Soil Landscape Reconnaissance Mapping

NSW WESTERN REGIONAL ASSESSMENTS

FINAL REPORT JUNE 2002

Brigalow Belt South

Stage 2

Resource and Conservation Assessment Council

Soil Landscape Reconnaissance Mapping—Brigalow Belt South (BBS) Interim Bioregion

NSW WESTERN REGIONAL ASSESSMENTS

FINAL REPORT JUNE 2002

A project undertaken for the Western Regional Assessment of NSW
for the Resource and Conservation Assessment Committee

Project number: WRA 21

ORTA · RECENS · QUAM · PURA · NITES

NEW SOUTH WALES GOVERNMENT

For more information and access to data contact:

Resource and Conservation Division
Planning NSW
GPO Box 3927
SYDNEY NSW 2001
Phone: (02) 9762 8207

Department of Land and Water Conservation
Soil Information Systems Manager
soils@dlwc.nsw.gov.au
PO Box 3720
PARRAMATTA NSW 2124
Phone: (02) 9895 6172

This document should be cited as:

Soil Information Systems Unit (2002), *Soil Landscape Reconnaissance Mapping - Brigalow Belt South Stage 2*, Department of Land and Water Conservation, Sydney, NSW.

ISBN 1 74029 181 6

CONTRIBUTIONS AND ACKNOWLEDGMENTS

This program was funded by the New South Wales Government and managed through the Resource and Conservation Division, Department of Urban Affairs and Planning (DUAP). Thanks go to Ric Noble, John Ross and Emma Mason for their advice and assistance with many aspects of this project.

Many Department of Land and Water Conservation (DLWC) staff also contributed to this project and their efforts are greatly appreciated. The following soil surveyors provided soil survey and soil attribute data for the project: Rob Banks, Karl Andersson, Mark Blair, Brian Jenkins, Brian Murphy, Dacre King and Evan Pengelly. Katie-Jane Nixon, Mark Young, Nicole Simons, Kate Secombe and Alex McGaw provided technical assistance.

Thanks also go to Ross Beasley, Jeremy Black, Jeanine Murray, Gary Coady and Paul Wettin for making staff available for this project and to the Barwon Region's Resource Information Unit (RIU), in particular their GIS unit, for the provision of spatial information and expertise.

Simon Ferrier and his team at National Parks and Wildlife Service (NPWS) provided important technical advice on gap analysis and provided their code for use in the flexible gap analysis tool developed by Dr. Geoff Goldrick.

Special thanks are extended to Dr. Geoff Goldrick who set-up and produced a series of sophisticated soil landscape maps using geographic information and statistical modelling systems, as well as developing the flexible gap analysis software. This project could not have been completed without his significant inputs. Geoff received assistance in constructing the GIS modelling environment from Paul Rampant at the Centre for Land Protection (Bendigo, Victoria) and Neil McKenzie, Chris Moran, John Gallant and Elizabeth Bui from CSIRO Division of Land and Water.

DLWC's Barwon Region Resource Information Unit and the NSW Department of Mineral Resources provided many of the underlying datasets.

Casey Murphy, Katie-Jane Nixon, Alex McGaw, Nicole Simons and Mark Young constructed and populated the database, with Nicole preparing the accompanying user guide. They also loaded soil profile information into the NSW Soil and Land Information System (SALIS), and calculated many of the soil attributes. Jonathan Gray provided the lithology classification used in the database. Mark Young, Andrew Murrell and Hugh Jones completed the model accuracy checks.

Special thanks are also extended to Leanne Wallace, Peter Gallagher and Dorothy Mullins for their advice and support, within both DLWC and the WRA steering committee.

This project was initiated and managed by Greg Chapman, with technical supervision and quality control provided by Greg Chapman, Casey Murphy and Brian Jenkins.

This report is presented as per those prepared for previous Comprehensive Regional Assessments. It also builds on work presented in Goldrick *et al.* 2001. The report was written by Greg Chapman and was edited by Anne Bell, Brian Jenkins, Humphrey Milford, Casey Murphy, Mark Young and Andrew Murrell.

DISCLAIMER

CONTENTS

Figures

Tables

1 EXECUTIVE SUMMARY

1.1 WHAT HAS BEEN PROVIDED?

The NSW Department of Land and Water Conservation (DLWC) has provided reconnaissance-level soil landscape mapping and soil attribute information for the Brigalow Belt South (BBS) Interim Bioregion, an area of approximately 52,400km^2. This information is to be used in native vegetation mapping and to assist with the determination of soil and land capability for public land use allocation.

Starting in late January 2001, and with datasets made available in February 2002, this project provided fundamental soil and other biophysical attribute data for many modelling projects within the NSW Western Regional Assessment process. These projects include assessments of the distribution of individual plant and animal species, the extent and pre-European distribution of vegetation communities, and site quality and associated fertility for timber resources. Furthermore, this project provided information concerning soil properties and limitations that will prove useful for sustainable land use planning.

1.2 HOW WAS IT DONE?

Innovative computer modelling methods were used to map the BBS in a more rapid, quantitative and repeatable manner than could be achieved by using traditional reconnaissance land resource mapping methods. In the southern portion of the mapping area the technique used involved recursive partitioning of available environmental variables trained on existing soil landscape mapping to develop rules for allocating soil mapping classes to adjacent areas. The environmental variables included digital elevation models (DEMs) and climate interpolative drapes as well as geology and soil parent material maps and gamma radiometric imagery.

In the northern and western parts of the BBS, where soils information was not available, nine smaller training areas were chosen for representative soil mapping, which was then modelled throughout those areas based on the rules developed in the training areas.

Several key environmental variables such as surface geology and gamma radiometrics were neither continuous nor wholly available at the time of survey, which hampered modelling progress. Combined with limited digital elevation information in areas of lowest relief in the west of the study area, this presented challenges that need to be addressed before starting similar surveys.

Initial outputs from the model were originally applied universally over the entire area. It was found, however, that this over-represented the landscape-forming processes that dominate the south-eastern part of the study area but are not as prevalent elsewhere. The study area was subsequently subdivided into seven provinces. This allowed geographic differences in landscape and soil formation processes in the bioregion to be modelled individually. This approach was adjusted with further refining of province boundaries and subdivision of mapping areas for individual modelling.

Training area information, and other compiled soil data, were used for iterative refining of the soils map over the remainder of the area by using various combinations of repeated recursive partitioning.

Advice on model success or otherwise was only possible using critical feedback from the soil surveyors and technical officers who extensively sampled soils and landscape conditions in the study area. Location of sampling points was guided by the use of a flexible gap analysis software package running on laptop computers. The model-feedback approach was developed, in some cases, down to individual soil landscapes. In some areas, the model was modified more than eight times until further improvement could not be obtained within the constraints of the project. In some instances, modelling reverted to the use of radiometrics classifications, existing mapping and hand-drawn polygons because the available environmental variables could not sufficiently delineate soil boundaries.

A smoothing and editing process was employed to resolve edge matches caused by discontinuities in the environmental surfaces.

Mapping was conducted at a technical standard consistent with national agreements and standards developed under the Australian Collaborative Land Evaluation Program (ACLEP) by DLWC's Soil Survey Unit team of trained and qualified soil surveyors.

Soil and land descriptions follow the guidelines of the *Soil and Land Survey Field Handbook* (Macdonald *et al.* 1990). Soil and land data collection used a combination of integrated and free soil survey and is a synthesis of methods outlined in the *Australian Soil and Land Survey Handbook—Guidelines to Conducting Surveys* (Gunn *et al.* 1988), with the exception of the 15 km buffer.

The map unit database was joined with map coverage to enable various soil and landscape parameters to be portrayed.

1.3 WHAT ARE THE PRODUCTS?

Over 3,500 soil profiles were collected from existing records or described as part of this project and entered into DLWC's NSW Soil And Land Information System (SALIS). This compares with around 1,300 soil profile records available for the area at the start of the project. An independent sampling set, which is being finalised, will feature around 80 sites taken from the same locations as fauna and flora descriptions. This data will be valuable for double-checking against vegetation mapping predictions.

The main product is a reconnaissance Soil Landscape Map for use in Geographic Information System (GIS) format. Each mapping unit is linked to a corresponding record in an MS Access database. Main attributes include:

- Soil type;
- Soil depth;
- Drainage;
- Estimated rooting depth;
- Fertility;
- Estimated plant available water capacity; and
- Soil regolith stability class.

Further, each map unit has been assessed for localised or widespread presence of:

- Seasonal waterlogging, flood hazard, high run-on, groundwater pollution hazard;
- Gully erosion risk, sheet erosion risk, wind erosion hazard;
- Shallow Soil, non-cohesive soil, complex soil;
- Seepage scalds, saline discharge, potential recharge;
- Foundation hazard, surface movement potential; and

- Steep slopes, mass movement hazard.

Soil attribute information can be linked to Digital Elevation Models (DEMs) to produce a higher resolution of soil attributes at more intense scales.

A feature of the project was the use of 180 stratified random soil profiles to compare with predictions of the mapped surfaces. For the five major attributes tested the mapping was found to be correct to category level in more than 70% of cases. Soil Landscape mapping which is almost three times more expensive is expected to be accurate to category level in 85% of cases

Beyond the scope of this project, DLWC undertakes to make these and other map views available over the Internet in the future.

This mapping is provided at reconnaissance level and should be used only as a guide to the distribution of specific soil attributes identified for the purposes of this project.

2 INTRODUCTION

2.1 BACKGROUND

This report describes a project undertaken as part of the Western Regional Assessments (WRA's) of public lands in western NSW. The WRA's provide the scientific basis on which rational decisions are made to determine the best use of public land in western NSW. The Western Regional Assessment process is based on consecutive assessment of specified target areas.

The Brigalow Belt South (BBS) project is the second WRA project area. It follows Stage 1 assessments in forests in the east of the Southern Brigalow Interim Bioregion. Future assessment areas are likely to include land in the Nandewar and New England Tableland Interim Bioregions that were not studied as part of the North-East Comprehensive Regional Assessment (CRA) undertaken in 1997.

2.1.1 The Brigalow Belt South Interim Bioregion

The Brigalow Belt South Interim Bioregion (BBS) covers over 52,400km^2, or 6.2% of NSW. It extends from the Queensland border south to Narromine and east to Merriwa (see Figure 1). Gunnedah, Coonabarabran, Dubbo and Narrabri are its major town centres.

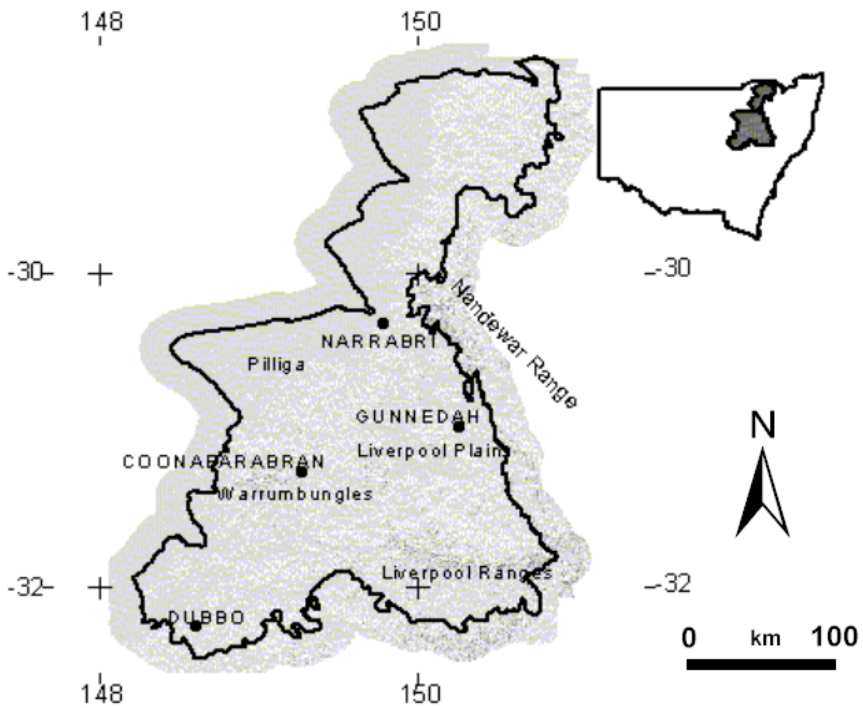


Figure 1: Location and extent of the Brigalow Belt South Interim Bioregion

The area is covered by extensively cleared eucalypt woodland and lies in a sub-humid zone with summer-dominated rainfall. Land uses include extensive grazing, cropping and forestry. The landscape consists broadly of basalt plateaux in the south with rolling hills and low hills in the east and north, descending to subdued rises and outwash fans in the west.

The underlying geology is diverse, including extensive areas of sediments dating from recent back to Jurassic times, large tracts of Pilliga and related sandstones in the west, and volcanic rocks forming mountains such as the Warrumbungles and the Liverpool Ranges. Soils range from highly fertile basalt-derived soils of the Liverpool Ranges and plains to very poor soils and outwash materials of the Pilliga sandstones. Soil erosion, salinity and soil structure decline are amongst the main degradation issues within the BBS.

2.1.2 Study rationale

This project was intended to provide essential soil and landscape information for input to vegetation mapping, determination of plantations potential and assessment of limitations to sustainable land use and land capability in the context of salinity, soil erosion and other forms of land degradation.

The Resource and Conservation Assessment Council (RACAC) of the NSW Government commissioned a series of studies in the BBS so that the best possible decisions regarding the long-term use of State-owned land (for both conservation and production purposes) could be made. These decisions will affect regional land use planning as well as conservation and land use management. The studies were to include assessments of geology, mineralogy, soils, vegetation, fauna, socio-economic factors and wood production. The soils inventory was considered important for the modelling of individual plant and animal species distributions, extant and pre-European distribution of vegetation communities, and site quality for wood resource productivity.

Mapped soil attributes (including depth, fertility, waterholding capacity and stability) are important for the vegetation mapping project, particularly as the terrain within the BBS is mostly subdued and soils assume a greater influence in determining plant species distribution. Soils information is a fundamental precursor to subsequent vegetation mapping so it was considered essential that soils information be provided as soon as practicable.

Mapped soil attributes were required as inputs to a number of modelling projects. Prior to this project, the only reliable soil information available was an incomplete coverage of 1:100 000 and 1:250 000 scale soil landscape mapping along with sporadic coverage by other mapping. None of these provided complete coverage of soil properties at an appropriate resolution across the entire area as required for vegetation modelling purposes. In particular, the distribution of soil attributes such as estimated soil fertility, soil drainage, effective rooting depths (ERD) and estimated plant available waterholding capacity (EPAWC) were considered essential parameters for effective vegetation modelling.

It should be noted that limitations regarding ERD and EPAWC exist in described locations where the substrate is not reached and the soil surveyor must estimate the soil depth using previous experience. This estimate may be inaccurate and can introduce significant errors into the calculated EPAWC values.

The particular aim of this project was to supply soil attribute data for inclusion in the vegetation modelling process coverage for the entire BBS. However, the spatial and temporal scales of the BBS Soil Landscape mapping project made it impracticable to use conventional soil landscape mapping techniques. The constraints of the project demanded the use of rapid, automated techniques with only limited reliance on time-intensive techniques such as air photo interpretation and map unit correlation.

It is stressed that the automated methods used require careful checking against soils and landscape conditions. The need for expert input and soil observations was recognised as fundamental to provision of a quality product. To meet these constraints it was initially decided to augment preliminary soil landscape models using GIS and existing data with traditional soil survey techniques of air photo interpretation and field sampling.

Predictive modelling of soil attributes has been shown to be a cost-effective means of improving the resolution and coverage of soil attribute mapping within narrow timeframes. The approach is based on work undertaken by McKenzie and Austin (1993); Moore *et al.* (1993); and Gessler *et al.* (1995). It involved the modelling of soil attributes recorded at field survey sites within each mapped parent material, climatic and topographic class (or soil landscape, if available) in relation to fine-scaled terrain and climate variables derived from DEMs.

This project covers the original BBS Interim Bioregion WRA project area as specified in the approved project proposal. The mapping is at reconnaissance level and should only be used as a guide to the distribution of specific soil attributes identified for the purposes of this project. One can expect that the accuracy of the model predictions will decrease in environments dissimilar to that in which it was trained.

Subsequent modifications were made to the WRA boundary to include more land outside the original project area. The project includes a 15 km buffer around the bioregion which was mapped on request for NPWS at a very coarse level (see Section 5.4.12 for more details).

2.2 PROJECT OBJECTIVES

The project objective was to expand the existing soil landscape coverage of soil attributes where little or no data was available, thus developing a mapped coverage of soil attributes across the entire region to assist with vegetation modelling and assessment of soil and landscape capability. This included:

- Using algorithms previously developed during the Upper and Lower North-East and Southern CRA projects (DLWC 1999a;b) and site criteria for collection and ranking of relevant parameters, including fertility, soil depth and soil waterholding capacity;
- Fitting of specific soil attributes to the existing Soil Landscape framework;
- Extension of the soil and landscape map framework over the remainder of the area; and
- Providing potential for greater resolution of soil attributes by making provision for allocation of directly unmapped subdivisions of soil landscapes (facets) that can be linked to digital elevation models and their derivatives, thus allowing them to be mapped individually at more intense scales.

3 METHODS

3.1 SCOPE OF PROJECT

The DLWC project requirement was for the provision of soil attribute information across the BBS Interim Bioregion. The Soil Landscape concept was used in conjunction with geographic statistical modelling to deliver the soil attributes. Existing soil landscape information was used wherever possible, with new reconnaissance soil landscape mapping and modelling undertaken in areas without existing soil landscape coverage.

A range of soil data was collected including soil drainage, soil depth and soil types. From the soils data a number of soil attributes were calculated or modelled including effective tree rooting depth, estimated soil available water holding capacity and soil fertility.

Soil landscapes were broken down to their constituent soil landscape facets, which were assessed for the full range of soil attributes. This partitioning provides the potential to predict soil attributes at resolutions approaching 1:25 000 scale.

Note: The soil landscape mapping provided by this project is at reconnaissance level and should be used only as a guide to the distribution of specific soil attributes identified for the purposes of this project.

3.1.1 Project planning and management

RACAC convened a meeting in Dubbo in November 2000 to determine the scope and linkages for the BBS reconnaissance soil landscape mapping project. The linkages discussed and subsequently adopted for the project plan included:

- Sharing of vegetation and soil data sites where practical;
- Collection of vegetation data at soil project sites, and collection of soils data at vegetation project sites;
- Training of vegetation field scientists in soil data collection, and training of soil surveyors in vegetation attribute data collection; and
- Sharing of geological and soil information between projects, particularly radiometrics and geological map outputs.

Specifications and plans for the DLWC soils component of the project were refined after this meeting. The proposal was developed by Greg Chapman (DLWC) on advice from Dr. Geoff Goldrick and based on modelling work presented by Rampant (2000).

The DLWC project was initially costed and planned using MS Project software. The plans (including initial deadlines and staff requirements) were then communicated to all involved.

The initial DLWC project proposal required a second costing due to budget cuts. As a result, the 15 km buffer around the study area was not to be used (a buffer zone was subsequently re-included at the request of vegetation modellers in February 2002).

The DLWC project was subsequently approved by the DLWC Executive and RACAC.

Monthly reports on progress were provided to the WRA Steering Committee. This reporting mechanism proved to be well suited to assisting with project linkages, providing background for project variations and communicating between projects.

The project steering committee, which included representatives who were at the initial project planning meeting from RACAC, State Forests of NSW (SFNSW), NPWS and DLWC, met at Gunnedah in May 2001. At this meeting changes in project methodology and linkages with other projects were discussed.

A briefing on geology of the mapping area at the NSW Department of Mineral Resources (DMR) was attended in March. Two further meetings were held in July and October 2001 with vegetation mapping staff. The July meeting discussed shared sampling site strategy and the second meeting focussed on landholder liaison and cross-project field sampling requirements.

A final project output training day was held in February 2002 at Dubbo. Here, the outputs from the project, the methodology used in the project, and the attributes recorded in the database, were explained to various agency staff.

3.1.2 Why the geographic statistical modelling approach was used

The spatial and temporal scales of the BBS reconnaissance soil landscape mapping project made it impracticable to use conventional soil landscape mapping techniques. Initial delays in project staffing and establishment further compounded the demand for rapid, automated techniques with limited reliance on time intensive techniques such as Aerial Photo Interpretation (API) and intensive field sampling.

To meet these constraints it was decided to derive preliminary soil landscape maps by using a geographic statistical modelling approach. This involved the use of a Geographic Information System (GIS) to extract existing soil landscape map unit classes and environmental values at selected sampling points. Rules for the environmental variable values were derived for each soil landscape mapping type. These rules were then applied to adjacent mapping areas and proto-soil landscape map units were derived. The process used is explained in Figure 2 (next page).

Small training areas were selected to provide initial training sets for the modelling where pre-existing soil landscapes did not exist. The models were then tested and refined through field observation and soil surveyor feedback.

3.1.3 Project variation

Project methodology was modified due to delays in both the project establishment and the delivery of datasets essential to the modelling.

Typically, conventional reconnaissance soil survey would consist of approximately one month of soil surveyor time per map sheet. Within this period a soil surveyor would undertake remote sensing and aerial photo interpretation, reconnaissance fieldwork and field data collection, record map unit descriptions, calculate required soil attribute parameters, edge match with adjacent map sheets, and transpose field sheets for digital scanning. Further time would be required beyond this period for editing, production of final maps, and the merging of the various map sheets into a single coverage.

By the time all the required resources were assembled the project was running approximately three months behind schedule. It was apparent that insufficient time was available to continue with conventional reconnaissance survey, supplemented by modelling, as per the original project brief. Aerial photo interpretation over the entire study area was abandoned in favour of further modelling based on the nine training areas and existing soil landscapes. A new work program was devised and is listed in section 4 of this report.

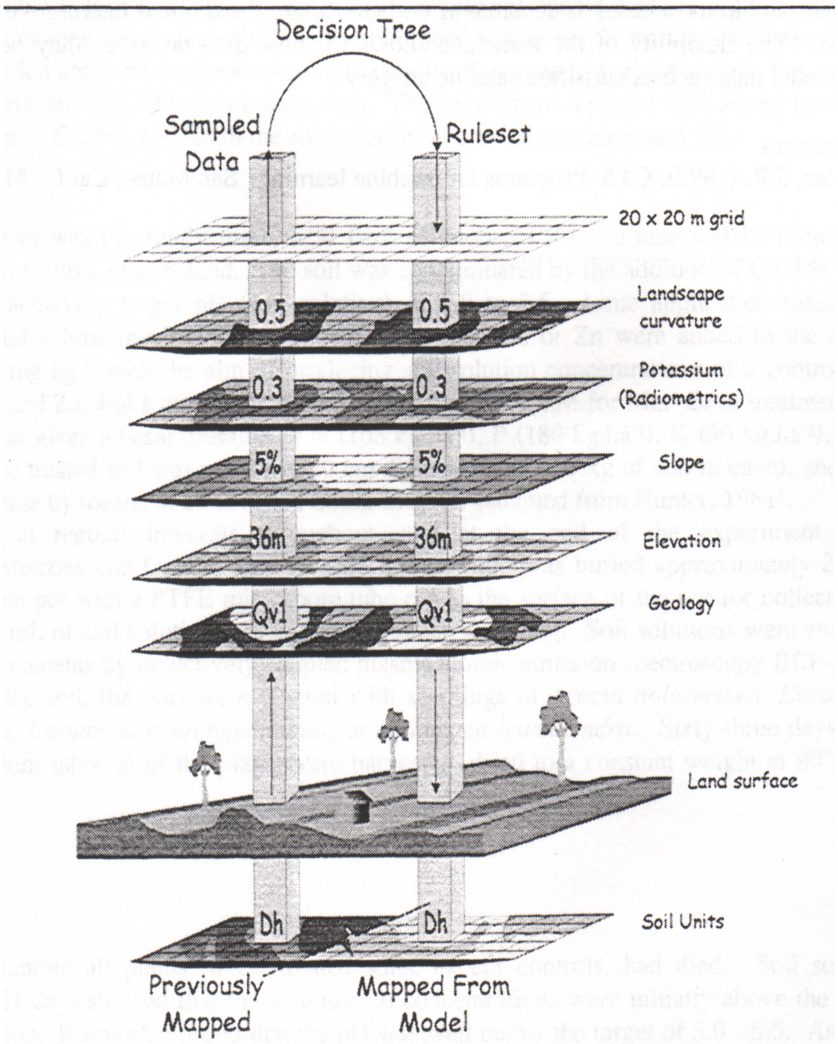


Figure 2: Diagrammatic representation of the use of regression models for predicting soil-landscape units

(Rampant 2000, p239)

3.1.4 What is a soil landscape?

Soil landscapes are defined as "areas of land that have recognisable and specifiable topographies and soils, that are capable of presentation on maps, and can be described by concise statements" (Northcote 1978). The mapping of landscape properties can be used to distinguish mappable areas of soils because similar causal factors are involved in the formation of both landscapes and soils. This allows both soil and landscape limitations and other parameters to be portrayed using the same map unit framework.

Traditionally, through remote sensing, interpretation of landscape features and the description of soils in the field, a soil landscape model can be built to predict the distribution and occurrence of different soil types within each landscape. To map soil landscapes, a set of rules are developed and applied based on landscape features. With the current project, the rules were sourced from existing mapping and training areas and then applied using GIS.

Different soil types have different soil attribute properties, and these can often be linked to digital elevation-based models for higher resolution of soil attributes. To gain this higher resolution, soil landscapes can be subdivided into smaller units known as Facets.

Facets are a way of dividing a landscape into discrete sub-units, each containing a distinct soil type or suite of soil types. Facets are mapped at a scale too fine for 1:100 000-scale modelling. Dividing a landscape into discrete units (or facets) is not a new idea, but has been used previously in Walker (1991). Figure 3 (below) shows an example of how a soil landscape can be subdivided into facets, referred to in this instance by landform element across the top of the diagram.

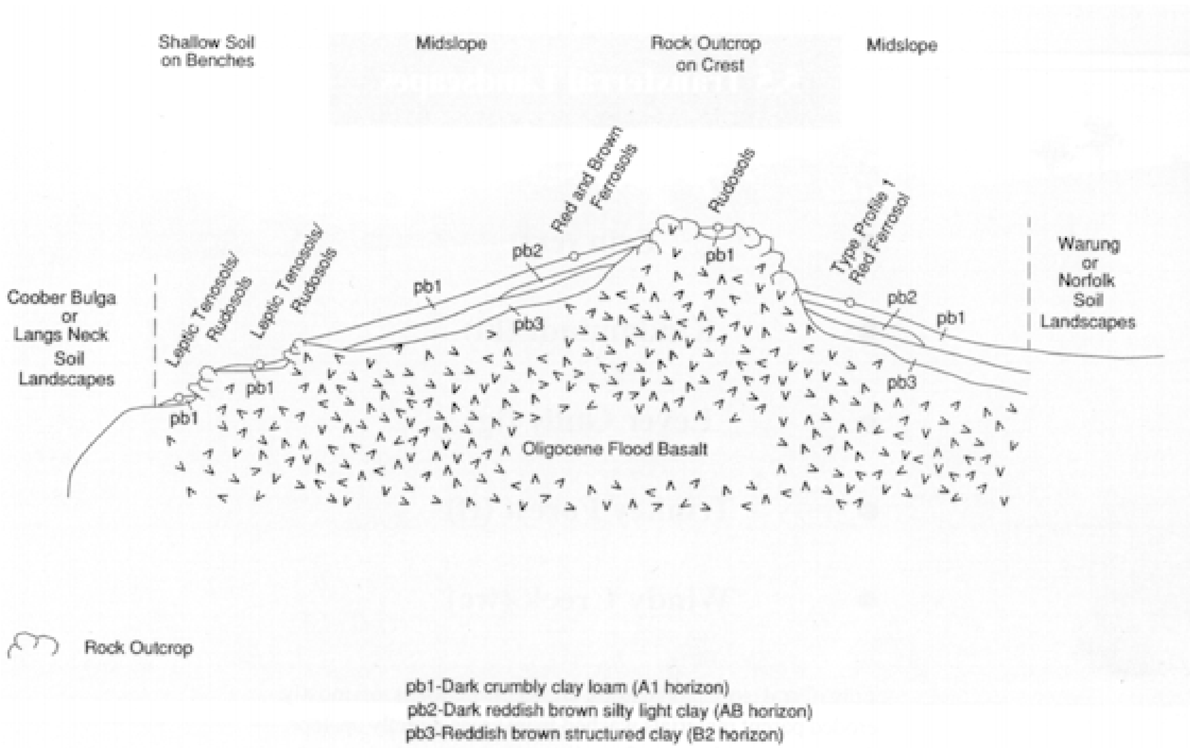


Figure 3: Cross-section of a soil landscape showing the relationship between the soil landscape and its facets

(Banks 1998, p85)

3.2 EXISTING DATASETS

3.2.1 Soil landscape maps

Existing soil landscape coverage for the BBS Interim Bioregion consisted of three published and five draft Department of Land and Water Conservation (DLWC) 1:100 000 soil landscape map sheets, plus two published and one draft 1:250 000 soil landscape map sheet. All existing soil landscape work was in the southern portion of the BBS (see Map 1).

Published soils information at 1:100 000 scale is available for the Curlewis (Banks 1995), Blackville (Banks 1998) and Tamworth (Banks 2001) 1:100 000 map sheets. This covers a portion of the Liverpool Plains and the Liverpool Ranges. The mapping of Hesse (2000) has covered a part of the Pilliga outwash. Published soil landscape mapping at 1:250 000 scale is available for the Singleton (Kovac & Lawrie 1991) and Dubbo (Lawrie and Murphy 1998) map sheets. Other sources of information include a CSIRO soils map of the Edgeroi 1:50,000 map sheet (Ward *et al*. 1992) and a reconnaissance soil map (Keady and Banks 1998) for part of the western extent of the study area.

Draft 1:100 000 scale soil landscape information and linework is available for the Murrurundi (McInnes-Clarke, in press), Boggabri (Banks, in progress), Baan Baa (Pengelly, in progress), Tambar Springs (Townsend & Pengelly, in progress) and Coolah (Townsend, in progress) map sheets. Soil landscape information and linework at 1:250 000 is available for the Narromine (Taylor, in press) map sheet. The mapping of many of these sheets was accelerated to provide data for the BBS project.

Extensive reconnaissance-level soil landscape survey and mapping needed to be undertaken for the remaining 21 (partial and complete) 1:100 000 map sheets not covered by the existing mapping.

3.2.2 Soil profile point data

Map 1 shows the distribution of the 1,364 soil profile descriptions available in the NSW Soil and Land Information System (SALIS) for the study area prior to the commencement of the project. This data was filtered using queries for their completeness to provide estimates of soil fertility, rooting depth and plant available water holding capacity.

During the course of the project, a further 1,990 soil profile descriptions were collected, either from fieldwork undertaken as part of the project or from data collected from other studies. This includes descriptions of soil samples collected as part of the vegetation plot assessments. The actual number of soil profiles described substantially exceeds the estimate of 1,550 profile points as per the original proposal specifications.

A separate set of 124 points was collected independently at various locations throughout the study area for checking purposes. Predominantly in the central and southern sections of the study area, the location of these points was based on stratified random sampling criteria based on 1:100 000 map sheets, geology and geographic spread. Approximately 90 further soil descriptions were located at fauna and flora description sites.

This data can be made available by sending an e-mail inquiry to soils@dlwc.nsw.gov.au.

3.3 CALCULATION OF KEY SOIL ATTRIBUTES

Key soil attributes for WRA modelling, namely soil fertility, soil drainage, effective rooting depth (ERD) and estimated plant available soil water-holding capacity (EPAWC) were identified and developed during the North-East CRA project and used in the Southern CRA project. The attributes selected were important for the 3PGR program that was used in those assessments to model and predict many aspects of plant growth.

A methodology for assessing the soil attributes required for vegetation modelling was developed for the North-East CRA project in consultation with Dr Phil Ryan (CSIRO Forests) and Dr Neil McKenzie (CSIRO Land and Water). The following outlines the methodology used to assess soil attributes within every facet of each soil landscape (i.e. unmapped partitions of the soil landscape) for both existing soil landscape information and for new reconnaissance soil landscape mapping.

Soil landscape descriptions including assessment of soil attributes for each facet have been entered into a specially designed MS Access database.

3.3.1 Modified fertility class

Five soil fertility classes (see Table 2) were developed, based on the *Great Soils Group (GSG)* classification (Stace *et al*. 1968), using a method outlined originally in Charman (1978) and subsequently in Murphy *et al.* (2000). A class of "1" indicates a soil of very low fertility, while a class of "5" indicates a soil with high fertility. *Modified soil fertility classes* were ascribed to each GSG based on Table 2. Fertility classes were raised or lowered dependant on positive or negative soil fertility attributes that differed form the nodal soil description. For example, a facet such as a crest with a Red Podzolic Soil has a fertility class of "3" (see Table 2), but this classification can be downgraded to a modified fertility class of "2" if the soil is excessively stony and/or very shallow. Conversely, a soil's modified fertility class may be upgraded if the soil has positive soil fertility properties, such as considerable depth, free drainage or high organic matter content in the topsoil.

Table 1: Fertility classes of Great Soil Groups

after Charman (1978)

Great Soil Group	Fertility Class
Solonchak	1
Alluvial Soil	5
Lithosol	1
Calcareous Sand	1
Siliceous Sand	1
Earthy Sand	1
Grey-brown Calcareous Soil	1
Red Calcareous Soil	1
Desert Loam	1
Red and Brown Hardpan Soil	1
Grey Clay	3
Brown Clay	3
Red Clay	3
Black Earth	5
Rendzina	3
Chernozem	5
Prairie Soil	5
Wiesenboden	3
Solonetz	2

Great Soil Group	Fertility Class
Solodized Solonetz	2
Solodic Soil	2
Soloth (Solod)	2
Solonized Brown Soil	2
Red-brown Earth	4
Non-calcic Brown Soils	4
Chocolate Soil	4
Brown Earth	3
Calcareous Red Earth	2
Red Earth	3
Yellow Earth	2
Terra Rossa Soil	3
Euchrozem	4
Xanthozem	3
Krasnozem	4
Grey-brown Podzolic Soil	2
Red Podzolic Soil	3
Yellow Podzolic Soil	2
Brown Podzolic Soil	3
Lateritic Podzolic Soil	1
Gleyed Podzolic Soil	3
Podzol	2
Humus Podzol	2
Peaty Podzol	2
Alpine Humus	3
Humic Gley	2
Neutral Peat	2
Alkaline Peat	2
Acid Peat	1

3.3.2 Drainage

Five drainage classes were defined based on the classes used by the SALIS Soil Data Cards (Abraham & Abraham 1992; Milford *et al*. 2001; and McDonald *et al*. 1990). They are:

1. Very poorly drained;
2. Poorly drained;
3. Imperfectly drained;
4. Moderately well-drained; and
5. Well-drained.

3.3.3 Effective rooting depth

Effective rooting depth (ERD) is an estimate of the soil and substrate available for tree roots to penetrate. It is an important factor in the calculation of *Estimated Plant Available Waterholding Capacity (EPAWC).* Where the parent material has not fractured, or where an impeding layer for tree roots exists (e.g., pan or rock), then an estimate of ERD has been calculated. This is based on the average depth of soil and regolith that tree roots are likely to penetrate (ERD). Where the parent material is fractured, tree roots will be able to penetrate both the solum and, to some extent, weathered parent material. Figure 3 shows how ERD is determined.

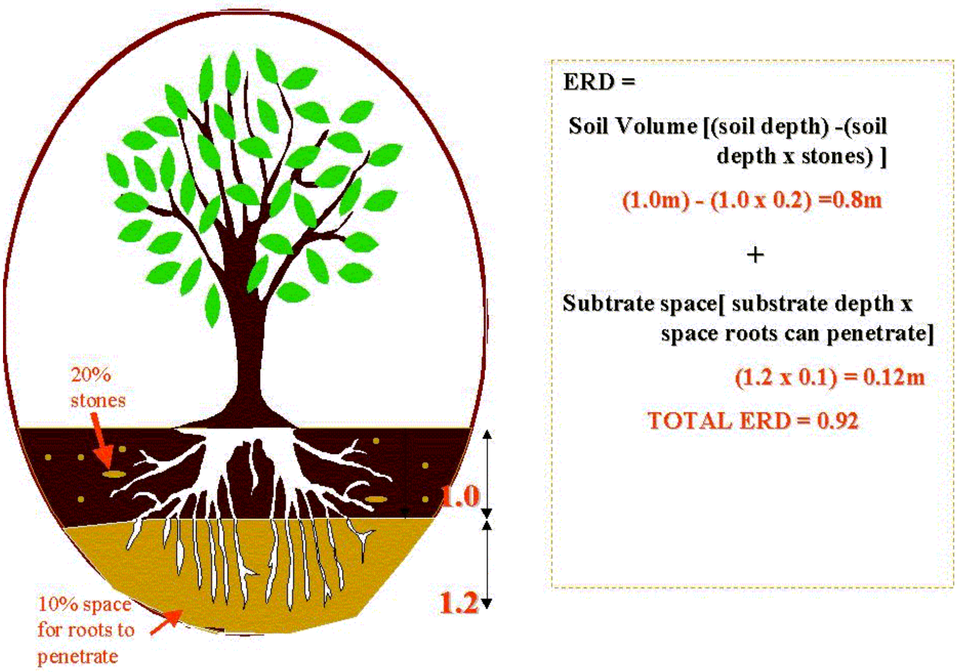


Figure 4: The calculation of estimated rooting depth

It should be noted that ERD can be a problematic attribute to measure in rocky soils and soils deeper than 1.5 m. In such instances a modified estimate is provided based on the properties of the lowest horizon and the probable solum depth. A maximum ERD of six metres is given as an arbitrary limit to root penetration in deep alluvial soils. It should be noted that there are instances of many trees sending roots far deeper than this to access water from aquifers.

To calculate the ERD:

Estimate the size, depth and number of fractures in the parent material and estimate an **average** depth that roots will be able to penetrate;
Add this to the depth of the solum; and
Subtract the Fragment Amount volume (see below) from the final calculation to get the effective rooting depth.
Example:

The soil depth is **1.2 m;** fragment amount is **10%**. The substrate is *fractured*, so roots will penetrate the substrate. The depth of the substrate to which the roots will penetrate is estimated to be **2.5 m**, but only **20%** of the substrate volume is available (i.e., cracks, etc).

Derivation of ERD is achieved through the following formula:

ERD = soil depth + (substrate volume available to roots % x root penetration into substrate) - (fragment amount % x soil depth)

Thus:

ERD = 1.2 m + (20% x 2.5 m) - (10% x 1.2 m) = 1.58 metres

3.3.4 Estimated plant available water holding capacity

Estimated Plant Available Waterholding Capacity (EPAWC) is an estimation of a soil's capacity to store water for use by plants. The methodology for the calculation of EPAWC is outlined by Greacen and Williams (1983) with reference to work undertaken by Salter, Berry and Williams (1966) and Salter and Williams (1963;1965;1967;1969). EPAWC is based on the strong relationship that exists between soil texture and available water holding capacity.

The EPAWC of a soil profile is calculated by multiplying the soil texture grade by the soil structure factor (1.2 for finely structured soils) which in turn is multiplied by the horizon thickness in metres. This is repeated for each horizon inside the ERD. The EPAWC for the soil profile is the sum of EPAWC calculated for all layers.

Table 2 : Estimated plant available water holding capacity (EPAWC) values for texture grades

(modified from Salter & Williams 1967, 1969, Gracen & Williams 1983 and Hazelton & Murphy 1992)

TEXTURE	EPAWC (mm of water stored per m of soil)
Sand	150
Coarse sand	80
Fine sand	200
Loamy sand	160
Loamy coarse sand	108

TEXTURE	EPAWC (mm of water stored per m of soil)
Loamy fine sand	217
Clayey sand	150
Light clayey sand	150
Heavy clayey sand	150
Clayey coarse sand	80
Light clayey coarse sand	80
Heavy clayey coarse sand	80
Clayey fine sand	215
Light clayey fine sand	215
Heavy clayey fine sand	215
Sandy loam	180
Light sandy loam	180
Heavy sandy loam	180
Coarse sandy loam	125
Light coarse sandy loam	125
Heavy coarse sandy loam	125
Fine sandy loam	192
Light fine sandy loam	192
Heavy fine sandy loam	192
Loam	180
Loam, fine sandy	185
Silty loam	200
Light silty loam	200
Heavy silty loam	200
Sandy clay loam	150
Light sandy clay loam	150
Light - medium sandy clay loam	150
Medium sandy clay loam	150
Heavy sandy clay loam	150
Coarse sandy clay loam	140
Light coarse sandy clay loam	140
Light - medium sandy clay loam, coarse sandy	140
Medium sandy clay loam, coarse sandy	140

TEXTURE	EPAWC (mm of water stored per m of soil)
Heavy coarse sandy clay loam	140
Fine sandy clay loam	180
Light fine sandy clay loam	180
Heavy fine sandy clay loam	180
Clay loam	180
Light clay loam	180
Medium - heavy clay loam	180
Heavy clay loam	180
Clay loam, coarse sandy	170
Clay loam, sandy	175
Light clay loam, sandy	175
Heavy clay loam, sandy	175
Clay loam, coarse sandy	170
Light clay loam, coarse sandy	170
Heavy clay loam, coarse sandy	170
Clay loam, fine sandy	190
Light clay loam, fine sandy	190
Heavy clay loam, fine sandy	190
Silty clay loam	190
Light silty clay loam	190
Heavy silty clay loam	190
Light silty clay loam, fine sandy	195
Sandy clay	140
Sandy light clay	140
Sandy light - medium clay	140
Sandy medium clay	140
Sandy medium - heavy clay	140
Sandy heavy clay	140
Coarse sandy clay	130
Coarse sandy light clay	130
Coarse sandy light - medium clay	130
Coarse sandy medium clay	130
Coarse sandy medium - heavy clay	130

TEXTURE	EPAWC (mm of water stored per m of soil)
Coarse sandy heavy clay	130
Fine sandy clay	150
Fine sandy light clay	150
Fine sandy light - medium clay	150
Fine sandy medium clay	150
Fine sandy medium - heavy clay	150
Fine sandy heavy clay	150
Silty clay	183
Silty light clay	183
Silty light - medium clay	183
Silty medium clay	183
Silty medium - heavy clay	183
Silty heavy clay	183
Clay	180
Light clay	180
Light - medium clay	180
Medium clay	180
Medium - heavy clay	180
Heavy clay	180

3.4 ANALYSIS

3.4.1 Geographic modelling tools

The geographic modelling environment for this study is summarised in Table 3.

Table 3 : The GIS software packages and their functions

Software	Function	Source
ESRI ArcView 3.2	Desktop vector GIS	www.esri.com
Spatial Analyst	Extension for ArcView 3.x enabling analysis of raster data (grids)	www.esri.com
ER Mapper	Image analysis software for converting and manipulating gamma radiometric and satellite imagery datasets	www.ermapper.com
Image Analyst	Extension for ArcView 3.x enabling analysis of image data	www.esri.com
S-Plus	Statistical analysis package	www.cmis.csiro.au/S-Plus
RPART	Routines for the construction of classification and regression trees based on recursive partitioning within S-Plus (Therneau & Atkinson 1997, Venables & Ripley 1999, Atkinson & Therneau 2000)	http://lib.stat.cmu.edu/DOS/S/Swin/Rpart.zip
SLAP	Extension for ArcView for terrain analysis, coordinate transformation, derivation of erosion indices and various raster and vector manipulations, developed by Dr Geoff Goldrick	Soils Information Systems Unit, DLWC

3.4.2 Soil landscape description recording

Soil landscape descriptions including assessment of soil attributes for each facet were entered into the MS Access database.

Enhanced resolution of the soil attribute information can be undertaken by linking DEMs with attributes of these facets (stored in the MS Access database) based on Compound Topographic, Elevation, Aspect and Solar Radiation Index and other factors of organisation that determine distribution of soil types within individual soil landscape map units.

4 WORK PROGRAM

The work program included the following steps:

4.1 PROJECT SET-UP PHASE

1. Marshall staff, initial training, obtain specialist software and equipment;
2. Literature search and entry of existing soil profiles into SALIS;
3. Establish map unit and facet strings;

Build MS Access database for soil landscape descriptions; design, test and populate;

4. Soil profile attribute data entry to SALIS;
5. Assemble predictor surfaces and prepare predictor surfaces for proto-soil landscape modelling;
6. Predictor variable preparation;
7. Initial data modelling;
8. Cross-validation;
9. Smoothing of the output prediction surfaces to become maps; and
10. Initial Model Evaluation.

4.2 TRAINING AREA PHASE

1. Compilation of independent check dataset;
2. Staff briefing and work allocation;
3. Training area selection;
4. Training area pre-field preparations;
5. Training area fieldwork; and then
6. Training area digitising and database entry.

4.3 MAIN FIELDWORK AND MODELLING PHASE

1. Post-training area modelling and model iterations;
2. Modelling by province and weighting of modelling within provinces;
3. Site selection and gap analysis;

4. Gap analysis;
5. Flexible gap analysis;
6. Main fieldwork; and
7. Landholder communication protocols.

4.4 POST FIELDWORK PHASE

Each of the steps is detailed below:

1. Final modelling;
2. Province boundary adjustments;
3. Line work adjustments;
4. Polygon tag changes;
5. Soil profile point data gaps;
6. Soil landscape database gaps;
7. Mopping up fieldwork;
8. Linkage of soil landscape polygons and database;
9. Attribution of soil landscape spatial summary parameters;
10. Database attribution with major variables and calculation of key soil;
11. Map and database checking;
12. Design of Web products;
13. Access to information; and
14. Final report.

5 PROJECT TASKS

The project comprised the following essential tasks:

5.1 PROJECT SET-UP PHASE

5.1.1 Marshall staff, initial training, obtain specialist software and equipment

Agreements were reached for the participation of DLWC's Barwon Region soil surveyors and GIS staff, as well as for Dr Geoff Goldrick of DLWC's North Coast Region. Two Technical Officers were recruited to provide field and office support for the project. Arrangements were made for Dr Goldrick to visit soil scientists at the Centre for Land Protection Research and at CSIRO Division of Land and Water for training in the use of recursive partitioning statistical techniques. Other key staff members were trained to develop the MS Access database to be used for storing map unit descriptions. A high capacity PC suitable for heavy GIS work was also ordered. This phase took four months to complete due to innumerable delays in gaining approvals.

5.1.2 Literature search and entry of existing soil profiles to SALIS

An extensive literature search was carried out. Soil profile descriptions and other data pertaining to the project area was collected and entered into SALIS, amounting to some 1, 364 profile descriptions.

5.1.3 Establish map unit string

A map unit code string that also includes information about unmapped facets (Table 4) was devised. Use of this map unit string ensured that the soil surveyors were describing and mapping soil landscapes consistently. The string provided a shorthand way to describe a landscape for correlative purposes, as well as to provide a unique linkage between the database and GIS coverage. The unique linkage was vital for populating database fields from the derived spatial surfaces.

The map unit string contains 13 alphanumeric characters and is described in Table 4. Only the last three or four characters of the soil landscape code were displayed on the maps— this ensured that map unit tags would fit comfortably within polygons.

Table 4: Map unit and facet string layout

Province Number (13-18)	Lithology Code (1st upper case 2nd lower case)	Landform Relief Modal slope (2-letter upper case)	Landform Attribute or Element (4-letter prefix lower, remainder upper case)	Soil Landscape Code (3 to 4-letter lower case)	Facet code (single numeral)
18 *Pilliga*	Vm *Volcanic, mafic*	RH *Rolling Hills*	pHIL *Hills*	ahz *Ant Hill*	2 *Second largest facet in landscape*

5.1.3.1 Provinces

To achieve the project's aims within the restricted schedule, a number of tasks were implemented to ensure orderly, on-time progress was achieved.

The BBS region was split into a number of *Provinces* (see Map 3), each of which corresponds to one of the Interim Bioregion subdivisions developed by Morgan and Terry (1992). Use of these subdivisions allowed the BBS to be divided into broadly distinct physiographic areas, each of which would be allocated to an individual soil surveyor. The provinces are identified by unique numbers as follows:

13 = Northern Outwash

14 = Northern Basalts

15 = Pilliga Outwash

16 = Liverpool Plains

17 = Pilliga

18 = Liverpool Ranges

19 = Talbragar

Province numbers less than 13 have already been used by previous projects.

The province boundaries were subsequently subdivided and adjusted using existing geology maps and the work of Hesse (2000) (see Map 4).

Some provinces were further subdivided to facilitate modelling as the survey was nearing completion. These changes did not require any further alterations to the string.

5.1.3.2 Lithology code

Lithology (as a substitute for soil parent material) has proven to be useful in mapping the soils of the Upper North-east and Southern CRA Areas. Lithology in these areas simplified a plethora of geological mapping units of slightly different ages but fundamentally the same geology and soils. The lithology code used was based on Gray and Murphy (1999) and is listed in the table below. However, in the case of the BBS Interim Bioregion, the lithology code occasionally did not distinguish between separate geological units with the same lithology codes, but appreciably different soils.

Table 5: Table listing types and codes for lithology used in the map string

Code	Description	Examples
Ac	Alluvial clays	
Ag	Alluvial gravels	
Al	Alluvial loams	
Ao	Alluvial organic	
Aq	Alluvial quartz sands	
As	Alluvial lithic sands	
Ax	Alluvial undefined	
Cc	Colluvial clays	
Cg	Colluvial gravels	
Cl	Colluvial loams	
Cq	Colluvial quartz sands	
Cs	Colluvial lithic sands	
Cx	Colluvial undefined	
Ec	Aeolian clays	
Eq	Aeolian quartz sands	
Es	Aeolian lithic sands	
Ex	Aeolian undefined	
Lc	Lacustrine clays	
Ll	Lacustrine loams	
Lo	Lacustrine organic	
Lq	Lacustrine quartz sands	
Ls	Lacustrine lithic sands	
Lx	Lacustrine undefined	
Ma	Metamorphic argillaceous	shale, slate, phyllite, schist
Mc	Metamorphic calcareous	marble
Mi	Metamorphic intermediate	granulite, eclogite
Mm	Metamorphic mafic	greenstone, amphibolite, serpentinite
Mq	Metamorphic quartzose	quartzite
Ms	Metamorphic siliceous	siliceous gneiss
Mx	Metamorphic undefined	hornfels

Code	Description	Examples
Oc	Marine clays	
Og	Marine gravels	
Ol	Marine loams	
Oo	Marine organic	
Oq	Marine quartz sands	
Os	Marine lithic sands	
Ox	Marine undefined	
Pi	Plutonic intermediate	granodiorite, syenite, monzonite, diorite
Pm	Plutonic mafic	gabbro, dolerite, wherlite
Ps	Plutonic siliceous	granite, adamellite, quartz porphyry
Pu	Plutonic ultramafic	dunite, peridotite
Px	Plutonic undefined	porphyry
Sa	Sedimentary argillaceous	mudstone, claystone
Sc	Sedimentary calcareous	limestone, dolomite
Sf	Sedimentary ferromanganiferous	laterite, ferricrete, bauxite
Si	Sedimentary intermediate	greywacke
Sl	Sedimentary lithic (coarse)	lithic sandstone and conglomerate
So	Sedimentary organic	coal, peat
Sq	Sedimentary quartzose	chert, jasper, silcrete
Ss	Sedimentary siliceous	quartz sandstone, quartz siltstone
Sx	Sedimentary undefined	turbidite, clastics
Ux	Unconsolidated undefined	
Vi	Volcanic intermediate	dacite, trachyte, andesite, intermediate tuff
Vm	Volcanic mafic	basalt, picrite, mafic tuff
Vs	Volcanic siliceous	rhyolite, dellenite, siliceous tuff
Vx	Volcanic undefined	tuff, pyroclastic, pumice, ash
Xx	Undefined	

5.1.3.3 Modal slope code

The two-letter modal slope codes used in this project are based on those of McDonald *et al.* (1990). Examples include **UR** - undulating rises; **SH** - steep hills; and **PM** - precipitous mountains.

5.1.3.4 Landform pattern or element code

These three-letter codes describe unique landform patterns or elements, as described in McDonald *et al*. (1990). These were used to determine similarities between map units described by different soil surveyors. Attributes of landform patterns cover such things as **HIL** - hills, **SAN** - sandplain, **ALP**-alluvial plain and **DUN** - dunefield. Where significant, landform element codes were also used—e.g., **HCR** - hillcrest, **HSL** - hillslope, **FOO** - footslope, **BRI** - beach ridge and **TAL** - talus, etc. Prefixes "p" and "e" identify landform patterns and elements respectively.

5.1.3.5 Soil landscape code

Soil landscapes were used to delineate areas of different soils by using distinguishing landscape features (especially those important to vegetation modelling). Each soil landscape was given a three-letter code. Three-letter codes allowed the inclusion of published two-letter soil landscapes codes (e.g., *ki* from the local geographic name Kindamindi) and their variants with three-letter codes (e.g., kia) to be given the same length character code. To achieve a consistent three-letter code system across all mapped units, normal soil landscapes were allocated an extra letter (using a descending series starting with *z*), which also allowed different soil landscapes from different areas with the same two-letter code to be distinguished (e.g., *ki* would become *kiz*).

Later, a four-letter code was added so that the originating soil surveyor could be linked with the modelled output and the map unit description.

A soil landscape variant usually has a different property than the parent soil landscape, e.g., shallower soils, but generally all other soil landscape features are similar. The letter *a* is used to identify the first soil landscape variant of the parent soil landscape. Any subsequent variants were given alphabetically ascending postscripts (e.g., *b*, *c*, *d*, etc). Although not mutually exclusive, the code was linked in the database to the long map unit string code, which was unique and ensures that each soil landscape is linked to appropriate data in the database.

5.1.3.6 Facet code

As mentioned earlier, facets are a way of dividing a landscape into discrete sub-units that each contains a distinct soil type or suite of soil types at a scale too fine to be depicted on 1:100 000-scale maps. This methodology has been used previously in generating soil landscapes in the BBS area—e.g., the soil landscape modelling by Banks (1998) uses facets and was included during modelling.

The soil surveyor nominates the percentage area of the soil landscape covered by each facet. The percentage of the parent soil landscape does not reflect any modelled probability of occurrence or a modelled spatial percentage.

Modellers can map facets at a finer scale by using collateral data sets like DEMs because each facet is given a description of the main landform, lithological, climatic or soil factor used to delineate it from other facets. This is known as a Factor of Organisational Grouping (FOG) value (see Table 7).

For each facet in any soil landscape, an identifier is used. Facets are generally numbered in accordance with the area that they occupy, with the number 1 being allocated to the most extensive facet, 2 allocated to the next most extensive facet, and so on.

Example of a long string map unit code. The map unit code *18VmRHpHILahz(2)* is used to define a map unit on the *Liverpool Ranges* province with *volcanic mafic* lithology, a *rolling hill* modal slope class, a *hills* landform pattern, an *ahz* soil landscape code and the *(2)* second largest facet. A record at facet level in the database shows that this facet occurs across about 10% of the soil landscape, its location is determined by elevation and that the elevation range is the highest in the landscape. This long string code is unique in NSW for all soil landscape map units.

5.1.4 Build MS Access database for soil landscape descriptions

An MS Access database similar to the database used for the Southern CRA soils project and other soils mapping projects in western NSW was developed and populated for this project.

The central database was used to set-up, correlate, store and keep track of the soil landscape information collected by the soil surveyors. A user-friendly data entry screen with pull down buttons for many attributes ensured that consistent information was entered, reducing the likelihood of errors in the dataset. The data entry screen is linked to separate soil landscape and facet tables that allow ready export and linking with other databases and GIS.

Each soil surveyor was given a copy of the database, which was to be filled in following fieldwork and its contents added to the main central database. To ensure version control and to prevent any confusion between map unit description updates, a file-naming protocol was implemented. Centralising the database allowed the data to be readily verified and controlled at a single location. Table 6 provides a list of the information recorded in the database for each soil landscape.

Table 6: Soil and landscape properties recorded for each soil landscape unit in the MS Access database

Database attribute	Description
Soil landscape name	The name of the soil landscape
Soil landscape code	Three to four-letter soil landscape string that occurs on the maps. It is not unique and the code can occur on numerous mapsheets, but is linked to the soil landscape string and secondary mapsheet units that identifies the correct dataset in the database.
Protosoil-landscape code	Original map unit code, derived from model.
Soil landscape string	13-character string contains code information for province, lithology, modal slope code, landform pattern or element and soil landscape/soil landscape variant code. It is unique in the database.
Completed by	Identifies the person responsible for the entry of the soil landscape into the database.
Date completed	Identifies the date the soil landscape was entered into the database.
Mapsheet number	Topographic 1:100 000 mapsheet number, e.g., 8838.
Process group	Soil landscape geomorphologic groups. Examples include residual, erosional, alluvial and aeolian.

Database attribute	Description
Province number	One of six province codes is entered. Makes up part of landscape string.
Lithology code	DLWC lithology codes.
Modal slope code	Modal slope code entered.
Landform pattern or element code	Landform pattern - element class.
Included soil landscape description	Lists soil landscapes included within the soil landscape being described.
Variant description	Variant description entered
Geology code	The geology code.
Minimum and Maximum elevation	Minimum and maximum elevation (in metres) of the soil landscape.
Minimum and Maximum slope	Minimum and maximum slope for the soil landscape.
Local relief	Relief of the soil landscape.
Rock outcrop	% of rock outcrop in soil landscape.
Dominant lithology	Dominant CSIRO lithology code.
Sub-dominant lithology code	Sub-dominant CSIRO lithology code.
Dominant structural formation class for vegetation	Dominant original structural class for vegetation.
Sub-dominant structural formation class for vegetation	Sub-dominant original structural class for vegetation.
Dominant land use	Up to three dominant land uses.
Sub-dominant land use	Up to three sub-dominant land uses.
Confidence ranking for soil landscape	A confidence ranking (1 - 9) that the soil landscape unit is applicable.
Count of total number of facets	A total count of the number of facet records within the selected soil landscape record.
Landscape limitations	Major soil and landscape limitations that are likely to be present and pose restrictions to urban and rural activities. Limitations are described as either localised or widespread. Limitations include steep slopes, mass movement hazard, rockfall hazard, flood hazard, waterlogging, permanently high watertables, seasonal waterlogging, water and wind erosion hazard, high run-on, mine subsidence district, shallow soil, non-cohesive soil, surface

Database attribute	Description
	movement potential, rock outcrop, high foundation hazard, groundwater pollution hazard, wave erosion hazard, gully and sheet erosion risk, complex soils, periodically frozen soils, potential saline aquifer recharge zone, saline discharge zone, salinity hazard, seepage scalds and woody weeds.
Notes and comments	Notes on the soil landscape.

The database also contains data collected and applicable at the facet level. The pertinent fields are detailed in Table 7. The table contains fields that relate to Factor of Organisational Groupings (FOGs). FOGs are the external attributes by which a facet can be delineated using GIS to map likely facets at intense scales of 1:25,000. The characteristics of the dominant facet can be used as an analog to the behaviour of its parent soil landscape, or alternatively weighted averages can be used to show the gross distribution of soil depths, fertility and plant available water capacities across the landscape.

Table 7 : Facet properties recorded for each soil landscape unit in the MS Access database

Database attribute	Description
Facet code	14-character string code (the soil landscape 13-character string with an extra number added as a postscript and starting with the number 1 and increasing consecutively for each new soil sub-landscape present), e.g., 18Vm RHpHILahz(1) is the first soil sub-landscape in the map unit 18VmRHpHILahz landscape. Every soil landscape has at least 1 and generally ≥ 4 soil sub-landscapes.
Additional facet notes	Additional facet notes.
% of soil landscape	the average % of area which the facet covers.
Nature of profile	The type of profile used. Includes published type profile, field observation, CRA type profile and other information.
SALIS type profile and associated SALIS survey number	SALIS profile number and associated SALIS survey number.
GSG and its associated fertility	Great Soil Group classification and associated fertility.
Soil depth	The mean estimated soil depth is entered for each facet.
Mean estimated effective rooting depth	Mean estimated effective rooting depth, i.e., the volume of soil/voids in substrate that are accessible by tree roots is provided.
Mean modified fertility	Mean modified fertility class for each soil sub-landscape if given .
Estimated plant available	EPAWC is provided for each soil sub-landscape.

waterholding capacity (EPAWC)	
Mean drainage	Mean drainage class of the soil sub-landscape.
Soil regolith class	Soil regolith stability classification after Murphy, Fogarty and Ryan (1998) classifies the stability of a soil for forestry uses into four classes. Class 1 is stable coherent soils with low sediment delivery potential to streams. Class 2 is non-coherent sandy soils with low sediment delivery. Class 3 is coherent soils with high sediment delivery. Class 4 is non-coherent soils with high sediment delivery potential.
FOG-landform element	Identifies the landform element attributes that were used to subdivide the soil landscape into facets, e.g., plain, hillslope, fan, ox-bow.
FOG-site morphology	Identifies the site morphology attributes that were used to subdivide the soil landscape into facets, e.g., upper slope, midslope, open-depression.
FOG-landform pattern	Identifies the landform pattern attributes that were used to subdivide the soil landscape into facets, e.g., alluvial plain, terrace, plateau.
FOG-relative elevation	Identifies the relative elevation attributes that were used to subdivide the soil landscape into facets, e.g., lowest, middle, highest.
FOG-site exposure	Identifies the site exposure attributes that were used to subdivide the soil landscape into facets, e.g., sheltered, moderately exposed, highly exposed.
FOG-substrate lithology	Identifies the substrate lithology attributes that were used to subdivide the soil landscape into facets, e.g., alluvial, basalt, sand, granite.
FOG-remote sensing pattern	Identifies the remote sensing attributes that were used to subdivide the soil landscape into facets, e.g., radiometric signature like high K.
FOG-other	Identifies the other attributes that were used to subdivide the soil landscape into facets, e.g., soil type.

The database was populated for all existing map units in the study area and tested against a preliminary map. A map unit string was used to link the database to the GIS map product.

There are over 500 different soil landscape map units and over 1,100 facets in the final database.

5.1.5 Build WRA Soil Data Card for SALIS with appropriate fields for western soil conditions

Soil conditions in the western part of NSW differ considerably from prevailing conditions in higher rainfall areas of the state. A two-page Western Regional Assessment Soil Data Card was designed with input from experienced soil surveyors that contained fields of value to the survey:

- Reason for profile selection: e.g., random; bulked sample; opportunistic; curiosity/unusual soil; boundary checking; pre-planned; transect/catena; grid; type profile. Type profiles are considered to be representative of the soil within a facet.
- Segregation types: carbonate and gypsum added to the existing set.
- Salinity test results.
- Crumb test (for dispersion and slaking).
- Testing for pH as alkalinity and acidity can vary throughout the area.

5,000 of these cards were printed. Delays were experienced in having SALIS programmed to accept the new fields on the cards because of other corporate priorities.

Many casual observations were also made and recorded on field sheets and in notebooks. Table 8 shows the field attributes recorded at each site on the WRA observation cards.

Table 8: Site and soil attributes recorded on WRA Soil Data Cards

Parameter field	Attribute
Landform	Site morphology Landform pattern Site disturbance Land use (site and general area) Erosion type, its severity and present condition
Topography	Slope gradient Aspect Microrelief
Lithology	Solum parent material Substrate Geology map code
Soil	A-horizon Solum Depth to impeding layer Estimated rooting depth Layer colour Layer soil texture Layer soil structure Layer grade of structure Fragment amount Segregation type and amount Soil Ph Crumb, AgNO3, HCl and field dilatancy tests Erosion hazard Surface salinity Layer permeability Surface condition (wet and dry) Australian Soil Classification Great Soil Group Classification Soil map code

Parameter field	Attribute
	Nature of exposure
Vegetation	Vegetation species
Hydrology	Soil permeability Profile drainage Run-on/runoff Presence of free water
Other	Easting and northing coordinates Mapsheet Type of profile assessment Type of sample taken

Soil and land parameter descriptions follow the guidelines of McDonald *et al.* (1990). Descriptions of the parameters used are included in Milford *et al.*(2001).

5.1.6 Soil profile attribute data entry to SALIS

Completed Soil Data Cards were collated and optically scanned into SALIS. Algorithms (developed for North-East CRA and used in the Southern CRA) were used to assess specific soil attributes (soil fertility, soil depth, effective rooting depth, drainage and estimated plant available waterholding capacity) from existing soil profiles, collected soil profiles and information in existing soil landscape reports (see Appendix 1 for further details).

5.1.7 Assemble predictor surfaces and prepare predictor surfaces for proto-soil landscape modelling

The base data for the project consisted of DEMs and derived attributes, airborne gamma radiometrics, interpolated climate surfaces, existing lithology coverage and existing soil landscape maps at scales of 1:100,000 and 1:250,000. Table 9 outlines the coverage and potential uses of these datasets for digital soil landscape modelling.

The base data was used to produce the environmental parameters (predictor variables) listed in Table 10. Table 10 also lists a brief description of each parameter, the source and method (usually the software) used to derive it and where appropriate, a reference to a more detailed discussion on the nature and derivation of the parameter.

Table 9: Base datasets used in the BBS study and their potential uses

Dataset (coverage)	Uses	References
Gamma Radiometrics (Missing from north of study area and from Warrumbungles)	Signal taken from dry soils to depths of up to 50 cm. Discrimination of soil parent materials and interpretation of landforming processes, interpretation of soil depth and veneers of aeolian materials	Wilford *et al.* 1997; Pickup and Marks 2000.
Digital Elevation Models (9 sec DEM complete; 25 m DEM complete except north-western portion of Narrabri 1:250,000 sheet)	Derivation of terrain attributes as predictors of soil landscape class; important predictor for drainage and depositional and erosional processes	Moore *et al.* 1993; Wilson and Gallant 1996; Wilson and Gallant 2000.
Interpolated Climate Surfaces (see Table 10) (Complete but accuracy of interpolations between surfaces is dependant on sparse datasets)	Climate as a predictor of biological activity, soil moisture regime and hence soil type; fire regime influenced by climate influences soil parameters	Hutchinson 1991; McKenzie and Ryan 1999.
Lithology (Complete but composed of tiles of maps of different styles and purposes - not continuous)	Major determinant of soil mineralogy	Gray and Murphy 1999; Paton *et al.* 1995.
Soil Landscapes and other Soil Maps (Limited coverage, mostly in south. Mapped at 1:100,000 and 1:250,000 scales with edge mismatches between mapping scales and different unit names - not continuous.)	Tested model of soil distribution in the landscape	Banks 1998.
Soil Profile Description Points (Limited, mostly distributed throughout areas of previous soil mapping)	Point locations of soil types	

The parameters and predictor variables obtained from these datasets were used to derive the following variables for use in the model.

Table 10: The environmental parameters (predictor variables) used for modelling and prediction of soil landscapes in the BBS

Parameter	Description (units)	Source/Method	Reference
Elevation	height above sea level (m)	DEM	
Slope	slope gradient (°)	DEM + ArcView + SLAP	Wilson *et al.* 2000
Compound topographic index (CTI)	measure of topographic control on soil moisture based on slope and downslope flow accumulation	DEM + ArcView + SLAP	Wilson *et al.* 2000
Profile curvature	measure of the downslope rate of change (slope/gradient)	DEM + ArcView + SLAP	Wilson *et al.* 2000
Tangential curvature	measure of topographic convergence and divergence	DEM + ArcView + SLAP	Wilson *et al.* 2000
Relative elevation index (REI)	local (300m radius) measure of position in the landscape	DEM + ArcView + SLAP	Goldrick (this study)
Temperature (mean annual)	°C	DEM + ANUCLIM	Houlder *et al.* 1999
Precipitation (mean annual)	Mm	DEM + ANUCLIM	Houlder *et al.* 1999
Moisture index (mean annual)	measure of the balance between precipitation and evaporation	DEM + ANUCLIM	Houlder *et al.* 1999
Moisture index (coefficient of variation)	measure of the seasonality of the moisture index	DEM + ANUCLIM	DEM + ANUCLIM
Radiation index	relative index of the total annual amount of radiation modified for slope and aspect	DEM + ANUCLIM + ArcView + SLAP	
Potassium gamma ray count			Wilford *et al.* 1997

Parameter	Description (units)	Source/Method	Reference
Thorium gamma ray count			Wilford *et al.* 1997
Uranium gamma ray count			Wilford *et al.* 1997
Parent material lithology	reclassification of geological units	DMR Geology	Gray *et al.* 1999

5.1.8 Predictor variable preparation

All base data and predictor variables were converted to ArcView grids for use in Spatial Analyst. Many datasets were not available in this form. Also, many datasets needed to be joined into single themes as they consisted of several tiles. Grids were then reprojected to the BBS-WRA standard (Australian Map Grid Zone 55 with a central meridian located at 147 Transverse Mercator Australian Spheroid scale factor 0.9996, False Easting 500000 and False Northing 10000000). These routines were performed using SLAP. The resultant surfaces consisted of grids with 25 m cells.

Digital climate drapes for all available climatic variables were prepared using the ESOCLIM 5.0 software. These were modified according to the effect of terrain as derived from DEMs.

5.1.9 Initial data modelling

To "train" the soil landscape model, 25,000 sample points were randomly generated for the area covered by existing 1:100 000 scale soil landscape data (Figure 5).

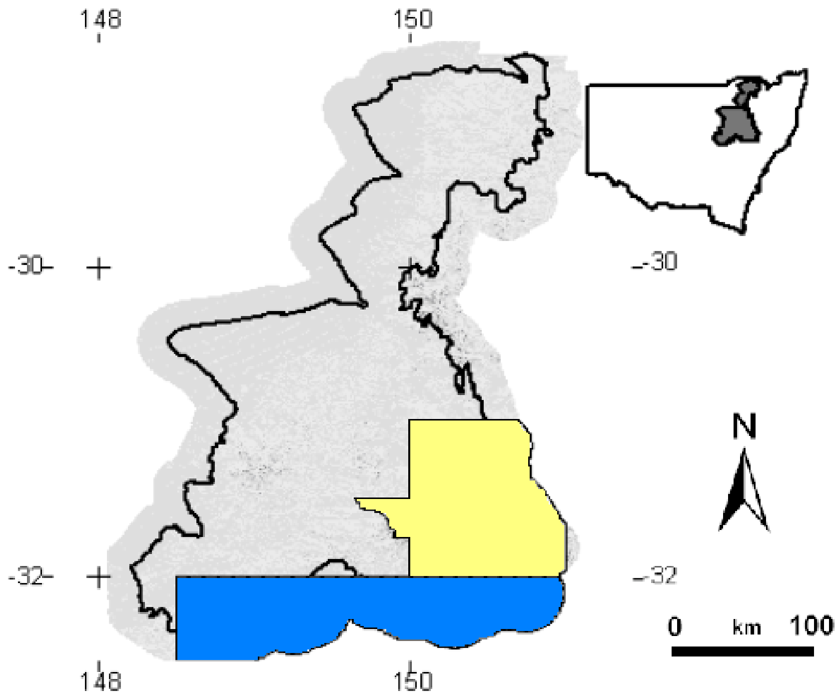


Figure 5: Extent of existing 1:100,000 (yellow) and 1:250,000 (blue) soil landscape mapping in the BBS

The 1:250 000 scale soil landscape data was excluded for this preliminary analysis due to incompatibilities in the soil landscape classes, but these are being resolved for future analyses. Use of 25,000 sample points is arguably excessive because if meaningful relationships do exist between soil landscapes and environmental parameters, they should become evident at much lower sampling densities. However, it was considered preferable to over-sample the training data and then reduce tree complexity (by the methods described below), rather than run the risk of under-sampling and missing important relationships.

A training data matrix was constructed from the sample points with soil landscape class as the dependent (y) variable and the environmental variables in Table 10 as the predictor (x) variables. Because radiometric coverage was incomplete, two training data matrices were used, one including radiometric data but omitting lithology, the other including lithology but omitting radiometric data.

These two datasets were analysed using RPART to create a classification tree (Figure 6). A classification tree is a method of hierarchically splitting a dataset into increasingly homogenous subsets. The criterion for each split is to minimise some measure of the impurity of the resultant classes, in this case the Gini index (see Therneau & Atkinson 1997; and Venables & Ripley 1999 for details).

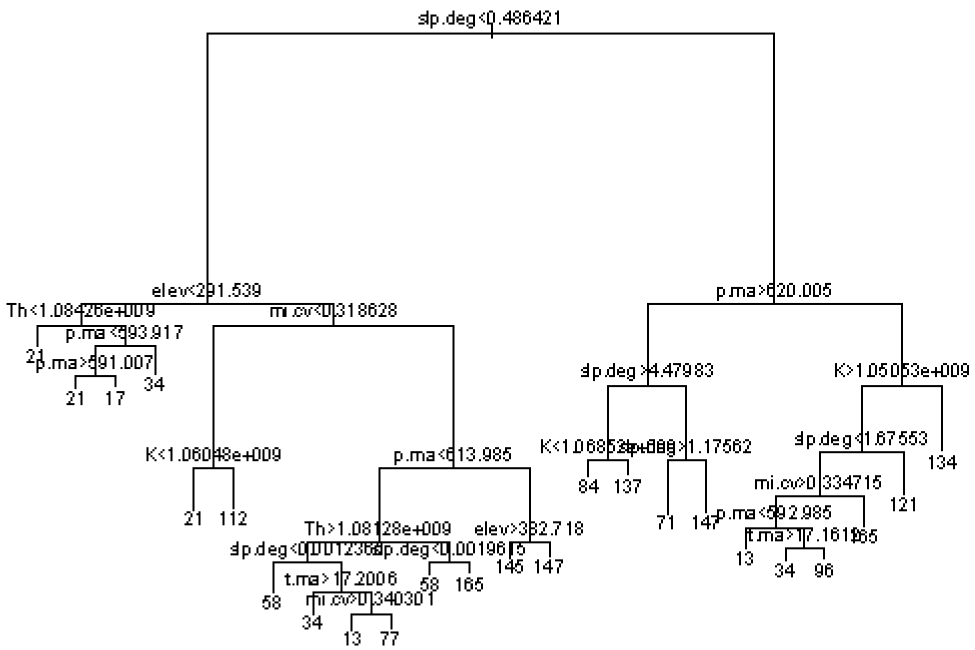


Figure 6: A partial classification tree (upper nodes and branches) from the RPART analyses of the training data including radiometrics for the BBS indicating the decision criteria

Note: the first modelled and pruned tree has 451 nodes and is too complex to illustrate.

Two important controls on RPART are "minsplit" and "cp". The former determines the minimum number of observations in a node for which the routine will try to compute a split. The latter is a measure of tree complexity (Venables & Ripley 1999) with values ranging from 0 to 1 and lower values indicating greater complexity. For the BBS datasets, "cp" was initially set to 0.000001 and "minsplit" set to 2 to allow a very complex tree to be generated.

5.1.10 Cross-validation

The classification trees were constructed with 10-fold cross-validation. Cross-validation involves splitting the training data into roughly equal-sized subsets and growing the tree on all but one of these subsets, and testing it with the unused portion of the dataset (Venables & Ripley 1999). In this case, the training data was split into 10 subsets, then 9 were used to grow the tree and the tenth used to test it. This was repeated 10 times (for each combination of subsets) and the results averaged to produce the final tree.

5.1.10.1 Tree 'pruning' to reduce complexity

A tree with few "leaves" (terminal nodes) is likely to "under fit" the data— i.e., it is unlikely to adequately reflect the complexity of the relationship between the predictor variables and the dependent variable. On the other hand, a tree with very many leaves may "over fit" the data— i.e., it reflects noise within the data rather than meaningful relationships. Cross-validation provides a measure for "pruning" the tree in order to reduce its complexity. As the size of a tree increases, the cross-validation error tends to decrease until it reaches a minimum. After this point, the cross-validation error may remain steady or even increase with increasing tree size (Figure 7).

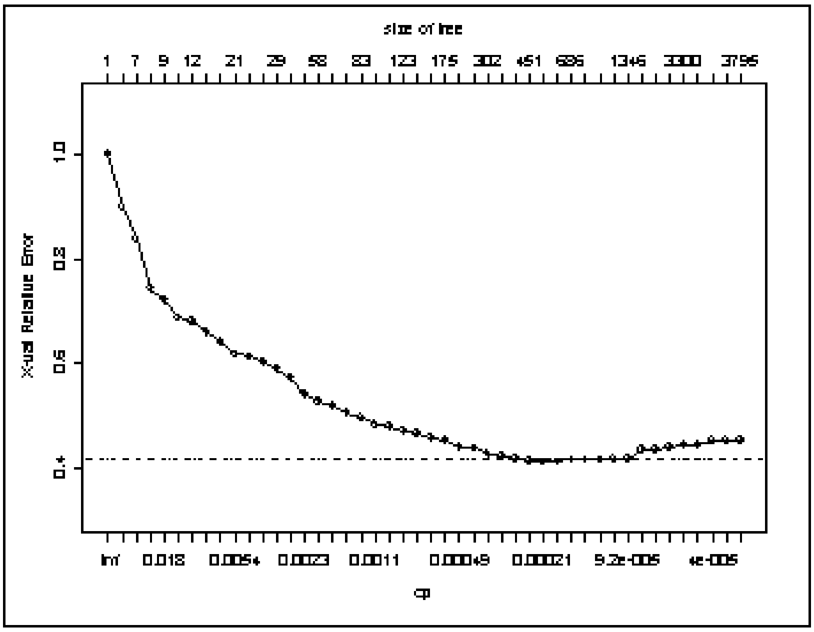


Figure 7: The relationship between tree complexity and cross-validation error for the classification trees of the BBS based on radiometrics

In other words, beyond this minimum increasing, the size of the tree does not increase the information about the relationship between the predictor and dependent variables. So this minimum provides an objective basis for tree pruning. Venables and Ripley (1999) suggest applying the 1-SE rule— i.e., to select the smallest tree with a cross-validation error within one standard error of the minimum cross-validation error. For example, the unpruned tree for the BBS training matrix with radiometrics had 3,794 leaves, but a plot of tree size versus cross-validation error indicates that if the 1-SE rule is applied, the tree may be pruned to 451 leaves with no meaningful loss of information (Figure 7). At this level of complexity (cp = 0.0002), the misclassification rate is about 26 per cent.

For the training matrix with lithology, the initial tree had 6,958 leaves, but this was pruned to 895 leaves with cp = 0.0001 and a resultant misclassification rate of 28 per cent. It is important to re-emphasise that a "perfect" fit could be achieved by setting cp = 0, but this would merely result in an exceedingly complex tree that reflects noise (inherent in any natural database), such as soil landscapes. The appropriate pruning allows description of meaningful relationships between the predictor variables and the soil landscapes.

The pruned trees were then translated into an Avenue script for generating a grid of predicted soil landscape classes. The two grids were combined such that the values from the radiometric tree took precedence over values from the lithology tree.

5.1.11 Smoothing of the output prediction surfaces to become maps

The combined grid was simplified by passing a modal filter over the grid and then removing regions less than 80 cells in size— about 20 ha, the minimum mapping area specified by Reid (1988). In the outwash areas of subdued topography and large-scale fan influences, the minimum mapping areas were increased to 40 ha (Reid 1988) by absorbing them into adjacent regions. The initial grid of predicted soil landscape class for the whole BBS is shown in Figure 8 on the following page.

5.1.12 Initial model evaluation

The grid model of the Curlewis 1:100,000 soil landscapes was smoothed to minimum mapping area of 20 ha (i.e., eighty 25 m *x* 25 m cells) and compared with Curlewis original soil landscapes. The model compared favourably with the original mapping (see Figure 8b).

The model was also compared with other portions of the BBS area against local knowledge. The model outputs did not predict well where other landscape and soil-forming processes become dominant and where the predictor variables are not sufficiently sensitive.

It was decided that a number of small training areas would be established to guide the model in areas that did not have representative soil mapping.

5.2 TRAINING AREA PHASE

5.2.1 Compilation of independent checking dataset

An independent soil data checking set of around 130 locations for the study area was considered to be sufficient, being around 10 per cent of the dataset expected to be collected as a result of fieldwork from the project (Reid 1988). The points sampled were confined to areas without existing (and checked) soil landscapes as these had already been thoroughly field checked.

The check sampling points were mostly restricted to public roads and other accessible areas.

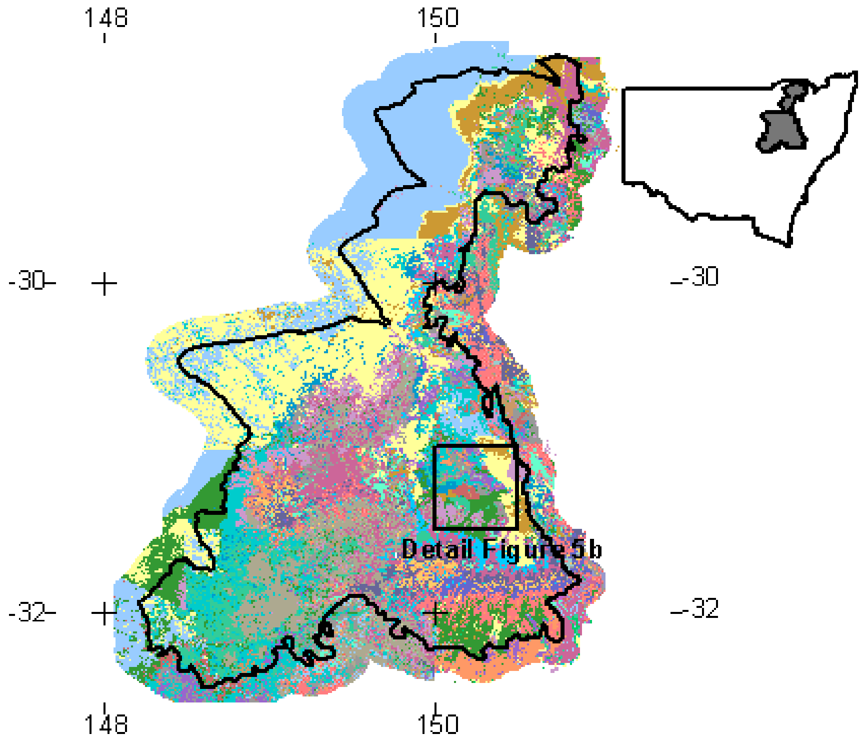
Sampling locations were initially stratified according to broad geological types and landform patterns being evenly distributed across 1:100 000 map sheets. It was later decided to complete the independent sampling set at the end of the main mapping phase at locations where both flora and fauna had been sampled.

The soil profile information from the independent check dataset was not used during the main mapping phase.

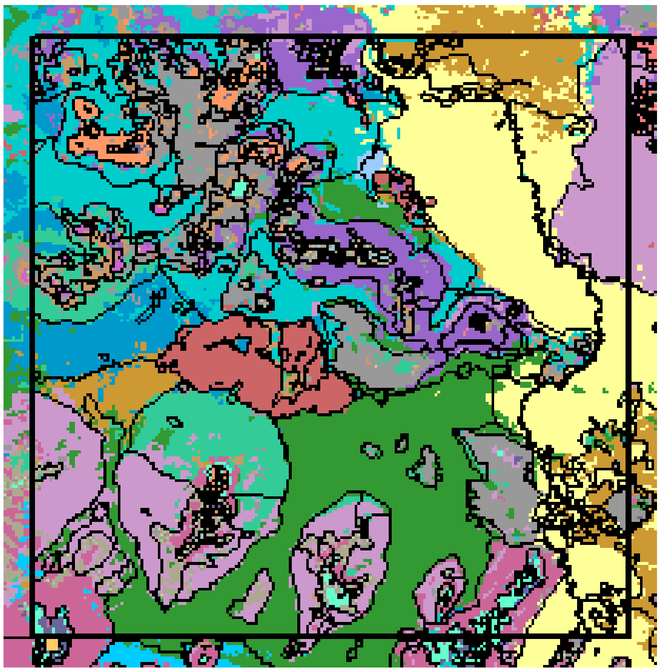
5.2.2 Staff briefing and work allocation

A meeting was held in May in Gunnedah where the project background, methodology, requirements and database were discussed. File name protocols and database version control rules were discussed. Work areas were allocated and training areas were selected.

Figure 8: Grid of predicted soil landscapes as initially modelled for the south-eastern portion of the BBS (a) and detail of a section with overlay of the training data line work (b)



(a)



(b)

The above analysis is based on soil landscape mapping on the Curlewis 1:100 000 map sheet (Banks 1995). Blocks of colour represent the modelled soil landscape outputs, whereas the black lines represent the original soil landscape mapping.

5.2.3 Training area selection

To overcome the problem of existing soil landscapes representing the rest of the study area, additional mapping using air photo interpretation and field observation for several small areas scattered throughout the BBS was required. The aim was to create training datasets with a broad coverage of various environmental conditions of the BBS and then interpolate between these training datasets, rather than extrapolate models outwards from a single training set.

Nine training areas were selected on the basis of geomorphic and radiometric patterning in data sparse areas. Each training area was located to maximise the greatest amount of topographic and parent material variation in a reasonably small and accessible area. Map 2 shows the locations of the training areas.

1:100 000-scale soil landscape mapping already in progress was accelerated. In particular, line work was completed and scanned to ensure that these preliminary maps could be provided and also used for training. It was realised that the 1:250 000 soil landscape mapping would confound the model because of the conflicting scale and potential overlap between the two scales of mapping in intermediate areas. Consequently the 1:250 000 scale mapping was not used for training until later in the project.

5.2.4 Training area pre-field preparations

Literature review and reconnaissance of each training area was undertaken. Base maps were prepared showing context information, navigational data and, where possible, radiometrics and topographic features. Each area was examined using radiometrics and air photos in stereoscopic pairs. Preliminary soil landscapes were delineated.

A copy of the database populated with existing soil landscapes was provided to each soil surveyor. To keep the model as simple as possible and, more importantly, to prevent map unit decision conflicts in the model, instructions were issued to ensure that no new units were to be produced that duplicated previous mapping.

5.2.5 Training area fieldwork

The soils and associated physiographic information for each facet within each soil landscape in each training area was observed and recorded on WRA Soil Data Cards.

5.2.6 Training area digitising and database entry

Training area field sheets were traced and despatched for scanning. The resulting digital soil maps were edited and despatched for use in the model.

Complete descriptions for all soil landscapes were entered into the MS Access database. Some communication problems and iterative changes to training area line work delayed this process.

5.3 MAIN FIELDWORK AND MODELLING PHASE

Expert knowledge can be incorporated into the tree-based soil landscape models at several stages, including the initial choice and derivation of predictor variables, pre-stratification, and assessment and modification of the decision rules produced by the model.

The success of the model was primarily dependent on the choice of appropriate predictor variables and the high-quality datasets that describe them. The range of possible predictor variables is very large, and expert knowledge was required to determine which of these would be the better predictors of soil landscape class in particular areas.

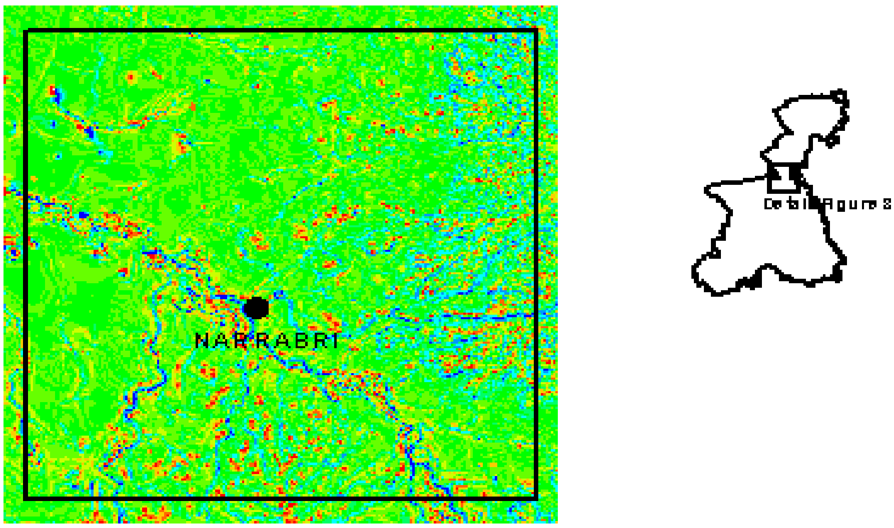


Figure 9: Detail of area around Narrabri illustrating the usefulness of the REI for identifying levees (red) and depressions (blue)

It was also possible to manipulate base environmental data to develop new predictor variables. For example, in the broad alluvial plains of the BBS it is known that the soils of elevated areas such as levees and dunes are different from the soils of the surrounding plain (Banks, R. 2002, pers. comm). However, the difference in elevation that defines such features is often subtle and hard to identify reliably from raw elevation data. This association led to the development of the relative elevation index (REI), which was designed specifically to identify such features. REI values were found to correlate well with the existence of levees and depressions (Banks, R. 2002, pers. comm.; Figure 9).

Another example of the use of expert knowledge to manipulate base datasets is the reclassification of geology into lithological parent material. This is important because geological classes may be differentiated on a variety of characteristics unrelated to their role as soil parent material. Reclassification of the geological database allowed the number of classes to be reduced from 409 to 24 in a way that emphasised this role. It was later found, in the West Pilliga, that lithology was insufficient to discriminate between some soils. As a result of field checking and soil surveyor feedback, the model was altered to include geology tags in these areas.

5.3.1 Post-training area modelling and model iterations

The model was run using the training datasets. Because of the relative size of the new training datasets compared to the original soil landscape trained areas, the model remained dominated by the weighting of parameters in the Liverpool Plains and Liverpool Ranges. For example, climate parameters that were important soil landscape discriminators in the Liverpool Ranges were not appropriate for this purpose in the topographically subdued Pilliga outwash.

5.3.2 Modelling by province and weighting of modelling within provinces

The decision was made to use the provinces of Morgan and Terry (1992) as separate modelling areas. The model was run separately for each area and adjustments made on the advice of soil surveyors. Soil surveyors spent at least one day providing input to their modelled provinces. A brief synopsis of each province is listed below.

5.3.2.1 Liverpool Ranges

The model worked well here due to excellent training areas, a single geological province (basalt and associated lithologies) and strong relief and terrain, which allowed DEM parameters to be used consistently.

5.3.2.2 Liverpool Plains

The southern portion modelled well initially, but prediction was poor in the northern part. Climate and elevation parameters were reduced in importance, but the model still failed to perform. Draft soil landscape mapping from the Baan Baa and Boggabri 1:100 000 map sheets was then scanned, checked and used directly.

This left a remaining small section of <500 km^2 in the north to be modelled. However, the model did not perform well on much of this area due to the strong influence of fan soil landscapes (originating in the Nandewar Ranges Interim Bioregion to the east of the BBS, for which no training areas were available). Further fieldwork was undertaken to describe these soil landscapes.

In summary, the Liverpool Plains was judged to be satisfactory, being based on soil landscape line work, except for a small portion in the extreme north-east.

5.3.2.3 Talbragar Valley

This province had been entirely mapped with 1:250 000-scale soil landscapes. After initial modelling, part of this area was checked. It was found that the initial model did not effectively discriminate between the Pilliga, Ballimore and Purlewaugh sandstones (all of which have been allocated the same lithology classification). Furthermore, the model under-weighed the influence of radiometrics. The model was subsequently refined on two occasions. Consequently the 1:250 000 line work was retained and facets determined and modelled according to estimates of their percentage area, position in the landscape, and relative elevation. This was judged to be a satisfactory outcome.

5.3.2.4 Pilliga Outwash

Soil variation in the training area failed to relate to any discernible environmental predictor parameter. It was found that the mapping of Hesse (2000) was the best predictor of soil distribution, especially on the alluvial systems, while the hillslope systems appeared to be better predicted using soil landscape distribution rules from the hillslopes in the main Pilliga province.

However, Hesse's map did not cover the entire study area and, appeared to ignore possible geomorphically related soil types (which were more prevalent to the north of the Hesse study area) when compared to gross radiometric features.

As a consequence, it was decided to:

- Alter the shape of the Pilliga Outwash province to include north-running alluvial drainage systems from the Pilliga province and to exclude hillslopes in the Pilliga Outwash (including them instead in the Pilliga province); and
- Use the Hesse line work where it existed, and run a six-class unsupervised radiometrics classification on the remaining areas. The six-class classification provided a strong geomorphic structure and related well to the current and previous alluvial systems. Subsequent fieldwork showed that the soil variation in all but one of the radiometrics classes was similar.

In summary, the Pilliga Outwash extended the mapping of Hesse (2000), augmented by further alluvial classifications from a radiometrics pattern associated with more recent alluvium.

5.3.2.5 Pilliga

The Pilliga province was subdivided into north-eastern, south-eastern and western sections. This was done for practical reasons as this large province contained diverse areas such as the Warrumbungles, which did not have radiometrics coverage and exhibited strong topographic, climatic and parent material differences in comparison with the rest of the Pilliga province. Also, mapping from the east, such as Tambar Springs, could have conflicted with mapping in the south.

In the north-eastern Pilliga, modelling was based on training areas and soil landscapes from the Tambar Springs sheet.

In the south-east Pilliga, modelling was based on training areas.

In the western Pilliga, modelling was based on training areas for portions remote from the Warrumbungles volcanics. The training areas were adjusted on several occasions and more emphasis was placed on radiometrics and topography than on climate variables.

In flatter areas in the extreme south-west, training based on the 1:250 000-scale Narromine soil landscapes and radiometrics were used.

In the Warrumbungles, a geological map from Duggan and Knutson (1993) was compared with soils and landscapes during a number of field trips. It was agreed that the Warrumbungle volcanics be subdivided on the basis of slope as this matched very well with the boundaries for geological units and Normalised Vegetation Difference Image signals. Soils were subsequently examined and described.

The west Warrumbungles was not well represented by the modelling process due to subdued topography and lack of training data. Fieldwork revealed broad and predictable patterns that were subsequently hand-drawn.

5.3.2.6 Northern Basalts

Various models were run based on serially input training areas. Once again, climatic variables were given less emphasis than topography and lithology. In the Northern Basalts province, the model mapped more intensely than for the training areas.

5.3.2.7 Northern Outwash

The mapping of this province, with its subdued topography, relied almost exclusively on radiometrics. However, it was found that much of the variation in radiometric signal proved to have no clear correspondence with soil attributes. The model "tried hard" in many areas, but succeeded in producing only a mass of meaningless polygons. In an attempt to resolve this clutter, the model outputs were smoothed beyond the 20 ha limit to create minimum mapped areas of 40 ha. This resulted in a clear and logical fan pattern that related much more closely to soil types. Smaller map units in drainage lines were hand-drawn.

The area to the north of the Moree radiometrics was not received in time to be used for this project, however, fieldwork showed very little variation from the ubiquitous grey clays of this area.

5.3.3 Site selection and gap analysis

Initially it was decided that as many soil and vegetation description/sampling sites as possible would be at shared locations. The vegetation sites would be selected using gap analysis software developed for the NPWS by Ferrier *et al.* (Ferrier, S. 2001, pers. comm). It was initially decided that one of the environmental variables used in the gap analysis would be the modelled soil map for the study area.

However, this approach proved to be impractical because:

- Original soil and vegetation sampling sites were concentrated in geographically separate parts of the study area. The under-sampled areas for vegetation were relatively over-sampled for soils and vice versa;
- The criteria for native vegetation sampling requires extant native vegetation, while the soils sampling regime does not have this requirement;
- Accessibility limitations for soil sampling (using trailer-mounted corer equipment and heavy four-wheel-drive vehicles for routine fieldwork) were generally greater than for vegetation sampling. However, it should be noted that profiles were also described from hand augers, batters and pits when cores could not be taken;
- The gap analysis software is designed for incremental input of single sampling sites to choose the next least sampled location. While 500 or more sites could be selected in a single session, the system required rigid rules for sampling, particularly at those sites. A more flexible system of sampling was needed to allow for a greater number of sites, soil disturbance (in some locations), restrictive deadlines, and to avoid the use of heavy and noisy hole-digging equipment in sensitive locations.

Discussion concerning practicalities such as location of sites with relation to pixel centres and the possibility of running Gap Analysis in the field raised the possibility of using lap-top computers in the field to sample site selection.

5.3.4 Gap analysis

To assess gaps in sampling all combinations of landscapes, the concept of environmental space is required. Environmental space is somewhat similar to geographic space. Figure 10 illustrates the case of three data points in a two-dimensional environmental space defined by elevation and precipitation.

In this example, the distance between points *A* and *B* is small relative to the distance between points *A* and *C*. The distance, *d*, between points A and B has a Euclidean distance, the square root of the sum of the squares of the distance in both dimensions:

$$d_{ep} = \sqrt{(e_A - e_B)^2 + (p_A - p_B)^2} \qquad (1)$$

For the soil landscape model, the environmental space is *n* dimensional, where *n* is the number of environmental parameters— e.g., 14 using radiometrics, lithology, geology and topographic variables. The resulting distance is based on Euclidean distance, where *d* is the square root of the sum of the squares of the distance in each of the *n* dimensions.

Unlike geographic space, each dimension of environmental space is expressed in different units. Each may vary in its importance to the soil landscape model— i.e., environmental variables may need to be normalised and weighted before calculating environmental distance.

Figure 11 illustrates one such measure, where all variables have equal weighting and are normalised by subtracting the mean and dividing by the standard variation. Figure 11 suggests that for much of the BBS, the training data does not adequately describe the environmental space.

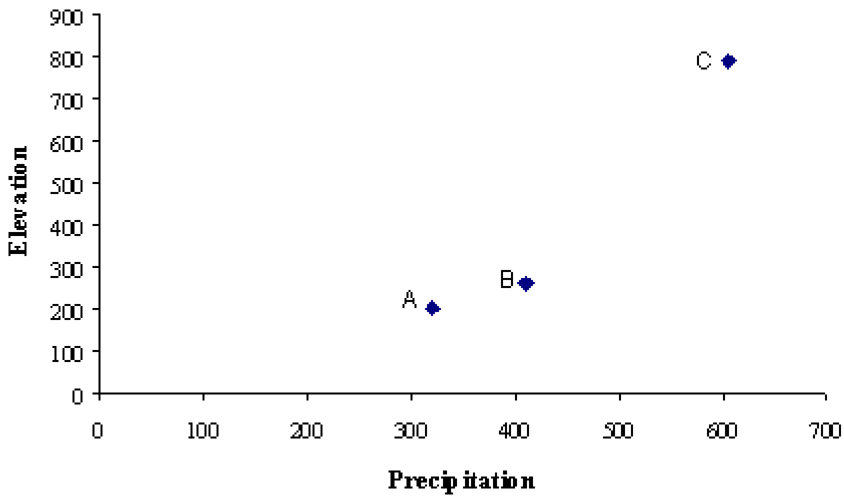


Figure 10: Three data points in the two-dimensional environmental space defined by precipitation and elevation

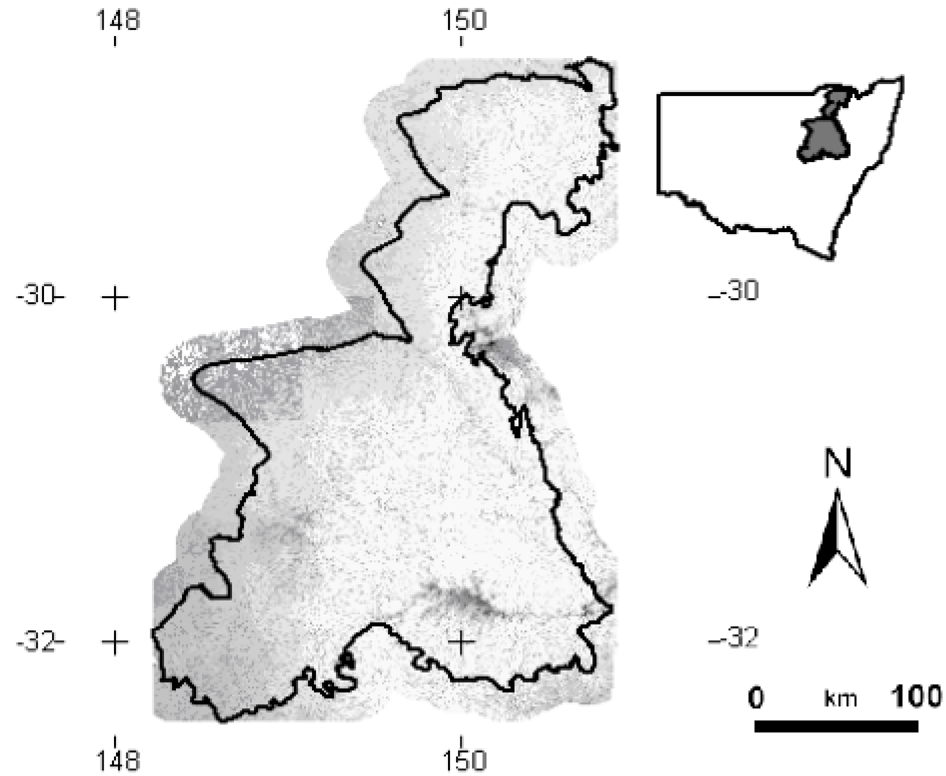


Figure 11: Estimates of the average distance in environmental space of all grid cells in the BBS from the points in the training data (excluding lithology and radiometric data)

Note: Darker tone indicates greater environmental distance from sampled environmental spaces. This forms the basis for gap analysis.

Considerable delays were encountered due to problems with joint gap analysis due to what transpired to be a minor data formatting problem that prevented input of the soils model to the gap analysis. This delayed the start of fieldwork for several weeks. In retrospect, it may have been better to start the fieldwork according to standard soil survey rule sets.

5.3.5 Flexible gap analysis

Due to time constraints, the flexible gap analysis tool was developed from foundations established by S. Ferrier (2001, pers. comm).

Soil profiles for the main fieldwork phase were selected using a portable version of the gap analysis method initially developed by Ferrier and used for vegetation mapping. The portable gap analysis software determines the locations of all existing sampling sites in the study dataset in terms of weighted environmental variable space. It then calculates the relative distance of all locations in the environmental space of the study area to all the sampled locations. A map of the area showing the degree of sampling in accordance with the environmental variables is then produced. Areas that are not well represented by the sampled dataset are then selected for sampling. As each sample is selected the portable gap analysis is run again and the map of degree of sampling changes. Essentially the flexible gap analysis tool takes the measure distance from all other sampling points in terms of environmental space (such as portrayed in Figure 11) and represents these distances on a map.

5.3.6 Main fieldwork

Main fieldwork was undertaken from October to December 2001. Extra soil survey staff were called in to accelerate fieldwork which had been delayed by data management problems.

Four trailer-mounted soil corers were used (where access was available) to excavate soil profiles for description to 1.5 metres. Each corer has a jackhammer, propelled by a hydraulic ram, that drives a 50 mm-diameter tube into the soil (see Figure 12). Once removed from the tube, each soil core is described using attributes on the WRA Soil Data Card. The corer is well suited to operations in flat accessible areas and a wide range of soils without very hard fragments.

Figure 12: A soil corer in use on the gently undulating plains of the Northern Outwash

Soil landscape details including the calculation for soil attributes (except EPAWC) were entered by the soil surveyors into the MS Access database for each map sheet area. These were e-mailed or sent by disc to the Soils Quality Officer who arranged for the calculation of EPAWC and entry into the main database.

Access to portable notebook computers with vehicle-based power supplies, and licences for ESRI ArcView 3.2 and Spatial Analyst, was organised. Brief training in the use of the SLAP flexible gap analysis module was arranged, and the system was pressed quickly into service. Many problems were encountered at first with running such complex software on relatively low-powered computer hardware. However, further development was undertaken, and by the end of the main fieldwork phase the gap analysis software was working successfully on datasets of around 1,000 km^2.

Soil profile points for around 200 locations were determined using flexible gap analysis. This would have been higher except for initial teething problems with hardware limitations and some initial software bugs.

An advantage of the use of in-field computers is that modelled outputs could be field-checked using lap-tops rather than the time consuming and more laborious approach of having draft maps printed.

5.3.7 Landholder communication protocols

The lengthy landholder consultation protocols used were not well suited to visiting up to 20 sites per day on private land. Consequently it was decided to locate as many sites as possible on publicly accessible land, with observations on private land to be undertaken only when communication protocols and assistance with making contact could be arranged. Delays were experienced in obtaining official maps showing publicly accessible land, but a travelling stock route and travelling stock reserve map was obtained from NSW Agriculture, which served as a useful guide to some accessible areas.

Soil surveyors were equipped with various BBS brochures and newsletters and other information about soil landscape mapping to aid them in explaining the aims of the project to landholders.

It was not until late September that a formal meeting was held with vegetation and soil survey field staff and the BBS Communications Officer. A protocol and supporting database to facilitate landholder contact was agreed to be set up, but no further communication was received on this matter.

Due to pressing deadlines, fieldwork went ahead despite the lack of the streamlined protocol, with soil surveyors instructed to approach landholders prior to entering private property and to use the ‘country code’.

One landholder rang to confirm that soil survey work was actually *bona fide* in the Coonabarabran area, as the soil surveyor and technical officer were working on a Sunday.

Most landholders were unaware of the BBS-WRA project, but access refusal was rare.

5.4 POST FIELDWORK

5.4.1 Final modelling

Following the main fieldwork phase, final models were built for each province. Each of the province models (except the Pilliga Outwash and Northern Outwash) was smoothed to minimum mapping areas of 80 pixels or 20 ha. This conforms with the minimum mapping area standards for 1:100 000-scale soil landscape mapping.

5.4.2 Province boundary adjustments

Province boundaries were adjusted so that alluvial, transferral and fan areas adjacent to alluvium and fans of the Northern Outwash and Pilliga Outwash were included in these provinces. Similarly, hillslopes that were included in the outwash provinces were included with hillslopes in adjacent provinces.

5.4.3 Line work adjustments

Line work was adjusted where necessary, mostly to remove vertical and horizontal boundaries caused by:

- Discontinuity in data, e.g., on the boundary between areas covered by radiometrics and areas without radiometrics;
- Discontinuities in the available geological mapping; and
- Edges of some existing soil landscape mapping and modelled areas.

Lines were relocated using radiometric and various digital elevation parameters as well as satellite imagery in consultation with soil surveyors and also, by adhering as much as possible to checked or richer datasets.

A listing of line work changes has been compiled.

5.4.4 Polygon tag changes

In some instances polygons were mis-tagged. This was a consequence of several factors, including:

- serial addition of training areas;
- modelled polygons adjacent to mapped soil landscapes;
- mis-linking of equivalent map units across province boundaries; and
- the model's occasional inability to always discriminate between whole soil landscapes and facets.

A listing of polygon tag changes has been retained.

5.4.5 Soil profile point data gaps

WRA Soil Data Cards were checked by soil surveyors and were sent to SALIS for scanning into the system. The soil profile information was later used to calculate EPAWC.

The locations of WRA Soil Data Cards entered into SALIS were superimposed on the last line work model and a count of soil profile observations by soil landscape was undertaken. This provided a basis for planning "mopping-up" fieldwork to address data gaps.

5.4.6 Soil landscape database gaps

A listing of map unit tags was compared with existing database map unit tags. Duplicate database entries and missing map unit descriptions in the database were addressed. A list was made of missing data at facet level for the main predictive surfaces.

5.4.7 "Mopping up" fieldwork

Of the >500 map units and >1,100 facets for the project, it was found that facets lacked soil profile descriptions. In most cases, soil type was known and confident estimates of important parameters were made. At the time of writing, "mopping up" fieldwork had been completed for accessible areas. This process was delayed due to staff leave, wet weather and vehicular breakdown. When this soil profile data becomes available, it will be added to the database and a further edition issued.

5.4.8 Linkage of soil landscape polygons and database

Polygon types were listed against and matched with map unit descriptions. This required correlation and unravelling of map units that were not updated during the modelling process. This was initially done on a province-by-province basis.

The map string from the database was then linked to each polygon according to its tag.

Once this process was completed for each province, the provinces were joined and merged into a final coverage. The map units were then linked to the database. The resulting outputs were checked for 'pinholes', overlaps and 'slivers' and were topologically corrected by a contractor using ESRI ArcINFO.

5.4.9 Attribution of soil landscape spatial summary parameters

A set of summary statistics was generated for each soil landscape based on the intersections of soil landscape polygon types with the various spatial surfaces. It included some initial modelling. Each of the following statistics is included in the database for generation of summary information and to facilitate modelling.

The mean positive and negative standard deviations are recommended for use in characterising the typical parameter range of the soil landscapes. This is because the intersect was completed on the final map, which has been smoothed for reliability at 1:100 000 scale. Use of range statistics necessarily includes some 'noise', which is a feature of natural resource datasets such as soil landscapes. In other words, small areas of included soil landscapes or mapping impurities result in outliers, which unrealistically extend the range for typical parameters.

For soil landscapes where the standard deviation around the mean is small, it is likely that the soil landscape may have been delineated according to that particular environmental attribute or another spatially correlated variable. Where there is a broad range in standard deviation around the mean, this can indicate that variation in the soil landscape is related to this or another spatially correlated variable. The soil landscape may have had a broad range due to inherent complexity. Alternatively, the extension of the area by modelling may have included areas that may have been better assigned to a different or new landscape.

The summary statistics include: Maximum, Minimum, Range, Mean, Standard Deviation, Standard Deviation plus Mean, Standard Deviation minus Mean.

5.4.9.1 Terrain attributes

- Slope: in degrees, multiplied by 100 and converted to integer.
- Relief: elevation difference within 300 m radius of each cell.
- Relative Elevation Index: pixel elevation less the minimum elevation within 300 m divided by maximum elevation minus minimum elevation (procedure developed by Dr Goldrick).

5.4.9.2 Temperature attributes

- Annual Mean Temperature ANUCLIM 5.0.
- Temperature Seasonality: coefficient of variation of weekly mean temperature. ANUCLIM 5.0.

5.4.10 Database attribution with major variables and calculation of key soil attributes

EPAWC was determined from type profiles (i.e., soil profiles that are considered to be representative of the main soil type of facets) and estimated using calculations and data in SALIS as described in the Method section of this report.

Analysis of gaps in the database and checking of outlier values was also completed.

5.4.11 Map and database checking

A number of methods of model checking were explored.

5.4.11.1 Visual comparison of the model with its source maps

The key soil attributes were linked to ArcView 3.2 and evaluated by various staff involved in the project.

A further planned stage is to have the various attribute surfaces scrutinised and verified by other soil scientists, particularly those familiar with the mapped areas. Initial scrutiny suggests that the surfaces are of sufficient quality for reconnaissance soil landscape mapping in and around the area of the training data (Banks, R., Senior Soil Surveyor, DLWC Gunnedah, 2002, pers. comm.). Figure 8 shows the comparison of initial model outputs with existing soil landscapes.

A visual comparison of the predicted and training maps for the BBS appears good and adjustments were made after examining discrepancies.

Visual examination, however, is not conclusive. Even a poor quality predicted map— e.g., one that has only about 50% agreement with the training map— can look generally acceptable, but even a very good model may produce a predicted map that differs from the training map when examined in detail.

Maps were produced and checked visually against expected outcomes. This resulted in a list of queries that were then checked against the database for clerical errors. Unsolvable discrepancies were directed to the soil surveyors for solution. In some cases, minor fieldwork was required to solve such discrepancies. Further fieldwork is required to solve all known discrepancies.

5.4.11.2 Logical consistency checking

A technical check for logical consistency of the datasets— e.g., production of pivot tables to check consistent features such as well-drained Solodized Solonetz and Lithosols with high EPAWC values— has begun. A thorough check will be undertaken by DLWC as part of a planned SALIS data audit. Discrepancies to date have been few and easily fixed by telephone queries.

5.4.11.3 Misclassification checking

Analysis in areas of existing soil landscape mapping in the Liverpool Plains and Ranges allowed the initial predicted surface soil landscapes to be compared with existing soil landscape maps of the Curlewis and Blackville 1:100 000 map sheets (Banks 1995;1998). In the case of BBS regression trees, the misclassification rates were 26 and 29 per cent for the radiometric and lithological trees, respectively. Given the 'noise' inherent in any natural dataset, and the expected variations within the soil landscapes (see below), this result seems acceptable. It should be noted that similar rates of misclassification are typical for land resource assessment and that soil landscape maps are typically found to predict most physical soil attributes in around 17 cases out of 20 (Dewar 1996).

There are several reasons to expect such discrepancies between the modelled and pre-existing maps, even in the best of circumstances. In most cases, a soil surveyor and a classification tree will predict the conceptual basis for each soil landscape in slightly different terms. For example, a soil surveyor might define a soil landscape with granitic parent material, steep to very steep slopes and low mean annual rainfall. The same soil landscape may be defined by the classification tree as those areas with a certain radiometric signal, slopes greater than 22° and mean annual precipitation below 587 mm. While these differences may seem trivial, they will lead to variations in the final product, most notably in the placement of boundaries.

Further, in delineating the boundaries and extent of this soil landscape, the classification tree follows precise, objective rules. The soil surveyor, on the other hand, is more likely to delineate the soil boundary through the subjective assessment of information from API, geology maps, topographic maps and climate maps. Moreover, it is often the case that two experienced soil surveyors will make slightly different assessments of the data and thus delineate the boundaries differently. This is not to say that the boundaries resulting from the classification tree are necessarily more accurate or better than those drawn by surveyors, but they are more precise and objective. On the other hand, it is also possible that where discrepancies exist, that the model *has* more accurately characterised the nature of a soil landscape than the surveyor. An advantage of the modelling approach is that the method is both quantifiable and reproducible.

In two areas on the Curlewis 1:100 000 soil landscape map sheet, the model appeared to predict soil landscape boundaries better than the original mapping. This was mostly the result of using radiometric data that was not available at the time of the original survey.

The predictive maps presented here are an evaluation of the technique based on presently available data in areas that are challenging for environmental modelling, especially where environmental surfaces are imperfect or of insufficient resolution or incomplete.

5.4.11.4 Evaluation against the independent dataset

Given expected differences between the training and predicted maps, it seems better to test the predictions through field observation and air photo interpretation. The question to be asked for each observation is— "are the characteristics of the site consistent with the predicted soil landscape?"

In evaluating this question, it needs to be recognised that within each soil landscape, there will be variations in most of its characteristics. For example, a single soil type may occur in a number of soil landscapes as either dominant or minor facets. The test of prediction must therefore be— "is it reasonable that the observed soil type should occur in this predicted soil landscape?", not— "is the observed soil the dominant type for this predicted soil landscape?" The latter criterion, used in a similar study by Bui *et al*. (1999), is too restrictive given typical heterogeneity of soil landscapes.

The independently gathered soil profile dataset is being gathered at various fauna and flora survey sites. This process has been delayed to allow completion of the main map and database delivery.

5.4.11.5 Evaluation of training area representativeness

No matter how well a model predicts soil classes in the vicinity of the training data, one can expect that the accuracy of the model predictions will decrease in environments dissimilar to that in which it was trained. This raises the problem of defining the degree of dissimilarity between the environment of the training area and the environment of the areas that were modelled. Some areas were modelled in much more detail than was warranted, and some tagged inaccurately. While field-checking has attempted to assign such areas to correct landscapes, not all can be checked off, so best estimates had to be made.

In future projects, the flexible gap analysis tool could be potentially used to model areas that are least similar to already mapped areas. In the current study, this opportunity was missed because gap analysis technology was not available at the time of training area selection.

5.4.12 Allocation of soil landscapes to 15 km buffer

DLWC is currently completing linework and database entry of soil landscapes for the 15 km buffer for NPWS. This buffer was dropped from the original proposal due to budget cuts to the project.

The 15 km buffer is being mapped and described independently from the main project using existing soil landscape information (where available) and on-screen digitising of estimated soil landscape boundaries using TM, geological, radiometric and contour information. Both soil landscape boundaries and information entered in the database, for new mapping areas, are estimates made by the soil surveyor and have not been field checked. ***As such, caution should be taken when using the coverages or soils information for these areas***.

5.4.13 Design of Web products

Preliminary design of a series of Internet-served maps of this study and their associated map unit descriptions are planned, but cannot be carried out efficiently at time of writing. This is based on a World Wide Web environment where the user can select various views of the database along with a number of checked maps. This process will take advantage of SALIS's developing ability to store spatial information and allow it to be viewed using an Internet spatial viewer. At present, this is awaiting approval for development of SALIS to accept the fields used in the MS Access database. As an organisation, DLWC is committed to producing interactive web products. It is expected that such a system will be in place early next financial year.

5.4.14 Access to information

This document is accompanied by a compact disk, which contains the MS Access map unit database, and shape file of the final mapped output that can be used in ESRI and other GIS platforms.

Distribution is provided on the basis that the data is not to be handed to third parties beyond the BBS-WRA project without written permission and appropriate acknowledgment.

For further copies of the CD or the dataset in other formats, please request via e-mail to the following address— soils@dlwc.nsw.gov.au.

Requests for access to publicly available soil profile information in SALIS should also be made via e-mail to the same address— soils@dlwc.nsw.gov.au.

5.4.15 Feedback

Feedback on this report or any aspect of the project is welcome. Such feedback should be sent to the following e-mail address— soils@dlwc.nsw.gov.au.

6 OUTPUTS

6.1 SOIL PROFILE INFORMATION

Soil profiles descriptions were collected or collated for the BBS area. Maps 1 and 2 show the distribution of soil profile sites from both newly mapped areas and existing soil landscape coverage. The data is held in the NSW Soil And Land Information System (SALIS) and is accessible by e-mail request to soils@dlwc.nsw.gov.au.

Once the final results of the project are made public, the soil point information will also be made available via DLWC's Soil Profile Attribute Data Environment (SPADE) spatial viewer on the Internet at the following Web address: http://spade.dlwc.nsw.gov.au.

6.2 MAP COVERAGES AND DATABASE

6.2.1 Soil landscape coverage

A complete, seamless coverage of soil landscapes for the entire BBS area (Map 5) has been completed. Over 500 soil landscapes have been compiled both from existing draft and published soil landscapes and from reconnaissance level 1:100 000 scale soil landscape mapping. Furthermore, over 1,100 facets with soil attribute data were described, which can be linked with DEMs to produce enhanced 1:25 000-scale coverage if required. This coverage and the matching database have been distributed to major stakeholders for modelling purposes.

6.2.2 Other coverages

Map coverages and surfaces used as part of the modelling program (see Tables 9 and 10) include Climatic Surfaces (Maps 6a through d), DEMs (Maps 7a through f), and derived Great Soil Group, GSG (Map 9). These were provided to Planning NSW staff and NPWS vegetation modellers in February 2002.

6.3 SOIL ATTRIBUTE THEMES

Examples of current soil attribute themes for Fertility (Map 11), Soil Profile Depth (Map 12), Effective Rooting Depth (Map 13), Estimated Plant Available Waterholding Capacity (Map 14), Drainage (Map 15) and Soil Regolith Stability (Map 10) for dominant facets have been generated. These themes were compiled from the soil landscape coverage and information in the map unit database and are displayed according to the dominant facet for each soil landscape. They are likely to be revised as the modelling process evolves.

6.4 SOIL LANDSCAPE DATABASE

An MS Access database was produced that detailed descriptions for each soil landscape in the BBS. It includes information on the main soil attributes useful for vegetation modelling (e.g., fertility, EPAWC, ERD and drainage). Other soil and landscape information (e.g., soil type, soil and landscape limitations) was also recorded for each soil landscape, making the data useful for many purposes other than vegetation modelling.

6.5 USE OF DATA

These maps and the associated database provide a guide to the distribution and assessment of soil landscape attributes across the BBS. The maps should be used only at 1:100 000 scale or smaller and should not be used for any other purpose than broad regional vegetation modelling purposes. Enhanced resolution of these soil attributes can be gained through linkage with digital elevation models. This will be undertaken by NPWS as required for modelling purposes. The soil attribute themes generated will assist in the modelling of:

- Biodiversity assessment;
- Potential and current forest community modelling;
- Fauna modelling;
- Rare flora species modelling;
- Centres of endemism;
- Response to disturbance;
- Plantation potential on cleared private land; and
- Industry development opportunities (e.g., intensification).

6.6 METADATA

Individual metadata statements conforming to the ANZLIC Page 0 standard have been prepared for the project, its associated MS Access database, and all maps in this report. These statements have been submitted for inclusion in the NSW Natural Resource Data Directory (available on the Internet athttp://www.canri.nsw.gov.au/nrdd).

6.7 OVERLAYS

A transparent overlay containing road and town information has been provided with the appendices. Issues of scale mean that this overlay only matches with maps 5 and 9 through 16.

7 DISCUSSION

7.1 ARE THE DATA AND CONCLUSIONS ROBUST AND RELIABLE?

No matter how well a model predicts soil classes in the vicinity of the training data, the accuracy of the model predictions is likely to decrease in environments dissimilar to that in which it was trained. This raises the problem of defining the degree of dissimilarity between the environment of the training area and the environment of the areas that were modelled. Some areas were modelled in much more detail than warranted, and some were tagged inaccurately. While field checking attempted to assign such areas to correct landscapes, not all could be checked in time, so best estimates were made. The results of the model were significantly accurate at the reconnaissance scale used. This is further outlined in Section 7.5 on the following page.

In future projects, the portable gap analysis tool could be used to model areas that are least similar to already mapped areas. In the current study, this could not be done because the gap analysis technology was not available at the time of training area selection.

7.2 HOW DO THE OUTPUTS FROM THIS PROCESS COMPARE WITH OTHER SOIL LANDSCAPE MAPPING PROJECTS?

The Brigalow Belt South Interim Bioregion was mapped faster and in a more objective manner than the Southern CRA and North-East CRA soil mapping projects.

The quality of outputs tends to be more variable than the CRA soil mapping projects, depending mostly on the quality of environmental variables available. The mapping technique is more rapid than soil landscape mapping, but does not result in as high a quality or as versatile a product. This is essentially because there is less time available for detailed checking that the traditional soil landscape mapping process enjoys. Soil landscape mapping also includes laboratory data and assessments of soil limitations. Neither of these was included in this project.

7.3 WAS THE PROJECT VARIATION SUCCESSFUL?

The project variation was successful in that it overcame time constraints and included the development of the flexible gap analysis tool, a hybrid approach in which:

- Soil surveyors are able to carry out their own sample site selection;
- Air photo interpretation is used as a moderating layer over the entire study area; and
- Soil landscapes can be modelled according to rule sets determined in-the-field.

This process is expected to result in better maps and datasets, and is recommended for future mapping projects of a similar nature. In particular, inclusion of limited soil testing for key plant growth variables is expected to increase the utility of the resulting dataset.

7.4 HOW WILL THE PROJECT BE USEFUL TO THE ASSESSMENT?

This project was undertaken to fill the substantial gap in soil knowledge within the BBS. This information was required by the joint BBS vegetation mapping project by the end of March 2002. As noted previously, mapped soil attributes are particularly important in terms of the vegetation mapping project in the BBS as the terrain is mostly subdued and soils assume a greater influence in determining plant species distribution. Soils information is a fundamental precursor to subsequent vegetation mapping so it was essential that soils information be provided as soon as practicable.

7.5 MODEL TESTING ACCURACY

7.5.1 Distribution and Descriptive Statistics

An assessment was undertaken in order to determine the accuracy of the model. The required outcomes of the assessment were to determine relationships between results, separated by into the provinces of Liverpool Plains, Northern basalts, Northern Outwash, Pilliga and Pilliga Outwash as outlined in pages 43-45 of this report.

In order to test the accuracy of the model, 180 independent field soil profiles were collected on a stratified random basis by 1:100 000 map sheets across all new mapping areas and compared to equivalent points determined by the model. The values of effective rooting depth, estimated potential available water capacity, fertility, A1 (topsoil) texture and regolith were extracted from both the profiles and the model and were compared. These were grouped into ranges as shown in Tables 11 through 15 with the exception of regolith, which is already grouped.

Table 11: Effective Rooting Depth ranges for calculation of accuracy
as per Isbell (1996)

Grouping (Range)	Value	Classification
1	0 – 0.25	Very Shallow
2	>0.25 – 0.5	Shallow
3	>0.5 – 1.0	Moderate
4	<1.0 – 1.5	Deep
5	>1.5	Very Deep to Giant
6	Disturbed	

Table 12: EPAWC ranges for calculation of accuracy

Grouping (Range)	Value	Classification
1	1-140	Very Low
2	141-300	Low
3	301-550	Moderate
4	551-800	High
5	800-1500	Very High

Table 13: Fertility ranges for calculation of accuracy

Grouping (Range)	Classification
1	Very Low
2	Low
3	Moderate
4	High
5	Very High

Table 14: A1 Texture ranges for calculation of accuracy
as per Isbell (1996)

Grouping (Range)	Texture Group	Texture Grades
1	Sands	Sand, Loamy Sand, Clay Sand
2	Loams	Sandy Loam, Loam
3	Clay Loams	Sandy Clay Loam, Clay Loam, Silty Clay Loam
4	Clay	Sandy Clay, Loamy Clay, Silty Clay, Medium Clay, Heavy Clay

Table 15: Regolith ranges for calculation of accuracy

Grouping (Range)	Value	Classification
1	R1	High coherence, low sediment delivery potential
2	R2	Low coherence, low sediment delivery potential
3	R3	High coherence, high sediment delivery potential
4	R4	Low coherence, high sediment delivery potential

A rating was given to point results on the basis of how well each matched the observed field result category. For all values other than regolith, a match was given a score of 1. If the groupings did not match but were one field removed from a match, a score of 0.5 was assigned. All other result's scored 0. Regolith values were assigned either 1 for a match or 0 for a mismatch due to the unranked nature of the data. The sum of these scores gave a possible overall score out of 5 for each point. Where limitations existed in the database, due to the absence of suitable facets in many of the soil landscapes, the soil type/facet that best matched the facet from the model was used.

Accuracy values and statistics were determined at the levels: a) match versus mismatch; and b) match versus partial match versus no match. These values were determined using the points ranking system above whereby a return of 1 indicates a match, 0.5 as a partial match and 0 as a non match. The Mode, Median, Mean and Standard Deviation were determined using the Descriptive Statistics tool in EXCEL. Mean was not applicable for Regolith as the data is nominal.

The overall accuracy using the match/partial match/no match system is shown in Table 21 on page 66. Accuracy for match/no match is also shown for each attribute in the same table. Accuracy for the different fields varied across provinces as can be seen in figures 13 through 17 on the following pages. All figures bar Regolith show match and partial match. The remaining percentage is for no match. Regolith values are only match and no match for reasons outlined above. The distribution of matches with regards to a score out of five is shown in Figure 18 on page 64. The distribution of results for match/partial match/no match is shown for each attribute in maps 16a - e in Appendix 3.

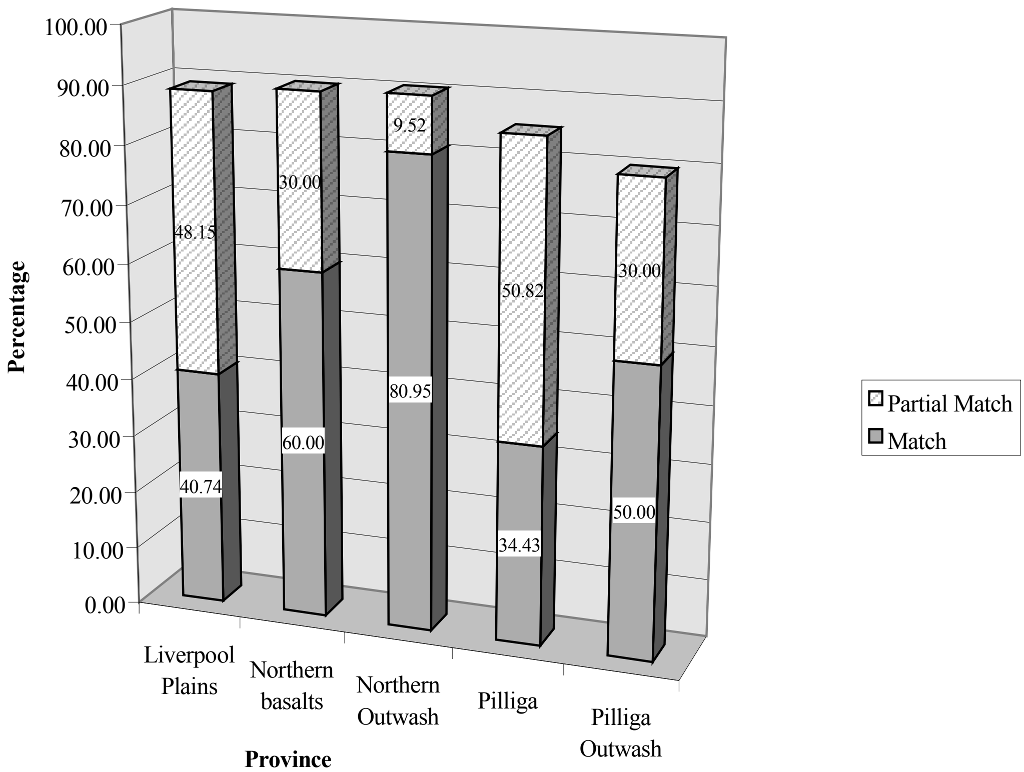


Figure 13: Percentage accuracy for topsoil texture across all provinces

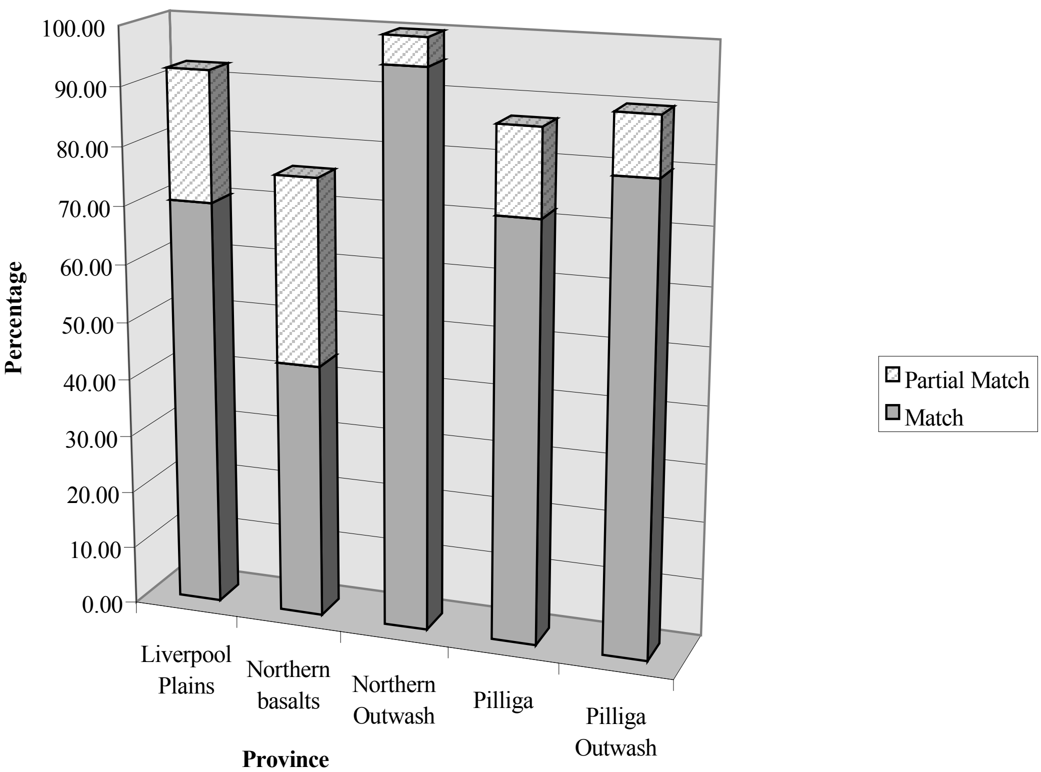


Figure 14: Percentage accuracy for Effective Rooting Depth across all provinces

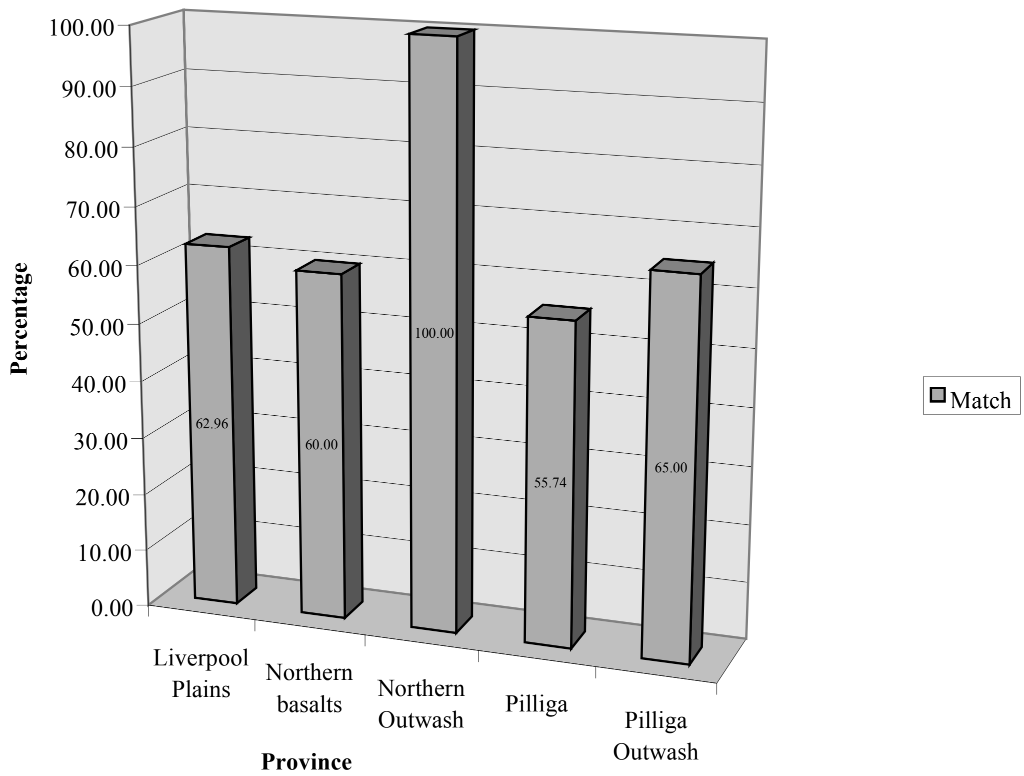


Figure 15: Percentage accuracy for Regolith across all provinces

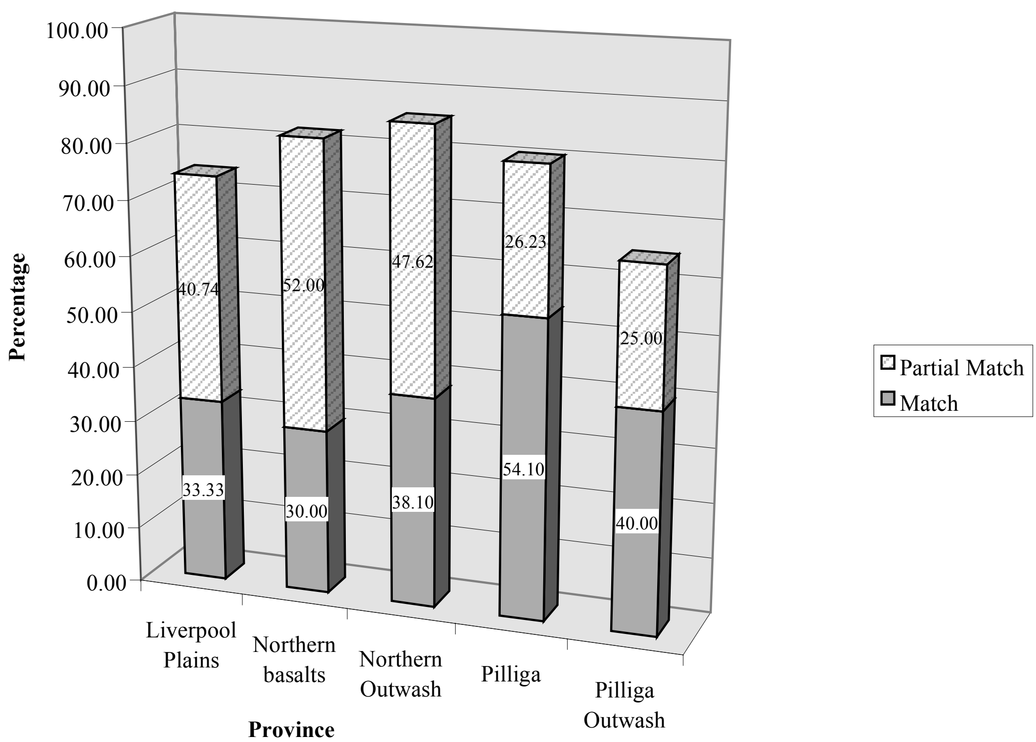


Figure 16: Percentage accuracy for EPAWC across all provinces

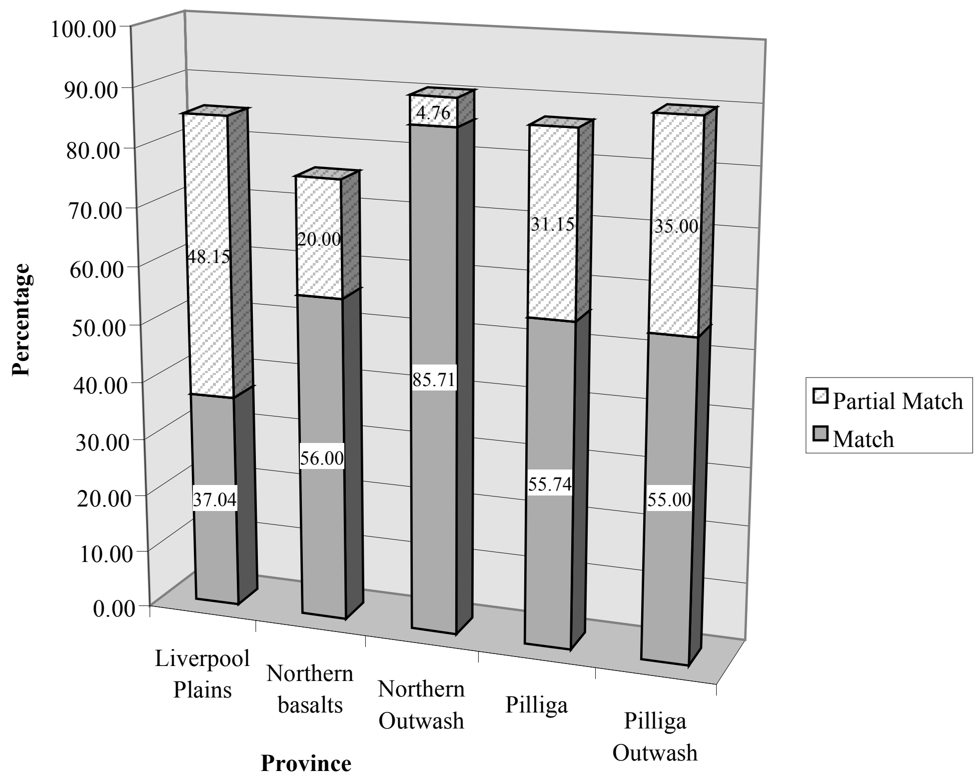


Figure 17: Percentage accuracy for Fertility across all provinces

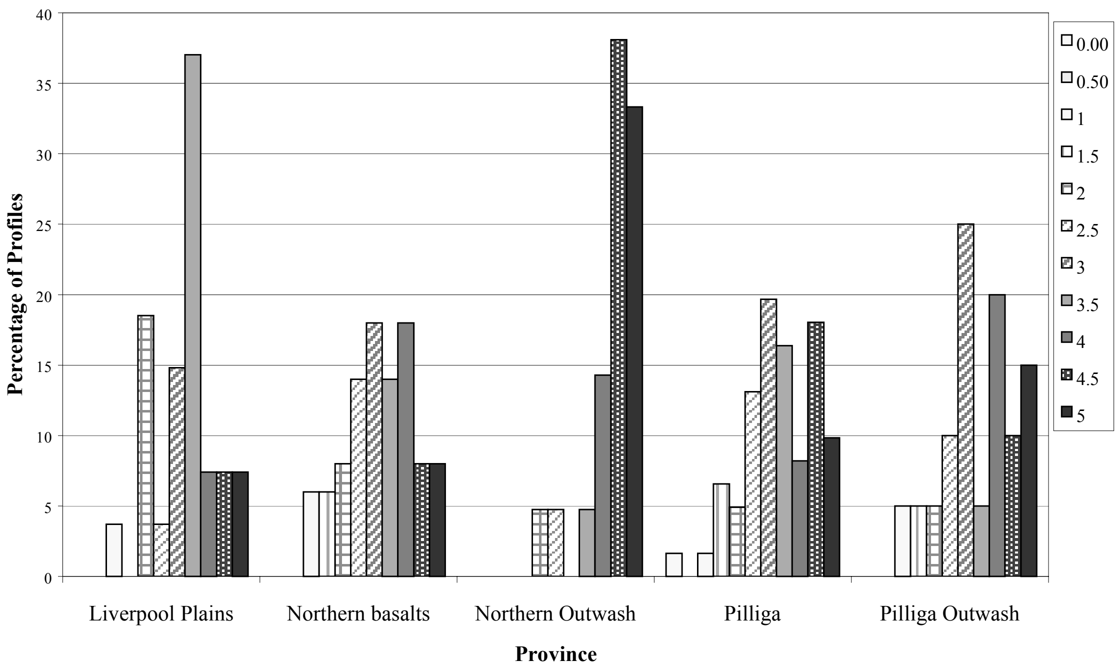


Figure 18: Percentage distribution of results showing percentage of matches for all for all main attributes from each province as a score out of 5. The score is given by the number of matches (1) and partial matches (0.5) added for each of the tested parameters; ERD, A1 texture, EPAWC, Regolith and Fertility as outlined above.

Table 16: Liverpool Plains province descriptive statistics for each of the test parameters and the final score. Test Parameter results are based on the groupings of data as outlined in Tables 11 through 15 and the final score as outlined on page 61

	Mean	Median	Mode	Standard Deviation
A1 Texture	2.89	3.00	4.00	1.09
ERD	4.96	5.00	5.00	0.19
EPAWC	3.52	3.00	3.00	1.34
Regolith Class	N/A	R3	R3	N/A
Fertility	2.89	3.00	3.00	0.97
Final Score	3.24	3.50	3.50	0.97

Table 17: Northern Basalts province descriptive statistic of each of the test parameters and the final score. Test Parameter results are based on the groupings of data as outlined in Tables 11 through 15 and the final score as outlined on page 61

	Mean	Median	Mode	Standard Deviation
A1 Texture	2.66	3.00	4.00	1.15
ERD	3.96	4.00	5.00	0.99
EPAWC	2.06	2.00	2.00	1.11
Regolith Class	N/A	R3	R3	N/A
Fertility	3.24	3.00	5.00	1.57
Final Score	3.17	3.00	3.00	1.10

Table 18: Northern Outwash province descriptive statistic for each of the test parameters and the final score. Test Parameter results are based on the groupings of data as outlined in Tables 11 through 15 and the final score as outlined on page 61

	Mean	Median	Mode	Standard Deviation
A1 Texture	3.81	4.00	4.00	0.60
ERD	4.95	5.00	5.00	0.22
EPAWC	4.05	4.00	4.00	0.74
Regolith Class	N/A	R3	R3	N/A
Fertility	3.14	3.00	3.00	0.48
Final Score	4.33	4.50	4.50	0.81

Table 19: Pilliga province descriptive statistic for each of the test parameters and the final score. Test Parameter results are based on the groupings of data as outlined in Tables 11 through 15 and the final score as outlined on page 61

	Mean	Median	Mode	Standard Deviation
A1 Texture	2.05	2.00	1.00	0.98
ERD	4.68	5.00	5.00	0.84
EPAWC	2.89	3.00	3.00	1.01
Regolith Class	N/A	R3	R3	N/A
Fertility	2.35	2.00	2.00	0.89
Final Score	3.35	3.50	3.50	1.13

Table 20: Pilliga Outwash Descriptive Statistic for each of the test parameters and the final score. Test Parameter results are based on the groupings of data as outlined in Tables 11 through 15 and the final score as outlined on page 61

	Mean	Median	Mode	Standard Deviation
A1 Texture	2.10	2.00	2.00	0.97
ERD	4.85	5.00	5.00	0.37
EPAWC	3.30	3.00	3.00	0.98
Regolith Class	N/A	R3	R3	N/A
Fertility	1.85	2.00	2.00	0.75
Final Score	3.40	3.25	3.00	1.15

Table 21: Accuracy for each of the provinces for all five parameters expressed as a percentage

	Liverpool Plains	Northern Basalts	Northern Outwash	Pilliga	Pilliga Outwash	Overall
Match and Partial Match	64.8	63.4	86.7	67.1	68	68.1
Match	48.9	60	80	54.83	58	55.9

7.5.2 Cohen's Kappa

A common measure of agreement is the sum of the diagonal entries in the error matrix divided by the total number of observations (in this case N=180). This is the proportion of samples placed into the same category by the model and from the field measurements, Po:

$$Po = \sum_{i=1}^{r} n_{ii} / N$$

However, this measure ignores agreement between the methods that may have been due to chance. The level of agreement due to chance is given by:

$$Pc = \frac{1}{N} \sum_{i=1}^{r} n_{i+} n_{+i} / N$$

This formula is based on the marginal totals n_{i+} and n_{+i} of the rows and columns. This formula also assumes that the marginal totals are independent which may not be the case for ordinal categories. There may be considerable difference between the observed and chance agreement. Cohen's Kappa, κ, is a statistic that measures this:

$$\kappa = \frac{Po - Pc}{1 - Pc}$$

When there is perfect agreement between the field and model values, Po = 1, so the maximum possible difference between the observed and chance agreements will be (1 – Pc). The kappa statistic thus gives the ratio of the excess of the observed agreement over chance agreement. Fitzgerald and Lees (1994) explain this as "The kappa statistic tests the null hypothethis that two independent classifiers do not agree on the rating or classification of the same object".

A rule of thumb (Landis and Koch 1977) for interpreting various kappa values is shown in Table 22 on the following page. The strength of agreement indicates the probability of a reference point value being the same as the model value. The weighted form of the Cohen's Kappa statistic is recommended for use with ordinal categories. This is described by Fleiss (1981). Everitt (1991) gives an estimate of the asymptotic standard error. This was used to calculate 95% confidence intervals and p-values.

Table 22 Strength of Agreement ratings for Cohens Kappa Statistic
from Landis and Koch (1977)

Kappa	Strength of Agreement
0.00	poor
0.00–0.20	slight
0.21–0.40	fair
0.41–0.60	moderate
0.61–0.80	substantial
0.81–1.00	almost perfect

Kappa was significant for all characteristics as shown in Table 23 below. Most of the soil properties measured for the value of kappa fell into the "fair" to "moderate" range. For ERD, the modelled values showed only slight concordance with the observed data.

Table 23 Summary of agreement statistics for the five soil characteristics. Po is the observed probability of agreement, Pc is the probability of agreement due to chance, kappa is the value of Cohen's Kappa statistic. This has been weighted for the first four ordinal categories. The unweighted form of kappa was applied to regolith as is the asymptotic standard error (ase).

	Po	Pc	kappa	ase	p-value	95% conf. interval
ERD	0.87	0.85	0.169	0.073	0.02	0.026 0.311
EPAWC	0.79	0.67	0.365	0.049	<0.001	0.269 0.460
A1 texture	0.79	0.58	0.494	0.047	<0.001	0.403 0.586
Fertility	0.84	0.68	0.511	0.049	<0.001	0.415 0.607
Regolith	0.64	0.43	0.367	0.055	<0.001	0.248 0.486

7.5.3 Synthesis and Conclusion of Accuracy Tests

The weakest Cohen's Kappa result was Effective Rooting Depth ("slight" range). This can be explained by a high probability of a match (85%) by chance due to the poor spread of the data. A large portion of the results fell in the very deep to giant range (soil depth greater than 1.5m). As the boreholes were dug to 1.5m, any sample with Rooting Depth greater than this fell into this range. All other results were in the "fair" to "moderate" range, the strongest match being Fertility (84% probability of a match, 68% probability by chance). Regolith had the lowest probability of a result by chance (43%) and overall (64%).

Results show that there is variation in the accuracy of the model for different provinces. There are many possibilities as to the reason that this occurred. Changed values due to incorrect classification are one possible explanation for incorrect results, particularly with regard to regolith values. This is currently under review and changes have been made. It was observed that results that were mostly incorrect had been extrapolated across large distances ie the model had used soil information described at one location to match soils several 1:100 000 map sheets away.

The quality control check will satisfy many of the concerns raised by various Stakeholders, who expressed various sentiments regarding the accuracy and usefulness of the soil maps produced by the project. Most areas of the model (eg Northern Outwash) have mapped accurately while other areas have not (eg Liverpool Plains). The majority of ground truthed areas (68.1%) were consistent with model values. The check was done using 180 soil profiles, which were selected on a stratified random basis in areas without Soil Landscape mapping covering each of the unpublished map sheets in the region.

This provides an indication of the accuracy of the model over different provinces, which were mapped using different environmental attribute weightings. Considering the broad scale reconnaissance level of mapping used (1:100 000), mapping could only be tested at a general landscape level as opposed to the 14,420 individual polygons, as excessively extensive fieldwork would be required to check the data at any resolution greater than that undertaken.

Soil landscape mapping is an interpretive process and as such, is based on the soil surveyor identifying the relationship between soils and landscapes properties. It is these key relationships that were identified by the soil surveyors and incorporated into the model.

Higher resolution data can be viewed for each province using the provided ArcVIEW shape files and attribute tables. All Individual profiles described in the course of this project will be made available for viewing through SPADE (located at http://spade.dlwc.edu.au) in the near future.

The process undertaken in this project is similar to that of a Soil Landscape Map. The production of a Soil Landscape map over this size area (54 000 ha) usually requires 21 person years to produce the final output while this project required approximately 5 person years to complete. The accuracy of Soil Landscape mapping ranges from 73 to 87 % accurate based on a range of soil properties including depth, texture, plasticity and Organic Matter content (Dewar *et al.*, 1996). The cost of this project worked out at around \$9.35 km^2. This equates favourably with figures provided by McKenzie (1991), who states the figure of \$28.00 km^2 (1988 value quoted) as being sufficient to provide a crude prediction of soil land characteristics at a nominated location

All minor alterations and typographical errors reported have been edited. An overlay has been included in the appendices showing road network and towns in the region.

The Department of Land and Water Conservation Soil Information Systems Unit are the custodian of this data set. As errors become known, efforts will be made where resources permit, to update the quality of the dataset. Users are requested to contact soils@dlwc.nsw.gov.au to provide feedback and information regarding inconsistencies, errors, omissions or updates that will enhance the data set.