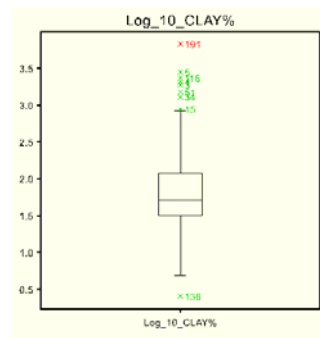
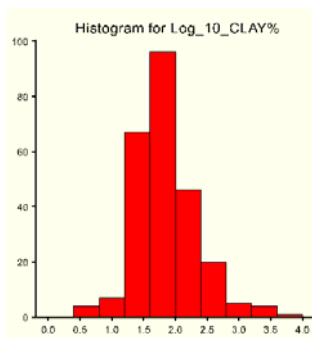
**10 Clay %**  
Untransformed



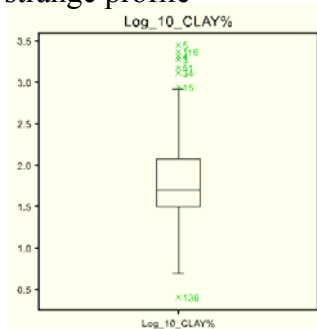
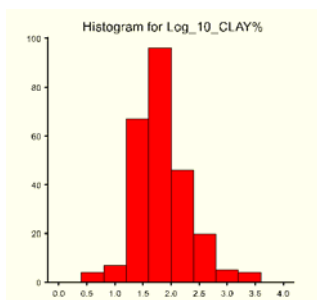
<b><i>n</i></b>	250
<b>Mean</b>	7.14
<b>Median</b>	5.50
<b>SD</b>	5.04
<b>Skewness</b>	3.57

$\text{Log}_{10\_Clay} = \log_e(10\_Clay)$

Transformed



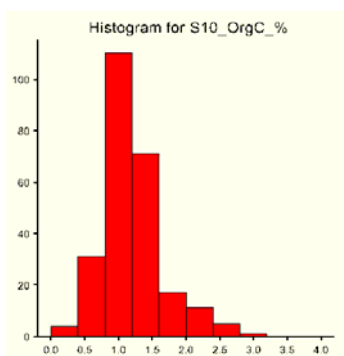
Extreme outlier removed as had a look at point 191 notes and remember it was a very strange profile



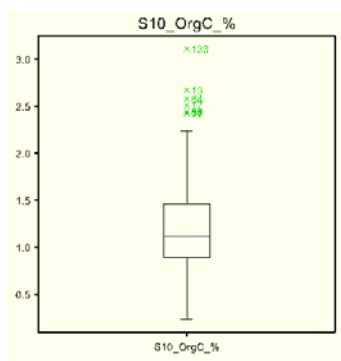
<b><i>n</i></b>	249
<b>Mean</b>	1.811
<b>Median</b>	1.705
<b>SD</b>	0.486
<b>Skewness</b>	0.714

<b>No. of Mild Outliers</b>	
<b>No. of Extreme Outliers</b>	0

**10 Org C%**



<b><i>n</i></b>	250
<b>Mean</b>	1.204
<b>Median</b>	1.120
<b>SD</b>	0.448
<b>Skewness</b>	0.944



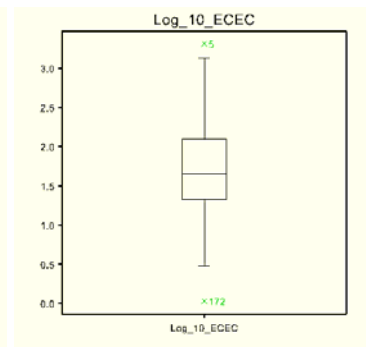
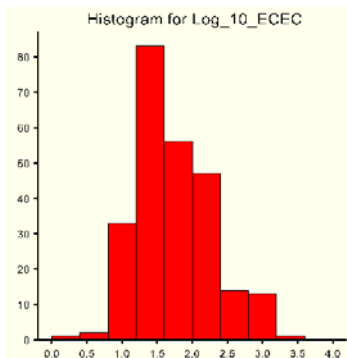
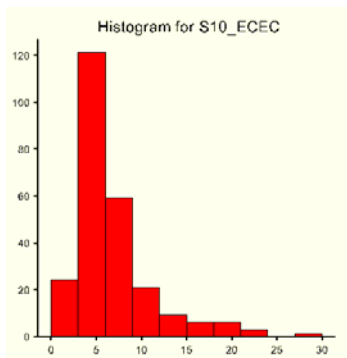
<b>No. of Mild Outliers</b>	4+
<b>No. of Extreme Outliers</b>	0



### 10 ECEC

Untransformed

Transformed



<b><i>n</i></b>	250	<b><i>n</i></b>	250
<b>Mean</b>	6.63	<b>Mean</b>	1.733
<b>Median</b>	5.20	<b>Median</b>	1.649
<b>SD</b>	4.30	<b>SD</b>	0.541
<b>Skewness</b>	2.02	<b>Skewness</b>	0.480

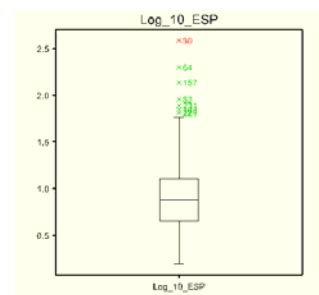
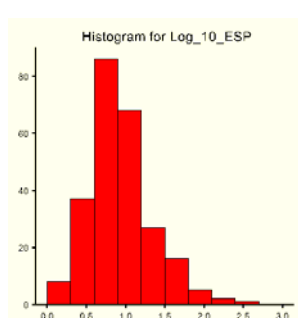
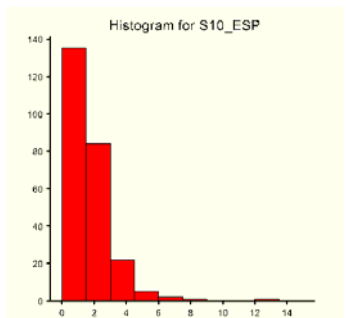
<b>No. of Mild Outliers</b>	2
<b>No. of Extreme Outliers</b>	0

$\text{Log } 10\_ECEC = \log_e(10\_ECEC)$

### 10 ESP

Untransformed

Transformed



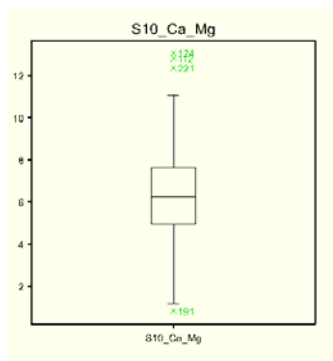
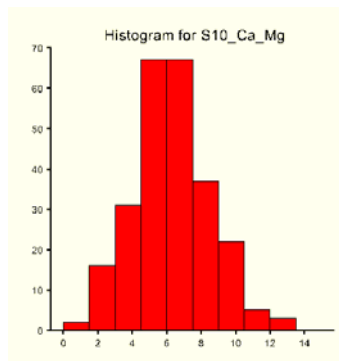
<b><i>n</i></b>	250
<b>Mean</b>	1.77
<b>Median</b>	1.40
<b>SD</b>	1.39
<b>Skewness</b>	3.14

<b><i>n</i></b>	250
<b>Mean</b>	0.934
<b>Median</b>	0.877
<b>SD</b>	0.387
<b>Skewness</b>	0.931

<b>No. of Mild Outliers</b>	
<b>No. of Extreme Outliers</b>	1

$\text{Log } 10\_ESP = \log_e(10\_ESP+1)$

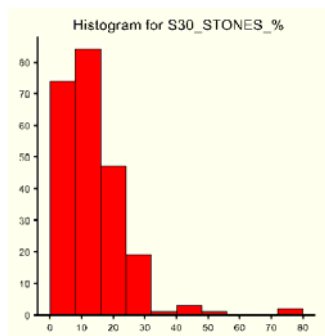
### 10 Ca/Mg



<b>n</b>	250
<b>Mean</b>	6.351
<b>Median</b>	6.261
<b>SD</b>	2.264
<b>Skewness</b>	0.188

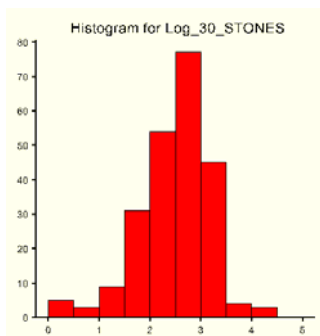
<b>No. of Mild Outliers</b>	4
<b>No. of Extreme Outliers</b>	0

### 30 STONES% Untransformed

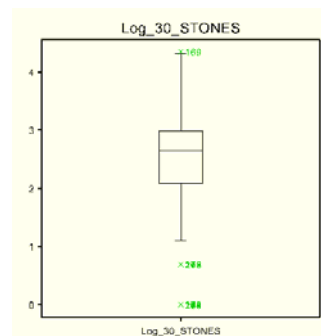


<b>n</b>	231
<b>Mean</b>	14.11
<b>Median</b>	13
<b>SD</b>	10.38
<b>Skewness</b>	2.34

### Transformed



<b>n</b>	231
<b>Mean</b>	2.488
<b>Median</b>	2.639
<b>SD</b>	0.735
<b>Skewness</b>	-0.908

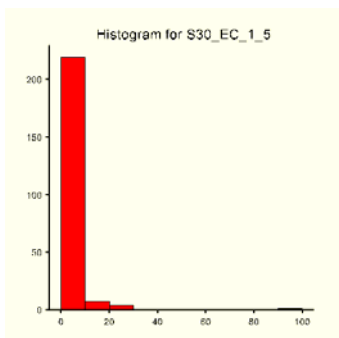


<b>No. of Mild Outliers</b>	2
<b>No. of Extreme Outliers</b>	0

$\text{Log } 30\_STONES = \log_e(30\_STONES+1)$

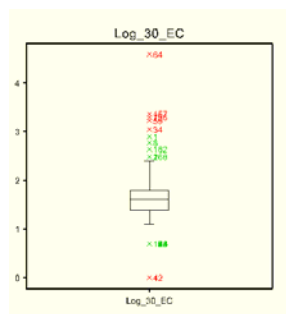
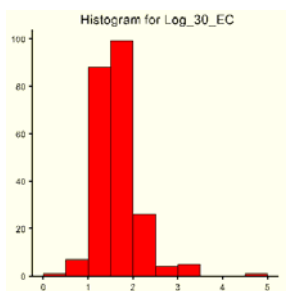
**30 EC(1:5)**

Untransformed



<b><i>n</i></b>	231
<b>Mean</b>	5.13
<b>Median</b>	4
<b>SD</b>	7.19
<b>Skewness</b>	9.59

Transformed

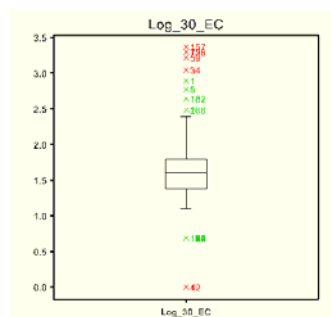
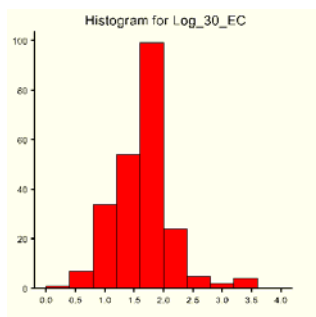


Pt 64 very extreme outlier removed.

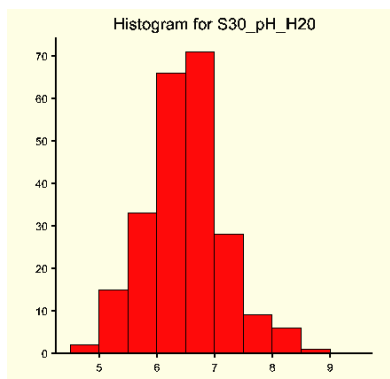
<b><i>n</i></b>	230
<b>Mean</b>	1.614
<b>Median</b>	1.609
<b>SD</b>	0.474
<b>Skewness</b>	0.830

<b>No. of Mild Outliers</b>	
<b>No. of Extreme Outliers</b>	

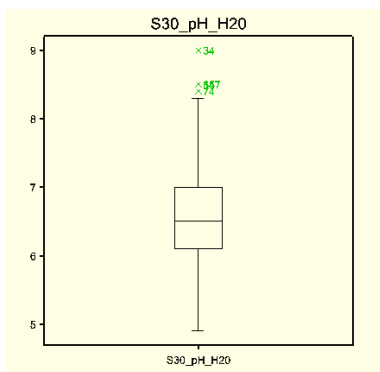
$\text{Log } 30\_EC = \log_e(30\_EC+1)$



**30 pH H<sub>2</sub>O**

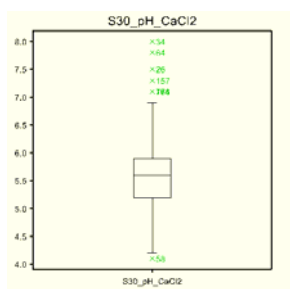
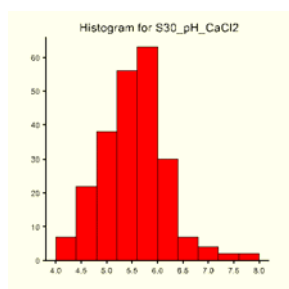


<b><i>n</i></b>	231
<b>Mean</b>	6.549
<b>Median</b>	6.500
<b>SD</b>	0.693
<b>Skewness</b>	0.418



<b>No. of Mild Outliers</b>	
<b>No. of Extreme Outliers</b>	0

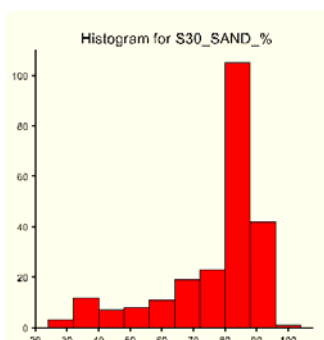
### 30 pH CaCl<sub>2</sub>



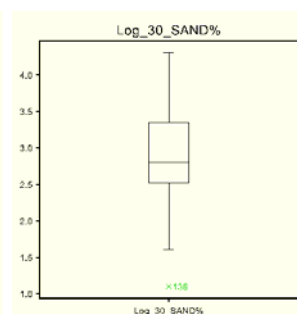
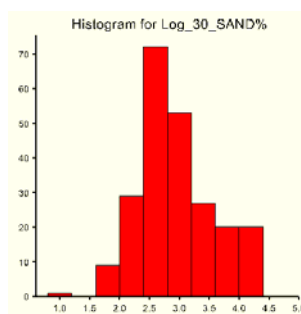
<b><i>n</i></b>	231
<b>Mean</b>	5.589
<b>Median</b>	5.6
<b>SD</b>	0.630
<b>Skewness</b>	0.494

<b>No. of Mild Outliers</b>	
<b>No. of Extreme Outliers</b>	0

### 30 SAND% Untransformed



### Transformed



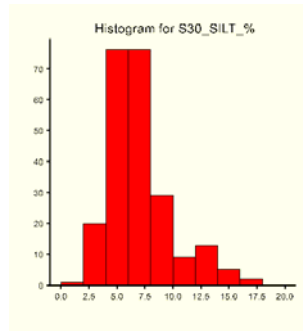
<b><i>n</i></b>	231
<b>Mean</b>	77.10
<b>Median</b>	83.50
<b>SD</b>	16.20
<b>Skewness</b>	-1.50

<b><i>n</i></b>	231
<b>Mean</b>	2.928
<b>Median</b>	2.803
<b>SD</b>	0.622
<b>Skewness</b>	0.348

<b>No. of Mild Outliers</b>	1
<b>No. of Extreme Outliers</b>	0

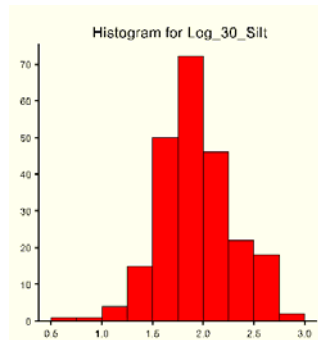
$\text{Log}_{30\_SAND} = \log_e(100 - 30\_SAND)$

**30 SILT%**  
Untransformed

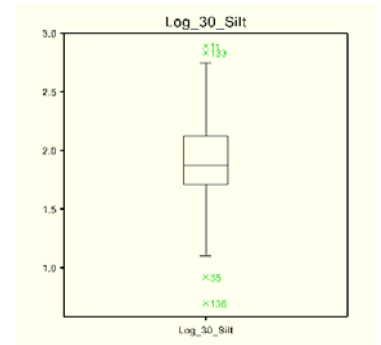


<b><i>n</i></b>	231
<b>Mean</b>	7.240
<b>Median</b>	6.50
<b>SD</b>	2.864
<b>Skewness</b>	1.221

Transformed



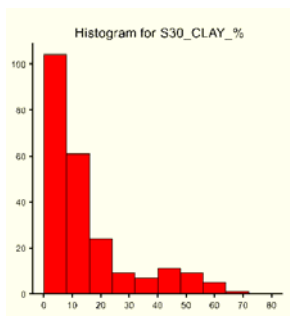
<b><i>n</i></b>	231
<b>Mean</b>	1.909
<b>Median</b>	1.872
<b>SD</b>	0.375
<b>Skewness</b>	0.078



<b>No. of Mild Outliers</b>	4
<b>No. of Extreme Outliers</b>	0

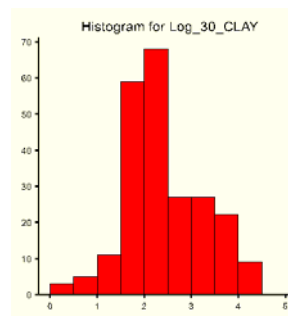
$\text{Log } 30\_Silt = \log_e(30\_Silt)$

**30 Clay%**  
Untransformed

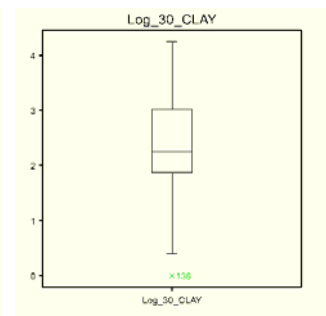


<b><i>n</i></b>	231
<b>Mean</b>	15.66
<b>Median</b>	9.5
<b>SD</b>	14.79
<b>Skewness</b>	1.68

Transformed



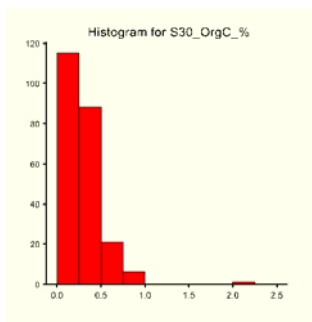
<b><i>n</i></b>	231
<b>Mean</b>	2.395
<b>Median</b>	2.251
<b>SD</b>	0.830
<b>Skewness</b>	0.253



<b>No. of Mild Outliers</b>	1
<b>No. of Extreme Outliers</b>	0

$\text{Log } 30\_CLAY = \log_e(30\_CLAY)$

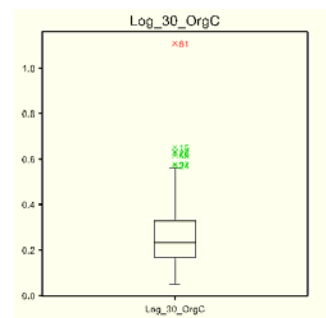
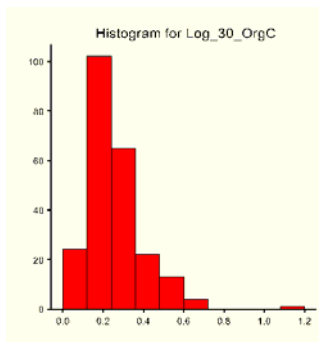
**30 Orc C**  
Untransformed



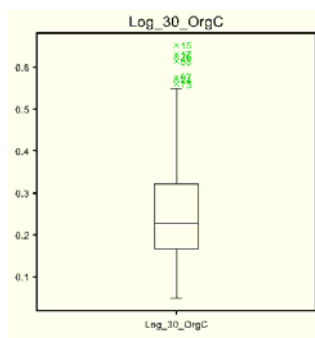
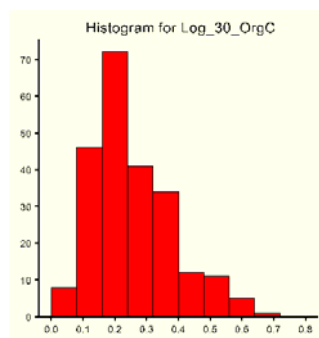
<b><i>n</i></b>	231
<b>Mean</b>	0.309
<b>Median</b>	0.260
<b>SD</b>	0.208
<b>Skewness</b>	3.140

$\text{Log } 30\_OrgC = \log_e(30\_OrgC+1)$

Transformed



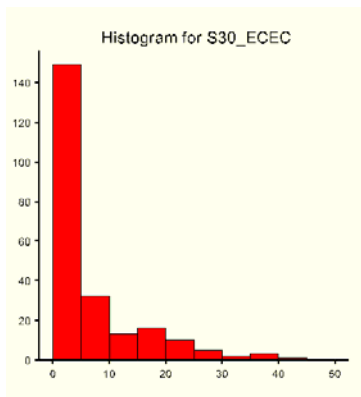
Remove 1 extreme outlier Pt 81:



<b><i>n</i></b>	230
<b>Mean</b>	0.255
<b>Median</b>	0.227
<b>SD</b>	0.126
<b>Skewness</b>	0.930

<b>No. of Mild Outliers</b>	
<b>No. of Extreme Outliers</b>	0

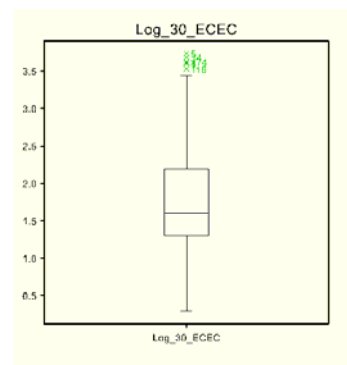
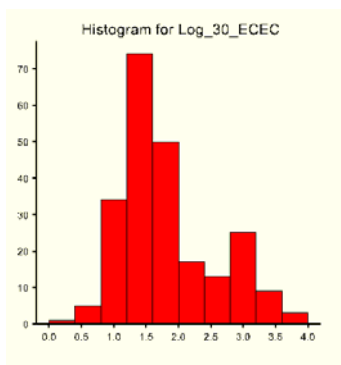
**30 ECEC**  
Untransformed



<b><i>n</i></b>	231
<b>Mean</b>	7.40
<b>Median</b>	3.98
<b>SD</b>	7.93
<b>Skewness</b>	2.01

$\text{Log } 30\_ECEC = \log_e(30\_ECEC+1)$

Transformed

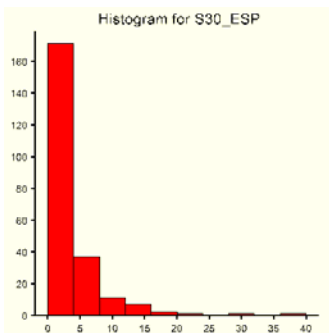


<b><i>n</i></b>	231
<b>Mean</b>	1.820
<b>Median</b>	1.605
<b>SD</b>	0.729
<b>Skewness</b>	0.814

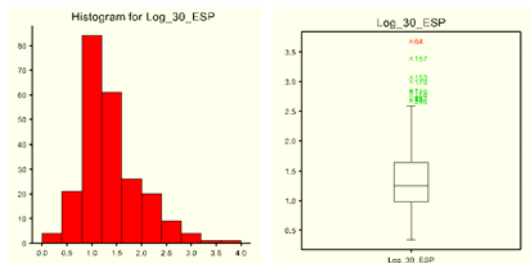
<b>No. of Mild Outliers</b>	
<b>No. of Extreme Outliers</b>	0

### 30 ESP

Untransformed



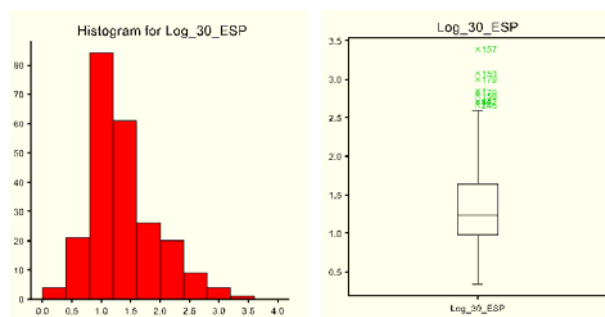
Transformed



Remove extreme outlier: Pt 64 ESP of 10 greater than other points. Pt along creek line also.

<b><i>n</i></b>	231
<b>Mean</b>	3.86
<b>Median</b>	2.48
<b>SD</b>	4.39
<b>Skewness</b>	3.94

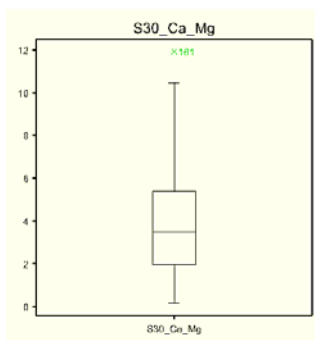
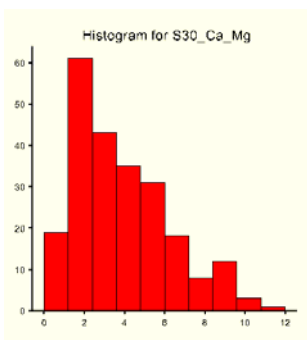
$\text{Log } 30\_ESP = \log_e(30\_ESP+1)$



<b><i>n</i></b>	230
<b>Mean</b>	1.356
<b>Median</b>	1.238
<b>SD</b>	0.570
<b>Skewness</b>	0.935

<b>No. of Mild Outliers</b>	
<b>No. of Extreme Outliers</b>	0

### 30 Ca Mg



<b><i>n</i></b>	231
<b>Mean</b>	3.887
<b>Median</b>	3.493
<b>SD</b>	2.376
<b>Skewness</b>	0.860

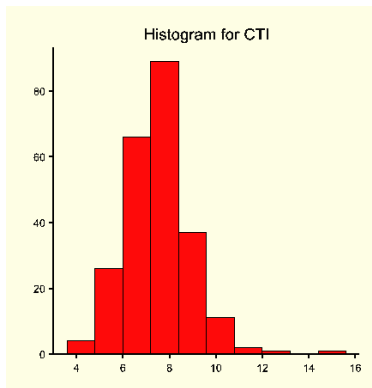
<b>No. of Mild Outliers</b>	1
<b>No. of Extreme Outliers</b>	0

**APPENDIX L****ANCILLARY VARIABLES TRANSFORMATIONS**

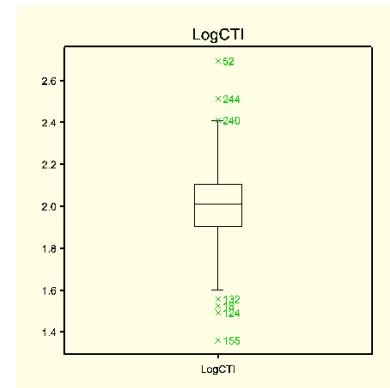
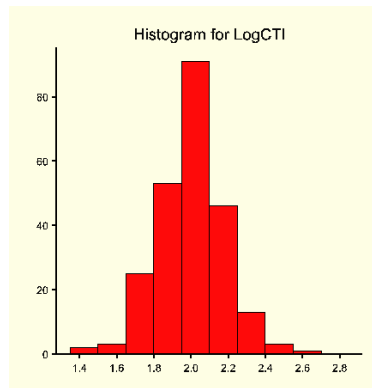


**CTI:**

Untransformed



Transformed



<b><i>n</i></b>	237
<b>Mean</b>	7.531
<b>Median</b>	7.450
<b>SD</b>	1.377
<b>Skewness</b>	0.843

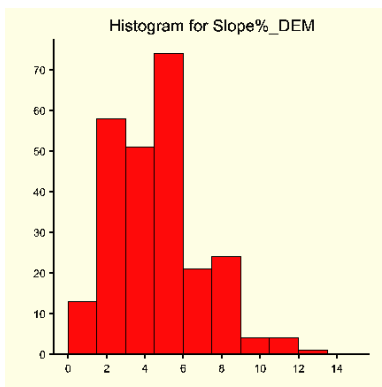
<b><i>n</i></b>	237
<b>Mean</b>	2.003
<b>Median</b>	2.008
<b>SD</b>	0.181
<b>Skewness</b>	-0.085

<b>No. of Mild Outliers</b>	7
<b>No. of Extreme Outliers</b>	

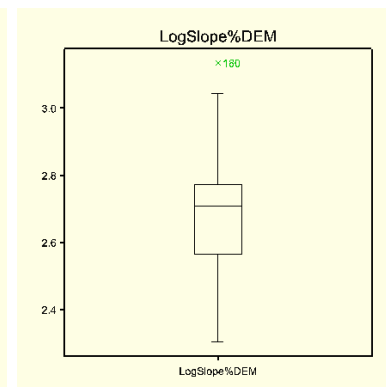
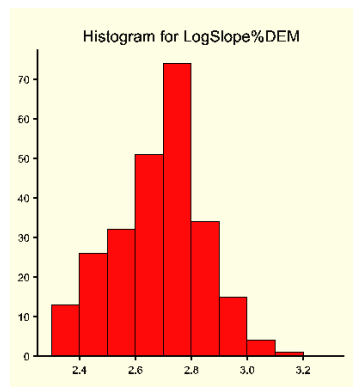
$\text{Log CTI} = \log_e(\text{CTI})$

**Slope%\_DEM:**

Untransformed



Transformed

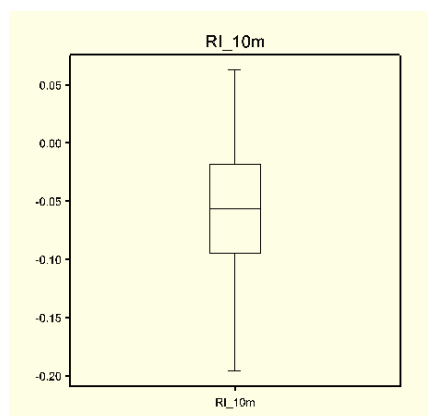
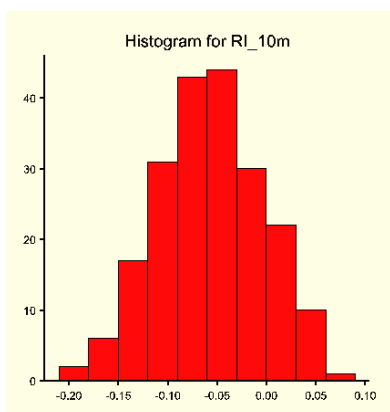


<b><i>n</i></b>	250
<b>Mean</b>	4.832
<b>Median</b>	5.000
<b>SD</b>	2.324
<b>Skewness</b>	0.591

<b><i>n</i></b>	250
<b>Mean</b>	2.685
<b>Median</b>	2.708
<b>SD</b>	0.154
<b>Skewness</b>	0.167

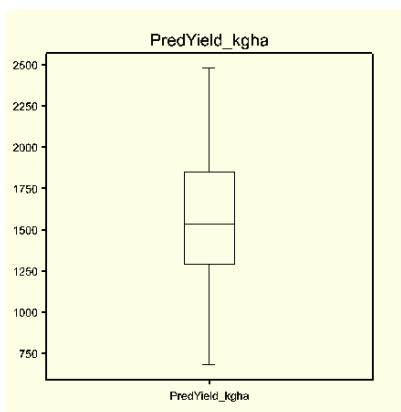
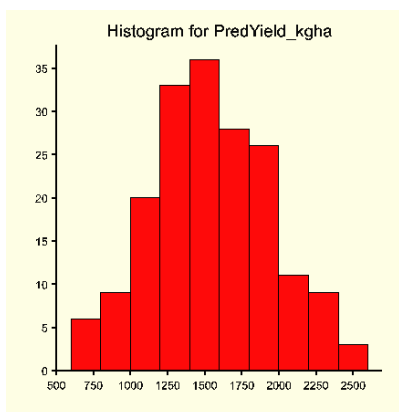
<b>No. of Mild Outliers</b>	1
<b>No. of Extreme Outliers</b>	0

$\text{LogSlope\%DEM} = \log_e(\text{Slope\%DEM}+10)$

**RI:**

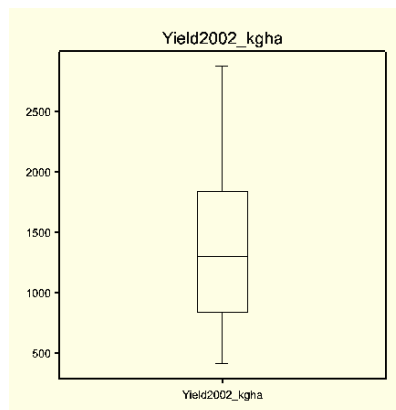
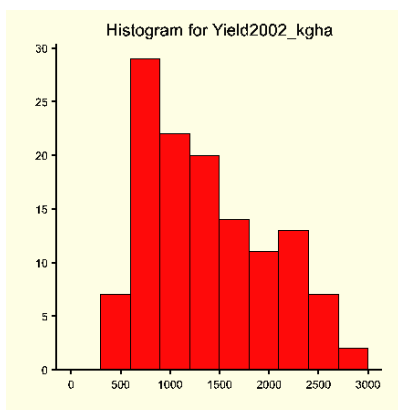
<b><i>n</i></b>	206
<b>Mean</b>	-0.056
<b>Median</b>	-0.057
<b>SD</b>	0.0535
<b>Skewness</b>	-0.0921

<b>No. of Mild Outliers</b>	0
<b>No. of Extreme Outliers</b>	0

**PredYield kg/ha:**

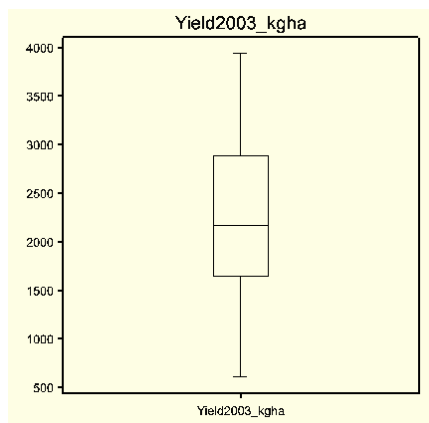
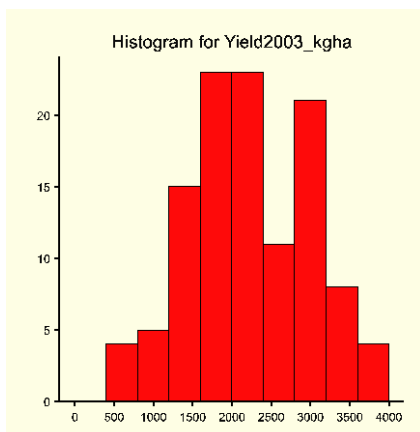
<b><i>n</i></b>	181
<b>Mean</b>	1549.0
<b>Median</b>	1532.2
<b>SD</b>	400.1
<b>Skewness</b>	0.1

<b>No. of Mild Outliers</b>	0
<b>No. of Extreme Outliers</b>	0

**Yield 2002 kg/ha:**

<b><i>n</i></b>	125
<b>Mean</b>	1382.1
<b>Median</b>	1298.8
<b>SD</b>	628.4
<b>Skewness</b>	0.5

<b>No. of Mild Outliers</b>	0
<b>No. of Extreme Outliers</b>	0

**Yield 2003 kg/ha:**

<b><i>n</i></b>	114
<b>Mean</b>	2231
<b>Median</b>	2169
<b>SD</b>	770
<b>Skewness</b>	0

<b>No. of Mild Outliers</b>	0
<b>No. of Extreme Outliers</b>	0

**APPENDIX M**

**CORRELATIONS BETWEEN SOIL VARIABLES**

Variable n=204	Log_A_HORIZ Correlation	Log_A_HORIZ Probability	RESTRICT DPTH_CM Correlation	RESTRICT DPTH_CM Probability	S10_STONES_% Correlation	S10_STONES_% Probability	Log_10_EC Correlation	Log_10_EC Probability
Log_A_HORIZ	1.000							
RESTRICT_DPTH_CM	0.163	*	1.000					
S10_STONES_%	-0.017	0.809	-0.310	**	1.000			
Log_10_EC	0.056	0.429	-0.213	**	0.265	**	1.000	
Log_10pHw	-0.279	**	-0.096	0.170	-0.070	0.322	0.285	**
Log_10_pHca	-0.165	*	-0.092	0.190	-0.049	0.488	0.445	**
Log_10_SAND% <sup>a</sup>	-0.128	0.069	-0.356	**	0.341	**	0.423	**
Log_10_SILT	-0.158	*	-0.359	**	0.318	v	0.404	**
Log_10_CLAY%	-0.088	0.211	-0.322	**	0.319	**	0.396	**
S10_OrgC_%	-0.051	0.469	-0.190	**	0.192	**	0.603	**
Log_10_ECEC	-0.163	*	-0.288	**	0.226	**	0.622	**
Log_10_ESP	0.098	0.164	0.072	0.308	-0.042	0.548	0.014	0.848
S10_Ca_Mg	0.053	0.449	0.025	0.720	-0.046	0.515	-0.170	*
Log_30_STONES	-0.175	**	-0.385	**	0.432	**	0.091	0.196
Log_30_EC	-0.003	0.969	-0.241	**	0.214	**	0.480	**
S30_pH_H2O	-0.160	*	-0.071	0.310	0.007	0.923	0.177	**
S30_pH_CaCl2	-0.134	0.057	-0.092	0.189	0.055	0.435	0.217	**
Log_30_SAND%	-0.085	0.225	-0.356	**	0.297	**	0.349	**
Log_30_Silt	-0.056	0.423	-0.393	**	0.238	**	0.258	**
Log_30_CLAY	-0.080	0.256	-0.324	**	0.284	**	0.346	**
Log_30_OrgC	-0.106	0.132	-0.363	**	0.354	**	0.390	**
Log_30_ECEC	-0.131	0.061	-0.346	**	0.277	**	0.348	**
Log_30_ESP	0.039	0.576	0.126	0.073	-0.134	0.055	-0.007	0.915
S30_Ca_Mg	0.109	0.121	0.020	0.782	0.016	0.818	-0.156	*

\*\* : Significant correlations  $\leq 0.01$

\* : Significant correlations  $\leq 0.05$

a: An inverse transformation was performed on Sand and subsequently correlations that would normally be expected to be negative are positive and vice versa.

Variable n=204	Log_10pHw Correlation	Log_10pHw Probability	Log_10_pHca Correlation	Log_10_pHca Probability	Log_10_SAND% <sup>a</sup> Correlation	Log_10_SAND% <sup>a</sup> Probability	Log_10_SILT Correlation	Log_10_SILT Probability
Log_A_HORIZ								
RESTRICT_DPTH_CM								
S10_STONES_%								
Log_10_EC								
Log_10pHw	1.000							
Log_10_pHca	0.960	**	1.000					
Log_10_SAND% <sup>a</sup>	0.291	**	0.304	**	1.000		1.000	
Log_10_SILT	0.242	**	0.248	**	0.935	**	0.797	**
Log_10_CLAY%	0.287	**	0.306	**	0.958	**	0.553	**
S10_OrgC_%	0.170	*	0.252	**	0.502	**	0.782	**
Log_10_ECEC	0.561	**	0.610	**	0.833	**	-0.059	**
Log_10_ESP	-0.166	*	-0.208	**	-0.032	0.654	-0.300	0.399
S10_Ca_Mg	-0.092	0.191	-0.054	0.446	-0.397	**	0.137	**
Log_30_STONES	-0.074	0.290	-0.087	0.216	0.099	0.157	0.329	*
Log_30_EC	0.307	**	0.335	**	0.394	**	0.433	**
S30_pH_H2O	0.244	**	0.219	**	0.470	**	0.424	**
S30_pH_CaCl2	0.273	**	0.262	**	0.451	**	0.636	**
Log_30_SAND%	0.253	**	0.266	**	0.721	**	0.717	**
Log_30_Silt	0.231	**	0.242	**	0.681	**	0.569	**
Log_30_CLAY	0.238	**	0.251	**	0.675	**	0.545	**
Log_30_OrgC	0.271	**	0.301	**	0.603	**	0.701	**
Log_30_ECEC	0.283	**	0.285	**	0.779	**	-0.062	**
Log_30_ESP	-0.026	0.717	-0.035	0.621	-0.004	0.954	-0.290	0.376
S30_Ca_Mg	-0.154	*	-0.133	0.059	-0.361	**	1.000	**

\*\* : Significant correlations  $\leq 0.01$

\* : Significant correlations  $\leq 0.05$

a: An inverse transformation was performed on Sand and subsequently correlations that would normally be expected to be negative are positive and vice versa.

Variable n=204	Log_10_CLAY% Correlation	Log_10_CLAY% Probability	S10_OrgC_% Correlation	S10_OrgC_% Probability	Log_10_ECEC Correlation	Log_10_ECEC Probability	Log_10_ESP Correlation	Log_10_ESP Probability
Log_A_HORIZ								
RESTRICT_DPTH_CM								
S10_STONES_%								
Log_10_EC								
Log_10pHw								
Log_10_pHca								
Log_10_SAND% <sup>a</sup>								
Log_10_SILT								
Log_10_CLAY%	1.000							
S10_OrgC_%	0.422	**	1.000					
Log_10_ECEC	0.793	**	0.700	**	1.000			
Log_10_ESP	-0.005	0.943	-0.046	0.517	-0.133	0.057	1.000	
S10_Ca_Mg	-0.433	**	-0.109	0.120	-0.251	**	-0.389	**
Log_30_STONES	0.057	0.418	0.044	0.531	0.068	0.336	-0.004	0.950
Log_30_EC	0.413	**	0.279	**	0.430	**	0.059	0.405
S30_pH_H2O	0.454	**	0.330	**	0.484	**	0.181	**
S30_pH_CaCl2	0.431	**	0.308	**	0.482	**	0.124	0.077
Log_30_SAND%	0.725	**	0.370	**	0.603	**	-0.028	0.686
Log_30_Silt	0.577	**	0.332	**	0.580	**	-0.043	0.545
Log_30_CLAY	0.705	**	0.353	**	0.566	**	-0.028	0.687
Log_30_OrgC	0.594	**	0.437	**	0.632	**	-0.204	**
Log_30_ECEC	0.771	**	0.373	**	0.709	**	-0.060	0.396
Log_30_ESP	0.046	0.517	0.090	0.200	-0.044	0.528	0.525	**
S30_Ca_Mg	-0.388	**	-0.223	**	-0.278	**	-0.247	**

\*\* : Significant correlations  $\leq 0.01$

\* : Significant correlations  $\leq 0.05$

a : An inverse transformation was performed on Sand and subsequently correlations that would normally be expected to be negative are positive and vice versa.

Variable n=204	S10_Ca_Mg Correlation	S10_Ca_Mg Probability	Log_30_STONES Correlation	Log_30_STONES Probability	Log_30_EC Correlation	Log_30_EC Probability	S30_pH_H2O Correlation	S30_pH_H2O Probability
Log_A_HORIZ								
RESTRICT_DPTH_CM								
S10_STONES_%								
Log_10_EC								
Log_10pHw								
Log_10_pHca								
Log_10_SAND% <sup>a</sup>								
Log_10_SILT								
Log_10_CLAY%								
S10_OrgC_%								
Log_10_ECEC								
Log_10_ESP								
S10_Ca_Mg	1.000							
Log_30_STONES	0.037	0.597	1.000					
Log_30_EC	-0.244	**	0.091	0.197	1.000			
S30_pH_H2O	-0.406	**	-0.040	0.568	0.356	**	1.000	
S30_pH_CaCl2	-0.338	**	-0.006	0.932	0.451	**	0.956	**
Log_30_SAND%	-0.324	**	0.026	0.717	0.507	**	0.436	**
Log_30_Silt	-0.277	**	0.135	*	0.238	**	0.427	**
Log_30_CLAY	-0.300	**	0.013	0.855	0.511	**	0.380	**
Log_30_OrgC	-0.178	**	0.120	0.088	0.276	**	0.168	*
Log_30_ECEC	-0.403	**	0.035	0.620	0.510	**	0.558	**
Log_30_ESP	-0.340	**	-0.032	0.649	0.288	**	0.438	**
S30_Ca_Mg	0.612	**	0.065	0.353	-0.275	**	-0.478	**

\*\* : Significant correlations  $\leq 0.01$

\* : Significant correlations  $\leq 0.05$

a : An inverse transformation was performed on Sand and subsequently correlations that would normally be expected to be negative are positive and vice versa.



Variable n=204	S30_pH_CaCl2 Correlation	S30_pH_CaCl2 Probability	Log_30_SAND% Correlation	Log_30_SAND% Probability	Log_30_Silt Correlation	Log_30_Silt Probability	Log_30_CLAY Correlation	Log_30_CLAY Probability
Log_A_HORIZ								
RESTRICT_DPTH_CM								
S10_STONES_%								
Log_10_EC								
Log_10pHw								
Log_10_pHca								
Log_10_SAND% <sup>a</sup>								
Log_10_SILT								
Log_10_CLAY%								
S10_OrgC_%								
Log_10_ECEC								
Log_10_ESP								
S10_Ca_Mg								
Log_30_STONES								
Log_30_EC								
S30_pH_H2O								
S30_pH_CaCl2	1.000							
Log_30_SAND%	0.405	**	1.000					
Log_30_Silt	0.396	**	0.613	**	1.000			
Log_30_CLAY	0.352	**	0.979	**	0.471	**	1.000	
Log_30_OrgC	0.171	*	0.600	**	0.448	**	0.594	**
Log_30_ECEC	0.539	**	0.888	**	0.629	**	0.852	**
Log_30_ESP	0.322	**	0.108	0.124	0.016	0.823	0.097	0.166
S30_Ca_Mg	-0.389	**	-0.500	**	-0.263	**	-0.489	**

\*\* : Significant correlations  $\leq 0.01$

\* : Significant correlations  $\leq 0.05$

a : An inverse transformation was performed on Sand and subsequently correlations that would normally be expected to be negative are positive and vice versa.

Variable n=204	Log_30_OrgC Correlation	Log_30_OrgC Probability	Log_30_ECEC Correlation	Log_30_ECEC Probability	Log_30_ESP Correlation	Log_30_ESP Probability	S30_Ca_Mg Correlation	S30_Ca_Mg Probability
Log_A_HORIZ								
RESTRICT_DPTH_CM								
S10_STONES_%								
Log_10_EC								
Log_10pHw								
Log_10_pHca								
Log_10_SAND% <sup>a</sup>								
Log_10_SILT								
Log_10_CLAY%								
S10_OrgC_%								
Log_10_ECEC								
Log_10_ESP								
S10_Ca_Mg								
Log_30_STONES								
Log_30_EC								
S30_pH_H2O								
S30_pH_CaCl2								
Log_30_SAND%								
Log_30_Silt								
Log_30_CLAY								
Log_30_OrgC	1.000							
Log_30_ECEC	0.680	**	1.000					
Log_30_ESP	-0.293	**	-0.009	0.900	1.000			
S30_Ca_Mg	-0.159	*	-0.447	**	-0.522	**	1.000	

\*\* : Significant correlations  $\leq 0.01$

\* : Significant correlations  $\leq 0.05$

a : An inverse transformation was performed on Sand and subsequently correlations that would normally be expected to be negative are positive and vice versa.

**APPENDIX N****REGIONALIZED VARIABLE THEORY**

The “regionalized variable theory” assumes that the spatial variation of any variable can be expressed as the sum of three major components (Burrough, 1996)

1. a structural component, associated with a constant mean value or constant trend;
2. a random, spatially correlated component
3. a random noise or residual error

If  $x$  is a position in 1, 2, or 3 dimensions, then the value of a variable  $Z$  at  $x$  is given by;

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon'' \quad (\text{A.4})$$

where  $m(x)$  is a deterministic function describing the ‘structural’ component of  $Z$  at  $x$ ,  $\varepsilon'(x)$  is the term denoting the locally varying, spatially dependent residuals from  $m(x)$ , and  $\varepsilon''$  is a residual, spatially independent noise (Burrough, 1996).

The variogram describes the spatial variation of a random variable, and estimating an experimental variogram is the first step in kriging interpolation techniques. A suitable function for  $m(x)$  is required. When no trend or drift are present,  $m(x)$  equals the mean value in the sampling area and the average or expected difference between any two places  $x$  and  $x + h$  separated by a vector  $h$ , will be zero (Burrough, 1996). The variability in space of the data is characterised by the semivariance, which is a measure of the deviation between pairs of  $z$  values at a certain distance and direction. Stationarity of difference and variance of differences, define the requirements for the intrinsic hypothesis of regionalized variable theory. *This means that once the structural effects have been accounted for, the remaining variation is homogeneous in its variation so the differences between sites is merely a function of the distance between them* (Burrough, 1996). If these conditions are met the semivariance can be estimated from sample data as follows:

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n \{Z(x_i) - Z(x_i + h)\}^2 \quad (\text{A.5})$$

where  $n$  is the number of pairs of sample points separated by distance  $h$  (referred to as the lag).

The function  $\gamma(h)$  is known as the semivariogram. The ‘semi’ refers to the fact that it is half of a variance; it is half the variance of a difference in this instance. It is the

variance per point when the points are considered in pairs. The semivariogram is now usually termed the variogram (Webster and Oliver, 2001b).

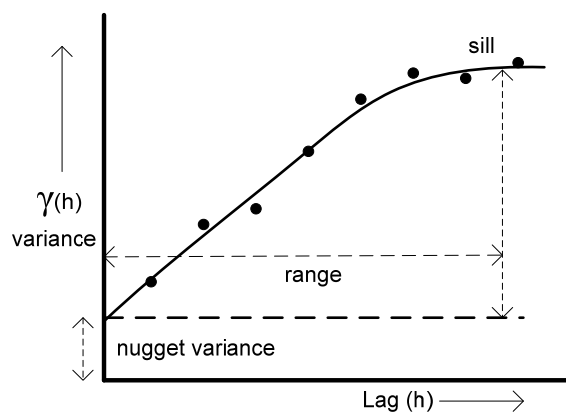


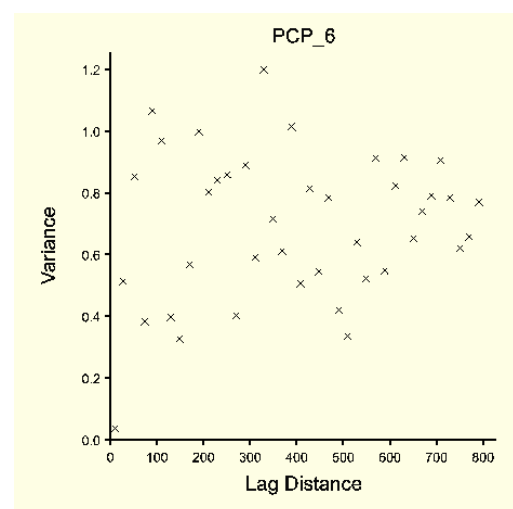
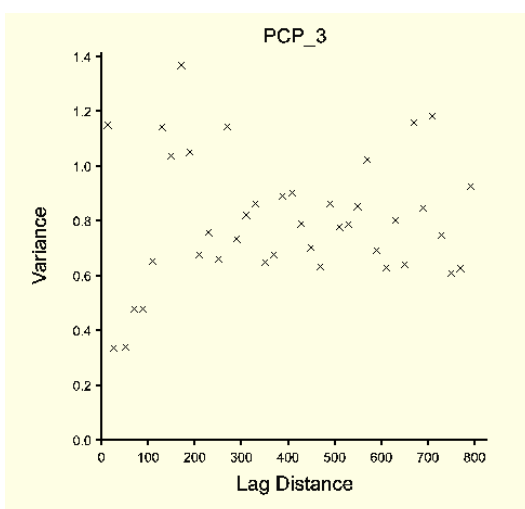
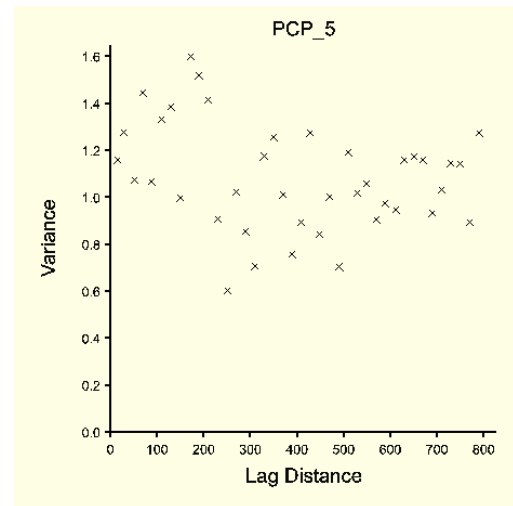
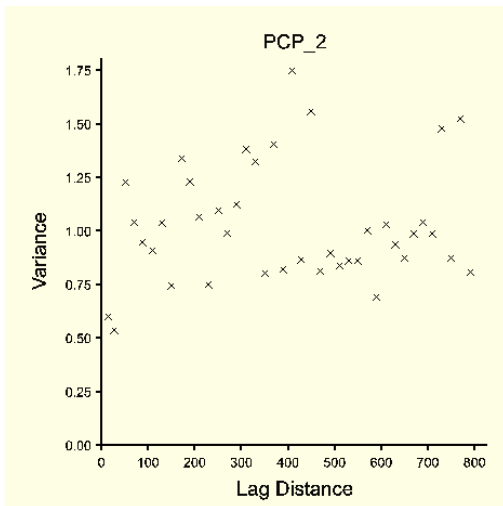
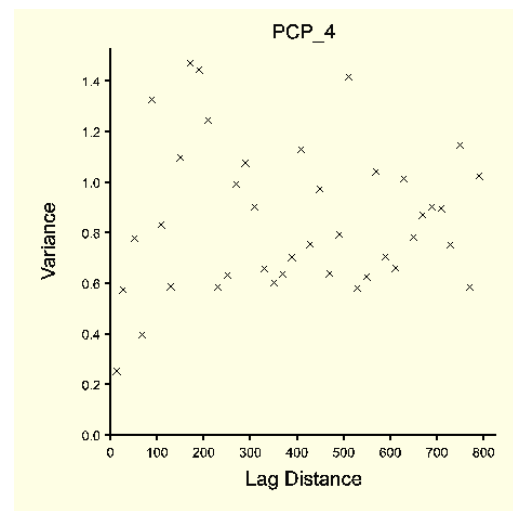
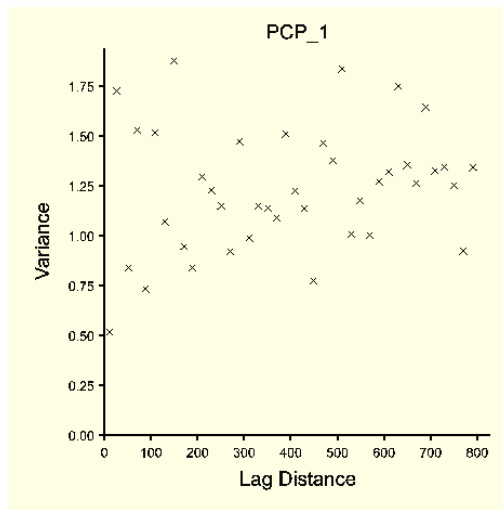
Figure A.3 General characteristics of a variogram

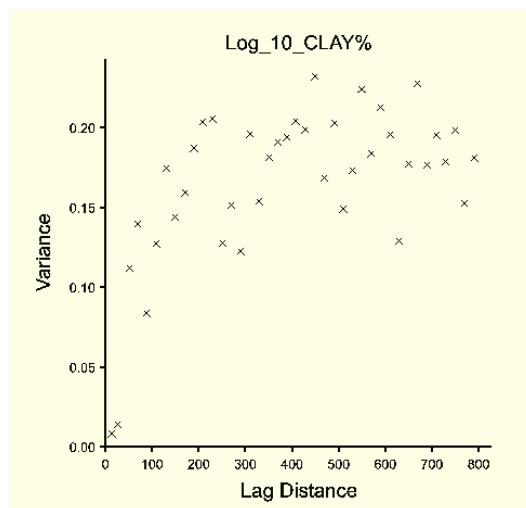
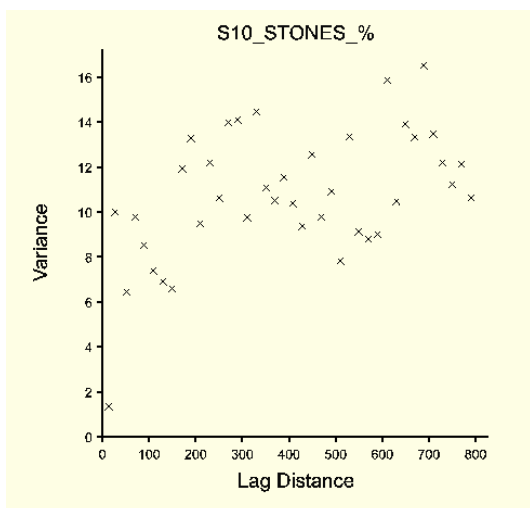
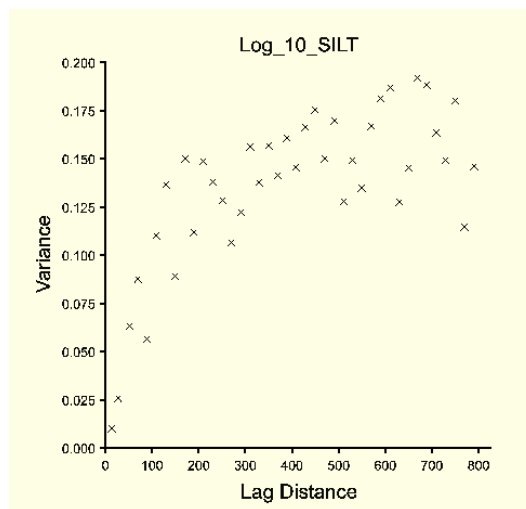
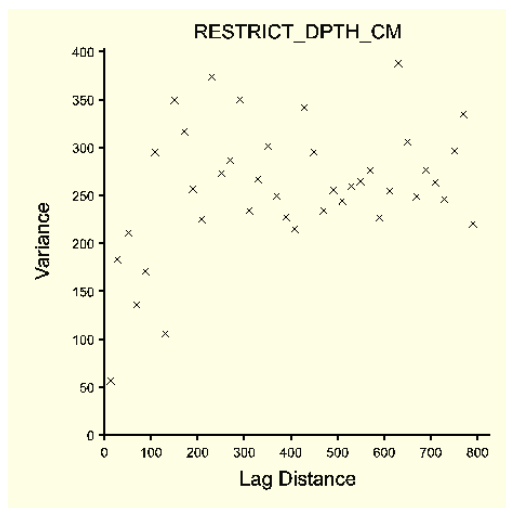
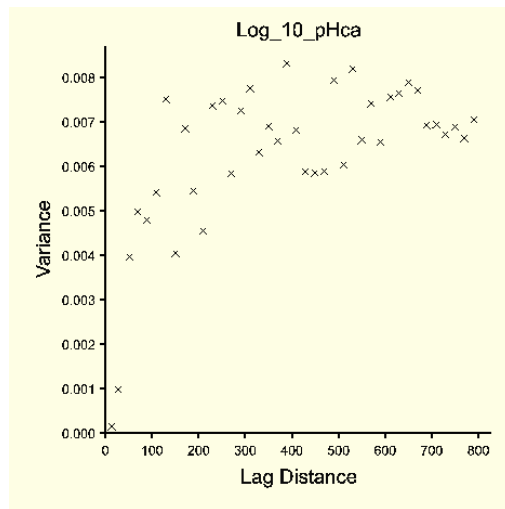
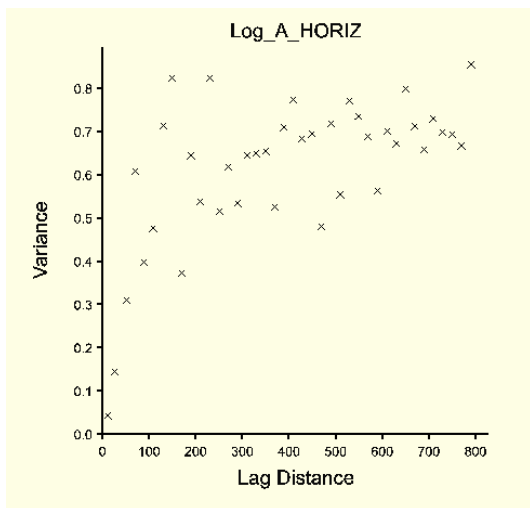
An experimental variogram is computed from the data and the best fit model selected. Commonly used variogram models include a) spherical, b) exponential, c) bounded linear and d) Gaussian. Variograms are used in interpolation methods such as kriging, co-kriging and other kriging derivatives.

**APPENDIX O**

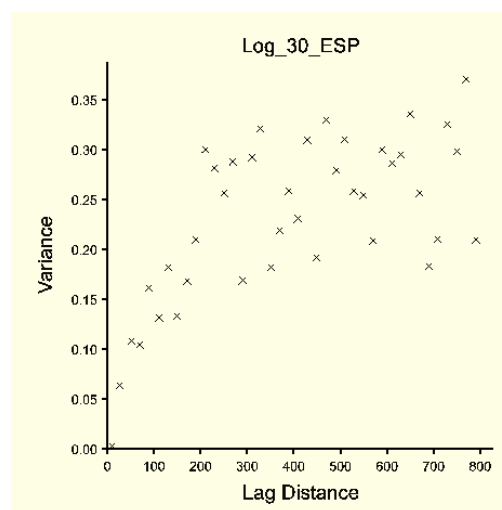
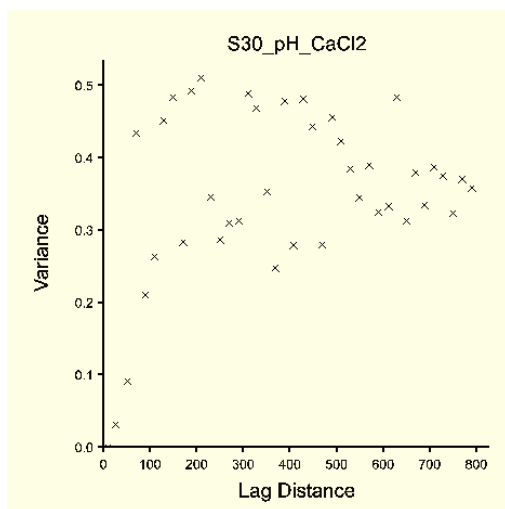
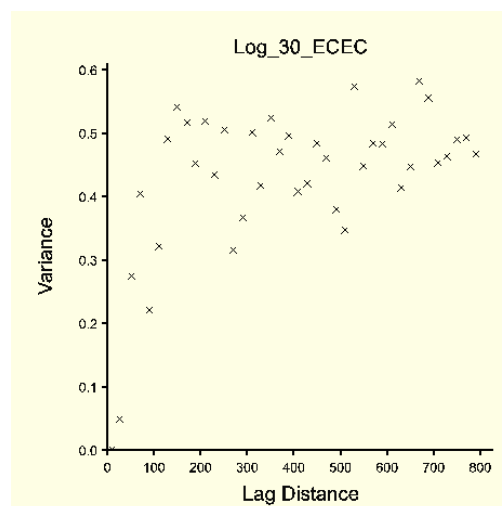
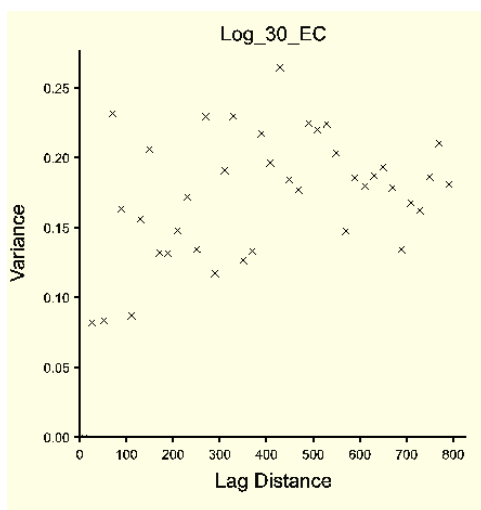
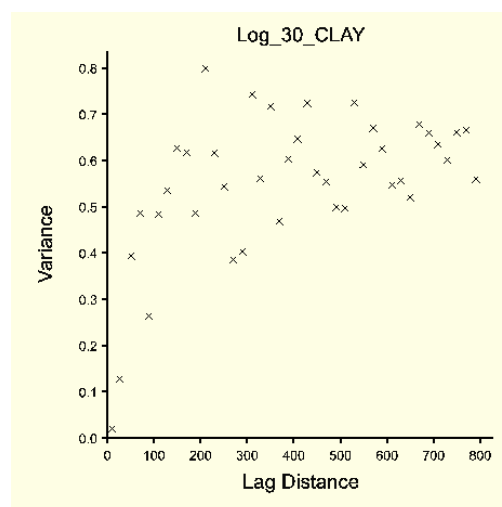
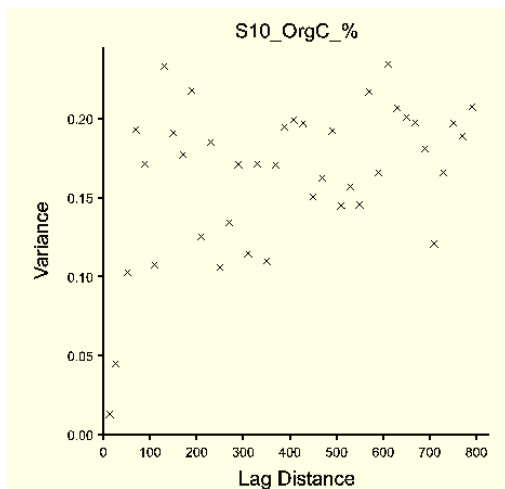
**EXPERIMENTAL VARIOGRAMS**

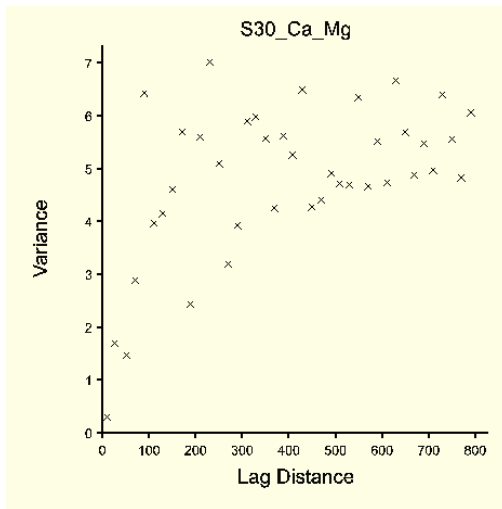
## Experimental Variograms with outliers removed via Variogram Clouds





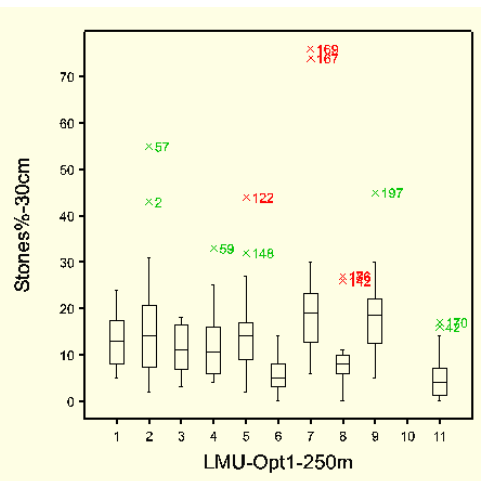
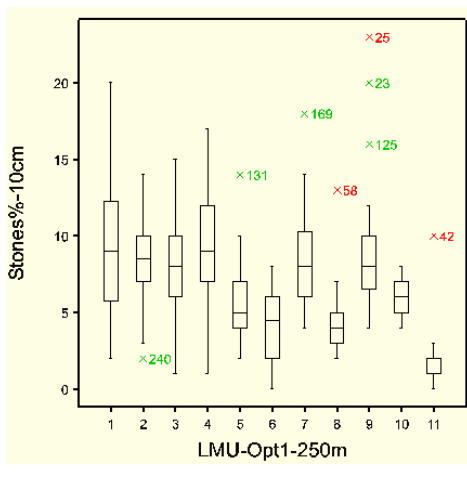
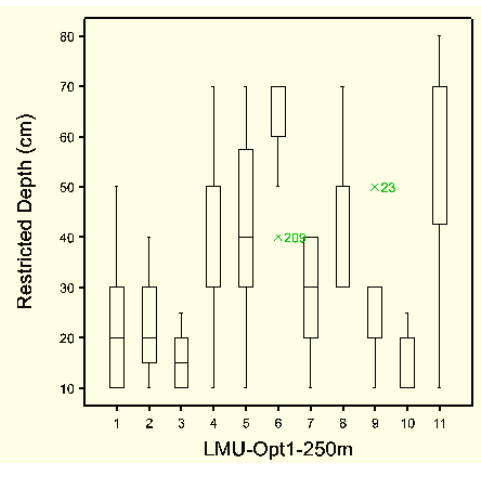
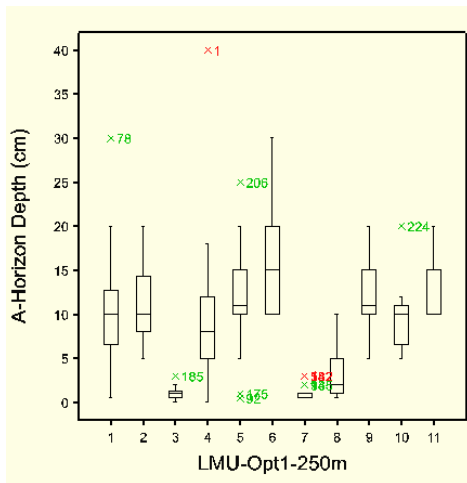
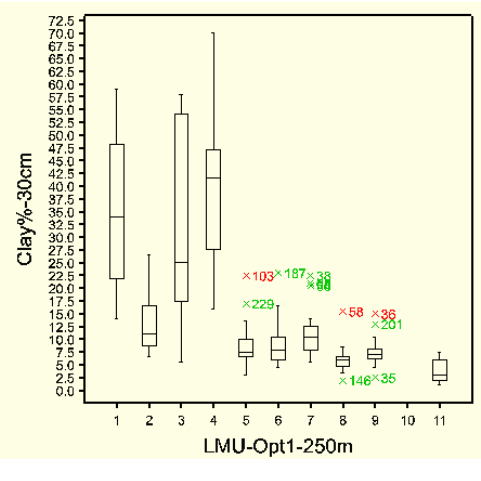
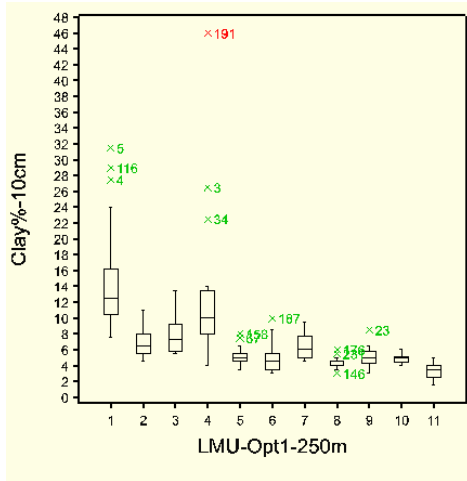


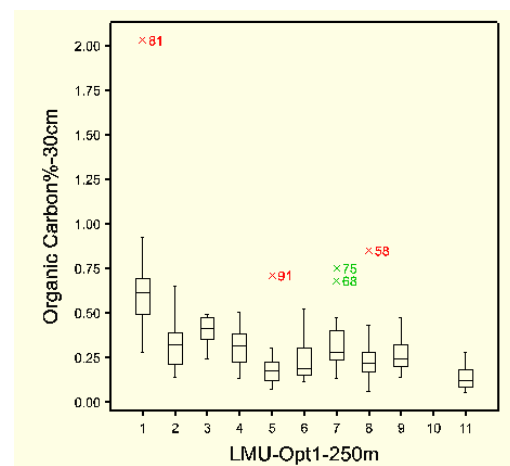
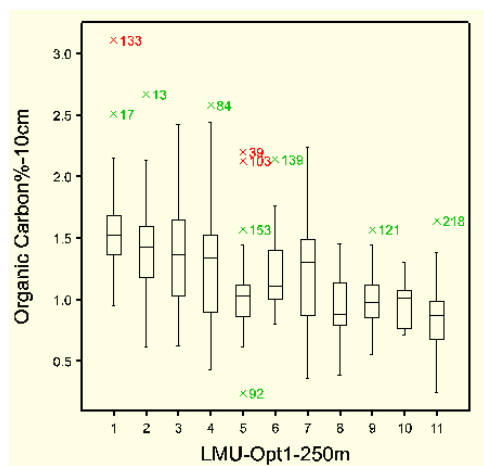
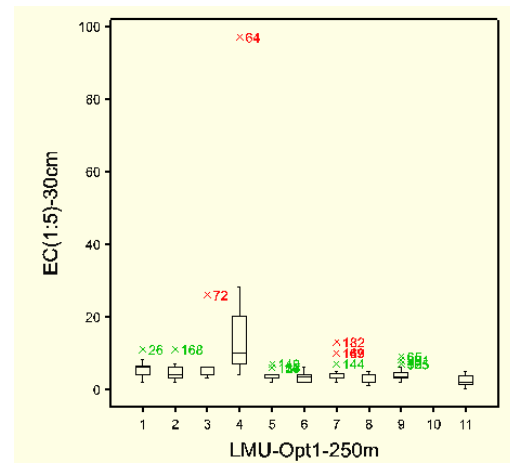
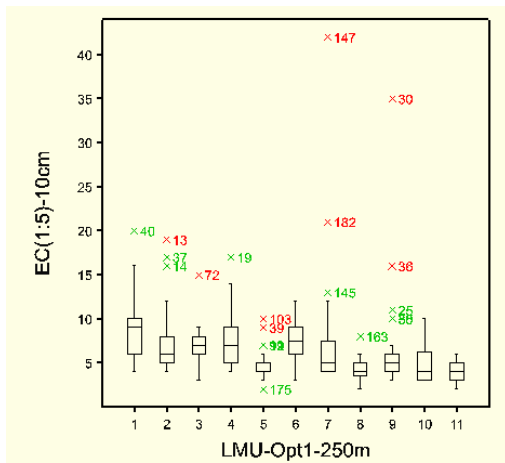
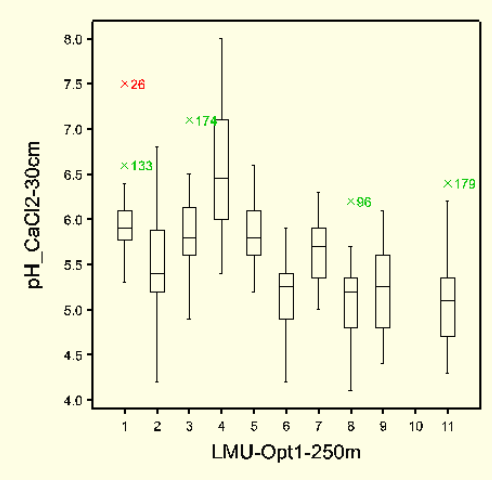
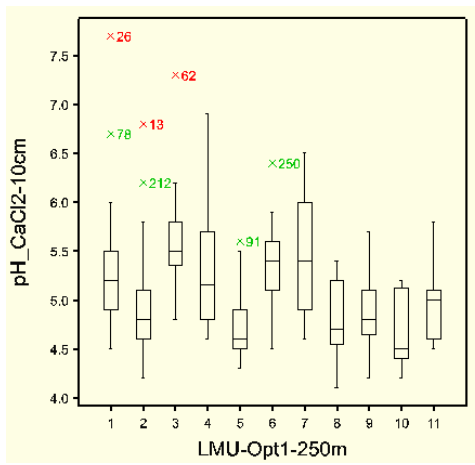


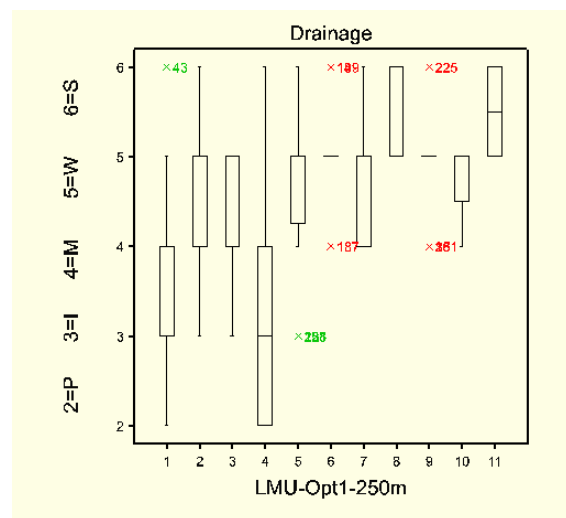
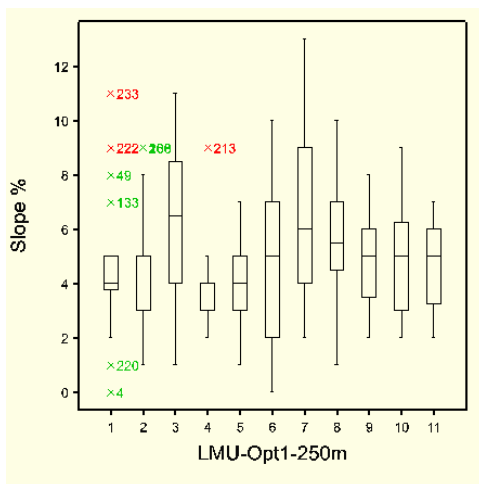
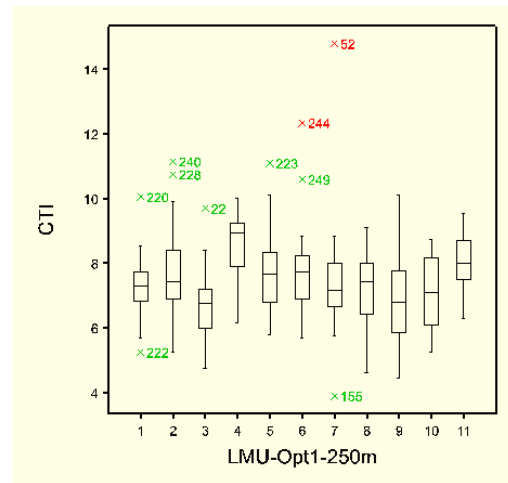
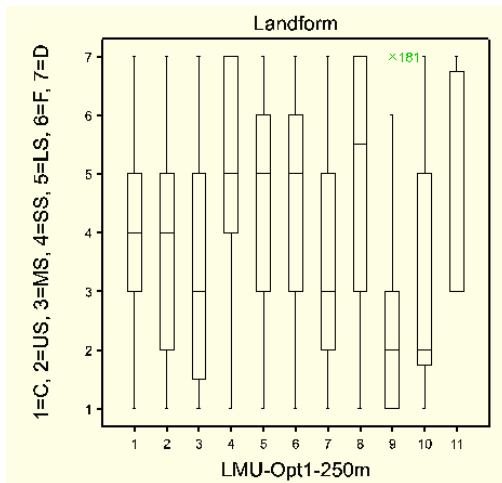
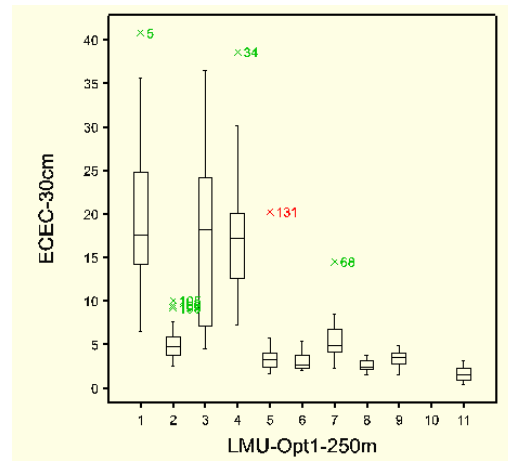
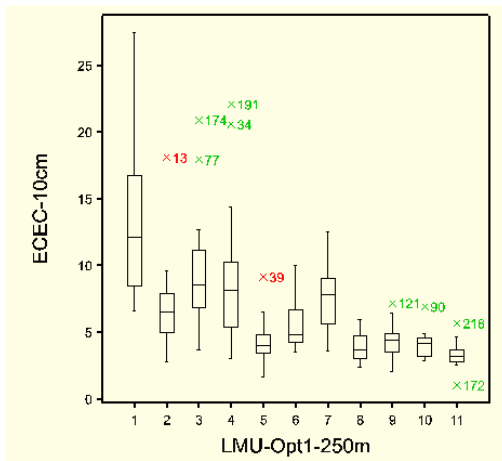


## **APPENDIX P**

### **BOX-PLOTS OF SOIL PROPERTIES AND ANCILLARY VARIABLES**







**RESULTS FROM KAPPA MAP COMPARISONS**

Method	Kappa											
Map1	Cluster250m											
Map2	Cluster200m											
Kappa	0.55											
KLocation	0.69											
KHisto	0.81											
Fraction correct	0.60											
	LMU 1	LMU 2	LMU 3	LMU 4	LMU 5	LMU 6	LMU 7	LMU 8	LMU 9	LMU 10	LMU 11	
Kappa	0.87	0.76	0.82	0.82	0.43	-0.01	0.82	0.60	0.84	0.76	-0.08	
KLoc	0.92	0.92	0.94	0.87	0.62	-0.02	0.83	0.81	0.85	0.86	-0.12	
KHisto	0.95	0.83	0.87	0.94	0.71	0.45	0.99	0.74	0.98	0.88	0.69	
Map 1 \ Map 2	LMU 1	LMU 2	LMU 3	LMU 4	LMU 5	LMU 6	LMU 7	LMU 8	LMU 9	LMU 10	LMU 11	Sum Map 1
LMU 1	8902	18	10	118	11	36	114	915	128	225	116	10593
LMU 2	99	9324	15	246	2355	130	97	819	145	143	235	13608
LMU 3	54	3	4740	92	0	58	772	291	180	120	94	6404
LMU 4	14	59	0	8265	12	60	55	643	119	52	52	9331
LMU 5	125	193	6	649	5513	345	604	1185	246	244	5334	14444
LMU 6	5	35	8	48	8	959	28	230	433	55	4663	6472
LMU 7	8	4	75	478	28	145	10808	620	289	286	20	12761
LMU 8	149	64	27	67	32	212	61	6008	392	58	185	7255
LMU 9	107	70	43	189	42	137	252	465	11647	291	132	13375
LMU 10	8	4	0	112	19	28	6	304	60	3690	10	4241
LMU 11	121	303	89	166	306	17767	115	351	109	168	850	20345
Sum Map 2	9592	10077	5013	10430	8326	19877	12912	11831	13748	5332	11691	950478



Method	Kappa											
Map1	Cluster250m											
Map2	Cluster300m											
Kappa	0.55											
KLocation	0.68											
KHisto	0.82											
Fraction correct	0.60											
	LMU 1	LMU 2	LMU 3	LMU 4	LMU 5	LMU 6	LMU 7	LMU 8	LMU 9	LMU 10	LMU 11	
Kappa	0.83	0.81	0.49	0.79	0.71	-0.01	0.33	0.08	0.83	0.82	0.13	
KLoc	0.92	0.88	0.84	0.81	0.76	-0.06	0.36	0.11	0.91	0.83	0.22	
KHisto	0.90	0.91	0.59	0.98	0.94	0.13	0.92	0.73	0.92	1.00	0.61	
Map 1 \ Map 2	LMU 1	LMU 2	LMU 3	LMU 4	LMU 5	LMU 6	LMU 7	LMU 8	LMU 9	LMU 10	LMU 11	Sum Map 1
LMU 1	9868	172	23	65	14	100	64	134	133	15	5	10593
LMU 2	267	12228	23	19	362	0	265	108	123	29	184	13608
LMU 3	295	40	5507	27	59	142	40	46	121	18	109	6404
LMU 4	186	429	92	7691	525	0	87	125	104	69	23	9331
LMU 5	79	88	1	105	10242	0	1774	161	145	29	1820	14444
LMU 6	70	335	4	0	162	0	14	1834	457	12	3584	6472
LMU 7	394	75	6810	83	325	5	4738	49	224	35	23	12761
LMU 8	459	387	492	404	232	219	2202	1419	687	77	677	7255
LMU 9	195	80	116	51	198	0	67	297	12278	81	12	13375
LMU 10	95	14	11	0	44	0	258	56	233	3514	16	4241
LMU 11	845	1980	1306	1191	834	0	1626	7769	948	345	3501	20345
Sum Map 2	12753	15828	14385	9636	12997	466	11135	11998	15453	4224	9954	950478

Method	Kappa											
Map1	Cluster250m											
Map2	Discrim200m											
Kappa	0.49											
KLocation	0.58											
KHisto	0.85											
Fraction correct	0.54											
	LMU 1	LMU 2	LMU 3	LMU 4	LMU 5	LMU 6	LMU 7	LMU 8	LMU 9	LMU 10	LMU 11	
Kappa	0.79	0.66	0.65	0.65	0.34	0.01	0.72	0.56	0.76	0.69	-0.07	
KLoc	0.80	0.75	0.69	0.82	0.39	0.02	0.73	0.62	0.80	0.72	-0.10	
KHisto	0.99	0.88	0.95	0.78	0.86	0.59	0.98	0.89	0.95	0.96	0.67	
Map 1 \ Map 2	LMU 1	LMU 2	LMU 3	LMU 4	LMU 5	LMU 6	LMU 7	LMU 8	LMU 9	LMU 10	LMU 11	Sum Map 1
LMU 1	8668	289	88	385	124	22	99	410	282	148	78	10593
LMU 2	253	8490	98	1118	2196	275	144	378	282	165	209	13608
LMU 3	191	31	4107	97	30	34	1010	236	515	125	28	6404
LMU 4	77	173	123	7872	426	7	97	339	118	61	38	9331
LMU 5	265	343	78	1398	5264	223	313	785	318	227	5230	14444
LMU 6	35	77	30	118	14	873	57	123	562	52	4531	6472
LMU 7	267	150	630	649	371	50	9381	530	485	243	5	12761
LMU 8	238	425	102	643	62	150	172	4723	493	142	105	7255
LMU 9	276	201	60	505	121	74	271	585	11014	213	55	13375
LMU 10	11	15	168	239	122	73	72	104	338	3097	2	4241
LMU 11	579	683	323	841	2566	12593	739	624	270	124	1003	20345
Sum Map 2	10860	10877	5807	13865	11296	14374	12355	8837	14677	4597	11284	950478

Method	Kappa											
Map1	Cluster250m											
Map2	Discrim250m											
Kappa	0.74											
KLocation	0.81											
KHisto	0.91											
Fraction correct	0.76											
	LMU 1	LMU 2	LMU 3	LMU 4	LMU 5	LMU 6	LMU 7	LMU 8	LMU 9	LMU 10	LMU 11	
Kappa	0.80	0.77	0.75	0.70	0.76	0.76	0.76	0.57	0.79	0.76	0.65	
KLoc	0.86	0.82	0.76	0.86	0.82	0.76	0.76	0.67	0.82	0.76	0.87	
KHisto	0.93	0.94	0.98	0.81	0.94	1.00	0.99	0.85	0.96	0.99	0.74	
Map 1 \ Map 2	LMU 1	LMU 2	LMU 3	LMU 4	LMU 5	LMU 6	LMU 7	LMU 8	LMU 9	LMU 10	LMU 11	Sum Map 1
LMU 1	9255	238	34	264	179	5	79	62	255	95	127	10593
LMU 2	214	11494	93	798	219	45	134	107	240	31	233	13608
LMU 3	180	82	4795	81	6	69	477	52	498	105	59	6404
LMU 4	101	216	170	8142	367	9	67	51	128	56	24	9331
LMU 5	99	323	47	1079	12152	141	63	85	148	76	231	14444
LMU 6	29	87	148	110	587	4987	14	58	265	33	154	6472
LMU 7	349	357	410	623	256	173	9904	251	276	160	2	12761
LMU 8	460	518	58	675	365	32	191	3764	527	126	539	7255
LMU 9	263	451	25	513	255	99	150	227	11315	50	27	13375
LMU 10	39	71	98	157	74	1	144	31	373	3216	37	4241
LMU 11	1067	1310	325	746	1692	937	1352	761	386	222	11547	20345
Sum Map 2	12056	15147	6203	13188	16152	6498	12575	5449	14411	4170	12980	950478

Method	Kappa											
Map1	Cluster250m											
Map2	Discrim300m											
Kappa	0.49											
KLocation	0.60											
KHisto	0.82											
Fraction correct	0.54											
	LMU 1	LMU 2	LMU 3	LMU 4	LMU 5	LMU 6	LMU 7	LMU 8	LMU 9	LMU 10	LMU 11	
Kappa	0.74	0.72	0.47	0.67	0.62	-0.02	0.29	0.06	0.75	0.72	0.14	
KLoc	0.82	0.77	0.71	0.79	0.67	-0.04	0.31	0.09	0.81	0.80	0.24	
KHisto	0.90	0.93	0.66	0.85	0.93	0.52	0.92	0.71	0.93	0.90	0.57	
Map 1 \ Map 2	LMU 1	LMU 2	LMU 3	LMU 4	LMU 5	LMU 6	LMU 7	LMU 8	LMU 9	LMU 10	LMU 11	Sum Map 1
LMU 1	8873	422	132	164	195	129	92	125	350	93	18	10593
LMU 2	390	10926	197	714	504	85	181	85	296	53	177	13608
LMU 3	377	172	4759	62	52	303	50	44	509	40	36	6404
LMU 4	144	445	352	7560	492	26	62	20	185	40	5	9331
LMU 5	166	395	85	1190	8981	55	1618	138	183	60	1573	14444
LMU 6	49	262	52	85	221	36	20	1425	635	17	3670	6472
LMU 7	573	248	5582	578	373	576	4273	96	339	114	9	12761
LMU 8	519	452	321	533	319	801	2148	1340	541	142	139	7255
LMU 9	303	383	55	463	358	135	287	157	11170	51	13	13375
LMU 10	73	155	117	68	81	117	286	72	448	2819	5	4241
LMU 11	1216	1427	632	918	1127	73	2075	8969	457	81	3370	20345
Sum Map 2	12683	15287	12284	12335	12703	2336	11092	12471	15113	3510	9015	950478

**APPENDICES' REFERENCES**

- Burrough, P.A. (1996) Principles of geographical information systems for land resources assessment, Clarendon Press, Oxford, UK. 194 pp.
- Caccetta, P.A. (2000) Technical note-A simple approach for reducing discontinuities in digital elevation models (dems), Report Number 2000/231, CSIRO Mathematical and Information Sciences, Wembley, Perth, Western Australia, 6 pp.
- FAO (1990) Guidelines for Soil Description,(3rd Revised edn), Food and Agriculture Organization of the United Nations, Rome, Italy. 70 pp.
- Gallant, J.C. and Wilson, J.P. (2000) Primary Topographic Attributes, in: Terrain Analysis: Principals and Applications, Wilson, P. W. & Gallant, J. C. (eds.), Wiley, New York, USA, pp. 51-85.
- Keigan Systems Inc. (1999) MFworks (Version 2.6), 633 Colborne St. Suite 250, London, Ontario, Canada, (program).
- Keigan Systems Inc. (2003) 'GeoMedia Grid Help Manual', 633 Colborne St. Suite 250, London, Ontario, N6B 2V3, Canada.
- Malczewski, J. (1999) Criterion Weighting, in: GIS and Multicriteria Decision Analysis, John Wiley & Sons, New York, USA, pp. 177-195.
- McDonald, R.C. and Isbell, R.F. (1990) Soil Profile, in: Australian Soil and Land Survey-Field Handbook, (2nd edn), McDonald, R. C., Isbell, R. F., Speight, J. R., Walker, J. & Hopkins, M. S. (eds.), Inkata Press, Melbourne, Australia, pp. 103-152.
- Moore, G. (1998) Distinctive Morphological Features and their Agricultural Significance, in: Soilguide: A handbook for understanding and managing agricultural soils, Moore, G. (ed.), Agriculture Western Australia Bulletin No. 4343, Department of Agriculture, Perth, Australia, pp. 43-50.
- Webster, R. and Oliver, M.A. (2001) Geostatistics for Environmental Scientists, John Wiley and Sons, Chischester, England. 271 pp.