Optimal prediction of coastal acid sulfate soil severity using geographic information systems

Marcus John Morgan
University of Wollongong

Optimal Prediction of Coastal Acid Sulfate Soil Severity using Geographic Information Systems

A thesis submitted in fulfilment of the requirements for the award of the degree

MASTER OF ENGINEERING (HONOURS)

From

UNIVERSITY OF WOLLONGONG

By

MARCUS JOHN MORGAN

B.Sci (Environmental), B.Comm (Economics)

FACULTY OF ENGINEERING
2006

CERTIFICATION

I, Marcus J. Morgan declare that this thesis, submitted in fulfilment of the requirements for the award of Master of Engineering (Honours), in the Faculty of Engineering, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Marcus J. Morgan

24 March 2006

Table of Contents

List of Figures

List of Tables

List Of Plates

Abbreviations

ACASS	Actual Coastal Acid Sulfate Soil
AGD	Australia Geodetic Datum
AHD	Australian Height Datum
Al	Aluminium
ALS	Airborne Laser Scan
AMG	Australian Map Grid
ANOVA	Analysis Of Variance
ANS	Australian National Spheroid
ANZECC	Australian and New Zealand Environment and Conservation Council
ARMCANZ	Australian Resource Management Council of Australian and New Zealand
ASRIS	Australian Soil Resource Information System
Cl:SO4	Chloride:Sulfate
CANSIS	Canadian Soil Information System
CASS	Coastal Acid Sulfate Soils
CASSOK	Coastal Acid Sulfate Soils Ordinary Kriging Model
CEC	Cation Exchange Capacity
CIS	Commonwealth of Independent States
CLT	Central Limit Theorem
DEM	Digital Elevation Model
DTM	Digital Terrain Model
EAA	European Environment Agency
EC	Electrical Conductivity
EM	Electromagnetic Survey
ESB	European Soil Bureau
EGII	European Geographic Inform Infrastructure
ESRI	Environmental Systems Research Institute
EU	European Union
EUSIS	European Union Soil Information System
ExAl	Exchangeable Aluminium
Fe	Iron
GCS	Geographic Coordination System
GDA	Geocentric Datum Australia
GIS	Geographic Information Systems
GPI	Global Polynomial Interpolation
GPS	Global Positioning System
GWR	Geographically Weighted Regression
FAO	Food and Agriculture Organization
IDW	Inverse Distance Weighting
ISG	Integrated Survey Grid
KCl	Potassium Chloride
LIDAR	Light Detection And Ranging
LPI	Local Polynomial Interpolation
LRR	Land Resource Region
LRRC	Land Resources Research Centre
LRU	Land Resource Unit
MGA	Map Grid Australia
MLRA	Major Land Resource Area

NASIS	National Soil Information System
NRCS	Natural Resource Conservation Service
NSW	New South Wales
OK	Ordinary Kriging
PCASS	Potential Actual Coastal Acid Sulfate Soil
PCS	Projected Coordination System
pH	$\log_{10}(H^+)$
RBF	Radial Basis Function
REMS	Reclaimed Water Management System
RMSPE	Root Mean Square Prediction Error
RMSS	Root Mean Squared Standardised
RV	Representative Value
Spocas	Sulfate –
Scr	Sulfur-Chromium Reducible
SAI	Space Applications Institute
SALIS	Soil and Land Information System
SOTER	Soil and Terrain Database
SPADE	Soil Profile Attribute Data Environment
SRADWG	Shoalhaven River Acid Drainage Working Group
SRTW	Self Regulating Tilting Weir
SSIS	Spanish Soil Information System
SSSD	State Soil Survey Database
SSURGO	Soil Survey Geographic database
STATSGO	State Soil Geographic database
SQL	Structured Query Language
TAA	Total (Titratable) Actual Acidity
TF	Thiobacillus Ferroxidans
TPA	Titratable Peroxide Activity
UK	Universal Kriging
USDA	United States Department of Agriculture
UTM	Universal Transverse Mercator
VIF	Variance Inflation Factor

Abstract

Coastal Acid Sulfate Soil (CASS) is a soil-water phenomenon that causes soil and water pollution resulting from the exposure, typically human-initiated, of pyrite to atmospheric and biotic oxygen. Structural deformation of capital works, combined with loss to flora and fauna (biodiversity) resulting from CASS has caused major concern to environmental managers, industries that rely directly on high quality water conditions for day-to-day operations, and landholders who experience characteristic scalding and other associated environmental problems on land adjacent to disturbed areas.

Areas of CASS in Australia have been identified by Department of Natural Resources (DNR) using a combination of expert knowledge, geomorphologic principles and Geographic Information Systems (GIS) known as Acid Sulfate Soil Risk Maps. These maps have been applied by local managers in planning and natural resource management to identify areas showing the highest probability of being severely affected by CASS.

In this project, with the DNR model as a starting point, the aim was to improve the way CASS severity is assessed. This included using five major soil-chemical parameters and/or relationships in a number of geostatistical models. The five parameters included were: Total Actual Acidity (TAA), pH, Chloride to Sulfate ratio (Cl^-:SO_4^{2-}), Depth to actual CASS layer (Jarosite layer), and Exchangeable Aluminium per cent of total Cation Exchange Capacity. Other parameters such as depth to Potential CASS layer (Pyrite layer) and Sulfur per cent (S%), also have weight but not as significant as the other parameters and were subsequently removed from further detailed analysis.

Ordinary Kriging (OK) was identified as the most suitable geostatistical method to predict CASS severity using the aforementioned soil-chemical principles. The resulting 3-Dimensional model was compared to the 2-Dimensional DNR Risk Maps with similarities in both models validating both approaches in determining severity using different methods. The CASSOK model put a greater emphasis on soil parameters down the soil profile and how they relate to surface elevation across a finite study area (Broughton Creek floodplain, New South Wales).

Applying the new CASSOK model to broader areas of New South Wales will be dependent on available data to input into the model. Using the current DNR risk maps is a broad indication of an area, using CASSOK will give a greater indication of what can be expected 2m below the surface. The ability to create a method that can be applied across the entire state of New South Wales, and then to a national level will be an invaluable resource to land managers in future planning and risk management.

Acknowledgements

I would like to acknowledge my supervisor Professor Buddhima Indraratna for his endeavour to understand the complex phenomena that are Coastal Acid Sulfate Soils and to support students who aim to understand more about Coastal Acid Sulfate Soils. I would like to thank Warwick Papworth from Shoalhaven City Council for his financial and practical support throughout my time as the Acid Sulfate Soils Project Officer, during which I completed this thesis. I would also like to thank Dr William Glamore for his patience in dealing with my fundamental questions about Acid Sulfate Soils and for his friendship and support over my candidature. My gratitude goes to Roy Lawrie, the true master of soil science. I also must thank Dr Phil Flentje and Dr Jeff Price for their advice, encouragement and stress-relieving golf games. Last of all I would like to thank my gracious, loving, and supportive partner Lynnette Hoffman who has had to endure the hardships that comes with supporting someone completing a research project, especially the irreplaceable time we have had from one another.
Soy Terminar! Gracias mi amore.

Dedication

This is dedicated to all the research students that undergo serious difficulties in trying to complete a research document due to reasons out of their control, I understand your grief.

Chapter 1: Introduction

1.1 Introduction to Coastal Acid Sulfate Soils

The development of coastal floodplain drains within areas underlain with pyritic soil has exposed these layers of soil to atmospheric oxygen (Pease, 1994, Sammut, *et al.* 1997, Wilson, *et al.* 1999). This has caused the chemical oxidation process to be initiated and the resultant generation of acid to be transported into nearby waterways. The negative effects on industries that depend on high quality waterways such as the fishing and agricultural industries can be seen in the loss of biodiversity, loss of flora and fauna, and loss of income.

In order to address these issues and to further understand the process of forming Coastal Acid Sulfate Soils (CASS), the development and implementation of remediation strategies to lessen or eliminate the negative effects of CASS has led to numerous research projects addressing this issue. The South Coast of New South Wales (Australia) was designated one of seven CASS 'hotspots' by Department of Natural Resources (DNR) after a number of investigative projects uncovered the initial problem (Norwood, 1975; Pease, 1994) followed by projects to address the problem, which focused on the experimentation and implementation of engineering strategies (Blunden, 2000; Tularam *et al.*, 2001; Glamore, 2003; Banasiak, 2004). These have shown positive improvement in water quality.

CASS are widespread throughout coastal areas of Australia, but have become a more severe environmental problem in southern Queensland and Northern New South Wales. The development of flood mitigation drains and other structures have exposed the soil to atmospheric oxygen, hence pyritic oxidation. In order to ensure future environmental problems with CASS are not encountered, a methodology to assess areas prone to developing CASS is necessary.

In this project, Broughton Creek floodplain, a floodplain on a tributary of the Shoalhaven River, will be assessed in terms of the potential risk that could occur if this land was further disturbed. A model will be developed using a Geographical

Information System (GIS) to help land managers identify areas where development should be avoided and to aid in future management of these sensitive areas.

1.2 Objectives

The project objectives included:

- The development of a CASS severity predictive model based on available secondary data. This approach focuses on risk associated with development in CASS areas, and the suitability and unsuitability of proceeding with development within these areas.

- Using more accurate elevation data and tools of a GIS to provide end-users with a representation of severity in an area typically affected by CASS; and

- The development of Best Management Practices (BMP) guidelines for the Broughton Creek Floodplain based on CASS risk management.

1.3 Thesis Structure

The organisation of this research project thesis is divided into three parts.

1.3.1 Section 1: Literature Review (Chapter 2, Chapter 3)

Section 1 assesses the current technological advances within CASS, the description of GIS, and its application to various environmental problems.

1.3.2 Section 2: Model Development (Chapter 4, Chapter 5, Chapter 6)

Section 2 collates existing data into a GIS to generate a statistically robust predictive model. The model pools information from research work previously performed within the Shoalhaven. The model is developed into a functional system and displayed in a spatial form that is user-friendly which will aid in future decision making.

1.3.4 Section IV: Best Management Practices and Recommendations for future development in CASS (Chapter 7, Chapter 8)

The development of a set of best management practices is the culmination of the preceding sections. Recommendations for future development are addressed in the best management practices. Alternate approaches to developing further capabilities within the model.

Chapter 2 Coastal Acid Sulfate Soils

Chapter 2 introduces Coastal Acid Sulfate Soil (CASS) as a phenomenon and details the soil chemistry behind the forming of these soils, as well as the change in soil and water chemistry through the process. The worldwide and regional distribution of CASS is shown. Important variables in distinguishing a CASS are explained and elaborated throughout the Chapter and new techniques and existing techniques in identifying CASS are analysed. A number of management strategies currently in place and proposed will be discussed and evaluated.

2.1 Introduction to Coastal Acid Sulfate Soil (CASS)

Coastal Acid Sulfate Soil (CASS) is a typically human-initiated, soil-water phenomenon that pollutes soil and water by exposing of the pyrite layer to atmospheric and biotic oxygen. Structural deformation of capital works combined with loss of flora and fauna (biodiversity) resulting from CASS has caused concern to environmental managers, industries that directly rely on high quality water for day-to-day operations, and landholders who experience scalding on land adjacent to disturbed areas.

Current research has centred on understanding and identifying CASS (Norwood 1975; Pease, 1994) as much as providing management solutions (Blunden, 2000; Glamore, 2003). This project aims to improve the process that identifies areas of CASS across spatially vast areas with only limited information about soil conditions in a given area.

Understanding CASS forming processes, determination of trigger levels of certain soil parameters, and an overview of the current methods describing the distribution of CASS, will build the foundations for the spatial statistical models presented in the outcomes of this project (see Chapter 6).

2.1.1 Thionic Fluvisol

CASS was first identified in the Netherlands almost 270 years ago (Pons, 1973). These soils are internationally recognised as a member of a group of soils termed Thionic Fluvisols (FAO, 2001). A Fluvisol (of which Australian Alluvial Soil is a typical

example) is a soil which shows the properties of a typical young soil, of which, when forming, received fresh sediment during floods that shows stratification of the organic matter. Its connotation (L. *fluvius*) explains a soil developed in alluvial deposits from a river although a Thionic Fluvisol can also be found in lacustrine and marine environments. The Fluvisol environment is typically threatened by flooding, such as alluvial planes which are normally planted with annual crops, orchards, or has some grazing. Fluvisols usually require flood control, drainage systems or irrigation systems to be put in place, before being used. Drainage canals have caused a rise in exposed sediments, and the discovery of Thionic Fluvisols.

Thionic Fluvisols differ from regular Fluvisols in that the former contains pyrite (FeS_2) and suffer severe acidity and high levels of noxious Al-ions affecting aquatic flora and fauna. The common name given to these soils is 'acid sulfate'. When forming, these soils must have a number of qualities that include (FAO, 2001):

- Soil must contain an Iron – iron oxides or hydroxides that can easily reduce Ferric (Fe^{3+}) ions to Ferrous (Fe^{2+}) ions.
- Soil must contain Sulfur found in sea or brackish water.
- Anaerobic Conditions – to enable reduction of sulfate (SO_4^{2-}) to sulfide (S^{2-}) and iron oxides as explained above.
- Iron and Sulfate Reducing Microbes
- Sufficient Organic Matter – available in environments with pallustric vegetation such as mangrove swamps, which forms bicarbonates (HCO_3^-).
- Tidal Flushing – to remove bicarbonates (HCO_3^-) formed during the formation of pyrite.
- Slow sedimentation.

Once the reduction of ferric ions to ferrous ions, and sulfate to sulfide has proceeded and the microbes and organic matter work to form bicarbonates and acid – a balanced chemical state. However when bicarbonates are flushed from the soil matrix by adjacent water the balance is altered and the result is acid pyrite. This is the commonly known 'acid sulfate soil' that affects the quality of water downstream after heavy rain when large amounts of H^+ and noxious Aluminium ions are released from the soil matrix,

damaging man made structures such as culverts and bridges, and natural habitats including fish, oysters and native aquatic plants.

2.1.2 Distribution

Throughout the world, soil maps and surveys differ in accuracy. Dent & Dawson (1996) estimate that from a compilation of a number of sources there are approximately 25 million hectares of coastal sulfidic sediments, or sediments that have the potential to develop into Actual Coastal Acid Sulfate Soils (ACASS), throughout the world. The most concentrated CASS areas are in the low lying coastal land of Asia, Africa and Latin America. Table 2.1 depicts the distribution of CASS within these regions, including Australia. As the tropical low lying parts of Asia (e.g. Indonesia, Thailand and Vietnam) are quite low and expansive, the majority of CASS, or 7.7 million hectares, are found in this region. In Australia, the majority of known CASS are found in coastal Queensland (77%) with New South Wales representing about 20% of the identified CASS in Australia.

2.1.3 Spatial Distribution in Australia

The extent of Potential Coastal Acid Sulfate Soils (PCASS) throughout the world was mapped by FAO-UNESCO (1978), but because it was based on an early Australian soil map that didn't recognise PASS it didn't identify any Thionic Fluvisols in Australia. Galloway (1982) produced a map that identified mangroves throughout Australia which gave a broad approximation of PASS areas. Chappell and Grindrod (1985) and Woodroffe and Chappell (1991) identified the 'big swamp' phenomena that hypothesised that in lower energy coasts around Australia, deposits from former mangrove areas were much greater than the distribution mapped by Galloway, hence a greater distribution of CASS/PCASS was possible. Detecting the geomorphology of PASS is complicated and an expert system designed by Dent & Dawson (1996) simplified the complex geomorphologic process by assimilating expert knowledge into a decision tree that would help identify the possibility that a CASS, either potential or actual was present. Figure 2.2 summarises the criteria used by Dent & Dawson in 'The Acid Test'.

Please see print copy for Table 2.1

Table 2.1 Estimation of CASS distribution by region (After Dent & Dawson, 1996).

Prob of ASS

Please see print copy for Table 2.2

Table 2.2 The Acid Test Criteria, with Expert Criteria^ from Broughton Creek (NSW, Australia). (After Dent & Dawson, 1996; Lawrie, 2002)

Areas of CASS in NSW have been identified by the Department of Natural Resources (DNR, formally Department of Land and Water Conservation) using a combination of expert knowledge and GIS. This mapping has been used by managers to target areas with the highest probability of being detrimentally affected by CASS. Atkinson *et al.*

(1996) explain the process, which is revisited in Chapter 3.8. Two main chemical properties of soil were assessed, which include TAA and pH. We will see later that pH and TAA exhibit auto-correlation (see Chapter 4.9), and therefore one must be eliminated from our analysis. Over the 6,000 hectares of PCASS mapped by DNR there were 840 samples of soil taken, which equates to 1 sample every 7.14 hectares. Due to the variability of soil this result was quite low, so the results should be taken as broad representation only.

Queensland's (QLD) Department of Natural Resources, Mines and Energy (NR&M) attempted to map areas of ASS based on extent, location, and risk level. The NR&M used geological layers (Quaternary (Qm), coastal and estuarine) to identify preliminary areas of ASS, which totalled close to 2.3 million hectares from the 1975 Geological Survey of QLD. 1:1,000,000 scale mapping was accomplished for all of QLD and then 1:50,000 or 1:25,000 for those areas found to be more severe or likely to contain CASS. Boreholes were taken approximately every 100 hectares in a 1:1,000,000 scale, every 25 hectares in a 1:50,000 scale and every 6.25 hectares in a 1:25,000 scale. The risk maps also contain the depth of PCASS and CASS layers.

This project aims to increase the accuracy of mapping areas based on a number of other soil chemical properties, rather than purely focusing on geological properties. This project will take an area that has already been identified from geological mapping and large scale borehole sampling, and try to predict the level of severity or risk it contains. However, an understanding of the geological forming processes and morphology of the landscape will aid in understanding how CASS are formed.

2.2 Formation – Geomorphology

2.2.1 Shoalhaven River: A Barrier Estuary

A case study of the Shoalhaven Floodplain, the study site for the project, is used to define a typical formation of CASS in an estuarine environment. The lower Shoalhaven River is located approximately 160 kilometres South of Sydney (New South Wales, Australia). The Shoalhaven floodplain is composed of Holocene sediments, the majority of which are estuarine, however some areas boast a thin layer of Pleistocene alluvium above them (Young *et al.*, 1996). The low-lying floodplain surrounding Broughton

Creek is marginally above 0 Australian Height Datum (AHD), with the majority of the land falling between 0 to 5m AHD. The levees of major waterways such as Broughton Creek and the Shoalhaven River have accumulated a thin layer of sediment above this estuarine sediment.

The lower Shoalhaven river is characterised by low lying alluvial plains, developed as a result of deltaic estuarine infill (Woodroffe *et al.*, 2000), as well as a late Quaternary sand barrier that separates the ocean from the deltaic flood plain. This estuarine infill progressed mainly through deltaic channels isolating the smaller basins (Woodroffe *et al.*, 2000). Wright (1985) describes the Shoalhaven floodplain as an example of a wave dominated delta whereby wave energy progressively reworks the sand barriers inland. The resultant formation is described by Roy (1980; 1984) as being one of three types of Holocene embayment fills, these being either a drowned river valley estuary, open ocean (saline coastal lakes) or barrier estuary. The Shoalhaven floodplain formed as part of a barrier estuary system, created during the Holocene period (last 10,000).

Barrier estuaries are normally characterised by narrow elongated channels, often with broad tidal and sand flats. (Roy, 1984). The Shoalhaven flood plain exhibits a very shallow deltaic surface based on infilling due to the inflow of sediments from tidal cycles. In terms of development, the flood plain has reached a mature stage and begun to re-deposit sand along the shores associated to the flood plain. Figure 2.2 shows the landform morphology of the Shoalhaven flood plain and Figure 2.2.1 shows the radiocarbon ages of the various landforms within the Shoalhaven river deltaic estuarine plain. The Shoalhaven River protrudes into the ocean periodically via the Shoalhaven heads but the majority of the time it only “exists” through an artificial canal created by Alexander Berry in 1822. The banks of this canal are prone to erosion and is continually widening.

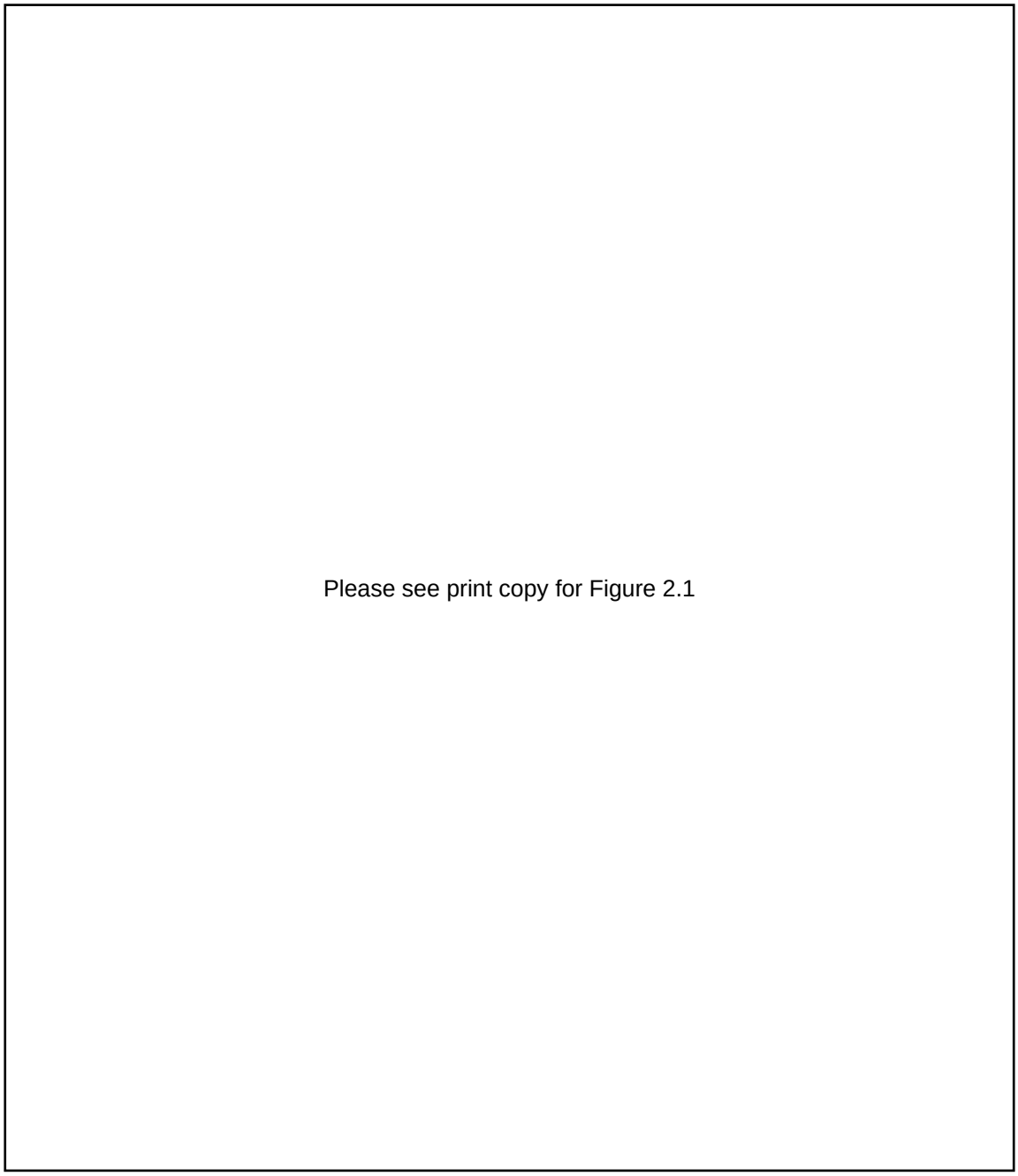

Figure 2.1 Landforms of the Shoalhaven River deltaic estuarine plains (Umitsu *et al*., 2001).

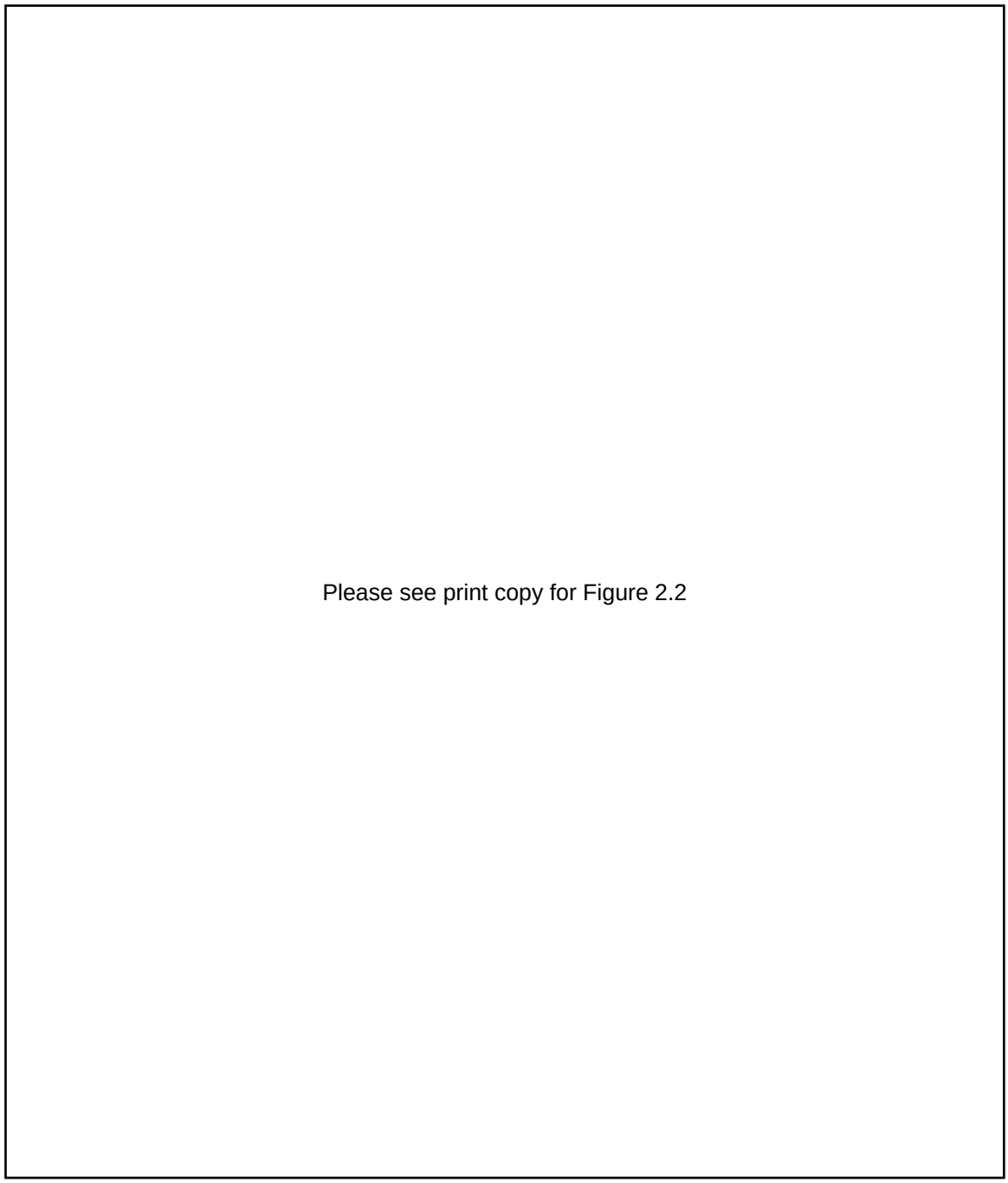

Figure 2.2 Stratigraphic profiles E and F, Shoalhaven River deltaic estuarine plains. (after Umitsu *et al.*, 2001).

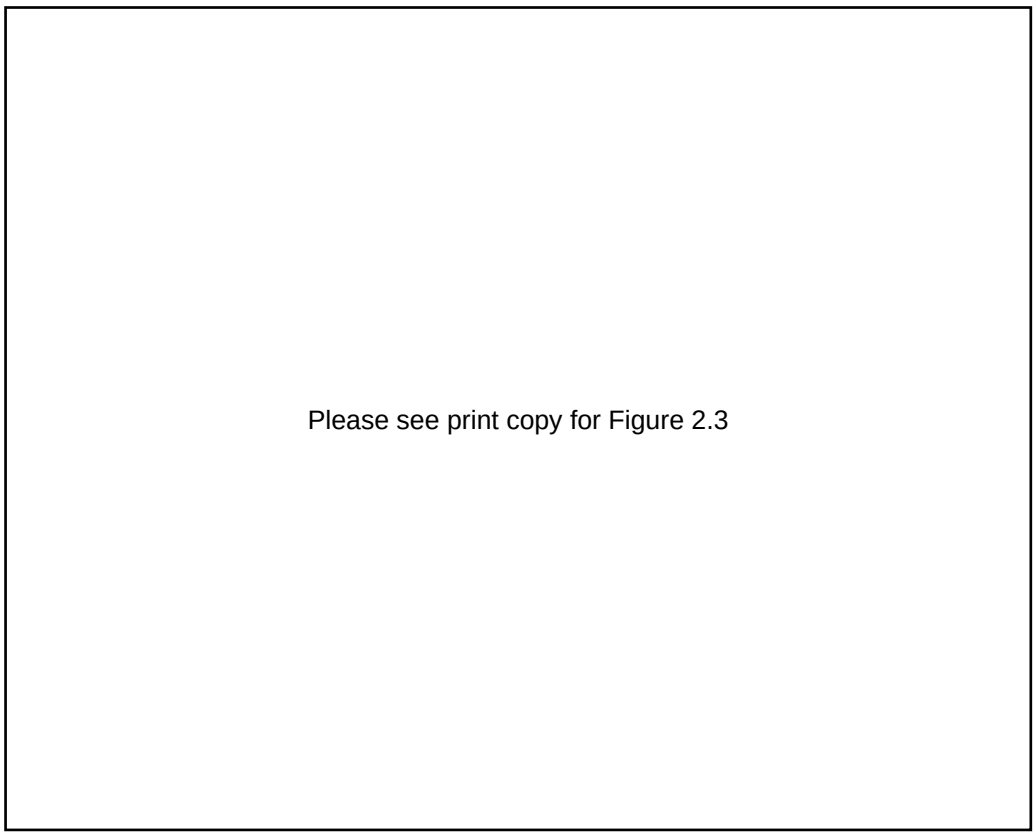

Figure 2.3 Typical Coastal Estuary (Brooke *et al*., 2005).

The formation of the barrier estuary has an adverse impact on the surrounding land. Due to the continual infilling of the barrier estuary during large rain events, there is a high probability that the sand barrier will be breached by the volume of flow from the river. In the case of the Shoalhaven River, a number of drainage channels were either implemented or upgraded (usually widened) in the early to mid 1970's. This had the affect of alleviating the build up of flow towards the mouth of the Shoalhaven River. Since these flood mitigation drains were implemented, there has been a large number of fish kills observed within the Shoalhaven floodplain (Lawrie, 2003). Norwood (1975) researched the Shoalhaven floodplain and reported on the severity of CASS conditions within this region. Due to excavation of the flood mitigation drains, the pyritic layer of the sub-surface soil is exposed to vertical influx of atmospheric oxygen which causes a take over of rapid oxidation. During heavy rainfall the soil releases large amounts of sulfuric acid into neighbouring waterways which affects the local fishing industry.

2.2.2 Influence of Floodplain Drainage on Landform

The layer of pyrite in a natural estuarine floodplain, is continually being submerged under saline or brackish water (Figure 2.4). Natural vegetation such as mangroves, salt marsh and Melaleuca's (tea tree) exist in an undisturbed environment. In this environment there is rarely any acid formation due to the equilibrium that exists as a result of buffering from bicarbonates in sea or brackish water, and the continual inundation of what is an acid forming layer.

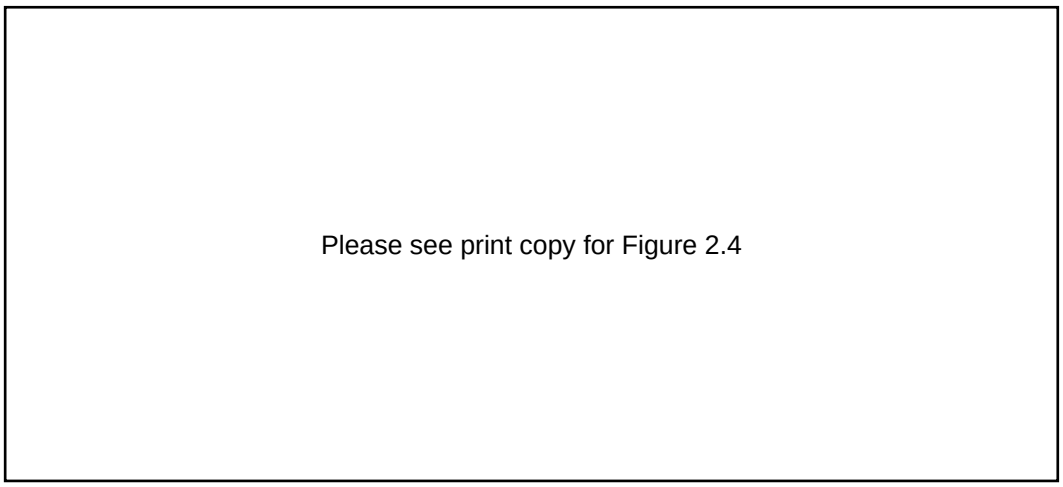

Figure 2.4 Natural Setting - low frequency, low magnitude, short duration acidity (Sammut & Lines-Kelly, 2001).

With many floodplains exhibiting ideal conditions for agricultural purposes such as grazing beef or dairy cattle, there was a landholder push to get local authorities to increase the flow from these areas after heavy rainfalls. In the late 1960's, early 1970's flood drainage channels were constructed, or existing drains were widened and deepened causing a change to the land, ground water drainage levels, and vegetation. As a result the groundwater table was lowered and the once inundate layer of pyrite became exposed, triggering a series of chemical reactions that produced acid and sulfate in what is now known as CASS. With increased drainage, there is increased runoff and with the development of PCASS into ACASS there is an increase in the frequency and severity of acid events (Figure 2.5).

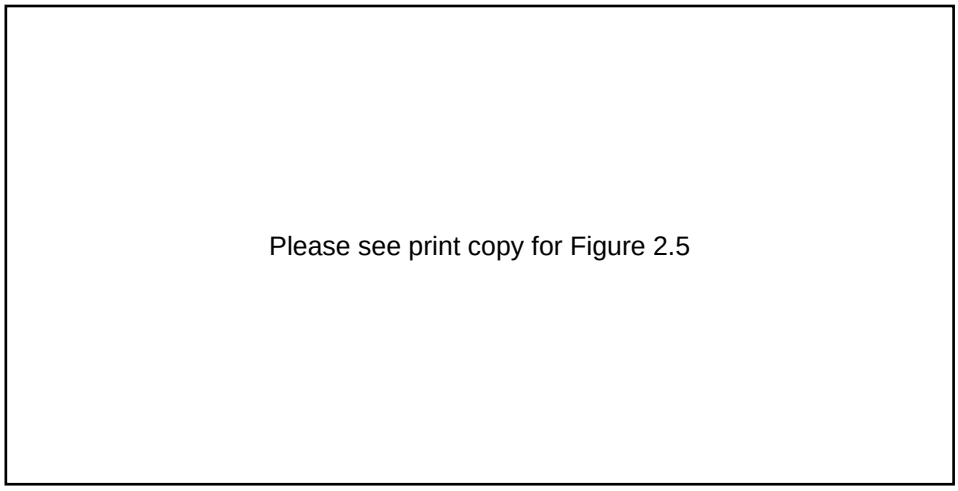

Figure 2.5 Post Drainage - High frequency, high magnitude, persistent acidity (Sammut & Lines-Kelly, 2001).

The chemical composition of pyrite can help in understanding how this acid forming process occurs and how these elements become harmful to a natural system. Iron Sulphide is the chemical component that is converted to form pyrite.

2.2.2 Iron Sulphide

CASS are formed when soil containing iron sulfides (iron pyrite) is exposed to air and oxidised. The generation of pyrite is explained by the chemical process described below (White *et al.*, 1996; FAO 2001):

$$4SO_4^{2-} + Fe_2O_3 + 8CH_2O + 1/2\ O2 \rightarrow 2FeS_2 + 8HCO^-_3 + 4H_2O(l) \qquad (2.1)$$

Ocean Water — Sediment — Organic Matter — Pyrite

Iron pyrite is benign and harmless in its reduced and undisturbed state but, when disturbed or exposed to atmospheric oxygen, it begins to oxidise, catalysed by bacteria that produces sulfuric acid. Ferrous iron is released into the waterways, as is soluble Aluminium if the soil matrix remains acidic, due to a lack of acid neutralising material in associated clay minerals (Melville, *et al.*, 1993).

2.2.3 Pyrite Oxidation

Pyrite oxidation is triggered when the water table is lowered below the upper boundary of this layer which is also known as the PCASS layer. Atmospheric oxygen (via root channels) and water alters the equilibrium by converting pyrite into ferrous iron and sulfate as represented in equation 2.2.

$$FeS_2 + 7/2\ O_2 + H_2O \rightarrow Fe^{2+} + 2\ SO_4^{2-} + 2\ H^+ \quad (2.2)$$

Pyrite Atmospheric Oxygen Iron (Ferrous) Sulfate Acid

However, the process does not show any acidity when carbonates are present, usually from brackish or sea water. The acid produced by 2.2 is stabilised with Oxygen molecules from Calcium Carbonate as represented in equation (2.3)

$$CaCO_3 + 2\ H^+ \rightarrow Ca^{2+} + H_2O + CO_2 \quad (2.3)$$

Calcium Carbonate Acid Calcium Water Carbon Dioxide

ASS is known to possess the properties of a silty clay loam and exhibit relatively moderate Cation Exchange Capacity. An ASS can be strongly calcareous when Calcium is widely distributed throughout the soil matrix. In this situation gypsum can be formed, as represented by equation 2.4.

$$Ca^{2+} + SO_4^{2-} + 2\ H_2O \rightarrow CaSO_4.2H_2O \quad (2.4)$$

Calcium Sulfate Water Gypsum

If the environment is depleted of acid neutralising carbonates which occurs during droughts or where the water table decreases enough to fall below the PCASS layer, the 2 mol of acid produced in equation 2.3 is no longer neutralised and the pH can fall below 3.5 which will mobilise other cations and heavy metals bound to the soil matrix. The Ferrous Iron in equation 2.2 will be oxidised to form jarosite (assuming an ample presence of K^+ ions), which is identified in the soil matrix by its yellow colour and strong sulfidic odour. This happens via process represented in equation 2.5:

$$Fe^{2+} + 2/3\ SO_4^{2-} + 1/3\ K^+ + 1/4\ O2 + 3/2\ H_2O \rightarrow 1/3\ KFe(SO_4)^2(OH)_6) + H^+ \qquad (2.5)$$

Iron (Ferrous) Jarosite Acid

Jarosite is an intermediary product formed in the process of pyrite oxidation. After Jarosite is formed the acidic component is eventually flushed from rainwater or floodwaters filtering down the profile and then transports it into drains and rivers downstream. Ferric Hydroxide ($Fe\ (OH)_3$) is formed from jarosite, which then converts to a more stable Goethite ($FeOOH + H_2O$). This process is known as 'acid at a distance' and forms a further 2 mols of acid (H^+), characterised by a red-brown flocculation in drainage channels (Glamore, 2003). The most commonly flocculated precipitate is iron mono-sulphide which can further oxidise to produce another mol of acid (H^+) if conditions become suitable (i.e. soil conditions are dry and oxygen is available).

The availability of oxygen in the soil matrix will determine the current state of iron, in the pyrite oxidation process. The redox potential and soil pH is a good measure of this. Figure 2.4 represents the various transitional states found in a typical CASS including the pH and redox potential when theses states are likely to occur.

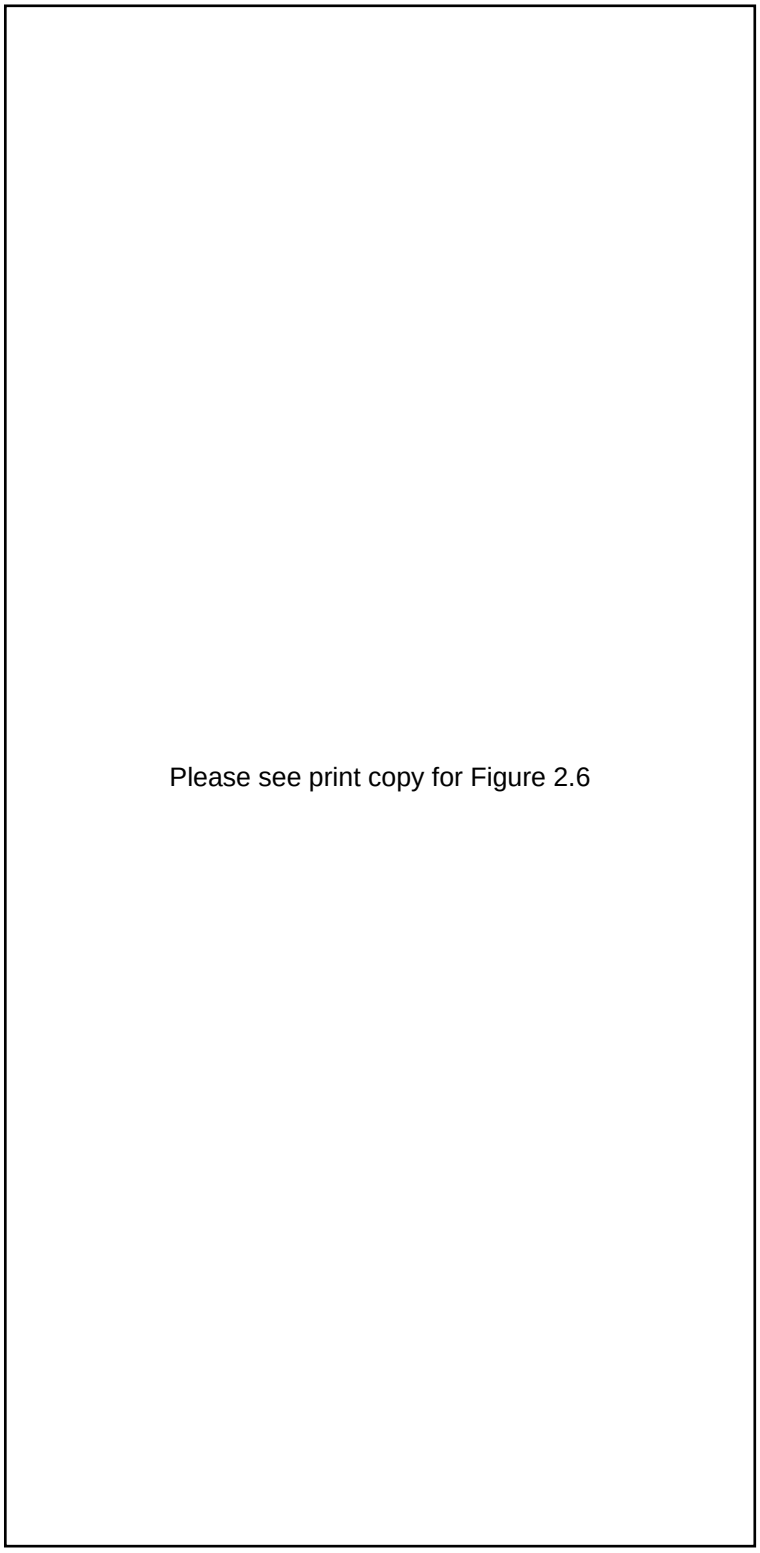

Figure 2.6 Redox potential v pH (stability) of compounds in a typical CASS (van Breeman, 1976).

2.2.4 Pyrite Oxidation influenced by Bacteria (Biological Oxidation)

After the initial formation of CASS from pyrite oxidation, the process continues with the presence of oxygen either from the atmosphere via root channels or through oxygen produced from a biological source.

Under reducing conditions, anaerobic bacteria (*Thiobacillus ferrooxidans* - TF) becomes an important step in determining the rate of pyrite oxidation process in low lying areas where the organic content is relatively high.

The rate of pyrite oxidation attributed to include biotic oxidation is the greatest in regions where organic matter exceeds 5%. Blunden & Indraratna (2001) developed a 3-Dimensional finite element model to analyse and contrast numerical and field data with an extension to include biotic oxidation. Rudens (2001) conducted a series of column experiments over a period of 56 days that emphasised the impact of TF on the rate of pyrite oxidation within Broughton Creek. The results from this were predicted to be comparable to soil conditions across the study area.

Dent (1986) and Chapman & Murphy (2000) further investigated the effect of bacteria acting as a catalyst in the process of pyrite oxidation. Favourable conditions for iron reducing bacteria include a soil pH less than 3.5, soil at a high temperature, and an organic content greater than 5%. In Blunden & Indraratna's Acid-3D model, the parameters were altered for the differing soil conditions. This is important when developing a model for predicting over spatially vast areas.

Rudens (2001) model was based on a 56 day field sample. The upper boundary of the pyritic layer was at a depth of –0.9m AHD. The influence of TF confirmed a strong correlation between anaerobic bacteria, pyrite oxidation, and organic matter.

2.3 Affects on Soil Parameters

The soil matrix consists of a number of chemical components that can indicate the presence of a CASS. Each of these parameters can be assessed individually or in combination to determine the toxicity or severity that a certain soil represents. Soil pH, Electrical Conductivity (EC), Total Actual Acidity (TAA) and the ratio of Chloride to Sulfate are seen as key indicators for determining the health of soil in an area with the geomorphological attributes of a PCASS (Dent & Dawson, 1996).

2.3.1 Soil pH

Indraratna *et al.* (1995) developed the inverse relationship between pH versus Al^{3+} and Fe^{3+} which explained that the more acid the soil, the higher the availability of Al^{3+} and Fe^{3+} ions released by the soil matrix (Figure 2.7). Using this relationship will help explain the significance that pH has in identifying and predicting CASS, but explaining pH will aid in understanding what is referred to as an acidic soil.

pH is a measurement of the acidity or basicity of a system. The pH scale is seen in Figure 2.8, which represents the typical values of pH in certain environments. pH is defined by the negative logarithm of the hydrogen ion concentration, or in soil, a measurement of hydrogen ions (H^+) where the higher the number of H^+ ions, the higher the acidity. Within the soil matrix pH often fluctuates down the profile due to factors such as amount of organic matter, soil type, groundwater level and quality, and cation exchange capacity within the soil.

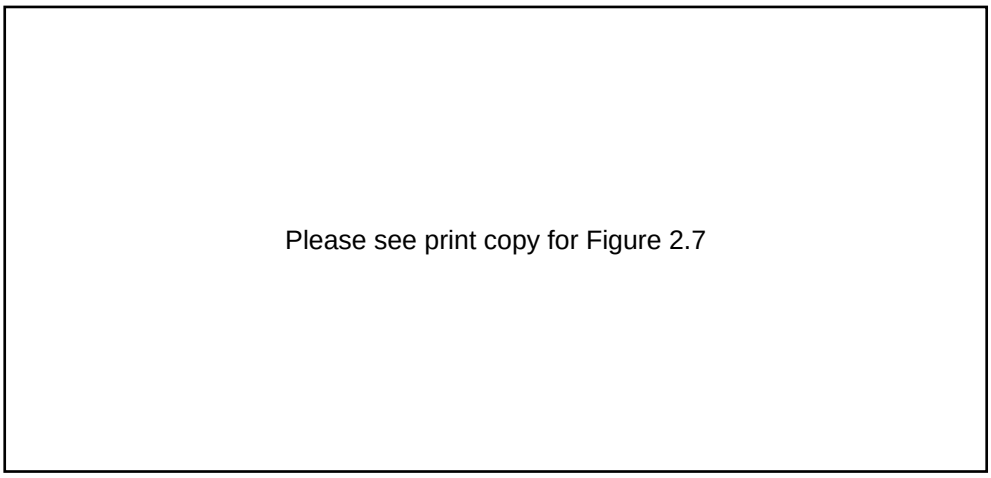

Figure 2.7 Relationship between pH and [Al^{3+}], and [Fe^{3+}] (Indraratna *et al.*, 1995).

ASRIS (3.6.3.4) developed a scale of pH acidity below 7.0. Each unit below pH of 7 is ten times more acidic. Table 2.3 divides the soil into four different acid classes based on pH (0.01M $CaCl_2$) method.

Soil Class	**pH**
Mildly Acidic soils:	pH 5.5 - 7.0
Moderately acidic soils:	pH 4.8 – 5.5
Highly acidic soils:	pH 4.3 - 4.8
Extremely acidic soils:	pH < 4.3

Table 2.3 ASRIS Soil pH Classification.

Using a datalogger or hand-held probe to measure pH is far more economical than collecting multiple soil bores throughout an area. For this project the aim was to determine the make up of soil pH in explaining acidity by using existing soil bores taken from other projects to show how data that was costly to obtain can be re-used.

In reviewing the available data, a number of methods were used for pH. The pH (water) method is less accurate than the pH Calcium Chloride (pH-$CaCl_2$) method and a general rule is to subtract 0.5-0.8 pH units from the pH (water) sample (Lawrie & Eldridge, 2002; DPI-Ag, 2004).

Dent and Dawson (1996), devised weights for a number of important parameters based on expert knowledge (Table 2.4) that would serve to identify CASS based on a number of selected soil and water parameters. Weights were assigned at certain intervals of a parameter to give a weighting of between 0 and 1 depending on the influence each parameter has in identifying a CASS.

Please see print copy for Table 2.4

Table 2.4 Dent & Dawson (1996) weights for Soil pH.

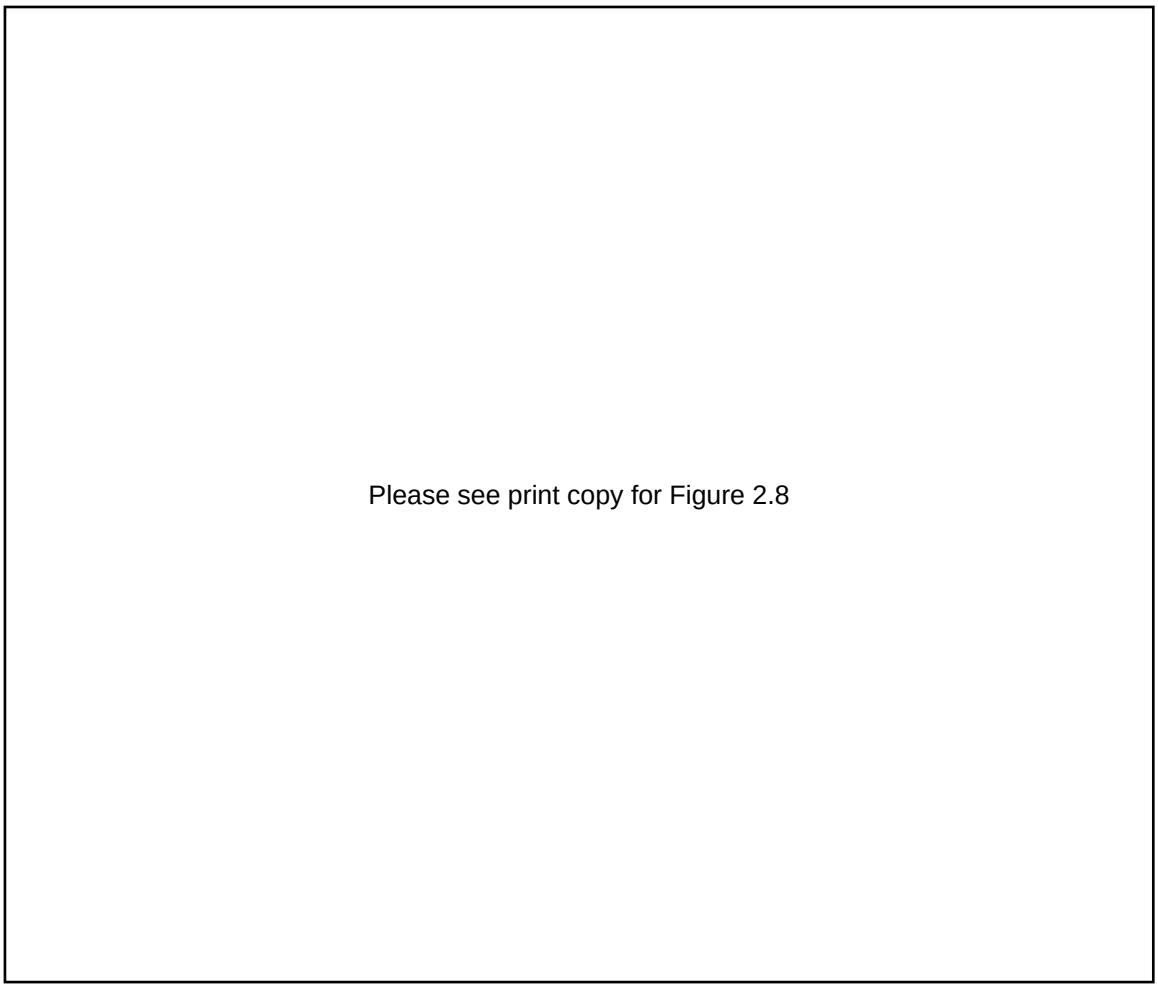

Figure 2.8 The pH log scale and examples of solutions found at each level (After DPI-Agriculture, 2004).

2.3.2 Total (Titratable) Actual Acidity

While soil pH is the most available measure of soil acidity it does not indicate the buffering potential of the matrix or enable the best predictions to be made regarding the quantity necessary to neutralise affected soil. As a result there are three main measurements taken in CASS to assess acidity of soil and predict any future acidity. Total Actual Acidity or Titratable Actual Acidity (TAA) is a measure of the soils acidity prior to oxidation of the sulfidic material.

TAA is usually determined by titration of a 1M KCl salt extract to generate a pH of 5.5 (Ahern *et al.*, 1998) although McElnea *et al.* (2000) suggested that where CASS was identified in low land rivers this should be raised to 6.5 to show its ability to combat

toxic elements in soil such as Aluminium. TAA is also described by Konsten *et al.* (1988) as the total amount of soil acidity freely available.

Titratable Peroxide Acidity (TPA) is a measure of the titratable acidity of a soil after peroxide oxidation. The original calculation by Konsten (1988), who coined the term Total Peroxide Acidity, was inaccurate because it did not include the full oxidation of pyrite (McElnea *et al.*, 2000)

McElnea *et al.* (2000) described TPA and TAA as 'Total' measures but coined them as 'Titratable' due to the significant omission of acid produced from pyrite oxidation. The remaining acidity generated by subtracting TAA from TPA is called Titratable Sulfidic Acidity (TSA). The Chromium reducible method of determining the Sulfidic content of soil (S_{CR}) was found by Ward *et al.*(2002) to be more accurate method than peroxide oxidation – combined acidity and sulfate (SPOCAS).

Calculating acidity in the soil can determine where the most hazardous sites are in the floodplain and this will also show the potential for a site to produce acid over a long time. Another measure of acidity is the content of Aluminium as a percentage of the Cation Exchangeable Capacity within a soil.

2.3.3 Cation Exchange Capacity

The Cation Exchange Capacity (CEC) is the ability of clay to absorb and exchange cations. As clay is negatively charged it usually attracts cations to the surface and maintains electro-neutrality. This is an electric double layer consisting of a negatively charged clay surface and positively charged cations attached. The soil particles where absorption and exchange takes place are known as exchange complexes. An exchange cation is defined as a cation held by electrostatic forces on a negatively charged colloidal surface.

CEC depends on the clay content, the type of clay and the total organic matter present. Clay content and CEC have a positively linear relationship in that the finer the soil particles the higher the CEC (e.g. clay). CASS are similar to a silty clay loam exhibiting a relatively moderate CEC.

In ASS, there are 4 mols of H+ produced with pyritic oxidation. The production of H+ replaces other cations in the complex and leaches the bases into the groundwater. A cation is replaced on the exchange complex stoichio-metrically. A clay surface has the ability to combine with many groundwater contaminants and pollutants due to its high electrical charge and surface area which acts as a purifying agent and buffer.

In CEC the most important bases exists as Na^{+}, K^{+}, Ca^{2+} and Mg^{2+} and the acidic cations consist of Al^{3+} and H^{+}.

2.3.4 Exchangeable Aluminium (Aluminium Saturation)

Exchangeable Aluminium (ExAl) is defined as the number of sites on the cation exchange occupied by Aluminium ions, relative to the valency of aluminium (Isbell, 2003). ExAl is calculated as a percentage of aluminium ions compared to CEC. It is an indicator for potential toxicity in terrestrial plants or in the case of ASS in aquatic habitats. Table 2.5 shows the ranges and toxicities in laboratory ExAl.

Please see print copy for Table 2.5

Table 2.5 Exchangeable Aluminium as Percentage of Cation Exchange Capacity (DNR, 2005).

High and/or toxic levels of ExAl occur on most soils with an acidity problem. If the percentage of Aluminium is greater than 5% it will affect most plants (Agricultural Bureau of South Australia, 2005). Depending on the vegetation and root mass of the species, this may directly affect the growth of the crop. However aluminium leaching into waterways severely affects aquatic species which may be linked to bacteria that

leads to diseases such as *Aphanomyces invadans* which forms the 'red spot' disease (Epizootic Ulcerative Syndrome). This causes skin ulceration and sometimes death (Plate 2.1).

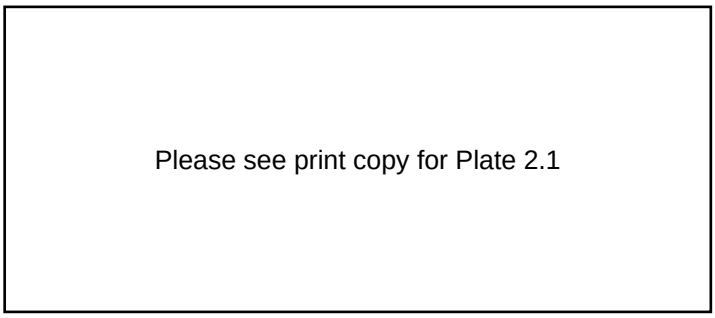

Plate 2.1 Red spot disease (Epizootic Ulcerative Syndrome) in silver perch (D. Callinan).

2.3.5 Chloride to Sulfate Ratio

The Chloride to Sulfate ratio is important for determining the presence of Sulfate produced as a result of pyrite oxidation. The lack of Chloride in a system indicates possible stagnation as well as a lack of influence from bicarbonates in downstream brackish water. Figure 2.9 shows a range of conditions and a range of Chloride to Sulfate ratios that correspond. For instance rainwater has a $Cl:SO_4$ of 5 and seawater has a $Cl:SO_4$ of 7. This is in contrast to a system which is seen as acidic. Dent & Dawson (1996) indicate a $Cl:SO_4$ of less than 4 indicates a potentially highly acidic environment. Figure 2.9 supports this showing that most acidic drains and creeks show a ratio of 3 or less.

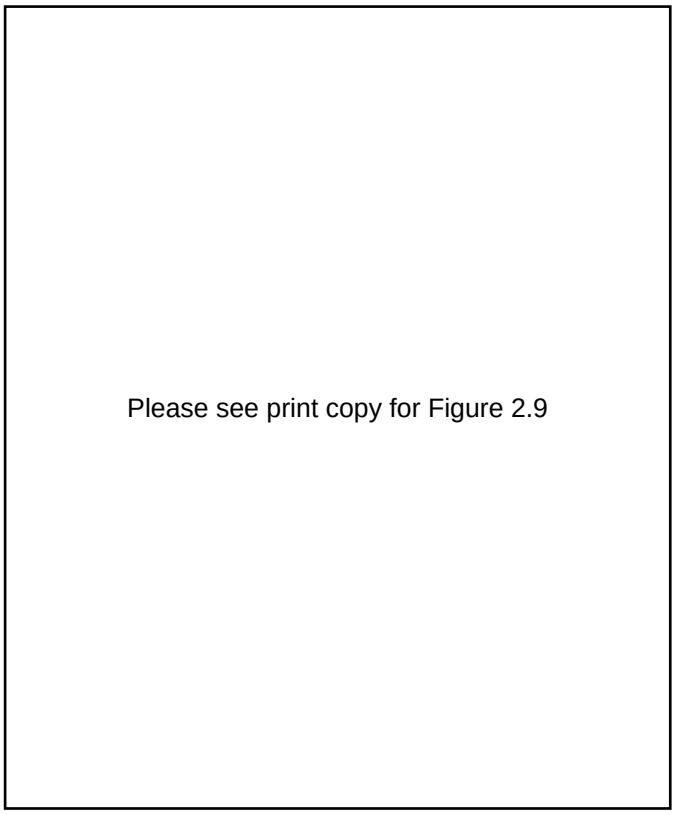

Figure 2.9 Chloride/sulfate ratios vs. log salinity for seawater, seawater diluted with increasing amount of precipitation, and acidified creeks and drains (Radke, 2000).

2.3.6 Electrical Conductivity and Salinity

Measuring Electrical Conductivity (EC) to determine salinity is a generally accepted and practiced technique (Webster, 2005). The presence of charged ionic species such as Na^{+}, Mg^{2+}, Ca^{2+}, K^{+}, Cl^{-}, SO_4^{2-} or $HCO_3^{-} + CO_3^{2-}$ in solution, enables water to conduct an electrical current and so salinity can be determined by the presence of these ionic species.

2.3.6.1 Electrical Conductivity in Soils

Measuring the Electrical Conductivity (EC) in soil can help determine its structure. Figure 2.10 shows the range of EC values found in typical soil. The larger the particle size and the more cohesion, the greater the EC, as is seen in clay which has an EC ranging from 10 mS/m to 1000 mS/m.

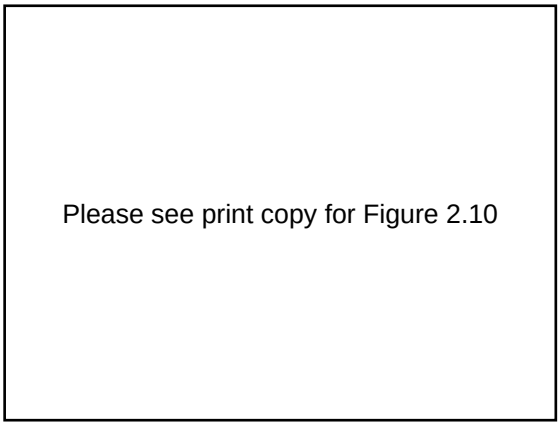

Figure 2.10 Electrical Conductivity of a soil and associated composition (Veris, 2005).

Veris (2005) developed a relatively inexpensive method for mapping EC over a broad area. The relationship between EC and soil structure makes it possible to apply this method to many applications including precision agriculture and other environmental management techniques. An EC cart containing two pairs of coulter-electrodes (Figure 2.11) is pulled behind a tractor and one electrode injects an electrical current into the soil while the other measures the voltage drop. The EC data is logged and geo-referenced. This process enables 20-40 points per hectare to be collected which makes it a more effective method for determining soil properties of a site rather than excavating a number of boreholes. The EC cart system measures accurately down to 90cm below the surface.

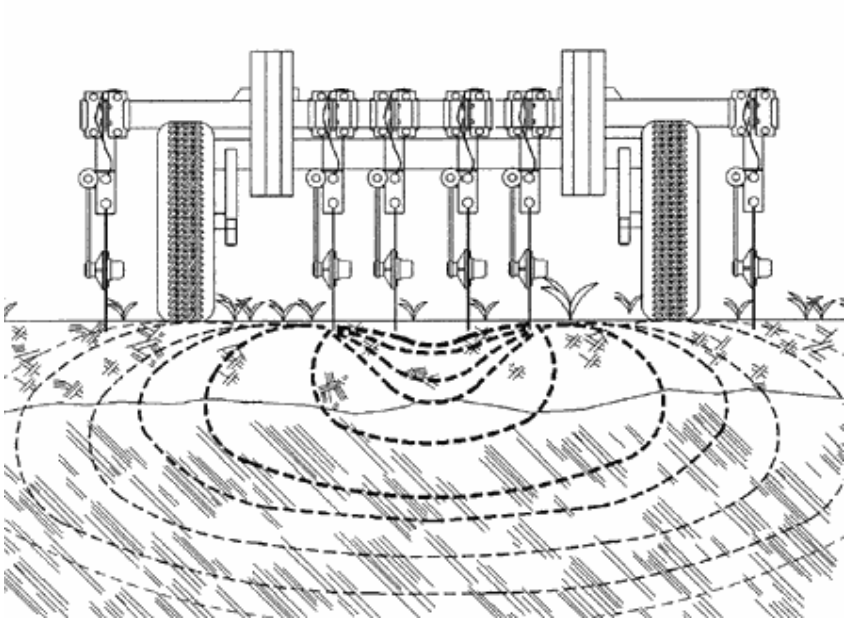

Figure 2.11 The Veris Sensor EC Cart.

More conventional accepted techniques include laboratory testing of EC where a 1:5 soil to water mix is measured out. The EC (1:5 soil to water) is used to determine the amount of soluble ions (salt) in soil (Abbott, 1985). This amount can be converted to represent EC in a saturated soil, which is the equivalent of soil conditions and represents its approximate salinity.

EC is measured in micro siemens per centimetre (μS/cm) or milli siemens per metre (mS/m). Low levels of natural salts found in waterways are vital for aquatic plants and animals to grow. However, when salts reach high levels in freshwater systems they can cause problems for aquatic ecosystems. Conductivity values can vary greatly within catchments due to the geology and soil types found. In some areas with flood mitigation drains and leaking floodgates, EC readings within the drains have been greater than a natural freshwater system. This becomes a major problem when the salts begin to seep into neighbouring land. However, Glamore (2003) and Ford (2002) modelled the effects of saline intrusion (seepage) into land adjacent to flood mitigation drains and found that salinity levels increase in land 5m adjacent to the drain but not enough to affect the growth rate of crops.

2.3.6.2 Soil Salinity

Taylor (1996) developed a guide for the effects on agricultural yield from different levels of salinity in Australian soils (Table 2.5). Many CASS environments consist of low lying coastal pastures (within floodplains) and any effects on plant growth is detrimental to the food supply that many graziers rely on for their stock.

Please see print copy for Table 2.6

Table 2.6 Values of Soil Salinity and Classes (After Taylor, 1996)

Table 2.6 is a basis to determine if a soil is highly saline and whether these ratings can be applied to land affected by CASS. Within the project area, it is predicted from past research (Glamore 2003; Ford 2002) that very little of the land within the top 2m of the soil will be greater than moderately saline.

2.3.6.3 Measuring Electrical Conductivity for Water Quality

A more economical method for measuring Electrical Conductivity (EC) is to determine the runoff EC or EC within a drain and around a given site. It is relatively inexpensive to log data over a long period of time and useful data can be collated, which was done while monitoring two upstream and downstream sites in Broughton Creek (Figure 2.12). However, there is little correlation between EC (1:5) measured in soil compared with EC (water) so this is just a rough indication of the current soil conditions at the site.

This existing data represents a saline environment that fluctuates over the year due to the rainfall, pyrite oxidation, and eventually acid run off. An influx of fresh water will reduce the EC towards 0 (i.e. to a completely fresh system).

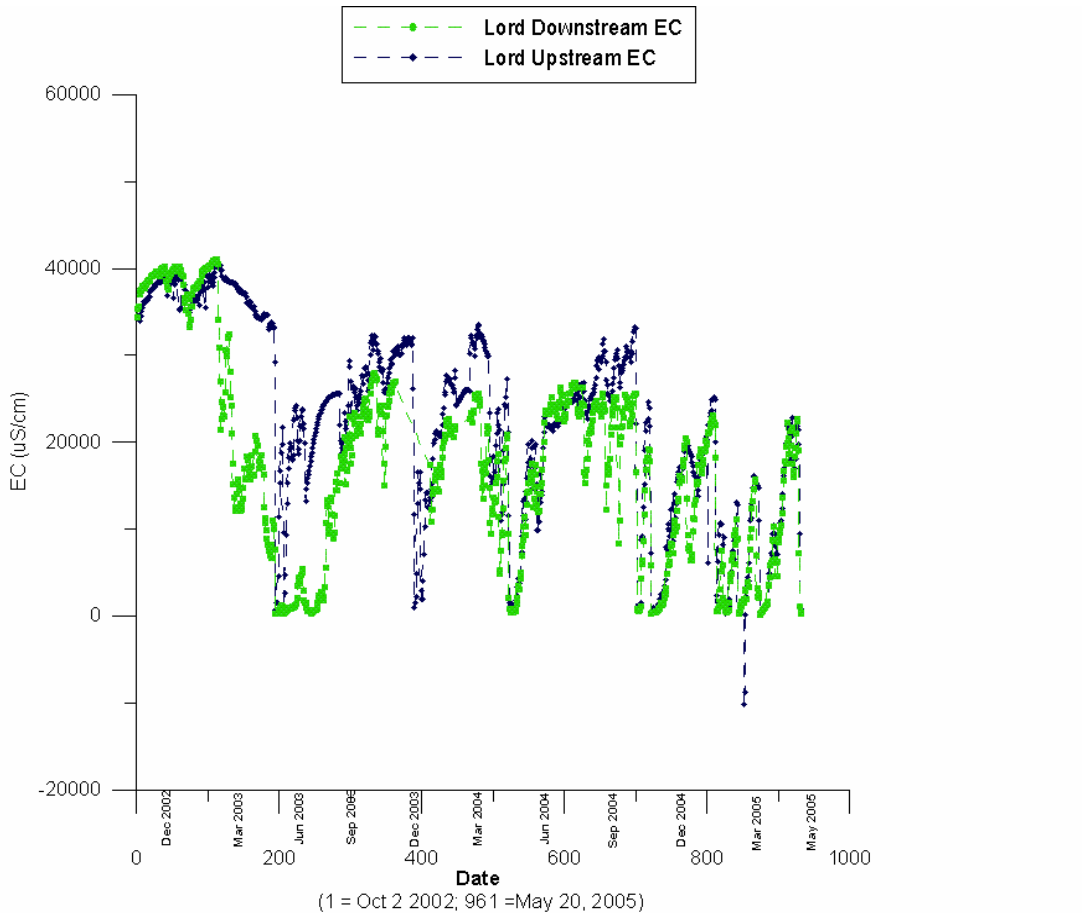

Figure 2.12 Electrical Conductivity at Broughton Creek Upstream and Downstream Datalogger (from October 2, 2002 to May 20, 2005).

ANZECC/ARMCANZ (2000) guidelines recommend that the values of freshwater conductivity be within 0 - 1,500 μS/cm, and lowland rivers to have an EC between 125-2200 (coastal rivers in the range 200-300 μS/cm). Estuarine Conductivity values

vary depending on the state of the tide and the amount of freshwater flowing into the system. Typical seawater has a conductivity value of approximately 45-50,000 μS /cm.

As aforementioned, comparing EC measured in water and in soil is quite different, however conclusions about the transport of salts through the soil can be inferred from an in-depth hydrological analysis as Glamore (2003) has shown. Conducting interpolation of soil EC over a wide area is a process that Veris (2005) and Electro Magnetic Surveying (Coram *et al.*, 2001) used with this information and data. A similar process of interpolation can be applied using a number of statistical and other geostatistical methods, which is the aim of this project.

2.3.6.4 Electromagnetic (EM) Surveys in determining EC

Within Australia, Electro Magnetic (EM) surveys have been used to determine outbreaks of salinity in actual and potential salt affected areas using either ground EM or airborne EM surveying techniques (Coram *et al.*, 2001). EM ground surveying techniques calculate bulk conductivity up to six metres down the soil profile, whereas airborne surveying can provide information up to 100m below the surface. In both techniques, but especially in airborne EM surveying, calibration is essential and it is necessary to install a number of boreholes on the ground for rectification. Bulk conductivity is proportional to EC, however the conversion process requires knowledge of the porosity of the soil and rocks in the saturated zone, and the volume (water content) of the unsaturated zone. Since porosity and volume are not usually recorded at a site, calibration of an EM survey should be done from porosity in the unsaturated zone and groundwater EC's in the saturated zone. One problem with EM surveys is that it doesn't differentiate between primary (or naturally occurring) and secondary salinity (human-induced salinity). Results generated from EM surveys require an expert hydrologist to differentiate between EC related to soil salinity and EC from groundwater conduits. Mapping salinity outbreaks is a process designed to monitor new outbreaks of salinity resulting from secondary salinity. The mapping of EC can be adapted to the mapping of CASS. ACASS are determined by the presence of an intermediate product commonly known as Jarosite. The mapping of Jarosite can provide a significant understanding of distribution throughout an area.

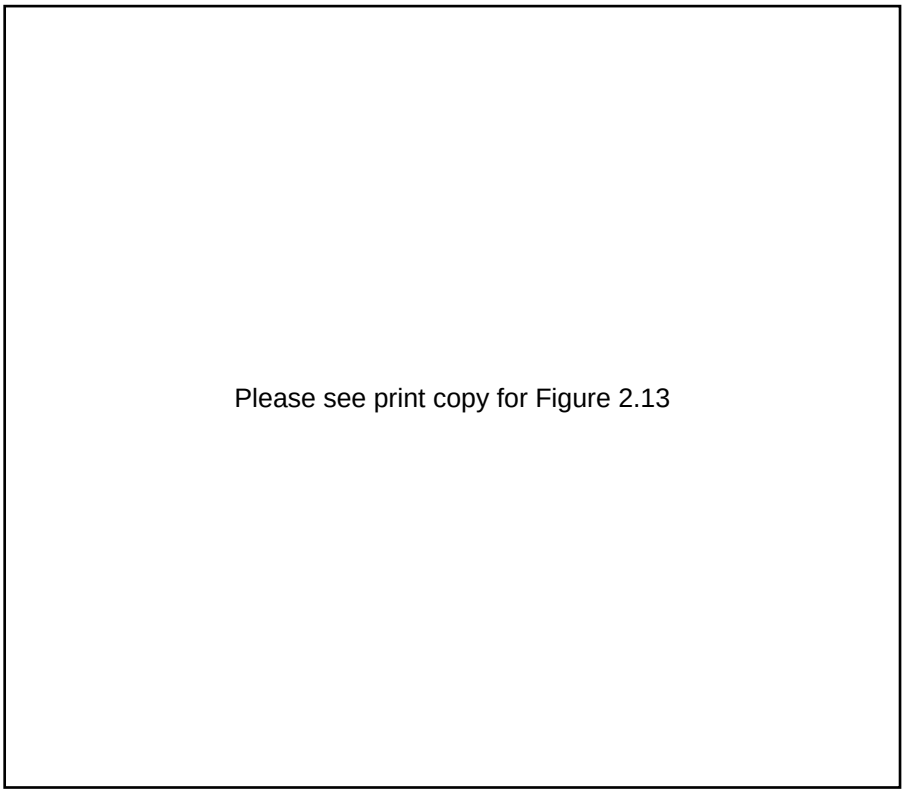

Figure 2.13. Electromagnetic Survey measuring Conductivity (Wynn, 2001).

2.3.7 Jarosite Level

A key element in determining whether a site is affected by ACASS is the presence or absence of Jarosite ($KFe_3(OH)_6(SO4)_2$). Jarosite is an intermediary product of the pyrite oxidation process and is normally found along root channels where the source of atmospheric oxygen is the greatest. Over time, the root channels affected by jarosite develop iron coatings and appear as red, mottled, peds.

Uncovering areas of Jarosite in the field is the best indication that the soil is an AASS. This is due to the conditions needed for Jarosite to form. Jarosite requires strong oxidising conditions, a potassium source, and a pH of approximately 3.7 or lower (Dear *et al.*, 2002). A soil pH of less than 4 under natural conditions in a floodplain is rare so experiencing a site with these conditions usually indicatives a CASS.

Land inundated by tidal water (i.e. elevation below sea level) or sea water would never have a chance to develop into an AASS. This doesn't mean there is no potential for an ASS to form. If grey black pyretic material is found, which usually occurs approximately 1m below the surface (but not always as this depends on elevation and landform structure), then there is a chance that in times of drought or a low water table, this layer could be exposed.

In the case study of Broughton Creek, the level of Jarosite varied over the floodplain but generally occurred at -0.2 to -0.9m AHD (Australian Height Datum). This was also seen on the Crookhaven-Shoalhaven floodplain, which is located to the south of the Broughton Creek floodplain, but bordered by the same river. What was noticeable in this case was the differences in drainage regimes, which over time can influence the amount and variability of Jarosite over the floodplain. For instance, from the samples taken in the Broughton Creek floodplain, significantly more was found than on the Crookhaven floodplain. Comparing the Digital Terrain Models (DTM's) of both sides can help clarify the reason behind this.

Plate 2.2 Oxidised PCASS, causing intermediate Jarosite product (Dear *et al.*, 2002)

2.3.8 Analysis of Variables

Each of the variables explained are important in identifying a CASS. In combination these variables can be used to determine an area that not only has a high probability of being a CASS, but can also show areas that will be the most toxic to the surrounding environment and where disturbance should be avoided at all costs. The development of high resolution elevation data has aided in the methods that can be used to depict where CASS are located.

2.4 Environmental Management Using Digital Terrain Models

Creating and using Digital Terrain Models (DTM) has become important for environmental managers who seek to have better control over the areas of land they are managing. A DTM enables them to perform analysis of the land to suit the needs of the manager. For instance, from a DTM a manager can understand changes in the slope, define the aspect, and create a hillshade of the terrain model to present to clients, or implement a new strategy that requires certain characteristics of land which must be held. Existing contour information is often very coarse within low lying areas. A DTM provides data on elevation to a much higher accuracy. This data enables certain parallels to be made with existing soils, vegetation and landform (morphology) information, to designate a severity rating to an area of land with the potential to cause environmental problems.

The use of DTM's in a Geographical Information Systems (GIS) has become a widespread environmental management and analysis tools. Much of the literature and research within Australia has focused on mapping and storing ecological data for use in the management of National Parks and identifying sensitive areas or species (CSIRO, 2000). However, there is a recent trend for local governments (e.g. Shoalhaven) to start accessing digital data for planning, knowledge, management, and to use GIS in daily data storage and operations.

GIS has become vitally important in the management of potential disasters throughout the world. For example GIS has been used in landslide assessment and planning in the USA (Hilton & Elioff, 2004), China (Xie *et al.*, 2003), Australia (Flentje *et al.* 2002),

and Jordan (Malkawi *et al.*, 2000), and in flood research and planning in China (Zhixiong *et al.* 2005), USA (Sands *et al.*, 2004), Canada (Ahmad, 2004) and Australia (Crowe *et al.*, 2003).

The use of DTM and GIS for direct CASS management is quite novel (Morgan *et al.*, 2003) and using this data to directly form a model that predicts severity is the aim of this thesis, and will aid in determining future management strategies to combat CASS runoff.

2.5 Acid Sulfate Soil Management Options – Broughton Creek

Glamore (2003, Section 2.6) assessed the historical development of CASS problems in the Broughton Creek floodplain from 1972 to 2003, including the initial discovery of CASS and the movement to understand the problem scientifically, including, the production of ground remediation works. As a result a local committee of relevant government and industry bodies was formed to implement some on-ground works and to promote further research. The group is known as the Shoalhaven River Acid Drainage Working Group, and it controls much of the progress in CASS management within the Shoalhaven.

2.5.1 Shoalhaven River Acid Drainage Working Group

The Shoalhaven River Acid Drainage Working Group (SRADWG) developed from the involvement of Council in cooperation with the University of Wollongong, NSW EPA, NSW Fisheries, NSW Agriculture, Department of Natural Resources (formally Department of Land and Water Conservation), volunteer groups (e.g. Shoalhaven River Watch) and landholders. SRADWG has been driving research in the Broughton Creek floodplain since the early 1990's when, after long periods of dry weather followed by heavy rain, extreme acidic conditions were identified in Broughton Creek. During the mid 1960's and early 1970's, a series of man made flood drainage channels were altered or constructed, in some cases dug deeper and wider, which have been a catalyst for pyrite oxidation and a significant problem of CASS runoff. (Pease, 1994; Buman, 1995; Blunden 2000; Glamore 2003).

Within Broughton Creek SRADWG has directed the research, planning, and strategies of a number of projects, with the goal of reducing the frequency, intensity, and duration of acid discharge in certain selected hot spot areas within the CASS Hotspot.

Remediation works that would influence the greatest area, while keeping in mind individual site characteristics focused on the management of ASS in the Broughton Creek floodplain. Some new and emerging remediation applications were implemented which included: (i) modified two way floodgates ('smartgates') for saline buffering, (ii) v-notched weirs and a self regulating tilting weir for groundwater manipulation, and (iii) deep sub-surface lime-fly ash injections and subsurface hydrated lime buffer strips for neutralising the acid produced.

All of these different management options have been applied to the Broughton Creek study site, to reduce the effects of acidic run off from CASS. Most achieved their aims, but some have been more practical than others as a remediation tool over a broader scale (i.e. shallow lime injection compared to the deep lime injection).

2.6 Management Applications in New South Wales, Australia

2.6.1 Two-Way Floodgates

Glamore's (2003) initial findings in the Broughton Creek floodplain showed that one-way floodgates:

(1) Increase acid production and transport
(2) Create acid reservoirs upstream of the floodgate
(3) Restrict fish passage into breeding areas
(4) Increase Aluminium and Iron flocculation
(5) Increase the growth of exotic flora species, and
(6) Deny the favourable process of tidal buffering

Glamore (2003) modified vertically lifting floodgates to permit two way tidal flow into an acidic drain. An initial investigation monitored data for a 12 month period and determined that tidal intrusion via modified floodgates dramatically improved water quality and generally reduced the impact of acid runoff from CASS. Based on these

findings the ‘smartgate’ was developed to optimise tidal intrusion. GIS was used as a method of optimising the amount of water into the flood mitigation drains.

Since the prototype gate was deemed a success in reinstating tidal buffering into drains in the Broughton Creek floodplain, four new smartgates were installed at new locations throughout the floodplain. The aim was to re-introduce tidal exchange, which buffered the acidic leachate from the flood mitigation drains and, increased ground water levels above the PCASS layer. The prototype smartgate was upgraded to a more permanent structure and standardised to the same advanced telemetry systems as the other Smartgate systems.

Since the smartgate was installed the water quality has shown pH falling under ANZECC guidelines for a relatively shorter period than in previous rainfall events (Figure 2.14). This technology can be applied to many areas and has more influence on the quality of water downstream than the other remediation methods.

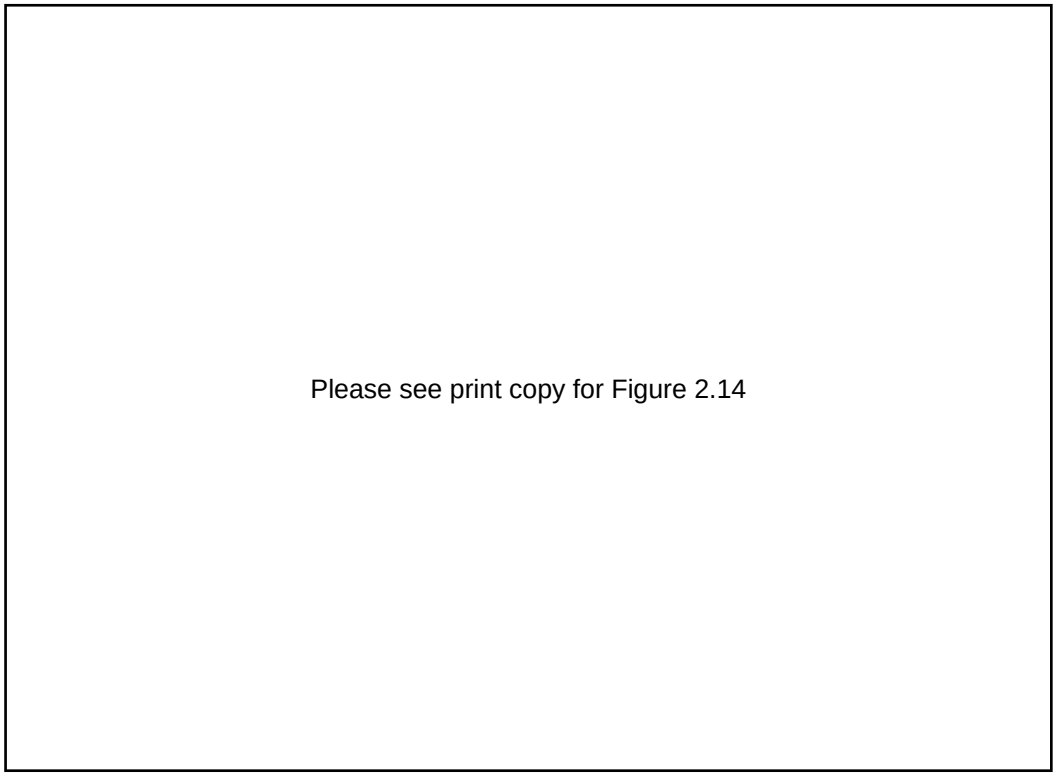

Figure 2.14: Upstream and Downstream: pH v Rainfall (from November 1, 2002 to May 20, 2005), (Morgan *et al.*, 2005)

2.6.2 V-notched Weirs

Blunden (2000) installed three v-notched weirs on the western side of the Broughton Creek floodplain to allow the groundwater table to be at, or above, the pyritic layer while adjusting to varying rainfall and climatic conditions during the year. This was to reduce pyrite oxidation process that would be initiated when the ground water table would fall below the pyritic layer.

It was discovered from baseline data that there was a strong hydraulic gradient between the ground and drain water at the field site. As a result, a number of weirs were implemented to elevate the ground water table.

Indraratna *et al.* (1999), estimated that in December 1997-February 1998 more than 1.3 tonnes of H_2SO_4 ha^{-1} was produced at the Berry field site. The hypothesis and realisation was that after installing the weirs this would be significantly reduced, which was seen at the conclusion of the study.

2.6.3 Self-Regulating Tilting Weirs

Weirs are another option for remediation that has been used in the Broughton Creek floodplain to reduce the time that PCASS is exposed to atmospheric oxygen. The inundation of the PASS layer over drier periods helps avoid pyrite oxidation from developing ACASS.

Weirs are likely to be a more successful management option in elevated parts of drained floodplains than low lying areas because there is a high risk from flooding. Some benefits of the weir include a reduction in poor quality water leaving the site. Agricultural productivity appears unaffected (Lawrie, 2001). The University of Wollongong (UoW) developed an automated structure known as a Self-Regulating Tilting Weir (Plate 2.3), and installed it in a flood mitigation drain on the western side of Broughton Creek floodplain (P6D8), after rigorous mathematical models validated its use as an option (Tularam, 2002).

Blunden (2000) reported that the installation of weirs at Broughton Creek promoted higher ground water elevation by reducing the influence of ground water drawdown from the drain. The lower hydraulic gradient established under the influence of the higher drain water level maintained by the weir, reduced the rate of acidic oxidation products from the ground water to the drain.

Plate 2.3 Self-regulating titling weir installed at drain P6D8, located on the western side of Broughton Creek.

2.6.4 Deep Subsurface Lime Injections

In low lying areas of the floodplain where weir operation may be impractical, the injection of lime slurry adjacent to major acid producing drains is one of a number of strategies in CASS amelioration. Despite altering drain water height in Broughton creek, pyritic oxidation was still occurring at low pH's, encouraged by *Thiobacillus ferrooxidans* bacteria.

A deep, local, lime-fly ash barrier injection was trialled by the University of Wollongong (Banasiak, 2004) aimed at neutralising the second peak of TAA, which at

120-130cm below the surface was 230 mol H^+/t. At the same time the injection of lime is to create an impermeable barrier to prevent further oxidation of the PCASS layer.

This site illustrated the positive effect the barrier had on the ground and drain water. Total Al concentrations in the ground water decreased slightly from the first stage of the injection and total Fe(3^+) concentrations in those holes influenced by them was less than those that were not. The success of the project needs to continue monitoring now that the barrier has been fully installed.

2.6.5 Subsurface Liming Buffer Strips

Lawrie (2003) conducted a lime ripping experiment on a flood mitigation drain that drains into Broughton Creek. The trial was to assess the viability of injecting lime slurry into the top 40-50cm of soil across a larger area (along the banks of the drains). It involved using a holding tank connected to 4 tines and a motorised pump, and an indicator to control the rate of injection. Nine tonnes of hydrated lime (CaOH concentration >90%) mixed with 27,000 litres of water in three separate applications was used. The rate of lime: water for each application was 22.68%, 27.18% and 25%. The drain was 680m by 5m wide and the injection took place on both sides of the drain.

The aim was to neutralise previously deposited acid in the soil, thereby preventing it from entering the flood mitigation drains and into Broughton Creek. The preliminary soil tests indicated a TAA of 370 mol H^+/tonne in the 60-70cm range below the surface, or the first peak of the bimodal distribution of TAA. These samples were included in the soil database used for this project.

2.7 Planning Remediation Strategies

All of these techniques have proved to be quite successful but choosing the right remediation method for the right area depends on a number of factors which will be addressed throughout this thesis. They include:

1. Selection of study area with the potential to be affected by CASS
2. Collection of available soil data and other information for that area

3. Analysis of available soil data
4. Need to obtain further data to fill in missing information about an area (e.g. further soil sampling)
5. Analysis and organisation of soil data in GIS, comparing and contrasting to DTM
6. Development of statistical models to determine the distribution of soil properties over an area.
7. Determination of the most severe areas or areas of need of remediation strategies
8. Choosing the most appropriate strategy for the study area based on severity in the floodplain and comparing it to the DTM.

Before a model that will provide a severity index over a floodplain can be created, the methods used in its development will be discussed. Specifically, a typical GIS will be explained in Chapter 3 that will show how information on soil can be analysed and developed into statistically robust models.

Chapter 3 Methodology

3.1 GIS Defined

A Geographic Information System is defined as a computer system for capturing, storing, checking, integrating, manipulating, analysing and displaying data related to positions on the Earth's surface. Typically, a GIS is used for handling maps of one kind or another. These might be represented as several different layers where each one contains data about a particular kind of feature (e.g. roads). Each feature is linked to a position on the graphical image of a map.

Layers of data are organised within a GIS in order to perform further statistical analysis (e.g. determining slope from an elevation layer). GIS is used in a variety of fields, the majority being in government, planning, public utility management, environmental, natural resource management, engineering, business, marketing, and distribution (FOLDOC, 1993).

A GIS can also collate large data sets to maintain and continually update a data set. In this project a GIS was used to organise, analyse and produce graphical output.

GIS have become increasingly important in the management of remote and spatially diverse areas. Since its beginnings in the early 1960's, as a research activity of the University of Washington (Nysteun, Tobler, Bunge, Berry) year, and the still existing Canada Geographic Information System, to the current programs used today, such as commercially available and licensed software including ArcGIS, ArcInfo, MapInfo, Manifold, Maptitude and other free GIS programs, GIS has developed from a concept to a viably useful application for environmental management.

The application of geospatial technologies such as GIS has become important when managing the complex nature of soil based problems (Kollias *et al.*, 1999). GIS enables current and future natural resource managers to have access to the tools necessary to answer questions about environmental problems that change across the landscape. Such technology enables complex statistical, geographical, environmental, chemical and physical information to be contained on a certain sized parcel of land. This is based on

the resolution of the raster based imagery for an area, limited only by the equipment (Aircraft, computer hardware) that is capturing the data. This data, collected as a series of cells, can be used to create predictive models of typical landscapes using GIS combined with suitable statistical software packages.

Looking at the development of GIS is important to show how it will be useful development CASS severity.

3.2 History of GIS

GIS has developed into a widely used management tool for spatial applications, albeit subject to change like any computer related phenomenon as older systems become outdated. A number of key organisations and individuals contributed to the development of GIS, including the Canadian Geographic Information System (CGIS) and ESRI. A GIS history project continues to be developed by the National Center for Geographic Information and Analysis at the University of Buffalo (http://www.geog.buffalo.edu/ncgia/gishist/). ESRI is still one of the lead developers in GIS platforms, their latest product released is ArcGIS.

3.3 ArcGIS

ArcGIS has been developed by ESRI from previous GIS programs such as ArcView 3.2 and ArcInfo. It is one of the more well known and widely used GIS products in the academic environment and will be used in this project. Within ArcGIS a Geodatabase helps manage the data enabling many users to access, analyse, and edit data simultaneously (see Appendices A13).

3.4 GIS in Soil Science

The application of GIS to soil science began in the 1960's when the Canada Land Inventory (CLI) created maps to classify the capability of soil for agriculture. During this period there was an increase in environmental awareness throughout the community. Management of the environment was becoming more important and a

number of commercial and governmental industries were developed to deal with this increase in demand (MacDonald and Kloosterman, 1984).

For example, ESRI primarily developed GIS software to be used in the forestry sector before applying its software to a variety of other spatially diverse environmental applications such as soil science. This helped to expand its operations from a local scale in California to a worldwide scale. ESRI's software has reduced the cost enough to enable use across a broader range of applications at a far lower cost than when originally distributed, however the cost is still significant and as a result many other smaller GIS suppliers such as MapInfo, Manifold and free GIS such as GRASS have emerged.

By the 1960's the soil survey discipline had developed into a recognised methodology, organising the science with a taxonomy system for classifying soils. For example, the Canadian Land Resources Research Centre (LRRC) was actively working to characterize its land resources with an emphasis on soils in agricultural regions. The data had become so large that the National Committee on Soil Survey recommended they be organised and stored within a computerised system. This began the digital soil data movement and introduction of GIS to soil management.

3.5 Case Studies

"Data constitute the raw material of scientific understanding. The World Data Centre system works to guarantee access to solar, geophysical and related environmental data. It serves the whole scientific community by assembling, scrutinizing, organizing and disseminating data and information" (ICSU World Data Centre System, 2004).

Soil information systems throughout the World have been formed to act as central repositories for data. Case studies from the LRRC to current day governmental organisations can explain how data is maintained throughout the world. Examples of three main regions (Australia, Spain, and Jordon) other than the original developers of GIS (USA/Canada) also show the development and use of GIS in soil management.

3.5.1 CANSIS

The Land Resources Research Centre (LRRC) is the Research Branch of Agriculture Canada, developed by the Canadian Soil Information System (CANSIS) in 1972. From 1975 to 1986 it was run with computer programs written in-house by LRRC. LRRC, with CANSIS was a world leader in the field of spatial representation of digital soil data. However, as other agencies developed their own systems, often based on commercial GIS software, it became difficult to exchange information. LRRC's original custom designed software was not compatible with the principal types of commercial software that were overtaking the market. It was no longer cost-effective to maintain the CANSIS style-software leading LRRC to adopt a commercially developed system to manage their data.

CANSIS merged data into the commercially available ARC/INFO software from ESRI in 1986. In 1994 CANSIS data was converted to hypertext, in order to be used in one of the first federally operated GIS websites in the World.

CANSIS was the original provider of digital soil map data in Canada up until recently, due to the increase in availability and reduced cost of commercial GIS software. Many provinces are developing higher resolution GIS capability. The greater access of GIS software to developers is the basis of the GIS revolution.

3.5.2 NASIS: National Soil Information System

NASIS (the National Soil Information System) was implemented more recently than CANSIS, and like CANSIS, acts as a tool that helps to create and maintain soil surveys within the USA. NASIS was formed in an attempt to standardize soil data throughout the USA. NASIS incorporates database technology to provide an automated means for storing all information about soil surveys (Figure 3.1). NASIS maintains the hierarchical structure of soil survey data, through the use of table-oriented editors. It is set to be the replacement for the USA's State Soil Survey Database (SSSD) program. The structure of NASIS is such that the data is organized in a way that will filter through the current land resource hierarchy (Figure 3.2) to be in a standard, accurate and secure form.

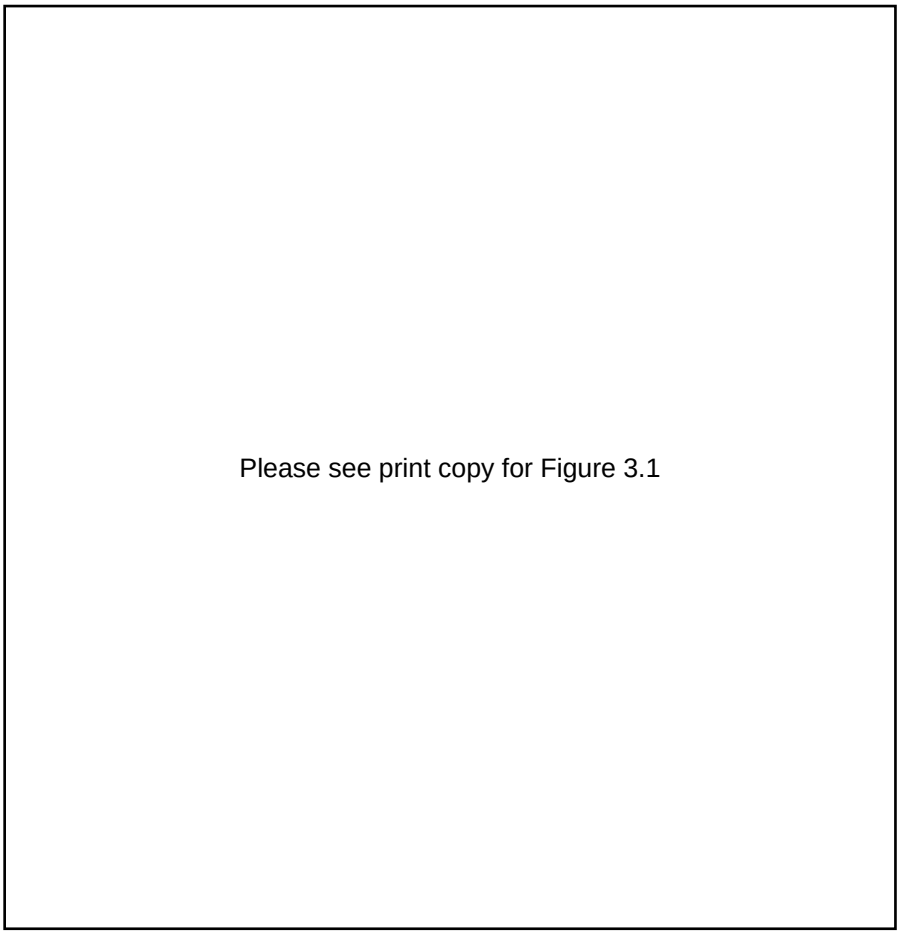

Figure 3.1: Structure of how NASIS organizes data to be used for other applications (USDA-NRCS, 2001).

The NASIS gives the process of collecting data from the Land Resource Region through to the pedon or profile some structure in sharing information. The major interfaces used by natural resource managers and stakeholders are the State Soil Geographic (STATSGO) database and the Soil Survey Geographic (SSURGO) database, which are being merged into the Soil Data Mart.

3.5.2.1 State Soil Geographic (STATSGO) Database

Soil maps for the State STATSGO database are produced by generalising detailed soil survey data. The mapping scale for STATSGO is 1:250,000 (with the exception of Alaska, which is 1:1,000,000). The level of mapping is designed to be used for broad planning and management uses covering state, regional, and multi-state areas.

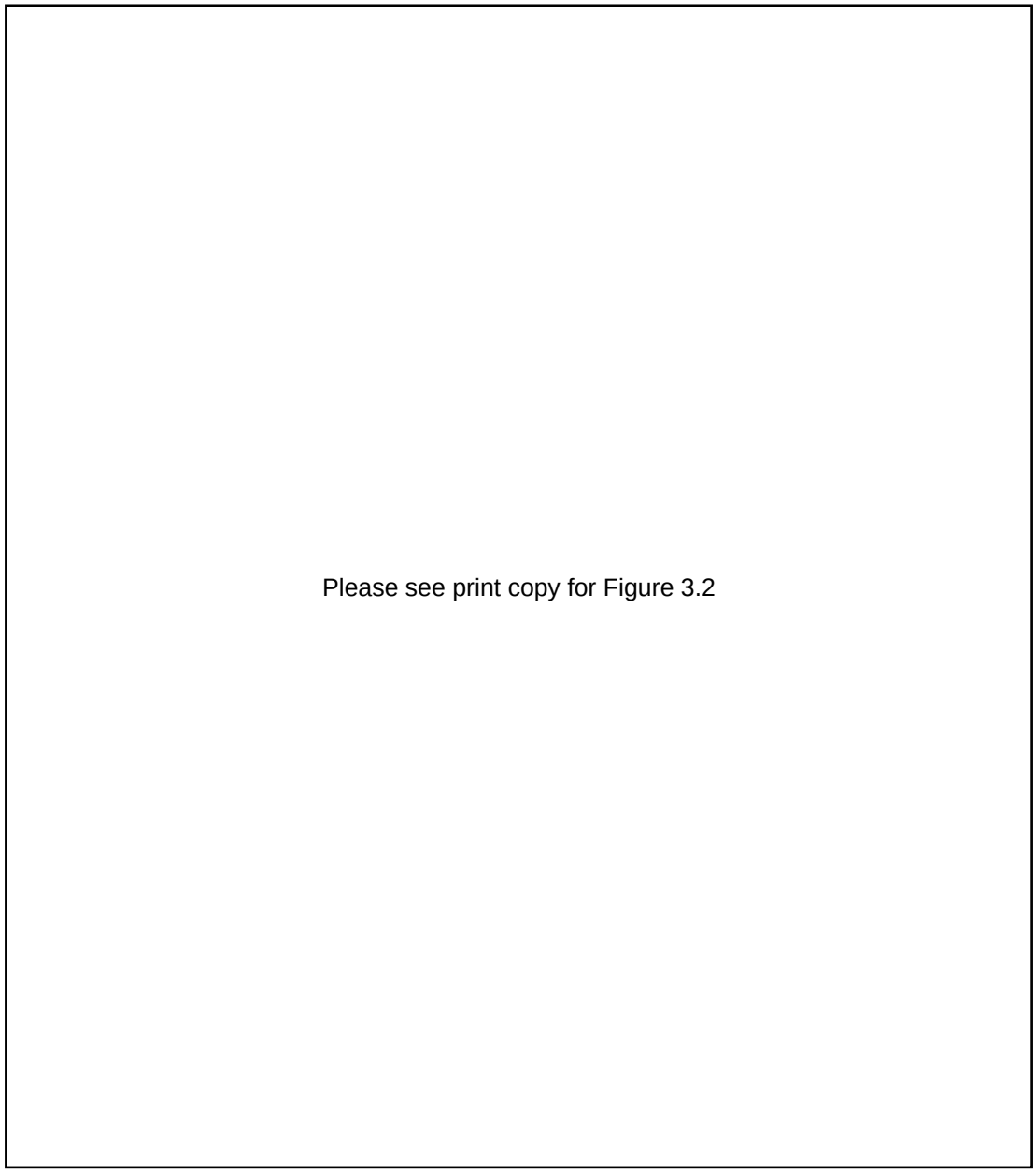

Figure 3.2 Hierarchical structure of soil data organisation in the USA (after USDA-NRCS, 2001).

3.5.2.2 Soil Survey Geographic (SSURGO) Database

The Natural Resources Conservation Service (NRCS) has developed the Soil Data Mart in an effort to improve soil data distribution. Currently, the national SSURGO Website is the primary source of on-line soil data in the USA. The Soil Data Mart will supersede

the National SSURGO Website. Completion of the SSURGO data digitizing is scheduled for 2008.

3.5.2.3 Soil Data Mart

The Soil Data Mart will eventually take over the SSURGO website and provide the following benefits to users of digital soil data:

- Determine where soil tabular and spatial data is available.
- Download data for one soil survey area at a time.
- Download a template Microsoft Access database for working with downloaded data.
- Generate a variety of reports for one soil survey area at a time.
- Find contacts for information about soil data for a particular state.

And

- Have a subscriber-based system for ease of updating.

Overall data is provided for SSURGO/Soils Data Mart by NASIS. NASIS main goals is to:

- Provide a dynamic and flexible system
- Support conservation assistance through improved data quality
- Provide improved automated map unit management, which involves, the correlation of map units in an on-going survey, the joining of map units between survey areas, sharing of map units between projects (e.g., and MLRA survey and county subsets), maintenance of multiple map unit legends (survey area, state, MLRA) and maintenance of complete correlation records and map unit data for map units correlated out of the survey area legend.

NASIS and CANSIS have been designed primarily for soil data standardisation throughout the USA and Canada respectively. Such systems have been the model for countries that lack a standardised structure. The World Data Centre, European Soil Information System (EUSIS), UNEP's Soil and Terrain database (SOTER) and the

Australian Soil Resources Information System (ASRIS) aim to produce soil data that is of the highest quality in a standardised format using similar structured systems as CANSIS and NASIS.

3.5.3 Europe: European Soil Information System (EUSIS)

The European Soil Information System (EUSIS) consists of a geographic data set, a semantic data set, a soil profile analytical database, a soil hydraulic parameter database, and a knowledge database in a fully integrated GIS within the European Geographic Information Infrastructure (EGII). It is part of the Agriculture and Regional Information Systems Unit (ARIS) of the Space Applications Institute (SAI). It is continuously maintained, updated and improved by a large network of national soil surveys operating under the umbrella of the European Soil Bureau.

The EUSIS's aim is the establishment of a common framework for Europe and to provide standardised soil information. The EUSIS has provided member nations in Europe with a tool comparable to other established systems in the United States (NASIS) and Canada (CANSIS). EUSIS, is also fully compatible with the FAO's World Soils and Terrain database, existing in the World Data Centre.

The development of this soil information system is continuing with the extension of the coverage to the Commonwealth of Independent States (CIS) (former Soviet Union) and to the Mediterranean basin. The main aim is the establishment of a common framework for the sustainable use of the soil resources in Europe, including the Mediterranean basin.

Within the EU region, Spain also has its own soil database on a nationwide scale for selected soils within the countries boundaries- SEIS (Sistema Espanol de Informacion de Suelos OR the Spanish Soil Information System).

3.5.4 SOTER: SOil and TErrain Database

SOTER uses current and emerging information technology to establish a World Soils and Terrain Database, containing digitized map units and their attribute data. This data handling system provides necessary data for the improvement of mapping and monitoring of changes of world soil and terrain resources.

The UNEP funded SOTER database for Jordan was compiled by the Soil Survey Section of the Forestry and Management Department of the Ministry of Agriculture, Amman. It comprises a standardised system of storing spatial and soil attribute data at a scale of 1:500,000.

SOTER also developed a spatial and attribute soil and terrain database at 1:1,000,000 scale for Argentina, Brazil and Uruguay. In Uruguay, the SOTER database has been used to prepare a water erosion risk assessment in the framework of UNEP's Global Environmental Outlook pilot project.

3.5.5 Australia: Australian Soil Resources Information System (ASRIS)

The ASRIS soil profile database contains over 160,000 soil profile descriptions in a standard format (Soil Information Transfer and Evaluation system - the SITES protocol), compiled from data held by State and Territory agencies, and CSIRO. The Australian Collaborative Land Evaluation Program (ACLEP) released the latest update of ASRIS in 2004, which included 5000 profiles with full quality assurance, and aim to have 10,000 profiles with similar standards by 2006.

The 164,030 soil profiles available were included if they met the SITES requirements, which required data to have:

- Map coordinates defining the site location
- At least one observation about the site which needed to include information on the horizon and sample taken

The majority of samples included had morphological descriptions of the profile and over 80% (131,605) had data on soil pH. Chemical and physical soil properties were

limited. For example, saturated conductivity was only measured in 0.3% of the samples included.

Data quality is a major issue with centralised databases and sufficient metadata often excludes samples from such systems. The ASRIS database contains information over a 50-year period. This data was collected by different agencies for different purposes and so there is little homogeneity of the data throughout the database. For example, the DNR collected a number of samples throughout New South Wales focusing on CASS soil parameters. Samples included presence/absence of ACASS, pH, and other morphological features. Another study on saline soils performed in Western NSW contained different fields such as sodicity and permanent wilting capacity, but may not have information relevant to acid soils, and so this makes it hard to compare data across a greater geographical area.

Other problems when collecting data over a large area includes: the variation in the description of the data (due to different surveyors interpretations), differences in laboratory methods when reporting on chemical data, and inconsistencies in reporting data resulting from a lack of understanding or ignorance of soil taxonomies. This makes reusing data for new research difficult, time consuming, and often cost ineffective.

Although soil data management is moving towards centralised depositories, state agencies still play a major role in coordinating database collection and maintenance. New South Wales, through the State government department have maintained an on-line soil database, accessible to all the public, the Soil and Land Information System (SALIS) database.

3.5.6 Region Specific (New South Wales, Australia): Soil and Land Information System (SALIS)

The ASRIS is the preferred central repository for maintaining soil data in Australia. However, state databases often contain more detailed local level information, and usually won't automatically be included into the national database.

The Soil And Land Information System (SALIS) of New South Wales is a database that contains descriptions of soils, landscapes, and other geographic features from across NSW. The NSW Department of Land and Water Conservation (now DNR) is the

custodian of this information. SALIS has been developed as a centralised repository of information, eliminating duplication of effort by storing all of NSW's soil and land information in a single system. SPADE is the on-line GIS interface that displays a map of NSW showing what soil data exists. Metadata about the boreholes is included on the web site, providing information about what was sampled at that location.

The soil information contained within SALIS includes physical, chemical, and morphological attributes collected at over 58,000 points across NSW (as of March 2006), this being continually updated. SALIS is continually being expanded using soil data cards, digital data files, photographic images, and maps.

3.5.7 Analysis of Databases

Maintaining a good structure from a local collection level to a national and international level will aid in making soil samples available for future generations and for future projects and all of these databases aim to achieve this. The problems with a lot of these databases are that although they aim to collect every single soil sample available in a region, in reality they only collect data that is relatively easy to collate. Soil data from consultant reports, university research projects or even from an individual landholder will most likely not be included. SALIS attempts to address this problem by having a system where an individual could provide information to the database in return for greater access to the available soil information on the site (i.e. descriptions would be more detailed with this exchange). Continual improvements to these databases should address this inadequacy in collecting information that is available and put in place future systems that will ensure all new information is forwarded to the central collection body. For this project SALIS was very useful, however there were some inadequacies that was addressed through using a multitude of sources to form the database used.

3.6 Visual Data Reference

Digital soil data becomes more valuable if there is a base reference layer to visually compare and reference points to. A typical project using soil data, will also use topographic imagery such as an aerial photograph. High-resolution digital elevation data is becoming more economically viable and may also be integrated into a project. Geographical layers such as geology, vegetation, and water can be referenced to digital

imagery such as aerial photographs or Digital Terrain Models (DTM's). The process of obtaining a DTM depends on the accuracy trying to be achieved, the budget in hand, and the equipment available at the time. Shoalhaven City Council used Airborne Laser Scanning to produce a more accurate DTM.

3.7 Airborne Laser Scanning (ALS)

Capturing change in elevation over a floodplain is quite difficult using standard topographic maps that provide elevation increments of 1m and above. CASS land is generally located beneath 4m in Elevation and finding data that describes the terrain beneath 4m is often non-existent or very primitive. In this project, accurate elevation was sought after in order to develop a model based on the most accurate available data. Airborne Laser Scanning is a tool that can capture this information, at a far lower cost than traditional survey methods.

3.7.1 History of ALS

Airborne Laser Scanning (ALS) or Light Detection and Ranging (LIDAR) was first used in 1993 by Geodan Geodesie B.V. as a cheaper alternate to collecting spatial information than traditional survey methods and photogrammetry. ALS has become important in creating DTM's with high precision at a far lower cost to other methods. In dealing with CASS problems and other environmental problems in low-lying areas, Shoalhaven City Council employed ALS for obtaining detailed survey information within budgetary constraints. The ALS of Broughton Creek operated as a pilot project in cooperation with AAM Geodan, a joint venture with AAM surveys and Geodan Geodesie B.V. The aim was to determine the effectiveness of using ALS for coastal and environmental management by testing the accuracy of ground level points against traditionally surveyed points. The effectiveness of the ALS was tested via a case study on Broughton Creek Floodplain. In gathering survey data via the ALS, the accuracy to which CASS is managed within the Broughton Creek Floodplain increased. The DTM produced from the ALS spot height points was used to delineate the drainage patterns of the Broughton Creek Floodplain, and to determine locations of remediation works. The

DTM was also rectified to a 1:25,000 aerial photograph (Figure 3.3) to obtain digital photographs of the study area. The image was of 0.5m pixel resolution (Figure 3.5).

3.7.2 ALS Process

ALS is a process of collecting point height data using aircraft that has historically been used for photogrammetry. In ALS, advanced Global Position System (GPS) satellites determine the location of the aircraft in respect to the ground location. A GPS receiver in the aircraft is referenced to the GPS positioning satellites above and GPS receivers on the ground. The attitude of the aircraft is determined by the inertial measurement unit of the plane and referenced in the process. During the flight, of which time is accurately measured for rectification purposes, a laser scanner emits laser beams of wavelength 1.047 microns, and collects the reflections. Some scanners can record the beam divergence of the scan, which can either be a wide or narrow beam and others use the pulse system, which measures ground heights (last pulse) and objects above the ground (first pulse). This method has been employed in determining the height of buildings in urban areas (Tao & Yasouka, 2002). The flying height of the aircraft is generally around 900 metres above the ground and the scan width is approximately 500 metres wide, dependent on the equipment used. There are up to twenty scanlines every second and up to 250 measurements every scanline, therefore making it possible to collect 300,000 points per minute.

In processing the data the attitude of the aircraft and the range of the laser scanners are used to produce the DTM based on ellipsoidal heights. The local geoid-spheroid separation factor is necessary to convert the points to orthometric heights. In the process of creating the DTM of the Broughton Creek Floodplain, the Australian Height Datum (AHD) heights were obtained by applying a correction with geoid information using AusGeoid 93 or AusGeoid 98.

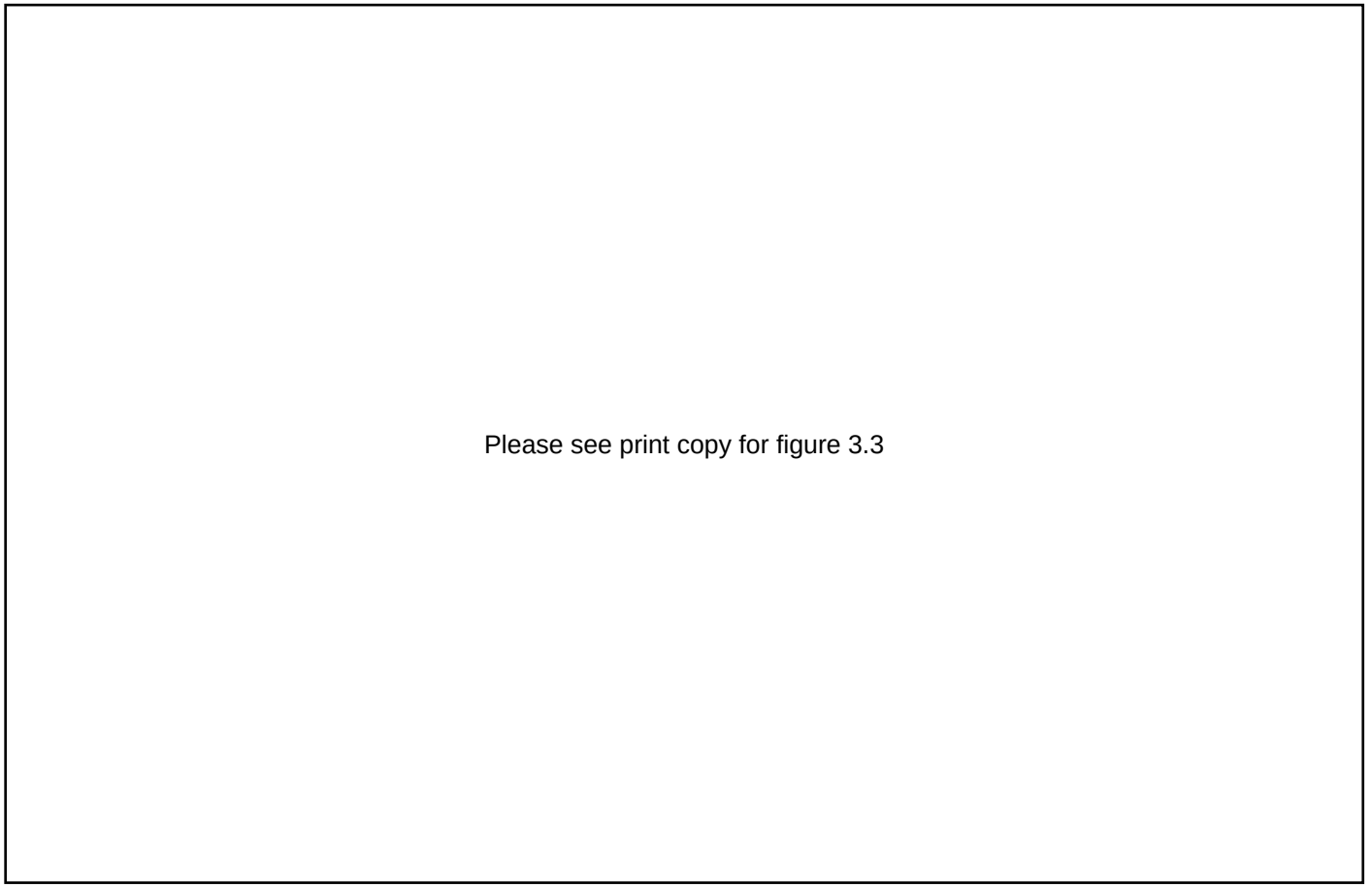

Figure 3.3 Aerial Photograph of Broughton Creek Floodplain: Elevation≤ 4m (Data Source: Shoalhaven City Council, 2002).

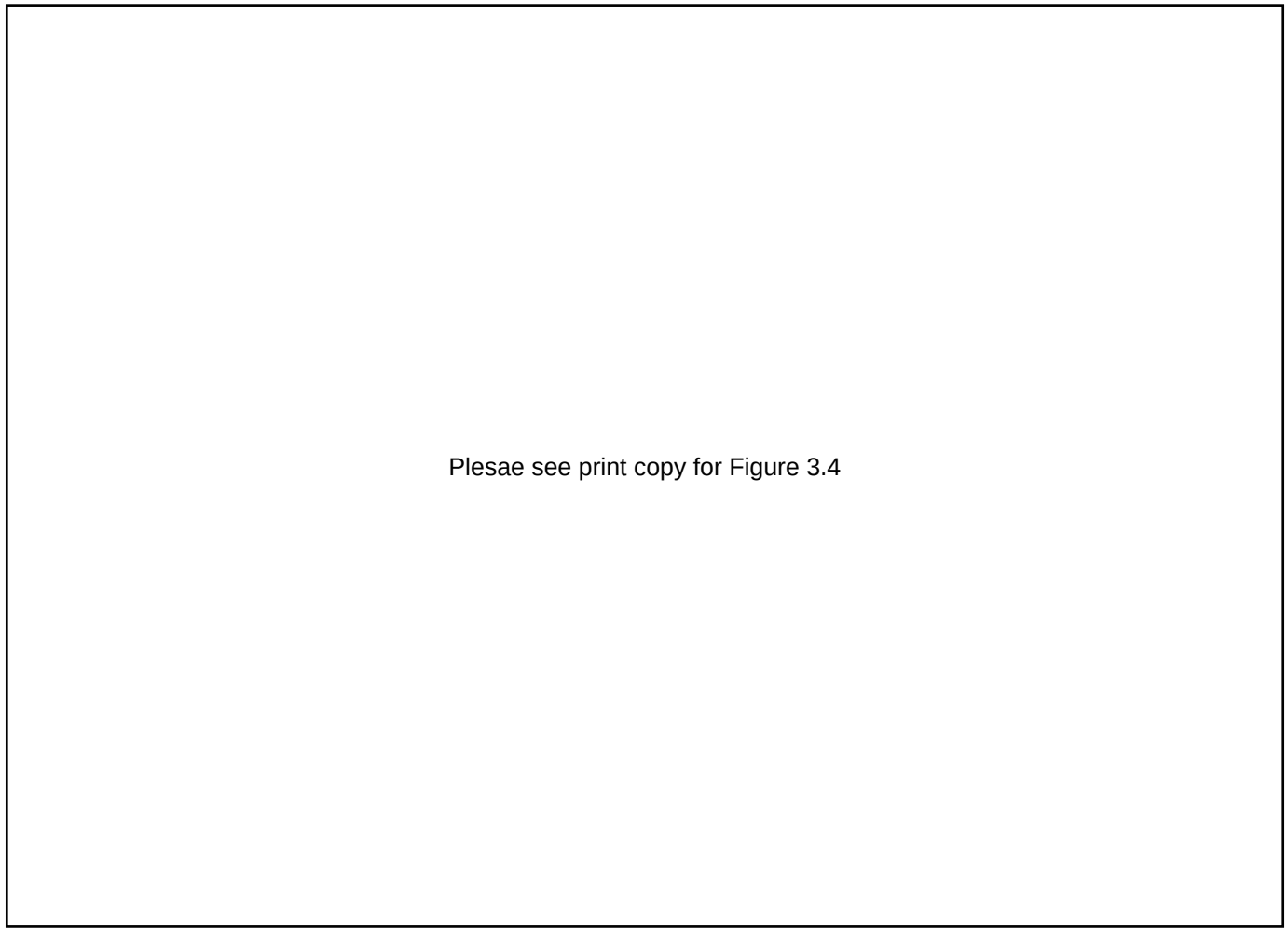

Figure 3.4 Three-Dimensional Flooding Simulation of Broughton Creek Drain (Data Source: Shoalhaven City Council, 2002).

3.7.3 ALS Accuracy

Most point densities in standard operations are seen to be too small (1 point per $10m^2$) and therefore not accurate enough to be used effectively in land management (Mass & Vosselman, 1999). The ALS survey taken by AAM GeoScan was designed to achieve an average spacing of four metres from fifteen overlapping swathes. This equated to 10,600,000 ground points and 3,600,000 non-ground points. Base station data from the GPS unit at the Shoalhaven Council offices was used as well as approximately 1500 ground truth points recorded in ISG projection (Zone 56/1, see section 3.9.5.3). The accuracy of the ground truth points was +/- 0.03m and the accuracy of the ALS data was a derived standard error of 0.16m.

A 0.5m and 1m DTM was generated by ESRI Australia from the ALS data, using Inverse Distance Weighted (IDW) interpolation of the some 11 million points. The IDW analysis used a high power (4) and a low number of points (6), with the aim of ensuring the grid cells, at a spot height location, reflected the measured value while maintaining good interpolation between points. Although this method of interpolation can cause a shift in object boundaries, by having greater point coverage, this shift can be reduced. The DTM was patched together from multiple $3km^2$ tiles.

3.7.4 Case Study: ALS Application to Coastal Acid Sulfate Soil

The pilot study site is located at Latitude –34.83, Longitude 150.66 (Easting 269295, Northing 114466 – ISG Zone 56/1), which is approximately 150km south of Sydney. The study site consists of approximately 230kms of drains and borders an area approximately $150km^2$. The study site is prone to large rainfall events and tidal flooding. Broughton Creek tributary feeds directly into the Shoalhaven River, which enters the Pacific Ocean via the Shoalhaven and Crookhaven Heads. The main source of industry within Broughton Creek floodplain is dairy and beef cattle and Manildra Starches (25%). Oyster and fishing industries within the area depend on the quality of the rivers and creeks. CASS have become a major problem in the Broughton Creek floodplain. With the installation of flood mitigation drains in the late 1960's, early 1970's, soil conditions have turned quite acidic and have affected these industries. Fish

and oyster populations have declined as a result of acid leachate entering Broughton Creek and eventually the Shoalhaven River.

Broughton Creek was designated as one of seven CASS hotspots in NSW. As a result funding for research and for the implementation of remediation structures became available. Remediation strategies available included those listed in Chapter 2.6. In order to determine the most suitable sites that would benefit most from such remediation options, Shoalhaven City Council used ALS to generate highly accurate ground elevation spot heights over a trial area including the Broughton Creek Floodplain. With this information a DTM was generated and used to determine the drainage pattern within the floodplain and the areas with the greatest likelihood of benefiting from remediation works (Figure 3.5). The placements of the SRTW and modified floodgates were dependent on the elevation of the land surrounding the flood mitigation drains. High-resolution 3D images with elevation intervals of 0.2m were developed for every drain in the Broughton Creek floodplain (Figure 3.6). Council together with the members of Shoalhaven River Acid Drainage Working Group (SRADWG) reviewed every drain and decided on the drains that needed further analysis to assess their suitability. Drains with very low-lying land adjacent to them were eliminated for the purposes of installing SRTW. Drains with levees high enough to avoid being overtopped were included for the possibility of having floodgates modified.

Two-dimensional modelling (Figure 3.4) was used to ensure that the amount of water that could be let into the drain would not overtop the levees. The drains to be modified were determined by the management committee (SRADWG) following more detailed assessment of the ASS layer in soils close to the drains and from cross sectional surveys of the drains themselves (Broughton Creek Management Plan, 2002.

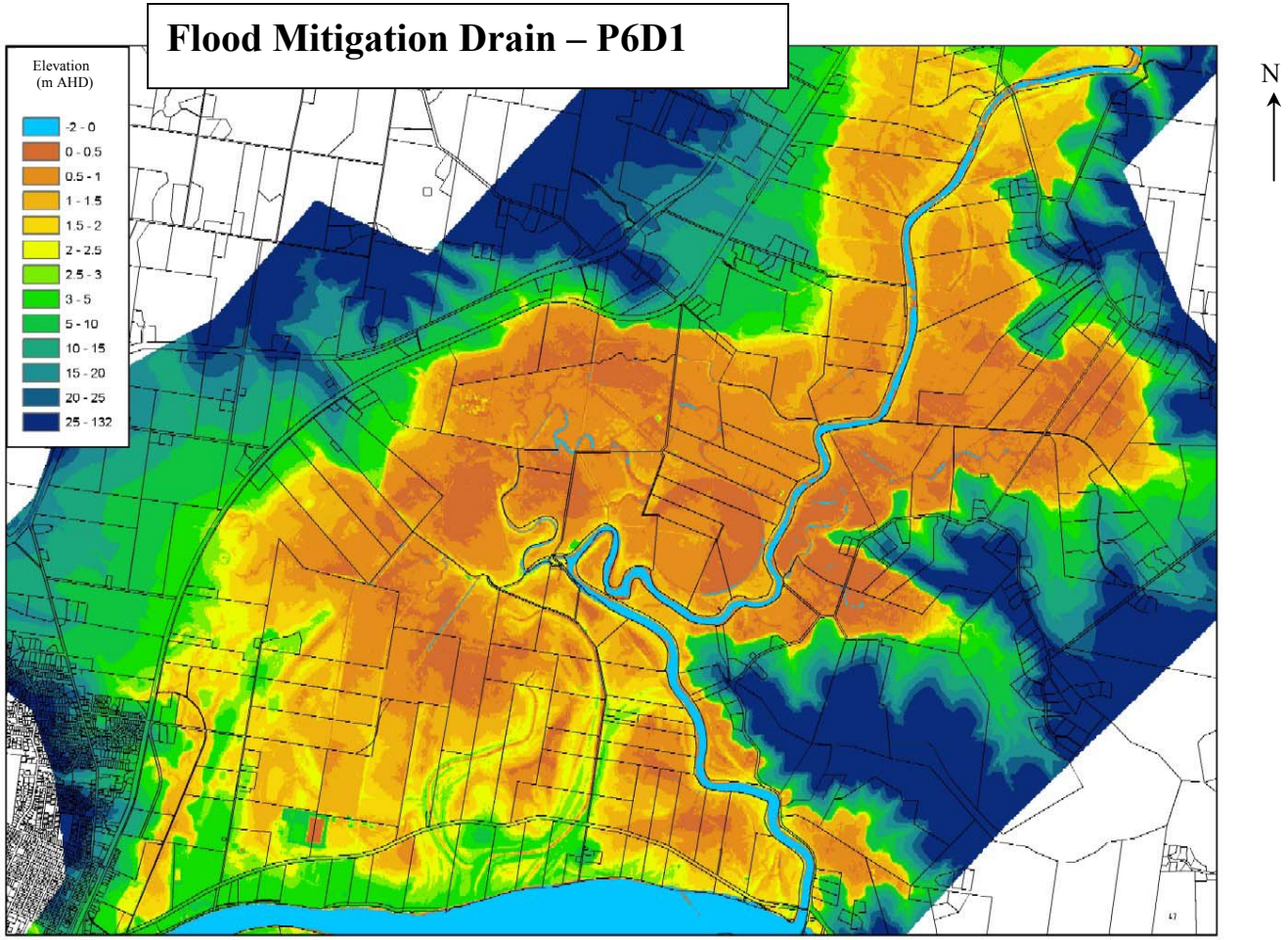

Figure 3.5 Digital Terrain Model of Broughton Creek Floodplain.

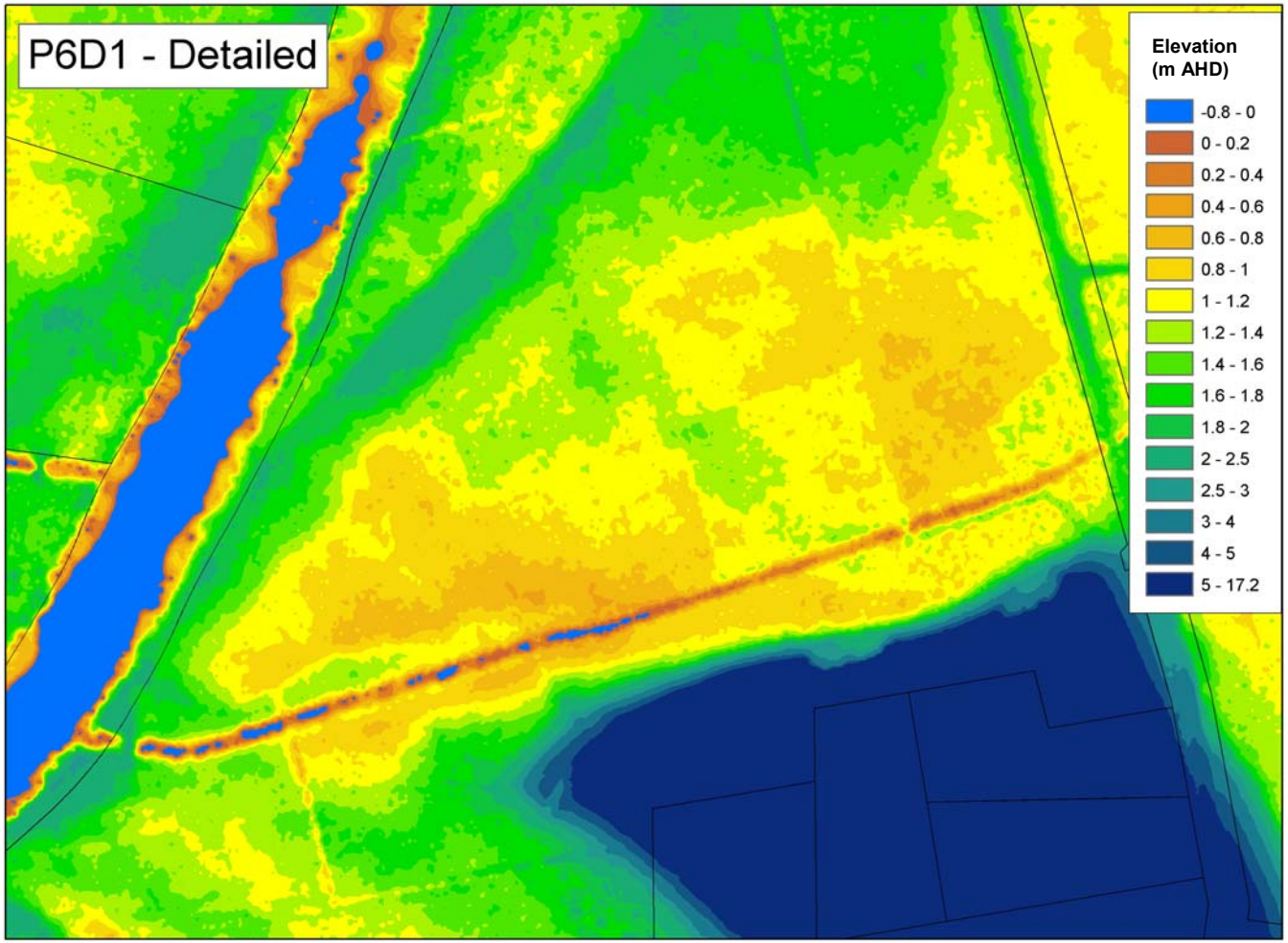

Figure 3.6 Flood Mitigation Drain Elevation Intervals on 0.2m scale.

3.7.5 Application of ALS data to Soil Data

ALS survey data was combined with existing borehole data to provide information regarding spatial positioning and elevation. Due to the available ALS data and soil data, Broughton Creek was chosen as the study site to develop a predictive model. The existing risk maps that predict locations of CASS through coastal New South Wales (see 3.8) are beneficial on a larger scale; however they use coarse elevation data and are not as accurate as a model that combines ALS with soil information.

3.8 Acid Sulfate Soil Risk Maps

In 1996 the Department of Land and Water Conservation (now Department of Natural Resources) prepared a series of 120 CASS Risk maps covering the entire NSW coastline. These risk maps predicted the distribution of CASS based on an assessment of their geomorphic environment. Three primary map classes; high probability, low probability and no known occurrence were mapped with codes indicating landform and depth class. Six thousand hectares of CASS were mapped. The risk maps were based on the premise that ASS distribution is strongly related to Holocene estuarine sediments and they do not occur at elevations above 1 m AHD.

3.8.1 Initial Mapping of ASS in coastal NSW

Initial maps were prepared by stereoscopic interpretation of 1:25 000 aerial photographs to identify landform elements in coastal environments with an elevation up to approximately 10 metres AHD. Landform elements were mapped onto 1:25 000 topographic maps. Where possible, air photos of the same area were used to take into account seasonal effects. Each map unit was allocated a landform process class, a landform element class, and an elevation class (Atkinson, 1996). Elevation data was taken from 1:4000 scale orthophoto maps when they were available and by extrapolation from known elevation points in other cases. Areas which had been mined, filled, or subjected to major soil disturbance were mapped as disturbed terrain. The elevation code was used to indicate the elevation of the present-day

ground surface. Water bodies such as rivers, lakes, creeks, and estuaries were mapped because of the likely occurrence of iron monosulphides within the bottom sediments. No elevation code was allocated to these areas.

Please see print copy for Table 3.1

Table 3.1: Landform Codes used on the Acid Sulfate Soil Risk Maps (Atkinson *et al.*,1996)

3.8.2 Risk Map Soil Sampling

Seven soil surveyors sampled a total of 840 sites over coastal NSW from March 1994 to April 1995. Soil Landscape Maps were used to provide additional data to the sampled

locations. Sampling was conducted over a selective scale, after the aerial photograph interpretation phase established a relationships between landform type, elevation and the occurrence and depth of acid soil materials. Further site selection was aimed at understanding the geomorphology and stratigraphic sequences in various landforms of each catchment. By doing this DLWC hoped to be able to predict the distribution and occurrence of pyritic sediments. At each profile soil morphological data, pH, and site information were recorded on NSW Soil Data Cards. DLWC used soil sampling equipment capable of sampling to 3 m.

Atkinson *et al.* (1996) found that the level of the pyritic/fluvial sediment interface occurred at less than 1 metre AHD. This information allowed estimation of the depth of occurrence of the pyritic sediment based on elevation of the ground surface and also resulted in no soil inspections on landforms higher than 4 metres AHD as soil sampling equipment did not penetrate to the anticipated level of the pyritic sediment layer.

3.8.2.1 Soil Sampling Techniques

Throughout the DLWC investigation between 1 and 4 soil samples of 300 - 500 grams from a number of selected sites were collected for laboratory analysis. Soil was taken from the soil profile that was suspected or considered to be ACASS or contain PCASS or pyrite. Over 1600 samples were collected for laboratory analysis. Samples were prepared to minimise contact with oxygen during time of transport and chilled to minimise oxidation of any pyritic material. The samples were tested for Electrical Conductivity, pH (1:5 soil water), pH (1:20) in H_2O_2, total actual acidity (TAA), and total potential acidity (TPA).

3.8.2.2 Risk Map Produced

Four significant map classes were developed through the DLWC's process. High and low probability classes were subdivided into depth categories. For each of the resulting categories the environmental risk associated with land use activities was described and typical landforms identified.

DLWC produced 129, 1:25 000 scale Acid Sulfate Soil Risk Maps, and 20, 1:100 000 scale catchment based maps including a set of guidelines (Naylor *et al*., 1995).

As aforementioned, DLWC selected a number of areas to designate as 'hotspots' within these areas mapped for CASS risk. 26 of the worst degraded 'hot spot' areas, totalling 55,000 hectares were identified from the Tweed River in Northern New South Wales (NSW) to the Shoalhaven on the South Coast NSW. In the Environmental Protection Authorities (EPA, now Department of Environmental and Conservation) Environmental Trust program, seven of these 26 were chosen as pilot projects to undergo in-depth monitoring of water and soil quality data, followed by the implementation of remediation strategies and other management options. Broughton Creek floodplain was chosen as one of the seven areas to be targeted first (Figure3.7).

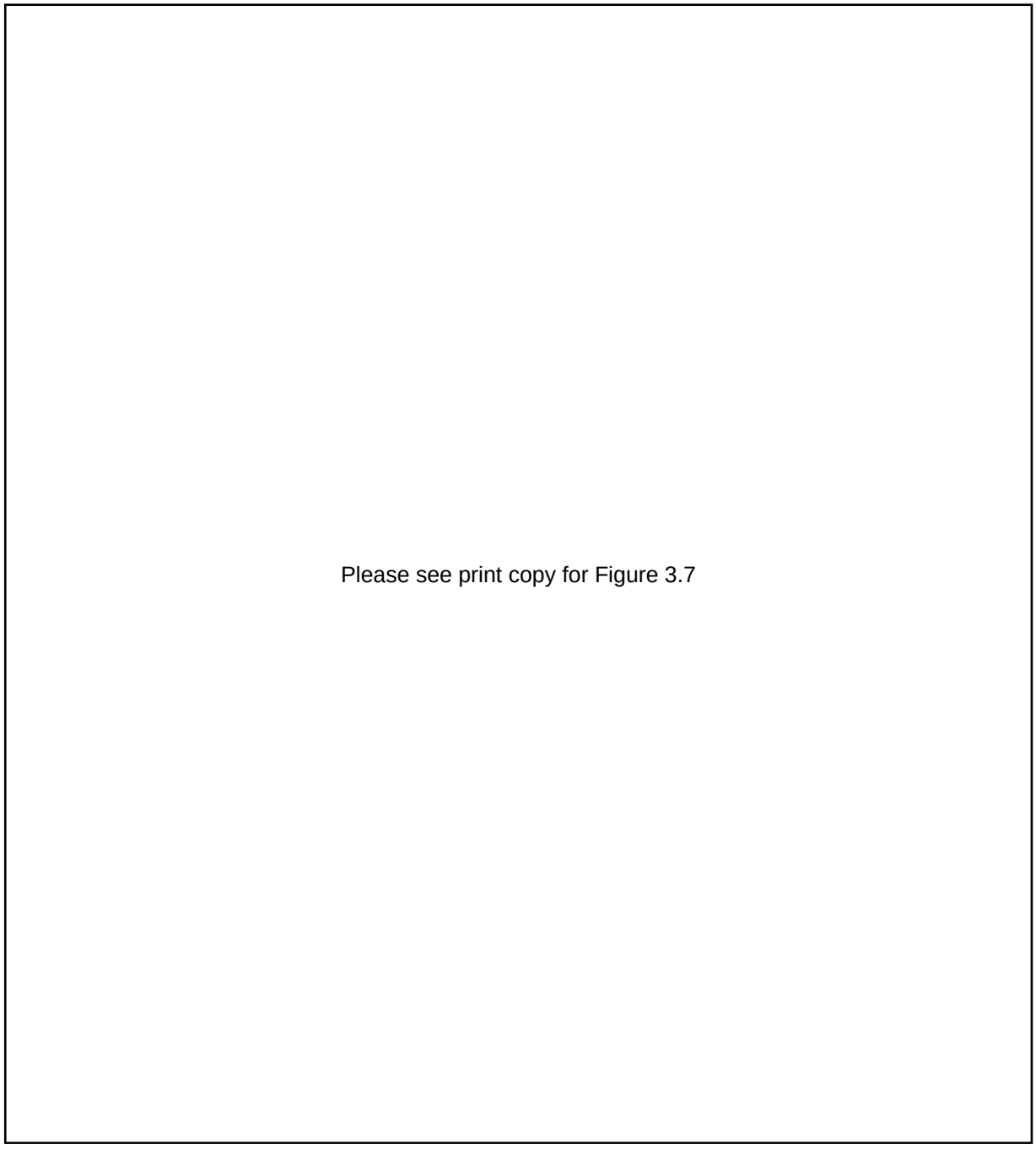

Figure 3.7 Acid Sulfate Soil Risk Map – Broughton Creek Floodplain (after Naylor *et al.*, 1995).

3.9 Soil Data

3.9.1 Locating Existing Soil Data

Existing soil data was collected from a number of different sources. As soil sampling is done for the majority of the time on a project basis, determining the projects that had been completed in the Broughton Creek floodplain required searching through a number of different sources:

a) Soil And Land Information System – As referred to in 3.6.3, The Soil And Land Information System (SALIS) is an online database that contains descriptions of soils, landscapes and other geographic features from across NSW. The aim of developing SALIS was to provide a centralised repository of information, which would eliminate the duplication of soil surveys (see 3.6.3). However as it became evident through the data collection phase this wasn't the all-encompassing soil database in NSW but was limited to selected soil samples. SALIS data contained multiple fields that described the soil sample in each of its profiles and totalled the majority of samples used as data for developing the predictive model.

b) University of Wollongong – Research into CASS in the Broughton Creek Floodplain started as early as 1975 (Norwood, 1975), which followed on from early research in estuarine soils in the Macleay River by Walker (1972). Norwood detailed the comparison between Broughton Creek Floodplain and that on the Southern side of the Shoalhaven River. 20 Soil data sampling points in the Broughton Creek Floodplain and 22 points within the Southern Floodplain were analysed for surface and subsurface acidity (pH), surface salinity (EC) and surface soluble Sulfates (and soluble sulfates as a percentage of total salts). Three profiles were analysed in Broughton Creek floodplain down to 1.80m below the surface, detailing the changes in the profile and the associated soil characteristics (pH, %soluble Sulfate, %sulfate salt, %soluble salt, %total sulfate, organic matter, % clay, %silt and %sand). It wasn't until 1994 that research from the University of Wollongong started up again around the Broughton Creek area. Pease (1994), detailed the effect of rainfall and drainage on CASS in Broughton Creek floodplain. Twelve soil samples had detailed profile analysis conducted on them including pH

and cation (sol. Al, Total Al, Total Fe) analysis. Water quality data (Cl, SO_4) was sampled in nearby bores and water bodies. Chapman (1995) and Sullivan (1995) sampled on private dairy land within the Broughton Creek floodplain. Eight profiles were sampled at varying depths, analysing soil acidity (pH), sulfate, jarosite presence or absence and level at to which the PCASS occurs in the profile. Blunden (2000) then conducted a more comprehensive analysis on a field site in a similar area as Chapman (1995) and Sullivan (1995). The sampling procedures that Blunden used included analysing down a number of soil profiles for soil acidity (pH, Total Actual Acidity), soil salinity (EC, Sulfates, Chloride and Cl-SO_4 ratio), soil cation content (Al, Fe^{2+} , Fe^{3+}) and detection of jarositic material or the PASS-pyrite layer. Toniato (1997), Thong (1998) and Mo-Ane (1998) analysed water quality data and Thong (1998) analysed a disturbed sample in laboratory conditions, however no data was applicable for this project. Glamore (2003), sampled 9 soil profiles across a transect, detailing soil acidity, salinity, cationic composition and anionic relationships as well as jarosite and PASS-pyrite detection. This field site was on adjacent land to that of Blunden (2000) and similar techniques were evoked. Rudens (2001) and Ford (2002) provided a number of soil samples with limited analysis but tested for soil acidity (pH) and salinity (EC) and other descriptive characteristics of the profile (colour, hue, jarosite depth/existence).

c) Agriculture NSW: Two major reports were compiled on the impact of irrigating CASS land with effluent from Manildra Starches company. These two reports comprehensively detailed cationic and anionic components of the soil profile down to depths of the profiles as far as 3.5m down the profile both before and after irrigation. A total of 32 samples were analysed for soil acidity (pH, Total Actual Acidity), soil salinity (EC, Sulfates, Chloride and Cl-SO_4 ratio), soil cation content (exchangeable Al as per cent of cation exchange capacity - %ExAl/CEC, Mg, Fe, K) and detection of jarositic material or the PASS-pyrite layer.

d) Shoalhaven City Council (SCC): NSW Agriculture and SCC undertook soil sampling, primarily for the 2001-04 ‘hotspots’ project. This sampling aimed at determining which locations within Broughton Creek floodplain should have remediation works. Such works included modified floodgates to allow tidal

buffering, installation of weirs to increase the ground water table or subsurface lime injections to act as a neutralizing barrier (Blunden 2000; Rudens 2001; Glamore, 2003). The main soil properties tested were soil acidity (ph and TAA), soil salinity (Cl/SO_4), height of the watertable on the soil profile, detection and height of jarosite in the profile, and the detection of the PASS-pyrite layer. There were 48 samples taken across the floodplain, with 13 analysed for the all the properties above.

Combining all of the samples into one database created a collection of 1377 records from 1970 to 2004, grouped into 16 different elevation intervals as represented in TABLE 3.2. Grouping of data was used as a method to eliminate the differences in sampling techniques (in terms of difference in elevation measured) used by different personnel that occurred over the 34-year period.

Elevation Below Surface (m AHD)	**Interval Number**	**Interval Range**
0	0	0
>0.00	1	0.001-0.049
>=0.05	2	0.05-0.099
>=0.10	3	0.10-0.199
>=0.20	4	0.20-0.299
>=0.30	5	0.30-0.399
>=0.40	6	0.40-0.599
>=0.60	7	0.60-0.799
>=0.80	8	0.80-0.999
>=1.00	9	1.00-1.249
>=1.25	10	1.25-1.499
>=1.50	11	1.50-1.749
>=1.75	12	1.75-1.999
>=2.00	13	2.00-2.399
>=2.40	14	2.40-2.799
>=2.80	15	2.80-3.499
>=3.50	16	3.50+

Table 3.2: Grouped elevation increments for data standardisation

After collating the data into a standardised database the data was exported into a GIS in the form of vector points, combined with the ALS DTM, the digital aerial photograph and other important feature layers such as cadastre, catchment boundaries, geology and vegetation. A GIS is one of the most efficient ways of coordinating data with spatial reference. It also enables the processing and manipulating of this data to produce the outcome desired by the user.

3.9.2 Organising Soil Data in a GIS

Soil records can be stored as spatially referenced files within a GIS. These are usually stored as vector data (see Appendices A1.1) and used with raster data (see Appendices A1.4) such as a DTM. A full explanation of the formats to which soil data can be stored in a GIS can be found in Appendices A1.

3.9.3 Data Coordinate Systems

Data stored in raster and vector formats must be referenced by a coordinate system, which explains how the data is viewed in reference to the earth. The data used in this project was generated under a number of different coordinate systems and then converted to the same coordinate systems to retain accuracy. There are two types of coordinate systems used in spatial data: geographic and projected. Geographic coordinate systems use latitude and longitude coordinates on a spherical model of the earth's surface. Projected coordinate systems use a mathematical conversion to transform latitude and longitude coordinates from a three-dimensional surface to a two-dimensional surface. These two coordinate systems are explained in more detail in Appendices A2 – A3.

Of the geographical coordinate systems, Universal Transverse Mercator (UTM) is the most commonly used and widely accepted (see Appendices A2.1).

3.9.4 Universal Transverse Mercator (UTM) Projections used in Broughton Creek Data

The previous datum used in NSW was the Australian Geodetic Datum 1966 (AGD66). Coordinates were expressed in terms of latitude/longitude, the Australian Map Grid (AMG) or the Integrated Survey Grid (ISG). Both AMG and ISG are Transverse Mercator projections. AMG uses 6º degree zones and ISG uses 2º zones.

The recently used datum is Geographic Datum of Australia (GDA) which is commonly projected as Map Grid of Australia (MGA). MGA is a Universal Transverse Mercator (UTM) projection using 6º zones. This ensures consistency between the states and will result in less zone boundaries than the ISG system.

Within Broughton Creek, Shoalhaven Councils data was primarily ISG in zone 56/1.

3.9.5 The Geodetic Datum

A datum is a framework to define coordinate systems. A surface, which can be used as a basis for referencing geodetic coordinates, is referred to as a geodetic datum. A datum is comprised of a number of elements - a spheroid (see Figure 3.8) that has a defined size and shape, a location, or origin, in three-dimensional space, and an orientation of each of its axes. The definition of these elements fixes the datum is space and enables users to reference points on the Earth to the defined coordinate reference system.

The datum used by the Global Positioning System (GPS) is the World Geodetic System 1984 (WGS84). This system is a geocentric based coordinate system with the origin of the defining spheroid located at the Earth's centre of mass (Figure 3.8).

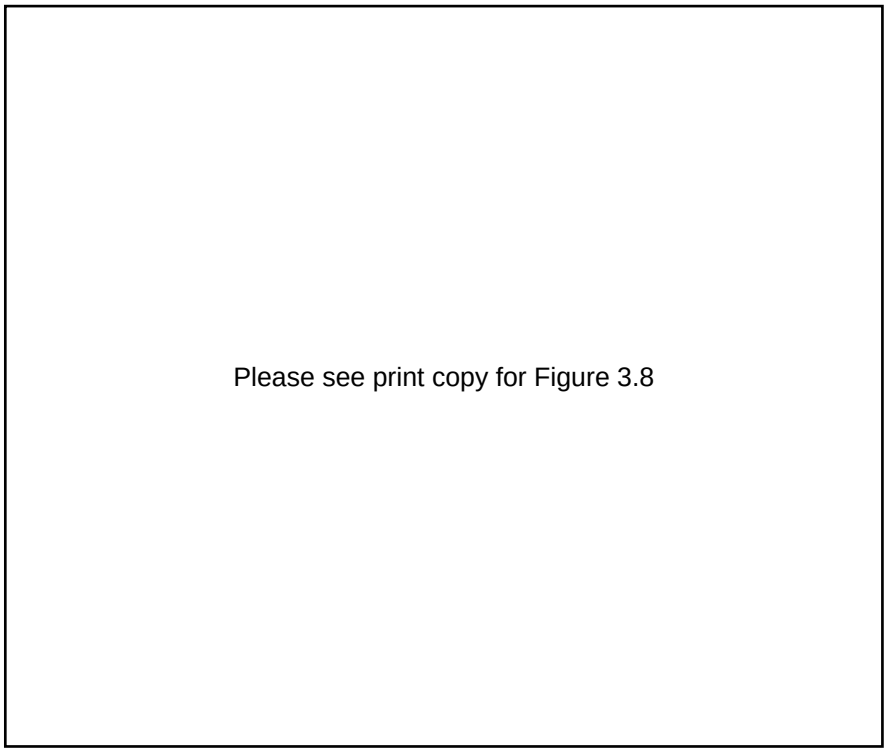

Figure 3.8 Spheroid shape as referenced to the earth (SBV, 2003).

The Australian Geodetic Datum 1966 (AGD66) is the commonly used coordinate reference system throughout Australia, however with the current move towards GDA, the former Geodetic Datum (AGD66) has become a Geocentric Datum, meaning that the shift will be in a North-Easterly direction to correct for this change in reference system (see Figure 3.9). The various types of Geodetic Datum are explained in more detail in Appendices A3.

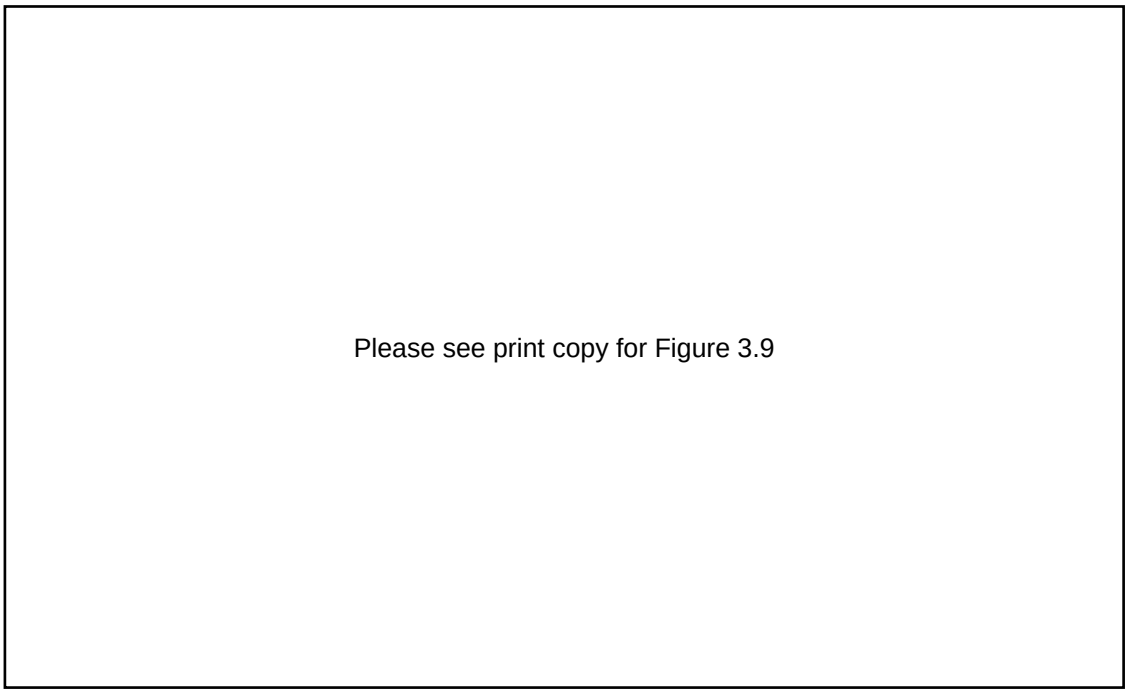

Figure 3.9: Geocentric Datum (WGS84) - Geodetic Reference System 1980 & Geodetic Datum (AMG66) – Australian National Spheroid (After ICSM, 2004).

3.9.6 Vertical Datum – Australian Height Datum (AHD)

A Vertical Datum is used to fix a position in the vertical direction, up and down the Z axis. A vertical datum is a line, value or set of values from which heights are measured. Australia's vertical datum is the Australian Height Datum (AHD) which approximates mean sea level and was determined by monitoring tide gauges around the Australian coastline. The change of horizontal datum to the GDA94 will not affect height the currently used Vertical Datum (ICSM, 2004).

Chapter 4: Preliminary Data Analysis

Converting data into a standard projection and organizing this data into a centralized database enables the data to be analysed more efficiently. It also helps determine whether or not the soil sample sites with selected parameters EC, ExAl/CEC, Cl-SO_4, TAA, and pH can be used to predict ASS severity.

There are 1377 samples that constitute up to 155 individual points sampled at various levels down the soil profile (Table 4.3). Within the analysis, the mean Easting and Northing (Projection ISG, Zone 56/1) was found to be at 268038.08 East, 1144579.1 North. This becomes important when applying geostatistical principles (see Chapter 5.5) to estimate an unknown point based on other known points. The mean ground elevation was 1.645 m showing that the samples were taken across an area of low elevation, which is consistent with previous research determining if CASS exist (Dent & Dawson 1996).

By applying the basics of data analysis which focuses on preliminary screening, the robustness of the data can be evaluated, which will determine whether the data set displays normality, linearity, homoscedasticity, and non-multicolinearity. These criteria must be met when applying the general linear model for evaluating these parameters.

4.1 Screening Data

The principle parameters – pH, TAA, Cl-SO_4, and ExAl% / CEC were analysed in a comprehensive database, and in a grouped database based on intervals (see Table 3.2).

Before applying a linear model to the data, the data went through a data screening process described by Tabachnick & Fidell (1989). Spatial autocorrelation (4.9) was also addressed as it often affects many geographic data sets (Wulder, 2002).

Data screening is important to determine how and if the data needs to be altered to fit into the mathematical assumptions of linearity, unbiased and normality. In order to apply multivariate statistics to the data set it must adhere to these assumptions. A step-by-step process of analysing the data can eliminate the unnecessary error associated with

transforming the data set before it is fully understood. Consideration and resolution of problems encountered in the screening phase is necessary to ensure a robust statistical assessment.

Tabachnick & Fidell (1989; 2001) describe an eight-step process of data screening which is has been slightly modified for the purposes of evaluating the data set in this project. By screening data the data will be in a suitable format to be used in multivariate analysis.

4.2 Univariate Descriptive Statistics

Descriptive Statistics explain the parameters to be used in the project in a general sense (Table 4.1). Calculating the mean, standard deviation, and range (minimum and maximum) can help determine whether outliers exist in the data, whether the mean value is suitable for the given parameter, and whether the standard deviations are acceptable. pH is the key variable in the determination of CASS severity. Comparing other variables to this is important, and as pH is measured in a logarithmic scale the other parameters were converted into a logarithmic scale to standardise the data (Table 4.2). Converting data into the same scale enables us to compare the standard deviations of the mean to determine if we are dealing with similar data sets, hence a plausible data set. This method of data transformation will be explored further in Section 4.5.2.

From this comparison, pH has a standard deviation less than 1, as do all of the other variables analysed after transformation to a logarithmic scale. This indicates that the data points do not have any severe outliers, however this must be evaluated further using confidence intervals (see Section 4.6).

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Sample No.	1377	17493506.8	28354862.5	0.0	92001991.0
Easting (ISG)	1369	268038.1	2258.8	264698.2	272771.3
Northing (ISG)	1369	1144579.1	1922.3	1141936.0	1149469.3
Elevation (Raster 9)	1369	1.646	3.246	-0.117	62.160
Elevation (SCC DTM)	1369	1.644	3.247	-0.002	62.160
Upper Layer Boundary	1377	0.616	0.586	0.000	3.800
Lower Layer Boundary	1373	0.771	0.632	0.000	3.950
pH (pH units - $CaCl_2$)	607	4.312	0.926	2.500	7.420
EC (ds/m)	536	3.284	5.295	0.020	41.030
EC (uS/cm)	538	3255.340	5297.910	21.400	41030.000
Bray Phosphate (mg/Kg)	239	30.583	44.444	0.000	235.000
Carbon (%)	216	3.999	3.623	0.131	20.000
Nitrogen (%)	207	2.073	3.726	0.000	20.930
Cation Exchange Capacity (cmol(+)/kg)	184	47.720	77.506	1.110	350.800
ExAl (cmol(+)/kg)	169	3.359	3.098	0.000	20.000
ExAl (%)	169	32.122	25.499	0.000	77.586
Total Actual Acidity (moles H+/ tonne)	154	94.338	81.603	1.900	400.000
Chloride:Sulfate	150	2.262	6.211	0.009	70.000
Sulfur - Chromium reducible (%)	93	0.631	0.889	0.005	3.800
Hydraulic Conductivity (ums^{-1})	87	10.910	8.069	0.980	39.370
Bulk Density - air dry (g/cm^{-3})	74	1.412	0.218	0.910	1.910
Jarosite Depth (to detection)	48	1.177	0.462	0.000	1.950

Table 4.1 Summary Statistics of All Variables.

LOG Variables	N	Mean	Standard Deviation	Minimum	Maximum
pH (pH units - $CaCl_2$)	607	4.312	0.926	2.500	7.420
Log EC (uS/cm)	538	3.097	0.639	1.330	4.613
Log Bray Phosphate (mg/Kg)	232	1.145	0.571	0.000	2.371
Log Carbon (%)	216	0.404	0.456	-0.883	1.301
log Total Actual Acidity (moles H+/ tonne)	212	1.835	0.462	0.279	2.602
Log Nitrogen (%)	201	-0.185	0.742	-2.398	1.321
Log Sulfur - Chromium reducible (%)	93	0.161	0.202	0.002	0.681
Log Hydraulic Conductivity (ums-1)	87	0.932	0.313	-0.009	1.595
Log Bulk Density - air dry (g/cm-3)	74	0.144	0.070	-0.041	0.281

Table 4.2 Summary Statistics of Log Variables.

4.2.1 Data Standardisation

The separation of the data into intervals removes the differences in soil sampling techniques and enables one soil sample to be compared to another layer (Table 4.3). The distribution of the data can be investigated using this method to standardise the soil sampling, as it represents a normally distributed sample when plotted as a graph (see Figure 4.1).

Interval	Depth Below Surface	Samples
1	>0.00	40
2	>=0.05	125
3	>=0.10	125
4	>=0.20	112
5	>=0.30	136
6	**>=0.40**	**155**
7	>=0.60	109
8	**>=0.80**	**152**
9	>=1.00	145
10	>=1.25	112
11	>=1.50	88
12	>=1.75	46
13	>=2.00	17
14	>=2.40	9
15	>=2.80	5
16	>=3.50	1
TOTAL SAMPLES		**1377**

Table 4.3 Number of sample points per Soil Layer.

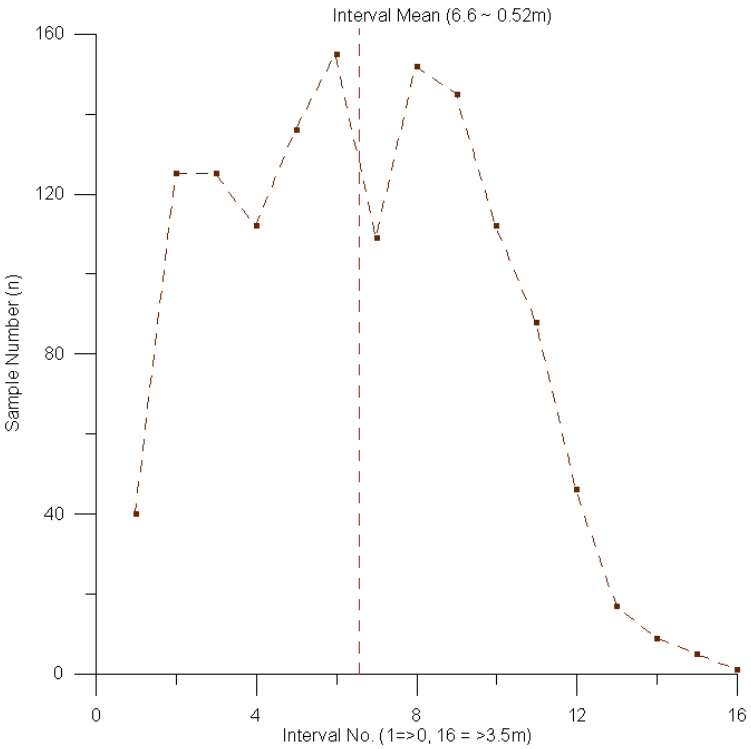

Figure 4.1 Normal Distribution of intervals in project data set.

4.2.2 Coefficient of Variation

The bivariate correlations between all variables can show inflated or deflated correlation (Table 4.4). This may be due to repetition of a variable in the data set. Correlation analysis can also uncover any discrepancies in relationships as defined in theory.

The majority of the correlation indices in this project are low. This helped to identify the variables that require transformations, if they didn't correspond with theoretical relationships. For instance Table 4.4 showed the relationship between pH and TAA of 0.56, which immediately signals some error in the data set caused most likely by mistakes in sampling. As pH and TAA directly measure acidity, the relationship should be a lot closer to 1.0. The relationship between Exchangeable Al% / CEC and pH generated a correlation of -0.73.

As aforementioned, correlation analysis aids in identifying relationships between variables that should be looked at in greater detail. The bivariate relationships may also uncover variables, which are multi-collinear or singular (see 4.8). From the bivariate correlation analysis, the variables with the strongest correlation were pH and log (TAA) which are theoretically more accurate. Since TAA includes pH (H^+ ions) in the summation of all (total) acidity within a system it should be auto-correlated or have a correlation close to 1. The predicted versus actual values of pH from the linear regression relationship of pH and log(TAA) is represented in Figure 4.2. Another variable of note was pH and log (ExAl%/CEC), represented in Figure 4.3, which shows that the lower or more acidic values of pH are more accurate in predicting than the upper values. Since we are dealing with an issue focusing on a highly acidic environment, this is a positive relationship.

	X_ISG	Y_ISG	PH	EC	TAA	CLSO4	EXAL_CEC
X_ISG	1	0.8066	-0.300	-0.425	0.363	-0.349	0.535
Y_ISG	0.806	1	-0.227	-0.217	0.392	-0.213	0.117
pH	-0.300	-0.227	1	0.444	-0.559	0.546	-0.626
EC	-0.425	-0.217	0.444	1	-0.071	-0.116	-0.335
TAA	0.363	0.392	-0.559	-0.071	1	-0.234	N/A
CLSO4	-0.349	-0.213	0.546	-0.116	-0.239	1	-0.301
EXAL_CEC	0.536	0.111	-0.626	-0.336	N/A	-0.301	1

Table 4.4 Pearson Bivariate Correlation Coefficients of Raw Data.

	X_ISG	Y_ISG	PH	logEC	logTAA	logCLSO4	logEXAL_CEC
X_ISG	1.000	0.806	-0.300	-0.446	0.399	-0.642	0.547
Y_ISG	0.806	1.000	-0.227	-0.262	0.497	-0.615	0.094
pH	-0.300	-0.227	1.000	0.367	-0.798	0.414	-0.737
logEC	-0.446	-0.262	0.367	1.000	-0.125	-0.248	-0.385
logTAA	0.399	0.497	-0.798	-0.125	1.000	-0.273	N/A
logCLSO4	-0.642	-0.615	0.414	-0.248	-0.273	1.000	-0.418
logEXAL_CEC	0.547	0.094	-0.737	-0.385	N/A	-0.418	1.000

Table 4.5 Pearson Bivariate Correlation Coefficients of Log-Transformed Data.

Equation	**R^2**
pH = 5.37 – 1.05 (logTAA)	0.6370
pH = 4.88 – 0.67 (logExAl/CEC)	0.5425
pH = 3.70 + 0.44 (ClSO4)	0.2982

Table 4.6 Simple regression equations between pH and important variables.

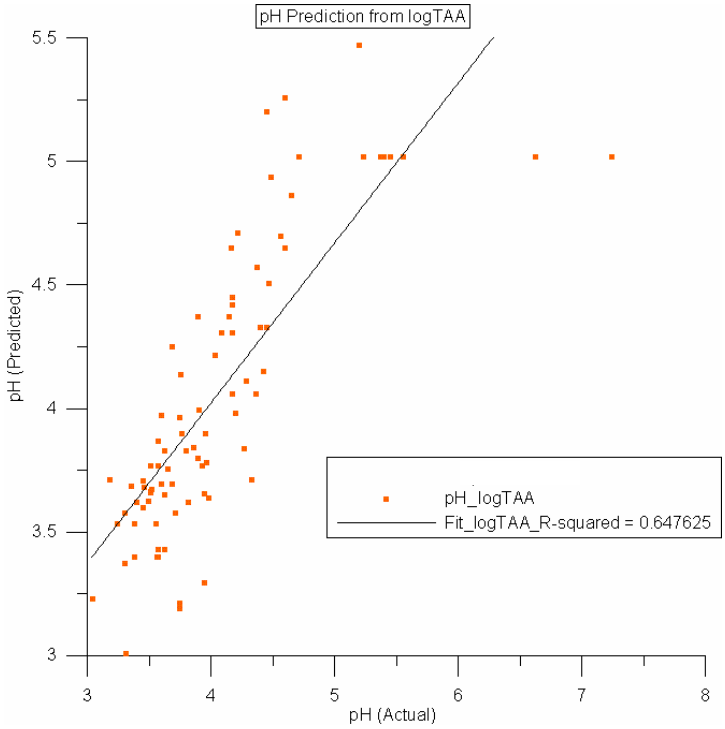

Figure 4.2 Pearson Correlation pH and log(TAA).

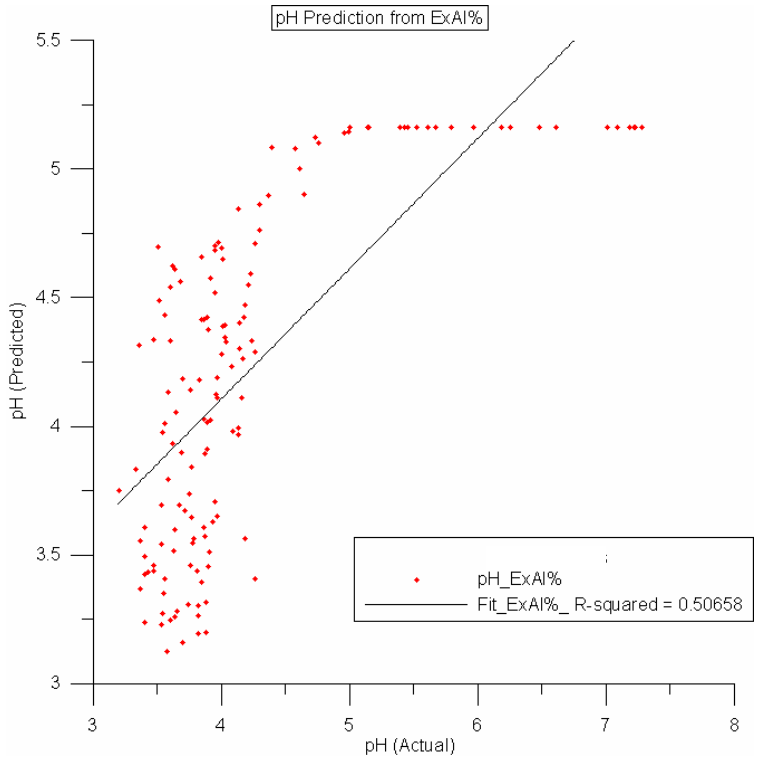

Figure 4.3 Pearson Correlation pH and log (ExAl%/CEC).

4.3 Missing Data Analysis

Not all variables were available at all sample points due to the number of data sets combined over the 35-year period. This means the data set has a number of sample points with missing data points. Using a geo-statistical interpolation method can help create a more comprehensive data set, using known data points to determine unknown points (see 5.5). With any modelling there is a certain error generated. Minimising this will increase the confidence in the results, hence the data set.

4.4 Variable Independence

The only dependent variables included in the primary analysis are pH and TAA. More sample locations have measured pH than TAA, and so using the relationship function between pH and TAA will help generate TAA from pH and create a larger data set. Within the preliminary analysis Cl-SO_4 and CEC had an R-value of 0.038. As these two variables are independent of one another, the R-value in accordance with the assumption of Orthogonality is close to 0 (perfect orthogonality).

4.5 Assessment for Normality

As previously mentioned, grouping data into intervals helps to remove the problem of having many different sampling techniques when using data from multiple soil surveyors. Samples are related to a height range within the profile which reduces data processing and makes it easier to analyse the data.

A soil sample can only be used in the determination of CASS severity if it has been tested for soil chemical properties. Determining whether the sample set exhibits a normal distribution will help determine if any transformations are necessary. For multivariate analysis, all variables and all combinations of the variables need to be normally distributed to generate useful predictive models. In many situations, data is not normally distributed around the mean and the data needs to be transformed (see 4.5.2) in order to generate a normal distribution. Figures 4.4 and 4.5 represent the variables used in this project and their associated means and distribution of the data around the mean.

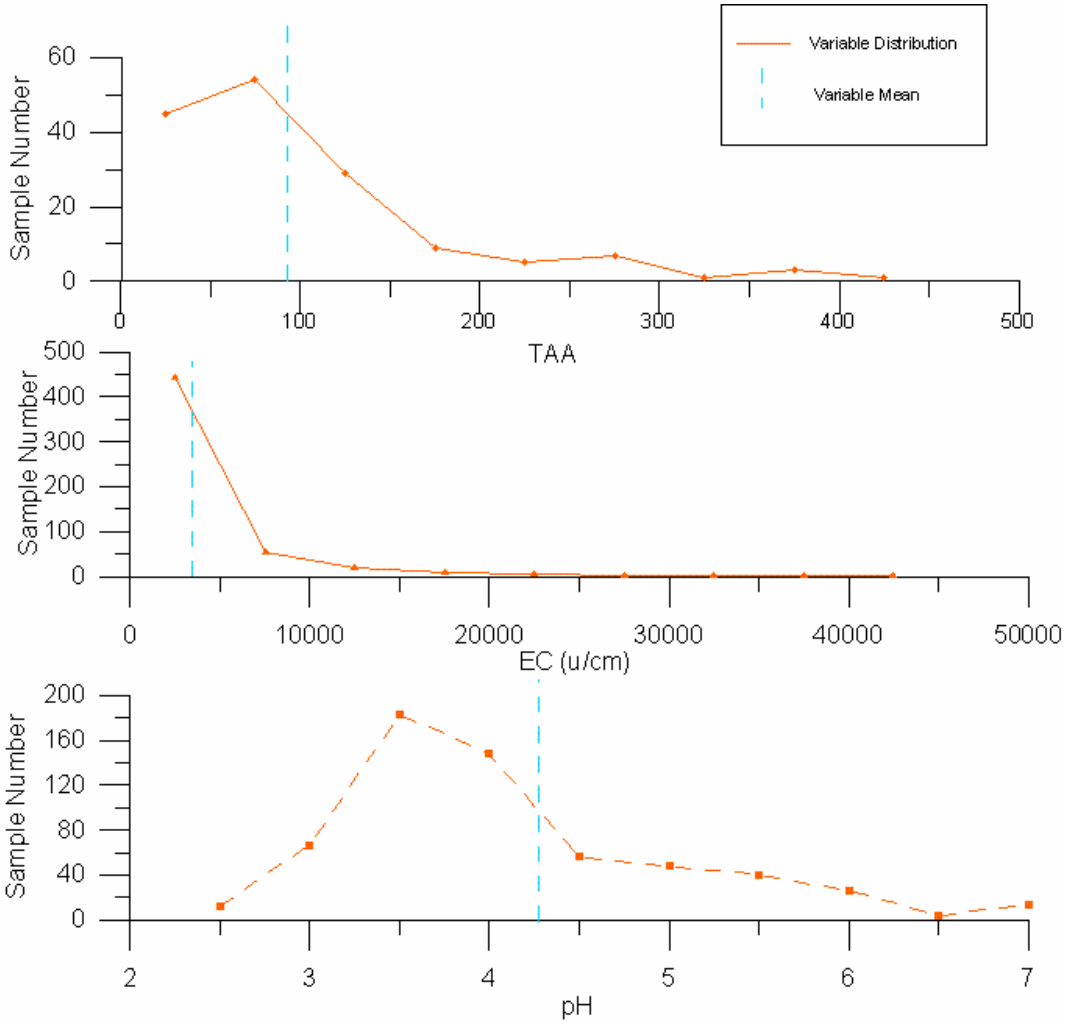

Figure 4.4 Distribution of variables (TAA, EC, pH) used in the project.

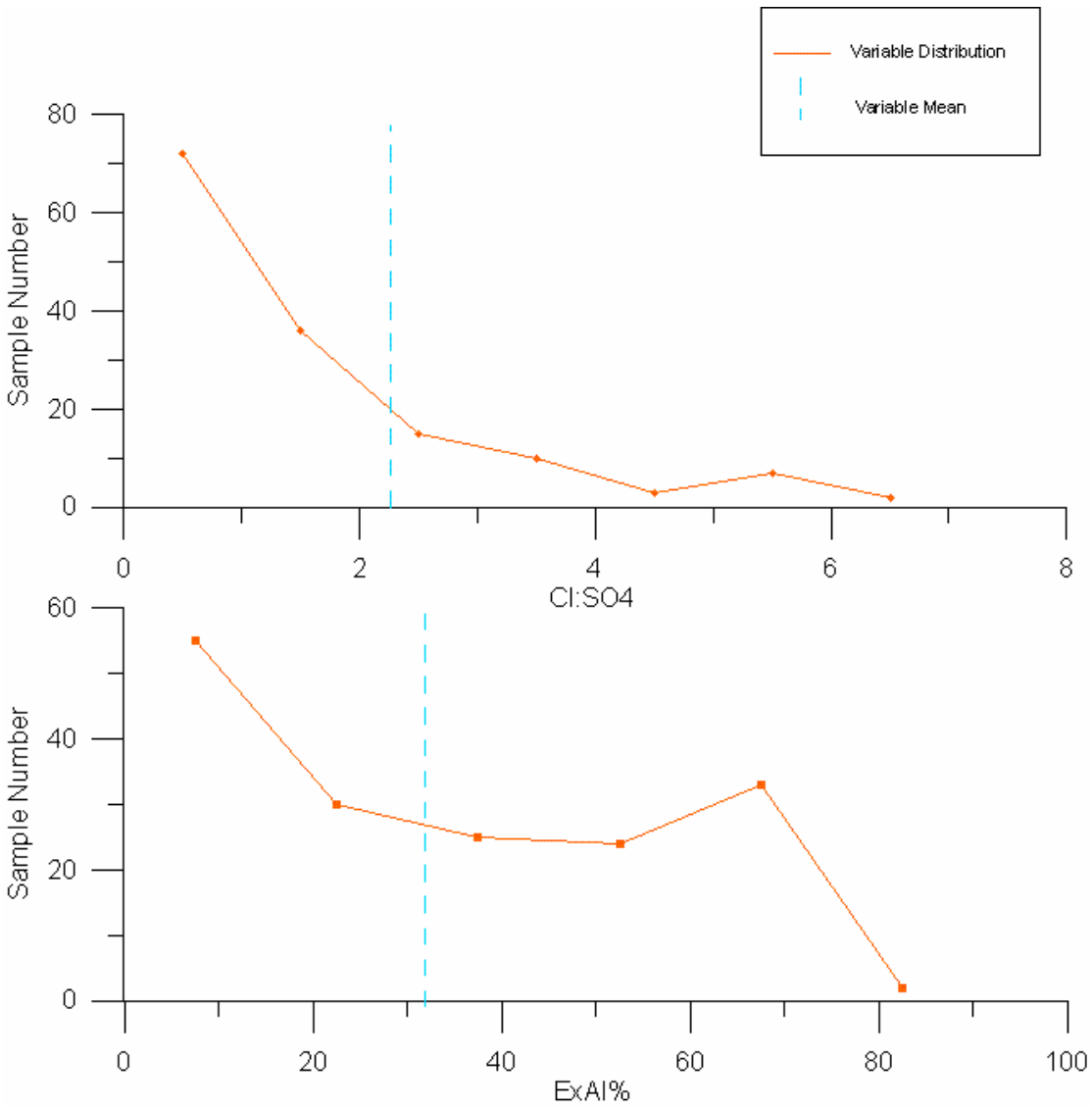

Figure 4.5 Distribution of variables (ExAl%, Cl:SO_4) used in the project.

4.5.1 Individual Variable Analysis

Each variable used in the analysis should have a normal distribution, in order to use multi-variate analysis methods. When the data is truly normally distributed, the residuals are also normally distributed and independent, not forming any particular pattern around the mean (when residual is equal to zero). From the data set used, Figures 4.4 and 4.5 show how the data is distributed.

Data that is not represented graphically by the bell shaped normal-curve, as is the case of all the variables in Figure 4.4 and 4.5, may need to be transformed from another distribution, for example from a negatively skewed distribution to a normal distribution. Figure 4.6 represents the possible distributions that a data set can show, and Figure 4.7 the possible transformations.

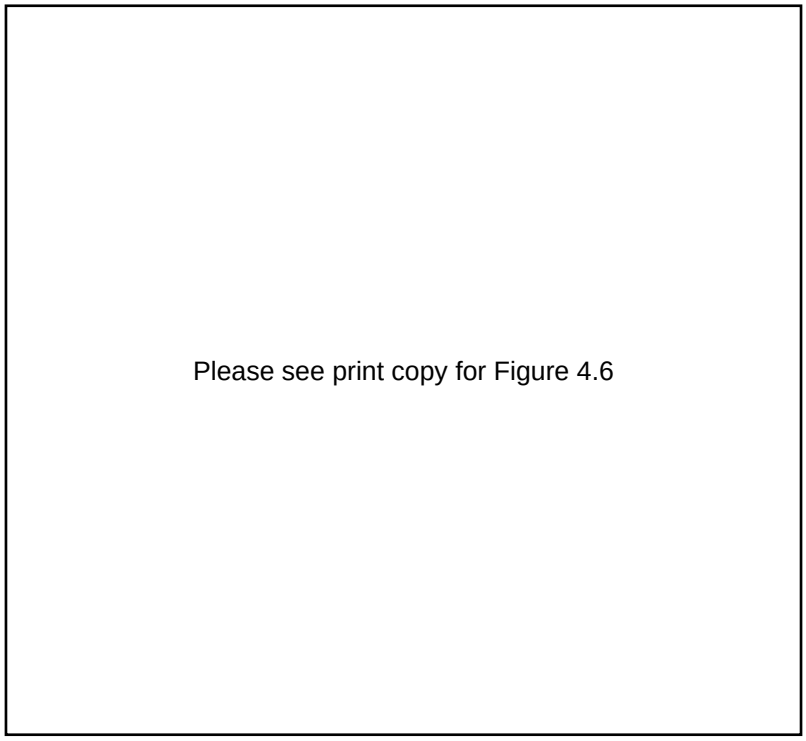

Figure 4.6 Various types of data distributions (Wulder, 2002).

4.5.2 Skewness, Kurtosis and Probability Plots

Figures 4.4 and 4.5 represent the variables used in the project and their distribution around the mean. The majority of variables used in this project show a distribution around the mean that is positively skewed, but in the case of ExAl%, the data shows negative kurtosis.

4.5.3 Data Transformations

Data not normally distributed about the mean can be transformed to rid breaches not only in normality but also in non-linearity. This is the case when data contains noticeable outliers and does not represent homoscedasticity. However, after transformation (see Figure 4.8 and Figure 4.9) the data should be assessed for normality.

Interpreting data post-transformation can be difficult. If the scale of the original data is arbitrary, interpretation will only be marginally hindered, yet if the scale is meaningful the transformation may be confusing. Transformations are appropriate when the non-linearity is monotonic throughout the data set. However, if the distribution plot is non-monotonic it may be necessary to use another variable in the model (Tabachnick & Fidell, 1989).

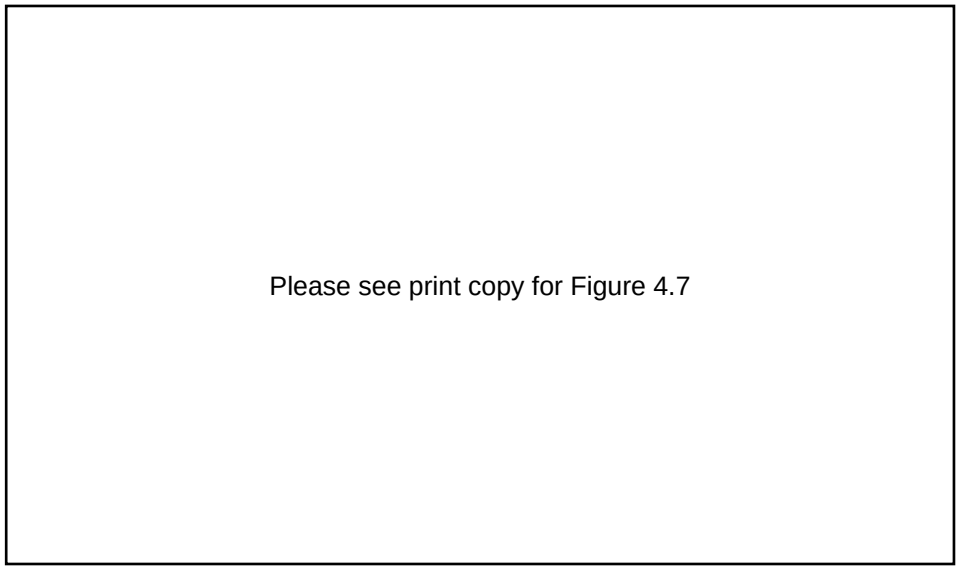

Figure 4.7 Possible Data Transformations to meet Normality Assumption (Wulder, 2002).

It was necessary to transform all variables (Figure 4.8 and Figure 4.9) due to their non-normal distribution. This was done using the log transformation.

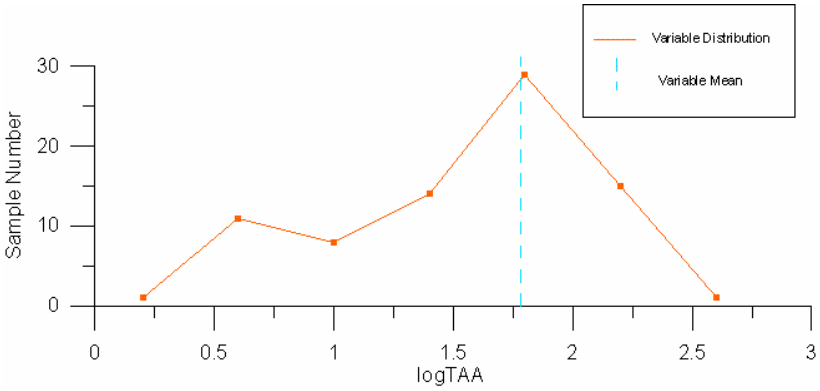

Figure 4.8 Log-Transferred Distribution of TAA.

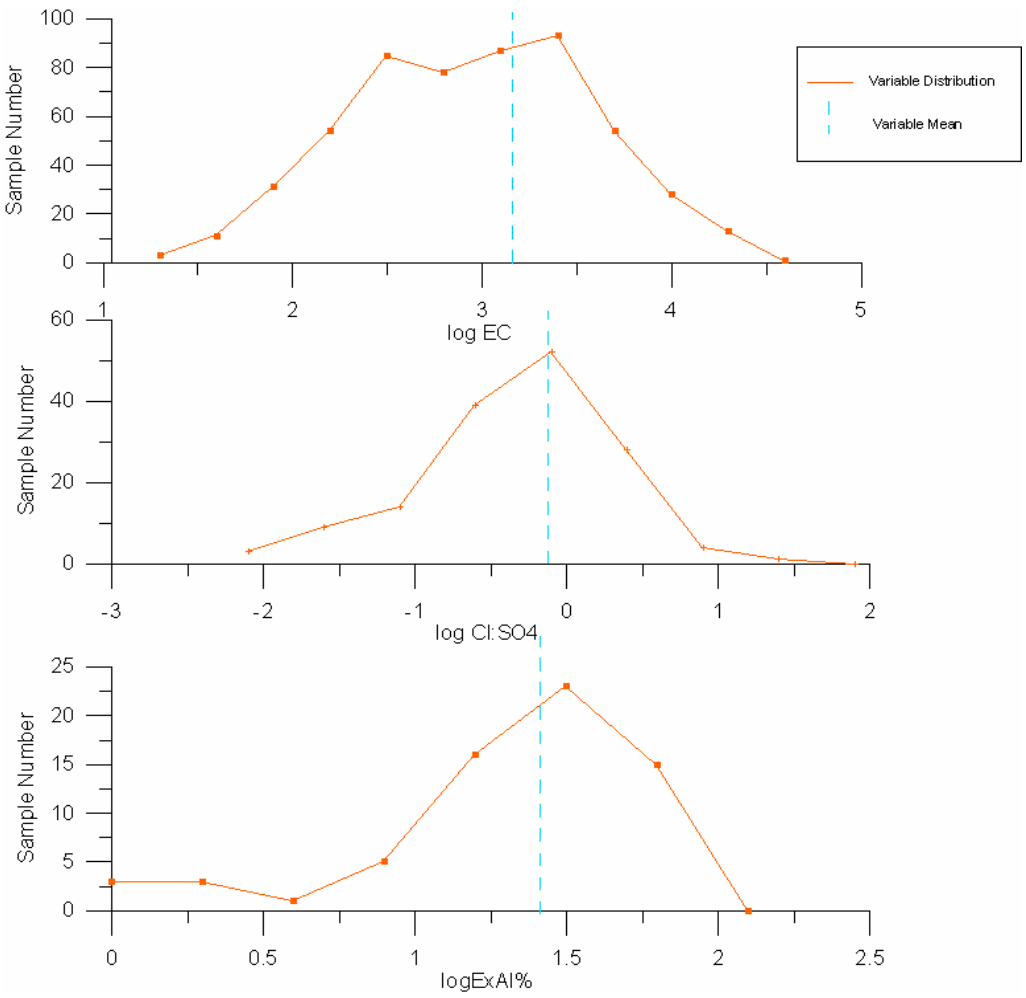

Figure 4.9 Log-Transferred Distribution of EC, Cl:SO_4 and logEXAl%.

4.5.4 Justification for Transformation

To ensure transformations have improved the relationship and the data set is normally distributed, the transformation must be substituted back into the original data set and compared to the actual values. This was done post multivariate analysis (see 5.2).

4.6 Outlier Identification

There are a number of reasons for outliers in the data. Such inconsistencies can cause the results of the analysis to be spurious, causing inconsistent interpretations.

Determining significant outliers from the sample population involves either assessing data mathematically or in a graphical representation. Theoretically there appears to be no major outliers in the data (Table 4.1), however the data must be evaluated using mathematical functions to grasp its distribution.

The Central Limit Theorem (CLT) is a suitable method to eliminate outliers and is theoretically described by Kallenberg (1997) that:

- The distribution of a sample mean is normal or approximately normal, even in cases where the parent population does not show normality
- When the mean of the population is μ and the standard deviation is σ, the mean of the sample should also be μ, and the standard deviation $\sigma / \sqrt{n}$ (where n is sample size from the population).

- 95% of the sample means lie within 2 standard deviations of the population mean or

$$\mu \pm 2.0 \left(\frac{\sigma}{\sqrt{n}}\right)$$

- Confidence intervals can be altered depending on the level of confidence sought (Table 4.7). Within the standard normal distribution, the equation is represented as

$$\bar{X} \pm z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right)$$

When n > 30, the z or the standard normal distribution is used to well approximate the t-distribution. In this case, Z is equal to one of the following scores as depicted in Table 4.7 depending on the level of confidence sought.

Level of Confidence	90%	95%	99%	99.9%
z-value	1.645	1.96	2.576	3.291

Table 4.7 Critical z scores (normal) for certain levels of confidence

The 95% confidence interval for pH is represented below, and the box-whisker plot (Figure 4.11) represents the visual outliers:

95% Confidence Interval (pH) = $\overline{X} \pm z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right)$

95% CI (pH) = 4.312 $\pm$ 1.96 (0.926)

= 2.498 $\leq \overline{X} \leq$ 6.127

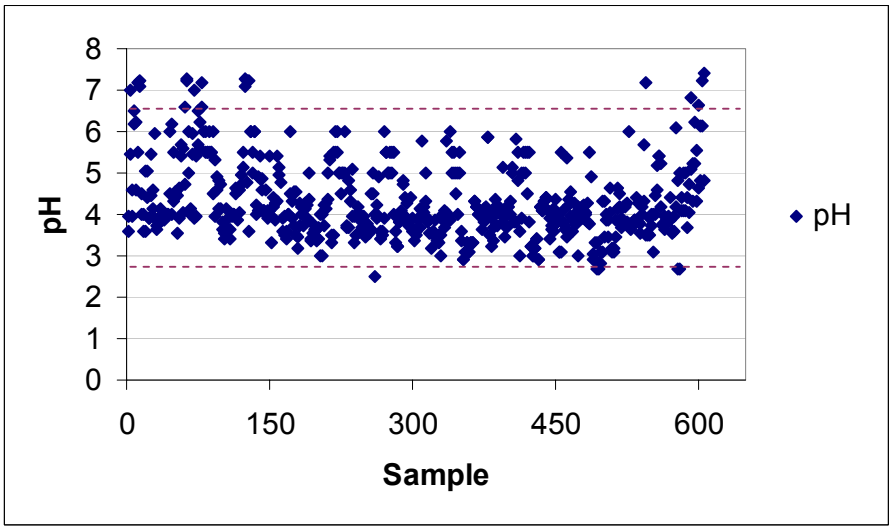

Figure 4.10 Distribution of pH (from 607 samples with the Confidence Intervals represented by the dashed lines (n-1 = 606).

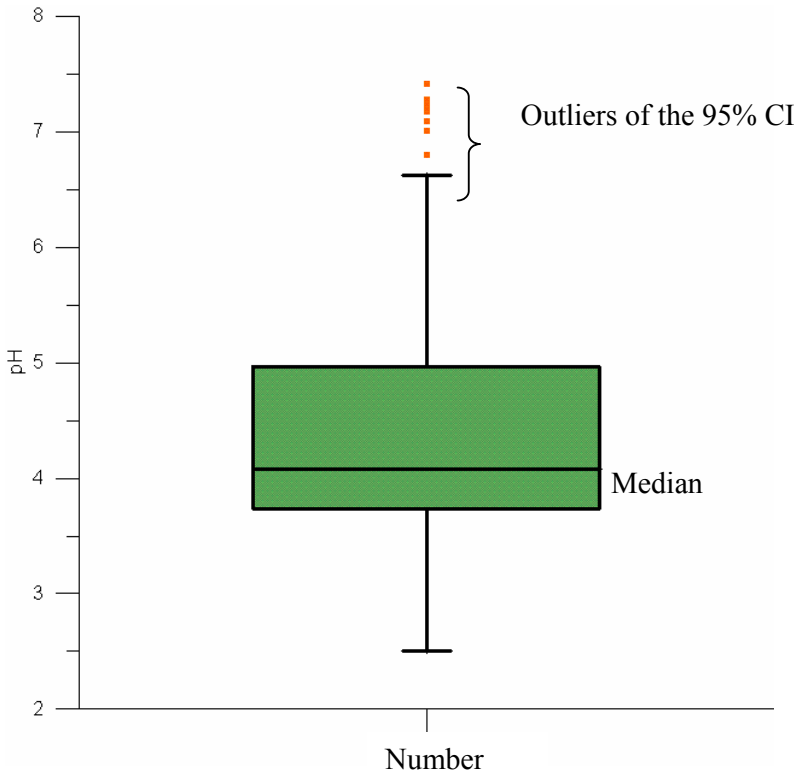

Figure 4.11 Box-Whisker Plot of pH (with 95% Confidence Interval, $2.498 \leq \overline{X} \leq 6.127$).

4.6.1 Univariate and Multivariate outliers

A Box-Whisker Plot (Figure 4.11) represents outliers based on the analysis of one variable (pH) in a data set. When trying to determine if there are outliers existing in relationships between 2 or more variables, it is necessary to analyse the data set with multivariate analysis tools such as scatterplots. Each variable was compared in the manner with the outliers in the bivariate relationship evident in Figures 4.12 and 4.13.

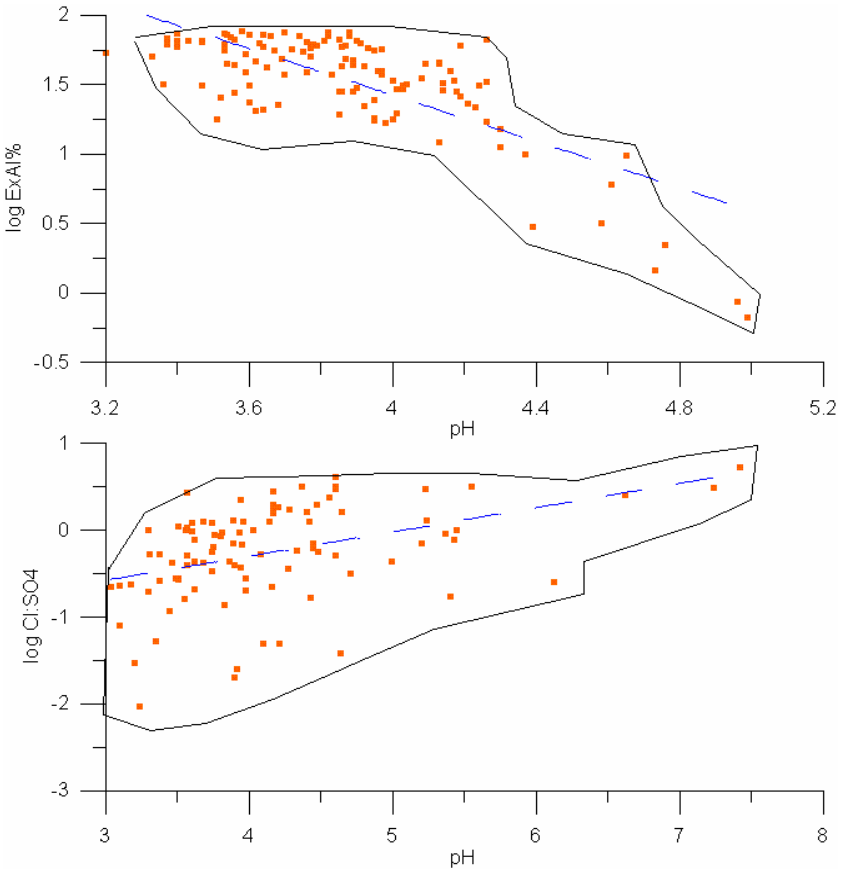

Figure 4.12 Homoscedasticity with both variables normally distributed (pH v logExAl, pH v log$Cl:SO_4$).

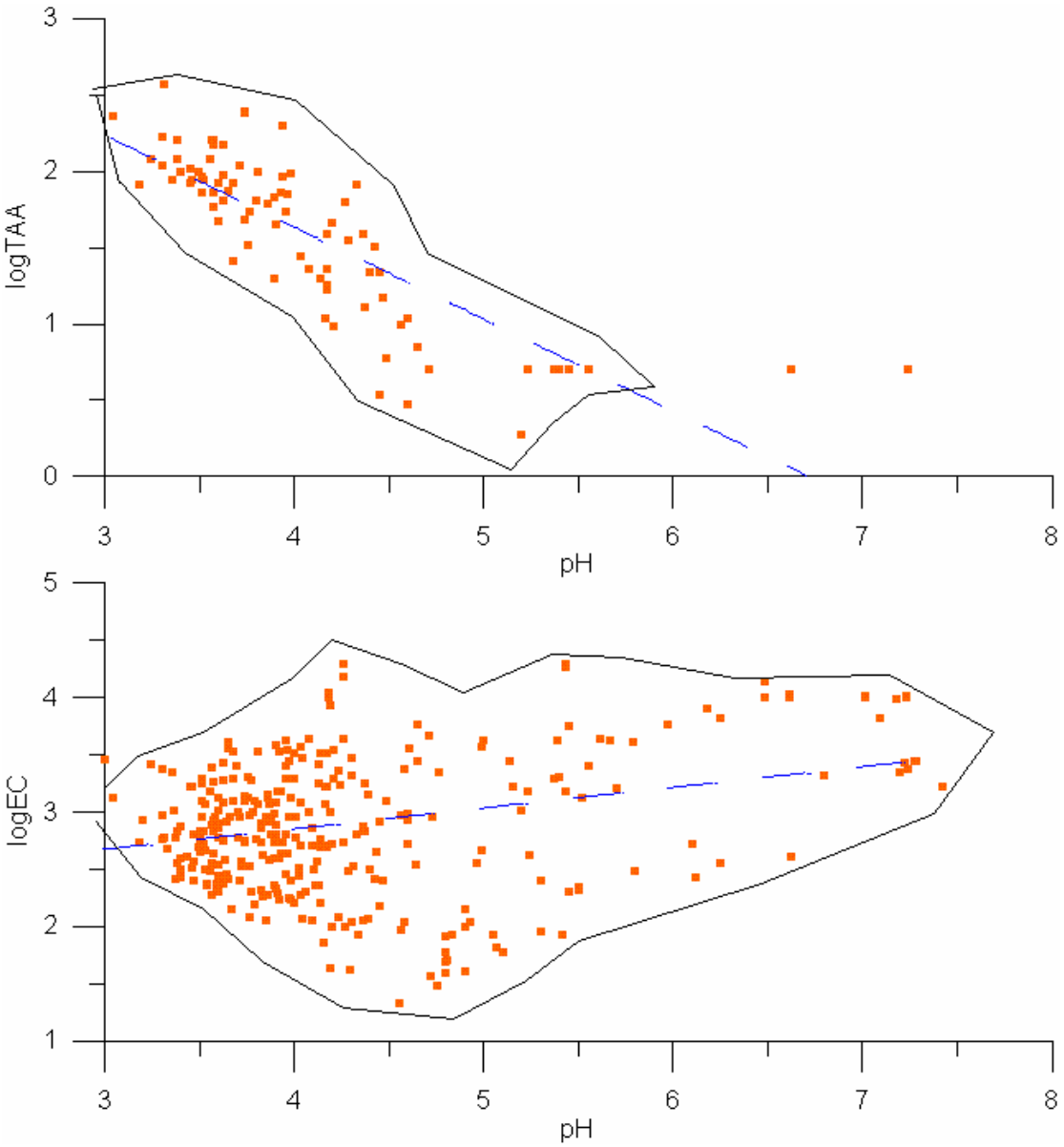

Figure 4.13 Homoscedasticity with both variables normally distributed (pH v logTAA, pH v log EC).

4.7 Nonlinearity and Heteroscedasticity

Homoscedasticity (Figure 4.14) assumes that the variance around the regression line is the same for all values of the predictor variable (X) whereas heteroscedasticity (Figure 4.15) violates this assumption. For the lower values on the X-axis, the points are all very near the regression line. For the higher values on the X-axis, there is much more variability around the regression line, whereas in homoscedasticity, the points remain central to the regression line.

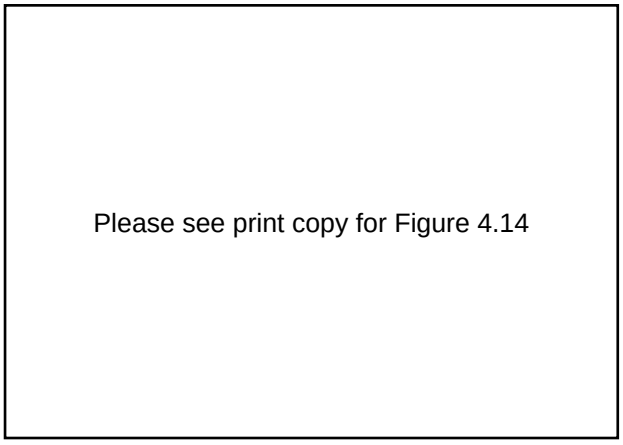

Figure 4.14 Homoscedasticity.

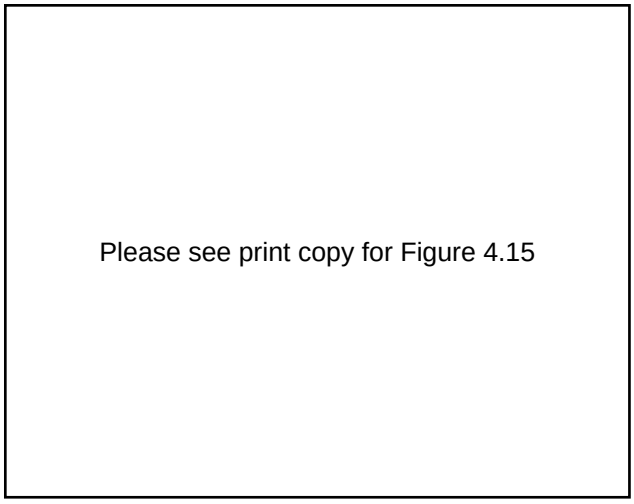

Figure 4.15 Heteroscedasticity (After Wulder, 2002).

For the data set in the project, heteroscedasticity was evident in most relationships prior to the log-transformation. However, after the log-transformation of the variables, the level of heteroscedasticity fell and the distribution became closer to exhibiting homoscedasticity (Figures 4.12 and 4.13).

4.8 Multicolinearity and Singularity

When multicolinearity or singularity exists within a data set there is a need to remove one of more variables from the data set (Wulder, 2002). Multicolinearity occurs in a data set when variables have a high correlation to one another, usually when $r > 0.90$, however Wulder (2002) suggests removing variables that exhibit bivariate correlation of above 0.70. Singularity occurs when the variables are perfectly correlated ($r = 1$). If both of these variables are included in a data set then this can significantly affect the results by reducing the degrees of freedom.

Looking at the correlations of the data set used in this project (Table 4.5) it is evident that from the log transformed correlations that there are no two variables that fall into the true definition of singularity or multicolinearity. However, logTAA and pH are correlated above $r = (+/-) 0.70$, and so one of these variables needs to be eliminated from the analysis before we can proceed further. As mentioned previously in this study, pH values are much more readily available than TAA, and the cost of obtaining further pH is lower than obtaining TAA. As TAA is a more predictable measure of acidity within a system, the relationship between TAA and pH will be used to predict unknown TAA values from known pH values.

Within SAS (or SPSS) multicolinearity is represented by a high Variance Inflation Factor (VIF). The VIF ranges from 1 to infinity. Tolerance of a variable is another statistic that is generated in SPSS/SAS which also is related to the level of multicolinearity. A tolerance of 1 indicates independence, whereas a high tolerance indicates multicolinearity between the variables. Belsley *et al.* (1980) suggest that a variable with a VIF greater than 10 should be excluded from the model (i.e $r^2 > 0.9$) as multicolinearity exists between the two variables.

	pH	logTAA	logEC	logExAl/CEC	$logCl:SO_4$
pH	-	2.838	1.088	2.231	1.207
logTAA	2.838	-	1.017	N/A	1.097
logEC	1.089	1.017	-	1.169	1.001
logExAl/CEC	2.231	N/A	1.169	-	1.325
$logCl:SO_4$	1.207	1.097	1.001	1.325	-

Table 4.8 VIF values between variables used in the project

The highest possibility of multicolinearity exists between pH and logTAA (VIF = 2.8377) as represented in Table 4.8. Although the value is not above the suggested VIF of 10, it is theoretically significant enough to eliminate one of these variables. Therefore, pH will be used in the predictive model:

logTAA = 4.04 – 0.60 pH (r^2 = 0.648) ……………………... ………….4.1

However, when using this model it became evident that the outliers exposed by Figure 4.11 would affect the predicted logTAA values. Therefore, these points were eliminated from the data set and the new VIF was 3.2787. The function to predict logTAA was represented as:

logTAA = 4.37 – 0.70pH (r^2 = 0.695) ………………….…………………4.2

This was further improved with the elimination of those points not included in the 95% Confidence Interval represented in 4.6. Any point with a pH greater than 6.127 was removed, with the resulting relationship:

logTAA = 4.81 – 0.80pH (r^2 = 0.733)……………………...……………4.3

As a result the VIF increased to 3.74 between these two variables. So after the removal of outliers, the sample size changed from 1377 to 1351. The prediction of logTAA is represented in Figure 4.16.

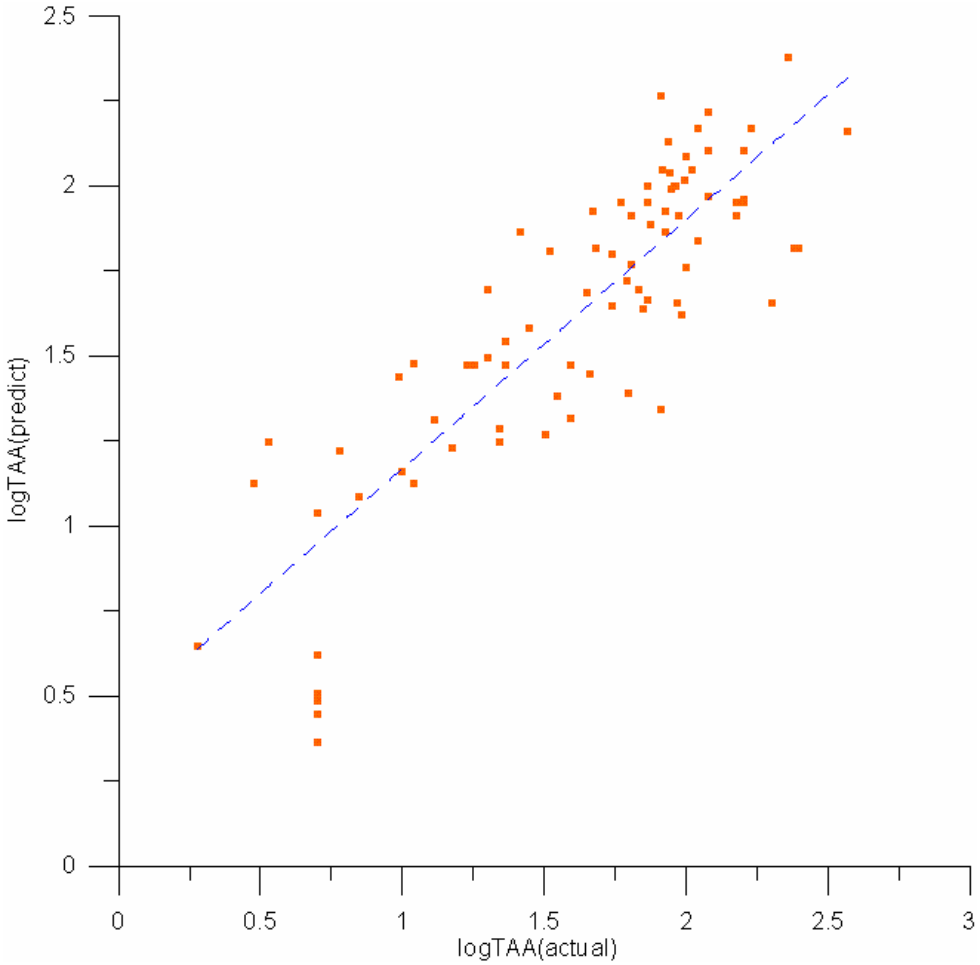

Figure 4.16 Prediction of logTAA using Equation 4.3.

4.8.1 Other Variable Transformations - ExAl/CEC

Compared with other variables, ExAl/CEC had little overlap at the same location, hence it is difficult to use the actual values of ExAl/CEC in multivariate regression. So that ExAl/CEC can be used in a multivariate regression, the prediction of log (ExAl/CEC) from log X and log Y (see Table 4.5) will generate 1351 predictive points of ExAl adhering to the Best Linear Unbiased Estimator principles. The predictions are compared to determine how well the linear regression model predicted (Figure 4.17 and 4.18). Figure 4.19 represents the residual error, showing the predictive model meets the assumptions of normality (Moore & McCabe, 1993). Outliers with ExAl/CEC above 100% were removed and the predictive data set reduced to 1274 samples.

$$\log (ExAl/CEC) = 5111.62 + 246.879 * \log (X) - 1064.70 * \log (Y) + e \quad \text{.....4.4}$$

$r^2 = 0.518$ ($F = 74.77$, $pr < F\ 0.0001$)

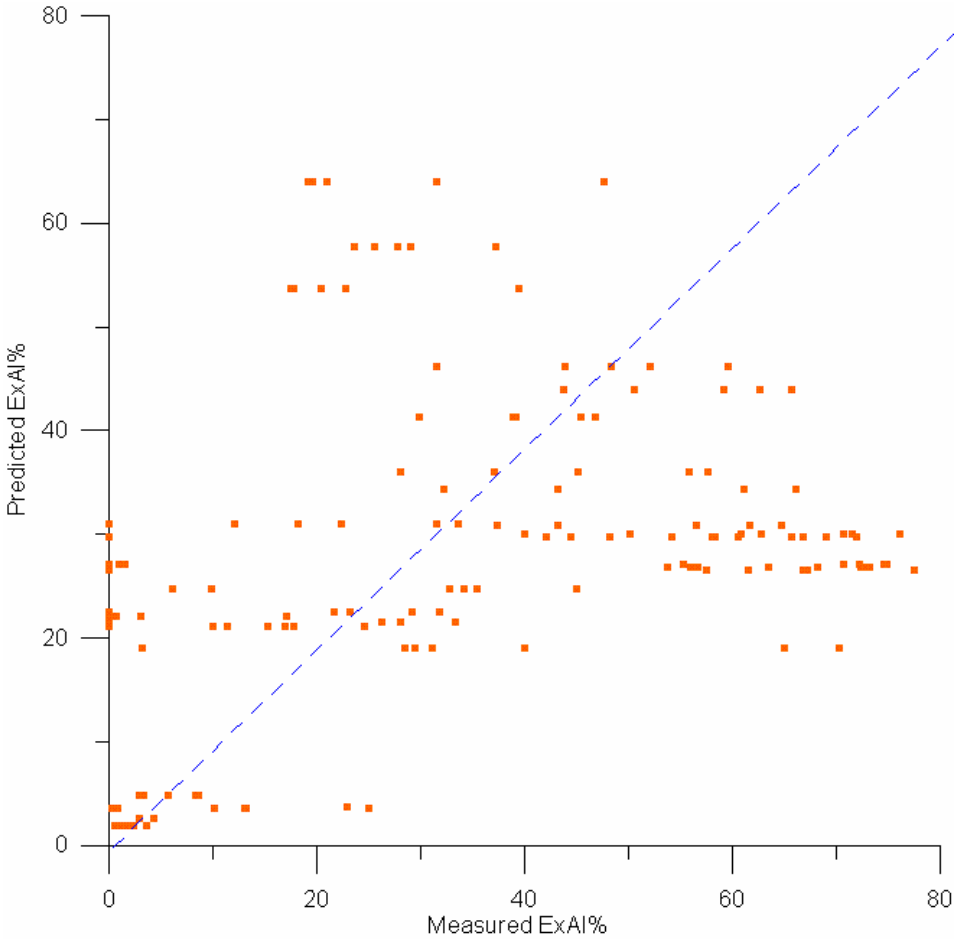

Figure 4.17 Prediction of ExAl/CEC using Equation 4.4.

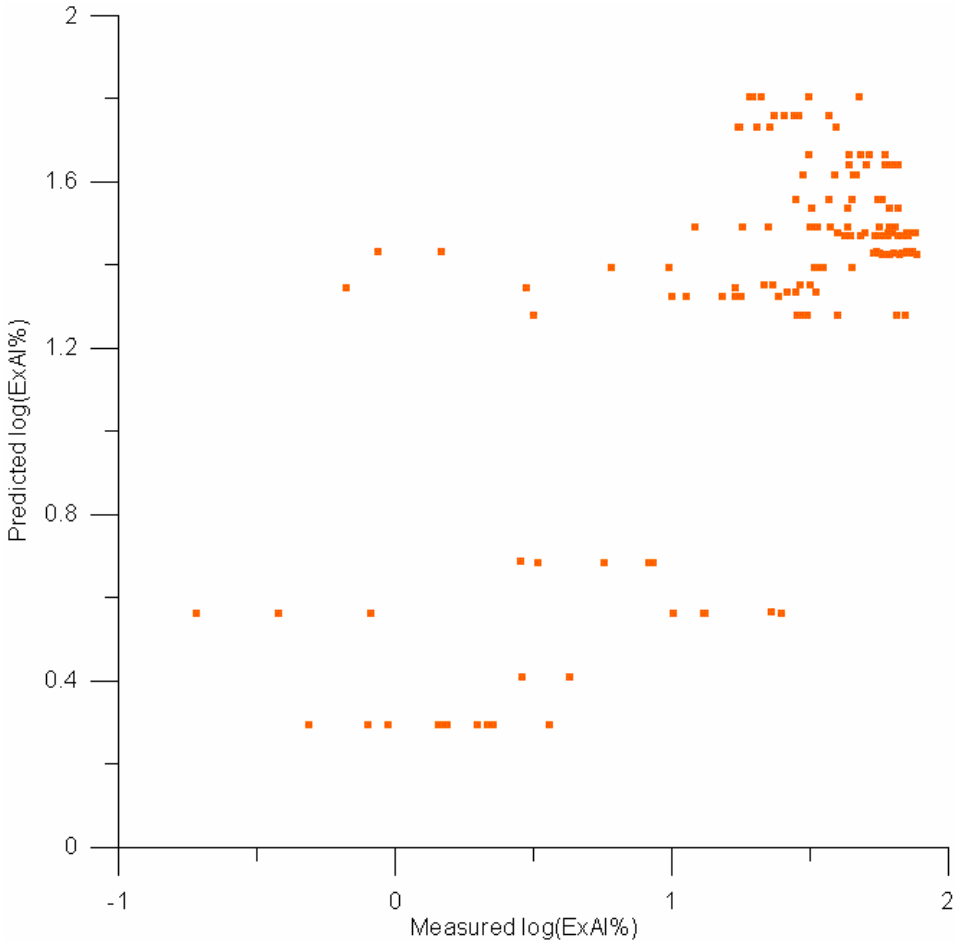

Figure 4.18 Prediction of log (ExAl/CEC) using Equation 4.4.

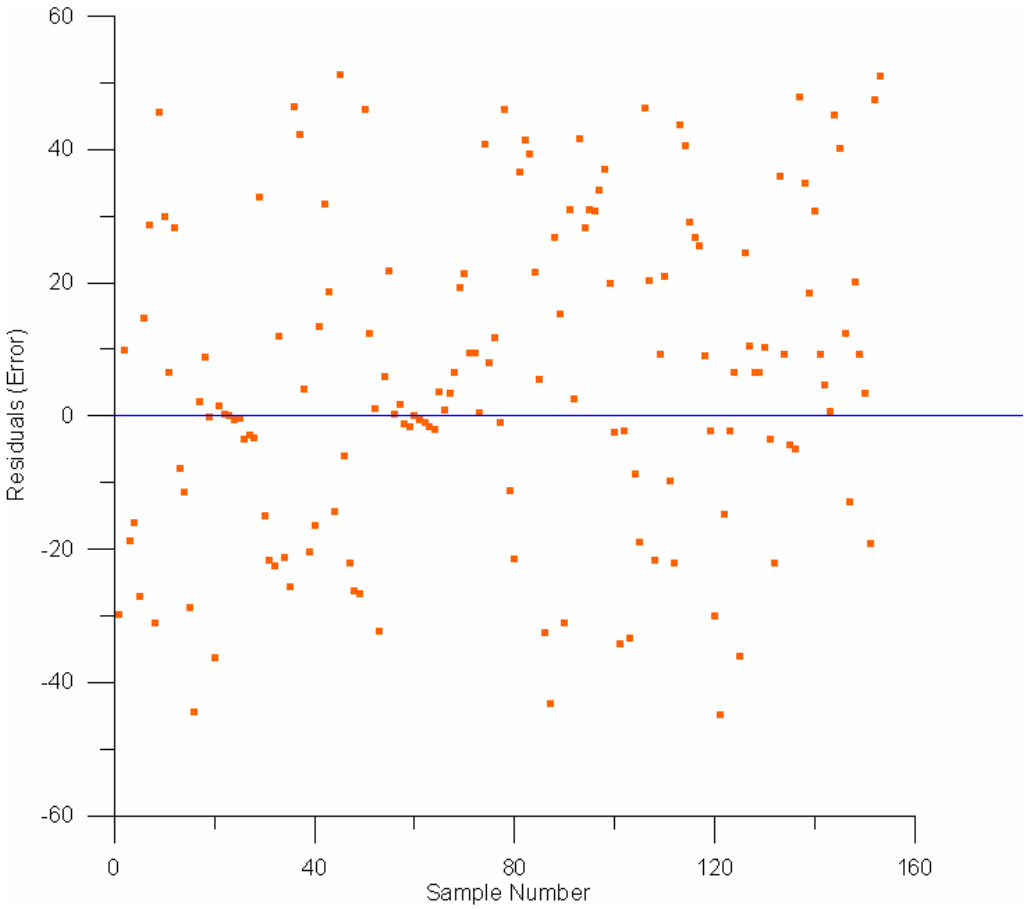

Figure 4.19 Residuals (Error) of ExAl/CEC using Equation 4.4.

Therefore, two scenarios will be introduced. One will include the prediction of ExAl/CEC using log (X) and log (Y), and the other scenario will leave ExAl/CEC out of the analysis.

$\log Cl{:}SO_4 = -0.532 + 0.617 \log (EXAL) - 0.097 \log (TAA) - 0.200 \log (EC) + e$…….4.5

$r^2 = 0.378$, ($F = 16.44$, $pr < F\ 0.0001$)

$\log Cl:SO_4 = 0.0788 + 0.0229 \log (EC) - 0.2501 \log (TAA) + e$ ……..4.6

$r^2 = 0.066$, (F = 2.9, pr > F 0.0609)

4.8.2 Data Summary

The data has been analysed, transformed and altered to adhere to normality as well as the best, linear, and unbiased predictor assumptions. As a result the number of samples to be used in the multivariate analysis and the general descriptive statistics of a number of these variables has changed. Table 4.9 represents the number of samples and associated descriptive statistics. The data range has significantly improved as has the number of classes in the EXAL_P and TAA_P variables.

Variable	N	Mean	Standard Deviation	Minimum	Maximum	Median
INT1	1345	6.616	3.115	1.000	15.000	7.000
X_ISG	1345	268042.270	2261.970	264698.170	272771.340	266849.510
Y_ISG	1345	1144599.500	1920.010	1141935.980	1149469.270	1144173.340
EXAL_P	1267	15.259	13.237	0.351	77.322	11.733
TAA_P	572	56.194	65.359	0.812	445.010	40.476
EC	504	3145.050	5335.790	40.000	41030.000	1235.000
CLSO4	144	1.517	1.755	0.009	10.000	0.988
logX	1345	5.428	0.004	5.423	5.436	5.426
logY	1345	6.059	0.001	6.058	6.060	6.058
logExAL	1267	1.000	0.435	-0.455	1.888	1.069
logTAA_pred	572	1.440	0.608	-0.091	2.648	1.607
logEC	504	3.069	0.638	1.330	4.613	3.092
logCl_SO4	144	-0.125	0.603	-2.029	1.000	-0.005

Table 4.9 Variables to be used in multivariate analysis

4.9 Spatial autocorrelation

Spatial autocorrelation shows the correlation of a variable in relation to its location. It measures the level of interdependence between variables in a geographical area. Variables can be seen to exhibit positive, negative or no spatial autocorrelation. Positive spatial autocorrelation aggregates similar values into a certain location, whereas negative spatial autocorrelation groups dissimilar values together. Negative spatial

autocorrelation is more sensitive to changes in scale. Geographical data is usually seen to exhibit positive spatial autocorrelation (Wulder, 2002). Spatial autocorrelation is useful in determining the effectiveness of an interpolation technique (see section 4.3).

Within the data set used, there is a positive spatial autocorrelation of 0.806.

Now that the data has been scrutinized and aptly transformed, it is ready to be used in various regression processes and in predictive modelling equations.

4.10 Spatial Visualisation of Data

Displaying data in a map form can give a better overview as to how the data is distributed spatially and will determine if there is any bias in the distribution. This is very important for further spatial data analysis. A dot map (Fotheringham, 2002) is produced for the soil sampling locations in Figure 4.20.

4.11 Evaluation of Preliminary Data Analysis

The preliminary data analysis process explained by Wulder (2002) enables a data set to be critically analysed through a structured process. This process points out the inadequacies in the data set and can suggest some ways of improving the dataset (e.g. through transformation). Running through a preliminary data analysis is vital so that when putting the data into a more intensive statistical analysis method the results generate the lowest error, with the least bias as possible.

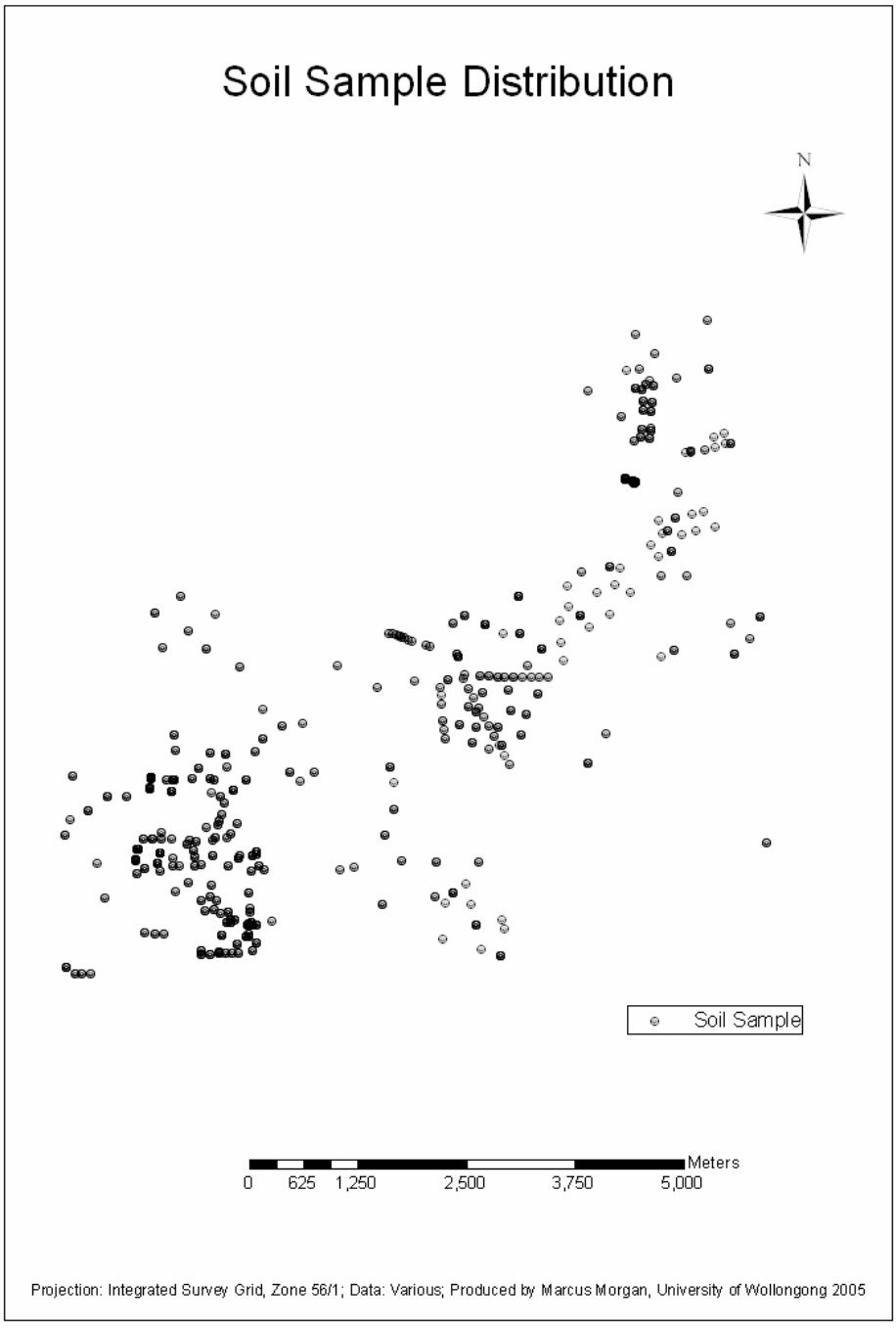

Figure 4.20 Map of soil sampling locations in Broughton Creek.

Multi-parameter analysis requires assessment of the data to ensure that certain assumptions such as linearity, unbiasedness, non-multicolinearity and homoscedasticity are met. The three main techniques explored to assess the data are:

1. Multiple Correlation and Regression (see 5.2)
2. Geographically Weighted Regression (GWR) (see 5.3)

And

3. Geostatistical Analysis (see 5.4)

These are analysed in the following Chapter for suitability to this application. Chapter 6 will build on from this by using the most appropriate tool, either the global analysis multiple correlation relationships, the local analysis tool of GWR, or interpolating between available points to build a model(s) to predict the severity of the Shoalhaven floodplain.

Chapter 5: Spatial Statistical Analysis – Choosing the Optimal Method

There are a number of statistical methods that further explore multivariate data sets. However there are relatively few that take into account the spatial positioning of the data, as well as attempting to predict the unknown from known parameters. A number of suitable multivariate tools were used including classical multiple correlation and regression, geographically weighted regression, and geostatistical analysis. Other suitable tools for spatial analysis include principal components and factor analysis, discriminant analysis, and cluster analysis. These latter methods were deemed unsuitable for this project and therefore were not used. For instance, one of the main aims of using principal component and factor analysis is to reduce the number of variables to determine the most influential component. In determining CASS severity, each variable used in the project has an equal weighting in terms of its influence on predicting whether CASS exist at that location. Other more applicable methods that look at all the data on a local level or across an entire area were considered more suitable.

5.1 Local versus Global Statistical Analysis

The focus of this project is to use GIS to predict the severity of CASS in un-sampled areas of a floodplain. Local statistical methods are important to GIS because they produce values for each location, and these values can be displayed in map form. Compared to global statistical methods, which look at generating an average set of results from the data, local analysis or local modelling focuses on testing the presence of differences across space, instead of assuming they don't exist (Fotheringham, 2000).

Local statistical analysis enables the discovery of hot spots in the data, or data that cannot be generalised over an area. This is relevant to the determination of CASS severity over a significant area of land and differs from Global statistical analysis that summarises data for a whole region. Table 5.1 summarises the differences between local and global approaches to statistical analysis.

Please see print copy for Table 5.1

Table 5.1 Comparisons of Local versus Global Statistical Approaches (After Fotheringham, 2002)

By analysing a multivariate data set, the aim is to determine whether each variable has a relationship with one another and if so, the strength of this relationship. This is typically shown in regression analysis. However, classical multivariate analysis doesn't look at how the changes in geographical location affect the fitting capacity of the model. Classical multivariate regression also doesn't take into account the three main reasons why relationships between variables differ over space. These include:

- Random sampling causes spatial variations by the differing sampling regimes

- Relationships may be intrinsically different across space due the nature of the variables (e.g. variations in geology causing variations in soil morphology).

- The representative model formed is erroneous or misrepresents the actual relationships seen in theory and practice.

The problem represented in this project is a spatial multivariate soil science problem. Soils are rarely uniform across large areas and past research within the study area (Blunden, 2000; Lawrie & Eldridge, 2002; Glamore, 2003). This research has supported

the theory that CASS can vary from paddock to paddock, or spatially separated by only 10's of metres. Contrasting global methods of multivariate analysis using local methods can show how geographic location of a sample can affect the outcome of a predictive model.

5.2 Multiple correlation and regression

Multiple regression is the combination of a number of independent variables (>2) that are combined to predict a dependent variable. The form of multiple regression is:

$$Y = A + B_1X_1 + B_2X_2 + ... + B_nX_n + e \quad \text{.....5.1}$$

where:

Y = Predicted Dependent Value,
A = the y-intercept
X = Independent Variable
B = beta-coefficients (slope) of Independent Variable (Regression Coefficients)
e = an error term.

In fitting a multiple regression model, it is assumed that observations are independent of one another and that the structure of the model is constant over the study area. This is not the case for most spatially referenced data. This will be explored further in 5.3, where these assumptions no longer hold true.

The predictive function (5.1) represents the strength of the independent variables compared to the dependent variable. An optimal model exhibits low residuals between predicted and actual Y-values and is usually represented graphically for validation.

Within a multiple regression analysis, the dependent variable can be predicted from the combination of all other variables (multiple regression) or a relationship between a certain independent variable that controls the other independent variables (partial regression).

Wulder (2002) suggests that in a typical multiple regression analysis, there should be a minimum 20 samples used for there to be a meaningful outcome, although states that 5

can be used if the parameter being measured does not vary too much over the selected area.

Since the predicted value is different depending on the combinations of independent variables used in the model, the aim of a regression is to determine the regression coefficient (B). The major advantage of determining B is to:

- Minimise deviations of residuals between the predicted and observed values for the dependent variable
- Optimise the correlation of predicted dependent variables and actual values of the dependent variable.

When generating the B coefficients, T-tests are used to evaluate the significance of the individual B coefficients, testing the null hypothesis that the regression coefficient is zero. Variables that are not significant at the p = 0.05 level are normally eliminated from the regression model.

The main focus for this project is using three independent variables to develop a model for the dependent variable that is most difficult and costly to obtain. The three variables were determined to be independent after being log-transformed and further analysed in the data screening process (see Chapter 4). From the bivariate relationships in (4.6.1)

$$\log (Cl{:}SO_4) = B_0 + B_1 \log(EXAL) + B_2 \log (TAA) + B_3 \log (EC) + e \ldots\ldots\ldots\ldots 5.2$$

5.2.1 Multivariate Linear Regression Models

Two multivariate linear regression models, defining log ($Cl{:}SO_4$) and log (ExAl), that is the dependent variables that are most difficult and costly to obtain are represented below:

$$\log (Cl{:}SO_4) = -0.968 - 0.493 \log (EC) + 1.604 \log (EXAL) - 0.160 \log (TAA) + e \ldots\ldots\ldots 5.3$$

($r^2 = 0.252$, F = 6.95, pr > F 0.0004)

$$\log (ExAl) = 0.515 + 0.300 \log (Cl{:}SO_4) + 0.153 \log (EC) + 0.059 \log (TAA) + e \text{…….} 5.4$$

($r^2 = 0.542$, F = 24.48, pr > F 0.0001)

$$\log (Cl{:}SO_4) = 0.0788 + 0.0229 \log (EC) - 0.2501 \log (TAA) + e \text{……..} 5.5$$

$r^2 = 0.066$, (F = 2.9, pr > F 0.061)

When comparing the two models (5.3 and 5.4) with log ($Cl{:}SO_4$) as the dependent variable of equation 5.3 and log (ExAl) as the dependent variable of equation 5.4, it becomes clear that these two variables have the greatest influence when they act as the independent variable in the two models. This is supported by equation 5.5 which uses only two independent variables to predict log ($Cl{:}SO_4$), removing log (ExAl) from the model. As a result r^2 in equation 5.5 is 0.066 compared to 0.542, and the model is no longer statistically significant as the p-value of the F-test is greater than 0.05.

Overall the model predicting logExAl has the better goodness-of-fit. From these relationships we can conclude that on 25% of the time, the variables in equation 5.3 will predict log ($Cl{:}SO_4$) correctly and on 54% of the time they predict log (ExAl) correct in equation 5.4. Both of these models are statistically significant.

5.2.2 B-Coefficients

The B-coefficient is important in determining the strength of the variables and relationship to the other variables – that is the higher the B-coefficient (partial coefficient), the stronger the variable.

In 5.3, log (ExAL) has the greatest effect on the dependent variable log ($Cl:SO_4$), followed by log (EC) and log (TAA).

In 5.4, predictably, log ($Cl:SO_4$) has the greatest effect on the dependent variable log (TAA). Once again, log (EC) has the second greatest effect followed by log(TAA).

Testing theoretical relationships against the above relationships will also determine the validity of our models in predicting the dependent variable. Table 5.2 explains the relationships found in typical CASS environments as determined by past and on-going research (Blunden 2000; Lawrie & Eldridge, 2002; Glamore,2003).

Independent Variable	**Dependent Variable**	
	$Cl:SO_4$	ExAl
ExAl	Inverse below 3	-
TAA	Inverse below 3	Positive
EC	Positive	Inverse
$Cl:SO_4$	-	Inverse below 3

Table 5.2 Theoretical relationship between variables

Assessing the B coefficients show that in the first model (5.3), log$Cl:SO_4$ has a positive relationship with logExAl and logEC, and an inverse relationships with logTAA. That is when $Cl:SO_4$ is high, EC and ExAL are high and TAA is low. Theoretically when $Cl:SO_4$ is above 3 the quality of the soil-water is good and therefore TAA would be low. ExAl should also be low, however the relationship generated from 5.3 shows the opposite, a one unit increase to $Cl:SO_4$ produces a 1.604 unit increase in ExAl. EC should also be positively related to $Cl:SO_4$, representing when there is more Cl in the system there are more salts and a higher conductivity. This indicates that this model may not be very useful in predicting $Cl:SO_4$.

In the model represented by the function in 5.4, a one unit increase in logExAl would decrease logTAA by 0.1126. This is contradictory to table 5.2 representing a model that is not very accurate. This is also the case for $Cl:SO_4$ and ExAl, representing a positive relationship in the model, yet representing an inverse relationship in theory (see Table 5.2).

The results above indicate the importance of the B-coefficients in determining whether a model is theoretically sound. Although the goodness-of-fit is better for 5.4 than 5.3, the model is theoretically incorrect and cannot be used as a plausible estimate of ExAl. Both models have flaws which will be addressed by other statistical models.

5.2.3 ANOVA – Analysis of Variance (F-test)

The ANalysis Of VAriance (ANOVA) tests the null hypothesis that all the population means are equal. The F-ratio is computed for the ANOVA test of significance and represents the following function:

$F = MSB/MSE$……..5.6

Where: MSB = Mean Square Between

MSE = Mean Square Error

The p value when reported with the F-ratio is representative of a test that states if there are two populations that had identical variances, what chance is there of obtaining a bigger F-ratio. In this case the null and alternate hypothesis would be:

$H_0: \mu_1 = \mu_2 = \ldots = \mu_{a\ldots}$(i.e. F-ratio will be close to 1)

$H_1: \mu_1 \neq \mu_2 \neq \ldots \neq \mu_{a\ldots}$(i.e. F-ratio will be a lot greater than 1)……..5.7

The F-ratio determines the significance or the hypothesis (as represented in equation 5.7) of the models in equation 5.3 and 5.4. In the two combinations tested in equations 5.3 and 5.4, the models were significant at the p = 0.05 level. The F-ratio was high in both (6.95 and 24.48 respectively) and the probability of greater than F-ratio was 0.0004

and 0.0001 respectively. This indicates that for equations 5.3 and 5.4 we reject the null hypothesis at the 0.05 level, and conclude that the model is significant. For equation 5.5 we cannot reject the null hypothesis as p is greater than 0.05 (0.061) and therefore this model is not statistically significant at the p = 0.05 level.

5.2.4 Limits to Multiple Correlation and Regression Analysis

Multiple Correlation and Regression analysis is a global statistical method that can predict for a dependent variable based on a number of input independent variable. However, the output doesn't produce a comprehensive layer which is the aim of the project - to show the most probable and severe acid sulfate soil areas in a floodplain. Interpolation methods such as Inverse Distance Weighting (IDW) can address these problems but this will require further analysis and further tools which are covered by Geographically Weighted Regression and Geostatistical principles. Global models such as those built on multiple regression analysis also don't take into account the variations seen in the data set on a local level.

5.3 Geographically-Weighted Regression (GWR)

After analysing the multiple regression methodology and applying such principles to the data set, we return to a local statistical analysis tool - Geographically-Weighted Regression (GWR). This method also includes general principles from multiple regression but aims to focus at the individual level.

As spatial data analysis can often be influenced by location or a geographically referenced area, other methods that incorporate both this fact and use traditional tools of multivariate analysis would be more appropriate for the analysis. GWR uses a methodology similar to that used in spatial interpolation which says that nearby individual samples (cases) should influence the regression equation more than individual samples (cases) further away. This creates a more difficult analysis process compared to analysing non-spatial data sets where each individual case is entered and weighted equally into the regression function.

To explain GWR in mathematical functions it is easier to compare the classical multiple regression model. For instance, if we take equation 5.1 to be the defining global model, then in order to represent this model as a local model a number of changes would have to be made. As explained earlier, most geographical data doesn't have a string of independent variables that are truly independent, and most global models such as multiple regression don't take into account the local variations as well as global. Therefore, to address this as GWR does, equation 5.1 can be written to include local variation (Fotheringham *et al.*, 2002):

$$Y(G) = A(G) + B_1(G)\,X1 + B_2(G)X_2 + \ldots + B_n(G)X_n + e \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 5.8$$

Where (G) indicates that the parameters to be estimated at a location whose coordinates are given by the vector G.

In Ordinary Least Squares (OLS), the purpose is to minimise the Residual Sum of Squares while obtaining the estimator B. Applying this principle to GWR can define the parameters for the model in equation 5.6. Therefore the estimator is defined as:

$$B = (X'X)^{-1}X'Y \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 5.9$$

As Fotherinham *et al.* (2002) suggests there are problems in calibrating such a model because there are more unknowns than observed variables. Since GWR gives more influence to those samples that are closer to the estimated parameter than those further away, there is a problem of bias. Trying to validate the model involves a fine balance between standard error and the level of bias. That is, the greater the local sample size the lower the standard error of the coefficient estimates and higher the bias.

To reduce the effect of such bias and error, the Weighted Least Squares approach is combined with traditional OLS to generate the GWR approach which weights

observations in regard to their proximity to the estimated parameter, that is the closer the observation the greater the weighting. This generates the following:

$$B(G) = (X'W(G)X)^{-1}X'W(G)Y \qquad 5.10$$

Where W is represented by:

$$W(G) = \begin{bmatrix} w_{11} & 0 & \Lambda & 0 \\ 0 & w_{i2} & \Lambda & 0 \\ M & M & \Lambda & M \\ & & \Lambda & \\ 0 & & \Lambda & w_{in} \end{bmatrix} \qquad 5.11$$

Calculating the weights involves substituting a mathematical function. Fotheringham *et al.* (2002) defined the Gaussian weight for the i-th observation to be

$$w_i(G) = \exp(-d/h)^2 \qquad 5.12$$

Where d = Euclidean distance between the location of observation i and location G,

And

h = the bandwidth (which may be defined by the user or determined from cross validation (see 5.6)).

After determining the weights and substituting into equation 5.10, the local variations in the predicted outcomes can be viewed and analysed in ArcGIS. In using GWR, hypothesis testing is possible. The variations in a study area could be evaluated to determine whether they occur by chance or not. GWR 3.0 developed by Fotheringham *et al.* (2002) is a useful analysis tool that proceeds through the steps required to perform a geographically weighted regression on the data set.

5.4 Spatial Interpolation

There are two main types of spatial interpolation that aim to predict unknown values between known points. These are separated into deterministic and geostatistical methods, respectively (Johnston *et al.*, 2001).

5.4.1 Deterministic Interpolation Techniques

Deterministic Interpolation methods create surfaces from known points based on similarity or degree of smoothing (i.e. Inverse Distance Weighting (IDW) or Radial Basis Function (RBF) respectively). Deterministic methods of interpolation are grouped into local and global methods, and predict as either exact (IDW, RBF) or inexact (Local Polynomial Interpolation (LPI), Global Polynomial Interpolation (GPI)) interpolators. This project will explore the main techniques used in ArcGIS which include three local methods (IDW, LPI and RBF) and one global method (GPI).

5.4.1.1 GPI – Global Polynomial Interpolation

The Global Polynomial Interpolation (GPI) method fits a plane between the sample points based on the overriding trend by using a polynomial function. This plane and surface depicted by the polynomial changes gradually and captures coarse scale pattern in data. There are a number of polynomial functions that can be supported in ArcGIS. These include from first order polynomial all the way to 10^{th} order polynomials. GPI is an inexact interpolator as the functions rarely pass through existing or measured values. The GPI method is quite useful when attempting to fit a surface that gradually changes over a spatially vast area (e.g. elevation, air pollution) and is also used for investigating and removing the effects of long range global trends (Johnston *et al.*, 2001). This type of interpolation would not suite our analysis as soil properties can change quite suddenly over a small area.

5.4.1.2 IDW – Inverse Distance Weighting

Inverse Distance Weighting (IDW) interpolation is an exact interpolator that assumes measured values closer to the areas being predicted have more influence than those

further away. IDW assumes each measured value has a local influence on the predicted outcome, and this influence is reduced the further the predicted point is from the known point. This explains how weights are assigned in IDW, the larger the distance, the smaller the weighting.

Unlike Kriging (see 5.4.2), IDW doesn't make assumptions about spatial relationships, except for the basic assumption that nearby points should be more closely related than distant points to the value at the interpolate location.

The process in using IDW begins by optimising power values. The optimal "p" value is determined by the p value that produces the lowest root-mean-square prediction error (RMSPE). Within ArcGIS, the optimal p value can be automatically determined at an early stage when setting up the model.

IDW is represented by:

$$\hat{Z}(x_0) = \sum_{i=1}^{N} \lambda_i z(x_i) \qquad 5.13$$

Where: $\hat{Z}(x_0)$ = the predicted value at location x_0.

N = number of points used in the prediction surrounding x_0.

λ_i = weights of points used in prediction

$\hat{Z}(x_i)$ = observed value at x_i.

And the weights are defined by:

$$\lambda_i = d_{i0}^{-p} / \sum_{i=1}^{N} d_{i0}^{-p} \qquad \sum_{i=1}^{N} \lambda_i = 1$$

After determining the power value, the neighbourhood must be defined to show the extent of the search capacity for the prediction. As values far away have little or no weighting on the end prediction, these values can generally be excluded. IDW enables a geographical boundary to be set to look at the points of influence. A number of options are available when defining the neighbourhood boundary. If the data set exhibits a

directional influence then this would need to be included in defining the search boundary. If there is no directional influence then points in all directions should be equally considered. Therefore the neighbourhood shape would be a circle. This could then be broken into segments to put a greater emphasis on variations on the local level. Section 6.3 explores this method of interpolation further by comparing models created using IDW.

5.4.1.3 LPI – Local Polynomial Interpolation

Similarly to IDW, Local Polynomial Interpolation (LPI) uses the neighbourhood search pattern to fit many polynomials that may overlap between other neighbourhoods. As with IDW the shape can be altered and the number of points specified. The difference between LPI and GPI is that GPI produces smooth surfaces based on one main function over the whole area, whereas LPI uses a number of polynomial functions and is sensitive to neighbourhood distance.

Optimising the model involves cross-validating the output surfaces calculated using different parameters (Johnston, *et al.*, 2001). The parameter that minimises the RMSPE is chosen. This is similar to the process of determining the p value in IDW.

5.4.1.4 RBF – Radial Basis Functions

Radial Basis Functions (RBF) methods like IDW are exact interpolators, meaning that the predicted surface passes through the measured values with the predicted points equalling measured points. However, RBF can predict above and below the minimum and maximum measured values unlike IDW. RBF are considered a form of artificial neural networks (Johnston, 2001). Using the input data (x_1, K, x_n) a neural network could perform a traditional function approximation using RBF as the mathematically function to determine the output, $\hat{y}$, which Figure 5.1 shows.

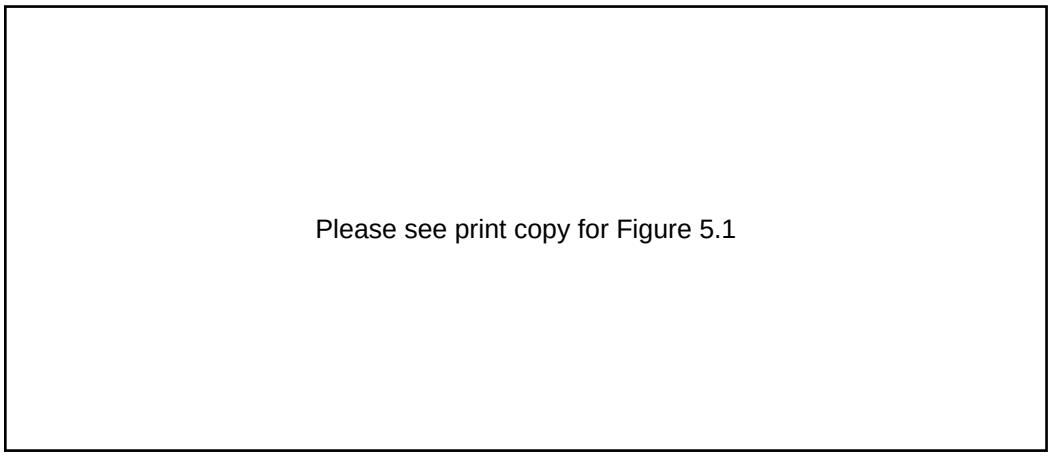

Figure 5.1 A RBF network (Gaussian bell function) with one output (Wolfram Research, 2005).

The RBF network, including a linear part, produces an output given by

by

$$\hat{y(\theta)} = g(\theta, x) = \sum_{i}^{nb} w_i^2 e_i^{-\lambda_i^2 \left(x - w_i^1\right)^2} + w_{nb+1}^2 + x_1 x_1 + \mathrm{K} + x_n x_n \qquad 5.14$$

Where: *nb* = the number of neurons, each containing a basis function.

w_i^1 = Basis functions

λ_i = Inverse of the width of the basis functions

w_i^2 = Weights in output sum

x_1, K, x_n = Parameters of the linear part (Wolfram Research, 2005)

The five main radial basis functions (w_i^1) that are input as the mathematical function to depict the predicted surface include:

1. Thin-plate Spline
2. Spline with Tension
3. Completely Regularised Spline
4. Multiquadric Function
5. Inverse Multiquadric Spline

RBF is quite useful in predicting slight changes over space such as small gradual elevation changes, as the radial functions overlap to form a gradual changing surface. For this project the scope is to determine the variety of changes in soil properties over a small area. Hence, RBF would not be suitable for our purposes.

5.4.2 Geostatistical Interpolation Techniques

Deterministic Spatial Interpolation tools are useful in a number of environmental scenarios such as with change in air pollution over a large area. However a lot of deterministic models fail to take into account data structure, as well as providing surfaces that give error across the whole geographical area. A number of the deterministic methods only account for gradual change over a large area which we saw in GPI and RBF. Geostatistical Interpolation methods on the other hand take into account local variation over a large geographical area and as well as producing predictive surfaces, also produce error or uncertainty surfaces, which indicate how well the predictors have performed. There are a number of geostatistical interpolation methods (Table 5.3) that can be applied to the data set. However, the most commonly used method, Kriging, is investigated for its applicability to predicting CASS severity.

Kriging & Cokriging	Predictions	Prediction Standard Errors	Quantile maps	Probability maps	Standard Errors of Indicators
Ordinary	X	X	X^	X^	
Universal	X	X	X^	X^	
Simple	X	X	X^	X^	
Indicator				X	X
Probability				X	X
Disjunctive	X+	X+		X+	X+

^Requires assumption of multivariate normal distribution +Requires assumption of pair wise bivariate normality

Table 5.3 Summary of the output of Kriging and CoKriging Methods (adapted from Johnston, 2001).

5.4.2.1 Kriging and Cokriging

Chrisman (2002) defines Kriging as a geostatistical technique used for interpolation that takes into account the spatial autocorrelation between points to give the optimal interpolation for that data set.

Cokriging is another interpolation technique that produces estimates of a distribution if a secondary parameter has been sampled more intensely than the primary parameter. If the primary parameter is difficult or costly to measure, then cokriging can improve interpolation estimates without having to more intensely sample the primary parameter (Isaaks & Srivastava, 1989).

There are a number of kriging techniques that can be used as predictors and to generate resultant standard errors of the prediction but most of the kriging techniques can produce probability maps, and Indicator, Probability and Disjunctive Kriging can produce Standard Errors of the Indicators (see Table 5.3).

To explain how a data set is processed through the kriging techniques, the theory of ordinary kriging is explained.

5.4.2.2 Theory of Ordinary Kriging (OK)

Before using Ordinary Kriging (OK), the data set must be analysed and parameters must be in a normal distribution.

The estimation of a random variable (Z) that is between many known variables ($z(x_1)$, $z(x_2)\ldots$, $z(x_N)$ at $x_1, x_2, \ldots, x_N$) over a spatial representation is the purpose for using kriging.

In OK the assumption is that the mean is unknown. In punctual estimation, estimating Z at a point say x_0 by $\hat{Z}(x_0)$ with the same support as the data by:

$$\hat{Z}(x_0) = \sum_{i=1}^{N} \lambda_i z(x_i) \qquad 5.15$$

With the expected error being

$E[\hat{Z}(x_0) - Z(x_0)] = 0$ and estimation variance

$Var[\hat{Z}(x_0)] = E\{[\hat{Z}(x_0) - Z(x_0)]\}^2$

$$= 2 \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_0) - \sum_{i=1}^{N} \sum_{i=j}^{N} \lambda i \lambda i \gamma(xi, xj) \qquad 5.16$$

Where:

$\gamma(xi, xj)$ = the semivariance of Z between points x_i and x_j.

$\gamma(x_i, x_0)$ = the semivariance of Z between the i-th point (x_i) and the target point (x_0).

For each kriged estimate there is an associated variance defined as σ^2 (x_0) which is defined in equation 1.2. Using a Lagrange multiplier (ψ) can determine the weights to minimize these variances.

By defining an auxiliary function

$$f(\lambda i, \psi) = \text{var}[\hat{Z}(x_0) - Z(x_0)] - 2\,\psi\left\{\sum_{i=1}^{N} \lambda i - 1\right\} \qquad 5.17$$

The partial derivatives of the function with respect to the weights will be

$$\frac{\partial(\lambda i, \psi)}{\partial \lambda i} = 0 \qquad \text{And}$$

$$\frac{\partial(\lambda i, \psi)}{\partial \psi} = 0$$

A series of N + 1 equations in N+1 unknowns is generated from the partial derivatives:

$$\sum_{i=1}^{N} \lambda_i \gamma(x_i, x_j) + \psi(x_0) = \gamma(x_j, x_0) \qquad \text{for all j} \qquad 5.18$$

$$\sum_{i=1}^{N} \lambda_i = 1$$

The weights generated by 5.18 can be substituted into 5.15. The estimation variance (prediction) can be obtained by the following equation:

$$\hat{Z}\ (x_0) = \sum_{i=1}^{N} \lambda_i z(x_i) \qquad 5.19$$

$$\sigma^2\ (x_0) = \sum_{i=1}^{N} \lambda_i z(x_1 x_0) + \psi(x_0) \qquad 5.20$$

OK is an exact interpolated as represented in an example with a target point (x_3). If (x_3) is a data point taken to be used to interpolate the target point (e.g. x_j) then $\sigma^2\ (x_0)$ is minimized when $\lambda(xj) = 1$ and all the other weights are 0.

OK represented in matrix form is as follows:

$$A\lambda = b \qquad 5.21$$

Where:

$$A = \begin{bmatrix} \gamma(x_1, x_1) & \gamma(x_1, x_2) & \Lambda & \gamma(x_1, x_N)1 \\ \gamma(x_2, x_1) & \gamma(x_2, x_2) & \Lambda & \gamma(x_2, x_N)1 \\ \mathrm{M} & \mathrm{M} & \Lambda & \mathrm{M} \\ \gamma(x_N, x_1) & \gamma(x_N, x_2) & \Lambda & \gamma(x_N, x_N)1 \\ 1 & 1 & \Lambda & 1 \end{bmatrix}$$

And $\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_1 \\ \mathrm{M} \\ \lambda_N \\ \psi(x_0) \end{bmatrix}$ and

$$\mathrm{b} = \begin{bmatrix} \gamma(x_1, x_0) \\ \gamma(x_2, x_0) \\ \mathrm{M} \\ \gamma(x_N, x_0) \\ 1 \end{bmatrix}$$

Inverting matrix A produces the Lagrange Multiplier

$$\lambda = \mathrm{A}^{-1}\mathrm{b} \qquad 5.22$$

There are only a certain number of semivariogram models available within ArcGIS to apply to kriging in order for the standard error to return a non-negative value. For the purposes of explaining OK we have used the exponential model. ArcGIS enables three different models to be added at the same time.

$$\gamma(h;0) = \theta_s\left[1 - \exp\left(-\frac{3\|h\|}{\theta_r}\right)\right] \qquad 5.23$$

Where $\theta_s \geq 0$ is the partial sill and
$\theta_r \geq 0$ is the range parameter

Figure 5.2 and 5.3 (3-dimensional) represents the changes in ExAl in the layer of soil from the surface to 50mm below the surface. This is an example of OK using the exponential semivariogram model as depicted in Figure 5.4.

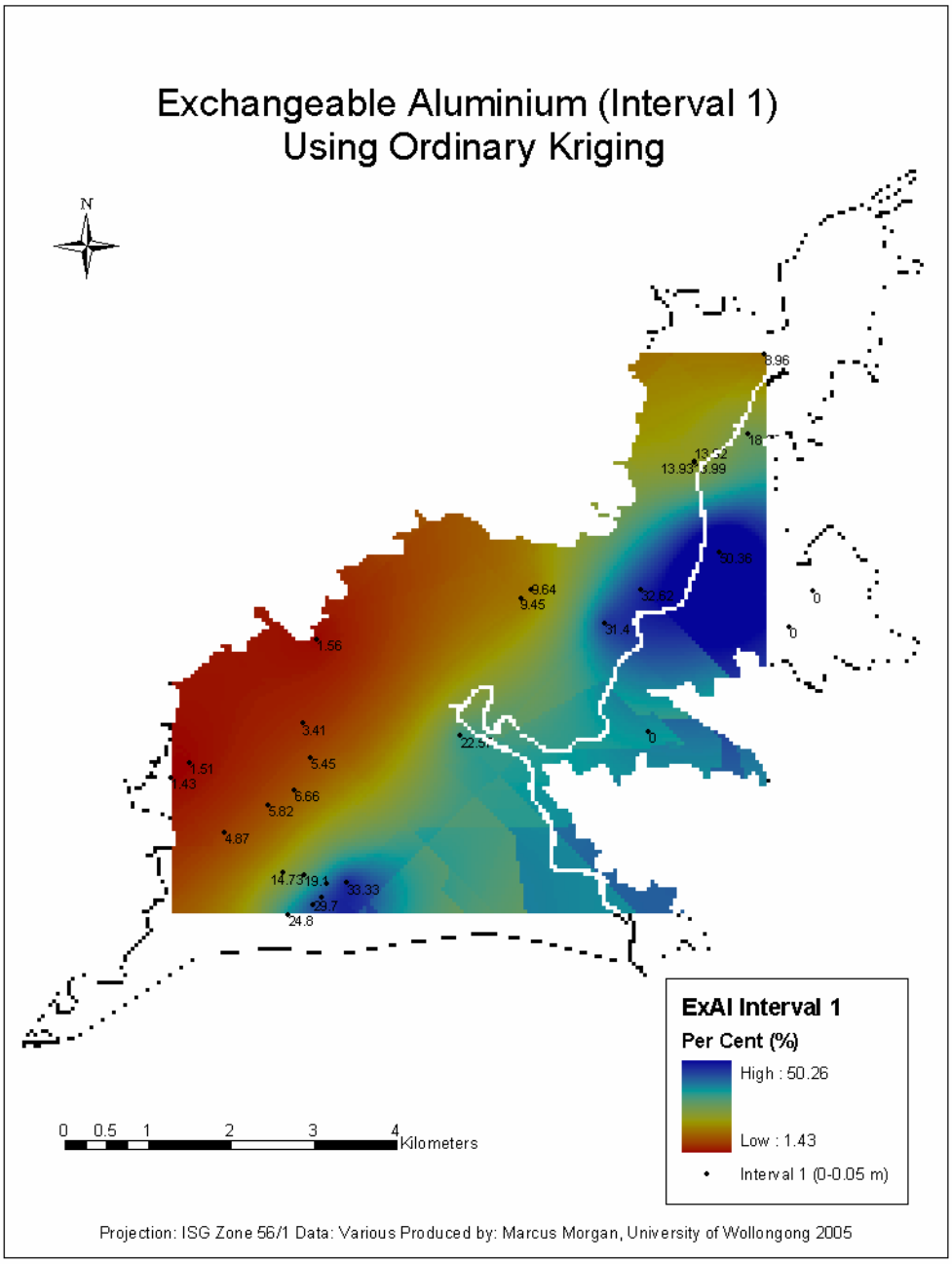

Figure 5.2 OK of ExAl for Layer 1 (0.00 – 0.05m AHD), Data Source: Wollongong City Council, 2002.

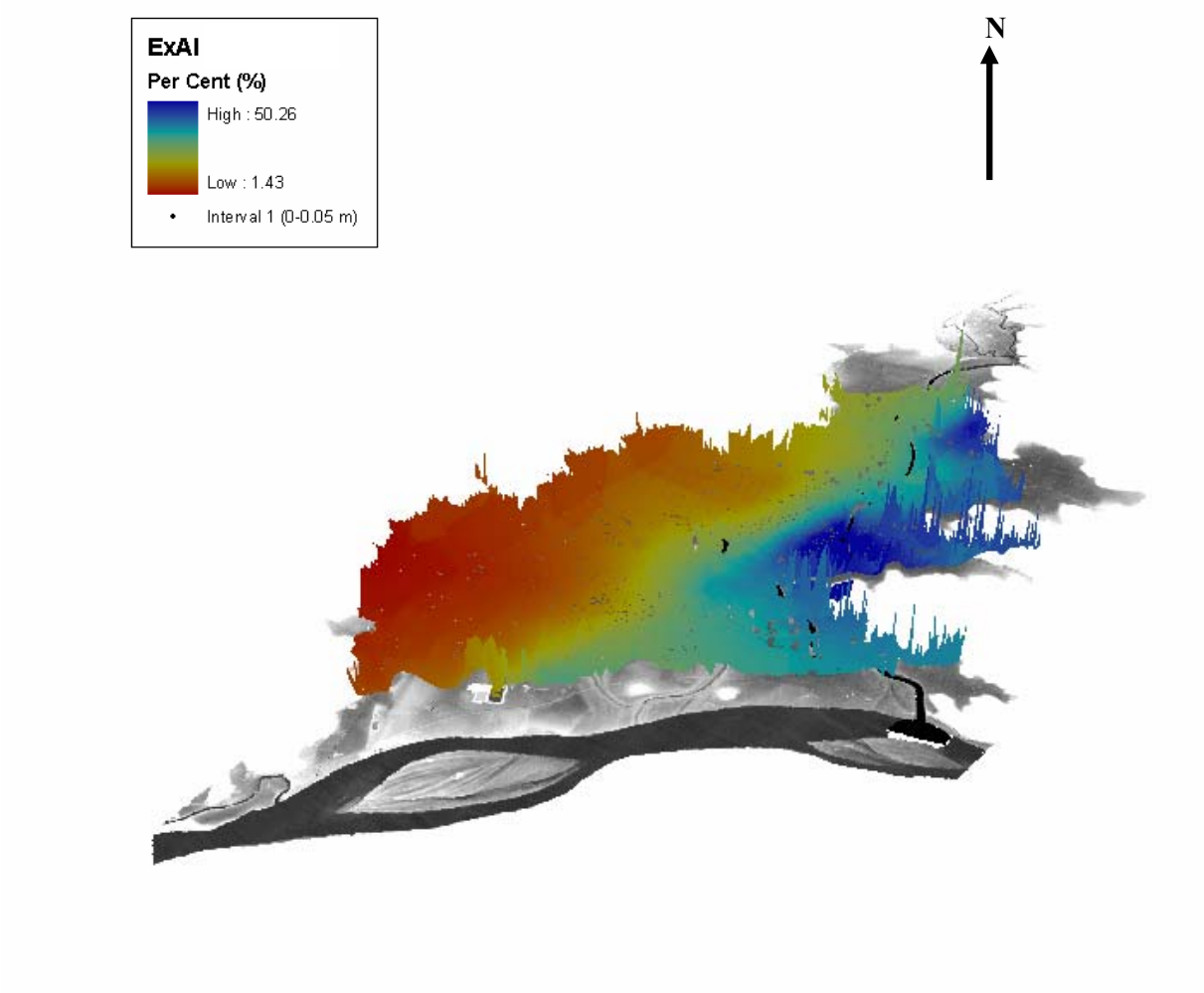

Figure 5.3 ExAl - Interval 2 on Catchment DTM (x100 exaggeration), Data Source: Wollongong City Council, 2002.

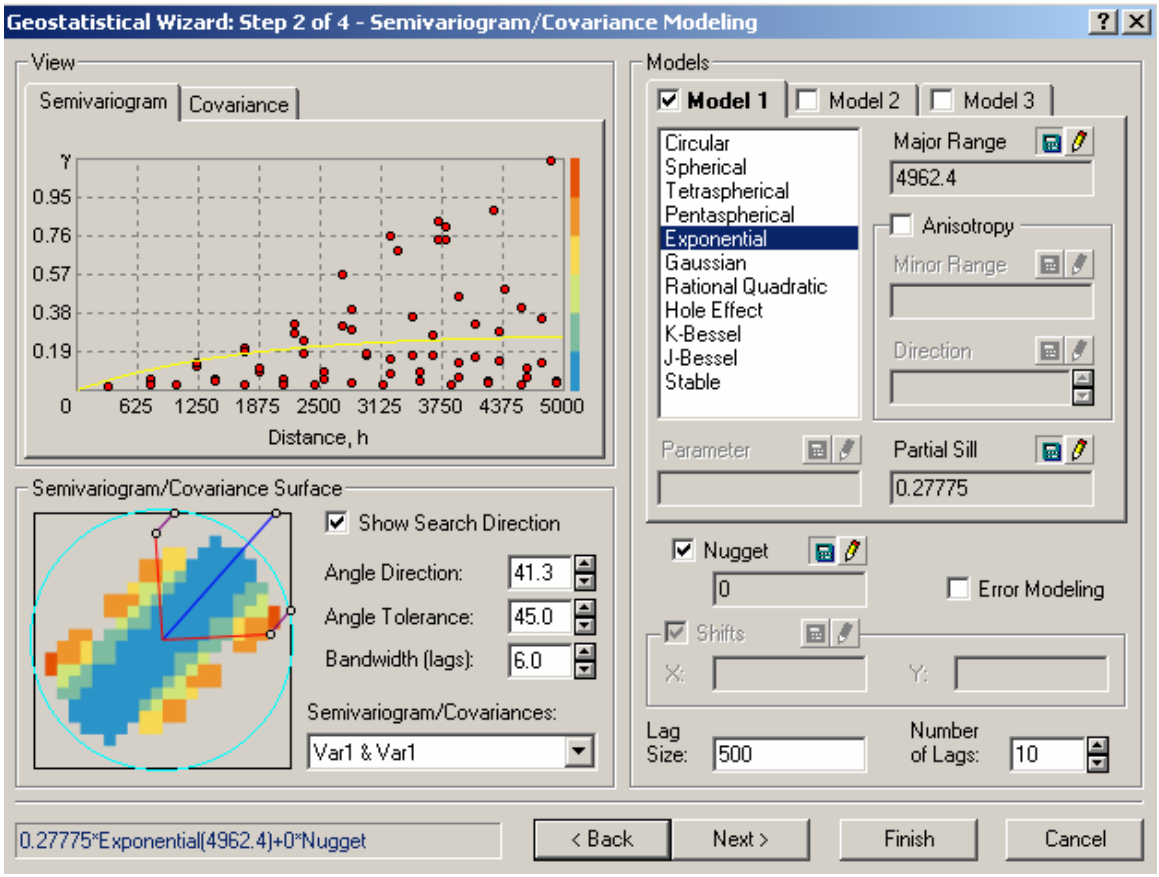

Figure 5.4 Semivariogram of ExAl for Layer 1.

Model Validity

The predicted values from the OK method can be compared to the measured values for verification of the model. Figure 5.5 represents that the model developed is close to being an exact interpolator as the predicted and measured values are almost equal (passing through the axis at y = x). Table 5.4 assesses the error generated for the models produced by the geostatistical analyst. Overall the error was low for the model developed for Interval 1, with the standardized mean error close to 0 (unbiased), and Average Standard Error slightly greater than the Root-Mean Squared (RMS) Error. This is a slight over estimation of the variability in prediction, as supported by the standardised RMS of 0.397 (Table 5.5).

Table 5.4 Error readings for a good geostatistical model

Root-mean-squared (RMS) prediction errors	Small
Average standard error ^	~ Same as RMS prediction errors,
Standardised* Mean Prediction Errors	~0
Standardised* RMS prediction[#]	~1

* Standardised Error = $\frac{\text{Predicted - Measured Values}}{\text{Estimated Kriging Standard Errors}}$

Where:

^Average Standard Error ~ Root-Mean-Squared Prediction Error[#] = Correct assessment of Variability in prediction

^Average Standard Error > Root-Mean-Squared Prediction Error[#] = Overestimation of the Variability in prediction

^Average Standard Error < Root-Mean-Squared Prediction Error[#] = Underestimation of the Variability in prediction

RMS Standardized Error[#] is ~ 1 Correct assessment of Variability in prediction

< 1 = Overestimation of the Variability in prediction

>1 = Underestimation of the Variability in prediction

Mean	-0.0086
Root-Mean-Square	0.161
Average Standard Error	0.289
Mean Standardized	-0.018
Root-Mean-Square Standardized	0.397

Table 5.5 Prediction Error Values for Interval 1 (0.00-0.05m) ExAl

5.5 Cross-validation and validation

Within ArcGIS, the geostatistical analyst enables the predicted models to perform a cross validation before the final predictive layer is produced. Figure 5.5 shows the resulting predictive points compared to measured values for ExAl across Interval 1. Each of these points produces an error unless they are exact interpolators.

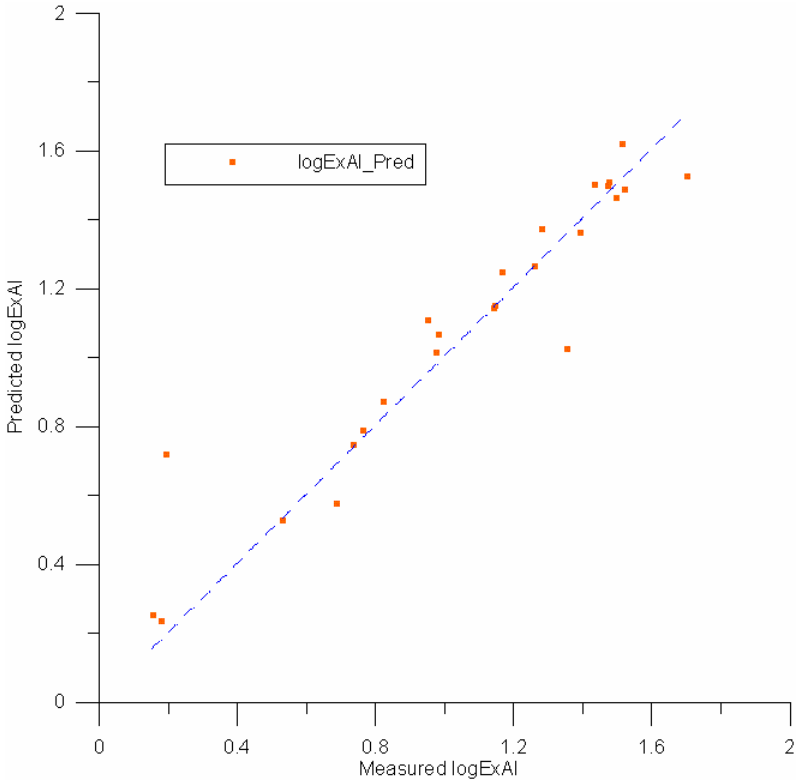

Figure 5.5 Cross Validation: Predicted v Measured logExAl for Layer 1 (0.00 – 0.05m AHD).

The problem with Kriging is that it often over estimates the lower predicted values and underestimates the higher predicted values. This is seen with two points in Figure 5.5.

Another method of testing the data is through creating sub-sets and using one as the learning set and the other as the validation set. By using a validation set the decisions made in choosing the semivariogram, lag size, and neighbourhood search size will be supported or rejected. If the validation set is supported then the model will be suitable for the remainder of the data.

Validation

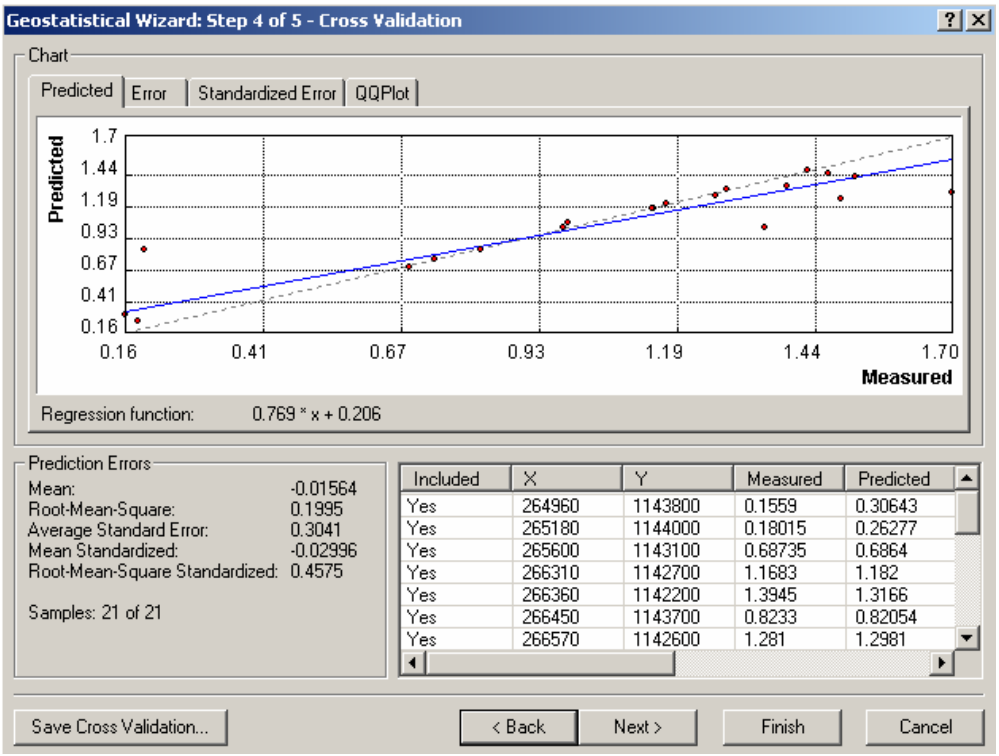

Figure 5.6 Cross Validation of the Training Set – Interval 1 (ExAl).

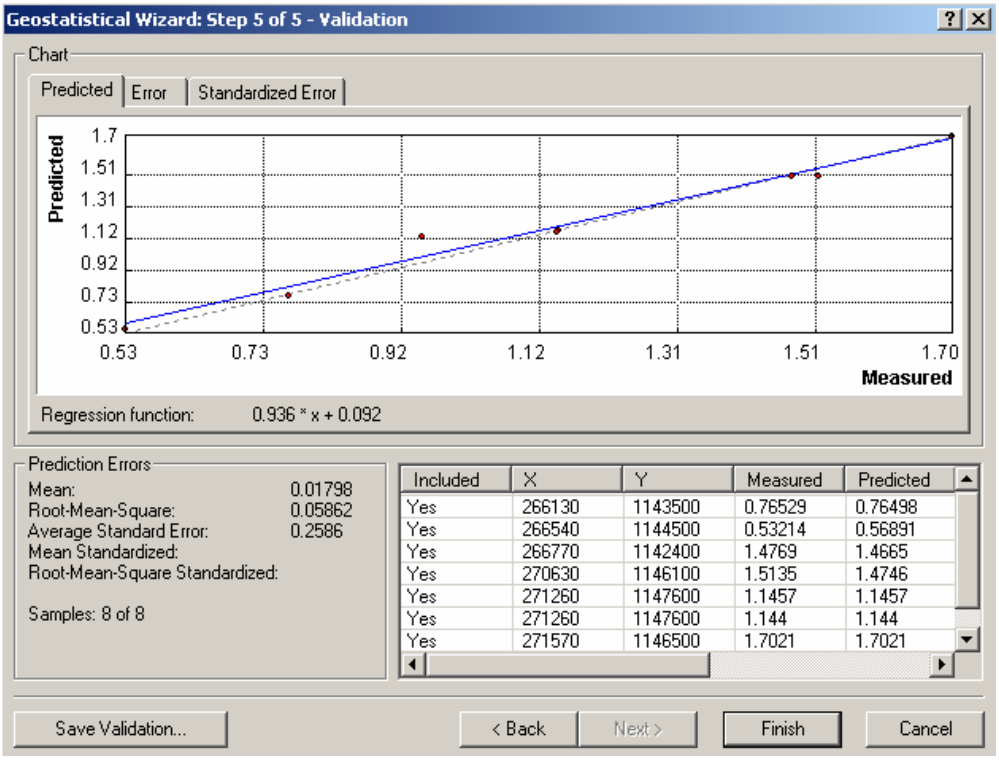

Figure 5.7 Validating the Validation Set – Interval 1 (ExAl).

From the validation data set, there is an overestimation of variability in prediction as the Average Standard Error (ASE) is greater then the RMS Error (Figure 5.6). This is more substantial in the validation data set as ASE is 0.259 compared to RMS of 0.059 (Figure 5.7).

The following Chapter (6) looks at these spatial statistical methods and develops models that predict the severity of soil across the Shoalhaven Floodplain. Three of the methods discussed in detail above to be used in the analysis include:

1. Ordinary Kriging
2. Universal Kriging

And

3. Inverse Distance Weighting

Chapter 6: Coastal Acid Sulfate Soil Severity Models - Prediction of Soil Variables

6.1 Model 1: Ordinary Kriging in Parameter Estimation

OK methods were applied to estimate all four parameters (ExAl, TAA, EC, and Cl:SO_4) across the study area. The aim of using OK is to generate the optimal interpolation from known points. Burgess and Webster (1980) used OK methods in their optimization of soil properties across a spatial setting. In more recent times OK has been used in precision agriculture in combination with remote sensing (Mulla, 1997), in determining the relationship between soil and crop attributes, and in determining the sampling intensity required across a vast area (Frogbrook, 1999).

6.1.1 Exploratory Spatial Data Analysis

In the primary data analysis (Chapter 4) all variables were assessed for the properties that validate a normally distributed data set, as most of the methods used require the data set to be normally distributed (OK, Multiple Regression, and Geographically Weighted Regression). This included analysis for linearity, bias, autocorrelation, the removal of outliers, and transformation of variables if needed. However, even though x and y variables were included for spatial reference, exploratory spatial data analysis can analyse the structure or determine whether any trends exist in the data, which cannot be analysed on an individual parameter basis. This will identify any global trends in the data.

The results of trend analysis are usually explained in terms of the relative strength of the polynomial function that can be fit through the data. Various relationships in the data can be seen to exist such as linear (gradual) and quadratic. This was performed on the data for the four main variables.

After assessing the data and removing possible trends, Kriging can then be used to interpolate between points to generate a predictive surface map or standard error probability map. There are a number of inputs that are required in OK that influence the model outcome. A semivariogram and covariance cloud is used to define the model.

6.1.2 Semi Variance Generation

Figure 6.1 represents the Semivariogram platform that is presented by ArcGIS, giving a number of options to generate the desired output. For ExAl, using the exponential model as defined in equation 5.23, a lag size of 500, and a search direction from North East to South West, the output was generated over Intervals 1-13 (Figure 6.3). After Interval 12 the size of the interpolation was reduced significantly and the number of samples taken at those depths was too few to interpolate (minimum of 10 sample points required). The general form of the semivariance is:

$$Y(s_i, s_j) = \tfrac{1}{2}\,\text{variance}\,(Z(s_i) - Z(s_j)) \qquad (6.1)$$

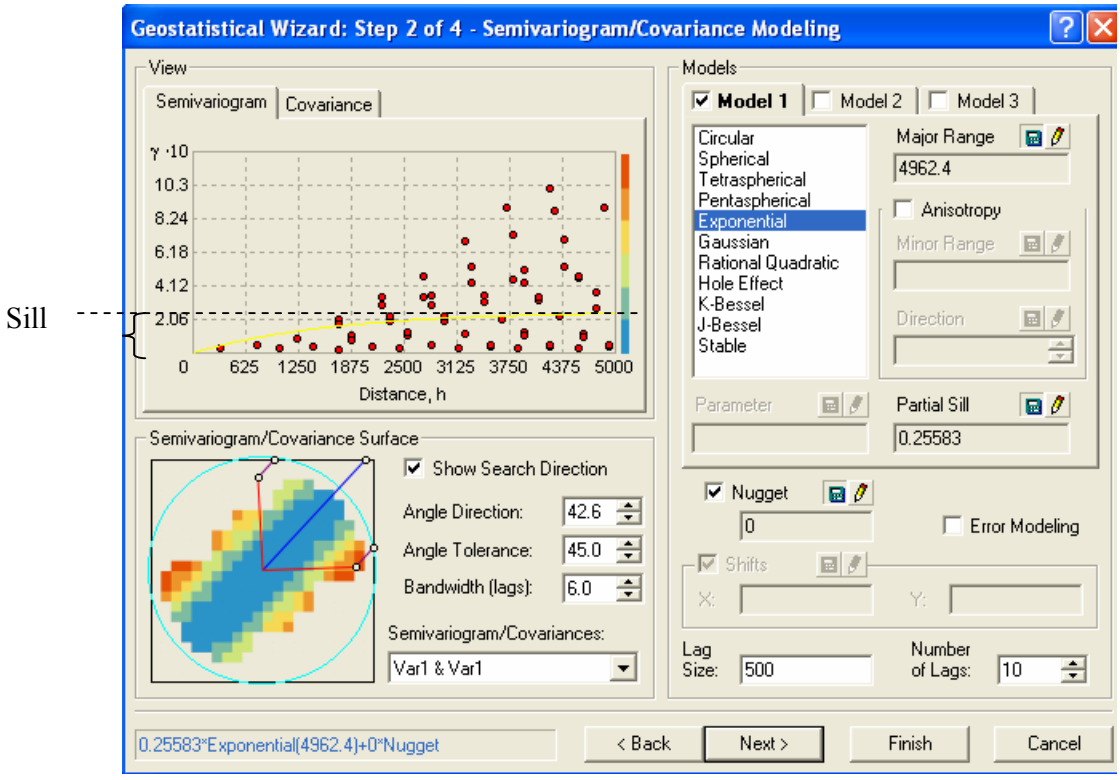

Figure 6.1 Semi Variance generation of Interval 5, ExAl Prediction Using OK.

If s_i and s_j are close together, $Z(s_i)$ and $Z(s_j)$ will be small. Figure 6.1 represents this in the semivariogram with values closer to 5000, inferring that s_i and s_j are far apart and

the corresponding variance is greater at these points. This explains the resultant exponential curve. The reverse is true as the distance (h) is smaller.

6.1.3 Covariance

The covariance also depicts the strength of correlation as a function of distance. The general form of the covariance function is:

$$C(s_i, s_j) = \text{covariance}(Z(s_i), Z(s_j)) \qquad (6.2)$$

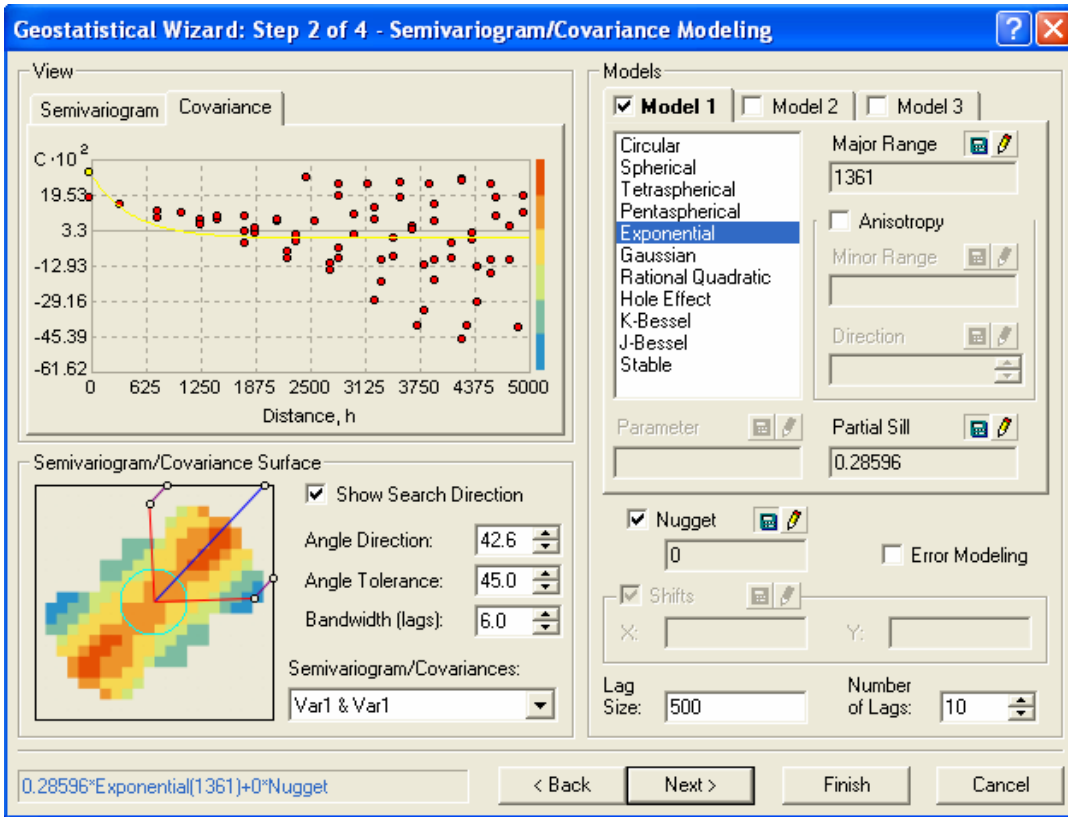

Figure 6.2 Covariance generation of Interval 5, ExAl Prediction Using OK.

In the covariance cloud, when two points (s_i and s_j) are closer together the correlation will be larger and their covariance will be larger. Figure 6.2 represents the covariance cloud of ExAl at Interval 5 (-0.30 to -0.40m) showing that the values closer to the y-axis

are those that have the smallest h (distance) between one another, and the larger covariance.

6.1.4 Model Generation

Each layer or interval of each of the variables was used to develop what collectively becomes a three-dimensional representation for that variable over the entire floodplain.

6.1.4.1 Exchangeable Aluminium (3-D Model)

Each layer was put through a process that optimised the model for ExAl. ExAl was represented in percent of Cation Exchange Capacity (ExAl/CEC). Figure 6.3 represents the stretched raster image of the multiple layers, with ExAl/CEC values ranging from 0 – 100 percent.

Figure 6.4 represents the three categories that ExAl/CEC was divided into. Table 2.4 describes the toxic levels of ExAl in a soil profile. These levels are represented in Figure 6.4 by three different colours. The results are clearer than in Figure 6.3, showing that closer to Broughton Creek ExAl/CEC levels are somewhat higher than in the backswamp, and on higher land to the edge of the floodplain.

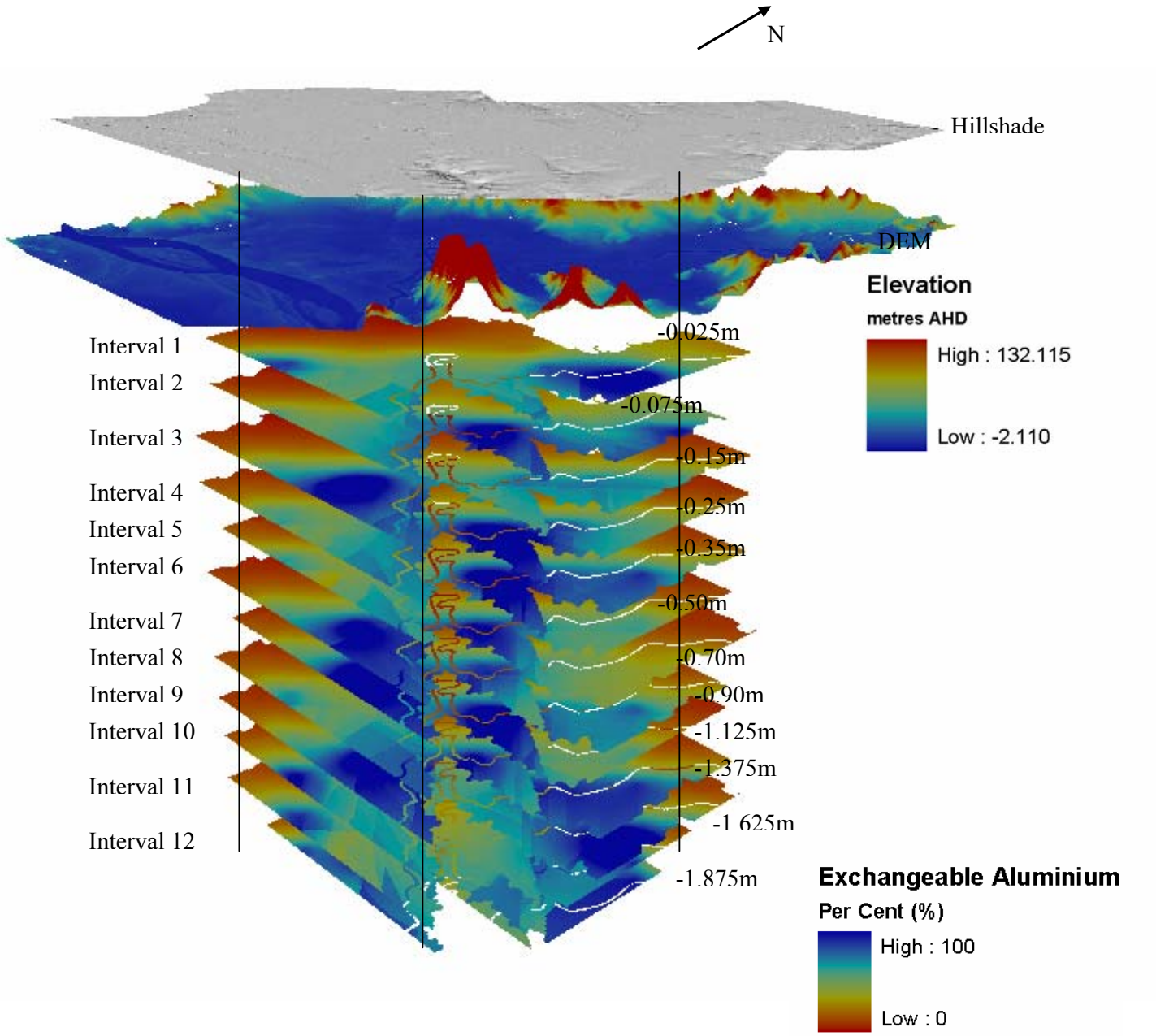

Figure 6.3 Exchangeable Aluminium Soil Profile – Stretched Raster.

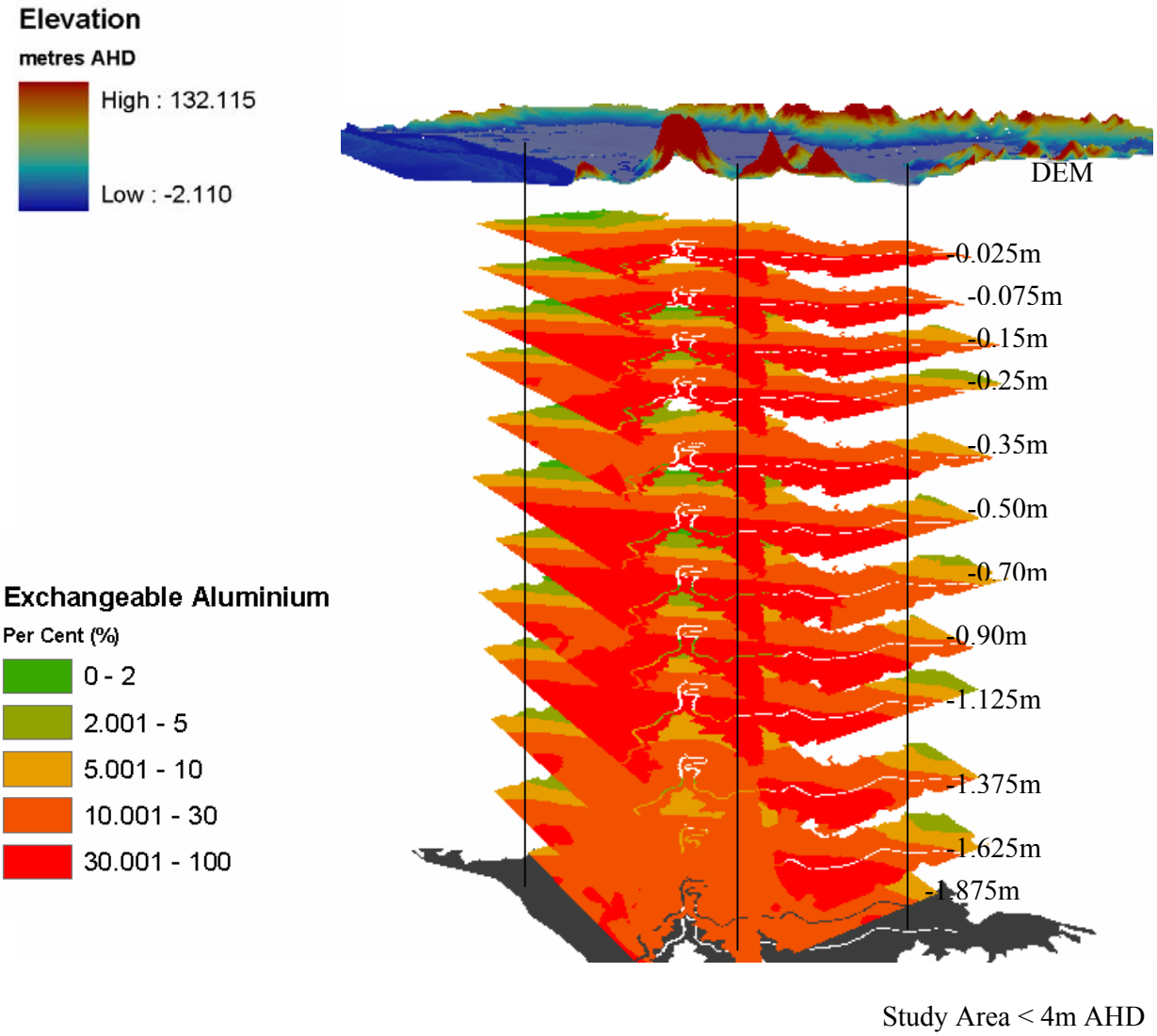

Figure 6.4 Exchangeable Aluminium Soil Profile – Classified Raster.

For the model created all intervals were cross-validated to existing points to determine the strength of the prediction. Figure 6.5 represents a very strong relationship between measured and predicted Exchangeable Aluminium for interval 5, which validates this layer.

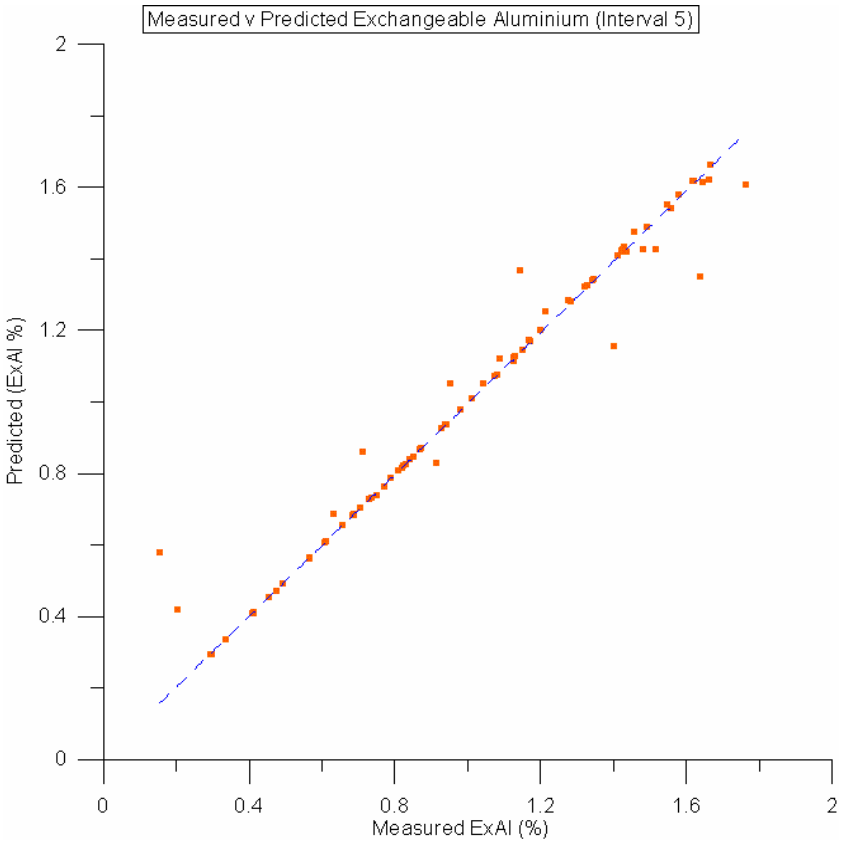

Figure 6.5 Measured v Predicted Exchangeable Aluminium (Interval 5).

6.1.4.2 Total Actual Acidity (3-D Model)

Each soil layer down the horizon was created from interpolating using OK, between points of known TAA throughout the floodplain. Figure 6.7 shows the stretched raster for each layer as generated in ArcMap. Although the scale is between 0 and 441 mol H+/tonne for all layers, the maximum and minimum values differ in each layer, so this

scale can not be taken to represent all layers, but rather it shows the possible values and approximate representation of TAA. Figure 6.8 corrects this error by using a standard classified scale.

Compared to ExAl, TAA is less severe in ratings shown by the classified raster in Figure 6.8. The severity of TAA is increased between 0.5 and 1.125m below ground surface level, whereas ExAl is consistently high between 0.0 and 1.125m below ground surface level. In both of the profiles the distribution of each chemical component was highest surrounding Broughton Creek, and in the backswamp areas in the north of the floodplain. This corresponds with previous research and management plans (Blunden 2000; Glamore 2003; Broughton Creek Management Plan 2004) for the floodplain.

The semi-variogram used a lag size of 500 with 10 lags to generate the optimal layers when using OK for TAA prediction. Figure 6.6 represents the cross validation of the TAA model for layer 5 at the 0.30m level below the surface, showing a good predictive model.

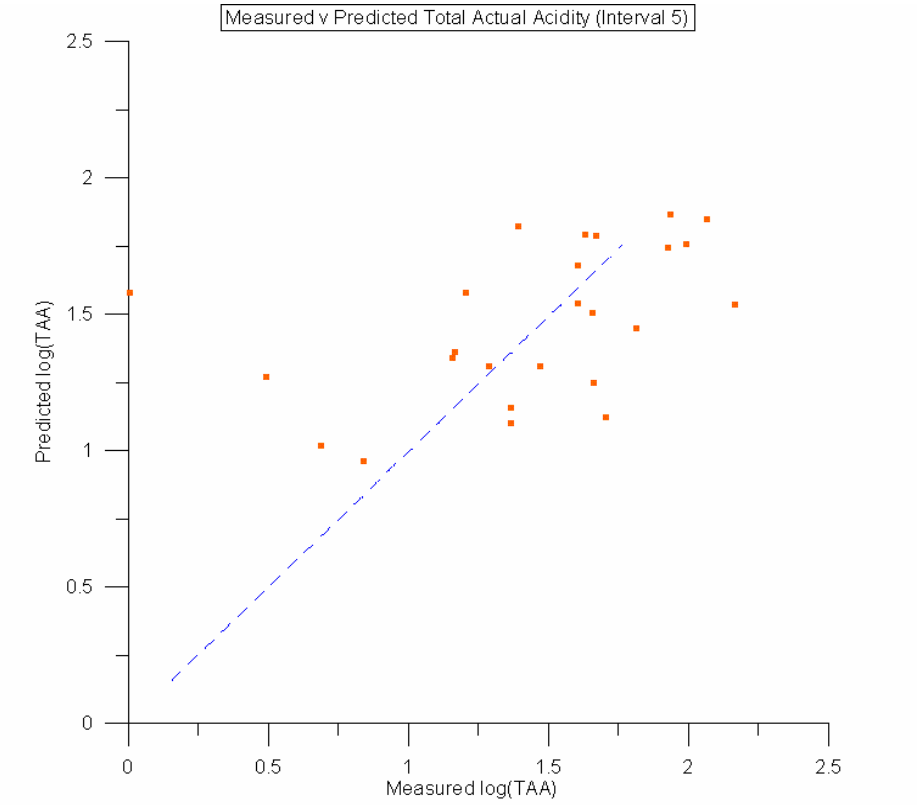

Figure 6.6 Measured v Predicted Total Actual Acidity (Interval 5).

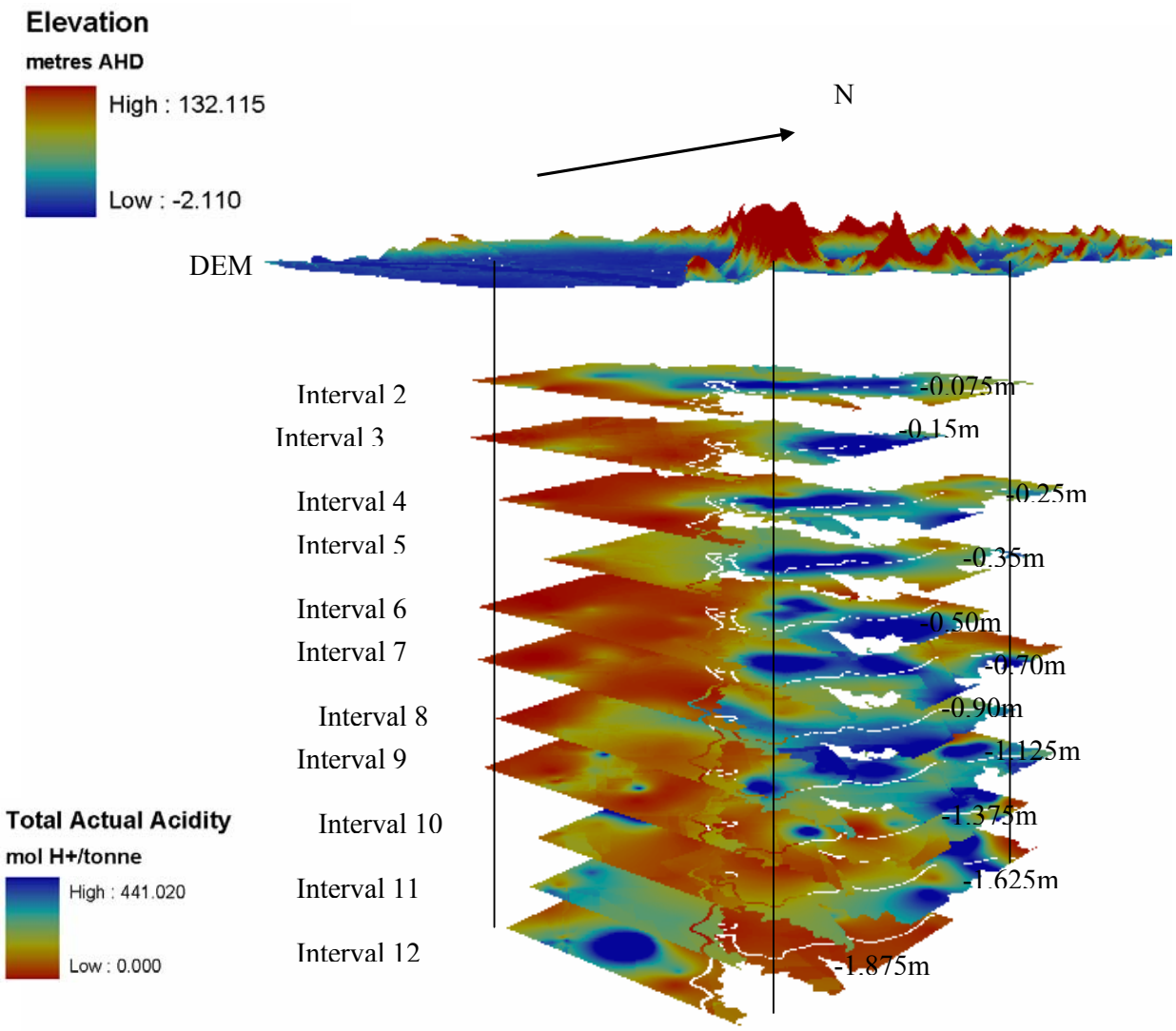

Figure 6.7 Total Actual Acidity Soil Profile – Stretched Raster.

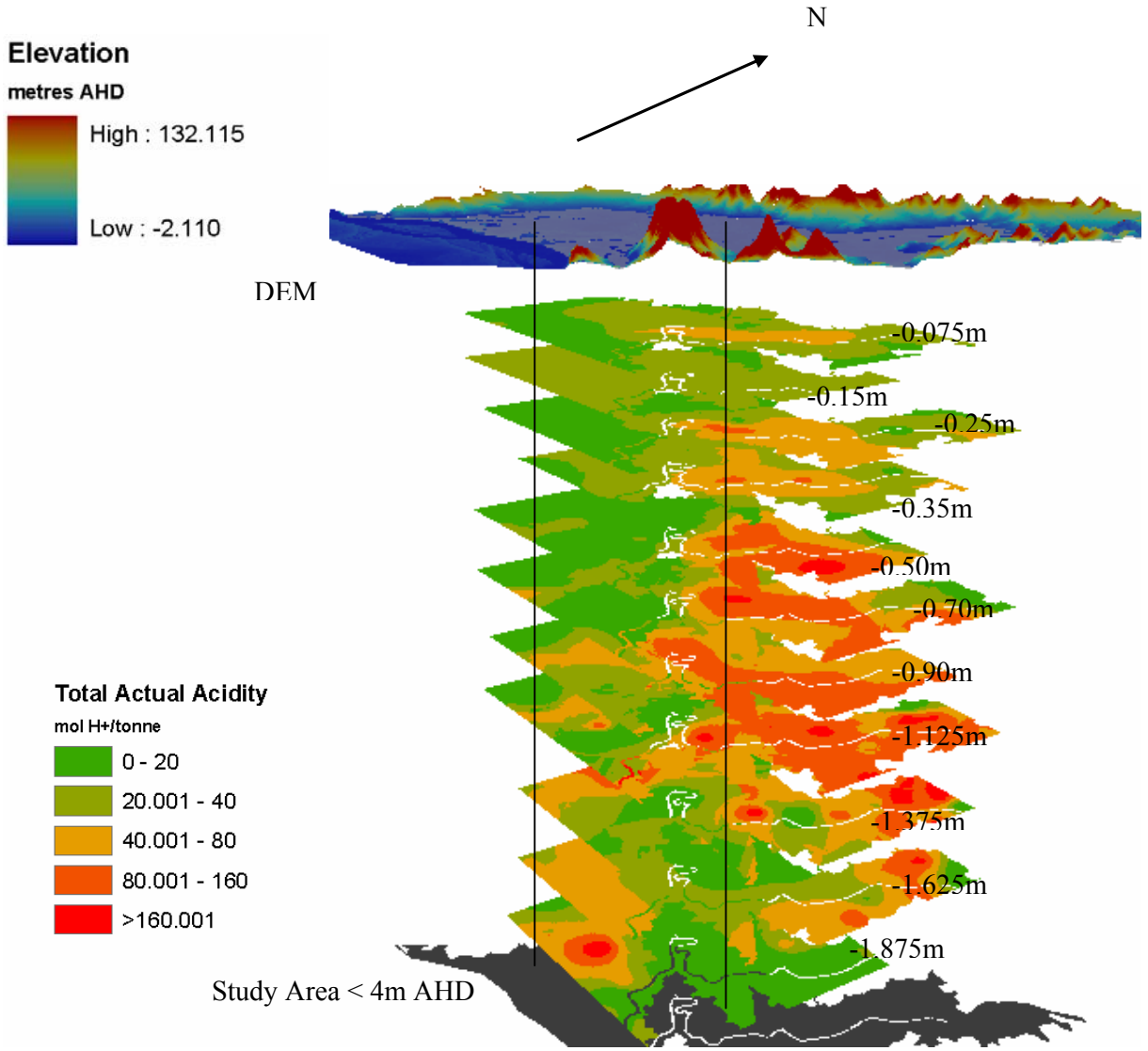

Figure 6.8 Total Actual Acidity Soil Profile – Classified Raster.

6.1.4.3 Electrical Conductivity (3-D Model)

The analysis of Electrical Conductivity (EC) is presented similarly to ExAl and TAA, by a stretched raster (Figure 6.9) and a classified raster (Figure 6.10), showing the variability in EC over the floodplain. The scale for the stretched raster is indicative of the range of values throughout all layers but cannot be used to infer which layer is the most severe. The classified raster (Figure 6.10) is more accurate and has a standard scale for all layers.

Studies by Glamore (2003) and Ford (2002) have found EC to be higher within layers close to the drains throughout the floodplain. However the levels were not significant to cause major agricultural damage. The distribution of EC throughout the profile is seen in Figure 6.10, which shows EC levels less than 4 dS/m for the majority of the layers. Close to the surface (-0.075m) the EC levels are high in the south west corner of the floodplain. This may have been a result of a flood in combination with a king tide, depositing salt on the surface.

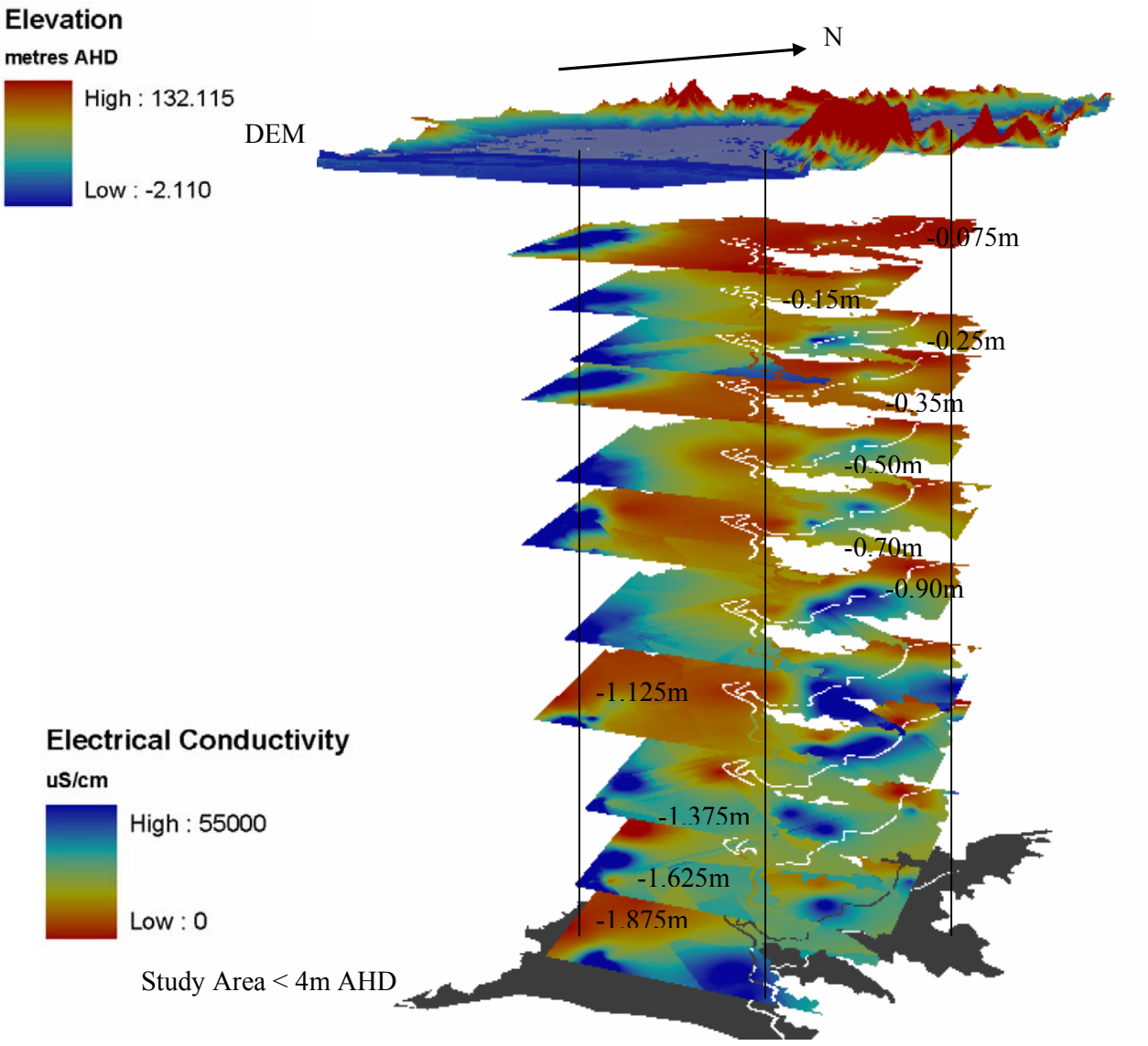

Figure 6.9 Electrical Conductivity Soil Profile – Stretched Raster.

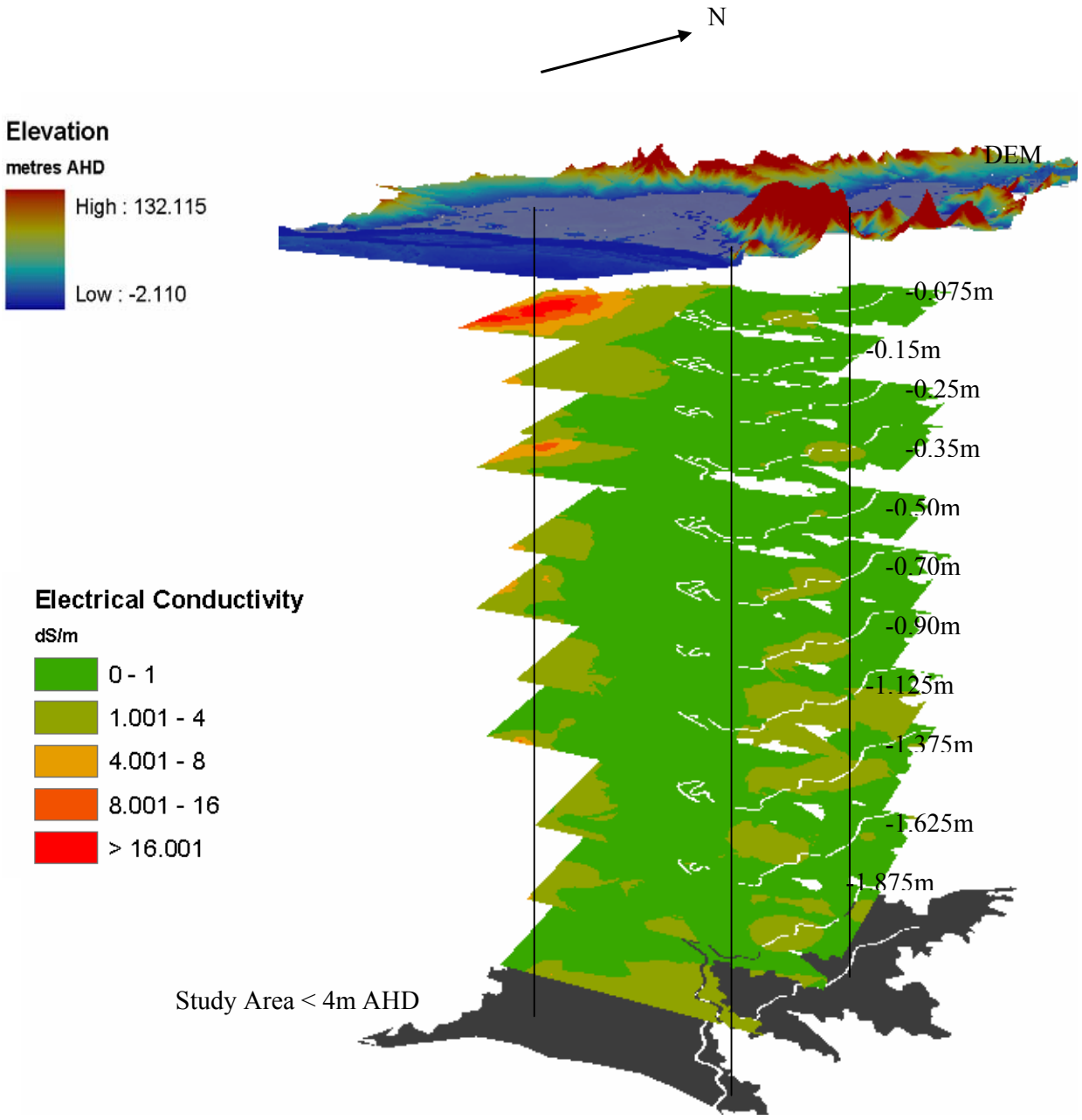

Figure 6.10 Electrical Conductivity Soil Profile – Classified Raster.

6.1.4.4 Error Generation

An optimal model should obtain the smallest error or the error that meets the best, linear, unbiased principles. In the models generated using OK, the optimal lags for the semivariogram model generated a number of errors. The Root Mean Squared Standardised (RMSS) error is optimised when it equates to one (see Chapter 5.4). Figure 6.11 depicts the RMSS error for all variables. There is a general trend between error and number of samples. Soil layers with a small number of points across the floodplain display the greater error. This is represented by layer 2 and layer 11 for TAA, which had 30 and 40 points respectively in the interpolation process, as compared to layer 8 which had 81 points. The resulting error shows RMSS close to 1 in layer 8 and further from 1 in layers 2 and 11.

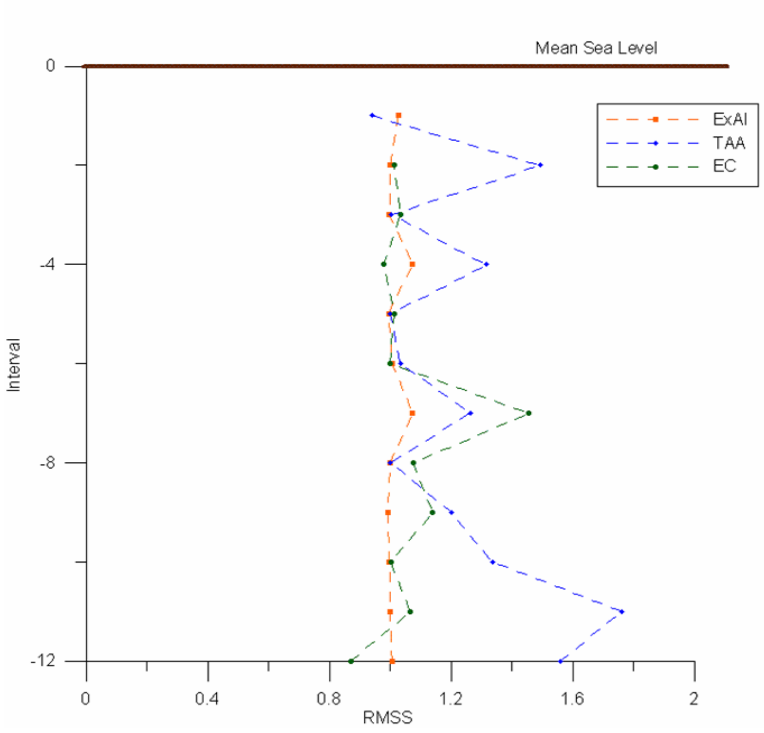

Figure 6.11 OK Optimal Lags – RMSS Error.

Root Mean Squared (RMS) and Average Squared Error should be equivalent to one another (see Chapter 5.4) when optimising the error in the model output produced by ArcGIS Geostatistical Analyst. These three errors will show the strength or lack of strength the model has in predicting between known values. Figure 6.12 compares RMS and ASE for all the variables. For EC and ExAl the errors are quite close. For TAA the variation is significant and the general fit is positively linear which is theoretically valid for TAA.

6.1.4.5 Weights

The elimination of $Cl:SO_4$ as a variable changed the composition of the severity model and the emphasis that each variable had on the final output of the model. The process of determining the significance of the other variables was bought about by a number of factors. These include a previous expert system by Dent & Dawson (1996), consultation with soil experts (Lawrie, Eldridge, Haddad, and Indraratna), availability of soil samples, and previous research that have viewed these variables in being important in understanding CASS baseline conditions (Blunden 2000; Glamore, 2003). As a result table 6.1 shows the weights generated and applied to each raster layer in developing the severity map (Figure 6.13).

Soil Parameter	Weight
ExAl	0.4
TAA	0.4
EC	0.2

Table 6.1 Weights used in the project to generate Severity Maps

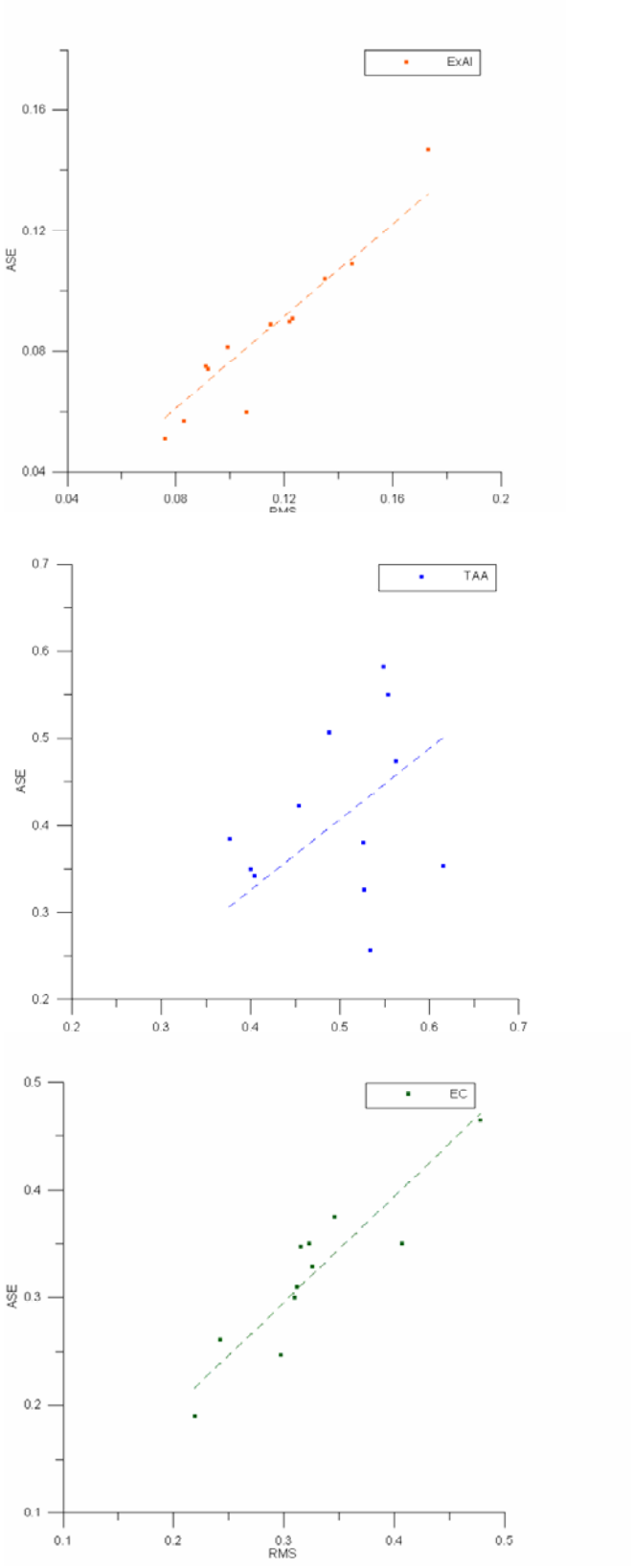

Figure 6.12 OK Optimal Lags – RMS v AS Error.

6.1.4.6 Ordinary Kriging Severity Map

The process in generating the CASS severity maps in ArcMap through Spatial Analyst extension is represented in Figure 6.13

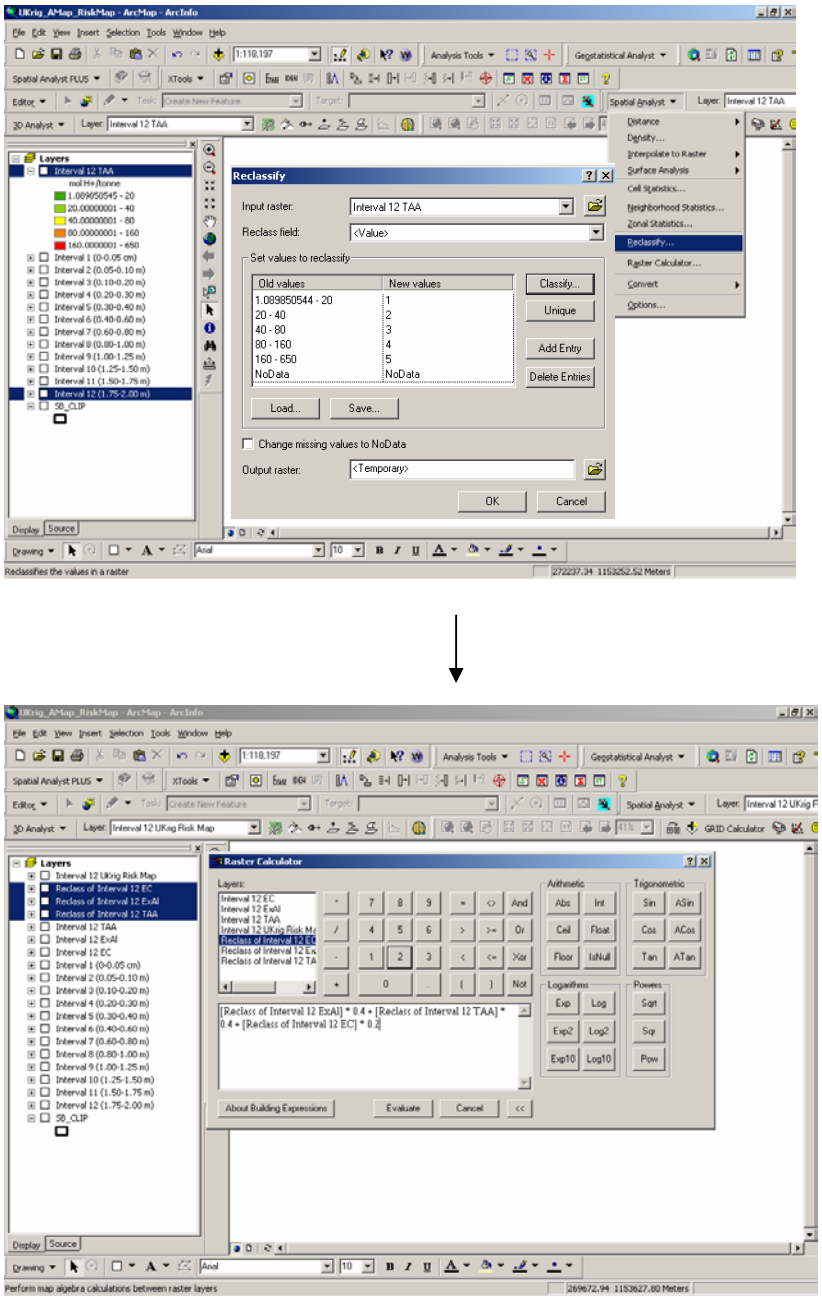

Figure 6.13 Process of Generating Severity Maps in ArcGIS.

ArcGIS Spatial Analyst extension was employed to develop the severity maps, layer by layer until a comprehensive 3-D profile was completed (Figure 6.14). The severity maps produced using OK for Broughton Creek floodplain show that from layer 5 (-0.35m) to layer 10 (-1.375m), there is an increased severity in the north-west section of the floodplain surrounding Broughton Creek. The areas of higher severity correlates with the backswamp, estuarine, and alluvial low-lying areas that were identified by the DLWC risk maps. Layer 12 produced an area that was classified as a high severity area. In this area the sulfidic layer was identified by Lawrie & Eldridge (2002) to be much further down the profile than sites in the north western and north eastern part of the floodplain. Two other models were developed and compared to the OK model.

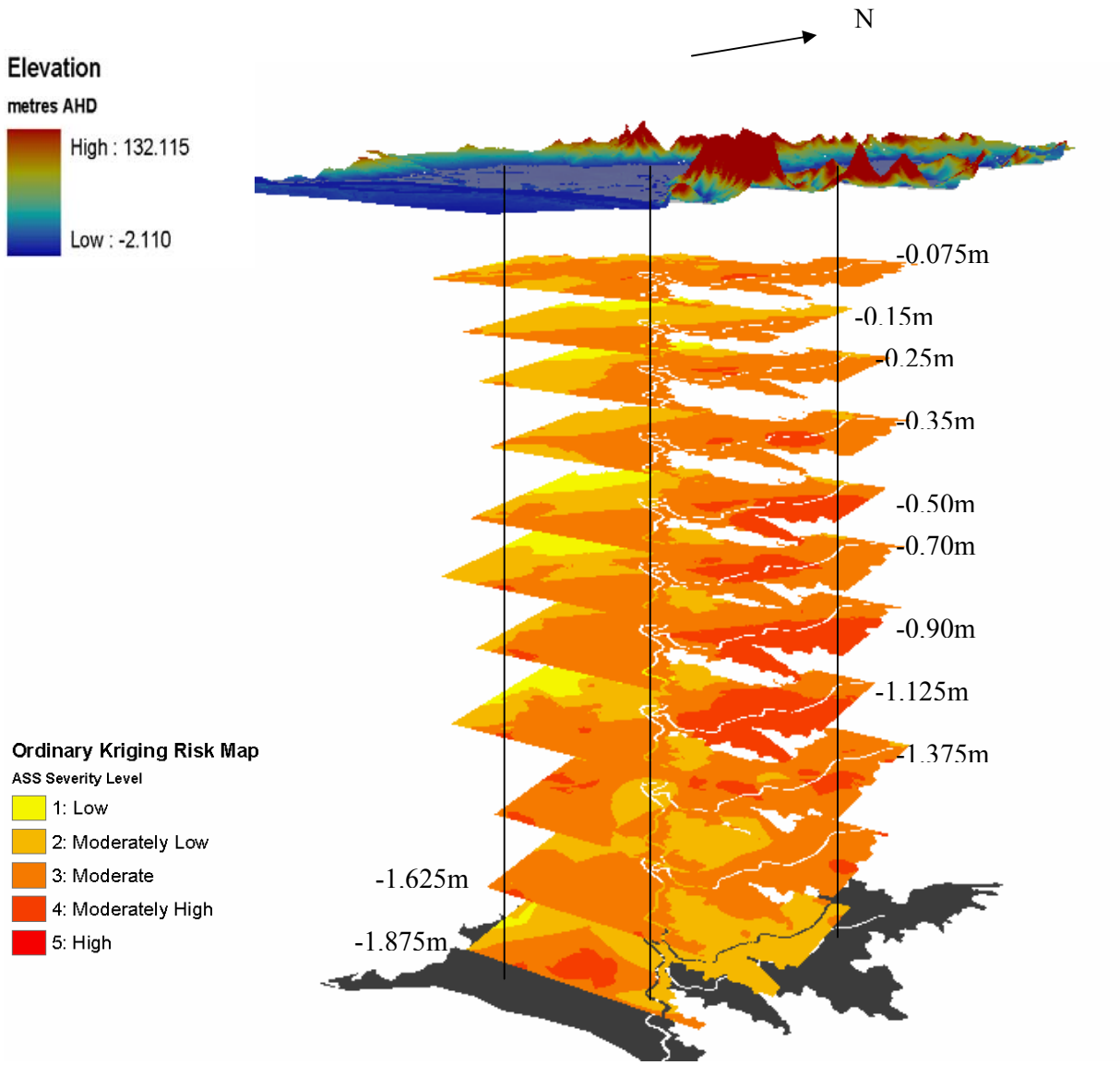

Figure 6.14 OK Severity Map.

6.2 Model 2: Inverse Distance Weighting in Parameter Estimation (Local)

The Inverse Distance Weighting (IDW) severity model is produced from combining Exchangeable Aluminium (ExAl), Total Actual Acidity (TAA) and Electrical Conductivity (EC), based on the weights defined in Table 6.1. The method of producing the IDW severity maps is similar to how the OK severity maps were produced (Figure 6.14).

6.2.1 Exchangeable Aluminium (3-D Model)

As in the case of OK, each individual ExAl layer was developed by a statistical process (IDW) that optimised the model. The layers were then combined to create a 3-dimensional soil profile.

The output of results support what was seen in Figure 6.4 (see Appendices Figure A5), that around the banks of Broughton Creek ExAl levels are slightly higher than further into the backswamp, or on higher land to the edge of the floodplain. The layers showing greater toxicity were from slightly below the surface (-0.075m) down to –1.125m, with random hotspots showing up in layers further down the profile as seen in layer 12 (-1.875m). This differed slightly from OK, which showed greater distribution of ExAl toxicity from the surface, down the soil horizon to –1.375m (layer 10).

6.2.2 Total Actual Acidity (3-D Model)

The TAA model generated using IDW (Figure A6, Appendices A6) shows that the principle of bimodal distribution of TAA down a soil profile holds in some areas. This is supported by layers 6 (-0.50m) and 8 (-0.90m) which show the highest readings of TAA. This was consistent with OK.

6.2.3 Electrical Conductivity (3-D Model)

As in OK, EC did not appear to generate the similar extent of severity as ExAl and TAA. However, the IDW model was consistent with OK in showing an area to the west of the floodplain that exhibited high EC levels near the surface (Figure A7, Appendices A6).

6.2.4 Error Generation

IDW generated a number of errors with the construct of the model. Root Mean Squared Error (RMSE) and Mean Error was estimated in the IDW model, and compared to RMSS, RMSE and ASE. The model error was low for IDW with the error increasing in layers of fewer points throughout the floodplain. TAA had the highest RMSE, followed by EC and ExAl was the most optimal model (Figure A8, Appendices A6).

6.2.5 IDW Severity Map

The same process described by Figure 6.13 was used to generate the severity map for the IDW model (Figure 6.15). The risk of uncovering a CASS environment with characteristics severe enough to affect the downstream environment was greatest in the soil profile between -0.50m to –1.125m, and in the northern section of the floodplain surrounding Broughton Creek. This is consistent with ExAl and TAA models developed and is also consistent with the severity map produced using OK. However, there are slight differences which will be discussed later in this chapter.

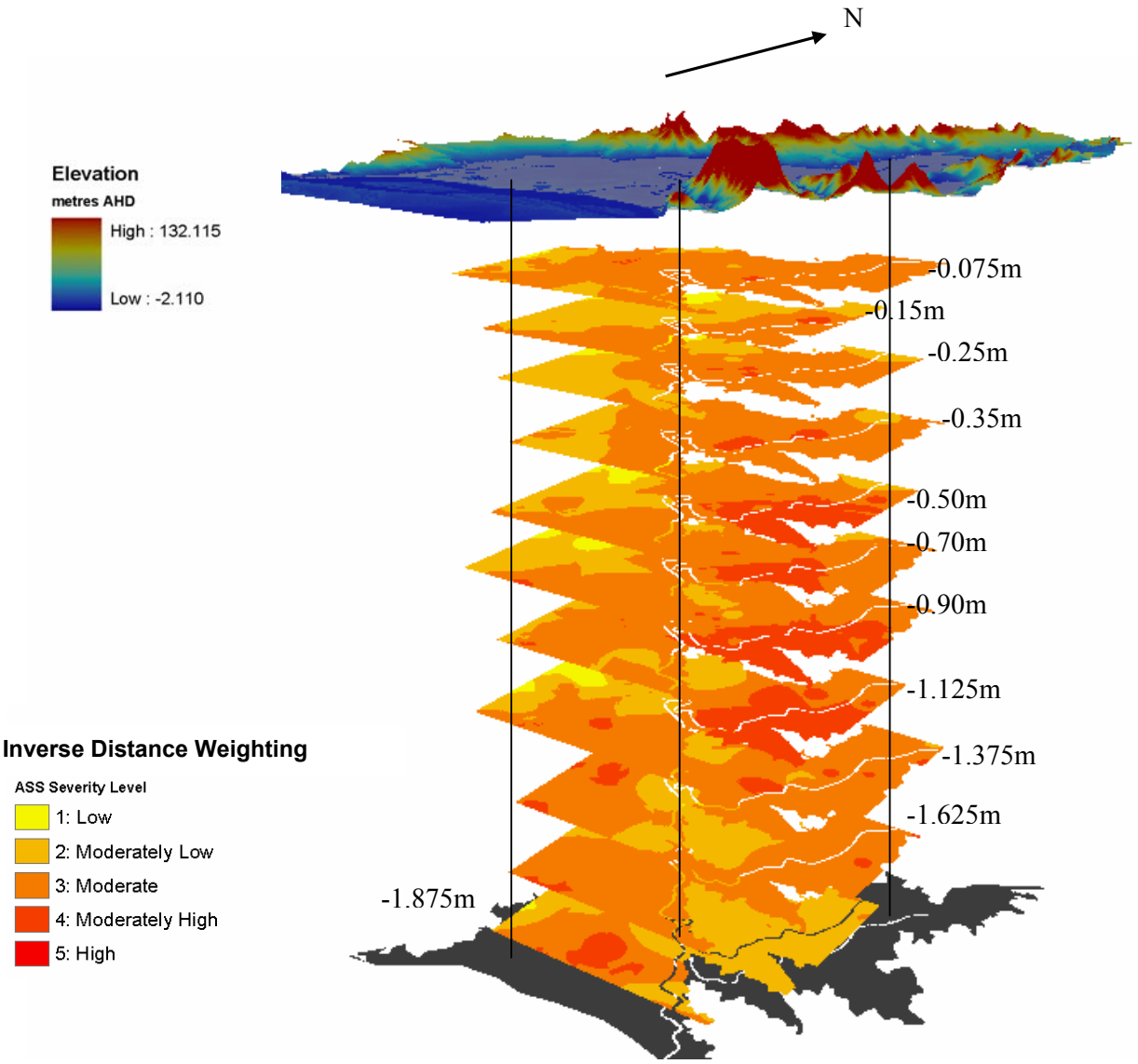

Figure 6.15 Inverse Distance Weighting Severity Map.

6.3 Model 3: Universal Kriging in Parameter estimation (local analysis)

The final method used was Universal Kriging (UK) due to the strength in prediction (low error) in a trial test of Interval 5, generated when selecting the methodology to explore further in the project. The same method was followed in ArcGIS however the local search was selected over a global or neighbourhood approach in all cases (Figure 6.16). This step (step 2) was defined before the optimisation of the covariance and semi-covariance (step 3). Each variable selected in the model was generated with the optimal semi-covariance and covariance and the errors generated were compared (see Chapter 6.3.4). In generating a severity map the process followed was that described by Figure 6.13.

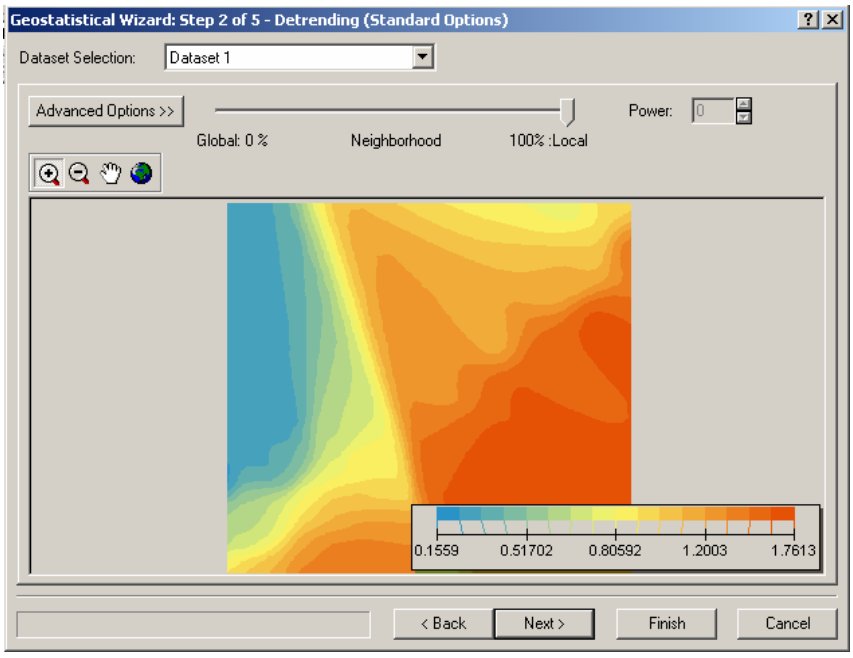

Figure 6.16 Universal Kriging Detrending – Local Search.

6.3.1 Exchangeable Aluminium (3-D Model)

In the ExAl model generated using UK, the toxicity was moderate to high throughout the majority of the lowest lying parts of the floodplain. This was the case from the profile closest to the surface down to the profile 1.125m below the surface (Figure A9, Appendices A6). Below this level there were a few hotspots, mainly in the backswamp areas. The results generated using UK was consistent with IDW and relatively consistent with OK.

6.3.2 Total Actual Acidity (3-D Model)

Total Actual Acidity (TAA) is representative of pH throughout this project and the distribution of TAA throughout the soil profile indicates the acidity held by the soil matrix. Using UK, TAA is highest around the North West and North East of the floodplain, and 0.70m and 1.125m below the ground surface level (Figure A10, Appendices A6). This shows there is a possibility that the bimodal distribution exists in this area but not throughout the floodplain. The results using UK slightly differ to the models using OK or IDW. With UK the upper peak of TAA is slightly lower at 0.70m compared to 0.50m in OK and IDW. However, the lower peak is consistent at 1.125m in all models.

6.3.3 Electrical Conductivity (3-D Model)

The EC model generated by UK is consistent with all other methods used (IDW, OK) in that it picks up the highly saline area on and just below the surface in the Western part of the floodplain (Figure A11, Appendices A6). In that same area, TAA and ExAl are low to moderately low. The highly saline environment within the first few layers of the soil surface could be attributed to seepage onto the land via flood drains that may have, at some time, experienced a king tide which deposited highly saline water in this area.

6.3.4 Error Generation

The error generated for the ExAl model was low except for layers 1 and 3 which appear high, most likely due to the low number of points available to interpolate (Figure A12, Appendices A6). Layers 3, 5, and 9 for TAA had errors lower than all the other layers. This showed that these layers have a good distribution of points, being representative of the distribution of TAA across the floodplain. Overall EC was the best predictor having the optimal Root Mean Squared Standardised (RMSS) error.

6.4 Universal Kriging Severity Map

The severity map generated using UK follows the same methodology as the IDW and OK statistical approaches, this being represented by the flow chart in Figure 6.13. In the UK severity map (Figure 6.17), layer 4 (0.35m below ground surface level) is moderately high in the Northern section of the floodplain around Broughton Creek. There are a few hotspot areas in the next two layers (5 and 6) but it is not until layer 7 and 8 (0.90m and 1.125m below ground surface level) that the distribution or severity increases across the floodplain. This differs slightly when compared to IDW and OK, where layers 6, 7, 8, and 9 are moderately high and layer 4 is moderate in levels of severity.

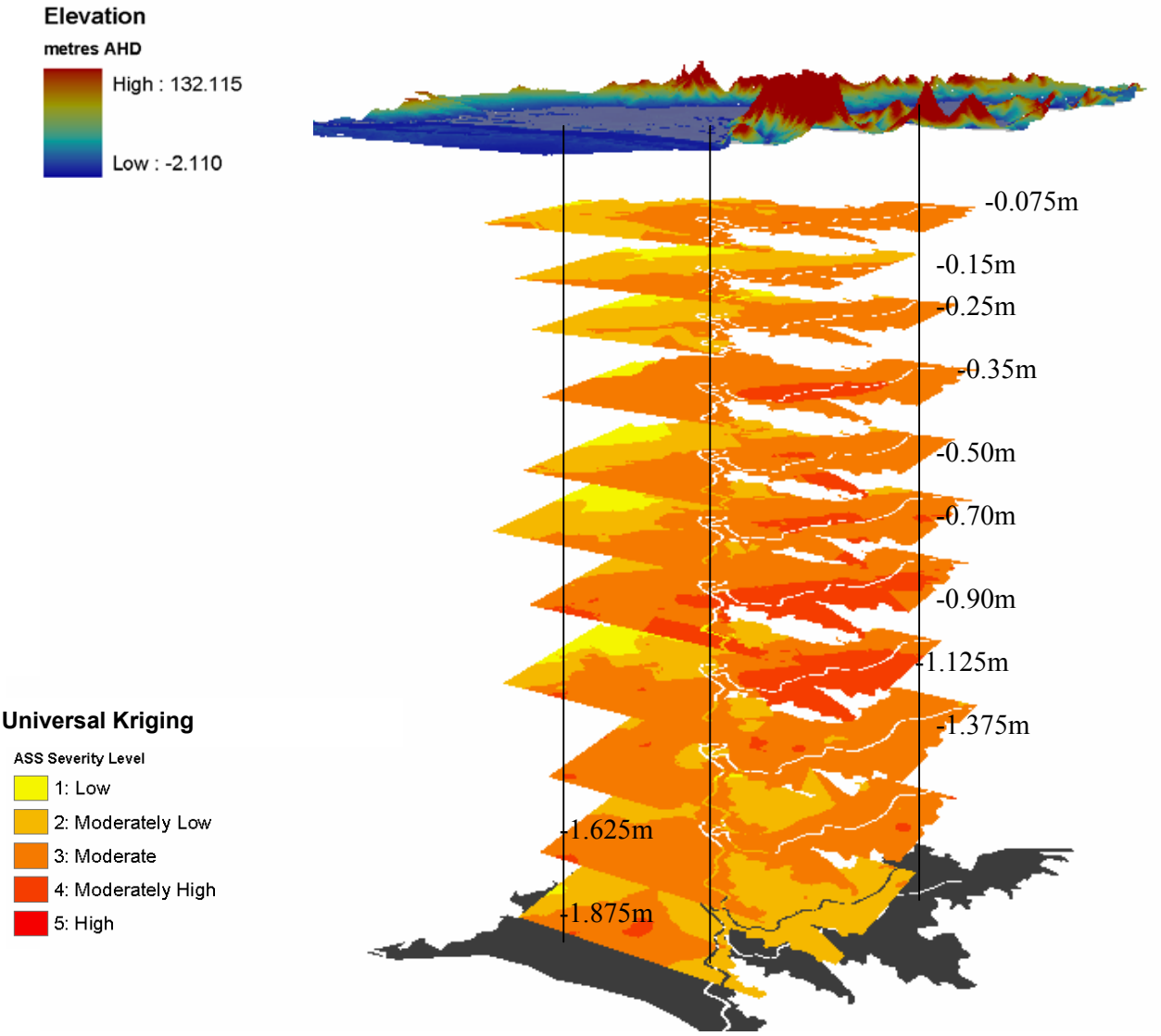

Figure 6.17 Universal Kriging Severity Map.

6.5 Model evaluation

6.5.1 Comparison of three generated models

Each model is compared in Figure 6.18 and each model generated a severity error (Figure 6.19) which can be compared and contrast. OK was closest to the optimal RMSE error for the model, with IDW giving consistent results throughout the soil profile and UK showing the highest error and most variable error down the profile. Therefore, from the three models generated the OK model was the model developed under the most optimal conditions. This model will be known as CASSOK (Coastal Acid Sulfate Soil Ordinary Kriging) for future development and explanation.

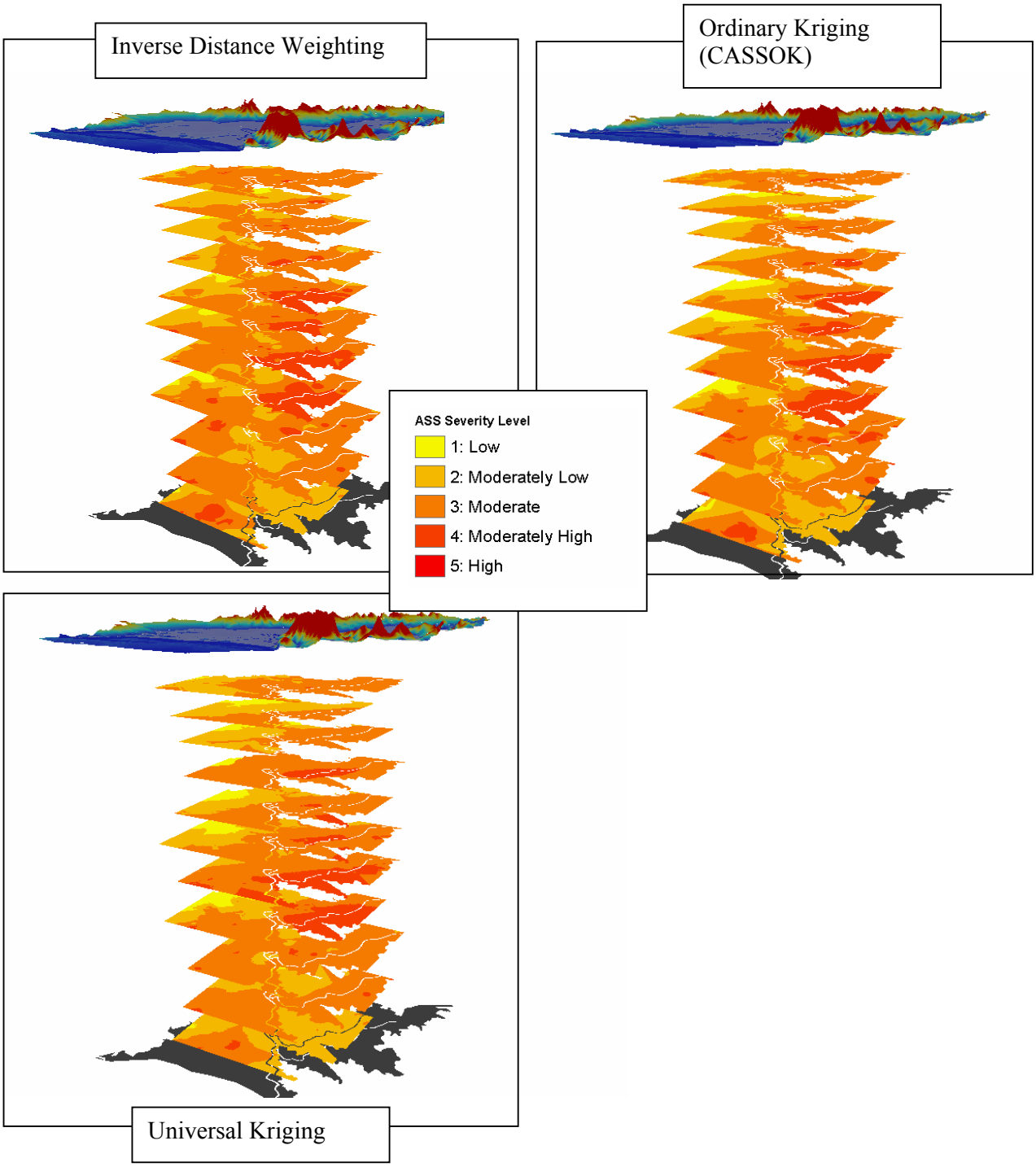

Figure 6.18 Severity Map Model Comparisons.

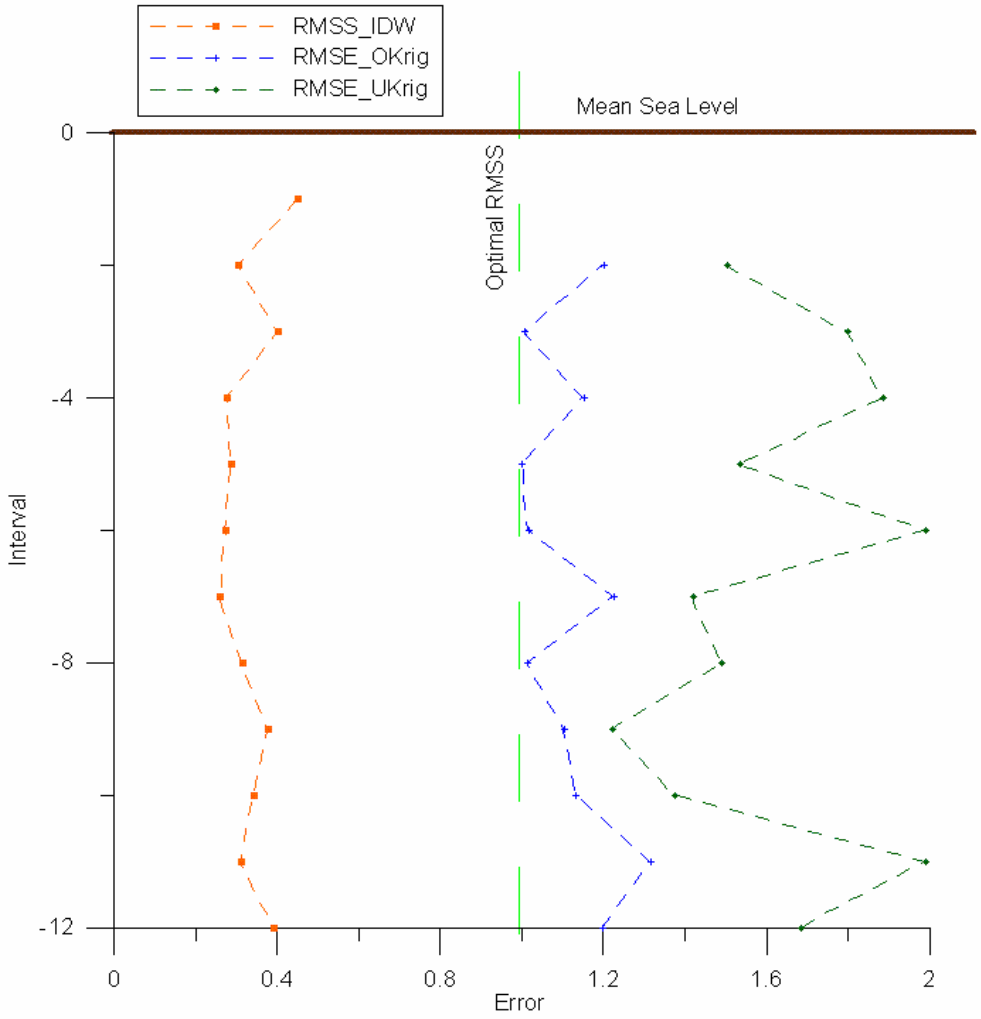

Figure 6.19 Severity Map Errors for all models (UK, OK, IDW)**.**

6.5.2 Comparison to DLWC/DNR Risk Maps

Comparing the CASSOK model to the CASS severity map (Figure 6.20) generates some interesting comparisons, but also emphasises the importance of having both maps to depict the areas that are severely affected by ASS. For instance, the low-risk alluvial plain in the DLWC/DNR risk map generally overlaps the low to moderately low severity area throughout the profile. However, what is not uncovered in the risk map are the acid hotspots further down the profile (layer 12, 1.875m below sea level). These areas will be more of an issue for land planners and developers and should influence land zoning by future land managers. Chapter 7 develops this idea further in a discussion on applying the CASSOK severity maps.

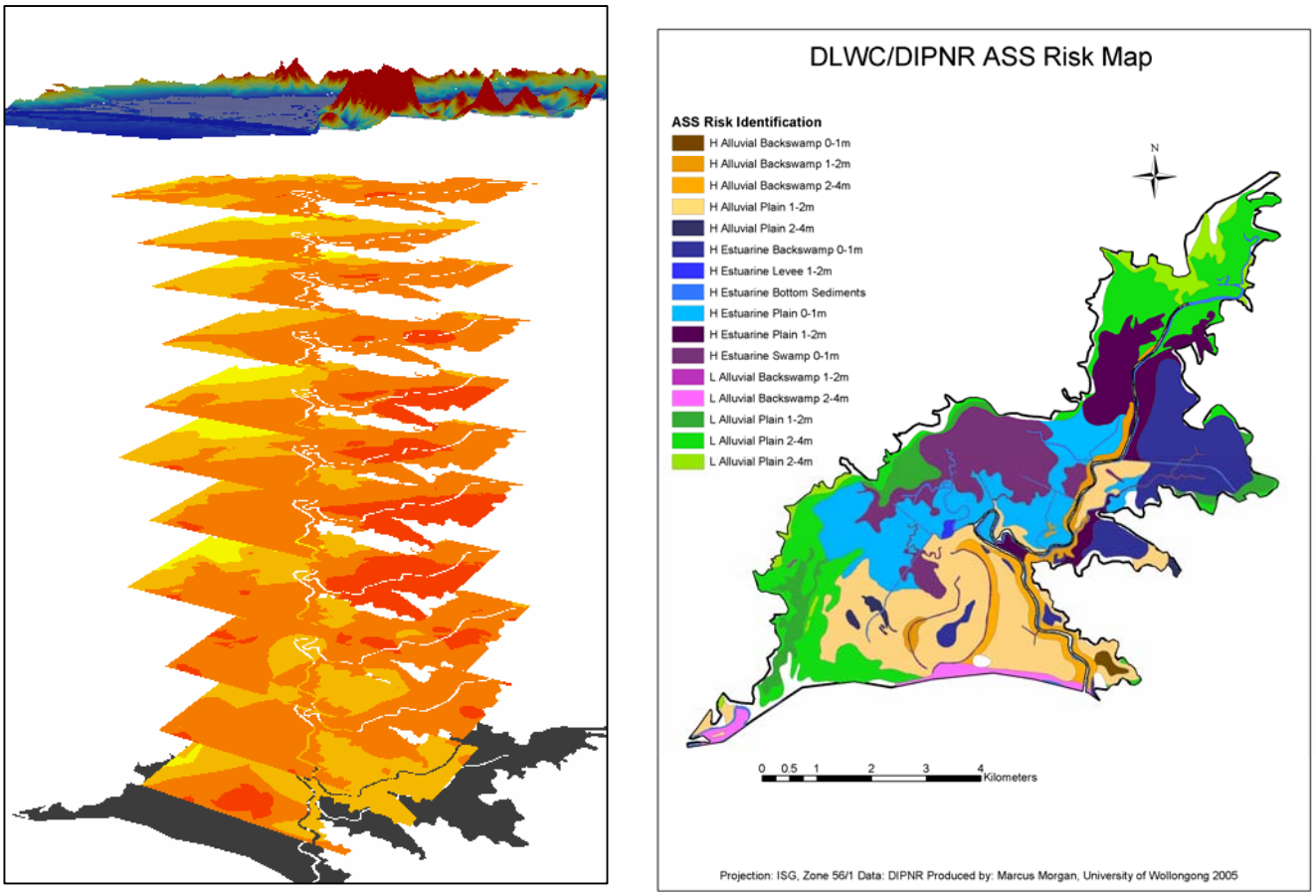

Figure 6.20 Comparison of CASSOK Severity Map and DLWC/DNR Risk Map.

Chapter 7 Application of Coastal Acid Sulfate Soil Severity Model

7.1 Using ASS Severity Maps

The process of developing a new and complimentary set of severity maps is an attempt to improve the understanding of how CASS varies over a floodplain and down a profile. The flow chart represented in Figure 7.1 explains the process from inception to completion for designing an optimal severity map in New South Wales (NSW), Australia using OK as the interpolation tool. The interpretation of results is supported by the severity map produced by DLWC (now DNR) which depicts the geomorphology of the floodplain. By understanding the geomorphology of the floodplain and also what has been assessed as high and low severity in this previous study, the results can be compared and contrast. As explained previously (Chapter 6.4.2) there are a number of similarities in the assessment using the DLWC method and the method employed by this project. However, the main difference is the uncovering of the local statistical variations seen down the profile. The OK method picks these up well in the deeper layers of the profile (see Figure 6.27), showing a number of relatively small high severity areas. For example, at layer 12 (1.875m below ground surface) there is an area to the South West of the floodplain that shows moderately high level of risk of causing adverse affects due to CASS runoff. The probability of these areas going on to form environmental problems caused by CASS will be dependent on how the land is managed in the future, and how it is zoned by local and state authorities.

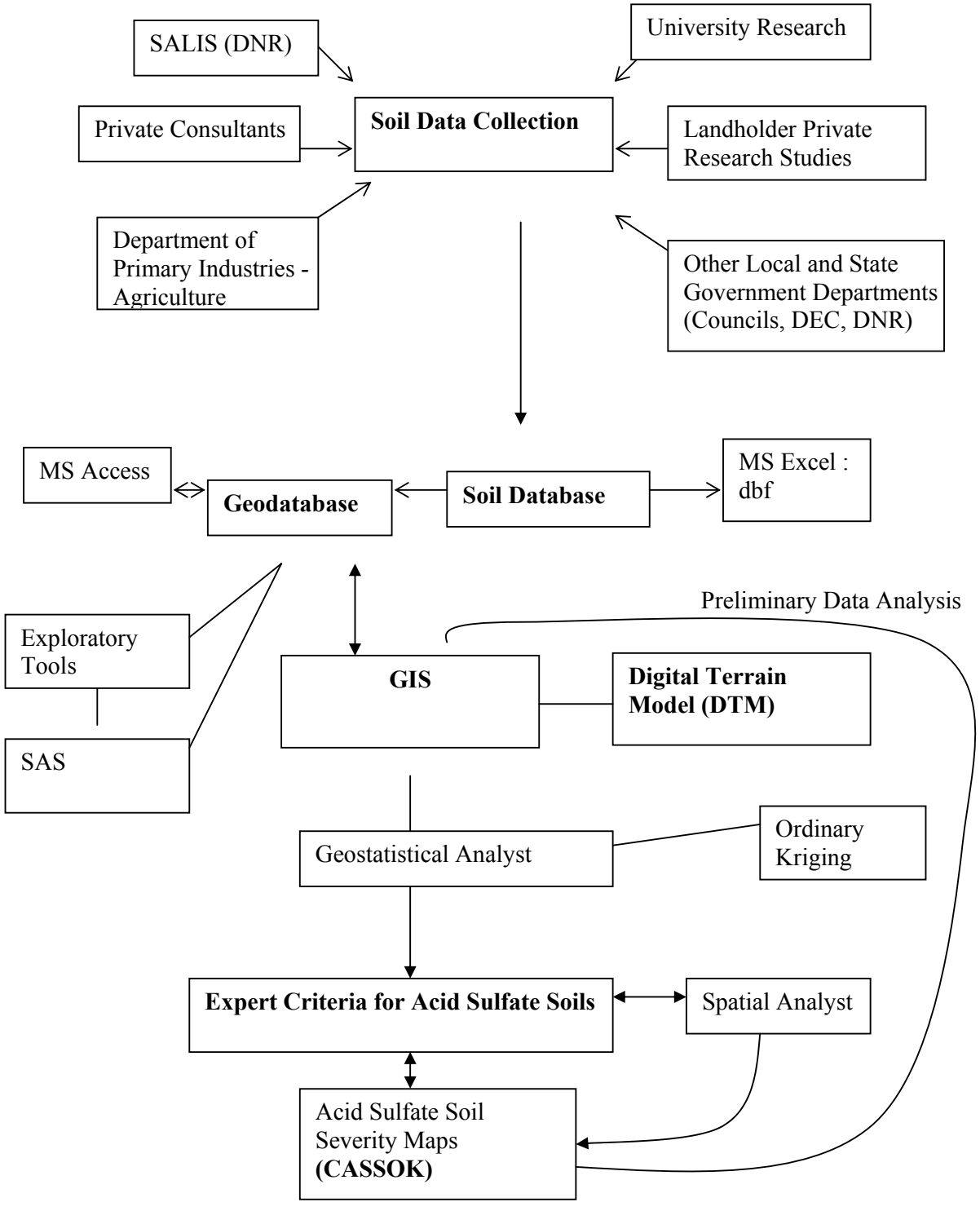

Figure 7.1 CASS Severity Map Schematic.

7.2 Validity of Severity Models

The validity of the CASS Severity maps is generally dependent on the level of error in the models and the future ground truthing that will occur when more soil samples are extracted. The optimal model was selected for the outcome of the project, but due to the number of different variables being input into the model, land managers should take the results as a preliminary indicator of what is most likely to be found when taking a soil sample in the Broughton Creek region. Further investigative boreholes should be taken when a decision that relates to uncovering the soil is made. For example, excavating land for new housing developments.

7.3 Applying Severity Models to Natural Resource Management

Identifying and defining CASS in Australia is the first objective of the National Strategy for the Management of CASS, developed by the Australian Resource Management Council of Australia and New Zealand (ARMCANZ), Australian and New Zealand Environment and Conservation Council (ANZECC), and the Ministerial Council Forestry, Fisheries, and Aquaculture in January 2000. The Management Strategy states the importance of the extent of CASS in Australia, estimating there are over 40,000 km^2 of CASS containing over 1 billion tones of sulfudic compounds (e.g. pyrite). Each tonne of pyrite when oxidized produces 1.6 billion tones of sulfuric acid.

The economic impacts are potentially high due to the high valued coastal land that has a demand for development. The management strategy estimates this waterfront investment and infrastructure is worth $10 billion. Many developments do not continue due to the potential damage caused by CASS and the necessary outlay to reduce the risk of ASS or the future risk in developing in a coastal ASS area.

The main aim of the National Strategy was to improve the management and use of CASS affected land by defining the role of governments (local and state), industry and the wider community in managing CASS.

Identification of these CASS problem areas is the first vital step in understanding the problem. This aids in the development of a management plan designed to combat the problem. This project has produced a set of severity maps for the Broughton Creek-Shoalhaven River floodplain to be used in combination with the risk maps produced by the State government (DLWC, now DNR) in 1996 as well as comparing to historic and current soil maps. As stated in the National Strategy, the extent of CASS needs to be established at a catchment and property level to detail the risks to areas under development pressure.

The suggestion by the National Strategy is to design a reliable property assessment method with an accurate environmental hazard assessment at a catchment level, to improve the water quality and productivity from downstream CASS. The basis of this environmental hazard assessment is the new severity maps created in this project. The severity maps were developed to detail the distribution and chemical composition of soils throughout the Broughton Creek floodplain and catchment.

This process can be applied to other catchments throughout NSW, Australia and in any region where there is enough available soil information. Implementing this approach for identifying severe areas will aid in faster decision making, more accurate analysis, and reduce the costs of alternate methods or approaches. The development of best practice for assessing the make up of soil in a catchment is the next logical step in the process of implementing this tool into management processes.

7.4 Best Practice Management for CASS Identification in a Catchment

Dent & Dawson (1996) developed a type of Best Practice Management (BMP) for CASS identification in their expert system that was interactive with the end user, which describes a variety of landscapes that could possibly be CASS affected areas. However, in this analysis there was no catchment analysis that could determine certain soil chemical properties and relate this back to the severe of the CASS. The likelihood of the land containing PCASS or ACASS, was estimated by many experts, but the questions were preliminary and again based mainly on a simple approach, assessing the geomorphology and making predictions of what this means on a broader scale. Other documents and manuals (e.g. Restoring the Balance – NSW Agriculture, November

2003) have described the process of site investigation and selection but none have looked deeply into using available soil chemical properties combined with highly accurate elevation data and applying this in a statistical model.

The schematic represented by Figure 7.1 becomes the basis for a BMP for CASS on a catchment basis. The CASS 'hotspots' project, funded by the NSW Environmental Trust, targeted Broughton Creek as one of the seven areas throughout NSW to implement on-ground management works, and to help reduce the affects of acid discharge into Broughton Creek and Shoalhaven River. This project went through a planning process described by the schematic in Figure 7.2, which is similar to the NSW floodplain risk management process.

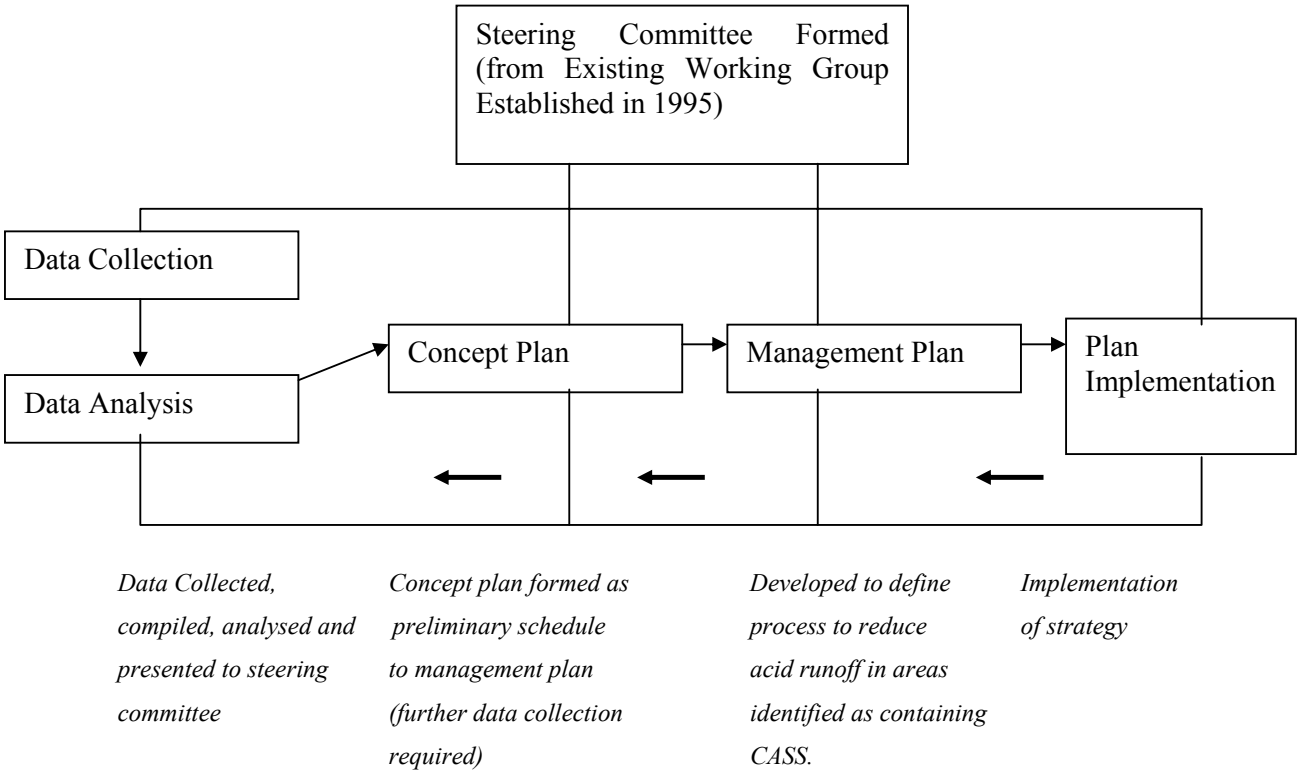

Figure 7.2 Outline of the process of Implementing CASS Management Strategies.

7.5 Coastal Acid Sulfate Soil Legislation

Implementing any best management practice is dependent on the structure of Federal, State and Local government legislation as to how it can flow through to the end users. In 2000, the creation of the National CASS body attempted to address the widening water quality problems attributed to CASS around Australia. However before this CASS problems were seen as a NSW or Queensland problem. Initiatives included in the Australian governments Ocean's Policy aimed at targeting some of the worst areas in NSW and developing on-ground remediation or management solutions. As a result $2.6 Million was apportioned to seven 'hotspot' areas under the National Heritage Trust's program.

Although initiatives such as this are being implemented, there are a number of documents that are being distributed and backed by different State and Federal departments that aim to assist land managers who deal with CASS on a regular basis, but often confuse land managers. For instance in 1997 there were 7 different documents being circulated to aid in the management of CASS (e.g. Environmental Protection Agency (EPA) guidelines from 1995 for assessing and managing CASS). In this period the Acid Sulfate Soils Management Advisory Committee (ASSMAC) was formed to coordinate a 'whole of government' approach to CASS management.

However, a major issue in forming a whole of government approach was the contradictory laws that would be against the implementation of the best management practice for CASS. For instance, the Drainage Act 1939, Rivers and Foreshore Act 1948, and the Water Act 1912 aim to give the landholder a greater drainage structure to alleviate flood or storm water from land. These laws were the catalyst for the implementation of drains in the 1960's, and suggestions (and even enforcement) for maintaining flow through regular clearing of vegetation from these drains.

7.5.1 Environmental Planning and Assessments Act 1979

The *Environmental Planning and Assessments Act 1979* (EPAA) has a provision for different levels of Environmental Planning Instruments (EPIs) through Local Environmental Plans (LEP's), Regional Environmental Plans (REP's) and State Environmental Planning Policies (SEPP's) which control the planning and development of environment issues.

LEP's are generally implemented by Council (for instance Clause 27 of the Shoalhaven LEP recognises that disturbance and identification of CASS as a major issue in planning), REP's (e.g. North Coast REP states that limits to development should be placed on LEP's when land is subject to Potential or Actual CASS) are prepared in NSW by Department of Natural Resources (DNR), and SEPP's by the Minister for the Environment (e.g. SEPP 14: Coastal Wetlands).

As a result of suggestions from various experts in the field such as ASSMAC, DNR developed the 1998 Acid Sulfate Soil Manual which centred on the Environmental Planning and Assessment Act, 1979. Unlike any other manual or publication prior to it, the manual tried to build a unifying set of principles for the management of CASS (Jones, 2000). However, it still appears that there are a number of documents that are being released that don't unify the state of NSW in CASS management. For a land manager to determine who the relevant body is, is quite overwhelming. An all encompassing approach such as the NSW floodplain risk management process which has combined two major documents used by floodplain managers in NSW should become a model for CASS management to work towards on a State level. Until such, more documents will be released being backed by local or regional agendas or by projects focusing on one part of the issue.

This thesis study has attempted to address an approach of improving the identification of the CASS problem and how this can be used in CASS planning. The hope is that will be added to what needs to be an all encompassing policy driven approach to managing CASS at minimum, on a state-wide level. Until policy is formulated that addresses these issues, land managers will continue to be confused about which legislation is relevant for CASS and which legislation is more important and over riding.

CH 8: Conclusion

A process of determining the optimal model to predict areas of CASS that are the most severe or present the most risk (CASSOK) involved trialling a number of statistical methods to determine the most suitable for the problem at hand. This process used simple statistical analysis (descriptive statistical analysis, outlier removal, confidence intervals etc.), compared local (e.g. GWR, IDW) and global (Multiple Regression, GPI) statistical analysis methods, and eventually used interpolation geostatistical techniques to develop a CASS model with the lowest error. OK became the most suitable method given the unique data set and aim of the project. The application of OK to generate a model using soil chemical properties for the prediction of CASS severity has come with some limitations.

8.1 Problems with Locating Data

A number of problems of particular concern when trying to find available data were encountered. The project involved mostly secondary data, including data from the period when records were written and not available in digital formats. Therefore, the data set may not be as comprehensive as it could be due to misplacement of data, lost records, or some other related issue.

8.1.1 Lack of Central Repository

Within the State of NSW there is little organisation of data. This is often the case when projects run throughout the State in different government departments with different personnel managing the projects. Although SALIS did provide some of the available soil data within NSW, this database is far from comprehensive and fails to take into account University research, results from consultant projects, and other tests that landholders may have had performed at their expense. To make the development of a soil database all encompassing there needs to be a central data repository.

8.1.2 Incomplete Information

The data set used in the project had a number of issues that limited the use of some data. These included:

- Missing Data or Incomplete Data: some data sets were incomplete due to recording or sampling error.
- Accuracy of Data: Some of the data was not used due to the suspect accuracy of where it came from or the lack of detail recorded (e.g. what the units of measurement were). Another issue of accuracy was the varying interpretations of a sample by those who had collected and analysed the sample. Some standardisation was necessary to fuse the differing analyses together.
- Metadata: Not knowing where the data came from, what it was representing, and who recorded it automatically eliminated certain data which could have enhanced the data set, or increased the distribution of sample points across the floodplain.

Other limitations included the limited time to upgrade the project into an automated, user-friendly system that could be easily upgraded with the advent of new data. This brings to the recommendations for future improvement of the model.

8.2 Recommendations

A number of recommendations for the improvement of the project are listed. Due to the scope of the project and the limited time to complete the research, some of these were not included or were seen as possibilities only if time permitted:

8.2.1 Online Interactive Tool.

The development of CASSOK in an online, automated environment will reduce the need for someone to know the architecture behind the model. The user will be able to upload the new soil data, it will appear in a GIS online environment and the user will be able to select the various tools which will process the model and generate a layer showing the CASS severity areas within the area that the user is interested in.

8.2.2 Hydrological Modelling.

The project did not look into aspects of hydrology and how this would affect or change the distribution of CASS over time in a given area, as this was beyond the scope of the project. This could be incorporated into the analysis in future projects.

8.3 Disclaimer Limitations

The CASSOK model is useful for determining areas likely to be affected by CASS, but also helps identify those areas that are most likely to be the worst affected areas within a floodplain. Any prediction comes with some uncertainty. CASSOK does not have 100% accuracy in determining CASS severity, but gives a better understanding of floodplains characteristics below the surface than many soil sample records on a piece of paper, or in a spreadsheet that have no similarities or standardisation. CASSOK can be used to save time and money by reducing the need for intense sampling in a certain area. It is useful for development planning, environmental planning, and for research institutions aiming to conduct more intense research on a certain site.

The process of merging all the data into a centralised data base can be used in any environmental project. This process is simple but effective and useful for future generations who are interested in gaining data about anything. Building these type of databases helps improve certainty in prediction – the more samples available for an analysis the less the uncertainty in prediction.

References

Abbott, T.S. (ed.) (1985), *Soil Testing Service - Methods and Interpretation*, NSW Department of Agriculture, Sydney.

Ahmad, S., Simonovic, S. (2004), 'Spatial System Dynamics: New Approach for Simulation of Water Resources Systems', *Journal of Computing in Civil Engineering*, 18(4), pp. 331-340.

Allen, J.R.L., (1964), 'Studies in fluviatile sedimentation: six cyclothems from the Lower Old Red Sandstone, Anglo-Welsh Basin', *Sedimentology* 3, pp. 163-198.

Allman, J.S. and Veenstra, C. (1984), *Geodetic model of Australia 1982*. Technical Report no. 33, Division of National Mapping, Canberra, 60 pp.

Ahern, C.R., Ahern M.R., Powell., B (1998) Guidelines for Sampling and Analysis of Lowland Acid Sulfate Soils (ASS) in Queensland 1998. Queensland Department of Natural Resources, Resource Sciences Centre, Indooroopilly, Queensland p. 42.

Atkinson, G. (1993), Potential acid sulfate soils risk maps, *Proceedings of the National Conference on Acid Sulphate Soils*, 24-25 June 1993, Coolangatta, p 149.

Atkinson, G., Naylor, S.D., Flewin, T.C., Chapman, G.A., Murphy, C.L., Tulau, M.J., Milford, H.B. and Morand, D.T. 1996, 'DLWC Acid Sulfate Soil Risk Mapping', in R.J. Smith & Associates and ASSMAC (eds), *Proceedings 2nd National Conference of Acid Sulfate Soils*, pp. 57

ANZECC (1992), *Australian Water Quality Guidelines for Fresh and Marine Waters,* National Water Quality Management Strategy Paper No 4, Australian and New Zealand Environment and Conservation Council, Canberra.

ANZECC and ARMCANZ (2000), *Australian and New Zealand Guidelines for Water Quality Monitoring and Reporting. National Water Quality Management Strategy Paper No 7*, Australian and New Zealand Environment and Conservation Council, Canberra.

Ball, J.E., Luk, K.C. (1998), 'Modeling Spatial Variability of Rainfall over a Catchment', *Journal of Hydrologic Engineering*, 3(2), pp. 122-130.

Belsley, D., Kuh, E. and Welsch, R.E. (1980), *Regression Diagnostics*. Wiley, New York.

Blunden, B. (2000), *Management of acid sulfate soils by groundwater manipulation*, Unpublished PhD Thesis.

Blunden, B., and Indraratna, B. (2001), 'Pyrite Oxidation Model for assessing ground-water Management strategies in acid sulfate soils', *Journal of Geotechnical and Geoenvironmental Engineering*, American Society of Civil Engineers, 127(2), pp. 146-157.

Bonham-Carter, G.F. (1994), *Geographic Information Systems for geoscientists: modeling with GIS*. Pergamon, Canada, pp. 389.

Bonham-Carter, G.F., Agterberg, F.P., and Wright, D.F. (1988), 'Integration of geological data sets for gold exploration in Nova Scotia', *Photogrammetric Engineering and Remote Sensing*, 54, pp. 1586-92.

Brinkman, R. (1982), 'Social and economic aspects of reclamation of acid sulphate soil areas' in H. Dost and N. van Breeman (eds.), *Proc. Bangkok Symposium on Acid Sulphate Soils*, Jan 8-24, 1981, ILRI Pub. No. 31, International Institute for Land Reclamation and Improvement, Wageningen, pp. 21-36.

Brooke, B., Powell, B., Preda, M., (2005), Coastal Acid Sulfate Soils: Oz Estuaries. Accessed 20 September 2005, <http://www.ozestuaries.org/indicators/>

Burgess, T.M. and R. Webster. (1980), 'Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging', *Journal of Soil Science* 31, pp. 315-331.

Burrough, P.A. (1986), *Principles of Geographic Information Systems for Land Resource Assessment.* Monographs on Soil and Resources Survey No. 12, Oxford Science Publications, New York, pp. 196.

Calkins, H.W., and Tomlinson, R.F. (1977), 'Geographical Information Systems: methods and requirements for land use planning', *IGU Commission on Geographical Data Sensing and Processing (RALI)*, USA, pp. 91.

Charman, P.E.V., & Murphy, B.W. (2000), *Soils – their properties and management*, 2nd Ed., Oxford University Press, Melbourne.

Chappell, J.M.A., Grindrod, J. (1985), 'Pollen analysis: a key to past mangrove communities and successional changes in north Australian coastal environments' in K.N. Bardsley, J.D.S Davie, and C.D. Woodroffe (eds.), *Coasts and Tidal Wetlands of the Australian Monsoon Region*. NARU Monogram. No. 1, NARU, ANU, Darwin, pp. 225-236.

Chrisman, N. (2002), *Exploring Geographical Information Systems*, 2nd Edn. John Wiley & Sons Ltd, University of Washington.

Collier, P.A. (2000), *Transition to the Geocentric Datum of Australia*. Office of Surveyor General (OSG), Melbourne, Australia.

Coram J, Dyson P & Evans R (2001), *An Evaluation Framework for Dryland Salinity*, Report prepared for the National Land & Water Resources Audit (NLWRA), sponsored by the Bureau of Rural Sciences, National Heritage Trust, NLWRA and National Dryland Salinity Program.

Crowe, P., Cinque, P., and Davies, B.K. (2003) *Geographic Information Systems and Spatial Data in the Service of Flood Management: a Report on an Initiative of the State Emergency Service*, Paper presented at the 43rd Annual Conference of the Floodplain Management Authorities of NSW, Forbes.

CSIRO (2003), 'Predicted Vegetation Cover in the Central Lachlan Region', Final Report, Project, AA 1368.97.

Dear,S.E., Moore, N.G., Dobos, S.K., Watling, K.M., & Ahern, C.R. (2002), 'Acid Sulfate Soil Technical Manual: Soil Management Guidelines, Version 3.8', Department of Natural Resources and Mines, Queensland pp 71.

Dent, D. (1986), 'Acid sulphate soils: a baseline for research and development'. International Institute for Land Reclamation and Improvement, publication 39,Wageningen, The Netherlands.

Dent, D., & Dawson, B. (1996), The Acid Test – An Expert System for Acid Sulfate Soils. Accessed 20 June 2004, <http://www.isric.org/Isric/Webdocs/AcidSKit/assweb2.htm>.

DMA (1991), 'Department of Defence World Geodetic System 1984', Technical Report 8350, 2, 2nd ed., U.S. Defence Mapping Agency.

Dobermann, A., and Fairhurst T. (2000), 'Rice, nutrient disorders & nutrient management', Handbook series, Potash & Phosphate Institute (PPI), Potash & Phosphate Institute of Canada (PPIC), and International Rice Research Institute (IRRI), pp 191.

DPI-Agriculture (2004) What is soil pH? Accessed 12 February 2005, <http://www.agric.nsw.gov.au/reader/soil-acid/ss392-soil-ph.htm>.

ESRI (Environmental Systems Research Institute) (2004) ArcGIS. Accessed 3 June 2004, <http://www.esri.com>.

Evangelou VP (1995), *Pyrite oxidation and its control: solution chemistry, surface chemistry, acid mine drainage (AMD), molecular oxidation mechanisms, microbial role, kinetics control, ameliorates and limitations, microencapsulation*, CRC Press, Florida.

FAO (2001), 'Lecture Notes on the Major Soils of the World,' FAO, Rome.

FAO-UNESCO (1978), 'Soil Map of the World,' 10, Australasia, UNESCO, Paris.

Flentje, P., Chowdhury, R., and Tobin, P. (2002), 'GIS Framework for Landslide Risk Management' in R.G. McInnes and J. Jakeways (eds.) *Proceedings of the International Conference on Instability – Planning and Management May 2002*, Isle of Wight, UK, pp. 237 – 248.

Fotheringham A.S., Brunsdon, C., Charlton, M. (2000), *Quantitative Geography: Perspectives on Spatial Data Analysis*. SAGE Publications, London 270 pp.

Fotheringham A.S., Brunsdon, C., Charlton, M. (2002), *Geographically Weighted Regression: the analysis of spatially varying relationships*. John Wiley & Sons Ltd, West Sussex 269 pp.

Fraser, C., Jonas, D., and Turton, D. (2000), 'Report on 1998 Airborne Laser Scanner Trials', 2, 11th May 2000. AAM Surveys Pty Limited, Queensland.

Galloway, R.W. (1982), 'Distribution and physiographic patterns of Australian mangrove' in B.F. Clough (ed.) *Mangrove Ecosystems in Australia*, ANU Press, Canberra, pp. 276-286.

Glamore, W. (2003), *Evaluation and Analysis of Acid Sulfate Soil impacts via Tidal Restoration*, Unpublished PhD Thesis.

Goodchild, M.F.(1993), 'The State of GIS for environmental problem-solving' *in M*.F. Goodchild, B.O. Parks and L.T. Steyaert (eds.)*, Environmental Modeling with GIS*, Oxford University Press, New York, pp. 8-15.

Goovaerts, P. (1997), 'Geostatistics in Soil Science: State-of-the-Art and Perspectives', in *Proceedings of Pedometrics '97, August 18-20, 1997*, Madison, Wisconsin, USA.

Graham T.L., and Larsen R.M. (2000), 'Coastal Geomorphology: Progressing the Understanding of Acid Sulfate Soil Distribution' in C.R. Ahern, K.M. Hey, K.M. Watling and V.J. Eldershaw (eds.) *Acid Sulfate Soils: Environmental Issues, Assessment and Management, Technical Papers*, *Brisbane, 20–22 June, 2000, 13, pp 1-10*. Department of Natural Resources, Indooroopilly, Queensland.

Hartsock, N. J., Mueller, T. G., Thomas, G. W., Barnhisel, R. I., Wells, K. L., and Shearer, S. A. (2000), 'Soil Electrical Conductivity Variability', in. P.C. Robert et al. (ed.) *Proc. 5th international conference on precision Agriculture*. ASA Misc. Publ., ASA, CSSA, and SSSA, Madison, WI.

Heuvelink, G. (1997), 'Spatial Aggregation and Soil Process Modelling', in *Proceedings of Pedometrics '97, August 18-20, 1997*, Madison, Wisconsin, USA.

Higgins, M. (1987), 'Transformation from WGS84 to AGD84', an interim solution. Department of Geographic Information memo, Queensland.

Hilton, B., & Elioff, A. (2004), 'California High Speed Train: A Statewide Geo-Seismic Evaluation Using GIS', in M.K. Yegian, and E. Kavazanjian (eds.), *GeoTrans 2004 Geotechnical Engineering for Transportation Projects (GSP No. 126)*, Los Angeles, California, USA.

Howe, D. (1993) FOLDOC: The Free On-line Dictionary of Computing, Accessed 8 June 2004, <http://foldoc.org/>.

Huber W.A. (2001), 'Geographical Information System Course Notes', Penn State Valley.

Indraratna, B., Sullivan, J. & Nethery, A. (1995), 'Effect of Groundwater Table on the Formation of Acid Sulphate Soils', *Mine Water and the Environment*, 14, pp. 71-83, Wollongong.

Isaaks, E.H., and Srivastava, R.M., (1989), *An Introduction to Applied Geostatistics*, Oxford University Press, NY, 561 pp.

Isbell, R.F. (1993), *A Classification System for Australian Soils,* 3rd Approximation, CSIRO, Australia.

Konsten, C.J.M., Brinkman, R., Andriesse, W. (1988) 'A field laboratory method to determine total potential and actual acidity in acid sulphate soils' in H.Dost (ed.), *Selected Papers of the Dakar Symposium on Acid Sulphate Soils*, pp. 106-134, ILRI: Wageningen, The Netherlands.

Johnson, K.M., (1987), 'Natural resource modeling in the GIS environment', *Photogrammetric Engineering and Remote Sensing* 53, pp. 1411-1415.

Johnston, K., Ver Hoef, J.M., Krivoruchko, K., Lucas, N. (2001), *Using ArcGIS Geostatistical Analyst*, ESRI, Redlands, 300 pp.

Jones, D. (2000), 'Acid Sulphate Soils: The Environmental Time Bomb', *Australasian Journal of Natural Resources Law & Policy*, 7 (1).

Kallenberg, O. (1997), *Foundations of Modern Probability*, Springer-Verlag, New York.

Kao, J-J., Lin, H-Y, (1996), 'Multifactor Spatial Analysis for Landfill Siting', *Journal of Environmental Engineering*, 122 (10), pp. 902-908.

Kirkby, S.D. (1994), 'Managing Dryland Salinization with an Integrated ES/GIS System', in E.G.Masters and J.R.Pollard (Eds.), *Proceedings Land Information Management, Geographical Information Systems*, 1, pp. 143-156.

Lawrie, R. (2003), *pers. comm.*, 28 August.

MacDonald, K.B., Kloosterman, B. (1984), 'The Canada Soil Information System (CanSIS) General User's Manual'. Land Resource Research Centre, Research Branch, Agriculture Canada Ottawa, Ontario, 56 pp.

Malkawi, A.I.H., Saleh, B., Al-Sheriadeh, M.S., Hamza, M.S. (2000), 'Mapping of Landslide Hazard Zones in Jordan Using Remote Sensing and GIS', *J. Urban Plng. and Devel.*, 126 (1), pp. 1-17.

Matthews K.J. (1997), 'Transfer of Australian Map Grid (AMG) to Geocentric Datum of Australia', *Surveying Australia*, September 1997.

Manning, J., and Morgan, P. (1992), 'What is IGS and why is it important to Australia?' *The Australian Surveyor*, 37(2), pp. 91-100.

Mass, H.G., and Vosselman, G. (1999), 'Two Algorithms for extracting building models from raw laser altimetry data,' *Journal of Photogrammetry & Remote Sensing,* 54, pp. 153-163, ISPRS.

McElnea, A.E., Ahern, C.R., Menzies, N.W. (2000), 'The measurement of actual acidity in acid sulfate soils and the determination of sulfidic acidity in suspension after peroxide oxidation,' *Australian Journal of Soil Research,* 40(7), pp. 1133 – 1157.

Melville, M.D., White, I., and Lin, C. (1993). 'The Origins of Acid Sulfate Soils', in R.Bush (ed.), *Proceedings of the First National Conference on Acid Sulfate Soils*, Coolangatta, pp 19-25.

Kollias, V.J., Kalivas, D.P., and Yassoglou, N.J. (1999), 'Mapping the Soil Resources of a Recent Alluvial Plain in Greece Using Fuzzy Sets in a GIS Environment,' *European Journal of Soil Science*, 50(2), pp. 261-273.

Milford, H.B., McGaw, A.J.E., & Nixon, K.J. (2001) *Soil Data Entry Handbook (3rd Edn.) for the NSW Soil and Land Information System (SALIS)*, NSW Department of Land and Water Conservation Resource Information Systems Group, Parramatta, 79 pp.

Montanarella, L. (2003), 'The European Soil Information System and its Extension to the Mediterranean Basin', Options Méditerranéennes, Série B, 34.

Morgan, M.J., Papworth, W., Aney, P., Perry, J., & Indraratna, B. (2003), 'Application of Airborne Laser Scanning for managing coastal acid sulfate soils in the Broughton Creek Floodplain' in C.D. Woodroffe, and , R. A. Furness (eds.), *Coastal GIS 2003: An Integrated approach to Australian coastal issues, Wollongong Papers on Maritime Policy*, 14, pp. 401-414.

Morgan, M.J. Indraratna, B.I., Golab, A. (2005), 'Acid Sulfate Soil Hotspots Data Analysis Final Report to Shoalhaven City Council, June 2005', University of Wollongong pp 77.

Moore, D., and McCabe, G. (1993), *Introduction to the Practice of Statistics*, W.H. Freeman and Company, New York, 854 pp.

Mulla, D.J. (1997), 'Geostatistics, remote sensing and precision farming,' in J.V. Lake, G.R. Bock and J.A. Goode (eds.) *Precision Agriculture: Spatial and Temporal Variability of Environmental Quality*, John Wiley & Sons, New York, pp.100-119.

Mulvey, P., (1993), 'Pollution, Prevention and Management of Sulfidic Clays and Sands'. in R. Bush (ed.), *Proceedings of the First National Conference on Acid Sulfate Soils*, Coolangatta, p116-129.

Naylor S.D., Chapman G.A., Atkinson G., Murphy C.L., Talau M.J., Felwin T.C., Milford H.B., and Morand D.T. (1995), *Guidelines for the use of acid sulphate soils risks maps*. NSW Soil Conservation Service. Department of Land and Water Conservation, Sydney, New South Wales.

Norwood, T.G. (1975), *Studies of the Acid Sulphate Soils of the Lower Shoalhaven River Floodplain*, University of Wollongong, Unpublished Honors Thesis.

Openshaw, S. (1990), 'Spatial Analysis and Geographical Information Systems: A review of progress and possibilities' in H.J. Scholten, and J.C.H Stillwell (eds), *Geographical Information Systems for Urban and Regional Planning*, Kluwer Academic Publishers, Dordrecht, Boston, MA, pp. 153-163.

Parsons, R.L., Frost, J.D. (2002), 'Evaluating Site Investigation Quality using GIS and Geostatistics,' *Journal of Geotechnical and Geoenvironmental Engineering*, 128(6), pp. 451-461.

Pease, M.I. (1994), *Acid sulphate soils and acid drainage, Lower Shoalhaven floodplain, New South Wales*, University of Wollongong, Unpublished Masters Thesis.

Pease, M.I., Nethery, A.G., Young, A.R.M. (1997), 'Acid Sulfate Soils and Acid Drainage, Lower Shoalhaven Floodplain, New South Wales', *Wetlands (Australia),* 16 (2), pp56-71.

Perry, J., Aney, P., and Turton, D. (2001) 'Shoalhaven Acid Sulphate Soil Amelioration Project,' *Scanning the Horizons*, 10, pp. 5-6.

Pons, L.J. (1973), Outline of the genesis, characteristics, classification and improvements of acid sulfate soils, in H.Dost (ed.) *Acid Sulfate Soils. Proceedings of the International Symposium, Acid Sulfate Soils, 13-29 Aug. 1972*, Wageningen, International Institute for Land Reclamation and Improvement, Wageningen, The Netherlands, p. 327.

SCARM & ARMCANZ, (1999), 'National Working Party on Acid Sulfate Soils: National Strategy for the Management of Coastal Acid Sulfate Soils,' NSW Agriculture, Wollongbar.

Stone, Y., Ahern C.R., and Blunden B. (1998), *Acid Sulfate Soils Manual 1998*. Acid Sulfate Soil Management Advisory Committee, Wollongbar, NSW, Australia.

Pons, L.J. and van Breeman, N. (1982), 'Factors influencing the formation of potential acidity in tidal swamps,' in H. Dost (ed) *Proceedings of the Second International Symposium on Acid Sulphate Soils*, ILRI Publ., No. 31, International Institute for Land Reclamation and Improvement, Wageningen, The Netherlands, pp. 37 - 51.
Radke, L.C. (2000), *Solute divides and chemical facies in southeastern Australian salt lakes and the response of ostracods in time and space*, Ph.D Thesis, Department of Geology, The Australian National University, 232 pp.

Rawlings, D.E. (1997), *Biomining: Theory, Microbes and Industrial Processes*, Springer Verlag, New York, 302 pp.

Rizos, C. (1999), 'Principles and Practice of GPS Surveying. Satellite Navigation & Positioning Laboratory (SNAP),' School of Geomatic Engineering, The University of New South Wales.

Roy, P.S, Thom, B., and Wright, L. (1980), 'Holocene sequences on an embayed high-energy coast: an evolutionary model,' *Sedimentary Geology*, 26, pp. 1-19.

Roy, P.S (1984), 'New South Wales estuaries: their origin and evolution,' in Thom, B.G. (ed.), *Coastal Geomorphology in Australia*, Academic Press Australia, Sydney.

Sammut, J., Lines-Kelly, R. (2001), 'An Introduction to Acid Sulfate Soils. NSW Department of Primary Industries, 23 November 2001, (Accessed 22 November 2004). <http://www.agric.nsw.gov.au/reader/soil-acidss/acid-intro.htm>.

Sands, R.J., Chang, C.C., and McDonald, J.M. (2002), 'Storm Water Management Study After Flooding of the South Bronx, NYC, New York,' in E.W. Strecker, and W.C. Huber, (eds.) *Urban Drainage 2002: Global Solutions for Urban Drainage. 9th International Conference on Urban Drainage (9ICUD), September 8–13, 2002*, Portland, Oregon.

SBV (Surveyors Board of Victoria) (2003), 'Surveying Using the Global Positioning System'. Melbourne, Australia, Accessed 20 October 2005, <http://www.surveyorsboard.vic.gov.au/gps.htm>.

Seppelin, T.O., (1974), 'The Department of Defense World Geodetic System 1972,' *Proc. Int. Symp. on Problems Related to the Redefinition of the North American Geodetic Networks*, Fredericton, New Brunswick, Canada.

Scott, H.D., and Udouj, T.D. (1999), 'Spatial and temporal characterization of land-use in the Buffalo National River Watershed,' *Environmental Conservation* 26 (2), pp. 94–101.

Tabachnick, B., and Fidell, L. (1989), *Using Multivariate Statistics*. Harper & Row Publishers, New York, 746 pp.

Taylor, S. (1996), *Dryland salinity,* Department of Land and Water Conservation, Bathurst.

Tao, G., and Yasuoka,Y. (2002), 'Combining High Resolution Satellite Imagery and

Airborne Laser Scanning data for Generating bareland DEM in urban areas,' in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science: Proceedings of the International Workshop on Visualization and Animation Of Landscape, 26 - 28 February 2002*, Kunming, China, 34 (5), W3.

Tomlinson, R.F., (1987), 'Current and potential uses of geographical information systems: The North American experience,' *International Journal of Geographical Information Systems* 1, pp. 203-18.

Turton, D. (2002), 'Introduction to Airborne Laser Scanning,' *11th ARSPC Conference Workshop on Airborne Laser Scanning*, *Brisbane 2-6th September 2002*.

Umitsu, M., Buman, M., Kawase, K., & Woodroffe, C.D. (2001), 'Holocene palaeoecology and formation of the Shoalhaven River deltaic-estuarine plains, southeast Australia,' *The Holocene*, 11(4), pp. 407-418.

USDA-NRCS (2001), 'Delivering Soil Data,' Briefing Note Number 1A, Accessed 12 December 2004, <http://nasis.nrcs.usda.gov/documents>.

van Breeman, N. (1973), 'Soil forming processes in acid sulphate soils,' in H. Dost (ed.) *Acid Sulphate Soils. Proc. Int. Symp., Wageningen*, ILRI, International Institute. for Land Reclamation and Improvement, Wageningen, 128(1), pp. 66-130.

van Breeman, N. (1976), *Genesis and solution chemistry of acid sulfate soils in Thailand*. Doctoral thesis, Wageningen, 263 pp.

Veris Technologies (2005), How does a Veris EC System work? Accessed 25 June 2005, <http://www.veristech.com>

Walker, P.H. (1972), 'Seasonal and stratigraphic controls in coastal flood plain soils', *Australian Journal of Soil Research*, 10, pp. 127-142.

Webster, I. (2005), 'The Salinity of Coastal Waterways', CSIRO Land and Water, Coastal Zone Australia, Canberra.

White, I., Melville, M.D., Sammut, J., Van Oploo, P.V., Wilson, B.P., and Yang, X. (1996), *Acid Sulfate Soils: facing the challenges*, Earth Foundation Australia Monograph 1, Sydney.

Wolf, D.C., and Skipper, H.D. (1994), 'Soil sterilization,' in R.W. Weaver, S. Angle, P. Bottomley, D. Bezdicek, S. Smith, A. Tabatabai, A. Wollum, and S.H. Mickelson (eds.), *Methods of soil analysis, Part 2: Microbiological and biochemical properties*, Soil science Society of America Inc., pp. 41-51.

Woodroffe C.D., and Chappell, J. (1991), 'Application of Holocene studies to conservation: the case of low-energy coasts,' In: G. Brierley, and J. Chappell (eds.) *Applied Quaternary Studies*, ANU, Canberra pp. 75-88.

Wulder, M. (2002), *A Practical Guide to the Use of Selected Multivariate Statistics*, Canadian Forest Service Pacific Forestry Centre, Victoria, Canada.

Wynn, J. (2001), 'A Ground Electromagnetic Survey Used to Map Sulfides and Acid Sulfate Ground Waters at the Abandoned Cabin Branch Mine, Prince William Forest Park, Northern Virginia Gold-Pyrite Belt,' U.S. Geological Survey Open-File Report 00-360, Reston, VA.

Xie,M., Esaki,T., Zhou,G., and Mitani, Y. (2003), 'Geographic Information Systems-Based Three-Dimensional Critical Slope Stability Analysis and Landslide Hazard Assessment,' *Journal of Geotechnical and Geoenvironmental Engineering*, 129(12), pp. 1109-1118.

Zhua, H.C., Charleta, J.M., and Poffijn, A. (2001), 'Radon risk mapping in southern Belgium: an application of geostatistical and GIS techniques,' *The Science of the Total Environment,* 272, pp. 203-210.

Zhixiong, D., Yalin, H., Jiren, L., and Shifeng, H. (2005), 'A Model of Flood and Waterlogging Disaster Loss Assessment Based on the Remote Sensing and GIS Spatial Information Grid,' in R. Walton (ed.), *EWRI 2005 Impacts of Global Climate Change, World Water and Environmental Resources Congress 2005*, *May 15–19, 2005*, Anchorage, Alaska, USA.

Appendices:

A1 Formats of GIS Data

A1.1 Vector

Vector data are defined by x, y coordinates that provide very accurate data with the benefit of limited data storage requirements. Features, events, and activities with a spatial component are modelled as points, lines or polygons to form a geographical relational database or geodatabase. Vector data can be modified, updated and analysed in the GIS to obtain the outcomes a user hopes to achieve. A typical example of a vector file in GIS is a soil borehole, represented as a point file that contains information pertaining to that geographical location. Polygon vector files are used to model areal features such as cadastral boundaries in local government planning. Data is acquired from one of a number of means but is generally either manually digitized or scanned from a hard copy of the data or acquired via analytical coordinate data, calculation or analytical geometry function solution. In vector files, a number of fields can specified (such as Date of Sample, pH of Soil etc.) with their corresponding row/attribute in a table, which can be collated into a geodatabase. Shape files and coverages are typically the file types that represent vectors.

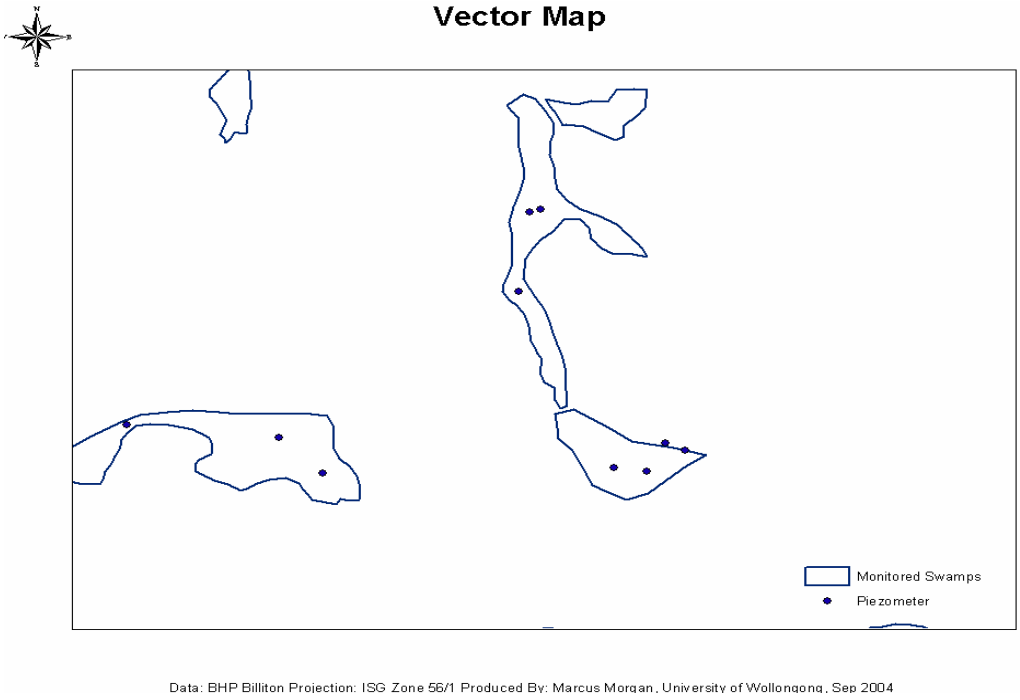

Figure A1 Example Vector Map with Point and Polygon Themes

A1.2 Coverages

A Coverage is a collection of geographic features and associated attribute information, analogous to a table in a database. Each feature class stores a set of points, lines (arcs), polygons, or annotation (text). Feature classes can have topology, which determines the relationships between features. The files describing topological and coordinate information for geographic features are stored in a single coverage sub-directory and additional attribute files in INFO sub-directory. Attribute information may also be stored in other databases such as MS Access or ORACLE. (ESRI, 2003; MASS.GOV, 2004). ArcGIS can export and import coverages in a geodatabase to be used in Microsoft Access and data can be manipulated or updated in the database interface.

A1.3 Shapefiles

Shapefiles store geographic features and their attributes. Geographic features in a shapefile are represented by points, lines, or polygons (see Figure 3.8). A folder might also contain database tables, which can store additional attributes that can be joined to a shapefile's features (ESRI, 2003). In a geodatabase, shapefiles and coverages are represented as feature classes. One of the additional benefits of ArcGIS is that when creating a geodatabase, it can be manipulated in MS Access, which is a lot easier when editing.

A1.4 Raster

Raster data composes of a series of grids or pixels that contain information about a specific geographical coordinate (e.g. landslide susceptibility). The sizes of these grids are dependent on the accuracy of the dataset. For instance a resolution of 0.1m (0.1m x 0.1m cell) is considered to be accurate to within 0.1m of an unknown point.

Raster data is generally divided into two categories, thematic data and image data. The values in thematic raster data represent some measured quantity or classification of a particular phenomenon, such as elevation, pH concentration, or slope. For example, in a soil class map the value 3 may represent silty-clay loam soil and the value 5 may represent clay soil. The values of cells in an image represent reflected or emitted light or

energy, such as that of a satellite image or scanned photograph. Other image processing software such as IMAGINE use raster image data over thematic

raster data. Within the ESRI line of software, the spatial analyst function uses thematic data as input enabling map algebra to be used, hence the transformation of data.

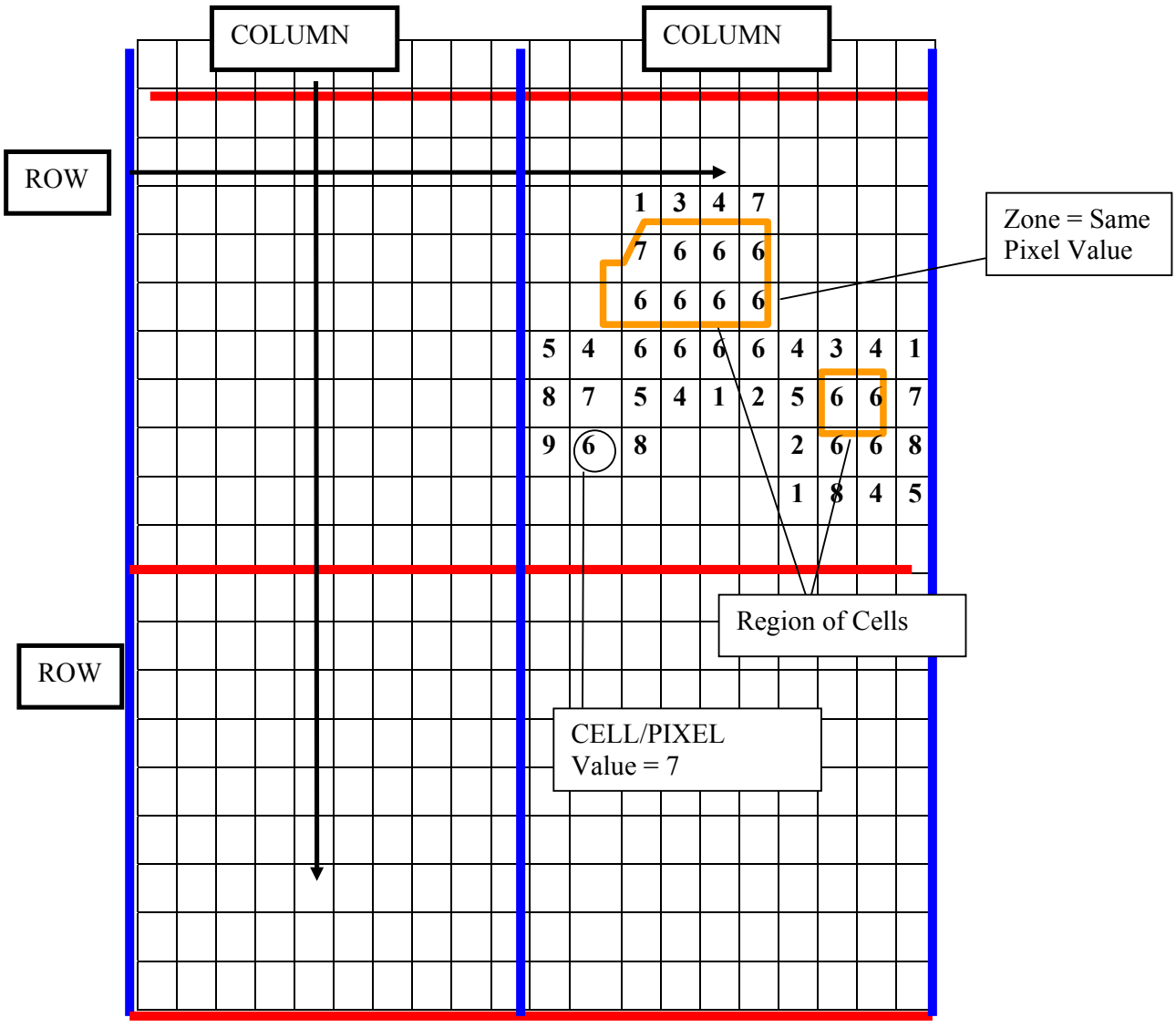

Figure A2 A typical example of a Raster thematic data.

Raster data are characterised by pixels or cells that are assigned a number for classification. Cells (as numbers) within proximity to each other form what is known as

a region of cells. A region can also consist of an individual cell if there are no other cells adjacent to the cells that are classified with the same number. A number of cells make up grids in rows and columns as in Figure A2.

A2 Geographical Coordinate Systems (GCS)

A Geographical Coordinate Systems (GCS) uses a three-dimensional spherical surface to define locations on the earth. A GCS includes an angular unit of measure, a prime meridian, and a datum, which is based on a spheroid.
Any point is referenced by its Longitude and Latitude values (or Easting and Northing values) of a particular coordinate system. Longitude and Latitude are often measured in degrees representing the angle from the located point to the earths centre, from between -90° at the South Pole to +90° at the North Pole for latitude and from -180° to 180° (traveling west-east) for longitude. Latitude is defined as 'Horizontal lines', or east–west lines of equal latitude, or parallels. Longitude is described as a number of 'vertical lines', or north–south lines, of equal longitude, or meridians. These lines encompass the globe and form a gridded network called a graticule where the origin is (0,0) located at the point where the prime meridian (Greenwich, UK) and equator intersect.

Although longitude and latitude can locate exact positions on the surface of the globe, they are not uniform units of measure. Only along the equator does the distance represented by one degree of longitude approximate the distance represented by one degree of latitude. This is because the equator is the only parallel as large as a meridian. (Circles with the same radius as the spherical earth are called great circles. The equator and all meridians are great circles.)

For example, due to the effect of the great circles (depending on the spheroid used for exact distances), one degree of longitude at the equator will be far greater in distance than at latitude 70. As the meridians converge toward the poles, the distance represented by one degree of longitude decreasing to zero. This is why you can’t display data accurately on a 1 or 2 dimensional surface and such there is a need for software and computer systems such as GIS to display the data accurately. Otherwise accuracy would be compensated. Of the geographical coordinate systems, Universal Transverse Mercator (UTM) is the most commonly used and widely accepted.

A2.1 Universal Transverse Mercator Grid System (UTM)

The Universal Transverse Mercator (UTM) grid system is a specialized application of the transverse Mercator projection, which is both cylindrical and conformal. It was first used by the US Army Map Service in 1947 for their use in worldwide and strategic mapping and has been adopted throughout the world. UTM divides the world into 60 numbered zones, both north and south, separated by the equator. Each of the 60 zones is 6 degrees of longitude (total 360) and each has its own central meridian. In Australia, New South Wales falls into UTM zones 54,55,56. The study site falls into UTM zone 56.

UTM generates linear parameters so that the origin of each of the zones is the intersection of its central meridian and the equator. Each parameter used to reference data is applied to this origin, to limit the range of the data and to limit the data to positive values (ESRI, 2003).

False Easting:	500000
False Northing:	1000000
Scale Factor:	0.9996

Table A1 UTM Linear Parameters

Table A1 represents the values that are applied to data in UTM. A False Easting is a linear value applied to the origin of the x-coordinates (i.e. central meridian). A False Northing is a linear value applied to the origin of the y-coordinates (the equator) and the scale factor is a value that is applied to the center point or line of a map projection to reduce the distortion (usually less than one). The other coordinate system that is used to represent data is the Projected Coordinate Systems (PCS).

A3 Projected Coordinate Systems (PCS)

In order to display data taken from the earth, which is in a three-dimensional format in to being able to view this accurately on a two dimensional surface in the form of a map, it is necessary to apply a projected coordinate system (PCS) to ensure the data remains accurate and is distorted the least. However, some map projections minimize distortion in one property at the expense of another, while others strive to balance the overall distortion but sacrifice accuracy taking mean or average values rather than actual. It will depend on the user of the data and the outcomes that are sought after as to how data is projected. Some of the possible reasons for using a PCS are:

- The desire to reuse data in a spatial analyst calculator. Comparatively, Latitude/Longitude is a good system for storing spatial data but not very good for viewing, querying, or analyzing maps. Degrees of latitude and longitude are not consistent units of measure for area, shape, distance, and direction and therefore are not recommended for use in analysis environment.
- For the production of a map in which you want to maintain one of more of these properties: area, shape, distance or direction.
- To achieve a certain output when creating a map on a small scale (e.g. world map).

A PCS is useful when needing to make precise measurements on maps and data should be projected when any calculations are to be performed. There are a number of projection types that will influence area, shape, distance and direction.

A3.1 Projection Types

There are several different types of map projections that can be applied to data. These projections are generally used when the user is trying to preserve a specific spatial attribute of the data (area, shape, distance, and direction), of which the projections will do with accuracy.

- *Equal Area* projections preserve **area**. Many thematic maps use an equal area projection. In the USA, Albers Equal Area Conic projection is widely used.
- *Conformal* projections preserve **shape** and are useful for navigational charts and weather maps. Shape is preserved for small areas, but the shape of a large area such as a continent will be significantly distorted. Common examples of conformal projections are Lambert Conformal Conic and Mercator projections, the later of which will be used throughout this project.
- *Equidistant* projections preserve **distances**, however it should be noted that no projection can preserve distances from all points to all other points. Instead, distance can be held true from one point (or a few points) to all other points or along all meridians or parallels. Equidistant map projection should be used when trying to find features that are within a certain distance of other features.
- *Azimuthal* projections preserve **direction** from one point to all other points. This quality can be combined with equal area, conformal, and equidistant projections, as in the Lambert Equal Area Azimuthal and the Azimuthal Equidistant projections (ESRI, 2003).

Other projections minimize overall distortion but don't preserve shape, area, distance or direction.

A3.2 Australian Geodetic Datum 1966 (AGD66)

The AGD66 is a regional datum, designed to best fit the surface of the Earth over the Australian continent. The Australian National Spheroid (ANS) is used and has a different size and shape to the GRS80 spheroid (Figure 3.9). The AGD66 is located by the coordinates of the Johnston Geodetic Station in the Northern Territory. The coordinates of the Johnston Station have been derived from astronomical observations. The orientation of the AGD66 is defined by the BIH (Bureau International de l'Heure) zero meridian and the Earth's mean axis of rotation at epoch 1962.0. This definition is different from that of the WGS84, therefore, transformations must be performed to convert GPS coordinates to the AGD66 coordinate system. Because of un-modelled systematic error, AGD has distortions of up to several metres. The UTM projection of AGD66 is the Australian Map Grid 1966 (AMG66).

The AGD66 coordinate set is the official basis for the geodetic control network in New South Wales as well as Victoria, Tasmania and the Northern Territory. An approximate transformation model AGD66 to WGS84 has been derived by the Land Information Centre (LIC) (see Table 3.5).

> AGD66 was derived from a Least Squares adjustment of the Australian geodetic network performed in 1966 and was used until the new adjustment was performed in 1984 (AGD84). Because of the limitations of the adjustment model used, and the systematic errors present in some of the observations, the AGD66 coordinate set suffers significant distortions (Rizos, 1999). Allman & Veenstra (1984) compared AGD84 and AGD66, which showed the clearly visible distortions in datasets. The maps of displacement vectors given in Allman & Veenstra, 1984, define the transformation model AGD66 to AGD84. States were reluctant to convert to AGD84 due to these reasons.

A3.3 ISG Integrated Survey Grid

New South Wales adopted a system of "Survey Integration" in the 1970's, where the Integrated Survey Grid (ISG) was introduced to minimise projection corrections for cadastral surveys. It was determined that corrections for a 2° zone were less than 1:8000 and could therefore be ignored. AGD66 remained as the datum. This was useful for local governments and became the standard (Manning, 1992)

Within Shoalhaven City Council and a number of local governments around Australia, ISG has previously been used due to its location specific accuracy. However, Australian standards are trying to convert all data throughout Australia into GDA for standardization purposes. Shoalhaven City Council has the issue of having 95% of the existing data still in ISG projection (see Table 3.6) and this causes problems when trying to upgrade there whole data system into an enterprise wide geodatabase style system, using ArcSDE (see 3.4.1).

A3.4 AMG Australian Mapping Grid

The least squares adjustment of the Australian geodetic network performed in 1966 used the AGD. This adjustment produced a set of coordinates in the form of latitudes and longitudes known as the Australian Geodetic Datum 1966 coordinate set (AGD66).

The grid coordinates derived from a UTM projection of the AGD66 coordinates, using the Australian National Spheroid (ANS), is now known as the Australian Map Grid 1966 coordinate set (AMG66), as represented by table 3.7.

Within the GIS each data set must have its projection defined in order to line up the data in the correct location. If the data is not defined, there will be no spatial overlay between data. In order to make this process in the future a lot less confusing, the Surveyor General of Australia develop a projection standardization method based on geocentric datum's. The conversion to GDA is managed through state agencies such as the NSW Department of Lands (formally as the NSW Land Information Service).

A3.5 Towards GDA94: Geocentric Datum of Australia

Australia began to move from the regional AGD to a geocentric datum known as the Geocentric Datum of Australia 1994 (GDA94) in 2000. This new datum is compatible with the WGS84 datum, which makes it easy for users in Australia to quickly transfer data by directly obtaining GDA94 coordinates from GPS satellite measurements. This process is due to take some time with the amount of data that each regional body has to process. For instance the Shoalhaven City Council has 95% of its data still in ISG, the remaining in AMG as of 2004, 4 years after the suggested date to begin the conversion to GDA. When converting, the only loss of accuracy is due to rounding errors in the calculations (generally <<1mm).

The former Office of Surveyor General (Collier, 2000) published a report containing the details of a number of transformation options for converting between coordinate systems. It also provides definitions and examples of ellipsoidal (spheroidal) coordinates defined by latitude, longitude and height; Cartesian coordinates defined by three X, Y, and Z axes; and grid coordinates defined by east and north.

Converting to GDA (horizontal Datum) caused the following approximate coordinate changes in NSW:

- If using Geographical reference points, then the numerical value of the Latitude will decrease by ~5.6" and Longitude increase by ~4.2"
- If using AMG, then both Eastings and Northings will increase by approximately 105m and 190m respectively (approximate North-East shift of 200m, see figure 3.12)
- If using ISG, then the numbers will be completely different due to different zone widths and different false coordinates for the origin used by the MGA.

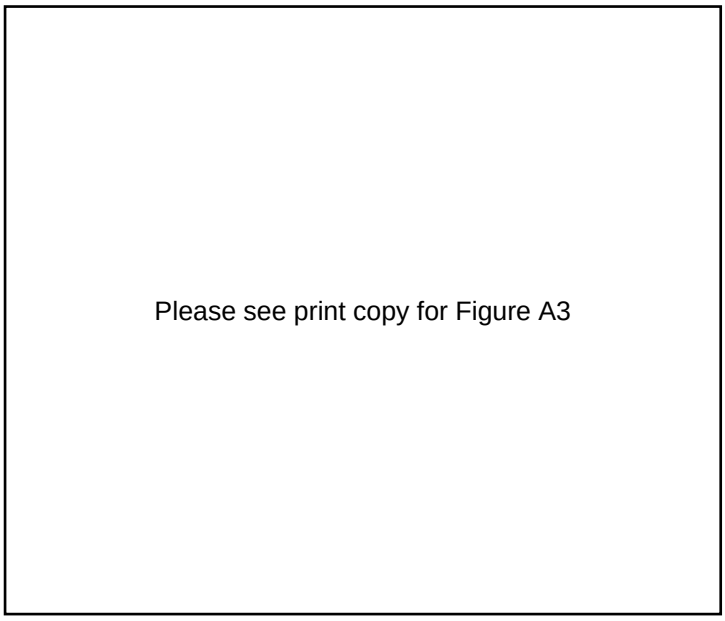

Figure A3 Change in AGD to GDA, in metres (ICSM, 2004)

A3.6 Project Coordinate System

Data for the project was obtained in AMG and ISG under AGD66. The majority of data used from State government departments was under AMG, and the majority of local government data was projected in ISG. For simplicity purposes, the project is projected in AMG to coincide with the large data sets that have been projected in AMG (e.g. ALS Digital Elevation Model) and with the limitations of computer hardware to process the data to GDA94. The predictive model generated (see 6.1.7) will need to be converted to

GDA94 to be useful in future accurate predicting. All ISG data was converted using Geod (Figure 3.13). After the data was converted to the same coordinate system, data analysis becomes easier. Having a standardized dataset eliminates the unnecessary error that could be generated from having data in two different projection systems.

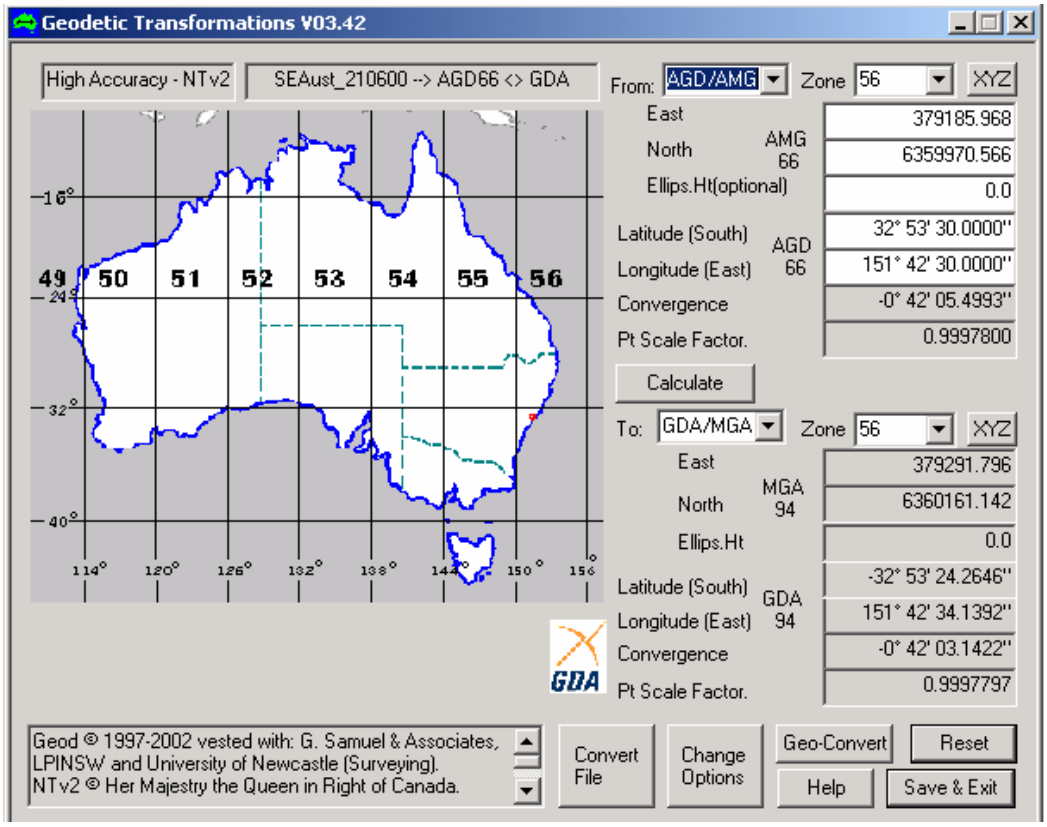

Figure A4 Geod – Geodetic Transformation Tool Version 3.42

A5 Coordinate System References

Australian National Spheroid (ANS)	
Semi-major axis (a):	6,378,160.0 metres
Semi-minor axis (b):	6,356,774.719 metres
Flattening (f):	1/298.25
Johnston Geodetic Station:	S 25° 56' 54.6", E 133° 12' 30.1"
	571.2m (ellipsoid height)

Table A2 AGD origin reference station and ANS parameters

Please see print copy for Table A3

Table A3: Transformation Parameters of relevance to Australian Surveys (Rizos, 1999): In this transformation model only the estimated standard deviations of the parameters are given (shown in brackets).

Ellipsoid (spheroid)	**Australian**
Semi-major axis	a = 6,378,160 m
Flattening	1/f = 298.25
Prime Meridian (Greenwich)	0
Projection	Transverse Mercator
False Easting	300,000 metres
False Northing	5,000,000 metres
Central Meridian	151
Scale Factor	0.99994
Latitude of Origin	0
Unit (meter)	1

Table A4 ISG origin Parameters

Ellipsoid (spheroid)	**Australian**
Semi-major axis	a = 6,378,160 m
Flattening	1/f = 298.25
Prime Meridian (Greenwich)	0
Projection	Transverse Mercator
False Easting	500,000 metres
False Northing	10,000,000 metres
Central Meridian	153
Scale Factor	0.9996
Latitude of Origin	0
Unit (meter)	1

Table A5 AMG origin Parameters

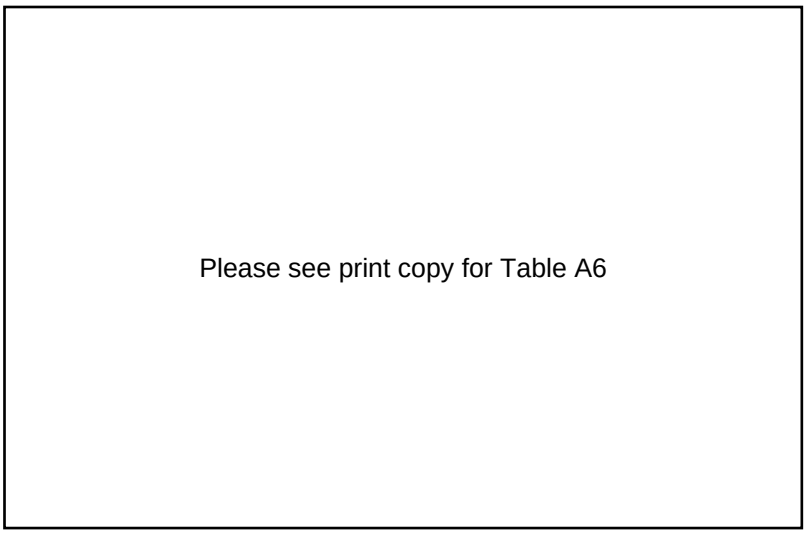

Table A6 GDA Projections (from NSW Department of Lands, 2003)

Parameter	AGD84	AGD66
DX (m)	-117.763	-117.808
DY (m)	-51.510	-51.536
DZ (m)	139.061	137.784
Rx (secs)	-0.292	-0.303
Ry (secs)	-0.443	-0.446
Rz (secs)	-0.277	-0.234
Sc (ppm)	-0.191	-0.290

Table A7 National Parameters – AGD84, AGD66 to GDA94

A6 Model Output

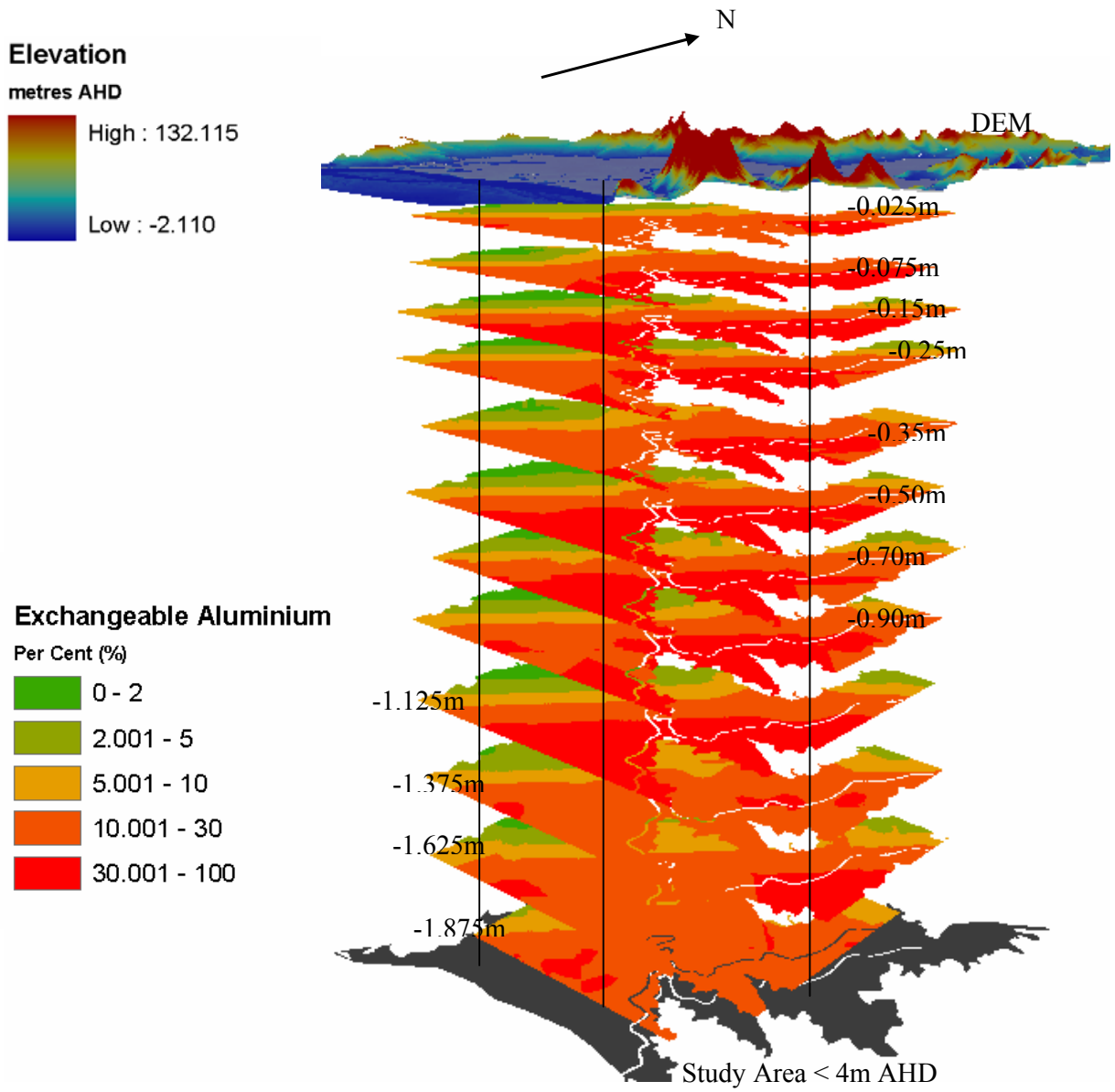

Figure A5 Exchangeable Aluminium Soil Profile – Classified Raster

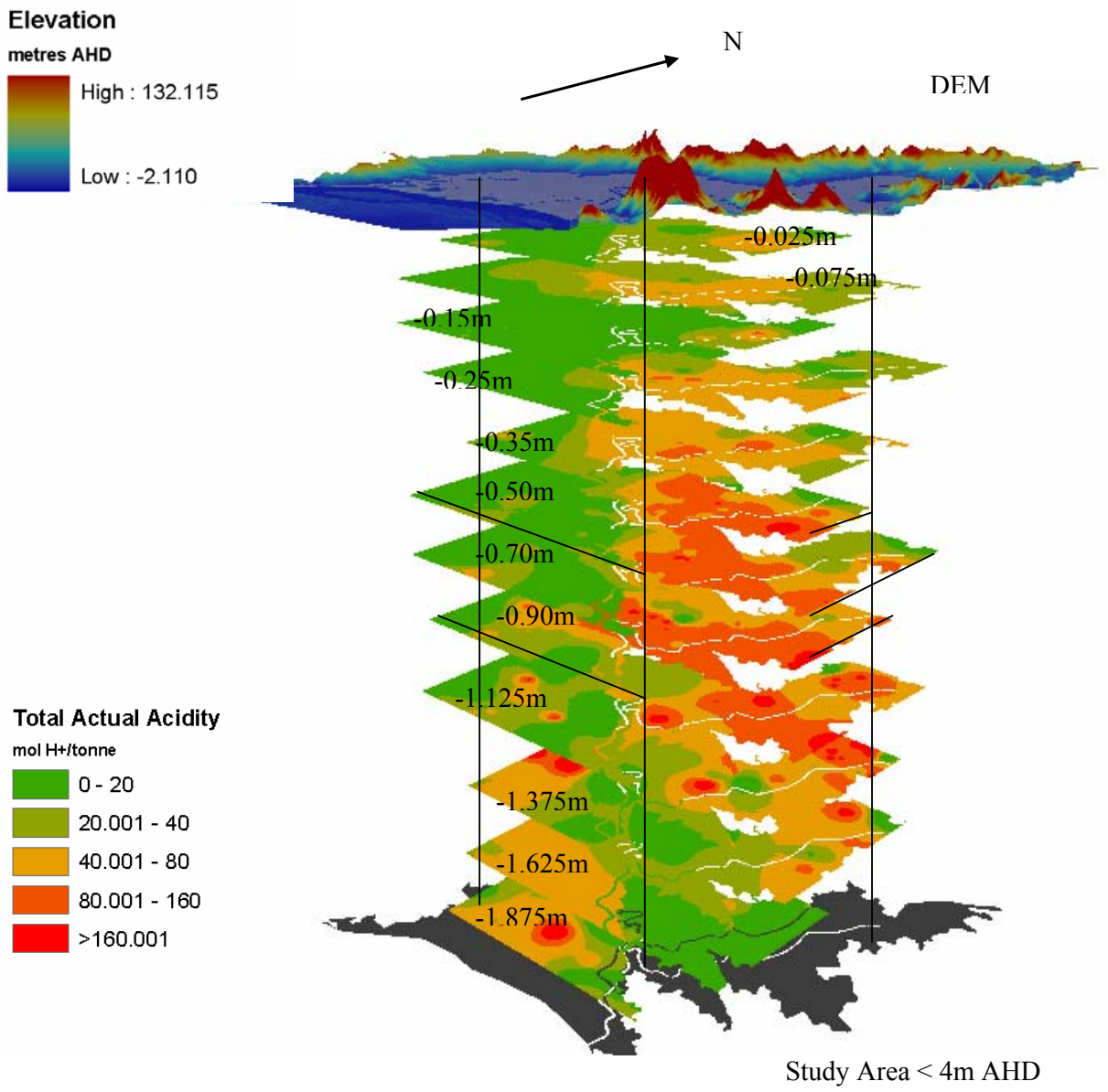

Figure A6 Total Actual Acidity Soil Profile – Classified Raster

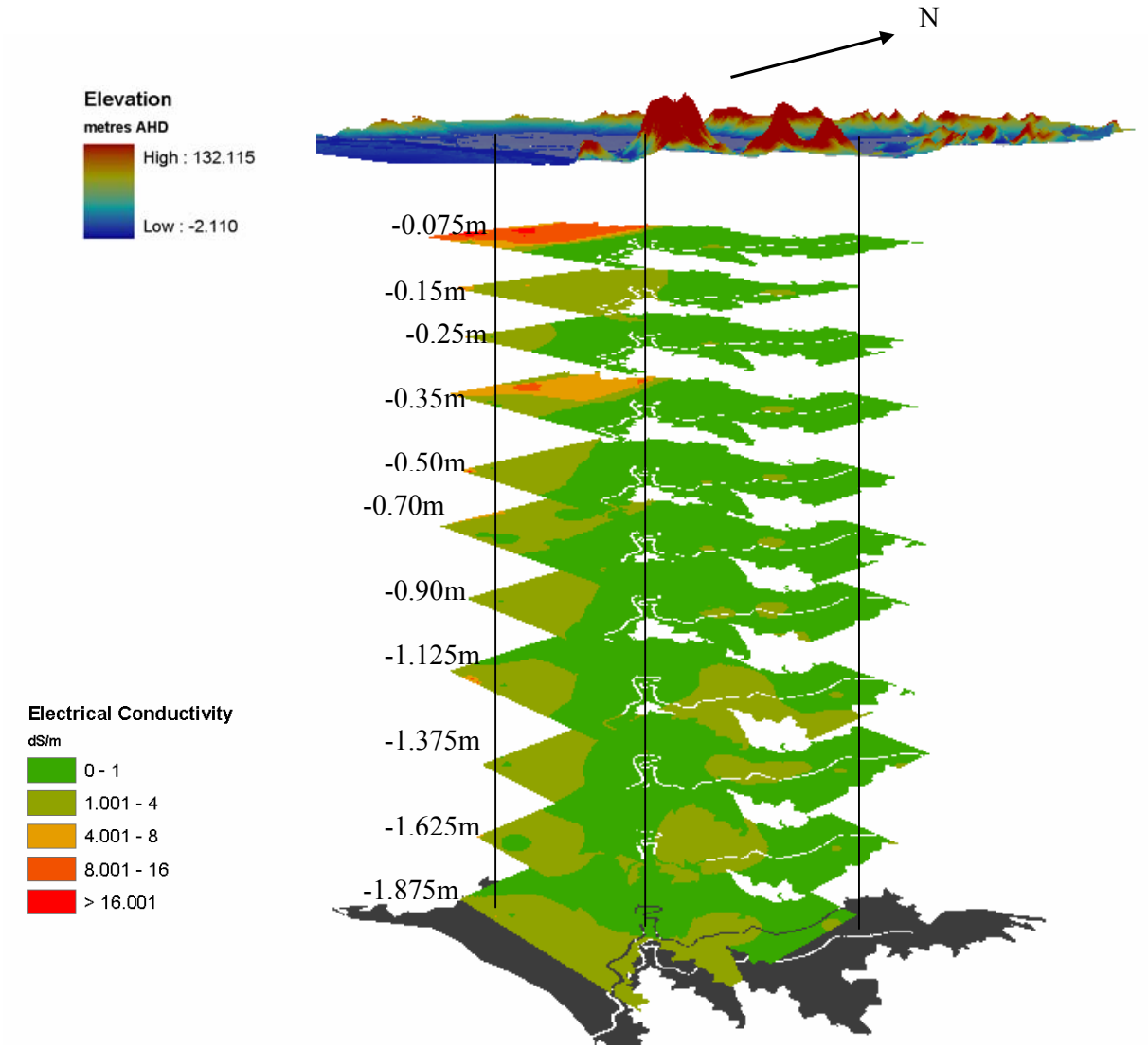

Figure A7 Electrical Conductivity Soil Profile – Classified Raster

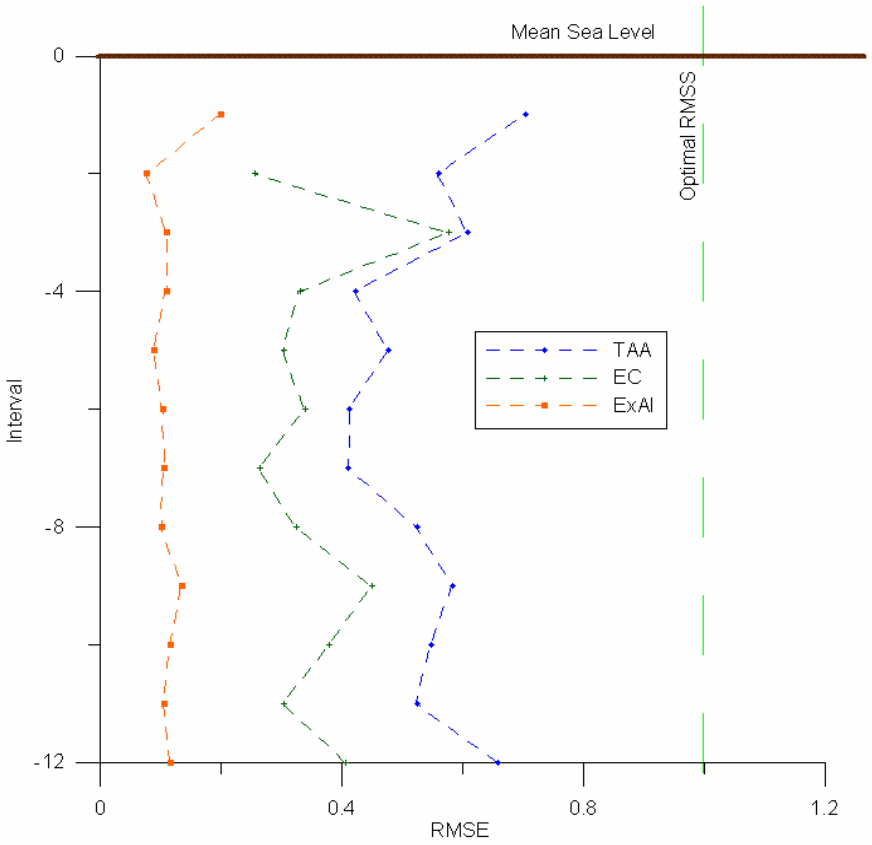

Figure A8 Root Mean Squared Error (RMSE) for the IDW model

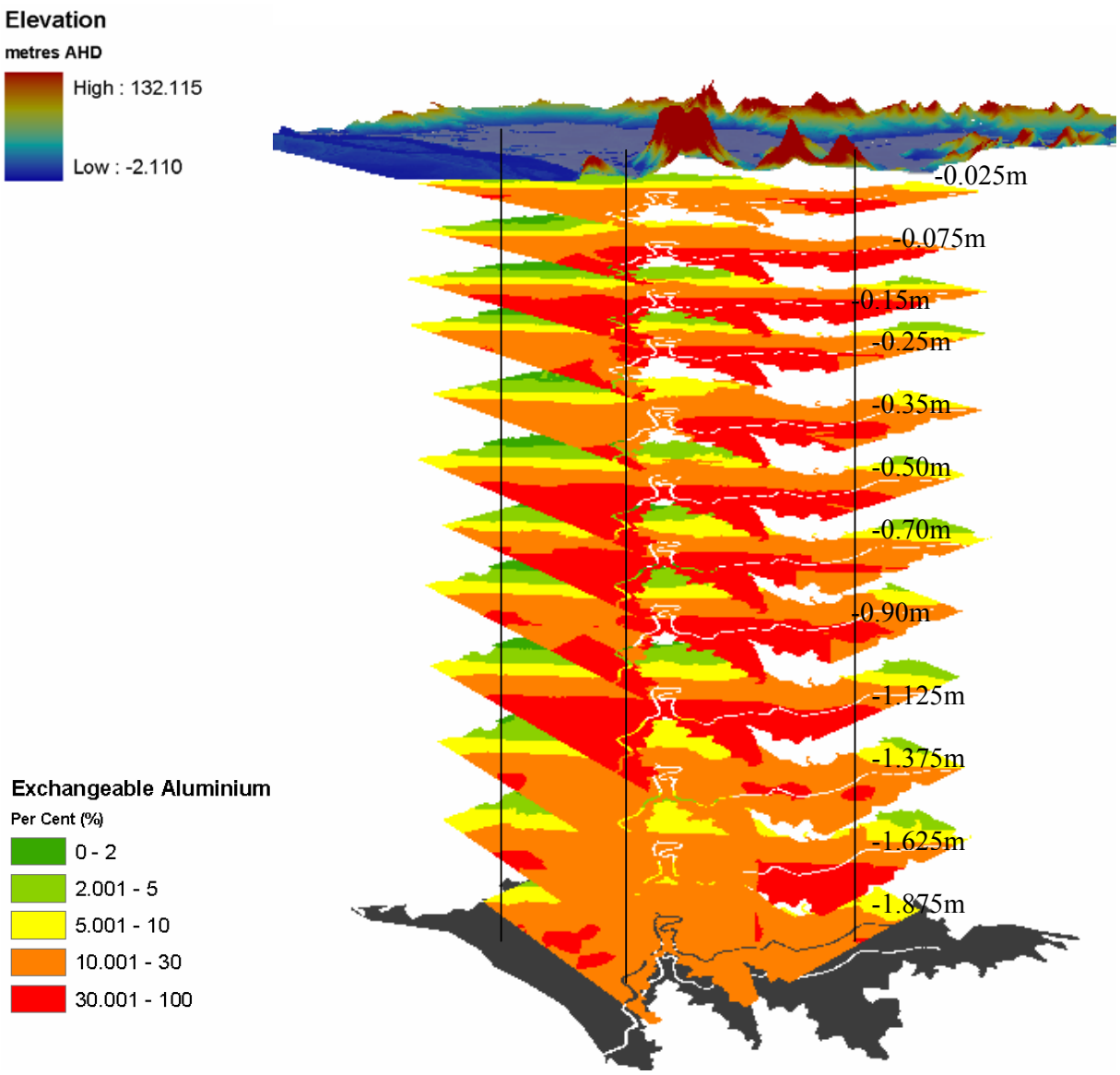

Figure A9 Exchangeable Aluminium Soil Profile – Classified Raster

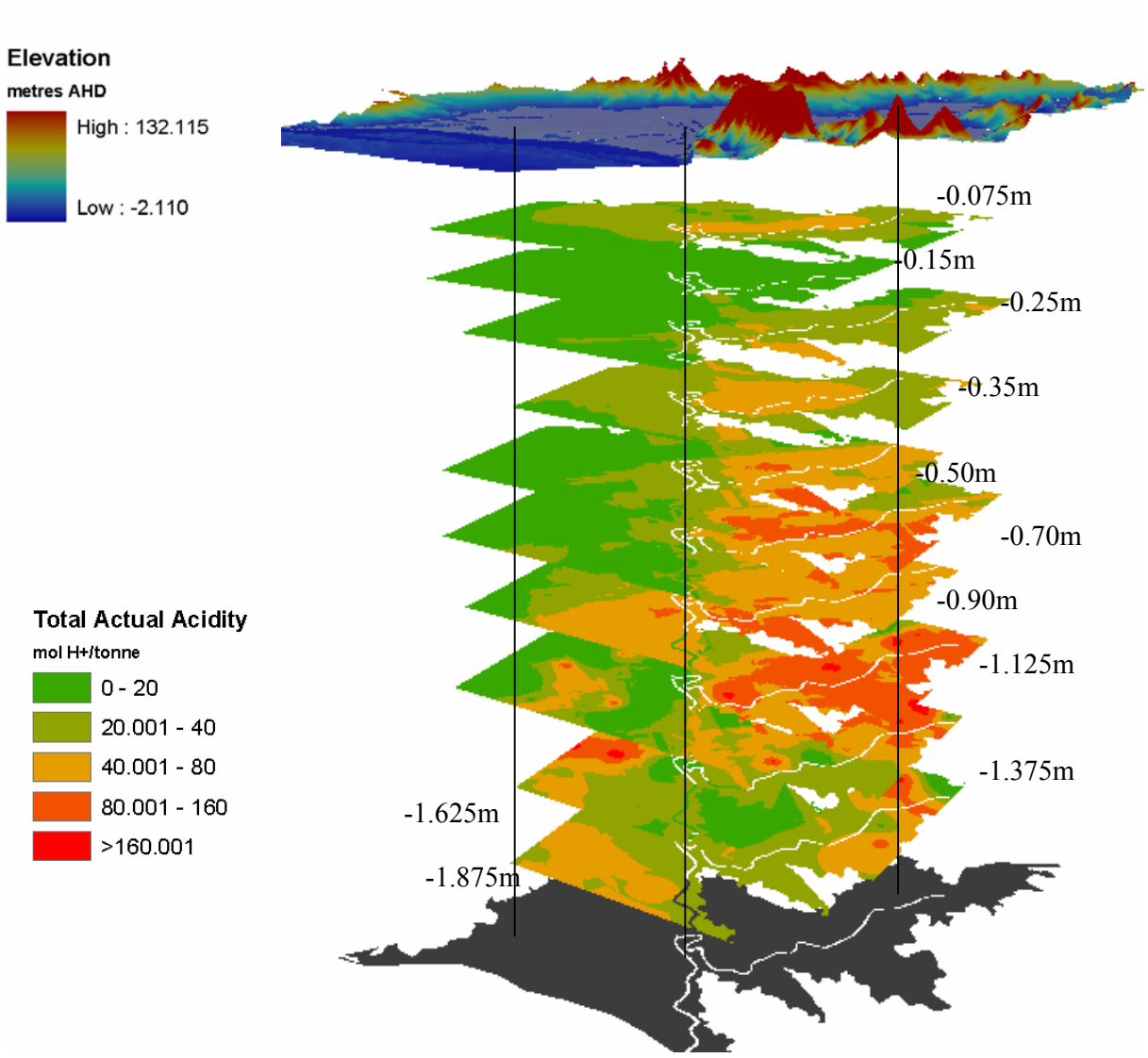

Figure A10 Total Actual Acidity Soil Profile – Classified Raster

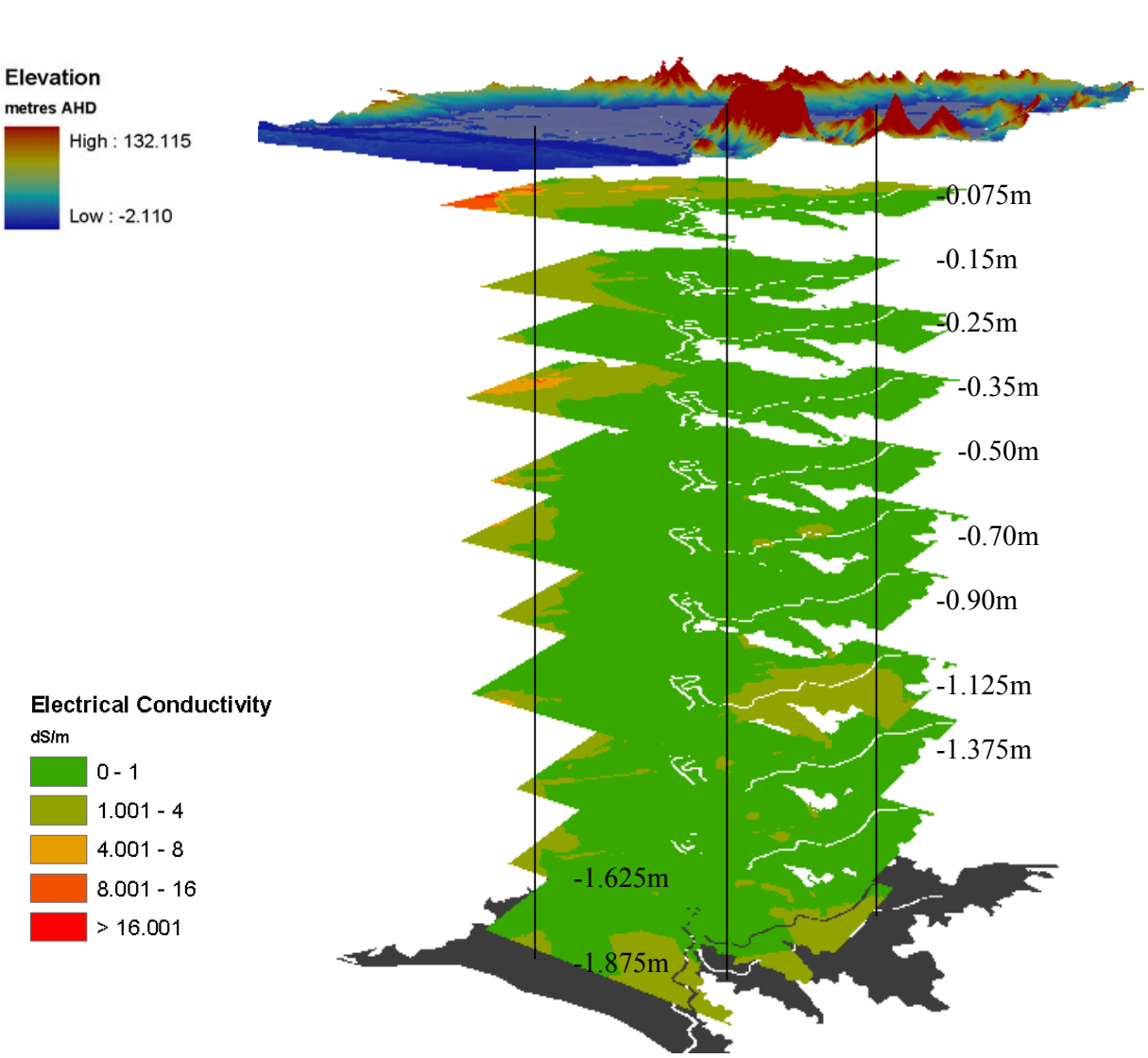

Figure A11 Electrical Conductivity Soil Profile – Classified Raster

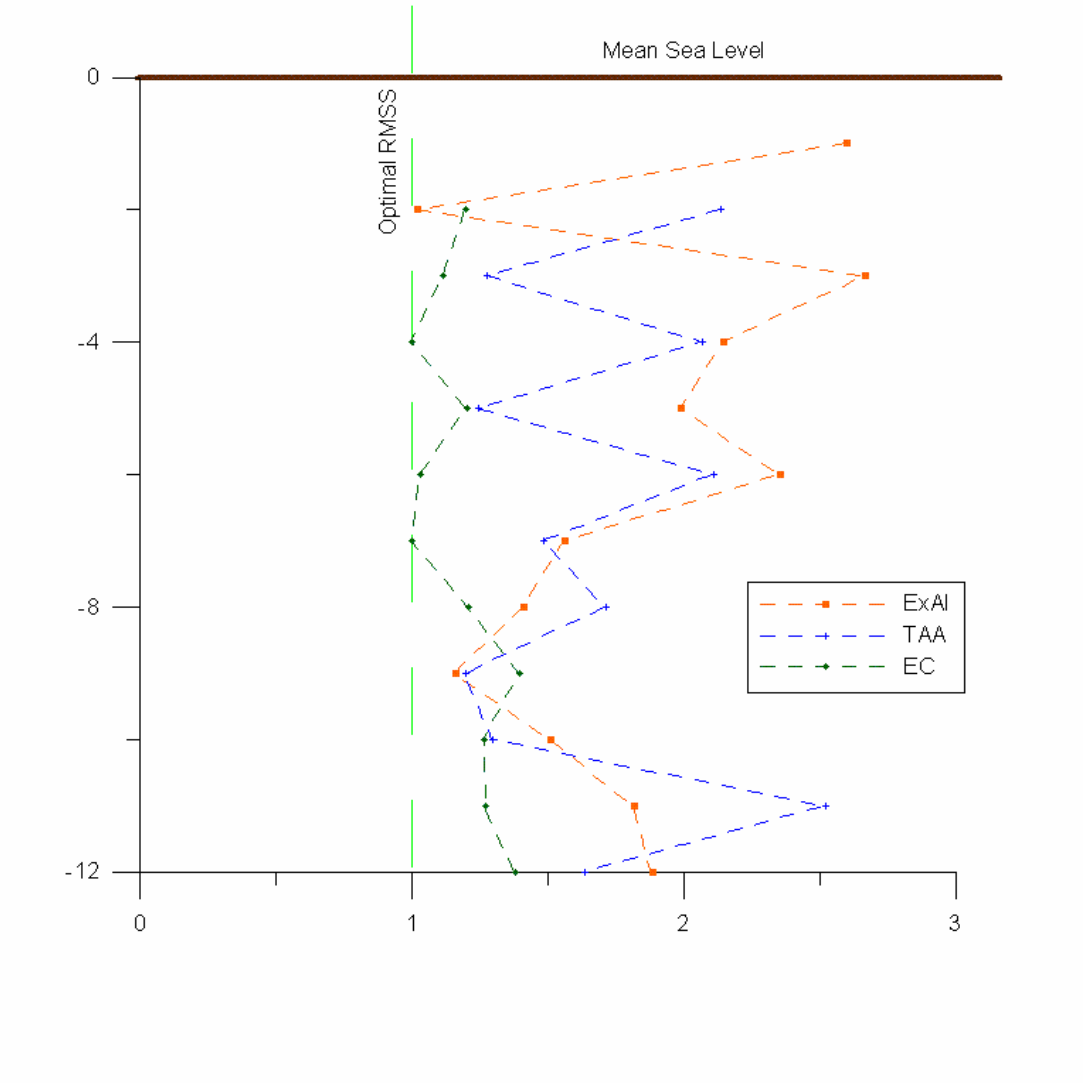

Figure A12 Root Mean Squared Standardised Error for all variables using UK

A13 The Geodatabase

The geodatabase model developed by ESRI in the ArcGIS suite of GIS software defines how a GIS is intertwined with a database management system to organise data. There

are two geodatabases which are used to define a variety of user data models; personal geodatabases and ArcSDE geodatabases that perform different functions.

A geodatabase supports topologically integrated feature classes which can be organised into feature datasets or exist independently in the geodatabase. Feature classes store geographic features represented as points, lines, or polygons, and their attributes. Feature classes can also store annotation and dimensions. All feature classes in a feature dataset share the same coordinate system. Tables may contain additional attributes for a feature class or geographic information such as addresses or x, y, and z coordinates.

Many objects in a geodatabase can be related to each other. For example, tables containing information about the same layer of soil but measure different attributes are related. A relationship class must define the relationships between objects in a geodatabase. This lets you use attributes stored in a related object to symbolize, label, or query a feature class. Feature classes in a feature dataset can be organized into a geometric network. The network combines line and point feature classes to model linear networks and maintains topological relationships between its feature classes. ArcSDE geodatabases also have the ability to contain rasters and geocoding services.

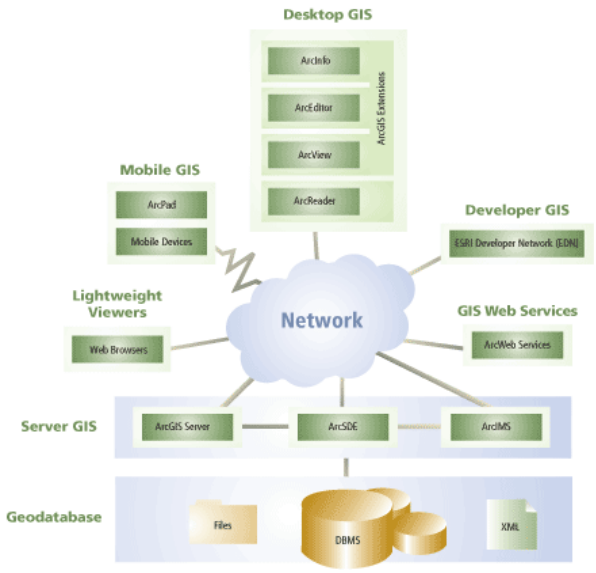

Figure A13 Geodatabase as represented by ArcGIS.

Shoalhaven City Council has employed an organization-wide GIS that includes all aspects in planning and management that can be spatially referenced. For example,

threatened species, soils, water and roads are managed via GIS however, only water, wastewater and REMS (REclaimed water Management Scheme) have been migrated into ArcSDE, running on SQL Server 2000. To be compatible it is necessary to re-design the existing datasets to suit the format of the geodatabase. An initial model has been complete for some of the cadastral layers.