

Inferring the Spatial Distribution of Regolith Properties Using Surface Measurable Features.

Shawn William Laffan

A thesis submitted for the degree of Doctor of Philosophy
of The Australian National University
December 2001.

Except where otherwise indicated, this is my own work.

Acknowledgements.

It has been a long gestation and there are many people to thank.

First and foremost, I owe a great deal to Brian Lees for his constant support in all matters academic over an extended period of time.

To the other members of my supervisory panel. Tony Eggleton for his support on fieldwork and draft reading and Bob Wasson for thought provocation.

COMALCO at Weipa provided much of the base data used in this thesis, an invaluable dataset to be able to use. Added to that were some most useful discussions with the mine geologists, Mike Morgan and Adam Crossing. The Cooperative Research Centre for Landscape Evolution and Mineral Exploration gave financial and fieldwork support.

To the numerous people in what began as the Geography Department and finished as the School of Resources, Environment and Society. Cartographically, to Clive Hilliker, Val Lyon and Kevin Cowan. Clive drafted many of the line drawings in the thesis and gave most valuable advice on diagram construction. Computationally, there was constant assistance, code creation and debugging advice from Steve Leahy, Karl Nissen, David Houlder and Tim Wilson. I am sure that, deep down, Steve and Karl really liked the Fortran code I sometimes asked them to translate. Administratively, Helena Wolf and Nikki Sekavs provided a grounding force, particularly when they wouldn't just give me money without a good reason.

To those many people who have shared the PhD journey with me. Luke Foster understands what Big Snowflake means, shared much of the burden of fieldwork, provided a fount of knowledge geochemical, and paid the rent on time. Also to Diane Pearson and Simon Bengner for not rubbing it in when they finished first. More recently to Graham Hammond, Kimberly Van Niel, Bruce Doran and Zhi Huang for northern experiences.

In the past beer has been the primary thought lubricant, but as one furthers one's research career one is drawn increasingly to the benefits of the grape. So, to red wine for its nourishing benefits (but also good whiskey, whisky and brandy).

Also, to the soldiers of Exercise Kangaroo 95 for not mistaking me for a fictitious invading Orange person and arresting me. Field work would have been difficult under such circumstances.

I am also indebted to my sister, who didn't make me give her car back when I was close to completion.

Importantly, to Phillippa, who has been putting up with this project almost as long as she's been putting up with me, for which I am grateful.

And finally, to William Gerard Tennyson Laffan. One tough old bugger.

Abstract.

The aim of this research is to determine to what extent properties of the regolith may be inferred using only features easily measured from the surface. To address this research question, a set of regolith properties from Weipa, Queensland, Australia, are analysed. The set contains five variables, oxides of Aluminium, Iron, Silica and Titanium, as well as Depth to Ironstone. This last represents the depth of the layer from which the oxides are sampled.

The research question is addressed in two ways. First, locations where the properties are related to modern surface hydrology are assessed using spatially explicit analyses. This is done by comparing the results of spatial association statistics using geometric and watershed-based spatial samples. Second, correlations are sought for between the regolith properties and geomorphometric indices of land surface morphology and Landsat Thematic Mapper spectral response. This is done using spatially implicit Artificial Neural Networks (ANN) and spatially explicit Geographically Weighted Regression (GWR). The results indicate that the degree to which regolith properties are related to surface measurable features is limited and spatially variable.

Most locations in the Weipa landscape exhibit some degree of modern hydrological control of the oxide variables at lateral distances of 120 m. This control rarely extends beyond 300 m laterally, although such locations occupy distinct positions in the landscape. Conversely, there is an extensive part of the landscape where Depth to Ironstone is under hydrological control. This occupies most of the lower elevations in the study area. Depth to Ironstone represents the depth to the redox front where iron is precipitated, but may in some parts of the landscape control the distribution of the watertable by being impermeable.

For the correlation analyses, the highest correlations are found with those oxides most mobile in solution. The spatially local GWR results also consistently outperform the spatially global ANN results, commonly having accuracies 40% higher at the error tolerance used. Much of this can be attributed to the localised effects of landscape evolution. Comparison of the GWR results against the local sample mean indicate that there is a relationship between regolith properties and

surface measurable features at 10-15% of sample locations for the oxide variables, and 22% for Depth to Ironstone.

The implications of these results are significant for anyone intending to generate spatial datasets of regolith properties. If there is a low spatial density of sample data, then the effects of landscape evolution can reduce the utility of any analysis results. Instead, spatially dense, direct measurements of subsurface regolith properties are needed. While these may not be a direct measurement of the property of interest, they may provide useful additional information by which these may be inferred.

Table of Contents.

Acknowledgements.	iii
Abstract.	v
Table of Contents.	vii
List of Figures.	xiii
List of Tables.	xvii
Chapter 1. Introduction	1
1.1. A Research Question.	1
1.1.1. The Regolith.	1
1.1.2. Why Regolith Properties and Surface Measurable Features?	1
1.2. Addressing the Research Question.	2
1.2.1. Previous Work.	3
1.2.2. Equilibrium.	3
1.2.2.1. A System of Cause and Effect.	4
1.2.2.2. Vegetation, Topography and Time.	6
1.2.2.3. Implications for This Research.	7
1.2.3. The Data Model.	7
1.2.3.1. The Dataset as a Communication Device.	8
1.2.3.2. The Dataset as a Data Storage Medium.	10
1.2.4. The Approach.	10
1.2.5. Data Requirements and Field Location.	11
1.3. Study Area.	12
1.4. Thesis Structure	13
Chapter 2. Literature Review.	15
2.1. Introduction.	15
2.2. Geomorphometry.	15
2.2.1. Before GIS.	16
2.2.1.1. Before the "Quantitative Revolution".	16
2.2.1.2. The "Quantitative Revolution".	17
2.2.2. The GIS Era.	19
2.2.2.1. Data Structures.	19
2.2.2.1.1. Finite Elements.	20
2.2.2.1.2. Triangulated Irregular Networks.	20
2.2.2.1.3. Raster Grids.	21
2.2.2.2. General Geomorphometry.	21
2.2.2.2.1. Calculation.	21
2.2.2.2.2. Alternate Indices.	23
2.2.2.3. Hydrology.	24
2.2.2.3.1. Flow Accumulation.	24
2.2.2.3.1.1. Convergent Algorithms.	24
2.2.2.3.1.2. Divergent Algorithms.	25

	2.2.2.3.1.3	Stream Tubes.	25
	2.2.2.3.1.4	Limitations of Flow Models.	26
2.2.2.4.		Analysis Resolution and Scale.	27
	2.2.2.4.1	Resolution.	27
	2.2.2.4.2	Multi-scale Indices.	28
	2.2.2.4.2.1	Larger Neighbourhoods.	28
	2.2.2.4.2.2	Generalising the Landscape.	30
	2.2.2.4.2.3	Hydrology and the Hillslope Scale.	30
	2.2.3.	Geomorphometry Summary.	31
2.3.		Mapping the Regolith.	31
	2.3.1.	Qualitative Mapping.	33
	2.3.1.1.	Classification Schemes.	34
	2.3.1.2.	Soil-Landscape Mapping.	36
		2.3.1.2.1 The Catena.	36
		2.3.1.2.2 Soilscales.	37
	2.3.1.3.	Regolith Mineral Mapping.	37
	2.3.1.4.	Limitations of Qualitative Mapping.	38
	2.3.2.	Quantitative Mapping.	39
	2.3.2.1.	Implicitly Spatial Methods.	40
		2.3.2.1.1 Predefined Classes.	41
		2.3.2.1.2 Soil Properties.	42
	2.3.2.2.	Explicitly Spatial Methods.	46
		2.3.2.2.1 Geostatistics.	47
		2.3.2.2.2 Incorporating Ancillary Data in Geostatistics.	49
		2.3.2.2.3 Artificial Neural Networks.	51
	2.3.3.	Mapping Using Remotely Sensed Data.	52
	2.3.3.1.	Pixel Unmixing.	53
	2.3.3.2.	Geobotanical Remote Sensing.	54
	2.3.4.	Regolith Mapping Summary.	56
2.4.		Implications for the Research.	58
Chapter 3.	Field Area Description.	61	
3.1.	Introduction.	61	
3.2.	The Weipa Bauxite.	61	
	3.2.1. Climate and Hydrology.	62	
	3.2.2. Geology and Geomorphology.	66	
	3.2.2.1. Regional.	66	
	3.2.2.2. Parent Material.	66	
	3.2.2.3. Local Geomorphic Units.	67	
	3.2.2.4. The Bauxite.	68	
	3.2.2.4.1 Spatial Variation.	68	
	3.2.2.4.1.1 Vertical Variation.	68	
	3.2.2.4.1.2 Lateral Variation.	71	
	3.2.2.4.2 Mechanical Reworking.	73	
	3.2.2.4.3 Melon Holes.	75	
	3.2.2.4.4 Age of the Regolith at Weipa.	75	

3.2.3. Vegetation.	77
3.2.4. Mining Operations.	80
3.3. Summary.	80
Chapter 4. The Dataset.	81
4.1. Introduction.	81
4.2. Dataset Documentation.	81
4.2.1. Drill Core Data.	81
4.2.1.1. Limitations and Known Errors.	83
4.2.2. Digital Elevation Model.	86
4.2.2.1. Limitations and Known Errors.	86
4.2.3. Remotely Sensed Data.	87
4.2.4. Other Data.	89
4.3. The Training Set.	89
4.3.1. Quantity and Distribution of Training Data.	89
4.3.2. Drill Core Data Preparation.	89
4.3.3. Independent Variables.	90
4.3.3.1. Remotely Sensed Information.	90
4.3.3.2. Geomorphometry.	91
4.3.3.2.1. General Geomorphometric Indices.	91
4.3.3.2.2. Hydrological Indices.	94
4.4. Chapter Summary.	96
Chapter 5. Review of Methods.	97
5.1. Introduction.	97
5.1.1. The Data Type.	97
5.2. Spatial Association Statistics.	98
5.2.1. Global Statistics.	99
5.2.1.1. Limitations of Global Statistics.	100
5.2.2. Local Statistics.	100
5.2.3. Anisotropy.	104
5.2.3.1. Dealing With Variable Anisotropy - Using the Landscape.	105
5.2.3.1.1. Watershed Sampling Schemes.	106
5.2.3.2. Multi-scale Geomorphometric Indices Using Landscape-based Sampling.	107
5.2.3.2.1. Calculating Geomorphometric Indices for Watersheds.	108
5.2.4. Spatial Association Summary.	111
5.3. Multivariate, Implicitly Spatial Analyses.	111
5.3.1. Modelling Approaches.	111
5.3.2. Decision Trees.	113
5.3.3. Feedforward Artificial Neural Networks.	115
5.3.3.1. Network Structure and Calculation.	116
5.3.3.1.1. Nodes and Layers.	116
5.3.3.1.2. Network Weights and Hyperplanes.	118
5.3.3.1.3. Calculating Network Output.	119

5.3.3.2.	Interpreting a Trained Network.	120
5.3.3.2.1	Using the Geographic Dimension.	121
5.3.3.3.	Using ANNs to Predict Interval Values.	123
5.3.3.4.	Bias in the Training Data.	124
5.3.4.	Multivariate, Implicitly Spatial Summary.	125
5.4.	Multivariate, Explicitly Spatial Analyses.	125
5.4.1.	Covariance.	125
5.4.2.	The Geographical Analysis Machine.	126
5.4.3.	Geographically Weighted Regression.	127
5.4.4.	Multivariate, Explicitly Spatial Summary.	128
5.5.	Assessing Multivariate Analysis Results.	129
5.5.1.	Interval Predictions.	129
5.5.2.	Confidence.	131
5.5.3.	Model Assessment Summary.	132
5.6.	Chapter Summary.	132
Chapter 6.	Methods Application.	133
6.1.	Introduction.	133
6.2.	Analyses.	133
6.2.1.	Spatial Association.	134
6.2.1.1.	Implementing the Watershed Sampling Method.	134
6.2.1.2.	Statistics.	135
6.2.2.	Artificial Neural Networks.	135
6.2.3.	Geographically Weighted Regression.	136
6.3.	Training Data.	136
6.3.1.	Dependent Variables.	137
6.3.1.1.	Spatial Extents.	137
6.3.1.2.	Numerical Scaling.	137
6.3.1.3.	Training Subsets.	138
6.3.2.	Independent Variables.	138
6.4.	Assessing the Models.	139
6.4.1.	Spatial Association Analyses.	140
6.4.2.	Multivariate Analyses.	140
6.5.	Chapter Summary.	141
Chapter 7.	Results.	142
7.1.	Introduction.	142
7.2.	Spatial Association Results.	142
7.2.1.	Semivariance.	142
7.2.1.1.	Directional Variograms.	144
7.2.2.	G_i^*	145
7.2.3.	Zones of Modern Hydrological Control.	150
7.2.4.	Multi-scale Geomorphometric Indices.	155
7.3.	Multivariate Analyses.	157
7.3.1.	Accuracies.	157
7.3.2.	ANN Analyses.	168
7.3.2.1.	Binary Predictions.	168

7.3.2.2.	Interval Predictions.	171
7.3.2.2.1	Numeric Range.	171
7.3.2.2.2	Spatial Distributions.	175
7.3.2.3.	Skewed Results.	177
7.3.3.	GWR Analyses.	180
7.3.3.1.	Numeric Range.	183
7.3.3.2.	Spatial Distributions.	183
7.3.4.	Spatially Adjusted ANN Results.	185
7.4.	General Implications.	188
7.4.1.	Modern Hydrological Control.	188
7.4.2.	Vegetation Spectral and Geomorphometric Relationships with Regolith Properties.	189
Chapter 8.	Discussion.	190
8.1.	Introduction.	190
8.2.	Model Error.	190
8.2.1.	Relative Enrichment.	191
8.2.2.	Spatial Association.	192
8.2.3.	The ANN Model.	193
8.2.3.1.	Binary Predictions.	193
8.2.3.2.	Similarity of Conventional and Multi-Scale Interval Analysis Results.	194
8.2.3.3.	ANN Analysis Summary.	195
8.2.4.	Geographically Weighted Regression.	195
8.2.4.1.	GWR Summary.	196
8.2.5.	Model Error Summary.	196
8.3.	The Landscape.	197
8.3.1.	Landscape Features and Processes.	198
8.3.1.1.	Vegetation.	198
8.3.1.1.1	Vegetation Disturbance.	198
8.3.1.1.1.1	Mining Activities.	198
8.3.1.1.1.2	Fire.	200
8.3.1.1.1.3	Cyclones and Tree-throw.	200
8.3.1.1.2	Other Vegetation Effects.	201
8.3.1.2.	Marine Incursions.	202
8.3.1.3.	Andoom Redsoil.	203
8.3.1.4.	Arcuate Features in Andoom.	205
8.3.1.5.	Parent Material.	206
8.3.1.5.1	Using TiO ₂ to Infer Variations in Parent Material.	207
8.3.1.5.2	Inferences.	208
8.3.1.6.	Melon Holes.	209
8.3.1.6.1	Present Day Control of the Regolith.	210
8.3.1.6.2	Formation Mechanisms of Melon Holes.	211
8.3.1.6.3	Palaeo-drainage Controls.	214
8.3.1.6.3.1	Antecedent Drainage.	214
8.3.1.6.3.2	Faulting.	215
8.3.1.7.	Landscape Features Summary.	216

8.3.2. Landscape Equilibrium.	217
8.3.2.1. Mineral Mobility and Exposure to Weathering Processes.	217
8.3.2.2. Spatial Patterns of Equilibrium Relationships. . .	220
8.3.2.2.1 Scale of Modern Hydrological Control.	221
8.3.2.2.2 Non-Hydrological Factors.	225
8.3.2.3. How Long Does It Take to Reach Quasi-Equilibrium?	226
8.3.2.4. Landscape Equilibrium Summary.	229
8.4. Broader Implications.	230
8.4.1. Comparison with Previous Work.	230
8.4.2. Other Landscapes and Properties.	232
8.4.3. Future Research Methods.	234
8.5. Chapter Summary.	235
Chapter 9. Conclusions.	237
Reference List	240
Appendix 1. Digital Data.	260

List of Figures.

Figure 1.1.	Spatial scales and time lags of geomorphic processes in channels.	5
Figure 1.2.	The entity and field data models.	9
Figure 1.3.	Study area location.	13
Figure 2.1.	General geomorphometric indices.	18
Figure 2.2.	Data structures used for elevation matrices.	19
Figure 2.3.	The elevation sub-matrix from which geomorphometric indices are calculated.	22
Figure 2.4.	The flow accumulation calculated using D8 can be half of what it should be.	26
Figure 2.5.	A system of the components of regolith mapping.	32
Figure 2.6.	Example variogram.	48
Figure 2.7.	Components of the kriging system.	49
Figure 2.8.	Examples of different approaches to regolith mapping.	57
Figure 3.1.	Study area features in 1995.	63
Figure 3.2.	Red-Green-Blue false colour composite of Landsat sub-scene..	64
Figure 3.3.	The water table fluctuations in the study area are extreme. . . .	65
Figure 3.4.	Layers within the bauxite.	69
Figure 3.5.	Generalised mineral profile in the Andoom Peninsula.	70
Figure 3.6.	Vegetation effects on the bauxite.	72
Figure 3.7.	Possible extent of a 5 m marine incursion with a 5 m storm surge superimposed.	74
Figure 3.8.	Vegetation in the study area.	79
Figure 4.1.	Drill core dataset and additional non-bauxite locations.	82
Figure 4.2.	DEM used in the analysis.	85
Figure 4.3.	General geomorphometric indices.	93
Figure 4.4.	Flow accumulation in cells.	95
Figure 5.1.	Comparison between watershed and circular sampling where there is hydrological control.	105
Figure 5.2.	Possible watersheds, and example geomorphometric values for two arbitrary distances in four of these watersheds.	110
Figure 5.3.	Example decision tree.	113
Figure 5.4.	Decision splits (lines) with boundaries parallel to axes (after Gahegan, 2000).	114
Figure 5.5.	Example feedforward neural network structure.	117
Figure 5.6.	Sloping hyperplanes used in ANNs.	118
Figure 5.7.	Geographic visualisation of ANN structure.	122
Figure 5.8.	Interval error matrix with tolerance levels.	130

Figure 7.1.	Omnidirectional, watershed and directional variograms.	143
Figure 7.2.	Gi* results for geometric and watershed sampling at the 300 m sampling distance.. . . .	146
Figure 7.3.	The percentage of the sample dataset where there is modern hydrological control.	151
Figure 7.4.	Zones of hydrological control at the 300 m sampling distance.. . . .	152
Figure 7.5.	Multi-scale elevation indices at the 150 m sampling distance.	156
Figure 7.6.	Error tolerance graphs for multivariate interval predictions. . .	158
Figure 7.7.	Error matrices for the interval predictions.	160
Figure 7.8.	Bauxite presence/absence analysis one, classification, possibility and confidence grids.	169
Figure 7.9.	Network structures for presence/absence analysis one.	170
Figure 7.10.	Interval ANN predictions, conventional indices.	172
Figure 7.11.	Network structure for interval ANN predictions using conventional indices, Al ₂ O ₃	174
Figure 7.12.	Error matrices for ANN analysis predicting all four mineral properties simultaneously.	178
Figure 7.13.	Error matrix for equalarea ANN, Al ₂ O ₃ , 500 bins.	179
Figure 7.14.	Ten cell GWR predictions, standard indices.. . . .	181
Figure 7.15.	GWR adjusted interval ANN predictions.	186
Figure 8.1.	Redsoil distribution in Andoom plotted against the G _i * results.	204
Figure 8.2.	Pipes and macropores below the bauxite layer, Andoom, through which groundwater flows preferentially.	212
Figure 8.3.	Weipa pH and element mobilities.	218
Figure 8.4.	Count of GWR <i>t</i> -scores with magnitude exceeding +/-2.	222

List of Tables.

Table 2.1.	Approaches to quantitative regolith mapping.	40
Table 3.1.	Dating techniques applied to the Weipa bauxite.	76
Table 3.2.	Vegetation ecosystems defined by Specht <i>et al</i> (1977).	77
Table 4.1.	Correlations between visible and infra-red bands in the Landsat image.	91
Table 5.1.	Calculation of network output.	119
Table 6.1.	Break points used to rescale the conventional independent variable data for training.	139
Table 7.1.	General response of omnidirectional variograms.	144
Table 7.2.	Spatial clustering of mineral and physical variables in the study area.	148
Table 7.3.	Error matrices of the binary predictions.	157
Table 7.4.	r^2 values for the interval predictions.	166