MANAGING DRYLAND SALINISATION WITH AN INTEGRATED EXPERT SYSTEM / GEOGRAPHIC INFORMATION SYSTEM

S.D. Kirkby (B.A Hons.)

Department of Geography
University of Adelaide

Submitted to fulfill the requirement for the degree of
Doctor of Philosophy
1994

Awarded 1995

ABSTRACT

To successfully manage dryland salinisation a coordinated regional management approach requires implementation. The aim of this research is to establish an appropriate interactive land classification methodology that identifies key land areas associated with the problem and then conveys to the end user information regarding the land classification decision making process. An Expert System (ES), a Geographic Information System (GIS), remotely sensed information and a relational database management system (RDBMS) have been utilised to construct the methodology.

Salt Manager represents the software system developed by this thesis to implement the new interactive land classification procedure. It is a system that conveys information regarding the expert's decision to the user in both a graphic and textual format, addresses the problem of data integration and suggests a methodology for dealing with uncertainty. The Salt Manager system has been applied to a salt affected region near Jamestown, South Australia.

The case study highlights the new land classification methodology, the powerful data processing and modelling capabilities and the data acquisition and update facilities. Weaknesses identified during the knowledge elicitation process and the GIS data collection process are addressed by the thesis via the application of machine learning techniques. Resultant methodologies concerning identification of redundant data, which is surplus to the requirements of the classification procedure, and the discovery of new domain knowledge are also considered.

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University of Adelaide Library, being available for loan and photocopying.

Stephen Denis Kirkby

ACKNOWLEDGEMENT

As a meteorite rocketing through the unknown void of research space the author has been fortunate to have been directed and influenced by a host of people. The author owes much to those who have provided him with an opportunity to gain a valuable education in the field of spatial information systems.

I owe an enormous debt of gratitude to Errol Bamford who has constantly identified the "central key thread in my ever evolving web of disorder". His contribution of many excellent ideas which have added to this thesis is greatly appreciated. Thanks Errol, for your patience and guidance. Similarly, Dr. Maureen Longmore's sensitive handling of my professional development and her ability to identify the "big picture" has provided me with invaluable experience. As my initial contact in the Department of Geography she provided the "spark and then maintained the fire of interest" in this research area. The intellectual stimulation provided by the brain-storming sessions which I had with Dr. Peter Eklund can also not be underestimated. He has opened up a pandoras box of exciting research ideas for me. Without his influence the value of interdisciplinary research would not have been so clear.

Special thanks also go to Frank Kurzel and Ahmed Salim, my two co-research students from the Department of Computer Science. Without their help I would probably never have completed the computer programming necessary for this research. It was a great opportunity to work in a team environment. Thanks guys.

I also wish to express my appreciation of the help provided by Cameron Humpheries from the Department of Computer Science and Jason Trenouth from Harlequin Ltd. Cameron's highly refined expertise as a systems manager enabled me to avoid the traps that await young players in this field. Similarly Jason's patience with my programming skills was greatly appreciated.

Many thanks are given to Dr. Dean Djokic at the Centre for Water Research, Department of Civil Engineering at the University of Texas Austin. I travelled to Austin to collaborate with Dean after 20 months of my candidature. He provided me with a number of excellent ideas which formed the basis for the interface between the different software systems.

The assistance and expertise provided by Chris Henschke from the South Australian Department of Primary Industries has been also invaluable. His ability to convey detailed knowledge of the study area and the salinisation problem was beneficial to the project. I express similar thanks to the other experts involved in the thesis.

An expression of gratitude must also be made to the National Wilderness Inventory group ("the militants") who, as the other G.I.S. group in the Department, have provided many stimulating discussions. Thanks Rob Lesslie, Danielle Taylor and Mike Maslen. In the same breath I must also thank my fellow Postgraduate students, in both the Department of Geography and the Department of Computer Science. The extensive discussions I had with Peter Deer from D.S.T.O. have also benefited me enormously.

I also express many thanks to Sue Murray, a cartographer in the Department of Geography, for the drawing, reproduction and layout of the diagrams within my thesis. Similar thanks go to Joanna Gagliardi for her advice on word processing techniques, formatting and cooking. Margaret Bruce must also be thanked for keeping a tight control of my research funding.

The influence of Professor Graeme Hugo must also be gratefully acknowledged. Professor Hugo provided the initial monetary and Departmental support for the project, without his backing the thesis would not have been completed on time. His attitudes and approach to research have also made an impression on me.

A very special thanks is deserved by Petra Heeg who is blessed with special talents. Her perceptive insights into my problems have always provided me with clear and concise answers. Without her support, warmth, love and friendship I would never have been in the position to complete this thesis.

Finally thanks must be given to Russell Flavel and Kym Perkin from the G.I.S. Group with the South Australian Department of Primary Industries. They provided excellent support with the development of the project, thereby formulating a strong working relationship between our respective Departments. Their open minded approach to research collaboration has taught me valuable lessons.

Table of Contents

Chapter 1

Chapter 2

Chapter 3

List of Figures

Appendices

List of Tables

Appendices

Chapter 1

INTRODUCTION

Currently, 300,000 hectares of previously productive non-irrigated soil in South Australia is too saline to permit growth of the usual crops and pastures (Policy Development Planning (PDP), 1992). The land degradation problem of dryland salinisation is considered to be the worst facing both South Australia and Australia. The problem is not however unique, other countries, such as the USA, Canada, Iran, Turkey and Latin America, are experiencing similar disturbing trends (Williamson, 1978).

Notably, the estimate of 300,000 hectares represents a significant increase from the estimates of 225,000 hectares in 1990 and 55,000 hectares in 1982 (State Dryland Salinity Committee, (SDSC), 1990). The dramatic differences in these estimates give rise to the assumption that the area affected by dryland salinisation is increasing and/or that the previous methods of recording salt affected land may have substantially underestimated the magnitude of the problem. Nevertheless, regardless of the exact figure, the trend indicates serious problems for land owners and rural societies in the near future.

Economic losses, stemming from the reduction in productive land due to salt affected soils, have currently been estimated at A$27M dollars per year (SDSC, 1990). With projections that at least a third of the Murray-Darling basin in South Australia will be lost to secondary salinisation within a 30 year period, further production and therefore economic losses will occur (Barnett, 1990). The long term economic losses are difficult to quantify.

Additionally, the social costs of this problem, e.g. how the loss of income in a region affects the generic small country town, have not been evaluated. As salinisation occurs on a regional scale, the loss of productive land for whole communities will eventually have devastating effects on local economies. A national review of salinity research released in December 1992 noted, "of the 126 projects in the dryland salinity area funded by national agencies, only three appeared to have a socio-economic focus ..... This neglect seems all the more remarkable when it is noted that all available evidence shows that the major determinant of success in managing dryland salinity will be landholder response to the problem (adoption of strategies)" (PDP, 1992, p. 6). From this comment it is clear that a key ingredient in the problem solution is the motivation of the landowner and central to this concept is the provision of information.

Dryland salinisation is a spatial phenomenon, i.e. the cause and subsequent effects occur within the spatial landscape. Logically, techniques best suited to determine the extent of salinisation will therefore employ spatial concepts. Two core techniques which deal with the acquisition and analysis of spatial information are remote sensing techniques and Geographic Information Systems (GIS).

GIS and remote sensing techniques developed separately but, over time, due to the logical compatibility of the two, they have been utilised together. Estes and Holz (1985) state that remote sensing in combination with the ever increasing functionality of GIS had "come of age". Recent literature from the GIS domain suggests that GIS is also moving towards a more advanced phase as it shifts from the initial data standard/formulation/acquisition phase to the analytical, data manipulation phase (Goodchild, 1993). An important component within this transition phase is the use of other analytical techniques such as Expert Systems (ES) (Skidmore, 1990).

It is proposed by this thesis that the integration of software systems such as ES with GIS systems, which contain remotely sensed and other environmental information, will provide an advanced analytical method largely independent of traditional techniques. Moreover, the use of other facilities provided by the ES will allow the decision making process to be conveyed to the end user. These techniques will be applied to the land degradation problem of dryland salinisation.

1.1 MOTIVATION

Reflection on the author's previous research conducted in this area provides an insight into the approach adopted for the current research. In 1990, the author conducted a study to determine the physical changes in a salt affected groundwater discharge zone using remotely sensed satellite imagery (SPOT 1), historical aerial photographs and a vector based GIS. Changes in the spatial extent of the salinisation were detected by registering photo-interpreted aerial photographs to a common scale within the GIS and subsequently initiating areal analysis. To determine the ability of the SPOT 1 data to discriminate salinisation a multivariate classification procedure was implemented. At that stage, software linking the output of the image processing system and the vector based GIS required development.[1] Although identifying 84.5% of the salt affected soils, the classification procedure would have been more successful if other environmental information had been included. Also, other crucial information, such as areas of groundwater recharge, could not be determined.

With the realisation that groundwater discharge (where salts rise through the soil profile to the surface) is only the expression of the problem occurring, it was decided to attempt to identify both major components of the salinisation process: recharge and discharge areas. Because the interaction of recharge and discharge is complicated, a flexible approach that simplifies relationships between different environmental datasets is proposed.

In the formulation of the methodology for this thesis two central issues were considered: firstly, the need to identify each component of the salinisation problem and secondly, the importance of conveying information regarding the land

---

[1] Gary Archer from the South Australian Department of Woods and Forests initiated the software development.

classification process to the end user, the landowner. *Salt Manager* is the system developed to satisfy these two major concerns.

Salt Manager will use heuristic knowledge elicited from dryland salinisation experts to analyse 13 environmental parameters contained within a GIS and a Relational Database Management System (RDBMS). The environmental data layers will be soil, geology, slope, elevation, depth to groundwater, electrical conductivity and a 7 band Landsat TM image. The result of the analysis will be two land classification maps, one indicating potential groundwater recharge and the other potential and existing groundwater discharge.

Using the functionality of the *Salt Manager* system the landowner will be able to interrogate the decision making process that has led to either of the two land classification maps. Information regarding the decision making process will be presented via the computer screen to the user in three ways: in a textual format, documenting the rule that led to the classification result; in a graphic format, displaying a scanned photograph of a representative area; and in the original map format. If the user disagrees with any of the information presented, they will be able to use the updating facilities provided by the *Salt Manager* to alter the relevant environmental information within the GIS and the RDBMS.

The updating facilities will be important, particularly for the environmental parameter "depth to groundwater". Groundwater will be an influential parameter in the assessment of potential groundwater discharge. As it is a dynamic parameter it will be important to collect the information regularly. By providing an updating facility it is expected that landowners who own land in "risk" areas will be likely to enter the "depth to groundwater" data from their properties. The assumption is that those landowners who enter data into the *Salt Manager* system are more likely to have confidence in the land classification result presented to them; especially as the interrogation facilities of the system will allow them to gain an understanding of the decision making process. This principle of conveying understanding regarding the decision making process will also apply to those landowners who do not enter data into the system.

Machine learning techniques will be applied in this thesis to decipher the completeness of the knowledge base and to improve system efficiency. Given that the GIS and the RDBMS will contain a large amount of data, the machine learning techniques will be also used to isolate relationships between the classification results and the environmental data parameters used in the decision making process.

1.2 STRUCTURE OF THESIS

To describe the proposed methodology this thesis will be divided into the following 9 chapters.

Chapter 2 provides the necessary background information regarding the conceptual approach adopted by the thesis. Specifically, the chapter considers the nature of salinity and the major factors that lead to the formulation of the problem. Traditional identification techniques will be considered along with the relevant management options.

Chapter 3 will briefly review remote sensing and GIS techniques and will then consider the approaches implemented to identify salinisation using these techniques.

Chapter 4 gives consideration to traditional land classification techniques and proposes, using analytical techniques such as ES, that a new land classification procedure can be implemented.

Chapter 5 is the first of the sections describing the methodology and results. Information utilised for the analytical procedure is documented. Specifically, this chapter considers the acquisition of field data for the study area.

Chapter 6 gives consideration to the design of the geographic database. The chapter is important as it justifies the data structures employed and provides an insight into the variability associated with spatial environmental data. A discussion of the computer software package utilised to model the spatial data is provided in Appendix 1.

Chapter 7 documents the knowledge acquisition phase. In order to use heuristic information to analyse the spatial information, the initial heuristic knowledge must first be acquired and then formulated into a logical system. There are three components associated with this phase: the initial interviewing process and rule base formulation; the evaluation of the rules using artificial intelligence techniques; and conveying the (un)certainty inherent in the decision making process.

Chapter 8 presents the software system, *Salt Manager*, developed by the thesis and provides an insight into the functionality of the methodology. The chapter is designed to provide an overview from the user's perspective describing each interactive function. The *Salt Manager* system is the result of the implementation of the proposed methodology.

Chapter 9 verifies and validates the results produced by the *Salt Manager* system. It compares the results obtained from field data collected in Chapter 5 to those derived from the *Salt Manager* system. An additional experiment, using machine learning techniques to identify relationships between environmental data and the final classification task, is also considered.

In Chapter 10 a discussion evaluating the results of the application of the methodology is carried out. A new land classification procedure is documented along with comments concerning technical issues, such as dealing with spatial error and verifying (un)certainty in the *Salt Manager* system. The chapter concludes by considering the major issues addressed by the thesis.

Chapter 2

THE NATURE OF SALINITY

Bell (1989) notes that soil salinisation in irrigated soils has been recognised by humans for thousands of years and claims that this problem probably contributed to the breakdown of Sumerian civilisation around 2500 BC. In the last 100 years, human induced salinity in non-irrigated areas has also become evident. Intensive research into salinisation in Australia commenced in the 1960s and now, in the early 1990s, a great deal of qualitative knowledge has been assembled regarding the problem.

The role of this chapter is to consider the qualitative information and ascertain why, after 30 years research, an empirical model explaining salinisation processes has not materialised. The chapter will first define salinity, then identify the causes and mechanisms which create saline seepage areas and finally consider management strategies.

2.1 WHAT IS SALINITY?

Salinity is related to the content of salts within soil and water bodies. These salts, when dissolved in water, disassociate into the common cations and anions listed in Table 2.1.

TABLE 2.1: Salinity Ions

Cations	Anions
sodium	chlorine
potassium	carbonate
calcium	bicarbonate
magnesium	nitrate
	sulphate

Source: Robbins et al., 1991.

The first two, sodium and chloride, account for 50 to 80 per cent of the total salt content in most of the Australian landscape (Williamson, 1990a).

Historically water salinity measurements were expressed as the total dissolved solids (TDS) (Bresler et al., 1982). Reflecting a change to summation of the individual ionic components, the term total soluble salts (TSS)(the total content of salts dissolved in water) is now utilised (Williamson, 1990a). As a simplified index to gauge TSS and therefore salinity in water, Scofield (1942) advocated the use of electrical conductivity (EC).

In North America, EC values are used to define soils that are salt affected but in Australia, due to the relatively high concentrations of sodium and chloride ions and hence extreme EC values, this is not common practice (Dowie, 1986). Instead, soil salinity is defined as the weight of soluble salts in a unit weight of dry soil. Surface soils are said to be saline when the sodium chloride content exceeds 0.1% for loams and more coarse textured soil, or 0.2% for clay loams and soils of finer texture (Peck, 1983).

2.1.1 The Origin of Salts in Saline Soils

Australian soils contain high quantities of stored salts, measured in the vicinity of 200 to more than 1000 tonnes per hectare (Williamson, 1990b). These salts may originate from the ocean via rainfall, weathering of soil and rock minerals and marine deposition from earlier geological periods (Williamson, 1990a). In southern Australia the atmospheric fall-out in the form of rain is considered to be the most important salt source (Bettenay et al., 1964; Peck and Hurle, 1973). Williamson (1990a) supports this notion by stating that the input via rainfall has been measured at 300 kg/ha.y^{-1} near the coast, approximately 30 kg/ha.y^{-1} 250 km inland, and 15 kg/ha.y^{-1} more than 600 km inland.

Considering that the annual rate of removal in even severely degraded catchments is often much less than 1% of the stored salt in the profile, the solution to the salinity problem lies, not with controlling the salt storage, but rather by understanding the mechanisms that move water through the landscape (Williamson, 1990b).

2.1.2 Movement of Salts in the Landscape

Soluble salts that are brought to the landscape are predominantly transported and redistributed in water. The water movement occurs as a result of:

1) pressure differences, due to water flowing from high points to low points in response to gravity;
2) moisture differences, where water moves via matric suction from wetter to dryer parts of the soil; and
3) humidity differences in the soil where water moves through the profile as vapour.

Pressure differences account for the greatest volume of water movement and the principal mechanisms are infiltration, deep percolation and/or surface run-off.

Infiltration is the term applied to the process where rainfall enters into the soil generally by downward flow through all or part of the soil surface (Hillel, 1980). This process accounts for a large proportion of salt movement as the salts stored in the soil profile are dissolved and mobilised by infiltrating water. The infiltrability of the soil depends upon:

1) time from the onset of rain (Infiltration rate is apt to be relatively high at first, then decreases to eventually approach a constant rate that is characteristic for the soil profile);
2) initial water content (The wetter the soil is initially, the lower the initial infiltrability owing to smaller suction gradients);
3) hydraulic conductivity (The higher the hydraulic conductivity the higher the infiltrability);
4) soil surface conditions. (The variability of the surface soil conditions ensures that some areas allow higher infiltration rates than others, e.g. areas with sandy soils, which have a coarse texture readily absorb water and allow deep percolation. Similarly, layered soils which have strata of varying permeability allow water movement more readily along those strata that provide the least resistance).

Another mechanism of high infiltration is deep percolation via macropores, holes in the soil created by root channels, soil cracks and insect activity (Nulsen, 1980). These areas are termed preferred pathways (Nulsen and Henschke, 1981). The soil water not utilised by plants and drains beyond the root zone continues downwards, as deep percolation, into the groundwater (Hillel, 1980).

High infiltration rates caused by surface run-off occur when the rainfall rate exceeds both the rate of infiltration and the surface storage capacity of the soil resulting in the excess water moving downslope (Cooke and Willatt, 1983). Salts transported by this mechanism are normally low in concentration except where the run-off encounters saline surfaces.

Where pressure differences are absent water can still move through the profile, as noted previously, by either moisture or humidity differences. Capillary rise, an agent in either of these two categories, occurs where the water table rises close enough to the soil surface for upward water movement to occur under the influence of capillary action. The critical depth to water table before capillary rise occurs is regarded to be about 1.5-1.8m depending on the soil texture (Hillel, 1980).

2.1.3 Types of Soil Salinisation

Eight different types of soil salinity are recognised in Australia (Table 2.2). This thesis is primarily concerned with a form of human-induced 'secondary' salinity; a term first used by Northcote and Skene (1972) to identify areas where human activities have resulted in adverse changes in the salinity of soil or water. There are two major types of secondary soil salinity; salt scalds (Type 5, Table 2.2) and saline seepage (Type 4, Table 2.2).

TABLE 2.2: Categories Describing Types of Soil Salinisation

Code	Classification	Description and Extent in South Australia
Type 1	Salt Marsh	Coastal land with water table, some 100,000 ha. chiefly around eastern Spencer Gulf.
Type 2	Salt Pan	Big inland Salt Lakes, e.g. Lake Eyre and Lake Torrens. 1.8 m hectares.
Type 3	Salt Flats	Land with water table below the surface. Naturally saline in subsoil or throughout, and carrying salt tolerant plants; 50,000 ha, chiefly east of Wilmington and Quorn.
Type 4.1	Saline Seepage (formerly 'Saline Seeps')	Seepage with water table within capillary range, resulting in a saline soil throughout the profile.
Type 4.2	Saline Seepage	Seepage with water table below the capillary range, resulting in a saline subsoil.
Type 5	Dry Saline Lands • Saline Loam (formerly Salt Scalds)	Land without water table, naturally saline in the subsoil or throughout, and with uniform or graditional light textured soils in areas with less than 375 mm annual rainfall. Extent very limited.
Type 6	Dry Saline Lands • Saline Clays	Land without water table, naturally saline in the subsoil or throughout, less than 300 mm rainfall; at least 5 m hectares adjacent to the N.T. border.
Type 7	Scalds	Where wind or water erosion has removed topsoil - common in pastoral areas, 1.2 m hectares. This is the focus of much of the contour furrowing and disc pitting work to aid revegetation in the Flinders Ranges and North East Pastoral Areas.

Source: Matheson, 1984, p. 2.

Salt scalds are formed by the erosion of topsoils to expose saline subsoils (Hill, 1988). The development of this feature occurs when small amounts of salt in solution enter the hydrological system each time it rains. Due to flat topography (little run-off) and high evaporation rates this salt remains in the sandy loam soil (Hartley, 1984). With future rain the salt is leached through to the impermeable clay where it accumulates. The vegetation, provided that its root zone is above the clay, is unaffected but once this cover is degraded by human interference (e.g. through overgrazing) the sandy loam is then prone to erosion resulting in a bare soil area with high levels of salt concentrations.

Saline seepage, which is the focus of this thesis, usually occurs in areas of high agricultural productivity where the land has been cleared (Hartley, 1984). Saline seepage is defined by Brown et al. (1976) as:

> " ..... a recently developed wet salty area in non-irrigated soil on which crop production is reduced or eliminated. The soil surface is intermittently or continuously wet and white salt crusts are often present."

2.2 THE CAUSES AND MECHNAISMS WHICH CREATE SALINE SEEPAGE IN AUSTRALIA

Saline seepage was first noted in South Australia in 1900 (Matheson, 1984) with other interesting observations noted by Perkins (1927, 1930). An early theory developed in Australia to explain the occurrence of increasing saline seepage areas was formulated by Wood (1924). He observed that, once the native vegetation was destroyed for agricultural purposes, there was a subsequent increase in salinity levels of the streams in the area. From this observation Wood hypothesized that the increase in salt levels was due to the rising water table level which brought salt to the soil surface (Wood, 1924).

Wood's enlightened hypothesis identified the basic cause of secondary salinisation, i.e. the removal of high water usage native trees and their subsequent replacement by crops and pastures which fail to use all of the water that soaks into the land after rainfall. Thus, soils remained wetter for longer periods, allowing water to percolate to the underlying groundwater table more frequently (Dyson, 1990).

To conceptualise this process Nulsen developed a simplified hydrological model (Nulsen, 1978). The model consists of three components:

1) an intake, or recharge area, where water enters the system;
2) a flow path of continuous water movement from intake to outlet (consisting of 3 component paths):
    a) vertical downwards (at intake area);
    b) horizontal; and
    c) vertical upwards (at outlet area); and
3) an outlet, or discharge area, where water leaves the system.

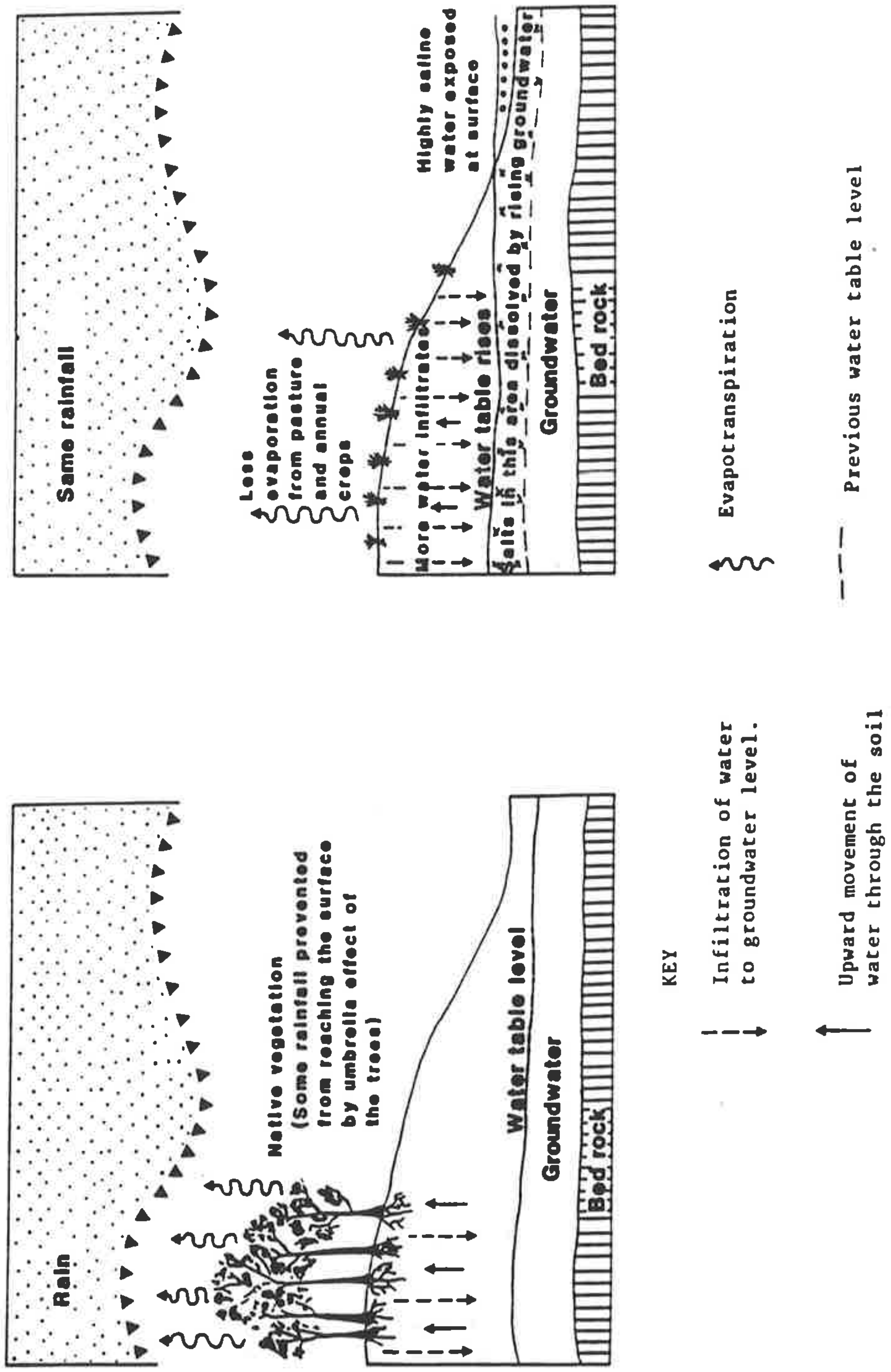
An important assumption of the model is that all components are inter-related, therefore, an increase in groundwater recharge will see a subsequent increase in groundwater discharge. Consideration of these three components follows.

2.2.1 Groundwater Recharge Areas

Before land is cleared a permeable topsoil exists, as a result of organic matter addition and micro-faunal activity, which allows water to infiltrate into the soil reducing run-off. Upon entering the topsoil, the water is drawn by matric suction into the subsoil where the soil moisture level is maintained in a relatively dry state by the constant uptake of water by the native vegetation, ensuring percolation of water downwards beyond the root zone is minimised. The low run off and low percolation in this situation leads to a relatively deep water table (Figure 2.1).

When land is cleared two hydrological changes occur: an increase in run off due to a decreased infiltration rate and an increase in percolation. The increase in percolation results from the change in vegetation cover. In an uncleared state, the perennial deep-rooted native vegetation communities transpire and utilise enough groundwater to maintain a hydrological balance (Greenwood et al., 1985; Borg et al., 1988).

FIGURE 2.1: How Clearance Causes Salinity Problems



Source: (Hartley, 1984, p. 12.

When land is cleared the native vegetation is replaced by annual crops and pastures which have comparatively shallow roots and transpire, although at variable rates, less moisture (McCarthy, 1990; Sharma et al., 1991). The loss of the native vegetation therefore results in a decrease of evapotranspiration and a subsequent increase in the amount of water available for groundwater recharge (via percolation) (Figure 2.1). As an example, estimates of recharge for the South Australian mallee will rise from 0.1mm.y^{-1}, under native mallee vegetation, to 1 to 30mm.y^{-1} under cropping and grazing (Walker, 1990).

This increase in groundwater recharge can result in a rise over time of the groundwater table; for instance Williamson and Bettenay (1980) measured a rise of 1.5m.y^{-1} after clearing a catchment near Bakers Hill in Western Australia. The rate of groundwater change varies according to the characteristics of the particular catchment. Hookey (1987) ascertained groundwater recharge moved quickly through the hydrological system to the groundwater discharge area in the Wights Western Australia catchment; but on a similar catchment in the same area, Hookey calculated it took 12 years for the groundwater recharge to move through the hydrological system (Hookey, 1987). The reason for this lag period relates to the preferred areas of recharge.

Some groundwater recharge is likely to occur over the whole area of a catchment, however the volume of recharge per unit area will vary greatly in both time and space (Morris and Thomson, 1983). Dyson's (1983) and Cook et al.'s (1989) research both support this claim. Dyson identified 30% of the Axe Creek catchment in Bendigo Victoria which supplied 83% of the groundwater recharge, whilst Cook et al. (1989) identified 20% of the Borrika catchment in South Australia which supplied 41% of total recharge. These preferred zones of recharge are due to the soil, geologic, land cover and topographic characteristics of the catchment (Jolly, 1988; 1989).

As already considered, the unique characteristics of each soil type influences infiltration rates. This principle of variability also applies to the geology of an area. Paleochannels of various sizes (ranging from 2 to 20 metres thick) at various widths and depths, and in various positions in the landscape (longitudinal in the valley to perpendicular) support both flow and reservoirs in the landscape (Williamson, 1990b). Similarly, recent research indicates that the zone of weathering immediately above the fresh basement rock appears to have a higher (by one or two orders of magnitude) hydraulic conductivity than the pallid and mottled clays above (Williamson, 1990b). The heterogeneous nature of geology, i.e. the presence of quartzitic veins, bedding planes and other structures, produces zones where preferred flow can occur.

The types of land cover, as briefly considered, also influence the amount of water available for recharge. The principal influence over land cover is the land manager. If they allow overgrazing to occur on recharge zones the leaf area of the existing vegetation cover will be decreased resulting in a subsequent decrease in water usage and therefore increasing groundwater recharge. Additionally, crop rotation is also very important. Research by Reu (pers. comm., 1992) indicates that recharge rates under the traditional wheat/fallow rotation system are 18mm/year whilst a rotation of wheat/medic (replacing fallow with a legume break crop) will result in 7.5mm/year. This data highlights how correct crop management will result in a reduction of recharge.

Topography also influences recharge. Slope affects recharge by its effect on groundwater tables. Bettenay and Mulcahy (1972) noted that, in south-western Australia, valleys with steep gradients tended to have water tables that lay deep in the landscape, conversely valleys of lower gradients had much higher water tables. This relationship is related to the amount of runoff that occurs on differing slopes; generally the steeper the slopes the greater the runoff and therefore the less water available to increase the water table level.

2.2.2 Throughflow Areas

In the 1970s two schools of thought developed to explain subsurface water movement. One group, Conacher (1975) and Conacher and Murray (1973), followed Cope's (1958) work and stated that water infiltrates to impermeable clay layers where it forms a perched water table and moves laterally as throughflow to emerge as saline seeps in valley floors. The other school of thought adopted the hypothesis that deeper percolation to groundwater bodies causes a rise in groundwater tables and discharge at the surface especially where groundwater slope and land surfaces intersect (Jenkin, 1979; Macumber, 1969 and Peck, 1978). The discussion as to which was the most important mechanism ended when it was accepted that both mechanisms could possibly operate at the same time in one area.

It is generally considered that the flow path in most groundwater systems consists of a shallow water table perched in sand (that is usually the dominant source of water) and a perennial deeper aquifer which is the major source of salt discharging to the saline seeps (Williamson, 1990b). The distance the groundwaters flow before emerging to cause salinity is determined by geologic, soil, land cover and topographic factors (Dyson, 1990). This is a critical issue, as it indicates that farm management practices in one zone may affect the development of salinisation in other areas.

Dyson (1990) identified three major types of groundwater systems. Firstly, local systems occur when the groundwater cannot flow over long distances, due to the geologic and soil conditions, and therefore flows from the slopes to the adjacent valley floors. With these systems the flow system boundaries generally correspond to surface water catchments (Dyson, 1990). The salinity occurring in these areas may be directly related to the land management within the catchment. Secondly, regional groundwater systems typically occur in sedimentary basins where the groundwater moves freely through the deeper perennial aquifer recharging to the regional groundwater table leading to its eventual rise and the subsequent saline seepage. Management of these systems is difficult as it requires a coordinated effort over a region where recharge is occurring. The third and final system is the intermediate groundwater system where water flows across catchment boundaries and throughout large areas, but is not regional as the process occurs over tens rather than hundreds of kilometres. The management of this groundwater system is also very difficult due to the need to implement rehabilitation programmes over the entire area.

The importance of this observation to the implementation of management strategies cannot be over stated.

2.2.3 Groundwater Discharge Areas

The discharge of the saline water from the groundwater system is by either flow into a stream or by capillary rise and subsequent evaporation from the soil surface (Nulsen and Henschke, 1981). With active salinisation in a saline seepage area the salt levels are expected to be higher near the surface where the soluble salt load is deposited, but where upward movement is not occurring (or occurring at a very slow rate) infiltration of rainwater tends to leach soluble salts from the surface into the subsoil increasing the salt content with depth (Bullock, 1991).

The increasing soil salinity in the discharge zone subjects the vegetation cover to increasing salt stress and suppresses plant growth. Two principal stresses result in characteristically dwarfed and stunted vegetation with reduced branching and leaf area:

1) drought stress, caused by an osmotic effect where the dissolved substances in the soil water are greater than in the plant cell; and
2) toxic stress resulting in disorganised cell structure, caused by the presence of specific ions; for example in barley, as the salinity levels increase the chloroplasts in the leaves swell, causing the internal components of the chloroplasts to become indistinguishable (Poljakoff-Mayber, 1975).

In such situations, yield reductions of over 20% can occur before salt damage is apparent to the land manager (Russell, 1973).

If the soil becomes increasingly saline, the stressed vegetation cover may perish paving the way for the establishment of more salt tolerant halophytic species (Chapman, 1975). If the groundwater continues to rise the salt tolerance levels of the halophytes may be exceeded resulting in their senescence. The removal of the vegetative cover from the discharge zone enhances the possibility for significant soil erosion to occur. This possibility is further increased by the dispersive nature of sodium enriched soils (Peck, 1977). As soil is eroded from saline areas all the normal complications of soil erosion are encountered. These include, silting of water ways, exposure of subsoil, loss of productivity and reduction of aesthetic value. It can also be noted that the rehabilitation of eroded saline ground is generally more difficult than other areas due to the problems associated with trying to reintroduce a vegetation cover.

There are three main topographic regions where saline seepage occurs. The first, valley salting, is noted to be the most widespread form in southern Australia (Working Party on Dryland Salting in Australia, 1982). It occurs where water movement through the soil continues unobstructed into the general groundwater store, causing an increase in the water table level to within capillary range of the valley soil surface. If the landscape is generally flat and poorly dissected the opportunity for salts to be drained out into the river system is reduced, resulting in an eventual build up of salts in the drainage area (Hartley, 1984). The second, hillside seepage salting, is found where the downward movement of water through soil on a hillside is obstructed by a dense impermeable layer, usually a rock, i.e. dykes or clay (Williamson, 1990b). The groundwater accumulates above this layer resulting in a perched water table that discharges to form a saline seep. The third, break of slope salting, commonly occurs at the base of slopes where a sudden decrease in the hydraulic gradient allows a discharge of saline water to the soil surface.

The role of soil and geology in the development of saline seepage relates, as mentioned, to the retardation of groundwater flow. Geological features, such as elevated basement rock, act as barriers to subsurface flow leading to an accumulation of subsurface groundwater and an eventual rise in the groundwater table. Similarly, soil types with low hydraulic conductivities may also inhibit the subsurface flow of groundwaters (Williamson, 1990b).

2.3 MANAGEMENT STRATEGIES

As it is virtually impossible to eliminate all the stored salt in the soil profile, salinity management strategies are aimed at either eliminating the source of water or modifying the mechanisms which redistribute the salt. In line with this premise there are two types of management strategies. Firstly, the long term management strategies address the basic cause of the problem, i.e. these aim at stopping excess groundwater recharge within the landscape from entering into the hydrological system. These are deemed long term strategies as the lags within the system ensure that even if groundwater recharge was halted there would still be water moving through the hydrological system. Secondly, short term strategies deal with the saline seepage areas where the groundwater discharges at the surface. Importantly, landowner management in groundwater recharge zones will impact on land owner management in groundwater discharge zones.

2.3.1 Groundwater Recharge Management

With the acceptance that tree clearance was the basic cause of the change in the hydrological system, the earliest groundwater recharge strategies were primarily concerned with replacing the cleared trees (Morris and Thomson, 1983). This resulted in research concentrating on the suitability of trees for certain areas, their water usage characteristics (Greenwood et al., 1985) and tree planting densities (Schofield et al., 1989). Recently, with the acknowledgement that rehabilitation programmes should not disadvantage the landowner economically, research has been primarily concerned with the use of both commercial tree species and agricultural strategies, i.e. planting high water usage crops such as lucerne to control groundwater recharge (Williamson, 1990b). Although research into maximising water usage within catchments is important, other research has concentrated on assessing what percentage of a catchment requires reforestation.

Three methods have been developed to determine the percentage of catchment that should be subject to rehabilitation strategies. The first, the regression method, relates the measured changes in groundwater level with the proportion of the cleared area that was reforested (Williamson, 1990b), e.g. to lower the water table at 200mm/year in a region which receives 700mm/year in rainfall, a catchment would have to be 91% reforested (Williamson, 1990b). Williamson notes that "this simple approach provides a guide, though it embraces many variable factors which suggest caution in its application" (Williamson, 1990b).

Another approach, the water balance method, first proposed by Peck (1978) and later refined by Schofield (1988), considers that forests remove water at the rate of pan evaporation, e.g. assuming a pan evaporation rate of 1870mm/year in an area of 750mm/year rainfall, 32% of the area of cleared land would need to be reforested to lower the groundwater level by 200mm/year (Williamson, 1990b).

The third method, developed by Hookey and Loh (1985) and Hookey (1987), is based on groundwater modelling using the Prickett and Lonnquist finite difference model (Hookey and Loh, 1985). This model varies in two ways from other models; firstly, it uses different data input values (i.e. hydraulic conductivity of 270mm/day, recharge rate of 40mm/year and a storage coefficient of 0.05) and, secondly, by utilising empirical parameters a spatial component which attempts to identify recharge zones is added to the model. The study by Hookey (1987) on two catchments in Western Australia determined that 33% to 48% of cleared land in the study region required reforestation if the saline area was to be controlled. An interesting finding, in relation to this thesis, was that the area and location of the reforestation were critical to the control of groundwater recharge. This highlights both the heterogeneous nature of the landscape in which subsurface water movement occurs and the important role of preferred pathways.

The spatial variability of groundwater recharge dictates the main limitation with this groundwater model and others such as *Topog* (Hatton and Dawes, 1991). These models rely on empirical data derived from drilling programmes to determine factors such as hydraulic conductivity and infiltration rates; this results in point data being extrapolated over a catchment area. Cook (pers. comm., 1992), states that on average 20 to 40 holes need to be drilled in any one catchment to determine the major recharge zones. As drilling is labour intensive, slow and expensive, both Cook (pers. comm., 1992) and Morris and Thomson (1983) acknowledge that on a regional scale it will, it seems at best, be some years before extensive recharge data is available for the major salt affected catchments of southern Australia.

Dyson (1990) and Hughes (1985) have recognised this problem and implemented regional landscape land classification approaches to identify areas of groundwater recharge. Although they have been relatively successful, weaknesses in the Landscape land classification approach (considered in section 4.1.2) imparts limitations on their methodology.

The area of research which has attracted the most recent interest has been the application of remote sensing techniques to the identification of both groundwater recharge and discharge zones. This area of literature will be considered in the following chapter.

2.3.2 Groundwater Discharge Management

Unlike groundwater recharge zones, groundwater discharge zones are easily identified by the surrogate variable of saline seepage; thus the main area of research has primarily concentrated on maximizing water use in the discharge zone. Clearly, if the location of potential groundwater discharge areas were known prior to the development of scalded areas, appropriately placed rehabilitation strategies would return a better investment on the money spent, i.e. the retention of maximum production potential. Both tree planting and agricultural strategies have been employed to rehabilitate areas.

The main supposition for planting trees near seepage areas is their usefulness to draw down the water table, to stop capillary rise and to allow the leaching of surface salts. Trees in this role are essentially seen as biological pumps (Greenwood and Beresford, 1979). Studies have shown that water tables can be significantly lowered by the use of tree species. Webster (1983) noted that planting of trees in the Wimmera-Mallee region successfully lowered the water table in some cases by up to 2 metres.

Initially, the lowering of the water table associated with a saline seep will be in the form of localised cones of depression around each individual tree; however Morris and Thomson (1983) notes that over time a general reduction can be achieved by further plantings into the centre of the seep. Obviously, this notion is dependent on the ability of the tree species to adapt to the high salinity levels. Research by Marshall

(1976), Sands and Rishworth (1978), Tulloh (1978), Thomson (1981) and Luard and El-Lakany (1984) has identified a number of species which are suitable for this task.

One problem with tree planting in discharge zones is that as well as reducing groundwater discharge they also reduce the volume of surface runoff. Thus water that would have normally run-off, now accumulates in the discharge zone further compounding the problem (Williamson, 1990a).

Agricultural strategies for controlling salinity in saline discharge areas have primarily concentrated on measuring evapotranspiration rates and salt tolerance levels of different pastures and crop species. Research has found that the salt tolerant halophytic saltbush species (Atriplex spp.) are well adapted to harsh saline sites (where trees are difficult to establish) and have been able to lower the water table in low rainfall (less than 500mm/year) areas (Schofield et al., 1989). Similarly, bluebush (Mariana brevifolia), and samphires (Habscurcia spp.) have been successfully grown and are providing valuable grazing for sheep during the autumn and early winter period. Other grasses, such as tall wheat grass (Agropyron elongatum) and puccinellia (Puccinellia cillata), with lower salt tolerant capabilities, have also been planted by landowners on salt affected discharge areas. The establishment of these salt tolerant species have enabled landowners to derive at least some productivity from salt affected lands.

To initiate rehabilitation strategies in areas of potential groundwater discharge they must firstly be identified. One model developed to solve this dilemma was formulated by Slessar et al. (1991). Their approach applies a three dimensional topographic model to divide the landscape into units conducive to groundwater discharge. Although Slessar et al. (1991) concluded that the project failed to meet expectations, it is considered by this thesis that if the topography was considered in relation to other environmental factors then the results would have been more successful. Remote sensing techniques, as previously mentioned, have been applied to the problem of identifying potential groundwater discharge zones. The results will be considered in section 3.4.1 of the following chapter.

2.4 SUMMARY

This chapter has highlighted the heterogeneous nature of both groundwater recharge and discharge and the extent of qualitative knowledge known about salinisation. It is noted that, due to the spatial variability of salinisation, regional assessments of the extent of the problem have not been conducted. This has enormous implications for management of both groundwater recharge and discharge zones. If accurate identification of relevant areas is not possible preventative measures cannot be implemented.

Importantly, this chapter has also shown the dynamic relationship between groundwater discharge and recharge areas. The implications of this relationship are of importance for this thesis. Management of the problem areas, both groundwater recharge and groundwater discharge once identified, must involve the landowners being proactive in key locations. It is a land degradation problem that cannot be acted upon in isolation. It requires a coordinated and integrated catchment approach.

By reviewing the literature on how remote sensing techniques and Geographic Information Systems (GIS) have been utilised to identify both potential groundwater recharge and discharge an understanding of the approach taken by this thesis will clarified.

Chapter 3

REMOTE SENSING AND GEOGRAPHIC INFORMATION SYSTEMS

Remote sensing and GIS have developed rapidly as research areas over the past two decades and are concerned with the handling and processing of spatial data. As the occurrence of saline seepage areas are related to spatial phenomena, both remote sensing techniques and GIS have been used to identify this problem, however due to the initial incompatibility of the two systems little success was achieved.

The objective of this chapter is to highlight how increased knowledge of remote sensing techniques and GIS has led to the symbiotic integration of the two and subsequently the methodology adopted by this thesis. The chapter consists of four sections. The first and second will define remote sensing and GIS, while the third will then analyse the integration of remotely sensed data with a raster based GIS. The fourth section will consider applications of remote sensing techniques and GIS to the identification of groundwater recharge and groundwater discharge areas.

3.1 REMOTE SENSING

Observation of the earth's surface by remote sensing instruments first began in France in 1859 when aerial photographs were taken by Gaspard Felix Tournachon from an air balloon over the village of Petit Becetre (Simmonett, 1983). From these inauspicious beginnings remote sensing has now evolved into a technologically orientated science utilising instruments located on platforms in both the air and space.[1]

Remote sensing can be broadly defined as "the use of electromagnetic radiation sensors to record images of the environment which can be interpreted to yield useful information" (Curran, 1985, p. 239). More specifically it can be defined as, "the acquisition of data and derivative information about objects or materials (targets) located at the earth's surface or it's atmosphere by using sensors mounted on platforms located at a distance from the targets to make measurements of interactions between the targets and electromagnetic radiation" (Harris, 1987, p. 2). Although different, the above definitions are fundamentally the same in the sense that they both employ information derived from the electromagnetic spectrum to increase the knowledge of the earth's processes.

Remote sensing systems include data acquisition systems such as sensors on aircraft or satellites, data transmission systems and data processing systems (Zhou, 1989). This thesis will concentrate on the information obtained from data acquisition systems and when considered will be referred to as remote sensing systems.

3.1.1 Remotely Sensed Data Types

Existing remote sensing systems produce large volumes of data in a format that can be displayed as an image. These images consist of small equal sized cells known as discrete picture elements, or pixels, arranged in regular lines and columns. Each cell, or pixel, is assigned a numerical value which represents the radiation or brightness level recorded for each defined wavelength band of the electromagnetic spectrum by the scanner. The value range for most remotely sensed imagery is from 0 to 255. This dynamic range, termed the radiometric resolution, is expressed in binary digits or bits, e.g. 256 levels of brightness has an 8 bit resolution, $2^8 = 256$.

---

[1] The term remote sensing was developed in the early 1960s by Geographers in the Office of Naval Research of the United States Navy (Simmonett, 1983).

In relation to data handling, the number of spectral measurements or bands determine the volume of data provided by a particular sensor, e.g. Landsat TM has 7 wavebands with 8 bit radiometric resolution, 6 of the bands have a 30m spatial resolution, the seventh has a spatial resolution of 120m. The spatial extent of the image is 185km^2 and therefore consists of 2.37 million pixels for the seventh band and 38 million pixels for the other six bands. At 8 bits per pixel a complete 7 band image is composed of 1.848 x 10^9 bits or 231 megabytes (Richards, 1986).

Analysis of this data involves the use of mathematical and statistical data processing techniques to transform the "meaningless" raw spectral information into "meaningful" information representative of geographical features. The data processing techniques can be divided into three groups. Image restoration processing methods compensate for data errors, noise and radiometric and geometric distortions introduced during the scanning, recording and transmission operations. The major processes are listed on Table 3.1.

TABLE 3.1: Image Processing Techniques

- Restoration
  - Radiometric correction (for system and environmental effects)
  - Geometric correction (image to map or image to image)
- Enhancement
  - Digital mosaics
  - Density slice
  - Magnificent or reduction
  - Transects
  - Contrast stretch
  - Image algebra (band ratioing, differencing, etc.)
  - Spatial and directional filtering
  - Edge enhancement
  - Principal components
  - Linear combinations (e.g. Kauth transform)
  - Texture transforms
  - Fourier transforms
- Information extraction
  - Supervised classification
  - Unsupervised classification
  - Contextual classification
  - Incorporation of ancillary data in the classification
  - Change-detection images

Source: Zhou (1989) (cited from Sabins, 1978; Jensen, 1986).

Image enhancements alter the visual impact that the image has on the viewer (Sabins, 1978). The enhancement does not produce information but rather enriches the existing information, resulting in the variations within the image being enlarged according to the user's specifications.

Information extraction utilises the decision making capability of the computers to identify and extract specific pieces of information from the corrected and improved image. In this procedure, human operators must provide the initial training data and instructions for the computer and must also be available to evaluate the significance of the extracted information (processes listed on Table 3.1) (Sabins, 1978). The output from this process is a classified image composed of a number of different themes representing particular spatial phenomena.

This brief analysis of remote sensing systems indicates that these systems produce large volumes of data which are processed by statistical and mathematical techniques to extract information to aid the decision making process. These systems do not provide database management capabilities and are therefore primarily considered to be data transformational.

3.2 GEOGRAPHIC INFORMATION SYSTEMS

In contrast, Geographic Information Systems (GIS) are database management orientated and are now designed to utilise information extracted by remote sensing systems. A GIS has been defined in various ways: "a special use of information systems where the database consists of observations on spatially distributed features, activities or events ....." (Dueker, 1979, p. 105); "a system, commonly computer based, for handling spatial data ..... It is designed to accept large quantities of spatial data derived from a variety of sources ..... and to efficiently store, retrieve, manipulate, analyse and display these data according to user-defined specifications" (Marble and Peuquet, 1983); or "a decision support system involving the integration of spatially referenced data in a problem solving environment" (Cowen, 1988, p. 1554).

A summary by Clarke (1986) considers the general characteristics of a GIS as:

1) a large body of data which have spatial or locational properties;
2) numerical or logical expressions of relations between these data;
3) a common file or data structure; and
4) the ability to perform the functions of data collection, storage, retrieval, analysis and automated mapping.

From the general definitions, a GIS should be able to input, store, retrieve and manipulate data and produce both tabular and cartographic outputs which reflect the selective retrieval and manipulation of entities within the database.

3.2.1 Input of Data

Data input is the operation of encoding spatial data and writing it to a database (Zhou, 1989). The creation of a precise digital database is an important task as errors introduced in this process propogate through the data analysis,manipulation and output stages. There are three stages in the data input process: data capture; data cleaning and editing; and geometric correction (Zhou, 1989). These stages will be considered in Chapter 6.

3.2.2 Database Management: Storage and Retrieval

Data storage and retrieval is carried out by a Data Base Management System (DBMS). There are a number of different DBMS systems, but this thesis will only consider a relational database

Relational databases, where the data and data relationships are normalised to a series of two dimensional tables with the rows being records that contain an ordered set of attribute values, are the most commonly used non-spatial DBMS. To link the attribute data with spatial data, where each entity is defined in terms of its location in space, the stored spatial data must be encoded in an accessible format.

There are two conceptual approaches to spatial data structures, the vector and raster approaches. Vector geocoding systems record data in point, line and polygon format with a high level of positional accuracy. These data structures maintain spatial data independence regardless of the change of scale and map projection and permit simple and efficient interfaces between the tabular DBMS records and their corresponding spatial data entities (Zhou, 1989). Vector data structures have traditionally been adopted by the major GIS software vendors, such as Environmental Systems Research Institute (ESRI.).

Problems inherent within these data structures are (Zhou, 1989):

1) considerable computational expense associated with spatial file editing and updating;
2) complex topology, involving calculating and encoding relationships between the nodes, arcs and polygons are required to achieve efficient data extraction; and
3) data processing algorithms are complex and inefficient for polygon overlaying and Boolean operations.

The raster approach is more appropriate for the three problems described above, as it allows geocoded data to be referenced to a grid cell data set. Thus, thematic maps or images in grid cell format taken from remote sensing platforms, can be processed with minimal problems into the spatial data structure. The raster system permits rapid data combination, easier mathematical operations for different data layers within each grid cell and fast algorithms for spatial modelling (Pequet, 1979). More advantages are provided relative to the vector approach in land resource management where intensive data combinations based on several data layers are required. Disadvantages with raster systems are (Zhou, 1989):

1) the data scan has to be resampled when changes in scale and map projection are required;
2) network analysis is difficult due to the lower spatial accuracy; and
3) due to the vast number of grid cells, one to one interfaces between the tabular DBMS and raster data is not always practical.

3.2.3 Analysis and Manipulation of Data

By definition a GIS must be capable of performing a series of manipulations on the spatial data held in its files (Zhou,1989). As spatial data are held in two formats, data processing algorithms vary accordingly. Vector data structures utilise vector-based algorithms for all internal data manipulations whereas raster system structures use image-based algorithms to mathematically analyse on a cell-by-cell basis several data layers. The major problem inherent in GIS systems derives from the limiting nature of software tools; most GIS do not employ advanced statistical analysis, clustering algorithms or knowledge-based software programmes to analyse data. In this thesis, the interfacing of GIS with external data processing procedures is considered to be the area where data manipulation will greatly improve.

3.2.4 Data Output Capabilities

Data output capabilities of GIS are usually a tabular report based on the results of the data analysis and a map output (Zhou, 1989). The map output is generally available, in accordance with the spatial data structures, in two formats. The cartographic approach produces a map output using combinations of vectors (points, lines and polygons), whilst the photographic approach converts the spatial data into a colour image where the brightness of each pixel (grid cell) represents the attribute value of the corresponding grid cell in each particular data layer (Zhou, 1989). In this thesis the majority of the output will employ the photographic approach to represent continuous environmental data.

3.3 INTEGRATING REMOTELY SENSED DATA WITH A RASTER BASED GIS

The integration of a raster based GIS with remotely sensed data results in a symbiotic relationship, whereby the data management capabilities of the GIS can fully utilise the data provided by the remote sensing system. As the two data structures are compatible, the data processing algorithms treat the data in both systems as being a Boolean surface. Therefore, data from a GIS may be incorporated into the conventional multivariate classification process within a remote sensing system or, conversely, "raw" or "classified" remotely sensed data may be analysed by statistical, logical or mathematical means within the raster based GIS.

Traditional raster based GIS, such as the Map Analysis Package (MAP), did not have the aspatial component linking each of the grid cells to its related attributes, as this was impracticable (Zhou, 1989). Thus no remotely sensed data was transferred into this system. To overcome this problem an Image Based Information System (IBIS) interface file was developed where for each labelled zone the number of image elements with the same attribute was recorded (Bryant and Zobrisk, 1981). Following from this simple tabular file more recent raster based GIS, such as RIGIS (Zhou, 1989), define and use one or more relational tables which are then handled by a RDBMS, allowing the integration of remotely sensed data with other environmental variables.

The development of raster based GIS has now reached the stage where both raster and vector data structures are within the one GIS system e.g. Arc/Info V.6.0. An integrated raster/vector based GIS has been selected for this project for three reasons:

1) it provides the opportunity to develop a fully integrated data set, of cadastral, environmental and remotely sensed data, which can be manipulated to identify key aspects of the regional salinity problem;
2) the grid cell size of a raster based GIS can be set at a high resolution (i.e. 30m^2) to provide information at the farm scale

whilst, with today's computing power, still providing the ability to conduct a regional analysis;

3) a raster based GIS allows the incorporation of raw remotely sensed data into the database, providing the option of identifying land cover types in relation to other environmental factors such as slope, elevation, soil type and geology without having to use probabilistic multivariate classification procedures. This thesis will integrate a raster GIS with an Expert System (ES) to analyse environmental datasets, including remotely sensed satellite imagery and Electromagnetic Induction (EM) data, to subsequently produce a thematic land classification map, identifying areas of existing and potential groundwater recharge and discharge.

3.4 APPLICATIONS OF REMOTE SENSING TECHNIQUES AND GIS TO THE IDENTIFICATION OF GROUNDWATER DISCHARGE AND GROUNDWATER RECHARGE

The separate development of remote sensing systems and GIS is reflected by a literature review of attempts to identify groundwater recharge and groundwater discharge. Early attempts utilised only remotely sensed imagery whilst later attempts attempted to incorporate remotely sensed imagery with other types of environmental information held within a GIS. This sub-section will consider attempts to identify both groundwater discharge and recharge.

3.4.1 Groundwater Discharge Identification

Remote sensing researchers have made use of the positive relationship between salt stressed vegetation, volunteer halophytic vegetation and the increasing rate of soil salinity to identify groundwater discharge areas (e.g. Kirkby et al., 1994a). Saline seepage areas, as already noted, are the visible expression of salinisation processes occurring in an area, therefore, in an attempt to determine the extent of salinisation processes, remote sensing techniques have been employed.

The traditional remote sensing system, aerial photography, has successfully been used to identify scalded saline seepage zones (Pankova et al., 1978; Mazikov, 1978; Kolarkar et al., 1980). Recording information extracted by these methods was time consuming and subject to the interpretation of the individual research worker.[2] In contrast, the modern image remote sensors such as Landsat Multispectral Scanner (MSS) and Thematic Mapper (TM), SPOT HRV, NOAA AVHRR and various radar systems all produce digital data which is processed using automated computer technology.[3]

The identification of dryland salinisation using modern image remote sensors first began in 1972 when Prentice (Richardson et al., 1976) demonstrated that he could detect conditions indicative of saline scalding. Honey et al. (1984) achieved similar results using Landsat MSS imagery but noted that, due to the coarse spatial resolution of the sensor, it was not suited to identifying saline seepage areas where scalding had not yet occurred.

In contrast, Chaturvedi et al.(1983) utilising colour infrared aerial photography and Hick and Russell (1988) using Geoscan MSS airborne data were both able to detect the scalded areas and the gradational response from healthy to stressed vegetation and then halophytic vegetation in relation to increasing salt levels. By analysing a spectral response curve of vegetation to increasing salt levels an understanding of their basic premise is attained.

Figure 3.1 is a typical spectral response curve for healthy vegetation. The visible green peak (point B:550nm) is due to the reflective characteristics of

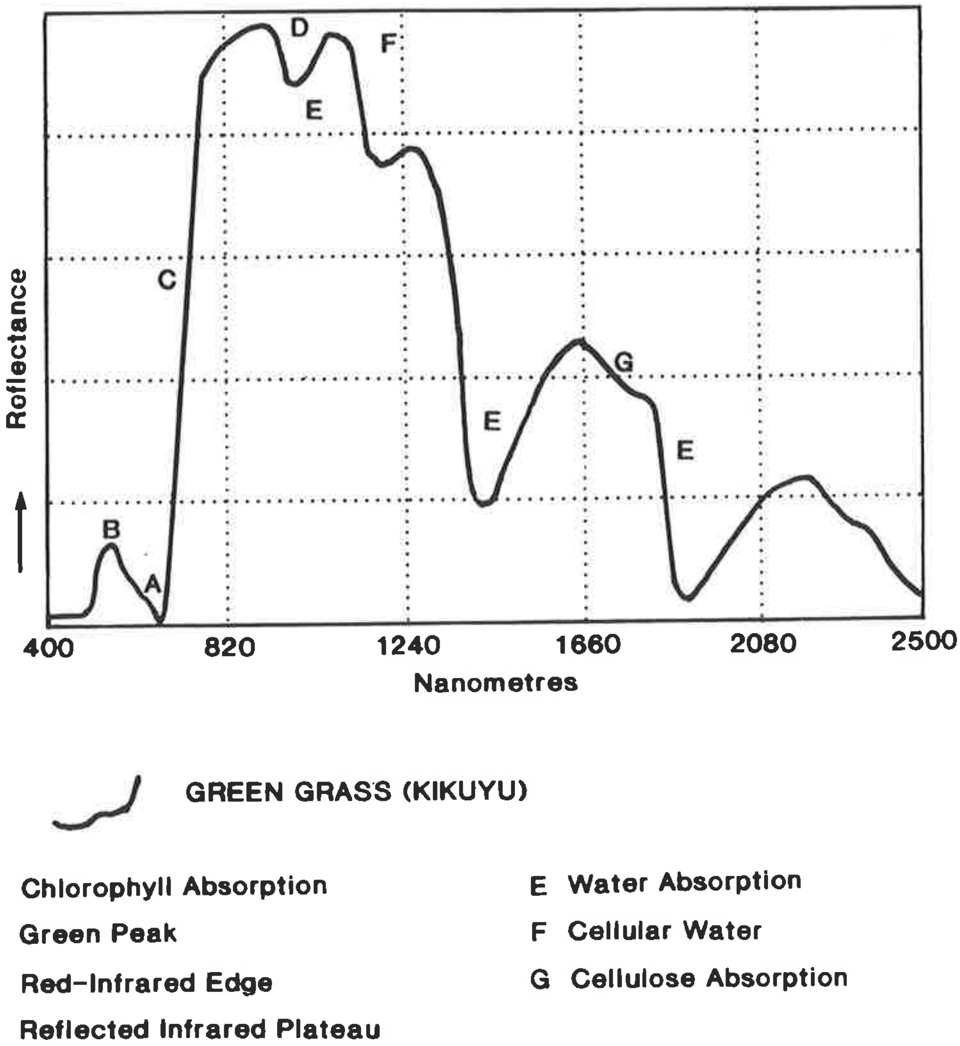
---

[2] This is not a problem when one research worker conducts a historical analysis and interprets the whole series of photographs, as done by Van Heusdan (1983) and Stephens et al., (1982).

[3] Due to the nature of current classification procedures this process is still open to the vagaries of the individual operator.

photosynthetically active vegetation dominated by carotenoid pigment absorption in the visible blue (point A:480nm) and chlorophyll absorption in the visible red (point C:680nm) (Hick and Russell, 1988). At 700nm (point D), the sharp rise in near infrared reflectance indicates the change from chlorophyll absorption to cellular reflectance (Hick and Russell, 1988). This dramatic increase in near infrared reflectance is determined by the cell wall and air space interfaces of the mesophyll layer, and to a lesser extent by the specific characteristics of these cells (Hick and Russell, 1988). Thus during the vegetation cycle, until maximum canopy density is reached, reflectance decreases at 680nm and increases between 700 to 1300nm. With senescence a reversal of this process occurs.

FIGURE 3.1: Spectral Response Curve for Healthy Vegetation



Source: Hick and Russell, 1988.

As the healthy vegetation in the discharge area experiences salt related stress, its spectral reflectance decreases in the near infrared. This occurs because the cellular structure of the vegetation in the mesophyll layer is affected by the increased salt concentrations. Additionally, the increased salt levels alter the capacity of the chloroplasts to absorb light, resulting in an increase in the spectral reflectance of the vegetation in the visible red (point C) area of the electromagnetic spectrum. These two facts enable an interpreter using the satellite imagery to differentiate between vegetation experiencing salt levels and healthy vegetation. Additionally, it has been observed by Graetz and Gentle (1982) that the leaves of the halophytic species Atriplex vesicaria (saltbush) are covered with collapsed epidermal hairs containing crystalline sodium chloride which substantially increase the reflectance in the visible wavelengths.

The three conclusions from the project conducted by Hick and Russell (1988) were:

1) the optimum seasonal time to discriminate between the stressed/halophytic vegetation and the healthy vegetation cover was during the spring flush, where the spectral variation of the vegetation cover was at its greatest and the influence of the soil background was diminished;
2) the best three band combination to identify groundwater discharge areas during the spring flush were: Band 3 (650-700nm); Band 4 (830-870nm) and Band 6 (1980-2080nm) of the Geoscan multispectral scanner;[4]
3) since existing satellite systems do not possess the optimal band combinations, there is a need to combine different types of environmental information to increase the probability of identifying saline discharge areas.

Nevertheless, research by Szilagyi and Baumgardner (1991), Sharma and Bhargava (1988), Rao and Venkataratnam (1991) and Kirkby (1993b) all successfully identified existing saline groundwater discharge areas using the 630 to 900nm range of the

---

[4] These bands correlate with two of the Landsat TM satellite bands; Geoscan Band 3 with TM Band 3 (630-690nm) and Geoscan Band 4 with TM Band 4 (760-900nm).

electromagnetic spectrum. Unfortunately, none of these studies were able to identify potential groundwater discharge zones.

Two other areas of the electromagnetic spectrum have been employed in an attempt to detect the formation of potential saline seepage areas prior to the vegetation cover being subjected to increasing salt levels. The first is the use of passive microwave sensors which have been either hand held or deployed from mobile platforms such as trucks. The theory for the application of this technique is that the water in the soil changes the microwave dielectric constant which in turn changes the spectral emission, therefore areas with increased soil moisture, such as saline seepage areas, should theoretically be detected. Early research by Jackson et al. (1983) and Chatuvedi et al. (1983) was unsuccessful with this technique, but later work by Jackson and O'Neill (1987) and Reutov (1989) using the L band (15.0 to 30.0 cm) and C band (3.8 to 7.5cm) was more successful. The major limitation with this technique relates to the limited spatial resolution of these remote sensing instruments. Further research on the dielectric constant is expected when active satellite radar systems such as ERS 1 and JERS 1 become operational.

The other area of the electromagnetic spectrum that has been utilised to identify potential saline seepage is the thermal infrared. Patterns which appear on a thermal infrared image are primarily a function of the temperature of the earth's surface (Huntley, 1978). Therefore, areas of high soil moisture, where groundwaters have risen and the soil surface is generally cooler, result in a distinct spectral response (Price, 1980). No conclusive results for potential groundwater discharge development have been produced by these papers.

Noticeably, all of these early attempts failed to incorporate the use of other environmental information held within a GIS during the classification process. This fact can be related to the initial incompatibility of the early vector GIS systems and remotely sensed data structures and to the ease in which groundwater discharge areas may be identified. In contrast, with the recent improvement of raster GIS systems and the realisation as to the difficulty of identifying potential groundwater discharge, some research work has concentrated on analysing remotely sensed imagery in relation to other environmental data held within a GIS. For instance Jupp et al. (1990a, 1990b, 1990c) combined AVHRR thermal data in a GIS to identify potential groundwater discharge areas within the Murray-Darling basin of southern Australia. The basic premise of this project was to isolate areas, by the use of ancillary geologic, soil and topographic data, where saline seepage may occur and then apply the thermal data to the specified areas. It was hoped that the reduced spectral range of the thermal image would enable enhancement and classification techniques to identify the saline seepage areas. As yet no conclusive results for the identification of potential groundwater discharge development have been produced, although recent papers have been provided interesting results on the identification of soil moisture changes (Jupp et al., 1994).

3.4.2 Groundwater Recharge Identification

To identify groundwater recharge zones two different approaches have been adopted. The first is based on the use of a range of remote sensing techniques to identify a surrogate variable representative of groundwater recharge. The second combines ancillary data from a GIS with remotely sensed data to identify groundwater recharge. The second approach reflects a gradual shift in remote sensing methodology to the analysis of integrated data sets within both remote sensing systems and GIS.

An empirical study conducted by Mackenzie et al. (1990) used thermal data to identify groundwater recharge zones. This study found a correlation between the thermal infrared band (8500-13000nm) on the Daedelus airborne scanner and low electrical conductivity readings measured by the Geonics EM34 conductivity meter. As the low EC readings are directly proportional to the coarseness of the soil, and as coarse materials imply a potential for rapid recharge or transmission of groundwater, these results indicate that high radiometric/spectral/spatial resolution thermal data could theoretically identify groundwater recharge zones. But as the final classification results were poor the conclusion stated "other data sets such as the geology of an area, need to be taken into consideration when using broad based data sets such as remotely sensed satellite thermal imagery so that areas can be eliminated from the classification"(Mackenzie et al., 1990).

A relatively new remote sensing sensor applied to the identification of groundwater recharge areas are the electromagnetic induction (EM) techniques, as briefly referred to in the Mackenzie et al. (1990) paper. These techniques measure bulk electrical conductivity of the soil profile. The response is due to a number of soil properties including salt storage, moisture content and clay content. Most EM sensors (Geonics EM31, EM39, EM34) are land based sensors (hand held or mounted onto a 4WD car), which measure electrical conductivity at points along a transect. They are, therefore, labour and cost intensive in terms of person hours for data collection over large areas. This problem has been recognised, resulting in attempts to mount EM instruments on airborne platforms (Cook and Kilty, 1994). It is clear from the literature that EM techniques are exceptionally useful for identifying groundwater recharge areas, yet authors, such as Williams and Arunin (1990), who have successfully used EM techniques for inferring groundwater recharge zones, state that other land system data must be incorporated into the construction of groundwater recharge and discharge maps (Williams and Arunin, 1990).

Hill (1990) combined Landsat TM data with ancillary information in a study of groundwater recharge areas in Bendigo, Victoria. This project used ancillary geologic and landuse information derived from a GIS to enhance the accuracy of the image classification. To determine spectral training classes, Hill traced the ancillary boundaries from the vector based GIS onto paper, manually overlayed these boundaries onto the TM image, chose the appropriate sites and then classified the imagery. Although this study indicated the effectiveness of combining relevant ancillary information, it failed to effectively utilise the combined potential of integrating remotely sensed data with GIS.

A brief summary of these remote sensing approaches indicates four points:

1) saline groundwater discharge areas can and have been detected by satellite based remote sensing systems;
2) potential groundwater discharge areas have not, as yet, been successfully identified by remotely sensed data;

3) groundwater recharge zones have been identified by combining remotely sensed electromagnetic induction (EM) techniques with ancillary data within a GIS. Due to the labour intensity of the data collection technique, it is currently not viable for regional studies. In the future, with attempts to mount the sensor of airborne platforms, this will problem will cease to exist. This point will be addressed later in the thesis;
4) most researchers consider the integration of remotely sensed data with ancillary data from a GIS to be a method which will successfully identify both existing groundwater recharge and discharge zones.

The last point raises the question: why, if remote sensing systems and GIS are regarded as the potential technological solution, has the problem not been solved ? Obviously the early problems associated with the compatibility of the remotely sensed data output and the early vector GIS systems contribute to this problem. Yet as these difficulties have now been addressed it is contended by this thesis that the main area which requires attention is the process of how to analyse integrated data sets held within a GIS.

3.5 SUMMARY

Researchers suggest that to identify potential groundwater discharge areas or recharge zones with remote sensing techniques, ancillary data is required to increase the probability of identifying the spectral responses representative of these features. It is considered by this thesis that instead of using ancillary information to enhance the remotely sensed information, the whole dataset including the remotely sensed information, should be analysed to develop a land class map indicating the key problem areas. This point reflects a change in methodology away from multivariate classification of remotely sensed imagery to the analysis of an integrated dataset within a raster based GIS. In the following Chapter the author will consider a brief review of land classification techniques, Expert Systems (ES) and the integration of ES with GIS to develop thematic land classifications.

Chapter 4

LAND CLASSIFICATION AND EXPERT SYSTEMS

Historically, land classification techniques have adopted a qualitative heuristic approach to determine land potential, due to the spatially variable nature of environmental parameters. With the development of new technologies, such as remote sensing and GIS to ascertain the spatial extent of environmental parameters, more quantitative approaches have been developed. It is contended by this thesis that the quantitative approaches developed are by no means an accurate and precise means by which to represent environmental parameters due to the uncertainty associated with the actual spatial location of particular attributes and the uncertainty associated with both the land classification decision making process and the data fusion process. As a result most land classification maps, produced using current quantitative methods, contain a high degree of uncertainty which, unfortunately, is not conveyed to the end user.

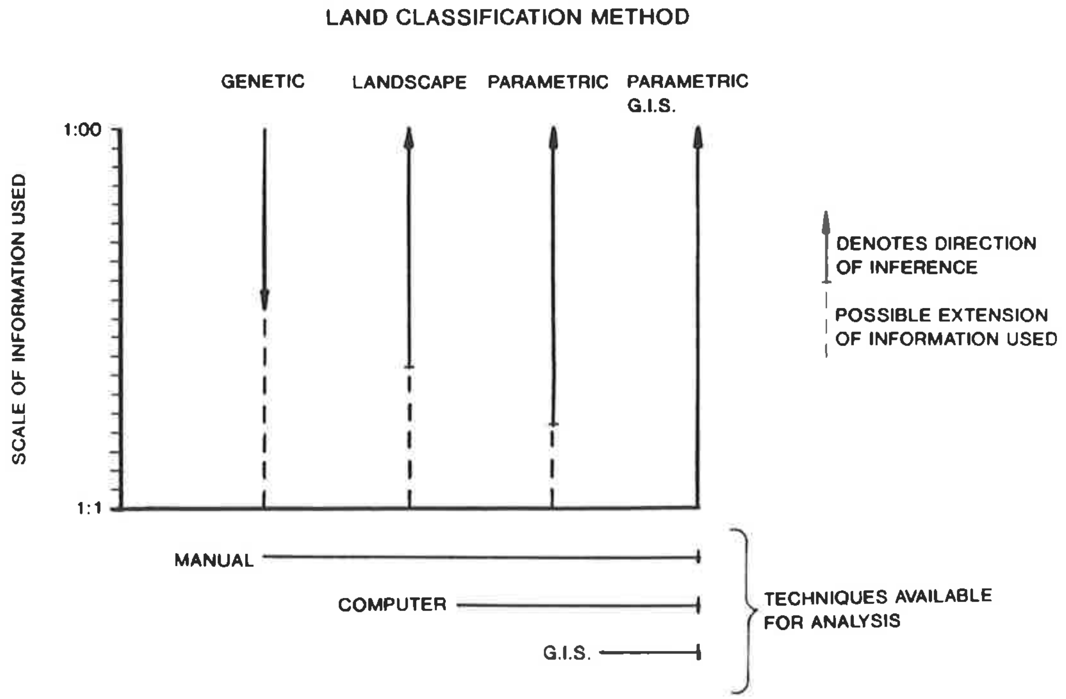
This chapter will describe a number of land classification methods, present a working definition of Expert Systems (ES) and review the literature on ES applications in land classification. Once this background information has been considered, along with the *Salt Manager* system being presented in Chapter 8, a new conceptual view regarding ES and GIS integration for land classification will be discussed in Chapter 10.

4.1 DEFINING LAND CLASSIFICATION

Land classification, as a means of recording land potential as a spatial phenomenon is central to the discipline of Geography. Hartshorne (1939) notes, "..... geography depends first and fundamentally on the comparison of maps depicting the areal expression of individual phenomenon or of interrelated phenomena" (Hartshorne, 1939, p. 463).

Land classification can be defined as providing a meaningful database for land use planning. The initial stage is to evaluate the land units, classified in terms of their capabilities for a range of potential land uses under consideration (Slessar et al., 1991). A brief summary of land classification techniques highlights three distinct approaches; genetic, landscape and parametric. The different approaches are summarised in Figure 4.1.

FIGURE 4.1: Conceptualised Diagram Describing the Evolution of Traditional Land Classification Approaches



Source: Kirkby et al., 1994b.

Each of these traditional land classification methods are based on a common method of decision making, where there are three basic stages experienced by an individual when making decisions to evaluate land units:

1) the receiving of stimuli e.g. the changes in electromagnetic energy recorded by the eye;
2) identifying the cause of variation in the stimuli e.g. parameters such as soil types or land covers; and
3) interpreting the phenomena which have given rise to the stimuli relative to some purpose of the interpreter (Gibbons et al., 1968).

The genetic, landscape and parametric land classification approaches all adopt this three stage decision making process when classifying land potential; yet all three approach the task in a different manner due to the variation in the availability of accurate spatially referenced environmental data and the associated technology to store, analyse and manipulate this data. Two of the techniques, the landscape and parametric approaches, evolved with the application of new technology to land classification problems. It is postulated that with the current developments in ES, a fourth approach is now underway (considered in section 10.1.1). The different land classification methods will now be considered.

4.1.1 Genetic Land Classification Approach

The genetic approach has its origins in the 19th Century when geographers, concerned with a holistic view of the terrain, acknowledged patterns of unity in the diverse landscape and, in line with the botanists and geologists of the day, began genetic groupings of natural phenomena (Mabbut, 1968). The approach aims to provide general descriptions of areas rather than precisely documenting what occurs. Two examples of this approach were Herbertson's (1905) study, which used the environmental criteria of temperature and rainfall to divide the world into natural regions, and Fenneman's (1916) study which divided the United States into physiographic divisions based on his morphologic interpretation of landforms. Both of these attempts arrived at their generalised land units by a repeated hierarchical subdivision from region down to section on the basis of inferred causal environmental factors (Mabbutt, 1968). Although the genetic approach used the same three step process, i.e. receiving stimuli, identifying parameters, and identifying suitable land classes, to evaluate land class potential; as a method it was restricted by the amounts of available environmental data at that scale and the methods of recording this data, by pen and paper. To compensate for these limitations, the genetic approach developed a hierarchical structure within which the ideal land units were designed to admit the largest possible number of general statements before details and exceptions were taken up (Fenneman, 1916). The level of generalisation or inference to determine the regions was therefore high, whilst the spatial accuracy and precision of the geographic areas represented on the maps was unknown.

A limitation of the method proposed by Mabbut (1968), related to the hierarchical structure. He stated that analysis of the landscape could only be conducted at the smallest nominated land unit, which in Fennemans' (1916) physiographic study was the section, and further analysis beyond the section would yield its constituent morphological elements. Mabbut (1968) noted that, in practice, the morphological or "atomic" components of the landscape were usually much too small to be mapped separately for any but the most detailed surveys. Today, in light of the rapidly expanding availability of large scale environmental databases (some derived from remotely sensed imagery) with detailed information regarding specific parameters such as soil and geology etc., Mabbutt's criticisms are no longer valid. Instead, a new problem emerges concerning the level of generalisation inherent in this classification procedure and the accuracy and precision of the final product.

4.1.2 Landscape Land Classification Approach

The landscape land classification approach, also known as the land system approach, was a quick, cost effective reconnaissance type survey used from the late 1920's onwards to evaluate land characteristics. As a concept, it is similar to the genetic approach as it uses a hierarchical structure and infers generalisations about relationships between parameters. There are however important differences. Firstly, it identifies, by visually interpreting aerial photographs or satellite imagery (both were considered to be revolutionary data sources when first utilised), distinctive patterns or objects in the observed environment. Secondly, it initiates an empirical study to determine the combination of factors influencing the character of each particular pattern, which is subsequently termed a land unit. It is an inter-disciplinary approach which assesses the components of the landscape and then devises land systems or associations based on which dominant component(s) are subjectively deemed as influencing the landscape character at a particular location. The approach uses the same three step decision-making process; receives stimuli, identifies the influential parameters and infers a land unit (Gibbons et al., 1968). An important assumption of the approach is that the factors composing the landscape are in a constant state of interaction, as any induced change to the unit changes its character.

Once the land units are defined they are grouped according to the landscape pattern into land associations, regions and provinces (Laut et al., 1982). This hierarchical structure differs from the genetic approach as it constructs the hierarchy from the "ground up" rather than from the region down, thus the level of generalisation is initiated at a local scale between observed empirical parameters (Figure 4.1).

At the time, the hierarchical approach was retained as it offered a relatively efficient means of storing data, an issue which had to be considered with the proliferation of environmental data. An example of this approach was Laut et al.'s (1982) study where the environmental descriptors were mapped, tabulated and described at several levels of generalization. The land units described surface water, soils, vegetative cover, and land use; the associations described mean annual rainfall and evaporation, groundwater resources and descriptors for the groups of environmental units; the regions described the climatic conditions, the monthly rainfall probabilities and monthly evaporation rates and the provinces summarized the classification. Information regarding the classification of the environmental land units and associations was presented to the user as a series of photographs and cross-sectional diagrams with associated tabular information. Documentation highlighting the rationale for the classification of each unit was not presented; instead the results were simply recorded. Similarly, information regarding the regions was presented in tabular format whilst the character of the provinces were described by a summary paragraph.

Two criticisms have been levelled at the landscape land classification approach. The first concerns the inference process. The initial identification of the land units stems from the identification of distinct patterns in the landscape. This process will vary according to the individual identifying the patterns on either aerial photographs or satellite imagery. Moreover, the process of inferring relationships between components of units and then inferring relationships between units and associations, compounds the subjectivity of the land classification procedure. Yet there is no mechanism to explain to the user why the classification decisions are made. Instead the results are presented without interpretation. The second limitation stems from recognising and defining the land components and then differentiating between them. Mabbutt (1968) argues that "the more fundamental the level at which we seek to understand the land, the more inseparable become the dynamic controls and thus to differentiate between the landscape components, ... a more detailed sample base will be called for, meaning a loss of those advantages one associates with reconnaissance survey" (Mabbutt, 1968, p. 21). This criticism is fundamentally the same as criticisms applied to the genetic land classification approach. Based on these criticisms, Mabbutt (1968) noted a new approach where "limiting values of significant attributes, rather than the complex as a whole, may provide a better basis for the classification of land" (Mabbutt, 1968, p. 21). This suggested land classification approach is known as the parametric approach.

4.1.3 Parametric Land Classification Approach

The parametric approach received greatest support from the U.S. Army Engineer Waterways Experiment Station at Vicksberg during the 1960's (Benn and Grabau, 1968; Grabau and Rushing, 1968). It's development coincided with the "Quantitative Revolution" in Geography when the improvement in computer technology gave geographers the ability to apply mathematical techniques to previously subjective problems. Geographers such as Stewart (1968) supported the trend towards quantitative models in land classification and stated, "an advantage of the deterministic and simulation models is that they should, by exposing areas of inadequate knowledge, stimulate basic research in areas of direct benefit to human need" (Stewart, 1968, p. 10). It is contended that an ES can be utilised to explain areas of inadequate knowledge.

As a concept, the parametric approach is fundamentally different from the landscape land classification approach as it is more concerned with large scale (i.e. <1:25,000) mapping. It divides the landscape, on the basis of selected attributes, into distinct data planes which are internally divided by incremental values scaling the degree to which that particular attribute contributes to the nominated land use. It therefore provides information about the land characteristics in a format, i.e. distinguishable parameters, similar to that used by humans in the decision making process identified by Gibbons et al. (1968). McHarg (1969) adopted this approach and overlaid the incremental data layers onto each other, but he did not account for variations in ecological character which render attributes in one area to be of more importance than others. To counter this problem the concept of weighting attributes to indicate their role in shaping the landscape character was developed.

There are two methods by which attributes are weighted. The first method is based on professional judgements using either empirical data or intuition which allocate numerical importance to an attribute. The second method is by numerical taxonomic means, such as paired comparisons, where each attribute is matched with each other sequentially, resulting after some manipulation in an intervally scaled set of land attributes (Lyle and Stutz, 1983). Both of these methods are a means by which environmental parameters can be systematically quantified and thus be analysed by statistical procedures. Moreover, the parametric land classification approach can be manipulated in any numerical fashion to account for inferred relationships, i.e. weighting values to express intuitive information. Therefore, it can be said that this approach attaches numeric values to qualitative information to produce a replicable method for land classification. No consideration is given to the uncertainty associated with the spatial representation of the environmental data layers, the uncertainty associated with the weighting process and the uncertainty associated with the actual land classification decision making process.

A major weakness of the parametric approach is the labour intensive nature involved with mapping environmental variables, weighting them and overlaying the data layers to determine relationships over large areas. However, with the development of GIS's this problem has been reduced.

4.1.4 Computerised Parametric Approach

GIS, as already considered, describes objects from the real world in terms of spatial location and attribute information. These software packages generally store spatial information and attribute information in separate databases, with the software package providing the links between the separate databases. Efficient data storage structures within GIS systems enable enormous amounts of environmental data to be stored. Goodchild (1993) notes that the Sequoia/2000 project based at the University of California has recently installed a 100 Terrabyte drive to handle spatially registered environmental information.

A GIS operates simultaneously on the spatial information (which can be held in either a point, line or polygon format known as vector or in a grid cell format known as raster) in association with attribute information e.g. soil colour, vegetation type. An important feature of a GIS is its ability to synthesize different data sets and create new information. It can in fact perform, via boolean overlay operations, the same functions as the parametric land classification approach. A GIS is, in a sense, a computerised parametric land classification approach.

GIS have successfully been utilised by researchers such as Compagononi (1986), Laut and Davis, (1989) and Bonham-Carter et al. (1990) to quantify, using parametric numerical weighting techniques, qualitative information into a replicable format to classify land potential. Although successful in classifying land potential, the same problems that apply to the parametric approach still apply namely, the inability to convey to the user the (un)certainty associated with accuracy and precision of the environmental data and the (un)certainty associated with the land classification decision making process.

The pragmatic need to use qualitative information is not questioned due to the spatially variable nature of relationships between parameters in the environment, i.e. at a large scale (1:2,500) within a 30m^2 grid cell of a raster GIS, quantitative empirical relationships are usually unknown between all environmental parameters as relationships vary according to spatial location. To compensate for the potentially unknown infinite number of relationships researchers infer relationships between certain nominated parameters[1].

To improve the reliability and therefore the usefulness of this land classification method, information concerning the methods used must be conveyed to any user. Currently this is not the case. A major limitation with the GIS parametric land classification approach relates to the separate output products: a GIS database, with spatial and attribute information, and a land classification map, with accompanying handbook, indicating the result of the land classification. If a "user", (defined as one who uses information derived from a land classification to assist in the decision making process), does not understand how a GIS works, then the information contained within the GIS database becomes irrelevant for that particular user. Similarly the map, without an accompanying handbook, does not explain why certain categories were identified and the handbook, when available, documents a specific methodology requiring a certain level of expertise before it can be understood. Additionally, neither of these two output products attempt to convey any degrees of uncertainty associated with either the environmental data parameters or the decision making process.

Densham and Goodchild support this notion: "For many spatial problems, Geographic Information Systems do not support decision-making effectively; analytical modeling capabilities are lacking and system designs are not flexible enough to accommodate variations in either the context or the process of spatial decision-making" (Densham and Goodchild, 1989, p. 708).

With current computer processing techniques a new, more widespread computer revolution is approaching Geography and subsequently a more informative type of land classification system using Expert Systems is being developed.

---

[1] In contrast, deterministic models, which are not scale dependent, using specific measured empirical data, have also been applied successfully to classify land potential (Millington, pers. comm., 1992)

4.2 EXPERT SYSTEMS

ES development began in the late 1960s emerging from attempts by researchers in Artificial Intelligence (AI) to build computer programmes that mimic human decision making processes. The earliest systems, such as MYCIN (Buchanan and Shortliffe, 1984; Shortliffe, 1976), were developed to solve problems in specific, narrowly defined areas such as medicine, which were well understood by a few experts. Initially, as the science of environmental problems was not well understood nor could a single expert solve an environmental problem, early ES's were not applied extensively as a tool for solving environmental problems. Hushon (1990) noted that only 21 environmentally based ES existed in 1987. Nevertheless, their usefulness is now generally accepted (Moffat, 1990; Starfield and Bleloch, 1983; Guillet, 1989; Robinson and Frank, 1987; Goodenough et al., 1987).

An ES can be defined as:

> 1) "..... a system of software or combined software and hardware capable of competently executing a specific task usually performed by a human expert" (Bowerman and Glover, 1988)
>
> or
>
> 2) " ..... knowledge-based programs with a performance, in a specific problem domain, intended to be comparable to that of a human expert" (Davis et al., 1987a).

According to Chard and Schreiner (1991) there are three types of ES: algorithmic, statistical pattern classification and production rule systems. This paper will only consider and discuss a production rule system. For a full explanation of the other two types of ES refer to Chard and Schreiner (1991).

4.2.1 Characteristics of a Production Rule System

Two distinguishing characteristics of a production rule based ES are their ability to record heuristic expertise in special data structures such as '*if/then*' production rules, rather than as mathematical models, and to store this heuristic expertise in a knowledge base data structure. The knowledge base data structure is separate from the part of the system, known as the inference engine, which infers logically valid conclusions from the knowledge base (Davis et al., 1989). Davis et al. (1987a) notes that due to these characteristics ES have been applied to problems where:

1) much of the information needed to solve a problem is heuristic (i.e. based on rules of thumb) rather than algebraic or algorithmic;
2) the information is likely to change either because of a need to explore alternative possibilities or because fresh information becomes available;
3) the information is incomplete or uncertain;
4) explanations of results/advice are required; and
5) a natural dialogue with the user is required.

There are two types of production rule ES systems: deductive and reactive. In deductive systems the *If* parts of some *If/Then* rules specify combinations of assertions, and the *Then* part specifies a new assertion to be deduced directly from the triggering combination (Winston, 1992). In contrast, in reactive systems the *If* parts specify the conditions that have to be satisfied and the *Then* part specifies and action to be taken. Sometimes the action is to add a new assertion; sometimes it is to delete an existing assertion. In deduction systems there is no conflict resolution strategy required as all rules should fire according to the *If* conditions. A deductive ES strategy has been employed by this thesis.

An ES can vary in the type of inference chaining used to solve a problem. Two predominant approaches are known as forward and backward chaining. If the search for a solution starts from a set of conditions and moves toward a conclusion, this is called forward chaining. Forward chaining starts with known data and infers all conclusions to conclude a hypothesis, thus the *If* (antecedent) component of an *If/Then* rule infers that the *Then* (consequent) part is true. Following from this, the programme looks for rules in which the *Then* condition of the first rule is an *If* condition in another rule (Hushon, 1990). Backward chaining starts with a final conclusion or goal, and attempts to determine a set of facts contributing to that goal by working in the reverse direction. It starts with rules containing the final goal in their conclusion and checks the premises in the antecedent of these rules to see if the truth of the premises can be deduced from facts in the knowledge base (Bowerman and Glover, 1988).

The criteria for deciding on forward/backward chaining strategies depends on the concepts of "fan-in/fan-out" (Winston, 1992). Whenever the rules are such that a typical set of facts can lead to many conclusions, the rule system exhibits a high degree of fan out, which argues for backward chaining. On the other hand, if the number of ways to reach the particular conclusion of interest using the facts is small, then a forward chaining strategy should be implemented. As this thesis aims to deduce particular conclusions from a large dataset the forward chaining strategy is more appropriate.

Both of these inference strategies are generally provided within a software environment known as an ES "shell". The great advantage of ES shells is that they provide a separate user interface, database and inference engine, which may, if additional knowledge is acquired, result in further rules being added more or less indefinitely to the knowledge base without altering the inference engine. Thus, unlike traditional quantitative scientific models, when new knowledge is formulated it can be easily incorporated into the knowledge base to increase the understanding about the phenomena being modelled (Davis et al., 1986). The knowledge base contained within an ES is constructed from the special knowledge of one or a number of "domain experts" who have gained their understanding from observation and experience.

Though substantial knowledge is collected in the ES production rules and is capable of being released appropriately to perform tasks, there is no ability to reason further with that knowledge as the basic semantics of the task domain are not entirely understood by the system (McKeown, 1987). This narrow characteristic is often described as "shallow" knowledge or brittleness (McKeown, 1987). It is this ability to assemble "shallow" bodies of knowledge, without any of the supporting reasoning and understanding of why the phenomena occurs, that contributes to the usability of ES as it results in qualitative knowledge, that cannot be encapsulated empirically, being applied to a problem. By adopting this conceptual stance an ES may be seen as filling a niche in situations where quantitative solutions to a problem have not yet been developed. This point is highlighted when considering how an ES can fuse information derived from different data sources, e.g. satellite imagery, airborne radar imagery, soil, and digital terrain data, into a series of *if/then* production rules that can be utilised to identify and monitor specific land features.

As ES's have evolved from research in AI they are commonly deemed to be "intelligent", but if intelligence is defined as "an attitude of the holistic functioning on human consciousness based on the subjectivity of the experiencer" then they clearly lack this ability (Bowerman and Glover, 1988, p. 39). The type of intelligence an ES does exhibit relates to behavioural intelligence, which is "behaviour in a manner that would be called intelligent had the behaviour been performed by a biological entity" (Bowerman and Glover, 1988, p. 41). This, therefore, relates more to the simulation or mimicking of behaviour. This notion is reflected in the explanation facilities of an ES that reiterate why a decision was made; i.e. once a decision has been reached, the decision is justified by noting the confidence level of the decision (a measure of the body of evidence supporting it) and listing each production rule supporting the final conclusion.

The reasoning facilities attempt to convey, within the interactive software package, a degree of understanding as to why decisions were made by the expert; therefore giving non-expert users an opportunity to understand the decision making process. This notion of attempting to explain the decision making process is crucial, as will be discussed in Chapter 10.

4.3 INTEGRATING EXPERT SYSTEMS AND GIS

The application of ES to the analysis of land suitability is a relatively new concept. However, the technique of Artificial Intelligence, especially ES and more recently Neural Networks are well documented. Chalmers and MacLennan (1990) have provided a comprehensive bibliography on ES in Geography and Environmental studies. This subsection will initially describe ES applications to environmental management; it will then consider the integration of ES and GIS and conclude with references to integrated ES/GIS systems which develop thematic land classifications.

Two early applications of ES technology to environmental management concerned fire management in National Parks. Starfield and Bleloch (1983) developed an ES to assist with determining when areas of the Kruger National Park in South Africa, should have been burnt off. Their system contained 38 production rules derived from formalising Park Manager's heuristic knowledge. Similarly, Davis et al. (1987b) describe an ES shell that was developed to aid fire management decision making in the Kakadu National Park in Northern Australia. Their system contained 105 production rules that were derived from both the heuristic knowledge of Park Managers, local Aborigines and records of fire events. A "burning index" with a corresponding index for assessing the likely ecological damage were the information outputs..

Robinson and Frank (1987) consider that there are four research areas when integrating an ES with a GIS. They are: map design; terrain feature/extraction; geographic database management and geographic or spatial decision support systems.

Map design, where the concept is to use an ES to capture expertise from a cartographer in order to produce a stylistically "good map", is an idea which initially was not successfully implemented. As indicated by negative comments regarding the formalising of the inconsistent cartographic knowledge made by Robinson and Frank (1987). However, more recent approaches have been successful. An example of a workable "smart mapper" was developed by Djokic and Maidment (1993). Their concept was based on the idea that the user had to define which of the possible graphical features were to be displayed on the map, with the ES providing the order and colours to be used to produce a consistent and legible map (Djokic and Maidment, 1993).

ES applications in the area of terrain feature extraction predominantly concern feature extraction from geographic data. Examples range from extraction of morphological characteristics such as slope, ridges etc., from digital terrain information (Hadipriono et al., 1990), to information extraction from remotely sensed data using ES techniques (Goodenough et al., 1990; Skidmore, 1990). This is an area where ES have been successfully applied to the analysis of geographic data.

In contrast, ES applications to GIS database management have achieved little success. According to Maidment and Djokic (1993) the goal of this integration is to reduce and streamline the operations that involve data queries and new data creation such as overlay or buffering analysis. In this case the role of the ES is, therefore to provide a user friendly graphic user interface for the GIS, to assist the user with their decision making. Unfortunately, Maidment and Djokic (1993) suggest that not many operational examples of this type exist.

The fourth category considers spatial decision support systems (DSS) incorporating ES. Twelve distinguishing characteristics of these systems have been listed by Wright and Buehler (1993). Essentially the difference between ES and DSS is the notion that the user of a DSS provides the methodology and experience to direct the system, rather than having it internally programmed as occurs with an ES (Maidment and Djokic, 1993). Fedra (1993) provides a number of examples of DSS where the user is required to provide information at critical stages.

The area of importance to this thesis is the integration of ES with GIS to classify land potential. By reviewing recent literature on the topic three different approaches may be identified. The first concerns the integration of ES with a GIS containing remotely sensed imagery. Skidmore (1990) used an ES to analyse remotely sensed data with other digital environmental data layers from a GIS in order to produce thematic maps of forest resources. The project utilised a raster based GIS, with the formalisms for knowledge representation and inference being prior probabilities and Bayes' Theorem[2]. The classification accuracy of the ES was 76.2%, an improvement of 25.8% on the maximum likelihood classification (Skidmore, 1989).

---

[2] Uncertainty management in this project used a so called quasi-probabilistic technique (Winston, 1992). This technique has no formal foundation in mathematics because the prior probabilities are assigned by the user and are only an indication of the true likelihood of an event occurring.

Another knowledge based image classification, which combined ancillary data, knowledge and spectral information to improve the accuracy of land cover classifications from remotely sensed imagery, was developed by Janssen and Middelkoop (1991). Their work used transition matrices to formalise the knowledge representation and Bayes' Theorem for the inference process. "Depending on the spectral discrimination, the ancillary data and knowledge, the overall accuracy of the classification increased by 4% to 20% with respect to the result based on only spectral information" (Janssen and Middelkoop, 1991, p. 3)

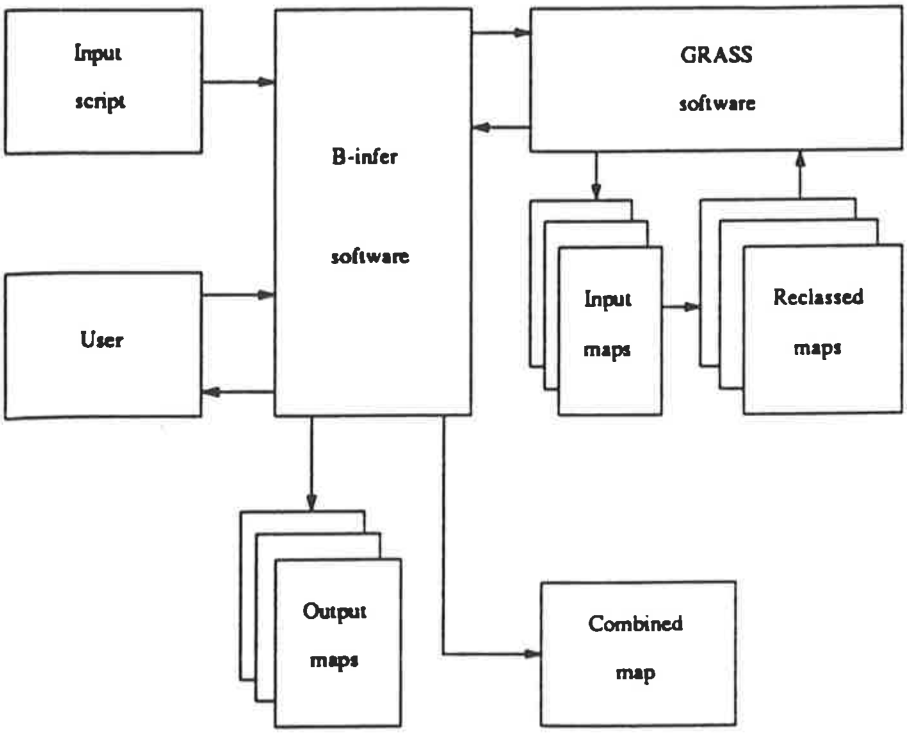
The second approach is the development of a generic integrated ES and GIS system. 'B-Infer' developed by Wright and Buehler (1993) is an example of this approach. This system utilises Bayes' Theorem and a raster GIS (GRASS) to ascertain land suitability for planning purposes. It is a generic computer system that aims to seemlessly integrate the ES and GIS computing environments in an easy to use "software package". The "focus of this research is the design and implementation of a shell development environment that can be used in the determination of suitability indices for a given land use allocation activity" (Wright and Buehler, 1993, p. 120). Central to the development of this system is the focus on determining the degree of uncertainty associated with the land classification process. Figure 4.2 displays the schematic structure of the system.

A similar system has been developed by Williams et al. (1994). Their package is able to accept image data from a GIS (GRASS), derive appropriate facts from the data and present these facts to an expert system shell. It then takes conclusions derived by the shell and uses them to construct maps of sea-ice regions which are subsequently stored back in the GIS to identify various categories of sea-ice (Williams et al., 1994). No inference mechanism assigns a degree of certainty to the classification process. Another system which does not deploy an inference mechanism has been developed by Luckman et al. (1994a).

The system developed by Luckman et al. (1994a,b) which represents the third category, uses an ES in conjunction with a GIS for automated land classification according to the New Zealand Urban Land Use Capability (ULUC) survey method.

Unlike the other two categories, there is no attempt to develop a generic system nor provide a probabilistic inferencing mechanism. Instead this system concentrates on developing an interactive system to advise the user regarding hazards affecting feasibility of urban development on different land types (Luckman et al., 1994b). In effect the system was designed to utilise the explanation facilities of the ES, but unfortunately the system was incapable of providing explanations for the capability class assessments beyond a statement of the constraint severities as assessed by the system (Luckman et al., 1994b). The authors deemed these explanations as "useful in assessing the credibility of the assessment" (Luckman et al., 1994b).

FIGURE 4.2: Schematic Representation of B-Infer



The integrated ES/GIS system developed by this thesis will enhance the explanation facilities provided by the ES in order to provide the user with an understanding of why decisions regarding certain land types were made. Stemming from this notion it will also use a probabilistic inferencing mechanism to indicate the degree of uncertainty associated with the rule formulation process. The third function of the system will be to indicate to the user the accuracy and precision of some environmental data layers held within the GIS.

4.4 SUMMARY

This chapter has conducted a review of traditional land classification techniques. Interestingly it has noted how the impact of new technologies have influenced methods used to implement the different land classification procedures. With the advent of different analytical tools, such as ES, the chapter postulates that another type of land classification method will be developed. This statement is considered further in Chapter 10.

The object of this thesis is to develop a land classification method which synthesizes remotely sensed information, geographical data and expert knowledge into an interactive software system, which can identify both existing and potential groundwater recharge and discharge areas whilst also providing an interactive capability to explain the decision making process, and the (un)certainty associated with this process, to the user.

As indicated by Chapter 2, heuristic knowledge regarding salinisation processes is well documented, yet few techniques, as indicated by Chapter 3, are available for a cost effective assessment of regional salinisation problems. Compounding this dilemma is the need to impart knowledge regarding the land classification process to the user.

The following three chapters will document the methods used in the formulation of this thesis. The chapters will cover the field data collection phase (Chapter 5), the geographical encoding of the data (Chapter 6) and the knowledge acquisition stage (Chapter 7).

Chapter 5

THE JAMESTOWN STUDY AREA, FIELD METHODOLOGY AND FIELD DATA

The town of Jamestown is located in the mid-north of South Australia, approximately 200 kilometres from the city of Adelaide (Figure 5.1). It is a highland region characterised by a series of alternating ridges and valleys trending in a north-south direction. In this area the North Mt. Lofty Ranges interface with the South Flinders Ranges. Elevations range from 420m to 660m above mean sea level. Surface drainage is predominantly from the north and is characterised by generally ill-defined surface waterways.

The area experiences a Mediterranean climate with cool to cold wet winters and hot dry summers. The average rainfall (112 years of record) at Jamestown Post Office is 460mm (Bureau of Meteorology, unpublished data). The average annual class A pan evaporation is 2369mm (estimated by the Bureau of Meteorology) and, as indicated in Table 5.1, evaporation exceeds rainfall in all months (Henschke et al., 1993).

Land in the Jamestown area was opened up for agricultural development in 1870 (Cooper, 1978). The old survey notes indicate that Gum Box, sheoak (Casuarina stricta), wattle (Acacia spp.), yacca (Xanthorrhoea quandrangulata) and spear grass occurred on elevated lands north of the town. However, the broad flat plains adjacent to the town were described as being well grassed with only a few sheoaks. The absence of tall trees may be attributed to the poor drainage characteristics and swampy nature of this land.

FIGURE 5.1: Location of the Study Region

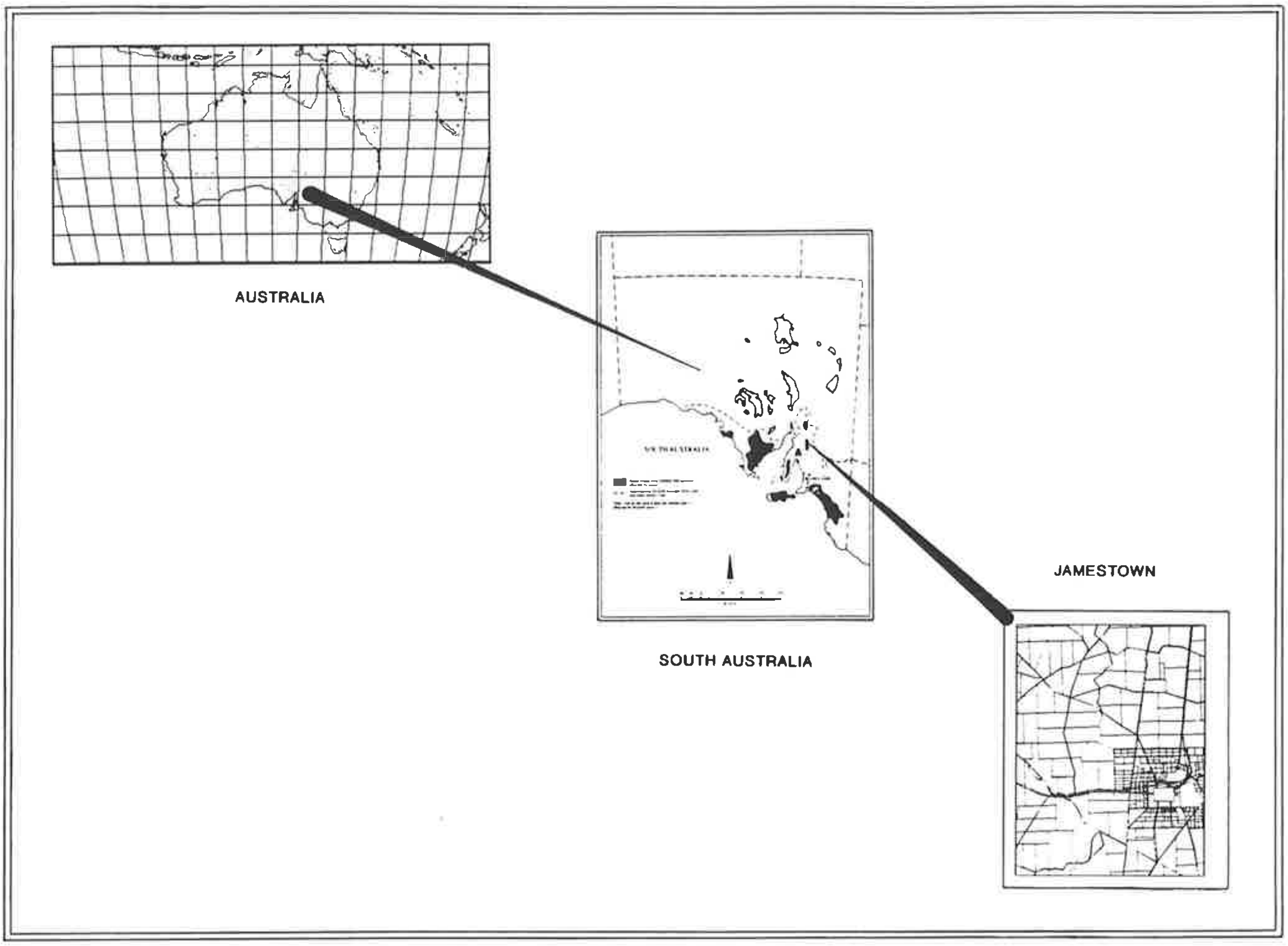


TABLE 5.1: Rainfall and Evaporation Data for Jamestown

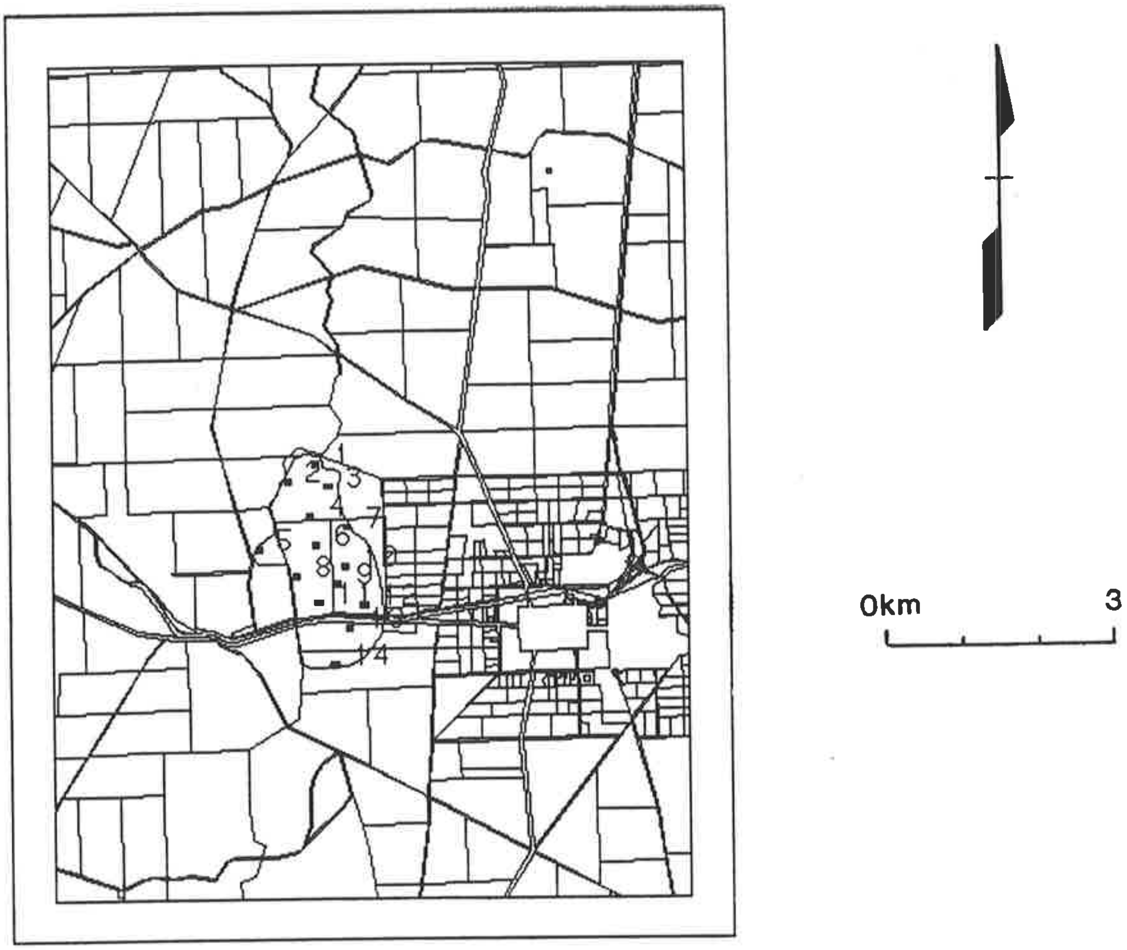
Month	J	F	M	A	M	J	J	A	S	O	N	D	Total
P(mm)	20	20	19	31	46	54	57	58	52	44	33	26	460
E(mm)	340	302	232	160	107	81	80	102	146	210	279	330	2369

Current land use comprises cropping and pasture rotations. The main crop grown is wheat with field peas becoming more popular in recent years. Sheep are grazed on stubble residues and mainly unimproved pastures. Dryland lucerne is grown along the broad valley floors. Jolly (1988) documented the history of lucerne growing in the Mid-North, noting a relationship between the decline in areas sown and the increasing incidence of dryland salinity. In recent years, salinity and waterlogging have caused a decline of many pastures around Jamestown.

Although Stephens et al. (1945) mentioned the occurrence of dryland salt scalds in the Jamestown area, it was not until the late 1970's that landholders "observed" the occurrence of dryland salinity associated with shallow water tables (Moore, pers. comm., 1993). It is uncertain if this represents a real appearance of salinity or an increasing awareness of the problem. In one particular 1x3km catchment west of Jamestown (Figure 5.2), mapping carried out in 1992 showed that 33 ha of land representing 11 per cent of the catchment was affected by dryland salinisation due to shallow watertables. The saltland was continuing to spread in a north-west direction in 1992/93.

The entire study area covers approximately 10x12km, with the 1x3km catchment to the west of Jamestown being approximately 301ha. Information regarding collection of environmental data for the study area will be considered at two scales. A general description of the parameters will be provided for each environmental feature within the region, whilst a more detailed analysis will be provided for environmental parameters occurring within the above mentioned 1x3km catchment. Economic constraints have restricted detailed data collection to within the 1x3km catchment. The majority of the detailed field data was derived from material collected during May 1991 to October 1993 by Henschke et al. (1993). The 1x3km catchment was a field site for the National Soil Conservation Program (NSCP) funded research on dryland salinisation processes. A collaborative agreement was established by the author and Henschke et al. (pers. comm., 1991) whereby the data collected by the Henschke et al. group was used to verify the accuracy of the land classification produced by the integrated ES/GIS.

FIGURE 5.2: **Location of 1x3km Catchment Subject to Intensive Field Survey**



In accordance with the findings of Chapter 2, the most influential parameters which affect the development of dryland salinisation are geology, soil, elevation, slope, hydrogeological characteristics and geophysical information, in particular electromagenetic induction (EM) techniques. Data has been collected for each of these parameters.

5.1 GEOLOGY

The generally broad north-south trending alluvial valleys of the Mid-north are underlain and flanked by folded metasediments of the Adelaidean System, including quartzites, sandstones, siltstones, slates and shales (Preiss, 1983). The regional geology of Jamestown is included on the BURRA 1:250,000, Geological sheet (S.A. Department of Mines, 1964) and in Preiss (1983).

The rocks which underlie the study region site are part of the Burra group of rocks formed in the Adelaide Geosyncline. A geosyncline is a large and often linear basin of deposition in which a great thickness of sediments has accumulated (Preiss, 1983). Typically these sediments are later folded, metamorphosed, intruded and uplifted into mountain chains. Zones of faulting also commonly develop during these sequence of events.

Structural geology of the area is dominated by an eroded anticline complex with the identified catchment area being located on the western limb of a spur anticline to the main anticline. The rocks are therefore dipping to the west (30-40^{0}).

The following formations are represented within the study region outside of the catchment:

1) Tapley Hill Formation (Pft) (Farina subgroup) (Umberatana Group): the dominant lithology of the Tapley Hill formation is a well sorted, dark bluish grey, slightly calcareous or dolomitic, often pyritic siltstone (Preiss, 1983). The grain size and carbonate content tend to increase in the upper part of the unit. Weathering tends to accentuate lamination, so that weathered outcrops commonly have a ribboned appearance (Preiss, 1983).
2) Appila Tillite (Pua) (unnamed subgroup)(Umberatana Group): earlier authors identified basal and upper greyish mudstones with abundant boulders (tillite), with respective thicknesses of 215 and 185m (Preiss, 1983). Separating the tillites is a 115m thick sequence of grey shale with occasional boulders and interbedded

brownish quartzite, grey dolomite, pebbly dolomite and thin boulder beds (Preiss, 1983).

3) Gilbert Range Quartzite (Plg) (Belair subgroup)(Burra Group): it is partly laminated and varies from feldspathic quartzite to arkose, of fine to very coarse grain size. Microcline and plagioclase have been altered extensively to clay; quartz is commonly stressed and recrystallised (Preiss, 1983).

The following formations are represented within both the study region and the catchment site. From older to younger, they are:

1) Saddelworth Formation (Pbs)(Umberatana subgroup)(Burra Group): consists of green and grey siltstones, calcareous in parts. Outcrops of this material occur along the eastern and northern boundary of the catchment. The outcrops reveal green, medium grey and dark grey laminated shale, slate and siltstone with less common dark grey dolomitic beds and laminated sandy siltstone or fine sandstone. The depositional environment was thought to be along a deeper off-shore subtidal shelf (Preiss, 1983).
2) Leasingham Quartzite Member (Pbal) (Umberatana subgroup) (Burra Group): a sandy arkose forming a prominent strike ridge along the western boundary of the catchment. In situ outcrop is found just north of the catchment where the dip of the rock is 30^0 west. It consists of fine to medium grained, feldspathic, laminated quartzite. The material is a shallow subtidal marine deposit (Preiss, 1983).
3) Mintaro Shale (Pbi) (Umberatana subgroup) (Burra Group): a grey laminated siltstone or slate occurring west of the Leasingham quartzite. Grain size varies from 0.03 to 0.3mm and mineral constituents include quartz, illite, kaolin, biotite, chlorite, dolomite, calcite, sodic plagioclase, muscovite, carbon, toumaline, apatite and rutile (Townend, 1965; Whitehead, 1978).

4) Within the valley floors superficial deposits of stream alluvium occur as recent alluvial plains (QrSPA, (Quaternary)) and (QrLAD, low angle slope deposits, (Quaternary)).

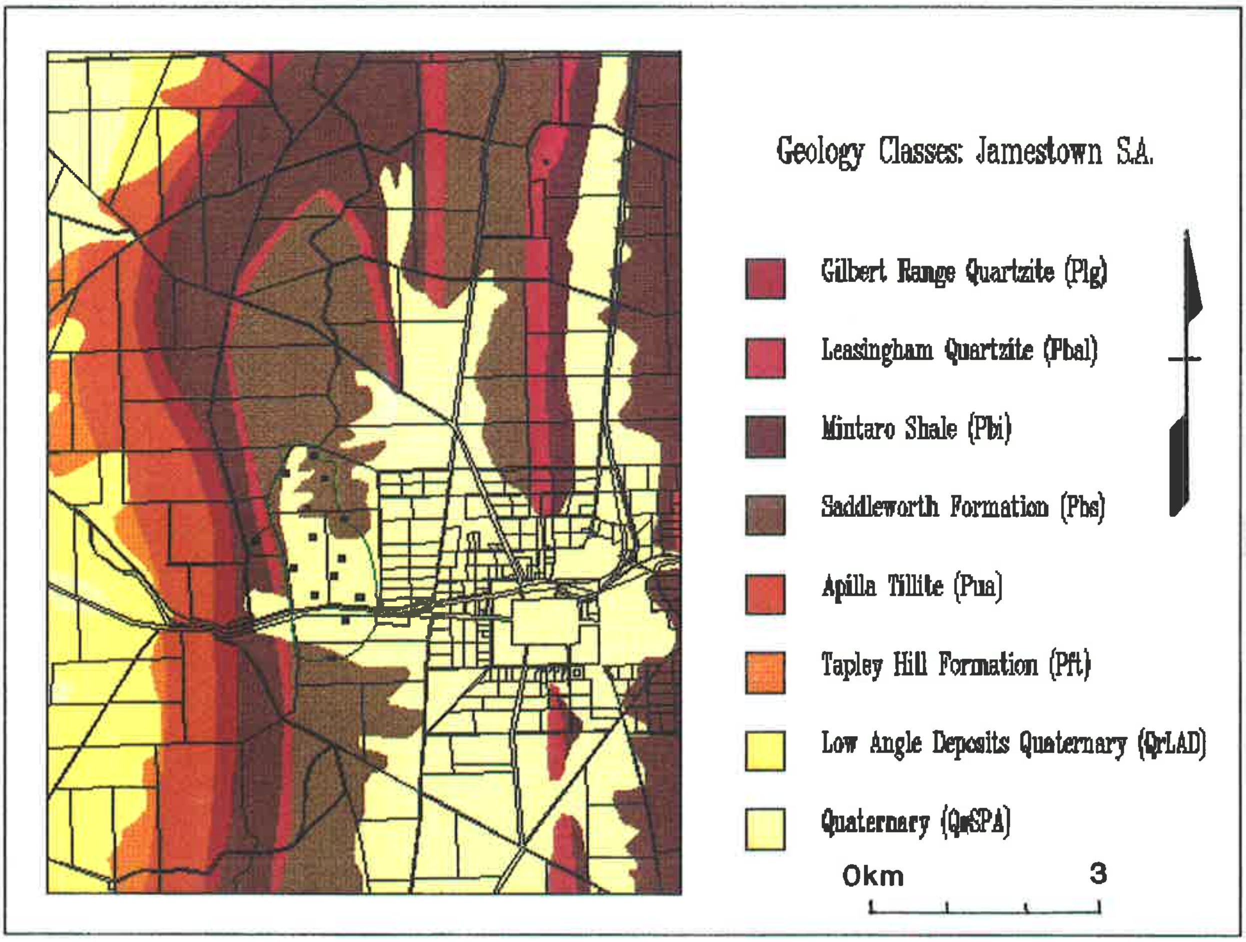
The above geological units were digitised into the Arc/Info GIS format from a mylar copy of a S.A. Department of Mines and Energy 1:50,000 Series 6631 IV (1975), first edition map sheet. Detailed description of the geographical data encoding will be considered in the following chapter. Field verification of the geologic boundaries occurred within the 1x3km catchment site to the west of the Jamestown municipality. The drilling logs and methods are summarised in Appendix 2.

Drill sites were located within the various geological units as follows: seven sites on the alluvial plains (Qr units); six sites in the Saddleworth formation (Pbs) and one site on the Mintaro shale (Pbi) (Figure 5.3). Although no sites were drilled directly into the Leasingham quartzite, one site was located on the western boundary of this formation.

According to Henschke et al. (1993), the researchers who analysed the drilling results, in general most profiles showed the following lithological characteristics:

1) a near surface clay layer consisting of hard setting top soil over soft lime, limestone or gravel lenses, underlain by mottled clays or light red to red-brown clay or tight clay. This layer was about 1m thick on the upper slopes and thickening to 5m in the valley;
2) a highly weathered zone, consisting of in-situ weathering products of the original parent material. This zone was around 10 to 15m thick and comprised pale green to grey talcy clays;
3) a weathering zone of bed-rock in a slight to moderate state of weathering. This zone consisted of around 10m of grey decomposed siltstone or shale;
4) fresh bed-rock was penetrated at two sites (3 and 7) and encountered fractures (or weathered zones) that were water bearing (Henschke et al., 1993).

FIGURE 5.3: Geology Map with Drill Sites



Drilling profiles within the alluvial plains (Qr unit) indicated the presence of sedimentary deposits including layers of plastic laminated clays and rubble/grit lenses which were encountered at various depths of up to 15m at sites 8, 10, 11, 12 and 13. This included calcareous and siliceous cemented zones. Some of this material was identified as a lacustrine algal limestone deposit of late Pliestocene age (Henschke et al., 1993).

Deeply weathered profiles occurred at sites 1, 11, 12 and 13, where hard bed-rock was not encountered by 36m. A bed-rock trough occurs in the alluvial valley between sites 11,12 and 13, and extending back up the catchment towards sites 6,4, and 1. The absence of a physical catchment divide between site 10 and south of 12, indicates an extreme erosive event in the geological past, whereby a large valley was gouged out and later infilled with sediments to form the present broad alluvial plains (Henschke et al., 1993). The bed-rock ridge/trough pattern may have some influence on groundwater flow and salinity development.

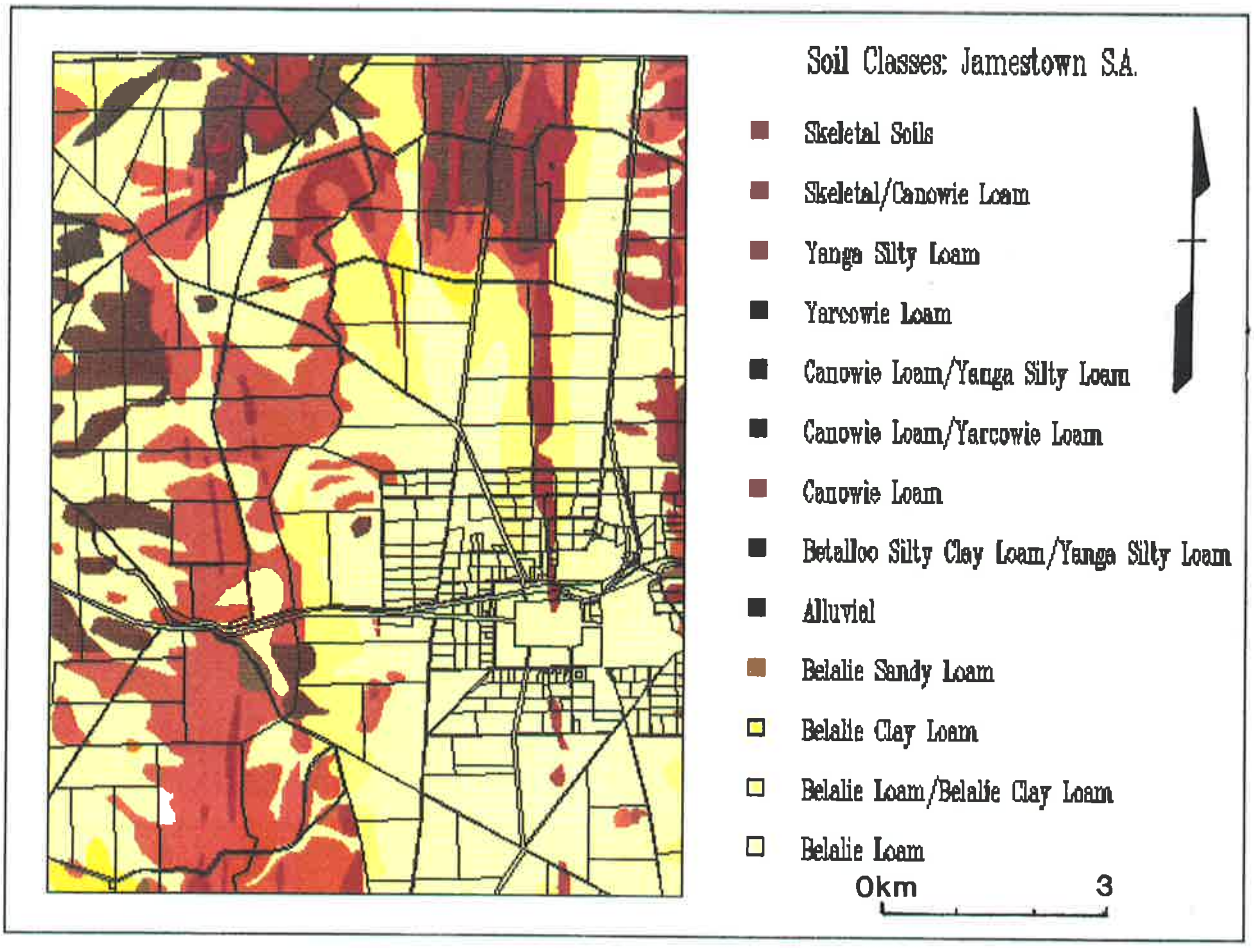
5.2 SOIL

A detailed CSIRO soil survey (Stephens et al., 1945) covering the Jamestown area at 1:50,000 was digitised into Arc/Info format. Again, a detailed discussion of this process will be undertaken in the following chapter. Two profiles were examined to 0.5-1m depth within each of the original soil mapping units for the local catchment to verify the accuracy of the map. Soil samples were collected by Henschke et al. (1993) using a hand auger and described the samples for colour, texture, structure and presence of coarse fragments. Following ground-truthing of this map slight alterations were made with the original soil mapping names being retained (Figure 5.4).

The following units are all variations of the Red Brown Earths and include the following mapping units: Belalie Loam (Bl), Belalie clay loam (Bcl), Belalie Sandy Loam (Bsl), Canowie Loam (Cl), Alluvial, Yarcowie loam (Yrl), Beetaloo Silty Clay Loam (BeSiCl), Yanga Silty Loam (Ysil) and their combinations, (Cl/Ysil), (Cl/Yrl), (BeSiCl/Ysil), (Yrl/Cl), (Bl/Bcl). The other soil type is a skeletal soil (Sk) occurring as small discrete pockets around rock outcrops, near to, or at the tops of ridges. Cracking clays associated with the Red Brown Earths occurred on the mid-slopes and flats, but were not large enough in area to map as a separate unit. A field description of soil units is provided in Appendix 3.

Belalie loam is a brown loam of usually weak structure overlying red brown clay subsoil. Moreover the subsoil overlies a deeper subsoil of light red brown clay. They have a relatively large thickness of clay and are poorly structured near the surface. This indicates a poorer infiltration capacity of these soils. Stephens et al. (1945) noted that this soil was subject to periods of inundation with some runoff. The Belalie clay loam has a description similar to Belalie loam except for the surface soil, which is a brown clay loam or light clay of moderate to good structure, occasionally self mulching.

FIGURE 5.4: Soil Map



The Canowie loam has a description similar to that of Belalie loam, but it is a shallow representative of the Belalie series developed as a residual soil on parent materials. It occurs on the slopes and ridges. This profile has a greater thickness of clay than Ysil but the soil is well structured indicating reasonable infiltration characteristics. Some surface runoff was noted by Stephens et al. (1945).

Yarcowie loam is a red brown sandy loam overlying clay loam subsoil that overlies a well defined nodular limestone layer. These soils first appear well up the slopes of the valley adjacent to the residual and skeletal soils.

The silty clay loam of the Beetaloo series is associated with highly calcareous materials that frequently exhibit crystalline structure. The series is restricted in the study area to a small area in the north west. This soil type was not drilled.

The Yanga silty loam is a residual soil consisting of light brownish grey silty loam over brown silty clay loam with slight loam and fine rubble. It is associated with the higher slopes and ridges and has a relatively thin clay B horizon. Stephens et al. (1945) observed minimal surface runoff indicating reasonable infiltration characteristics. Overall it is a soil type considered to have high groundwater recharge potential.

The skeletal soils are almost exclusively associated with the ridges which separate the valleys and generally with the top and upper slopes. Naturally, because of the parallel ridge and valley formation, the skeletal soils are distributed in elongated areas orientated like the Belalie soils in the intervening valleys in a north and south direction. They are dominantly shallow, stony soils of imperfect profile development with a large proportion of the surface occupied by out-cropping stone. The thin nature of the skeletal soils results in their contributing of high potential rates of groundwater recharge to the hydrological system.

Associated with watercourses lie areas of alluvial soils of varying depth, colour, and horizon sequence in the profile. In depth they vary from a metre to over ten metres, in colour through black, brown and grey.

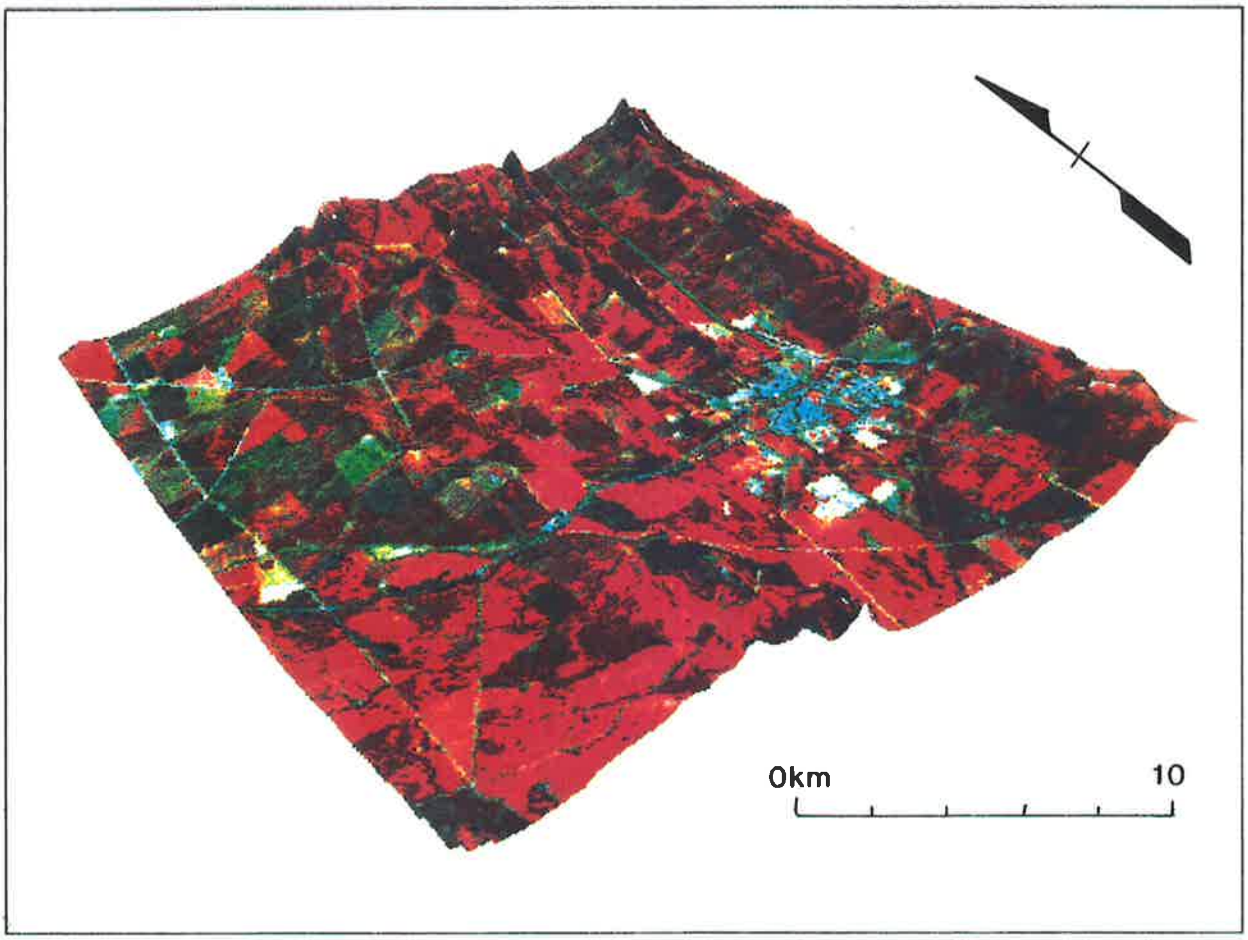
5.3 VEGETATION

There in no remnant vegetation cover remaining in the catchment area and there is very little in the region. It is unfortunate that neither the South Australian Native

Vegetation Authority nor the Jamestown council possess any information regarding the clearance history of the area. The only authority which attempts to propose a history, is the "Report of the Interdepartmental Committee on Vegetation Clearance" (1976). They suggest that the area was cleared and settled within 50 years of South Australia being founded.

Prior to clearance it is suggested by Peter Lang of the Native Vegetation Authority that the area had a sparse cover of Peppermint Gum (Eucalyptus odorata) with a stiff-matt rush (Lomandra dura) and other native grass understory (Peter Lang, pers. comm., 1991). The notion of an existing tree cover prior to human occupation is disputed by the current landholders in the region.

FIGURE 5.5: Landsat TM Image of Study Area



Current land use and vegetation cover was obtained from an enhanced October 24 1991, Landsat5 TM image (Figure 5.5).

Field work verifying specific spectral responses of various cover types was initiated on this date and two later days throughout the study area. In all 177 paddocks (selected via a random number generator), each containing 2x30m transects were analysed to verify the predominant spectral responses of the land cover. To record the radiometric readings, a Milton Series 100 Multiband radiometer was used. This hand held radiometer with four channels: 1) 400-500nm, 2) 500-600nm, 3) 600-700nm; and 4) 750-1100nm measures the spectral responses of each of the land cover types. The results were then calibrated with a known reflectance test card to determine a percentage reflectance. The results are shown in Figure 5.6. Analytical techniques used to process the remotely sensed imagery will be considered in the following chapter.

5.4 HYDOGEOLOGY

Piezometer networks, established by Henschke et al. (1993), were used to define the groundwater systems within the catchment. In addition, measurements were taken from bores throughout the region to verify the regional groundwater system. It was important to have a sufficient density of piezometers installed and bores monitored in order to determine if local or regional groundwater flow systems were responsible for the development of dryland salinity. Regular monitoring of piezometers and bores by Henschke et al. (1993) and the author was important to determine responses of the groundwater to rainfall events and the recharge potential across the landscape.

5.4.1 Piezometer Installation and Regional Bore Selection

Piezometer sites were selected by Henschke et al. (1993) so that a groundwater contour map and hydrogeologic cross-sections for the 1x3km catchment could be drawn. Sites were also selected to characterise groundwater systems beneath the various soil-landform associations in the catchment.

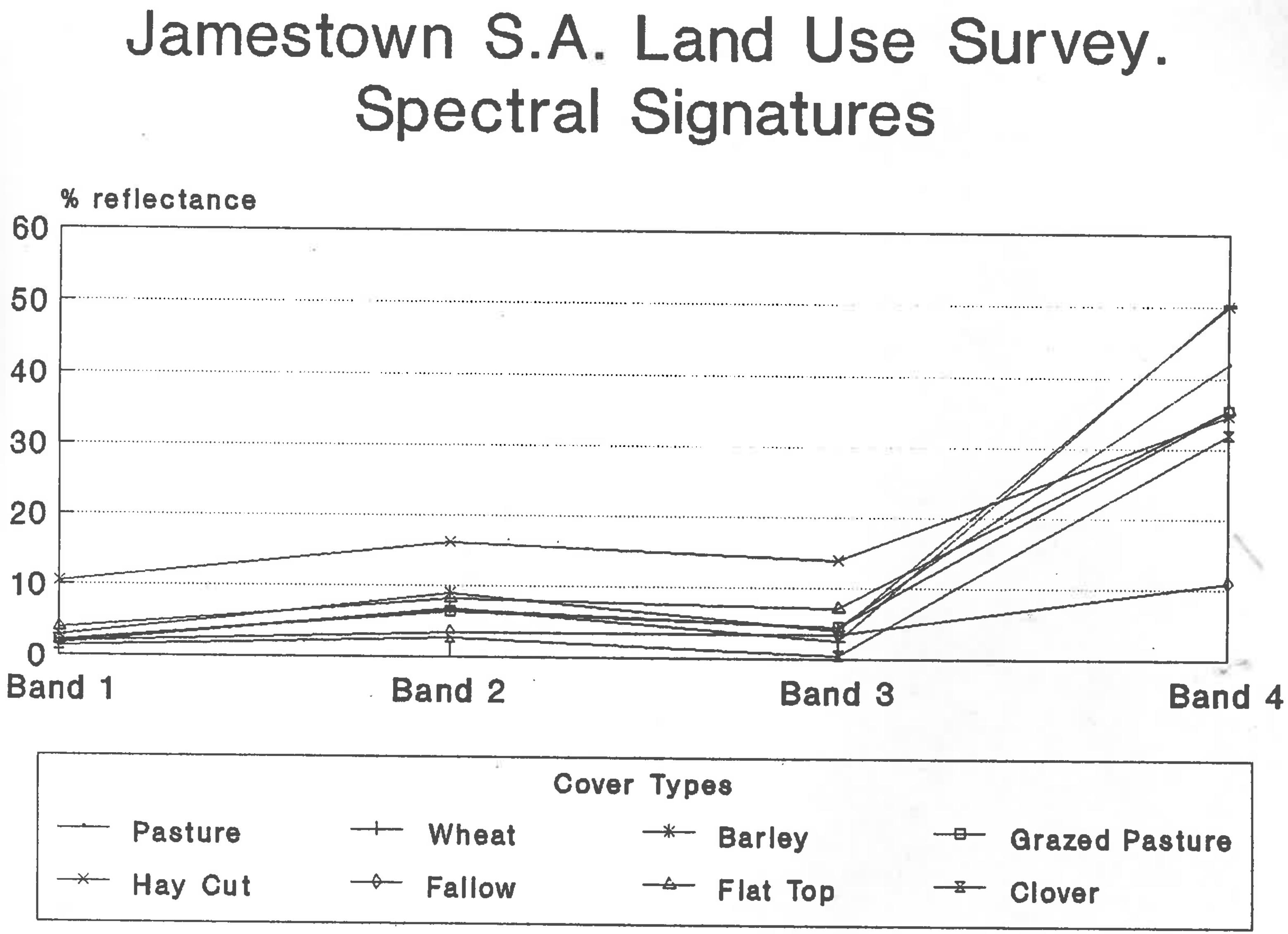
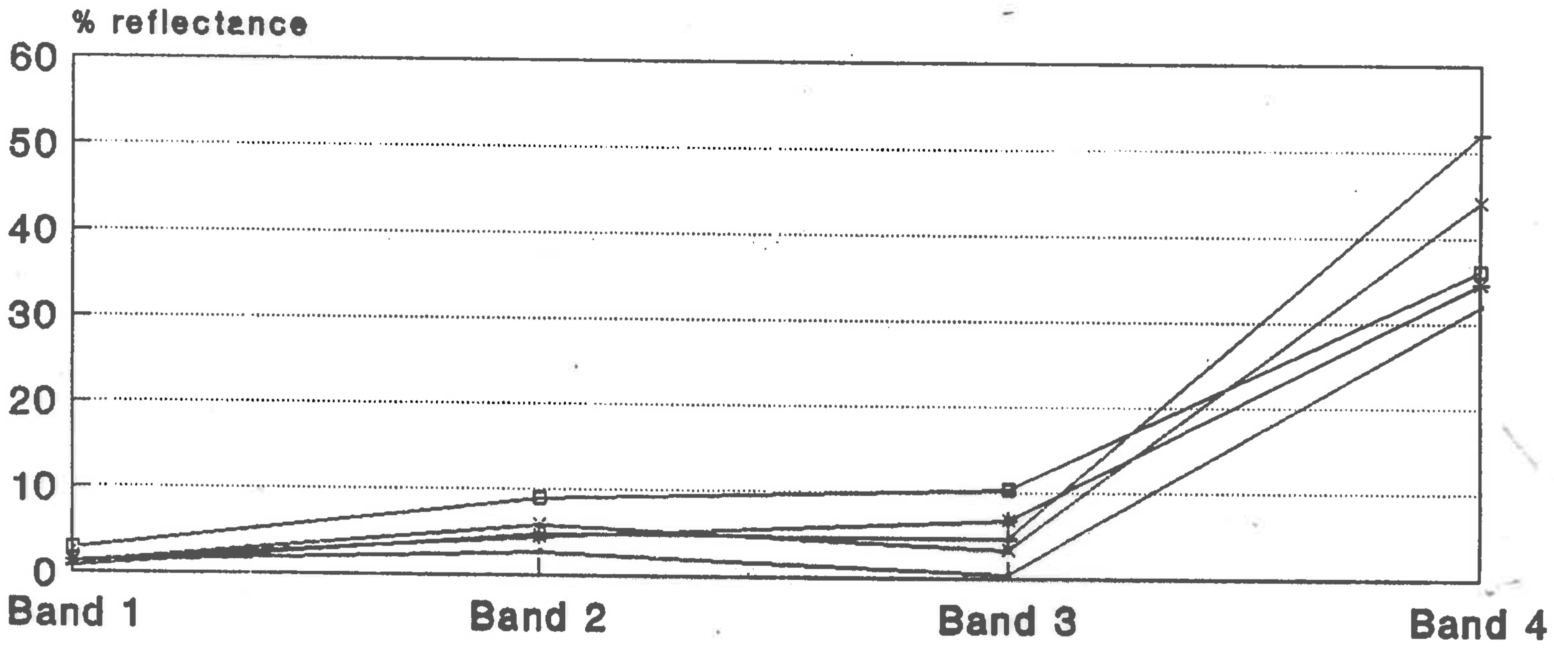


FIGURE 5.6: Spectral Responses of Land Cover Types


% reflectance
60
50
40
30
20
10
0
Band 1
Band 2
Band 3
Band 4

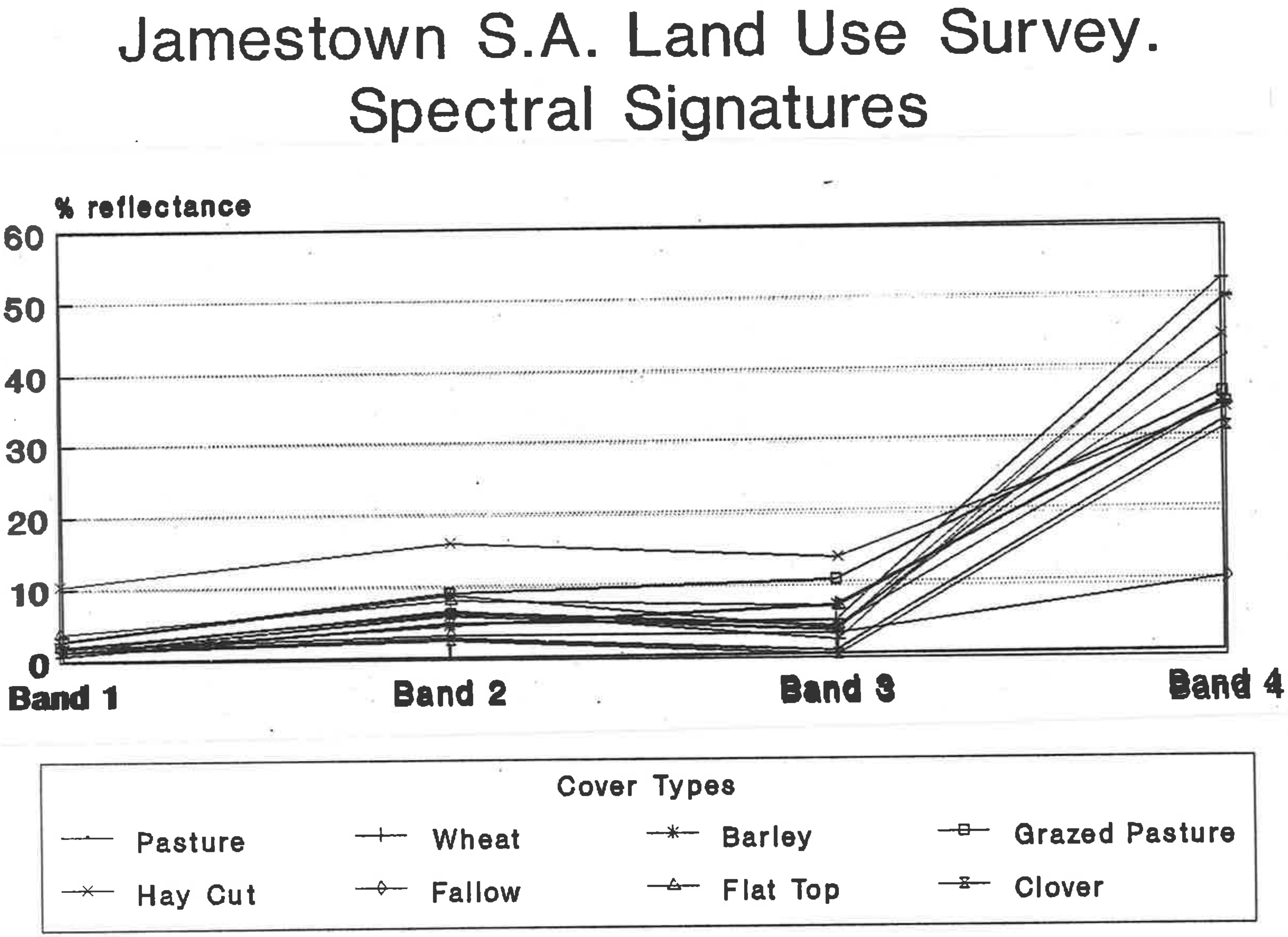


FIGURE 5.6: Spectral Responses of Land Cover Types

Piezometers were installed at 17 sites with multiple completions (nests) at 8 of these sites. Piezometer nests enable vertical gradients, and hence direction of groundwater flow, to be determined at each site. This information reveals whether the site is a recharge or discharge area (Henschke et al., 1993). A shallow observation well to 1-2m depth was installed at all sites to determine the presence of ephemeral watertables on hillslopes and the depth to shallow permanent water-tables in the valley.

Henschke et al. (1993) constructed the piezometers from 40mm of 50mm ID class 9 PVC casing with a bottom cap. The casing was slotted (hacksaw blade cuts) over the bottom 0.5 to 1.0m and lowered into the borehole. Washed and graded (8/16) sand was poured down the hole to fill the annular space over the slotted zone. Bentonite (granular or pelletized) was tipped on top of the sand to form an effective watertight seal above the screen. Holes were backfilled with a drilling spoil and a cement collar placed on the surface. Tubes were also reinforced above ground as protection from stock damage. This installation process was completed by Henschke et al. (1993). Completion details for each site are given in Appendix 4.

Within the region, bore sites were selected to produce a regular distribution across the landscape. All piezometer and bore hole locations were located onto a 1:10,000 ortho/topo/cadastral map using a Garmin GPS unit, by the author. Most bores were already marked as symbols on the topographic map but piezometer elevations required accurate surveying from a known benchmark, as Z values derived from the GPS unit were not considered to be reliable.

5.4.2 Monitoring Methods

Manual observations of standing water level were carried out by Henschke et al. (1993) on a six weekly basis for the piezometers from November 1990 to May 1993. Wesdata single channel loggers and capacitance water level probes were installed by Henschke et al. (1993) at site 8 in June 1992 and at sites 4, 7 and 14 in April 1993. The loggers were programmed to record a water level change of 10mm. Water level probes record short term responses in water levels which could easily be missed from monthly monitoring. Short term fluctuations help to determine mechanisms of recharge (ie. macropore flow), for example Jolly (1988) observed spikey responses in piezometers located near macropore channels, indicating a response time of a few hours.

Regional bore hole measurements were completed by the author and Henschke, over 4 days starting from 21st May 1992. Standing water measurements were taken whilst the bore was not in action. Depth to groundwater levels were determined by passing a hollowed copper weight with an attached tape measure down the bore hole cap until it reached the groundwater surface. Groundwater depth was ascertained from the tape measure. Because the groundwater system could be treated as a single aquifer, construction details of farm bores and wells was not as critical for construction of groundwater contours. Interpolation procedures utilised to derive the groundwater contour map were initiated in the Arc/Info GIS. Detailed explanation will be considered in the following chapter.

5.4.3 Hydraulic Conductivity

Slug tests were carried out by Henschke et al. (1993) to obtain estimates of the saturated hydraulic conductivity of aquifer materials adjacent to piezometer screens. A Unidata pressure transducer was lowered by Henschke et al. (1993) into the piezometer to 4-4.5m below the standing water level. An airline was connected between the top of the piezometer and an air compressor. The water level was lowered 2 to 4m by pressurising the headscape with the piezometer.

Hydraulic conductivity was calculated using the method formulated by Hvorslev (1951) - a summary of this method appears in Freeze and Cherry (1979). The following equation was used to calculate saturated hydraulic conductivity (K).

$$K = d^2 \ln (2ML/D)/8LT$$

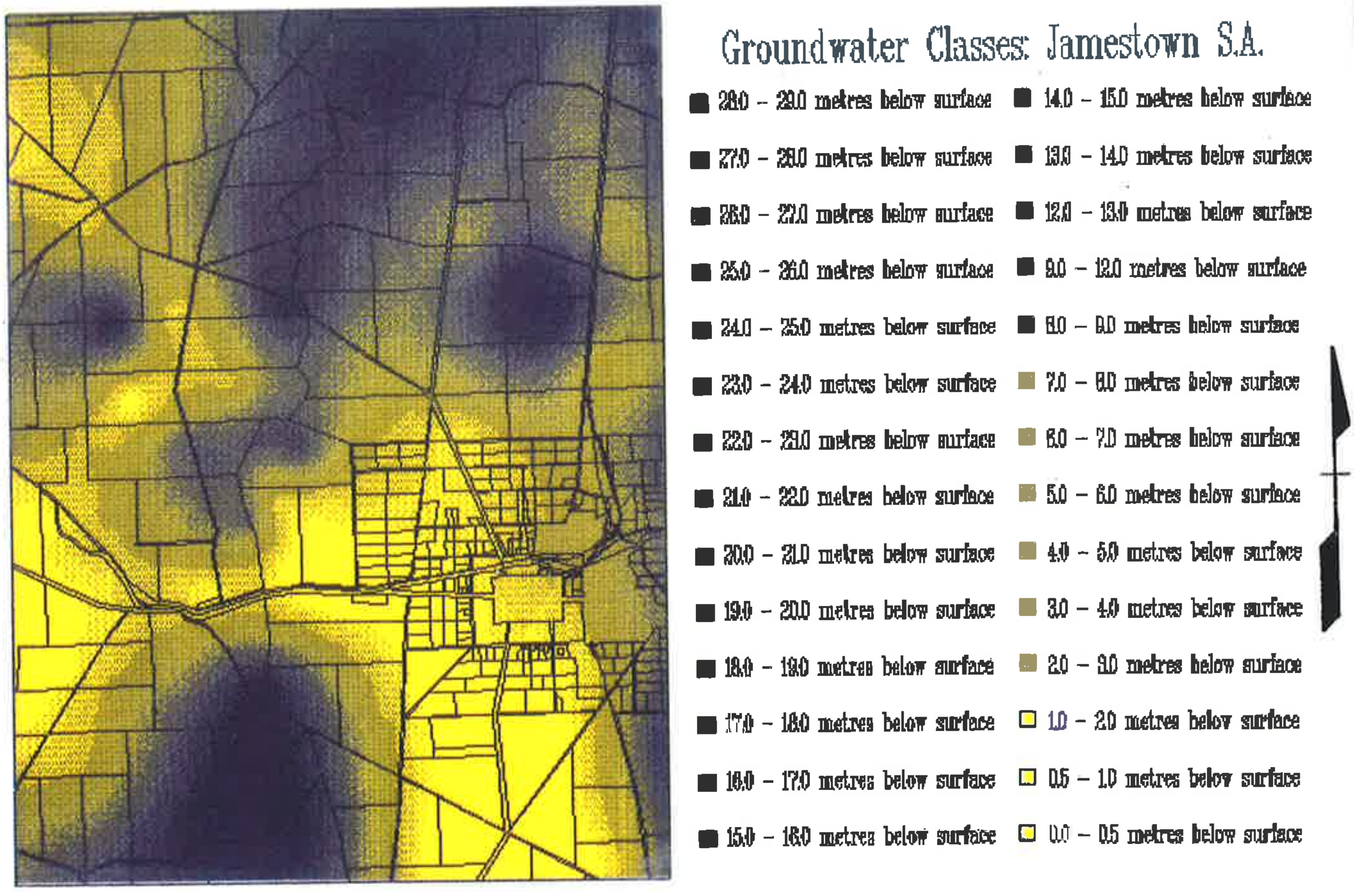
where d = diameter of piezometer, M = transformation ratio = 1, L = screen length. D = diameter of screen and T = basic lag time. A semilog plot of head ratio versus time was used to calculate the basic time lag (Henschke, et al., 1993). The basic time lag is the time at which the head ratio is 0.37, indicating that the water level has recovered by 63%.

5.4.4 Observations

Shallow observation wells installed at 1-2m depth rarely showed water table development except at sites on the valley floor where a shallow permanent water table was present. Development of winter perched water tables on the slopes was virtually nonexistent because of the absence of texture contrast soils in the catchment.

The Leasingham quartzite member on the western boundary of the catchment was considered to be a groundwater divide for the reason that this formation is not a pure quartzite but has a feldspar content of around 7%. Weathering products will therefore be clayey and this will tend to seal fractures, thereby reducing transmissivity of the formation (Henschke et al., 1993).

FIGURE 5.7: Regional Groundwater Map



Groundwater boundaries did not coincide with surface divides, particularly along the northern and southern divides of the catchment. This phenomenon thereby supported the development of a regional groundwater map (Figure 5.7). Interpolation techniques utilised to generate the regional groundwater map will be considered in the following chapter.

5.5 SLOPE AND ELEVATION

Slope (Figure 5.8) and elevation (Figure 5.9) data were derived from Digital Elevation Models (DEM) developed within the Arc/Info GIS. The original spot height and 2m contour information was provided by the South Australian Department of Lands. Discussion of DEM development will be considered in the following chapter.

5.6 GEOPHYSICAL SURVEY

Geophysical methods are used to indirectly infer the physical properties of surface and sub-surface features of the landscape. These methods have found application in dryland salinity investigations by providing information on sub-surface features that may influence development of salinity.

The geophysical survey carried out in the 1x3km Jamestown catchment area by Henschke et al. (1993) concentrated on surface electromagnetic induction (EM) using a Geonics EM31 conductivity meter. Electromagnetic induction techniques measure bulk electrical conductivity of the soil profile and the response is due to a number of soil properties including salt storage, moisture content and clay content. The surface EM31 survey was carried out to map the spatial variability of bulk electrical conductivity of the landscape and hence infer variability of potential groundwater recharge (Henschke et al., 1993).

FIGURE 5.8: Slope Categories for the Jamestown Study Region

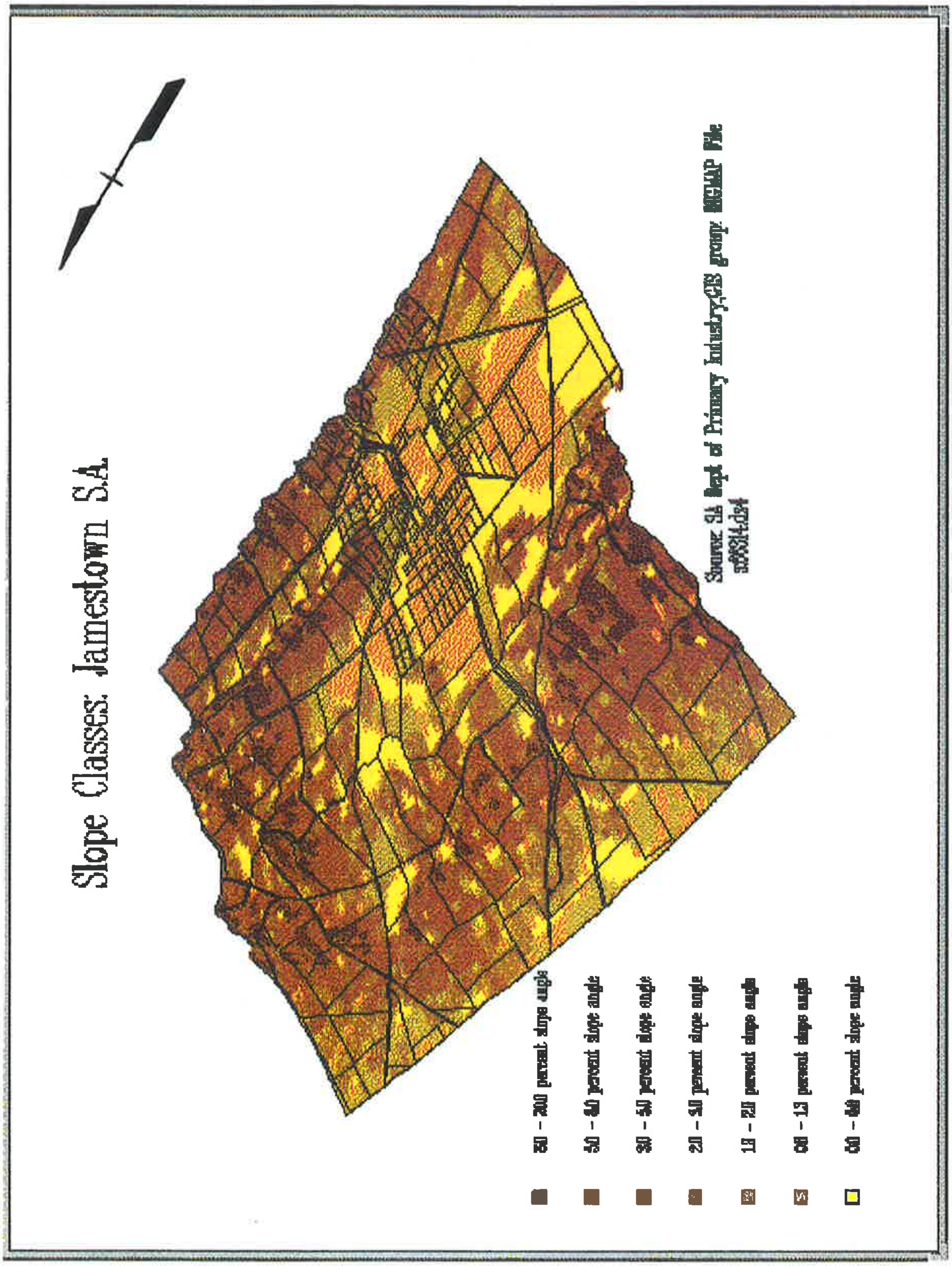
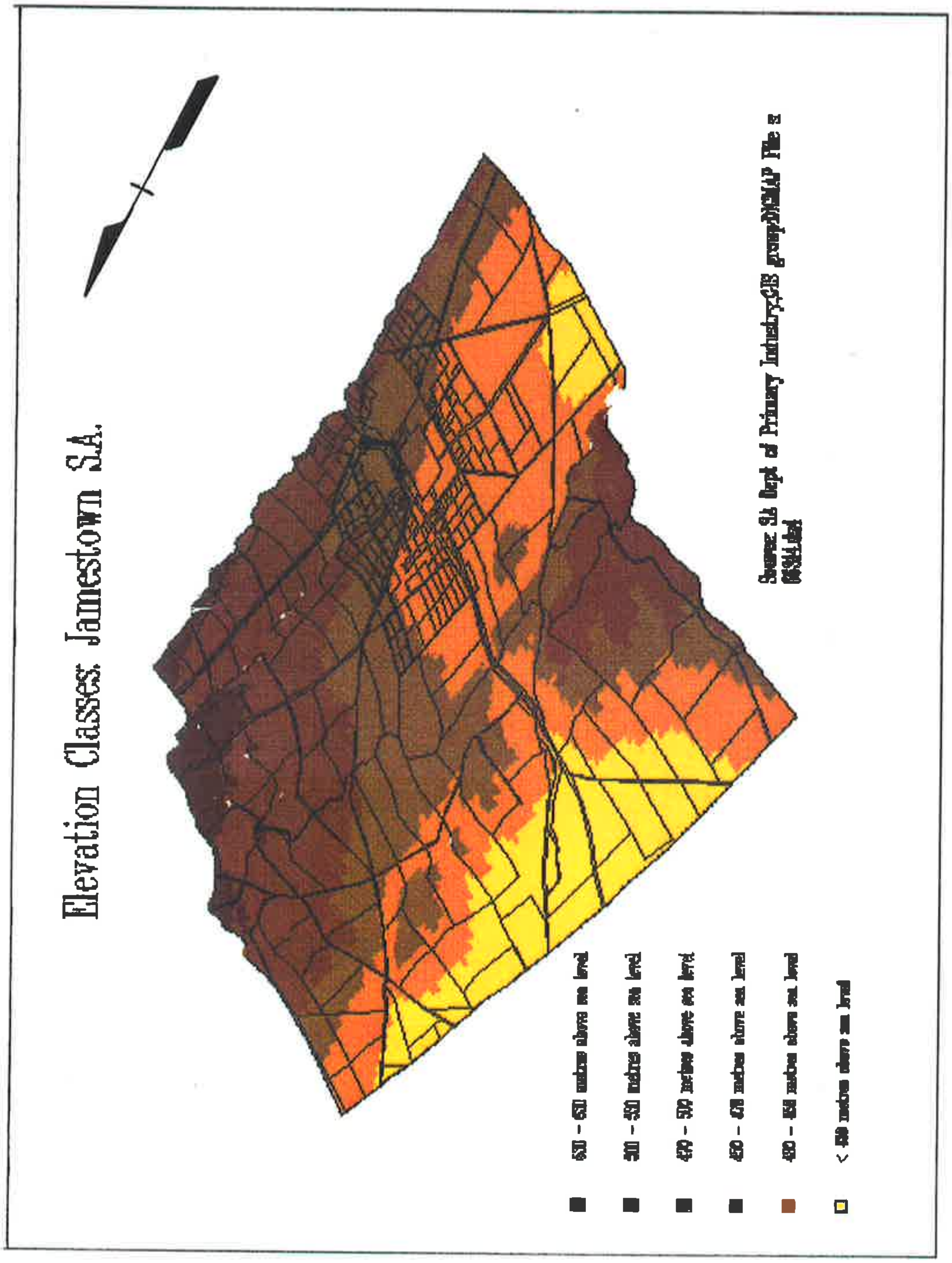


FIGURE 5.9: Elevation Within the Jamestown Study Region



Source: SA Dept of Primary Industry,GIS group/DIGMAP File = 06314.dat

EM induction can be used to define high recharge areas in the landscape. The method is based upon the assumption that piston flow (matrix-flow) is the predominant mechanism of recharge (Henschke et al., 1993). High recharge areas are associated with low clay contents and low salt storage (indicating vertical flushing of salts) and will therefore be reflected as zones of relatively low apparent electrical conductivity (ECa) in the landscape. Low recharge areas on the other hand will show a relatively high ECa due to high salt storage and high clay content.

Studies in the Mallee region of South Australia (Cook et al., 1989), indicated that variability in ECa was a function of depth to the chloride front which reflects recharge rate at that point in the landscape. Recharge rate was also found to be dependent upon clay content in the top 2m of the soil profile (Cook et al., 1992).

EM instruments which "sense" shallow rather than deeper are preferred for recharge are mapping. The EM31 senses to a depth of around 6m and was therefore seen to be the most useful instrument for this study.

The EM31 survey was carried out in January 1992 by Henschke et al. (1993) using a Geonics EM31 meter.[1] Eight traverse lines, spaced at 200m intervals, with readings taken at 20m intervals were established. In total 894 points were sampled from within the 1x3km catchment. Measurements were taken in both the vertical dipole mode at approximately 1m above the ground surface and in the horizontal dipole mode at ground level. No EM31 data was collected outside of the catchment area. This data collection stage was completed by two people over three days.

---

[1] A Geonics EM31 meter produces a low frequency alternating current in the transmitter coil which generates a primary magnetic field and this induces a small current in the soil (Henschke et al., 1993). This current then induces a secondary magnetic field in the ground and the ratio between the two fields determines the bulk electrical conductivity in milli-siemens per metre (mS/m). Depth of signal penetration can be altered by varying the height of the instrument above the ground and by taking measurements in the horizontal or vertical dipole mode (McNeil, 1980).

FIGURE 5.10: EM31 Horizontal Dipole

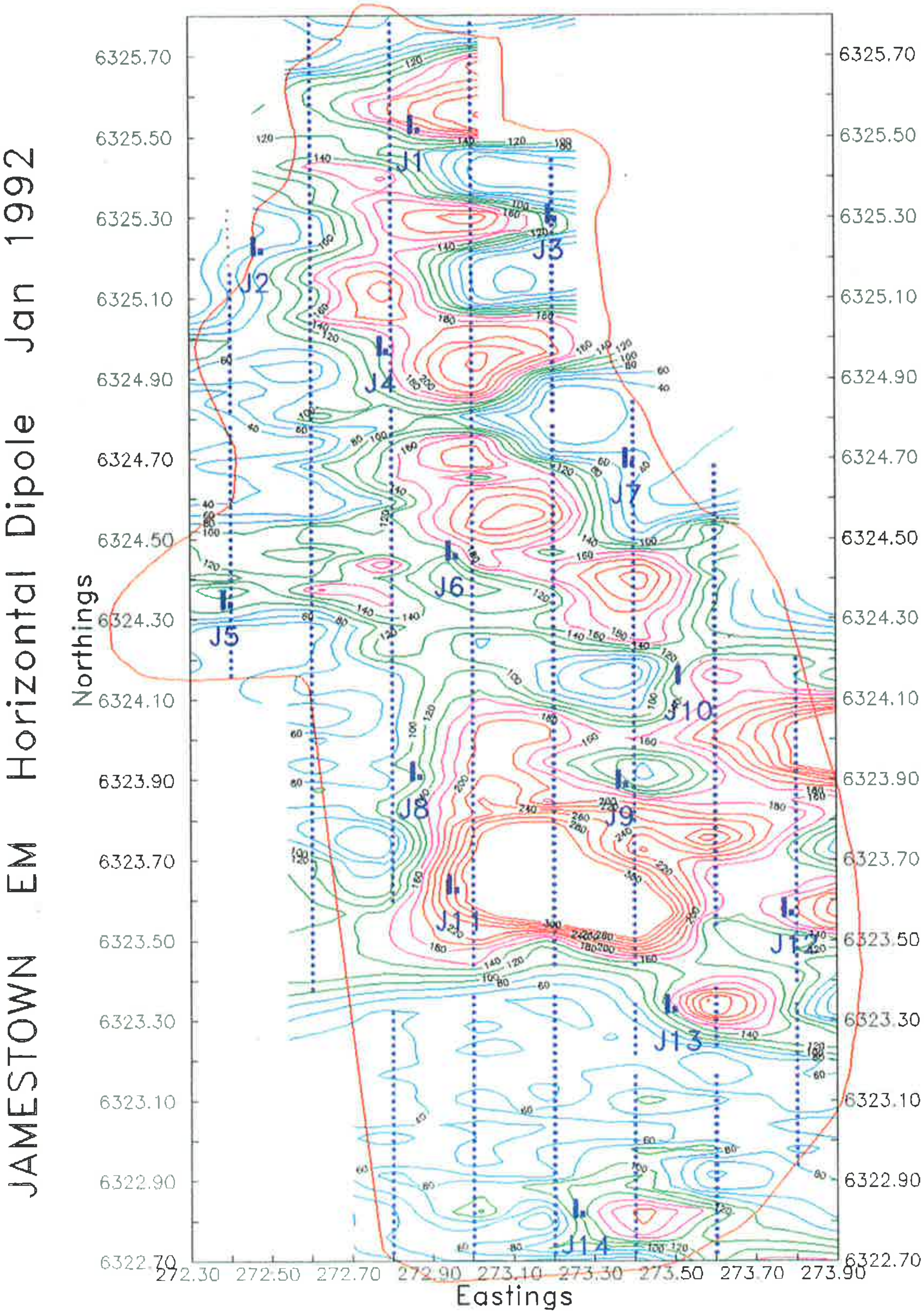
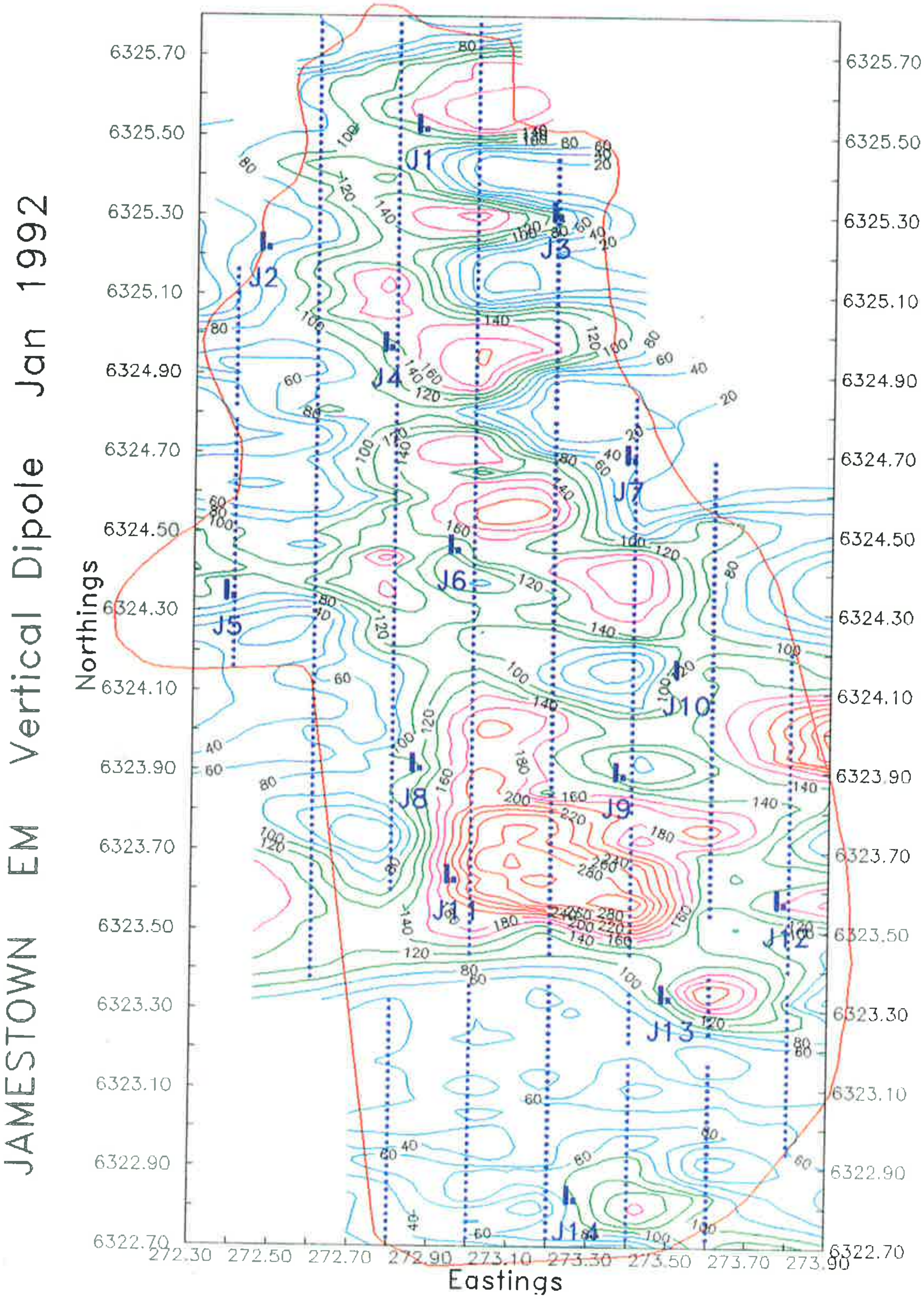


FIGURE 5.11: EM31 Vertical Dipole



EM conductivity maps for both horizontal (Figure 5.10) and vertical dipole modes (Figure 5.11) were generated using kriging interpolation techniques by Henschke et al. (1993) and provided in a digital format to the author. The geostatistical technique kriging, will be considered in the following chapter.

The following zones were identified based on the EM31 horizontal mode by Henschke et al. (1993):

1) very low conductivity areas (0-40 mS/m): occurred as small discrete elongated pockets mainly around the margins of the catchment especially on ridges where depth to bed-rock is shallower;
2) low conductivity areas (40-80 mS/m): a broad area occurs south of the highway/railway. Other areas occur along the western, eastern and northern boundaries;
3) medium conductivity areas (80-160 mS/m): the main areas being south of site 4 and site 6 and west of site 10 in the catchment;
4) high conductivity areas (160-200 mS/m): occurred through the centre of the catchment, bounded by sites 8, 9, 12 and 13 and also as a lineament extending from south-east of site 10 back up the catchment through sites 6 and 4 toward site 1;
5) very high conductivity areas (>200 mS/m): occurred as small discrete pockets in the discharge areas, near sites 9,11,12 and 13, and up along the centre of the catchment towards site 1.

5.7 SUMMARY

This chapter has provided a description of the environmental parameters collected and analysed for this project. Notably the detailed knowledge derives from the 1x3km catchment west of the Jamestown municipality. Financial constraints prevented further detailed studies being conducted within the region. This problem has resulted in the assumption that the conditions identified in the 1x3km catchment are representative of the relationships existing in the 10x12km region. Given the intensive nature of the data collection process and the precision to which the field collection strategy was implemented this assumption appears to be sound.

The benefit of the collaborative field work to this thesis lies with the identification of the groundwater recharge and discharge areas within the Jamestown study area. It will enable independent verification of the land classification map produced by the integrated ES/GIS.

The next chapter considers the database design and the processes involved in the capturing and converting environmental data into digital information.

Chapter 6

BUILDING THE GEOGRAPHIC DATABASE : DATABASE DESIGN, DATA CAPTURE AND CONVERSION

The construction of the digital geographical database for the Jamestown region commenced in 1991. In this chapter, the rationale for the database design and the techniques employed for geographical data encoding are examined. A description of the software modelling environment is provided in Appendix 1.

6.1 DATABASE DESIGN

Effective spatial database management evolves from well conceived plans. This is necessary as a spatial database should be designed to account for both present and future geographic applications. Zhou (1989) notes that a design of conventional databases includes the following important phases:

1) data definition: defining the kinds of information that must be stored in the database;
2) data refinement: refining necessary data to create accurate descriptions of the types of data that will be needed in the database; and
3) establishing relationships: between data attributes based on the likely query which will be applied to the database.

With the additional functionality of a spatial database, three other questions require answering:

1) What spatial data structure should be employed? This is the most common question when designing a spatial database. The answer to this question is dependent upon the proposed database application (Zhou, 1989).
2) What spatial resolution should be maintained? The answer to this question is largely dependent upon the spatial data accuracy of the source, as well as the requirements of proposed applications (Zhou, 1989).
3) What spatial referencing system should be used? This is often related to the choice of a geographical coordinate system such as map projection. The answer to this question is often dependent upon the areal extent and location of the geographical area of interest (Zhou, 1989).

6.1.1 Spatial Data Structure

A raster data structure is employed for the following reasons:

1) Soil, geology, groundwater depth, slope and elevation do not occur in the real world as discrete units where one data type is clearly distinct from another. Rather they exist as a continuum where there is a gradual change from one data type to another with the boundaries between the two being unclear. Raster data structures provide the ability to represent these continuous surfaces.
2) This project develops a land evaluation method, which relies heavily on Boolean and map overlay operations. Each gridcell in the raster file contains environmental information which can be analysed by the ES Thus in a sense, all data layers are overlayed and subsequently analysed.
3) Remote sensing provides the major data source to identify existing saline discharge. As this information is stored in its original raw form, the use of a raster based system enables integration with other geo-based information.

Although the raster data structure has been chosen as the primary data structure to form the basis for data processing and modelling, vector based spatial data are also kept in their original form and maintained by the Arc/Info GIS. The vector information retained includes, road, cadastral and drainage information.

6.1.2 Spatial Resolution

The choice of a raster data structure as the primary data structure leads to a decision as to which spatial resolution, or in other terms the grid cell size, is more appropriate? The choice of a gridcell size for the raster system is one between small cells which result in large data volumes, or using larger cells which incurs a loss of detail (Zhou, 1989). A 30m^2 gridcell size has been chosen as the primary spatial resolution based on the following considerations:

1) *Spatial resolution of the source data*: after detailed analysis of environmental data parameters occurring within the catchment: soil, geology, groundwater and EM31 data, the data sources on the sub-regional 12x10km scale were considered to be accurate enough to allow the formulation of raster coverages with 30m^2 gridcells. The Landsat TM remotely sensed data also has a 30m^2 spatial resolution. Thus the integrity of the Landsat TM spectral response was maintained by retaining a 30m^2 gridcell size.
2) *Requirements for the application*: both existing and potential salt affected groundwater discharge areas, and potential groundwater recharge areas are the features of interest in this study. They can be accommodated by a 30m^2 gridcell in the Jamestown region. Additionally, paddock sizes in this region are not large, as evidenced by the regional cadastral information.

6.1.3 Spatial Referencing System

The Australian Map Grid (A.M.G.) (Australian modification of Universal Transverse Mercator (U.T.M.) map projection system) has been chosen as the primary map projection.

6.2 DATA CAPTURE AND CONVERSION

Data capture and conversion has been conducted on a Sun IPX workstation and the MicroBRIAN image processing system. The Arc/Info GIS software has been used as the tool to digitise and edit maps, to convert them into raster data format.

6.2.1 Original Data Sources

The original source data for this study derives from thematic maps, digital information, point data collected in the field and Landsat TM imagery. The GIS coverages representing this information are, as described in Chapter 5: cadastral, roads, drainage lines, soil, geology, groundwater, EM31, elevation, slope, and an enhanced Landsat TM image indicating areas of salt affected land.

6.2.2 Thematic Maps

Thematic maps including, soil and geology were digitised using Arc/Info map digitising and editing routines. After data capture this data was initially stored in the Arc database as vector arcs and then at a later stage converted into grid format.

The conversion process used in this project relied on the "direct overlay" method. This method is based on the assumption that map themes digitised from source data represent a non-continuous surface in the world space (Zhou, 1989). Thus map units (or polygons) have no numerical relation to each other and the attribute values within a map unit are homogeneous (i.e. surface variance is nil within individual polygons)

(Zhou, 1989). This assumption, due to the original source data, is appropriate for the soil and geology map themes.

6.2.3 Digital Data

Data provided in Arc/Info digital format at 1:25,000 by the South Australian Department of Primary Industries includes: cadastral, road, drainage line and 2m contours plus spot height information. The spot height and contour information was used to establish a Digital Terrain Model (DTM) from which slope and elevation GIS coverages were generated. EM31 data was provided in digital Arc/Info format by the South Australian Department of Primary Industries.

6.2.4 Digital Terrain Models

A DTM has been defined as "any digital representation of the continuous variation of relief over space" (Burrough, 1986, p. 39). Over the past 15 years a variety of methods have been developed to encode terrain. Although the terms DEM and DTM are often used interchangeably; for the convenience of the discussion here, the term DEM is preferred for models which contain only elevation data, while the term DTM is used in reference to all spatial models derived from a DEM relating to attributes of a landscape other than the altitude of the surface. The term digital terrain model was originally coined by Miller and LaFlamme (1958).

There are two major methods by which to encode the terrain: mathematical models (using coefficients of polynomials or trigonometric series, approximating relief as smooth surfaces) or image methods (Dutton, 1984). The most common mathematical methods are based on continuous three-dimensional functions that are used to model terrain relief. To obtain the best generalised trend surface the function must have the maximum number of terms possible, thus there must be more point data than parameters. The accuracy of the generalised trend surface modelling is therefore very dependent on the amount of point data and the type of function used. Polynomial and Fourier series demonstrate high generalisation capabilities and have been used extensively for DTMs (Clarke, 1990).

The procedures for mathematical surface fitting fall into two main categories: Global and Local methods. Global methods model the entire distribution of the surface, whereas Local methods split the surface into small regular or irregularly shaped areas and work on each area at one time (i.e. the point elevations/observations within each area are used to fit the mathematical model for that area) (Burrough, 1986).

The major disadvantage of these methods is the reliance on the assumption that the point elevations are continuously distributed. This is not always the case for complex surfaces such as the earth's terrain. In cases where true representative models of terrain are required, mathematical methods are undesirable, due to the unacceptable errors generated during generalisation (Clarke, 1990).

There are two general types of image methods: gridded terrain models, known as lattices in Arc/Info, and triangulated irregular networks (TIN). A lattice is the surface representation of a grid, represented by equally spaced sample points referenced to a common origin and a constant sampling distance in the X and Y directions. Each mesh point contains a Z value (point elevation) of that location. The lattice resolution, that is the number of points in X and the number of points of Y, determines the accuracy of the lattice. The higher the resolution the more accurate the surface representation. Surface Z values of locations between lattice mesh points can be approximated by interpolation between adjacent mesh points. There are a variety of interpolation algorithms that are used, most of which are based on inverse distance mathematical theory and work as local operators (Burrough, 1986).

There are three interpolation techniques for lattices within Arc/Info: nearest neighbour assignment, bilinear interpolation and cubic convolution. Nearest neighbour assignment assigns the value of the nearest mesh point in the input lattice or grid to the output mesh point or grid cell. It is not a true interpolation technique as a value of 2 in an input grid will always be the value 2, never a 2.3 or 2.4. Bilinear interpolation computes the output mesh point value from the values of the four nearest input mesh points, based on the weighted distance to these points. Similarly cubic convolution calculates the output mesh point value in the same manner, except that the weighted values of the sixteen nearest input mesh points are used. Using this expanded neighbourhood the cubic convolution algorithm tends to smooth the data.

The alternative to representing the surface via a lattice is a system designed by Peuker (1977), known as TIN. This model approximates the terrain surface with a network of triangular facets all based on irregularly spaced Z data. The data points are connected by lines to form triangles and the surface of each triangle is a plane. These triangles form a continuous mosaic.

There are two interpolation techniques used in Arc/Info on TIN surface models, linear and quintic. The linear interpolation method considers the surface as a continuous faceted surface formed by triangles. With the slope facet being constant until another triangle is crossed, the normal slope facet then changes abruptly. The surface value to be interpolated is calculated based solely on the Z values for the nodes of the triangle within which the point lies. The surface value is thus obtained by intersecting a vertical line with the plane defined by the three nodes of the triangle (ESRI, 1991, pp. 2-19). The generalised equation of a point (X,Y,Z), in a single triangle facet is:

$$Ax+By+Cz+D-0$$

where A, B, C and D are constants determined by the coordinates of the triangles three nodes (ESRI, 1991, pp 2-19).

Similar to the linear interpolation method, quintic interpolation considers the surface model to be continuous (ESRI, 1991, pp 2-20). It provides increased functionality as it considers the surface model to be smooth, therefore the normal to the surface varies continuously within each triangle (ESRI, 1991, pp. 2-20). Moreover, there are no abrupt changes between triangle boundaries, instead the geometry of the neighbouring triangle when interpolating the Z value of a point is taken into consideration. Quintic interpolation employs a breakline bivariate quintic interpolation algorithm using a bivariate fifth degree polynomial in X and Y (ESRI, 1991, pp. 2-20).

The generalised equation (without breaklines) is:

$$z(x,y) = \sum_{j=0}^{5} \sum_{k=0}^{5-j} q_{jk} \, x^j \, y^k$$

This algorithm was originally developed by H. Akima (ESRI, 1991, pp. 2-20). There are 21 coefficients to be determined with the values of the function and its first-order and second-order partial derivatives being given at each node of the triangle, thus yielding 18 coefficients (ESRI, 1991, 2-20). The three remaining coefficients are determined by considering the surface as both smooth and continuous in the direction perpendicular to the three triangle edges (ESRI, 1991, pp. 2-20).

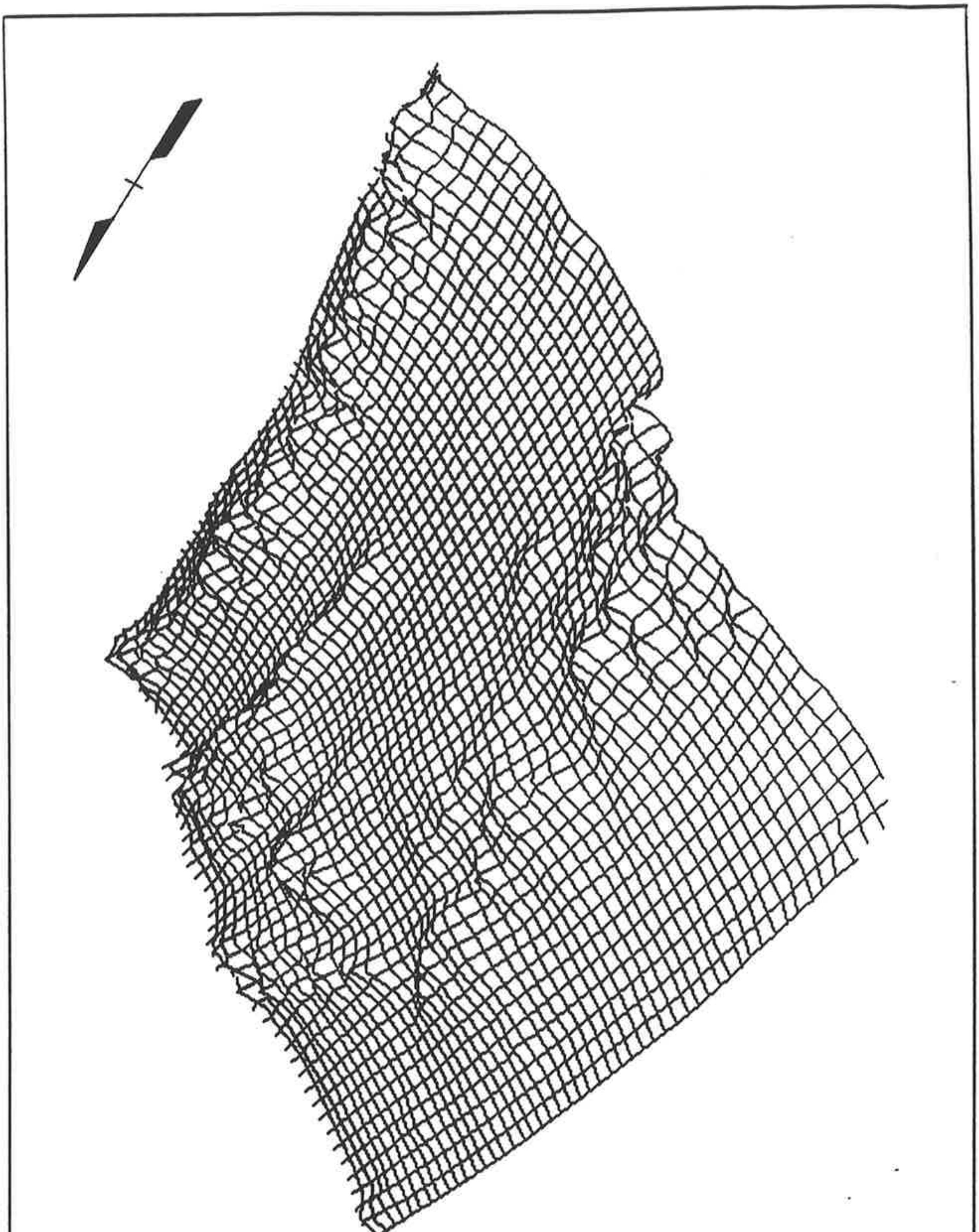
To create the initial DEM coverage the quintic interpolation algorithm was utilised. The resultant TIN coverage was reinterpolated to a lattice coverage containing $30m^2$ pixels (Figure 6.1).

Elevation and slope parameters were derived using commands from the Arc/Info GIS (Figures 5.10 and 5.11). Detailed explanation of the algorithms used by the ESRI software is not provided. The methodology provided by the GIS software for calculating these parameters is documented in the Arc/Info V6.0 user manuals.

6.2.5 Converting Point Data Collected in the Field to an Interpolated Coverage

Groundwater data was collected for 117 points in the Jamestown study area. To develop a groundwater map indicating the depth to groundwater another interpolation technique, kriging, was utilised. This spatial interpolation technique was first addressed by Matheron (1971) and D.G. Krige (both cited in Burrough, 1986). It is a geostatistical procedure that generates an estimated surface from a scattered set of points with z values (ESRI, 1991, pp. 5-5). "The method rests on the recognition that the spatial variation of any hydrological property, known as "regionalised variable", is too irregular to be modelled by a smooth mathematical function but can be described better by a stochastic surface. The interpolation proceeds by first exploring and then modelling the stochastic aspects of the regionalised variable.

FIGURE 6.1: DEM Lattice Jamestown Study Area



The resulting information is then used to estimate the Λi weights for interpolation" (Burrough, 1986, p. 155). In this project kriging is utilised as a method to obtain both the estimates of the values of the point map input variables and the kriging variances for each cell.

Within the Arc/Info GIS implementation of the kriging methodology is in accordance with techniques utilised by Burrough (1986) and Heuvelink and Burrough (1989). In simple terms, the variation is measured using the semi-variance, which is half the squared difference in Z value between pairs of the input sample points. According to ESRI, the semi-variance Y in Z values between all the pairs of points Z(x) and Z(x+h) separated by distance h (referred to as lag) can be estimated from the sample data with the equation:

$$\gamma(\mathbf{h}) = \frac{1}{2n} \sum_{i=1}^{n} \{Z(x_i)-Z(x_i+\mathbf{h})\}^2$$

where n is the number of pairs of sample points separated by distance h (ESRI, 1991, p. 7). A graphical plot of Y(h) against h is known as the sample semi-variogram. Development of this plot is essential to determining the optimal weights for the interpolation process (Burrough, 1986). The semi-variogram created by kriging the groundwater data is displayed in Figure 6.2.

In this case the nugget variance is very small compared to the spatially dependent random variation, thus in line with the methodology proposed by Burrough (1993), the Gaussian model was implemented. The Gaussian model is:

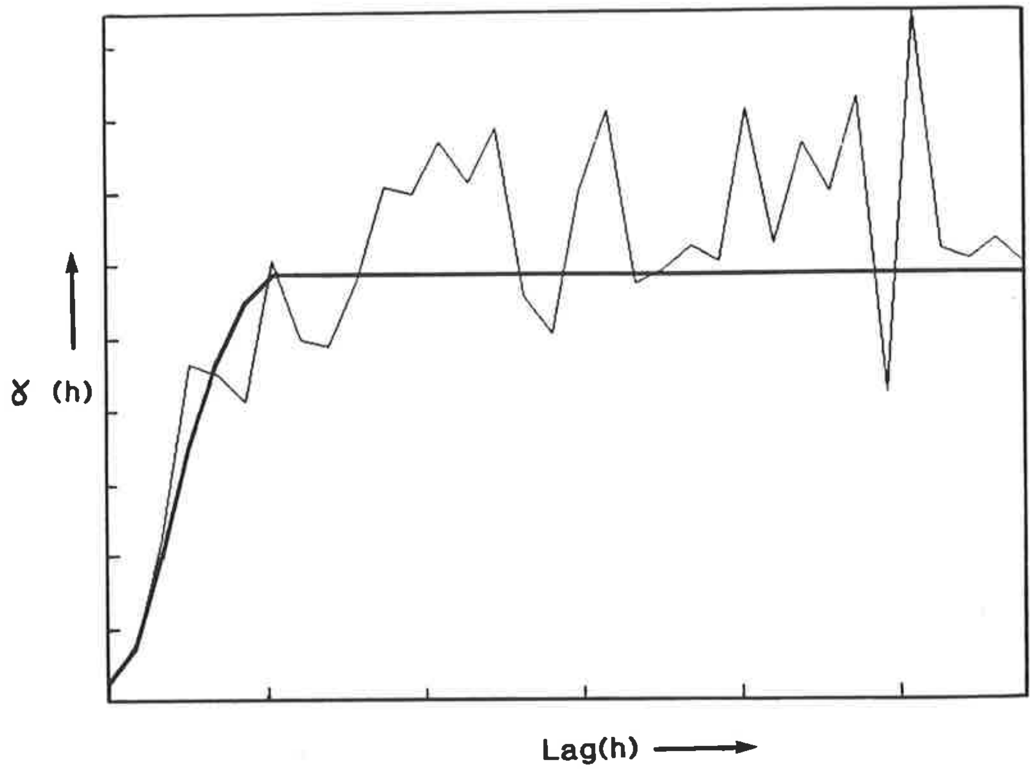
$$\gamma(h) = c_0 + c_1 \{1 - \exp(-h/a)\}$$

where a is the range, h is the lag, c_0 is the nugget variance and c_0+c_1 equals the sill (Burrough, 1993). In the above example, the values of the parameters to the fitted model are a = 1181.288, $c_0 = 2.388$ and the sill = 58.748.

Burrough (1993) writes that "smooth gradually varying attributes are often described by a Gaussian variogram" (Burrough, 1993, p. 22).

The Arc/Info GIS allows the variance generated by the kriging module to be plotted. The variance lattice contains the predicted variance at each "out lattice" mesh point.

FIGURE 6.2: Semi-Variogram Plot



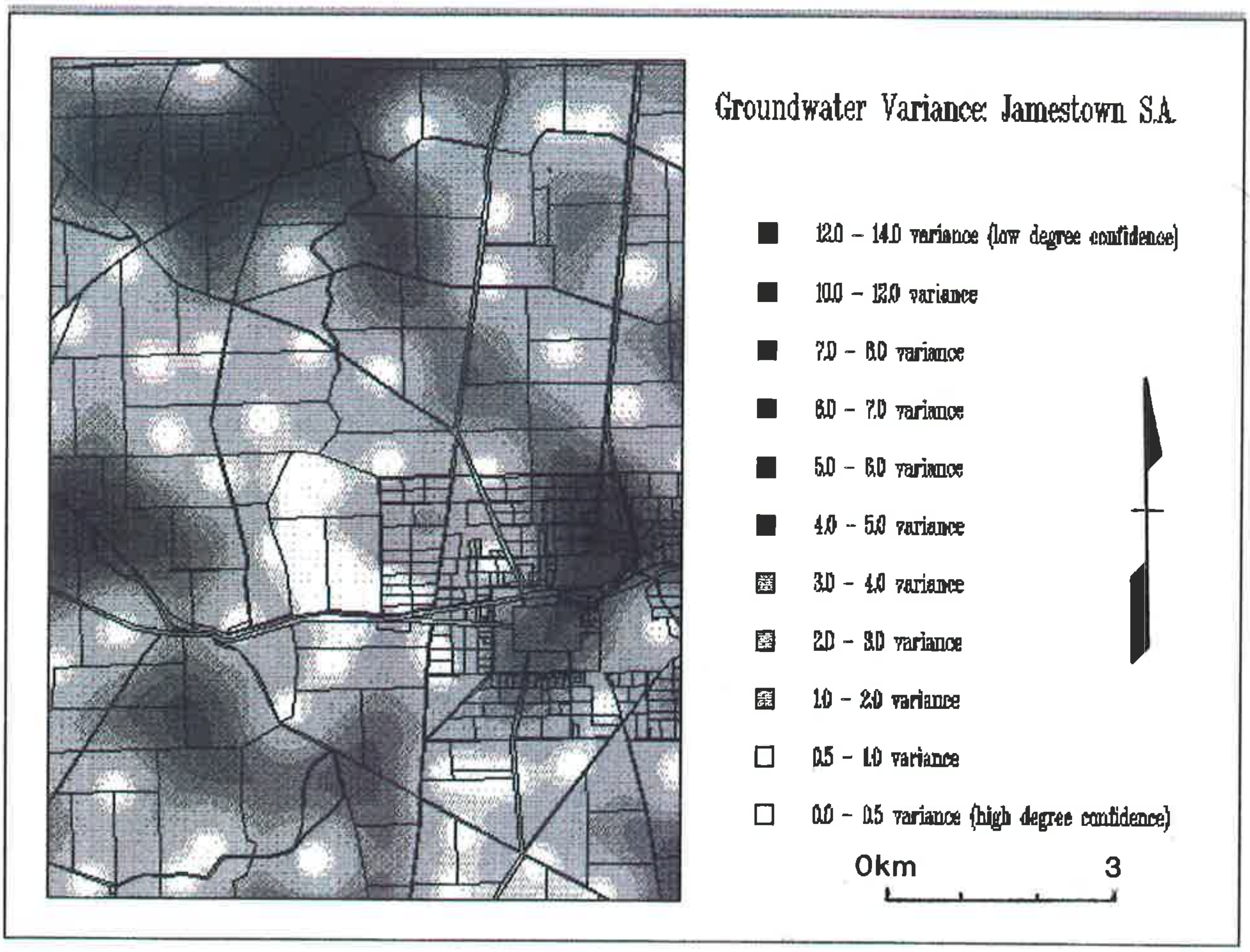
Assuming that the variance is normally distributed, there is a 95.5% probability that the actual Z values at the mesh point is the predicted "out lattice" value -/+ twice the square root of the value in the variance lattice (ESRI, 1991, p. 4):

$$Z = \langle out\_variance \rangle + 2 \sqrt{(variance\_lattice)}$$

The variance values produced by this procedure were used in developing a method for conveying uncertainty regarding the land classification process to the user. This process will be considered in the following section. The variance lattice produced is displayed in Figure 6.3.

Figure 6.3 highlights the spatial variation of the sample points. Notably sampling within the 1x3km catchment is of a dense nature. The other area of intensive sampling was conducted in the Bundaleer Valley south of the city of Jamestown. As the parameter, depth to groundwater is only taken into consideration to determine groundwater discharge potential (as indicated in the following chapter), the collection of sample data occurs in areas where discharge is likely to occur.

FIGURE 6.3: Spatial Distribution of Variance Generated for the Depth to Groundwater Coverage



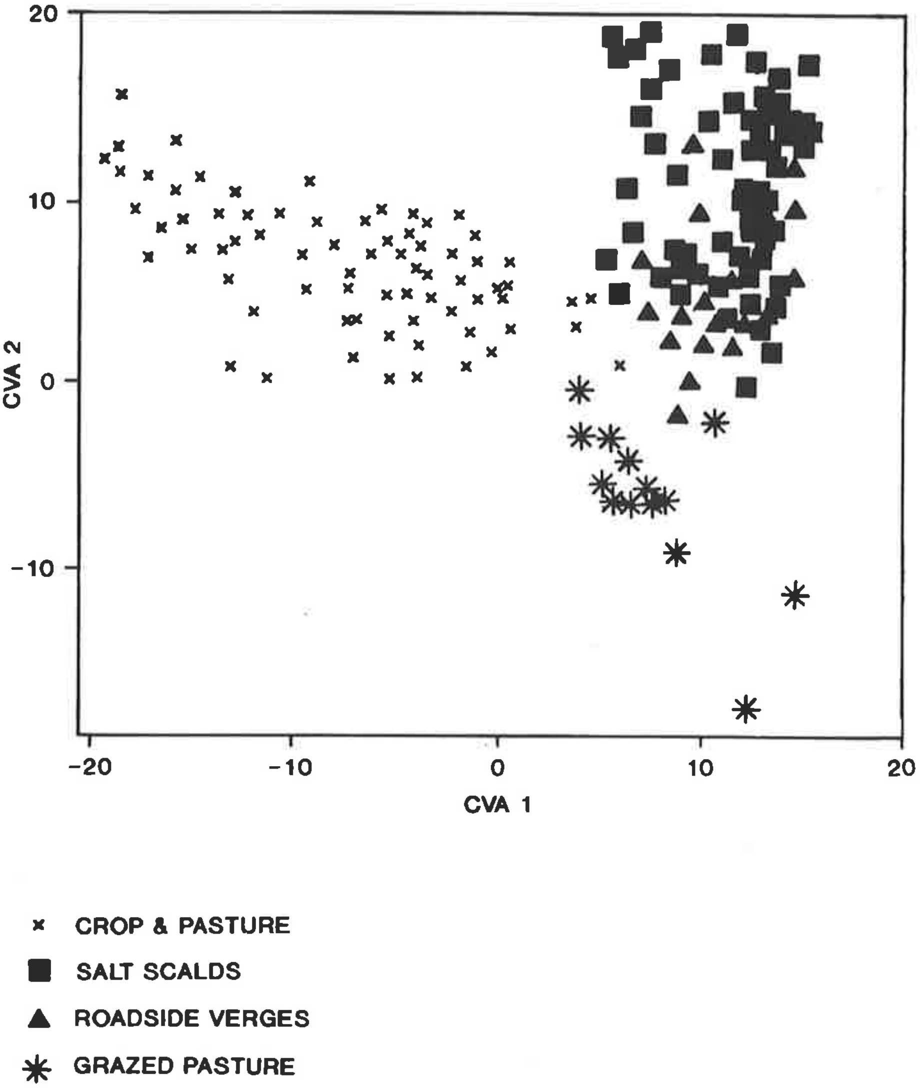
There is no need to have detailed sampling conducted along ridge lines where there it is unlikely that discharge will occur. Sensitivity analysis to verify the effect of the spatial distribution of the points on the resultant interpolated coverage was not conducted.

6.2.6 Landsat TM Imagery

Landsat TM imagery was acquired for the study region on the 24th October 1991 at 9.50am. Spring imagery was obtained as previous salinity detection research found it to be suitable to detect relevant spectral variations (Kirkby et al., 1992). The training sites, as considered in Chapter 5 were grouped into spectral classes based on the results of a canonical variate analysis (CVA) (Campbell and Atchley, 1981). In the CVA procedure successive linear combination of the spectral bands are found which maximise the ratio of separation between the classes to the pooled variation within the classes (Richards, 1986).

Figure 6.4 indicates the ordination plot of the 177 training sites using the spectral bands 2, 3, 4, and 5. Bands 1, 6 and 7 were excluded as their spectral responses were not found to be useful in discerning salinisation.

FIGURE 6.4: Results of the CVA Analysis



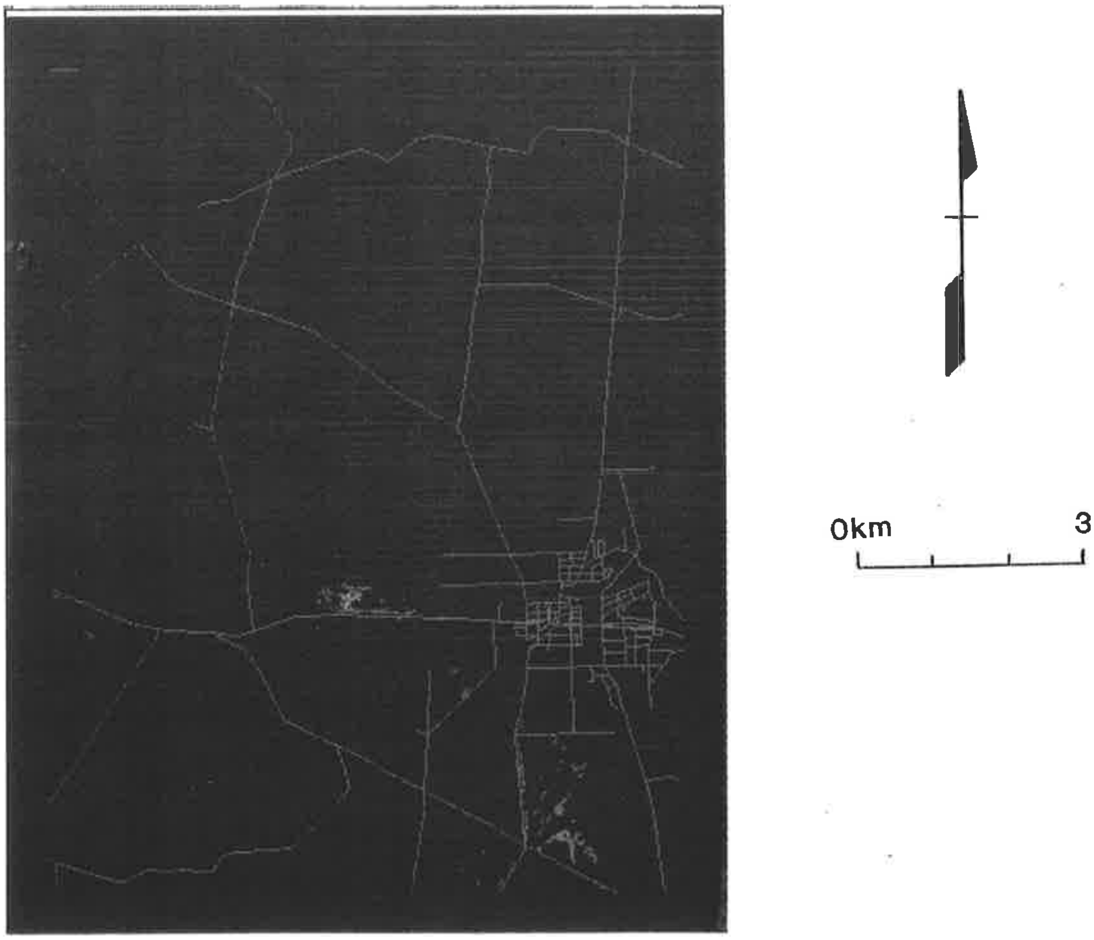
The sum of the canonical roots for the four band analysis was 152.64, with the first three roots, 75.7, 50.01 and 17.3 respectively, explaining 94% of the between site variation. Canonical variate analysis demonstrated clearly the spectral separability of the key cover types using Landsat TM data. Crop and pasture were separated from the saline sites by the first canonical vector (CV1). This CV1 from the above plot had the following coefficients for the four band analysis:

$$CV1 = 0.11B_2 + 0.18B_3 - 0.28B_4 + 0.06B_5$$

The second canonical vector (CV2) was not found to successfully discriminate between cover types. A second CV2 to discriminate between roadside verges and salt affected area would have been desirable.

The coefficients from CV1 were applied to the rectified and registered Landsat TM image. Figure 6.5 shows the salt affected groundwater discharge zones clearly in the enhanced display, notably the roadside verges also appear as salt affected.

FIGURE 6.5 **Spectrally Enhanced TM Image Indicating Areas of Salt Affected Soil**



The spectral values for the enhanced image were not directly utilised in the classification procedure, rather they were used to simply display the salt affected areas (Figure 6.5). Using the positional locating features of the MicroBRIAN system, the salt affected regions from the enhanced image were identified on the "raw" unenhanced Landsat TM image. Spectral range values for these salt affected areas were subsequently derived from the "raw" Landsat TM image using the interrogation facilities of the MicroBRIAN image processing system. The spectral ranges derived, i.e. Band 3 >40 and <58, formed the basis for the development of expert rules identifying groundwater discharge zones from the "raw" Landsat TM data (Chapter 7).

Conceptually, the thesis adopted the stance that only "raw" information should be retained in the GIS/RDBMS database. The rationale behind this decision stems from wanting to retain the maximum amount of information within the database, i.e. retain raw unclassified/manipulated information. This conceptual stance may have contributed to the low classification accuracy of the groundwater discharge zones attained in Chapter 9.

6.3 SUMMARY

Geographic data is often incomplete or imprecise. An understanding of the accuracy of derived results from GIS data coverages can only be determined by an assessment of the vagaries of the data acquisition process, for this reason specific attention has been paid to the digitising process, to the interpolation procedures and to the image processing stage. Consideration of errors introduced into the spatial database during the data capture and conversion process will be considered Chapter 10.

To complete the methods section the following Chapter considers the knowledge acquisition phase of the Expert System development. This following section is the final component of the data acquisition stage prior to the development and subsequent testing of the integrated ES and GIS system.

Chapter 7

KNOWLEDGE ACQUISITION

Knowledge acquisition is an activity that has traditionally been coupled with knowledge engineering, where the task is "to mine those jewels of knowledge out of experts heads one by one" (Feigenbaum and McCorduck, 1984, p. 107). However, knowledge is not a substance to be quarried (Young, 1988); rather it is something to be probed by knowledge engineers in an essentially social process (Goodall, 1985). As acquiring knowledge regarding the domain is difficult, this process of knowledge acquisition is considered to be the "bottleneck" in ES development (Gammack and Young, 1985).

This section initially considers the techniques employed by the project to derive information from the dryland salinity domain experts. Subsequently, it then considers the application of machine learning techniques to the ES rules as a means to improving efficiency of the knowledge acquisition phase. The final section defines certainty factors and considers their use in the quantification of uncertainty in the ES rule base.

7.1 KNOWLEDGE ACQUISITION: STAGE 1

There are three phases identified in the literature: domain definition or orientation; problem identification and problem analysis (Breuker and Wielinga, 1987). Each of these three phases were considered when the initial interviewing method was formulated for this project.

7.1.1 Interviewing Method

Five South Australian secondary salinity experts were interviewed using the focused interview method (Breuker and Wielinga, 1984). The first three were selected for their specific expertise in areas relating to the domain problem: Dave Maschmett for soils; Stewart Richardson for groundwater and Peter Smith for geology. In addition, Chris Henshcke, the geophysicist supervising the dryland salinity project being conducted in the small 1x3km catchment within the study region, and Mary-Anne Young, the Department of Primary Industries regional soils officer, were selected for their general knowledge regarding salinisation processes in the area.

The focused interview is where the interviewer prepares the topics in advance, and the interview is conducted in a manner similar to normal conversation (Breuker and Wielinga, 1984). The structure of the interview consists of three parts. First an introduction: to explain the goals; orient the expert to the role of the interviewee; and to agree to the sequence of topics. This step is designed to both motivate the expert and allow them time to plan strategies for responding (Breuker and Wielinga, 1984). Secondly, topics are addressed sequentially from a list, often in an associated manner determined by the interviewer, with various probing interventions employed to elicit further information about each topic. Finally, the interview is closed with a summary and evaluation of the information about each topic. According to Breuker and Wielinga (1984) each interview should be taped.

In accordance with this method, a series of interviews was conducted with each of the above mentioned experts over a three month period. Initially, each expert was interviewed separately, i.e soils, then geology and groundwater. Due to the unique relationship between geology and groundwater the two experts, Peter Smith and Stewart Richardson, were interviewed together. Once these experts had been interviewed twice each, a rule base was constructed and the three experts consulted again. Following from this the two local experts were consulted regarding the validity of the rule base. A typical rule was:

FIGURE 7.1: An Example of a Rule Elicited from an Expert

If Soil = Belalie Loam and
If Geology = Saddelworth Formation and
If Groundwater <= 2 metres below the earth's surface and
If Elevation <= 470 metres above sea level and
If Slope <= 3%
-->
Then land classification type is "high potential discharge"

After three focused interviews were conducted with both Chris Henschke and Mary-Anne Young a final rule base was established. This unstructured (or flat) rule base contained 139 rules and identified six thematic land classes.[1] Three classes represented groundwater discharge and the other three groundwater recharge (Table 7.1). At this stage the classes were treated as mutually exclusive.

TABLE 7.1: Land Classes Derived by the Experts

Existing Groundwater Discharge
High Potential GroundwaterDischarge
Low Potential Groundwater Discharge

High Potential Groundwater Recharge
Medium Potential Groundwater Recharge
Low Potential Groundwater Recharge

Once encoded into the knowledge base the rule base was condensed into 16 membership rules (Figure 7.2). This exploits the similarities of the rule base, i.e. if A

---

[1] It must be noted that at this stage of the rule base development no EM31 data was available. Once the data became available at a later stage the rule base was changed accordingly. This will be considered later in the chapter.

then C, if B then C, therefore if A and/or B then C. A decision tree was also produced. Once the rules were encoded into the rule base the Expert System was initiated to construct a salinity classification map. The first trial took approximately 4 hours to classify all cells in the study area and was therefore considered too inefficient. Subsequently, to improve efficiency the rule base was refined using the C4.5 inductive learning system (Quinlan, 1993).

FIGURE 7.2: An Example of a Membership Rule

If Soil = member (Belalie Loam, Belalie Clay loam) and
If Geology = member (Saddleworth, Tapley Hill formation) and
If Groundwater <= 2 metres below the earth's surface and
If Elevation <= 470 metres above sea level and
If Slope <= 3%
-->
Then land classification type is "high potential discharge"

7.2 KNOWLEDGE ACQUISITION STAGE 2: USING MACHINE LEARNING TECHNIQUES TO IMPROVE KNOWLEDGE ACQUISITION

In the development of the ES, the proper methodology for structuring knowledge was essential, given that the GIS contained large amounts of data. In the integration of the ES with the GIS for this project two specific inefficiencies were noted: block structuring of the classification rules so that they were mutually exclusive and the actual structure of the individual ES rules.

A method to overcome these problems is to pass classified data to an inductive learning program which can allow an ES's rule base to be reorganised to improve the performance and efficiency of the expert classification system (Eklund and Salim, 1993).

This subsection discusses several issues relating to GIS and machine learning techniques in artificial intelligence. It also explores the potential for machine learning in GIS.

7.2.1 Inductive Learning

Inductive learning is a process of acquiring knowledge by drawing inductive references from proven fact (in the simplest case, positive and negative examples of some events). Such a process involves the operations of generalising, specialising, transforming, correcting and refining knowledge representations. The study and modelling of this form of learning is one of the central topics in machine learning (Dietterich and Michalski, 1983).

There are several potential applications of inductive learning systems, but this thesis will focus on the automated construction of knowledge-bases for ES. As indicated in the earlier part of this chapter, the present approach for constructing knowledge-bases is a time consuming process of formalising expert knowledge and encoding it in some knowledge representation system, in this case production rules. Inductive learning programs can provide both an improvement on current techniques and a basis for developing alternative knowledge acquisition methods (Quinlan, 1983). In selected domains, inductive systems are already able to determine decision rules by induction from examples of expert decisions (Quinlan, 1979; Michalski and Chilausky, 1980), and there are many successful systems that employ these techniques (Quinlan, 1986; Cestnik et al., 1987). This process can greatly simplify the transfer of knowledge from expert to machine.

Another important application of inductive programs is in various experimental sciences. Here they can assist a user in detecting interesting conceptual patterns or revealing hidden structure in a collection of observations. Widely used statistical data analysis techniques are often insufficiently powerful for this task. Methods for conceptual data analysis are needed, not merely mathematical formulas, but logical style descriptions, characterising data in terms of high level, human orientated concepts and relationships (Quinlan, 1990). The best known example of such a system that uses this type of application are the ID3 system (Quinlan, 1979; 1986) and its descendants: Assistant Professional (Cestnik et al., 1988) and the C4.5 system (Quinlan, 1993). The most recent variant C4.5 is used in this project.

C4.5 is a system for automated knowledge acquisition for expert systems and other AI applications. The main vehicle for automatic construction of knowledge-bases is the induction of decision trees from examples. C4.5 is intended as an aid in solving decision problems that are hard (or impossible) by hand to improve an existing strategy for decision making (Salim, 1994). C4.5 is also able to use unreliable information and may also work with partial information which is often the case in real-world problems. The system points out the importance (or irrelevance) of certain data for a decision problem. If data does not appear in a decision tree then it is often the case that it carries no information relevant for the decision problem, or it is redundant (Quinlan, 1993).

From a generated decision tree a human expert can extract dependencies between data that show regularities in the domain of application. C4.5 can be used as an automatic ES rule generator as it has a facility to use generated rules for solving new decision problems. This method for constructing an ES was shown to be very efficient (Cestnik et al., 1988; Quinlan, 1993). Instead of formulating decision rules by hand, which may take many years of intensive work, it suffices to prepare a set of examples of solved decision problems which are then used by C4.5 to automatically generate decision rules which discriminate between the examples. Methodologically, this is of benefit because preparation of examples is less demanding work because examples may be retrieved from already stored cases.

In Appendix 5 consideration is given to the generation of a decision tree and the subsequent pruning strategies. Importantly, within this appendix, another experiment comparing the efficiency of inductive learning, Instance-based learning (IBL) and a neural network back-propagation (BP) algorithm is reviewed. The aim of this experiment was to verify how many data points were required to build an effective classifier. The result of this experiment, concluded that the C4.5 was a more efficient classifier for this task, thus supporting its use in this thesis.

7.2.2 Methodology

As previously considered the earliest version of the rule base consisted of 16 membership rules in total. These rules were fired over each of the 12-tuple cells (1 tuple for each coverage, no EM31 data) of the database to produce a salinity classification for each cell. Once classified by the ES this data was input to the inductive learning program C4.5 (Quinlan, 1993) as a training set.

7.2.3 Inductively Derived Rules

The *gain criteria*, considered in Appendix 5, selects ground water depth as the dominant discrimination. The training set is partitioned into two sets, one for training instances with groundwater depth <= 3 and the other for groundwater depth >3. Each of the partitioned training sets are considered for each subtree of the decision tree. In the subtree where ground water depth <= 3, slope is selected as the next node in the subtree. If the slope is >3 then all training instances with these features are of class 3. This equates to a single classifier as shown in Figure 7.3.

FIGURE 7.3: A Decision Tree Classifier Rule

```
Rule 1: If groundwater <= 3 then
            if slope <= 3 then
                if soil in (12,13,3,5,7) then
                    if tmband4 <= 100 then
                        if tmband3 <= 39 then
                            class 3 (Low potential discharge)
```

The above process is repeated, forming new training sets and selecting attributes for discriminating classes in the training set until all training instances classify to a single leaf node (which defines rules such as Rule 1 of Figure 7.3) (Salim, 1993). A decision tree is the output of this process.

It is often possible to prune a decision tree so that it is both simpler and more accurate. Even though the pruned trees are more compact than the originals, they can still be cumbersome, complex and inscrutable (Quinlan, 1987). Large decision trees are difficult to understand because each node has a specific context established by the outcomes of tests at the antecedent nodes. To provide insight, as well as accurate predictions on unseen cases, decision trees are rewritten as production rules.

Rewriting the tree to form a collection of rules, one for each leaf in the tree, would not result in anything much simpler than the tree, since there would be one rule for every leaf (Salim, 1993). C4.5 rewrites the decision tree to production rules eliminating irrelevant conditions in rules (Salim, 1993). It also makes the rules mutually exclusive and exhaustive using Minimum Description Length (Quinlan and Rivest, 1989).

In measuring performance of the generated rules against the correspondent decision tree the pruning and simplification criteria of Fayyad and Irani (1990) are considered. This defines a number of metrics for recording the relative quality of a decision tree. The salinity classification data was partitioned into two: 80% was used as a training set to construct a decision tree and the remaining 20% as a test data. Results are summarised below in Table 7.2.

TABLE 7.2: C4.5 Application Results

Performance Indicators					
	M1	M2	M3	M4	M5
Decision tree	**4** (0.00016%)	48	36	84	506
Pruned decision tree	**5** (0.00016%)	35	27	62	694
Production rules	**16** (0.00054%)	19	-	67	1277

Below is the list of metric measures of performance that were used in the above table:

1) M1: Percentage error on classifying unseen examples (actual numbers in bold);
2) M2: Number of rules in the rule-base (leaves in tree);
3) M3: Number of nodes in the tree (the nodes represent the antecedent conditions);
4) M4: The total number of preconditions in the rule-base. This measures the generality of the entire rule set. It is a more accurate measure of generality than the average number of preconditions per rule, since the latter is not indicative if the number of rules is relatively large when compared against the number of training examples (Salim, 1993);
5) M5: Average example support per rule (per leaf). This is the average number of examples in the test set on which a rule is applicable and correctly predicts the class. It is a measure of the applicability (generality or utility) of the produced rules.

The results are a classification accuracy that indicate a high degree of confidence in the decision tree. The results also show that the generated production rule accuracy is as good as that of a decision tree with the advantage of having fewer and more readable rules.

7.2.4 Inductive Learning and the Rule Base

Expert derived and inductively derived rules are similar in form; they both have simple structure of conjunctive attribute value tests. It is interesting to notice however, that the inductively derived rules when presented to an expert were judged quite favourably in terms of mutual exclusivity and structure of individual rules. In only a few cases the expert considered the rules to be confusing, i.e. presented combinations of antecedents which the expert had never considered.

There are two issues concerning the nature of the inductive rules generated by C4.5 in this domain. The first is the logical structure and readability of the induced rules. Are they readable? Does the rule structure present the most efficient means for determining a class at the expense of the domain logic? This concern was satisfied on the first run of the experiment. The rules could not only be read but the knowledge initially encoded into the ES was clearly distinguishable; the premise order was slightly different but the attribute value tests were identical. When the ES rules were restructured the classification process took 3 hours, an improvement of one hour.[2]

The second issue, was the discovery of "new knowledge", i.e. patterns in the data which were useful to the classification task but not represented in the ES rule-base. This will be discussed in the next subsection in detail.

Table 7.3 below gives the summary of the rule numbers associated with each class for the inductively derived and expert rules.

TABLE 7.3: Rule Base Comparison

Class Description	#Expert Rules	#Inductive Rules
Existing discharge (1)	1	4
High potential discharge (2)	1	5
Low potential discharge (3)	3	3
Low potential recharge (4)	4	4
Medium potential recharge (5)	4	5
High potential recharge (6)	3	1
Total	16	22

Table 7.3 shows that there are more inductive derived rules than those of expert. However, most of the inductive derived rules have fewer attribute value tests and

---

[2] Three hours is an unacceptably long period for data processing. The cause of the problem derives from an error associated with "garbage collection", strategies in Knowledgeworks to clear the cache were initiated but were not successful. This problem is currently be addressed by a postgraduate student from the Department of Computer Science.

therefore classify more rapidly. Also, some of the inductively derived rules are too general, therefore they are redundant and can be removed without affecting the overall classification performance.

7.2.5 Discovering New Domain Knowledge

The most interesting aspect of this experiment was to determine whether any new knowledge could be gleaned from the GIS database via inductive learning. "New knowledge" here means knowledge that was not explicitly used in the original classification task but is relevant to it (Eklund and Salim, 1993). A number of rules produced showed surprising results and are illustrated in Figure 7.4.

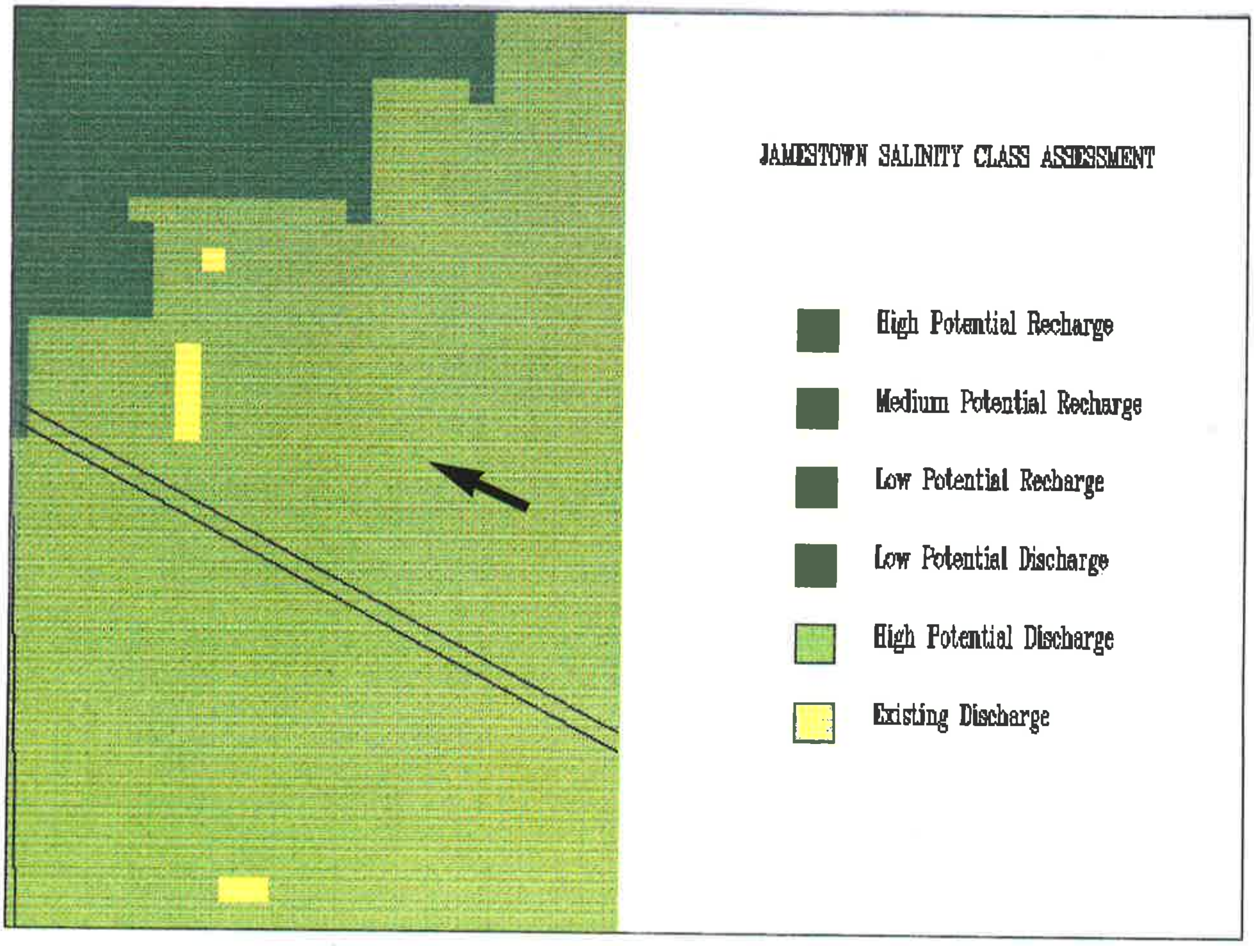
FIGURE 7.4: Sample Rules Generated by the C4.5

```
Rule 1:
        ground_water <= 3
        slope <= 3
        tmband3 <= 39
        soil in {12, 13, 3, 5, 7}
        tmband4 <= 100
        -> class 2

Rule 16:
        elevation <= 2
        ground_water <= 3
        slope <= 3
        soil in {12, 13, 3, 5, 7}
        tmband4 >100
        -> class 2

Rule 14:
        elevation <= 2
        ground_water <= 3
        slope <= 3
        tmband3 >58
        -> class 2
```

Rules 1 and 16 are consistent with the High Potential Discharge Class (HPD) (class 2) that has yet to become salt affected. This is inferred by TM bands 3 and 4 which indicate very healthy vegetation growth (see Figure 3.1, low and high spectral responses in each respectively, derived from the original Landat TM image). From this fact we can infer that the vegetation is experiencing flush growth, due to the high water table with a low salt content. Although there is a high potential for salt affectedness it has yet to manifest itself. On the other hand rule 14 indicates, via the remotely sensed information, that the vegetation is stressed. This is a cell that has realised its high salt discharge potential.

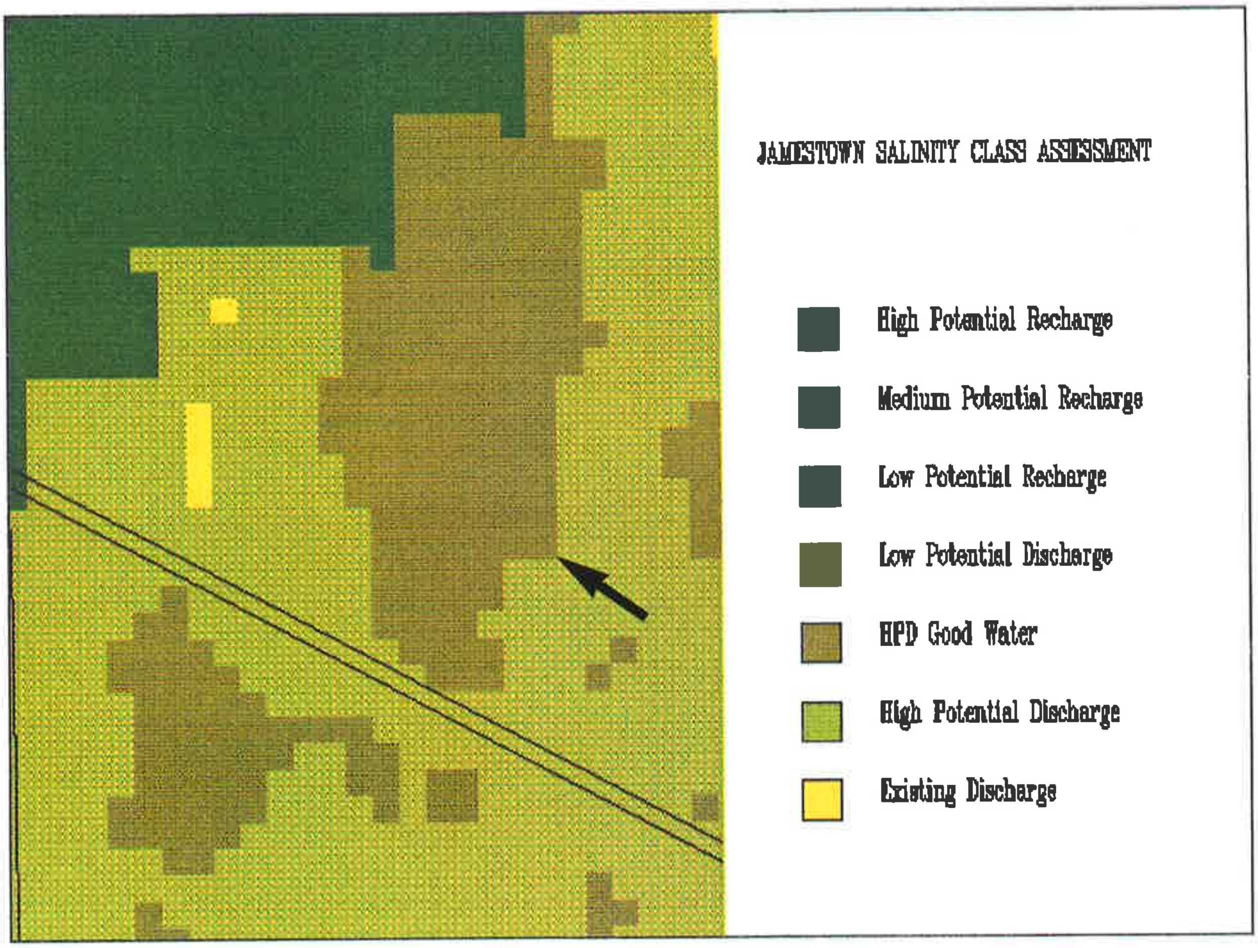
These two proper subclasses of HPD indicate that it is an aggregation of the two, and suggests a new class. The information was coded into the software system produced by this thesis, considered in Chapter 8, and a new land classification map was generated (Figure 7.5 and 7.6).

FIGURE 7.5: Original Classification



Note: Arrow indicates area of interest.

FIGURE 7.6: Classification with the Additional Class



Note: Arrow indicates area of interest.

From these two diagrams it is clear that the original vegetation High Potential Discharge (HPD) class was split into two classes: one indicating healthy vegetation where the groundwater was within two metres of the surface and the other stressed vegetation. Interestingly, the new class, where the vegetation is experiencing flush growth, occurs in the only two paddocks where contouring as a method of surface water control is implemented. From this observation we can imply that surface water engineering as a method of water management is useful in this area.

7.2.6 Adding Another Data Layer

Following from this experiment the experts were interviewed once again when a new EM31 horizontal/vertical partial data coverage was added to the database. It was at this stage that the experts, who were by now more familiar with the project and the whole process of knowledge elicitation, suggested that the rule base should be split into two, representing conditions conducive to both recharge and discharge. The experts did not want the two categories, recharge/discharge, to be treated as mutually exclusive as they had been classified in the first rule base. This notion was acted upon as it also enabled the opportunity to represent groundwater discharge as a continuum, rather than as discrete classes. In contrast, the classes for groundwater recharge remained discrete. The only change to the groundwater recharge rule base occurred with the addition of new rules to allow for the incorporation of EM31 data.

At this stage another experiment using the C4.5 program was conducted to measure the effect of the new data layer. The results from this experiment will be considered in Chapter 9.

7.3 THE GROUNDWATER DISCHARGE CONTINUUM

According to both the experts and the decision tree generated by the C4.5 program the parameter that has the highest degree of influence in the discharge classes is the depth to groundwater. As it is a dynamic parameter which can vary quickly over both space and time, it is logical to treat it as a continuum where information derived from the data is retained, rather than lost in the classification process.

To develop a continuum, 15 membership rules, plus the new rule identified by the C4.5 programme, were all considered individually in their relationship to 16 potential groundwater levels derived by the experts (Table 7.4). Thus new rules were generated according to the discharge potential (the continuum value) as assigned by the expert in response to a potential groundwater level for each membership rule. The 16th membership rule identifying existing groundwater discharge (using remotely sensed data) was also included in the new rule base of 257 rules.

TABLE 7.4: Knowledge Elicitation for the Discharge Continuum

depth to gw	soil	geology	slope	elevation	continuum
0.0 - 0.05m	bl	pbi	>4	>450m	55
0.05 - 0.2m	bl	pbi	>4	>450m	70
0.2 - 0.5m	bl	pbi	>4	>450m	120
0.5 - 0.75m	bl	pbi	>4	>450m	120
0.75- 1.0m	bl	pbi	>4	>450m	150
1.0- 1.25m	bl	pbi	>4	>450m	175
1.25-1.5m	bl	pbi	>4	>450m	195
1.5-1.75m	bl	pbi	>4	>450m	215
1.75- 2.0m	bl	pbi	>4	>450m	230
2.0- 2.25m	bl	pbi	>4	>450m	255
2.25-2.5m	bl	pbi	>4	>450m	280
2.5- 3.0m	bl	pbi	>4	>450m	320
3.0- 5.0m	bl	pbi	>4	>450m	510
5.0- 10.0m	bl	pbi	>4	>450m	1000
10.0-20.0m	bl	pbi	>4	>450m	2000
>20.0m	bl	pbi	>4	>450m	3000

The results from this table indicate a change in philosophy, where the critical depth to groundwater value, instead of being identified at 2 metres for each combination of environmental parameters, now varied according to the expert's opinion. This approach was adopted as it was thought to represent real-world occurrences more accurately. Importantly, this decision influenced the design of the data input facilities for the integrated system, i.e. as depth to groundwater data is crucial to the classification process a method for data collection should be incorporated. Additionally, a certainty factor, considered in the following subsection, for each of the rules for the groundwater discharge continuum was developed by the expert. This concept was developed as a means to convey (un)certainty regarding the decision making process to the user. Certainty factors, although easily implemented, were not developed for the three classes of potential groundwater recharge.

7.4 CERTAINTY FACTORS

A certainty factor is a measure of the extent to which the evidence that is described in the antecedent of the rule supports the conclusion that is given in the rules conclusion (Weichselberger and Pohlman, 1990). Certainty theory is based on a number of assumptions. The first is that in traditional probability theory the sum of confidence for a relationship and confidence against the same relationship must add to 1 (Luger and Stubblefield, 1989). However, it is often the case that an expert might have confidence 0.7 (say) that some relationship q is true given p and have no feeling about it being not true in p's absence. The second assumption is that the knowledge content of the rules is much more important than the algebra of confidences that holds the system together (Luger and Stubblefield, 1989). Thus, confidence measures correspond to the informal evaluations that human experts attach to their conclusions, e.g. "it's probably true" or "it is highly unlikely" (Luger and Stubblefield, 1989).

Certainty theory makes some simple assumptions for creating confidence measures and has some equally simple rules for combining these confidences as the program moves toward its conclusion. The first assumption is to split "confidence for" from "confidence against" a relationship:

Call MB(H|E) the measure of belief of a hypothesis H given evidence E.
Call MD(H|E) the measure of disbelief of a hypothesis H given evidence E.

Now either:

$1 > MB(H|E) > 0$ while $MD(H|E) = 0$, or
$1 > MD(H|E) > 0$ while $MB(H|E) = 0$ (Luger and Stubblefield, 1989).

The two measures constrain each other in that a given piece of evidence is either for or against a particular hypothesis. This is an important difference between certainty theory and probability theory (Luger and Stubblefield, 1989). Once the link between measures of belief and disbelief has been established, they may be tied together again with the certainty factor calculation:

$CF(H|E) = MB(H|E) - MD(H|E)$ (Luger and Stubblefield, 1989).

As the certainty factor (CF) approaches 1 the evidence is stronger for the hypothesis; as CF approaches -1 the confidence against the hypothesis gets stronger and a CF around 0 indicates that there is little evidence either for or against the hypothesis.

In the experiment when the rule base was being constructed experts allocated a CF to go with each of the premises of each rule within the groundwater discharge continuum rule set.[3] This CF reflects their confidence in the rule's reliability.

When a production rule is used, the certainty factors that are associated with each condition of the premise are combined to produce a certainty measure for the overall premise in the following manner:
For P1 and P2, premises of the rule,
$CF(P1 \text{ and } P2) = MIN(CF(P1), CF(P2))$, *and*
$CF(P1 \text{ or } P2) = MAX(CF(P1), CF(P2))$. (Lugerand Stubblefield, 1989).

The combined CF of the premises, using the above combining rules (min/max), is then multiplied by the CF of the rule to get the CF for the conclusions of the rule. As an example:

$$(P1 \text{ and } P2) \text{ or } P3 \rightarrow R1(.7) \text{ and } R2(.3)$$

where P1, P2, and P3 are premises and R1 and R2 are the conclusions of the rule having CF's 0.7 and 0.3 respectively. To extend the example, if the running program has produced P1, P2 and P3 with CF's of 0.6, 0.4 and 0.2, then R1 and R2 may be added to the case-specific results with CFs of 0.28 and 0.12 respectively. Implementation of this multiplicative strategy is considered in the following subsection.

Another factor that must be considered, (although not utilised in this thesis), is how to combine multiple CFs when two or more rules support the same result R. This is the certainty theory analog of the probability theory procedure of multiplying the

---

[3] It must be noted that the results from the C4.5 analysis were considered during the formulation of the CFs for each rule. The C4.5 provides strong evidence for the frequency of data categories and hence infers an almost probabilistic assignment.

probability measures to combine independent evidence (Luger and Stubblefield, 1989). By using this rule repeatedly one can combine the results of any number of rules that are used for R. This strategy is not implemented in this thesis since once a single rule fires over a cell no more rules are subsequently considered.

In support of the use of CFs, they have been shown to satisfy a limited form of "Turing test" when the ES is sufficiently well engineered. This was the case with MYCIN (Buchanan and Shortliffe, 1984). Another advantage is that the CFs that result from applying the combination rules are always between 1 and -1, as are the other CFs.

The major criticism is that it is not mathematically well-founded because the CFs must be hand allocated. It has therefore been criticised as being excessively ad hoc. Although it is defined in a formal algebra, the meaning of certainty measures is not as rigourously founded as in formal probability theory (Luger and Stubblefield, 1989). However, certainty theory does not attempt to produce an algebra for "correct" reasoning. Rather, it is the "lubrication" that lets the ES combine confidences as it moves along through the problem at hand (Luger and Stubblefield, 1989). Its measures are ad hoc in the same sense that a humans expert's confidence in his or her results is approximate, heuristic and informal. As the Harlequin ES shell provided a method for implementing CF's they were adopted by this thesis as means by which to represent (un)certainty. Other methods such as Bayes Theorem, Fuzzy sets, Theory of Evidence and Possibility theory although considered were not applied. Details of these strategies can be found in Zadeh (1965, 1983), Shafer (1976) and Dubois and Prade (1988).

7.4.1 Certainty Factor Application

In this thesis, CFs are used to determine the uncertainty associated with two areas of the system. The first is the uncertainty associated with the *if/then* production rule formulation process. As previously considered, rule formulation relies on the heuristic judgement of a series of experts. During this process, an expert is able to allocate a CF to each rule formulated. This CF conveys to the user the uncertainty within the expert's mind regarding the validity of the production rule. The second area where CFs have been applied is in the area of environmental data accuracy and precision. As considered in section 6.2.5., kriging allows an error surface to be generated from an existing point coverage. In this model, CFs are attached to the degrees of variance generated by kriging. Allocation of the CFs to the different levels of variance is again a subjective operation, i.e. a value less than 2 recorded for the variance is allocated a CF of 0.9. To determine the combined CF for each gridcell the CF attached to the uncertainty associated with the rule formulation process is multiplied by the CF factor attached to the error surface generated during the kriging process. The result is a CF value, stored in the RDBMS, for each gridcell which may be visualised using the GIS.

As kriging has only been applied to the depth to groundwater parameter, no error surface and therefore CFs have been generated, at this stage, for any other environmental parameter. Nevertheless, the depth to groundwater environmental data layer is the most important data layer in the decision to classify groundwater discharge potential. The decision not to apply CFs to the rules associated with the groundwater recharge classification process stems from wanting to implement uniform methods to both the groundwater recharge and discharge classifications. No error surface for the important environmental data layers associated with groundwater recharge were attained, thus no combined uncertainty values for both the rule formulation process and the GIS data parameters were developed.

Important to this discussion is the validity of combining CF values attributed to both the decision making process and the uncertainty within the GIS. Logically, when an expert formulates a CF for a production rule they intuitively incorporate the reliablity of the data into the CF allocation process. The method implemented here is an attempt to replace the uncertainty associated with the environmental data to allow the expert to focus on combining the data parameters.

In light of this notion to further increase the representation of the truth the multiplicative method was employed. When multiplying the CF for the rule with the CF from the data parameter the resultant CF must logically be lower than either of the two initial CFs. Consequently, a CF derived from the combination operation can never increase the degree of belief for the classification, rather this strategy will always decrease a CF for an area. This strategy is appropriate for combining only two CF's, it is not suited for combining multiple CF's. A strategy which is more viable, but not optimal, is the use of minimum functions where the CF allocated to the final result is the minimum CF in the data set, i.e. seven error surfaces have been generated for the seven data layers within the GIS and they have been combined with the CF generated for the rule by the expert; the final CF for the classification is the minimum of the combined CF data set. The research area concerning the fusion of uncertain data is one which is being increasingly reported on in relevant journals (Goodchild, 1993). This thesis does not aim to propose an optimum strategy to provide a solution to this problem, rather it aims to contribute to the discussion on the problem.

7.5 SUMMARY

The initial knowledge acquisition phase was in itself a learning process for the "new" knowledge engineer. This is reflected by the structured interview process that was initiated and the simplistic rules that were first generated. Once the initial rule base had been developed and subsequently tested by C4.5 the knowledge engineer and the relevant field experts became more confident about how to structure their heuristic knowledge. This increased confidence resulted in the formulation of two different rule bases representative of both groundwater recharge and discharge.

The role of the C4.5 program in this process cannot be underestimated. In a sense, it conveyed to the experts that the process of formalising their knowledge was valid. The discovery of the "new knowledge" also lead to a discussion among the experts and resulted in a "we can improve the methodology" attitude. Moreover, given the new found confidence in the structuring of the heuristic knowledge, the experts were more confident regarding the allocation of certainty factors to each production rule.

This chapter concludes the methods section, documenting the data collection, encoding and knowledge elicitation stage. The following chapter presents in detail the resulting software system produced by the implementation of the methodology considered in the previous chapters. Chapter 9 subsequently verifies and validates the accuracy of the classification produced by the new software system.

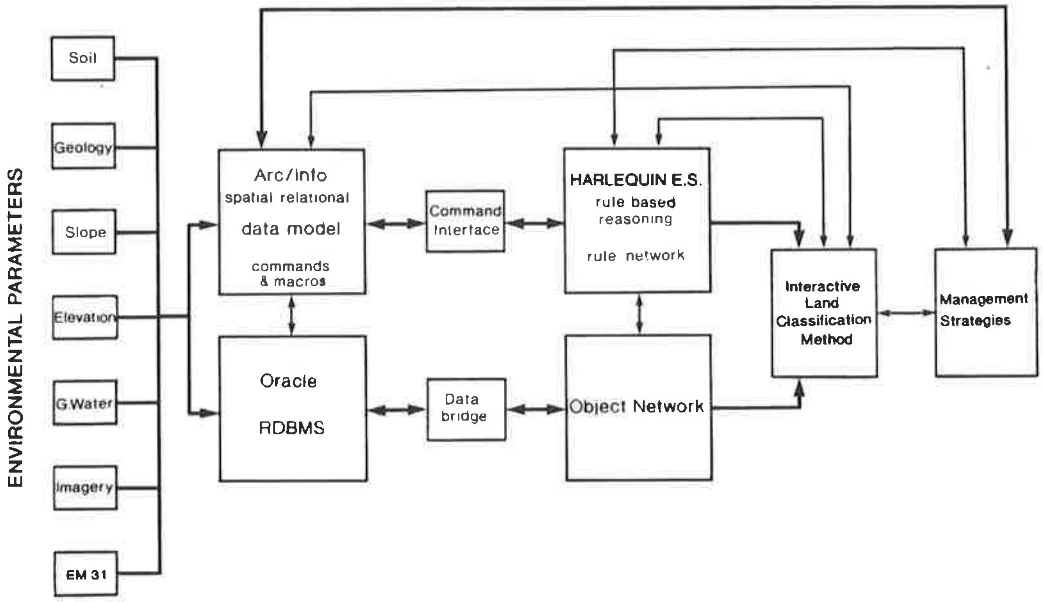
Chapter 8

SALT MANAGER: SYSTEM OVERVIEW AND APPLICATION

To make all land owners in a region more accountable for their management practices an interactive classification method for dryland salinity is necessary. The *Salt Manager* software system has been designed to be used in the regional agricultural field offices by the region's land owners and agricultural extension officers. It is unlike traditional parametric/GIS land classification approaches which deploy GIS digital maps, hardcopy thematic maps and handbooks to provide an explanation for the decision making process. These methods tend not to convey understanding of the decision making process to the user. A farmer who wants to know the rationale for why a paddock was classified as a certain type will, more likely than not, have no concept of how to operate a GIS; thus the information contained within the database and the method used to analyse the data will not be transparent to them. Similarly, the methodology to formulate a hardcopy land class map is documented in an accompanying handbook; but it is usually documented in a manner that conveys understanding only to an individual familiar with the knowledge domain and not to the final user who will implement this information. It is a supposition in this thesis that, if a user has an understanding of why a section of land was classified as a certain type, they will be more confident about the information and therefore more likely to apply it in their decision making process.

The *Salt Manager* system combines different types of environmental data within a GIS and RDBMS (Chapters 5/6); formulates heuristic knowledge known about salinisation from domain experts into an E.S (Chapter 7); integrates the three different software packages; and implements an interactive land classification method which provides an end user with data interrogation and update facilities. The components of the *Salt Manager* system are depicted in Figure 8.1.

FIGURE 8.1: Schematic Representation of the Salt Manager System



As the data acquisition and encoding has already been considered along with the knowledge elicitation and rule formulation process, this chapter will firstly consider the system design characteristics and secondly how to use the *Salt Manager* system. Information covered in the later section is further described in the *Salt Manager User Manual* (Appendix 1).

8.1 SYSTEM DESIGN CHARACTERISTICS

Salt Manager is a Unix based computer software system that integrates "off the shelf" commercially available GIS, RDBMS and ES computer software packages (Kirkby and Kurzel, 1993a; 1993c). It is a system that has been designed to model spatial phenomena using heuristic knowledge derived from a group of experts. Each commercial software package was selected for specific reasons; the GIS was chosen for its ability to represent and flexibly manipulate spatial topology representative of spatial phenomena, e.g. spatial extent of groundwater discharge (Arc/Info is also generally accepted as the most widely used GIS system); the RDBMS was selected for both its ability to store, manipulate and archive attribute information such as soil types and to provide a base platform that enables communication between all three packages (The Oracle system is also the RDBMS market leader); the ES was selected for its programming environment as it enabled a systematic means to apply heuristic knowledge to the analysis of environmental parameters, as well as providing a method for dealing with uncertainty and providing explanation facilities (The Harlequin system uses the Common Lisp Object System (CLOS) concepts within the Lisp programming environment). Detailed information of the three packages can be found in the *Salt Manager User Manual* (Appendix 1).

To enable flexible manipulation of the data by all three software packages, the system is based on a blackboard model (Engelmore and Morgan, 1989). The environmental data is held in the RDBMS within a blackboard table which replicates the raster (row/column/gridcell) structure of the GIS data file. Correspondingly a raster grid, with exactly the same configuration, exists within the GIS. A standard interface file, which allows the operation known as *relate*, is provided between the RDBMS and the GIS. This *relate* operation makes a connection between a record in the RDBMS blackboard table and the Info table associated with the GIS raster file.

To initiate the connection between the ES and RDBMS the ES provides an interface between the RDBMS tables and the objects used by the inference engine. This interface matches each raster cell, within each tuple of the blackboard table with an object, termed gridcell. In total, the blackboard table within the RDBMS consists of 302 columns by 402 rows, (121,404 gridcells), each, after analysis, containing 20 attributes

The third link in the *Salt Manager* system between the ES and GIS is provided by two functional processes. Firstly, the ES is able to initiate the GIS and subsequently run an Arc Macro Language (AML) program through a callback function. Secondly, communication between the GIS and the ES is conducted via an ASCII text file, where the GIS writes information in a format which can be read by the ES. Two programs are required to initiate this process.

To illustrate how the three systems combine, a functional task will now be considered, e.g. to fire the rule base over the environmental information a number of steps require implementation. The first is to establish the link between the ES and the RDBMS. This involves converting each raster gridcell from the blackboard within the RDBMS into an object containing 20 "slots", each pertaining to a particular environmental parameter, i.e. (121,404 objects, each with 20 slots). Once the inference engine is initiated, the rule base matches the values listed in the "slots" of each object with the appropriate antecedents of a rule and subsequently initiates the consequent[1]. In this case, the classification values are firstly written back into the object in a predetermined "slot", and then used to update the blackboard table within the RDBMS. As the RDBMS tables and the GIS topological coverages are related, the newly classified result can immediately be displayed by the GIS.

8.2 USING THE SALT MANAGER SYSTEM

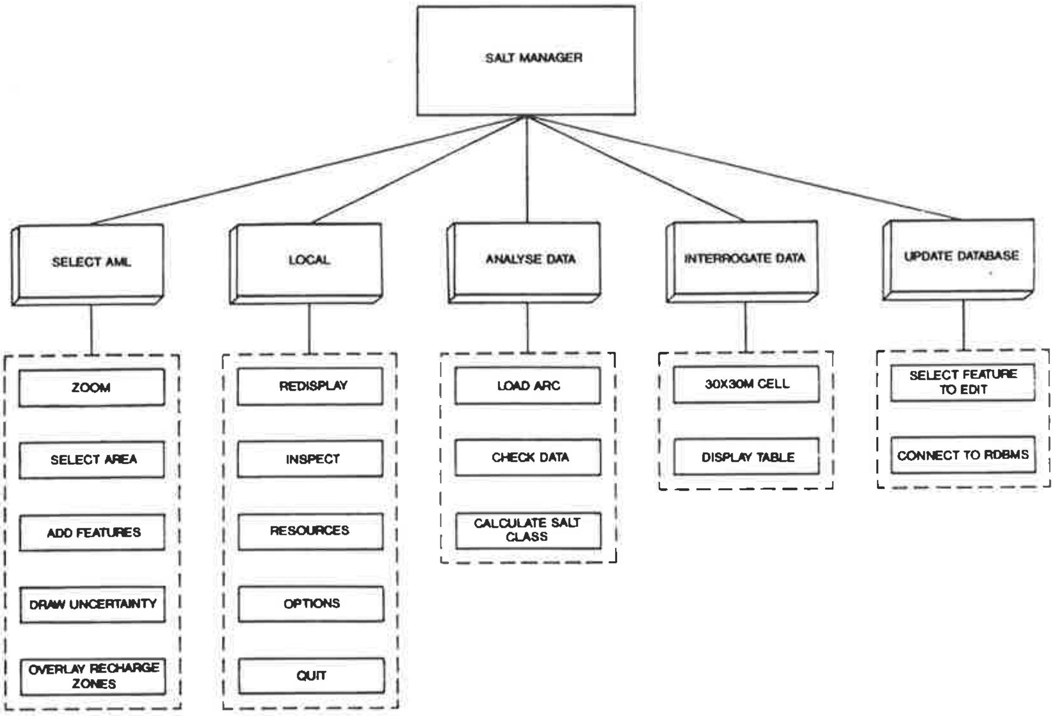
Operation of the *Salt Manager* system is via menu bars. A stylised diagram indicating the functionality of the system is displayed in Figure 8.2. There are five basic functions provided by the system. These are *Analyse Data, Interrogate Data, Update Database, Local* and *Select AML,* with each initiating, when selected by a user, a drop down menu with further selection choices.

There are three main tasks which the *Salt Manager* system performs, they are: (i) classification of the environmental data (*Analyse Data*); (ii) an interrogation procedure to ascertain the decision making process (*Interrogate Data*); and (iii) updating of the RDBMS (*Update Database*).

---

[1] For this process, either of the two rule bases (groundwater recharge or discharge), have to be manually entered into the Lisp code. This operation has, at this stage, been purposely kept manual to encourage active thought before the classification is re-initiated.

FIGURE 8.2: ***Salt Manager* Functions**



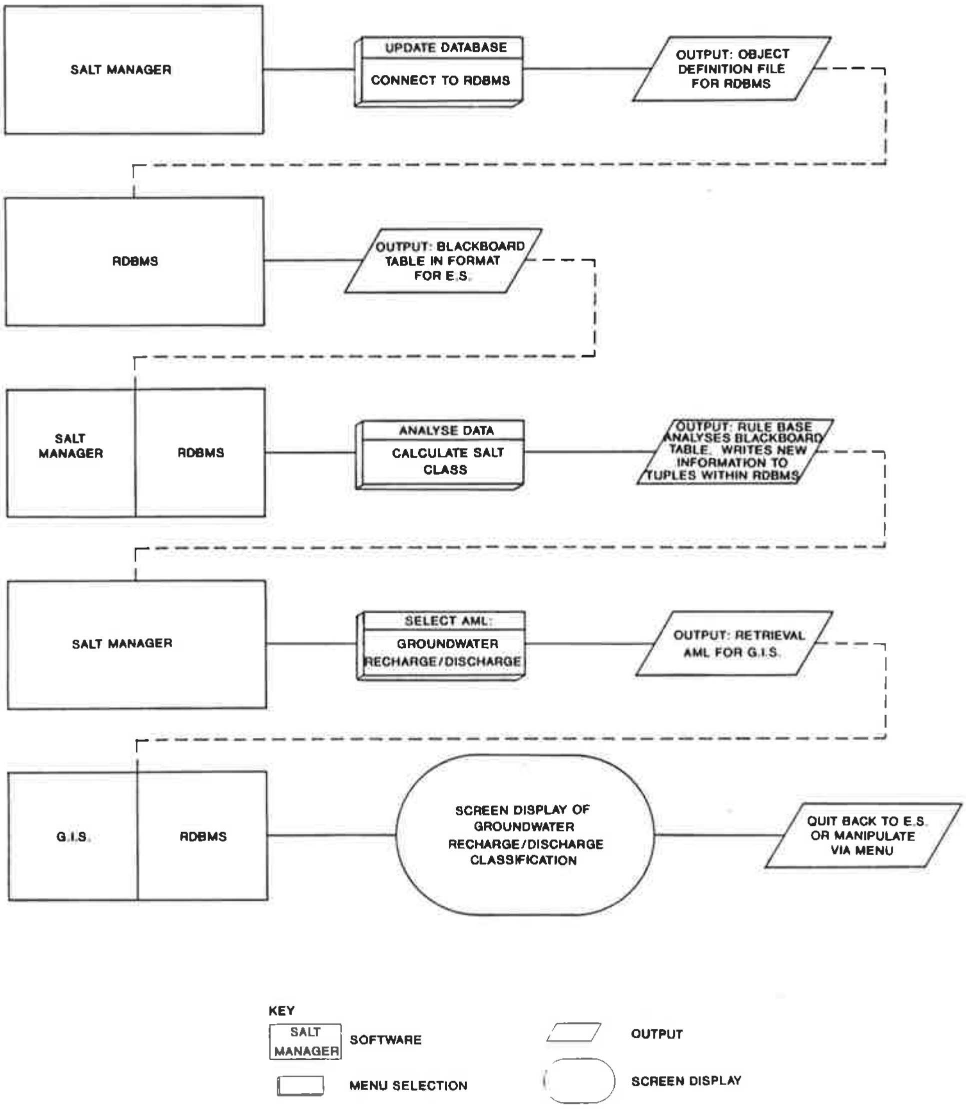
The fourth menu selection *Local* is the default option provided for each window generated within the Harlequin ES environment. The *Local* function provides the user with the opportunity to ascertain actions being initiated within the Harlequin ES. As this is the default option no further information will be provided in this thesis. The role of *Select AML* is central to the interrogation and updating components of the system and will be considered in relation to these functions.

8.2.1 Analyse Data

The *Salt Manager* system contains two separate rule bases which analyse environmental information to produce two types of land classification maps; an existing and potential groundwater discharge map; and a potential groundwater recharge map.

To initiate the land classification process the user first connects the *Salt Manager* system to the environmental data held within the database. This task is achieved by the user selecting, *Update database*, then *Connect to RDBMS*, from the menu bar. A flow diagram summarising the methodology is displayed in Figure 8.3.

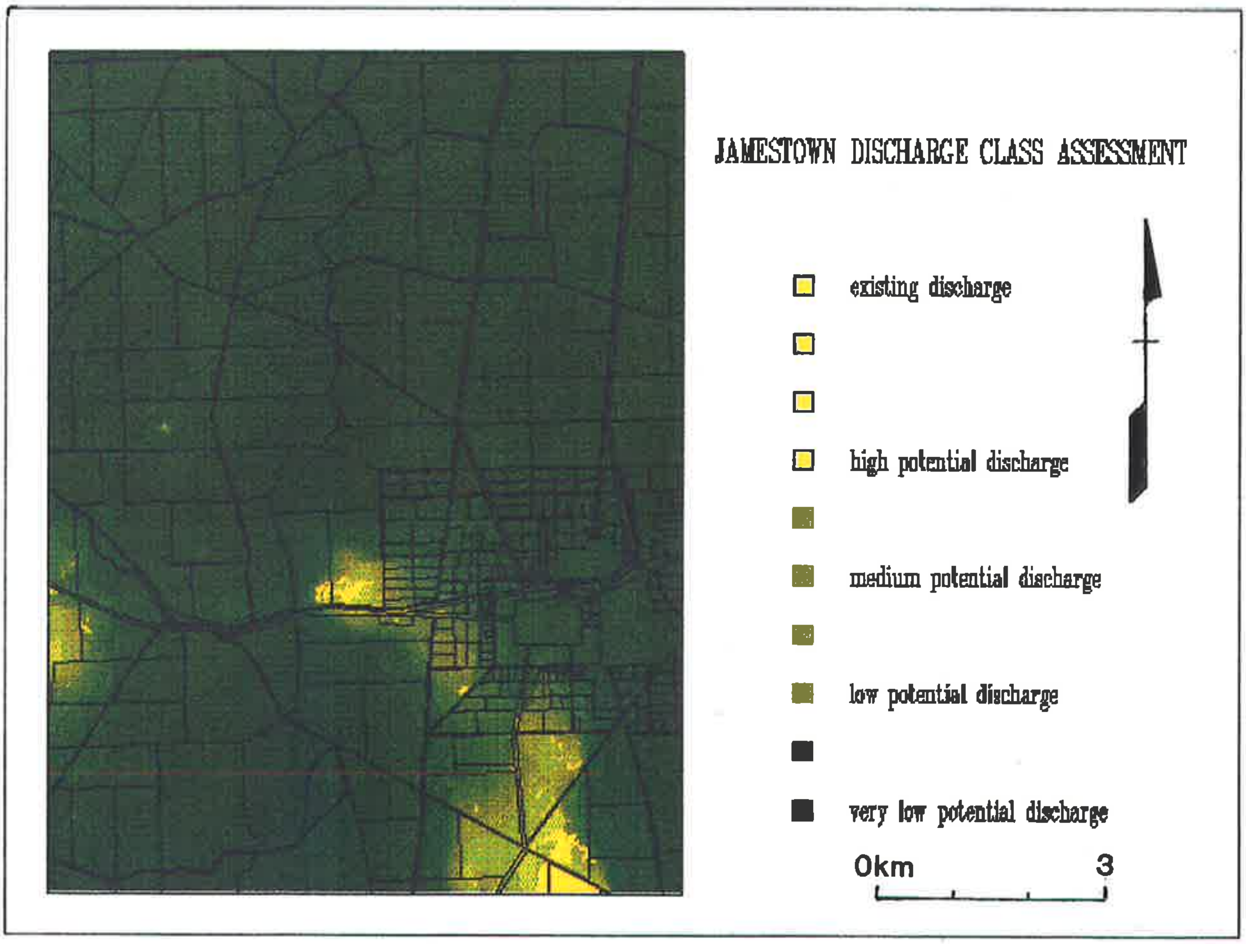
FIGURE 8.3: The Classification Process



To activate the inference engine the command *Analyse Data*, is then selected followed by *calculate salt-class*. Once the rule base has classified the environmental data the result can then be displayed.

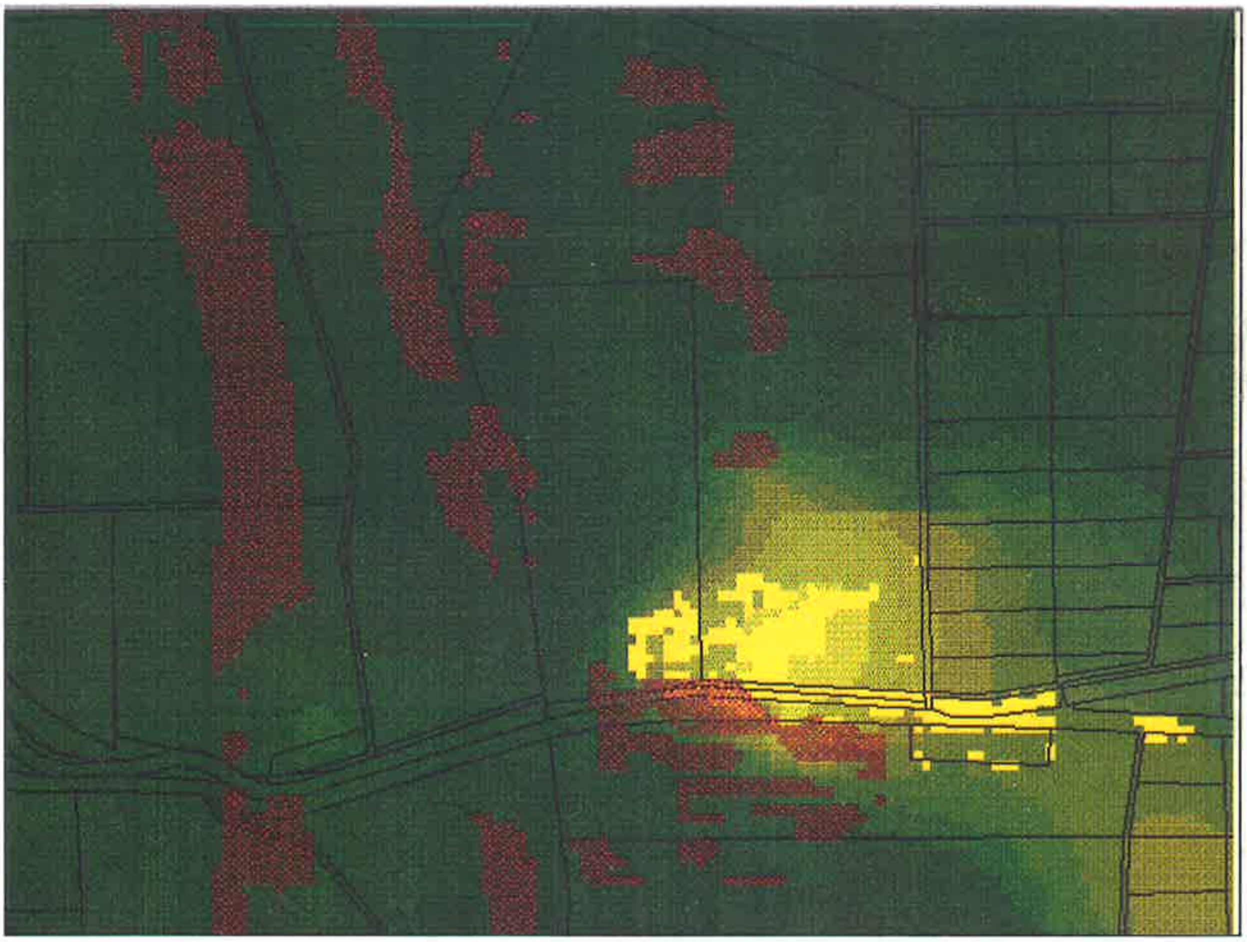
To initiate this process the user simply selects the relevant map to view, i.e. recharge/discharge, from the AML list provided within the *Salt Manager* and then initiates the GIS display function by selecting *Load Arc*, from the *Analyse data* sub-menu. The land classification map is then displayed. The thematic land classification map in Figure 8.4 is the classification of existing and potential groundwater discharge in the Jamestown study region.

FIGURE 8.4: Existing and Potential Groundwater Discharge as Determined by the Salt Manager



The areas of bright yellow are the existing salt affected areas as determined by the Landsat TM remotely sensed image. Extending away from these areas are potential groundwater discharge zones, the darker the yellow the higher the potential chance of groundwater discharge occurring, while the darker the green the lower the chance of discharge occurring. A potential groundwater recharge map generated by the *Salt Manager* can also be displayed via this process.

FIGURE 8.5: Existing and Potential Groundwater Discharge and High Potential Recharge Areas for a Number of Small Cadastral Units



Using the sub-menus designed for each of the AML's the user can manipulate and add other information to the displayed image. Options provided by this menu are listed in Figure 8.2. Using the added functionality of the AML menus, Figure 8.5 displays how a user can zoom on the displayed groundwater discharge map to identify a specific location. They can subsequently overlay cadastral information (a vector coverage within the GIS) or overlay potential groundwater recharge zones, areas of red on Figure 8.5. This added functionality of overlaying potential groundwater recharge zones is important as groundwater recharge and groundwater discharge zones are not mutually exclusive.

The land classification, identifying areas of either potential groundwater recharge and discharge or both, can then be interrogated by the user.

8.2.2 Interrogate Data

The user is offered the option of interrogating the decision making process which results in the land classification map produced by the *Salt Manager*. In this example, the groundwater discharge classification, will be interrogated. A flow diagram, Figure 8.6, has been developed to illustrate the process.

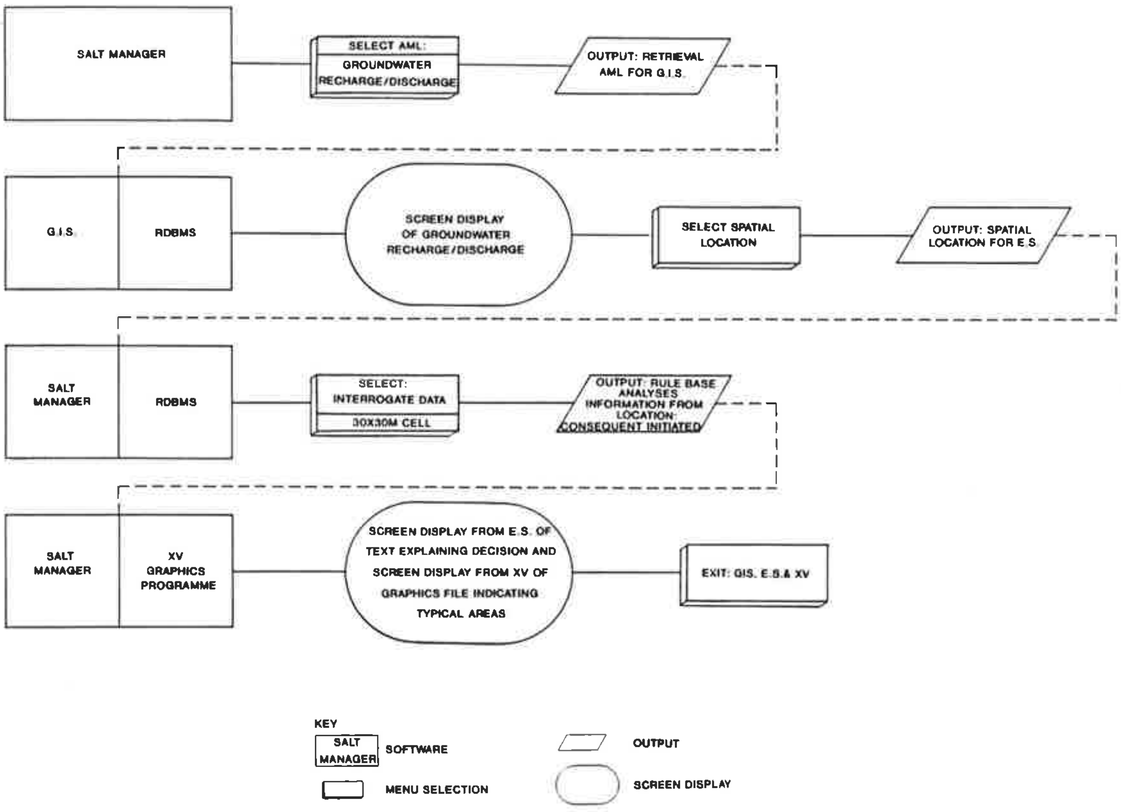
To commence this process the user simply selects the discharge/recharge classification to be displayed from the *Salt Manager*. This process is documented in Figure 8.6. Once the map is displayed the user may interrogate the map at the gridcell level ($30m^2$), by first selecting from the AML menu displayed by the GIS, the *select area* option. Subsequently, the user then selects the area they wish to interrogate by simply pressing the mouse button. Once selected, the user then presses *interrogate data/30x30m cell* from the *Salt Manager* menu bar.

Automatically, the rule base fires over the environmental information specified at that location[2]. Once the antecedent component of the rule has been satisfied the consequent is then initiated.

---

[2] Again, at this stage the type of rule base, either recharge or discharge must be first entered manually. Automation of this process is not difficult and will be initiated in the very near future.

FIGURE 8.6: Interrogating the Land Classification

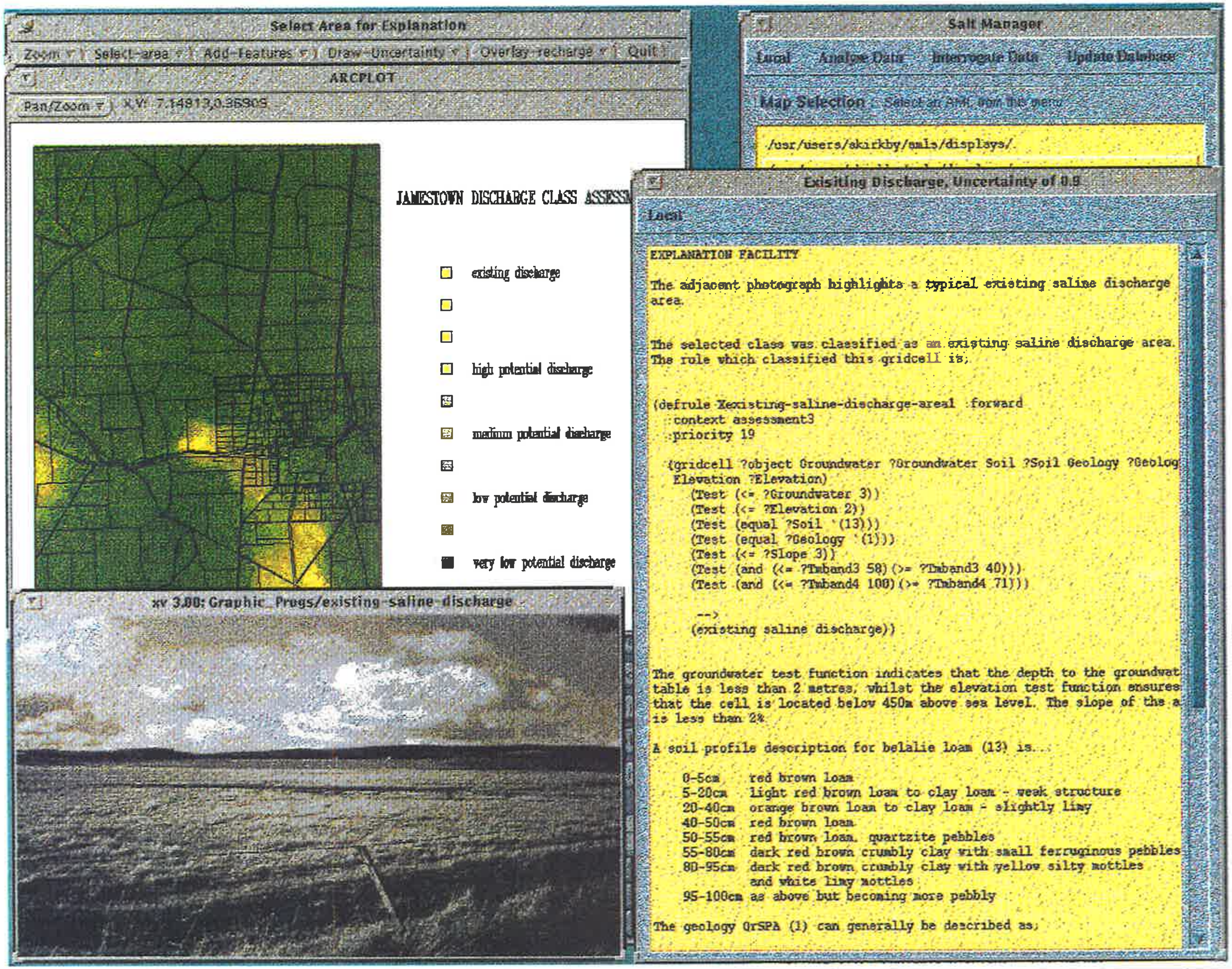


In this case, the different rule base initiates a contact file displaying text information that explains the decision making process and importantly provides a management strategy appropriate for the environmental conditions[3]. Concurrently, it also initiates a graphics file showing an area indicative of the classification. Figure 8.7 illustrates the results of the interrogation procedure. The Certainty Factor (CF) for the classification at the specific gridcell location is also displayed on the X-window title bar.

---

[3] It must be noted that there are four rule bases; two each for the groundwater recharge and discharge classifications. There is no difference between the antecedent of each of the two respective rule bases, where they differ is the consequents that they initiate. Obviously the rule bases utilised to interrogate the land classification initiate explanatory acts as their consequents, e.g. initiate graphics and explanatory text.

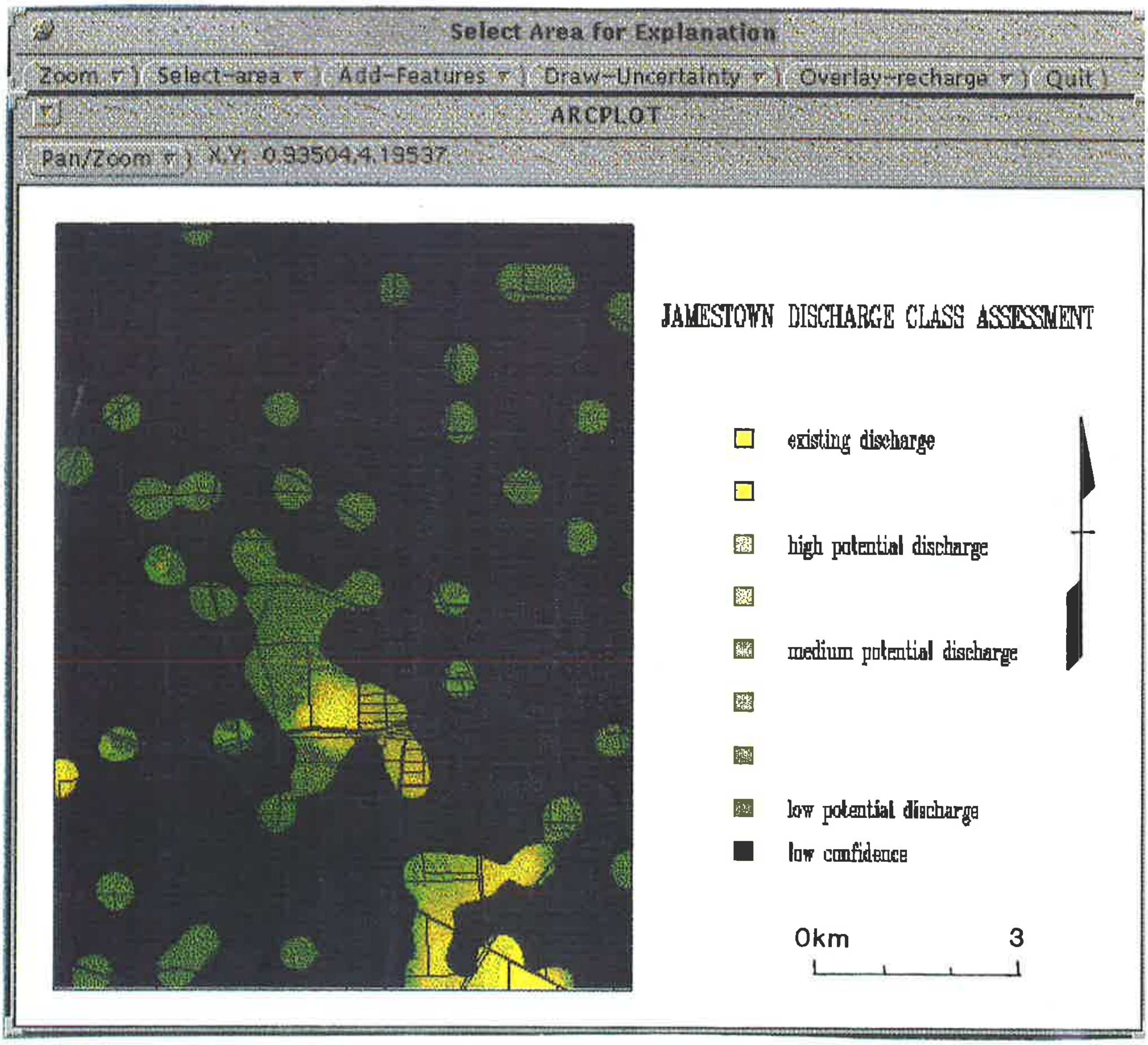
FIGURE 8.7: **Graphical Result of the Interrogation Procedure**



```
EXPLANATION FACILITY

The adjacent photograph highlights a typical existing saline discharge
area.


The selected class was classified as an existing saline discharge area.
The rule which classified this gridcell is;


(defrule Xexisting-saline-discharge-areal :forward
  :context assessment3
  :priority 19

  (gridcell ?object Groundwater ?Groundwater Soil ?Soil Geology ?Geolog
   Elevation ?Elevation)
     (Test (<= ?Groundwater 3))
     (Test (<= ?Elevation 2))
     (Test (equal ?Soil '(13)))
     (Test (equal ?Geology '(1)))
     (Test (<= ?Slope 3))
     (Test (and (<= ?Tmband3 58)(>= ?Tmband3 40)))
     (Test (and (<= ?Tmband4 100)(>= ?Tmband4 71)))

      -->
     (existing saline discharge))


The groundwater test function indicates that the depth to the groundwat
table is less than 2 metres, whilst the elevation test function ensures
that the cell is located below 450m above sea level. The slope of the a
is less than 2%

A soil profile description for belalie loam (13) is...

     0-5cm    red brown loam
     5-20cm   light red brown loam to clay loam - weak structure
     20-40cm  orange brown loam to clay loam - slightly limy
     40-50cm  red brown loam
     50-55cm  red brown loam, quartzite pebbles
     55-80cm  dark red brown crumbly clay with small ferruginous pebbles
     80-95cm  dark red brown crumbly clay with yellow silty mottles
              and white limy mottles
     95-100cm as above but becoming more pebbly

The geology OrSPA (1) can generally be described as;
```

CFs, as noted earlier (section 7.4.1), are only developed for the groundwater discharge classification. The results of the uncertainty analysis can be graphically presented to the user. The menu options provided by the AML allow the user to display the degree(s) of (un)certainty associated with the classification (Figure 8.8). This figure highlights areas where the land classification is believed to have greater than 0.85 CF degrees of confidence. The areas of yellow are existing discharge whilst the areas of darker green are considered to have lower discharge potential.

FIGURE 8.8: Areas of (un)Certainty Associated with the Groundwater Discharge Land Classification

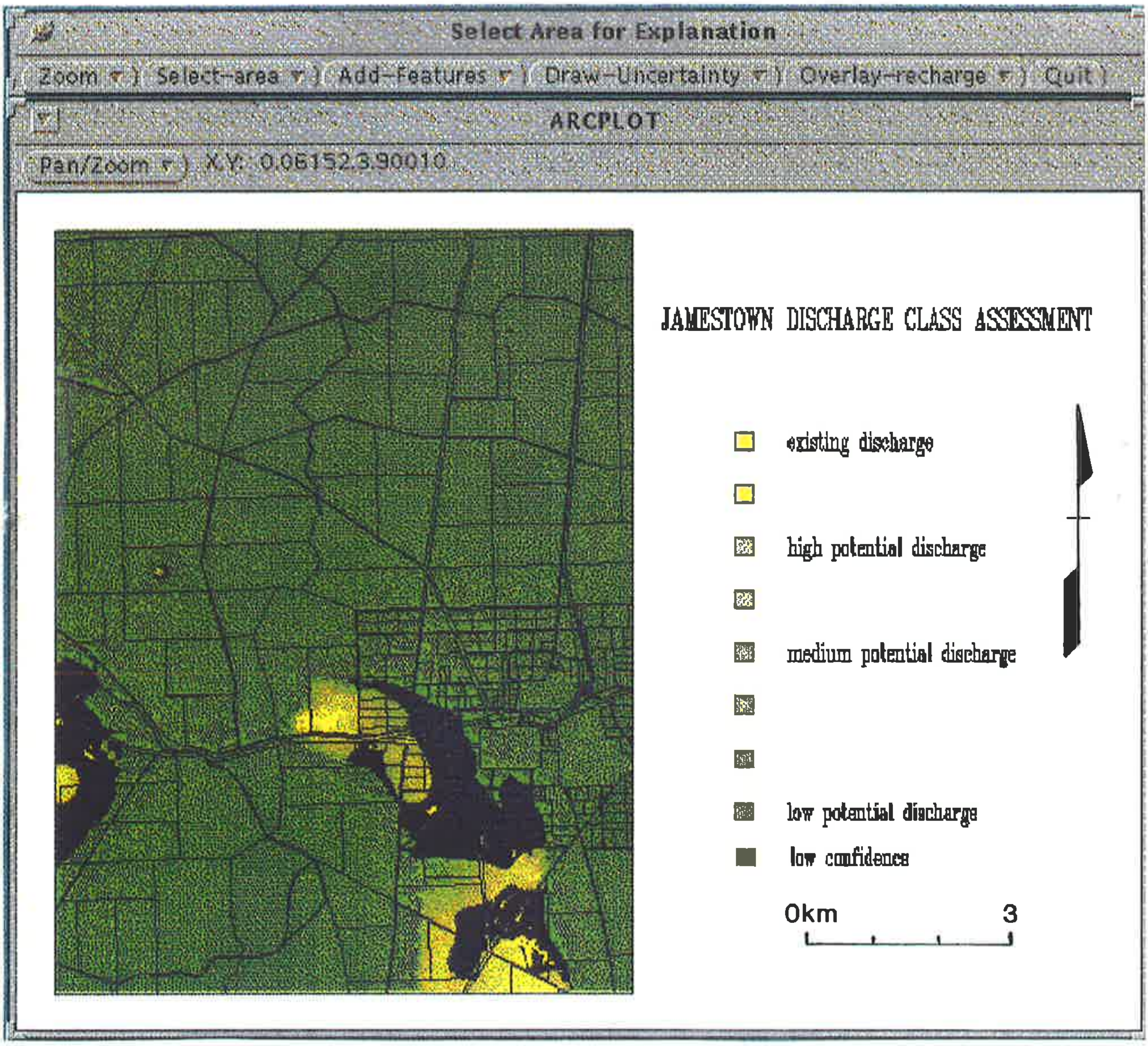


X,Y: 0.93504,4.19537

The areas of black are where the classification does not meet the selected confidence level. Noticeably, the areas where the classification does satisfy the confidence level occur directly around the sample points. Yet, logically, in relation to groundwater discharge other factors must be included.

According to the experts, groundwater discharge is most likely to occur when the groundwater is within two metres of the earth's surface.

FIGURE 8.9: Areas of (un)Certainty Associated with the Groundwater Discharge Classification (>0.85CF Confidence)



Thus, in the regional highlands in the north of the study area where, even though there may be a higher degree of uncertainty regarding the accuracy of the groundwater level, a kriged value of 20 metres can, with a high degree of certainty, infer a lower chance of groundwater discharge occurring. It therefore follows that all areas where the groundwater is below six metres from the earth's surface generally have a very low chance of groundwater discharge occurring. Figure 8.9 displays the inclusion of this logically sound concept.

As indicated by this diagram, the areas where the classification does not meet the confidence level require further intensive field investigation, especially for the "depth to groundwater" environmental data parameter. Once the user has completed interrogating the decision making process they may then update the database.

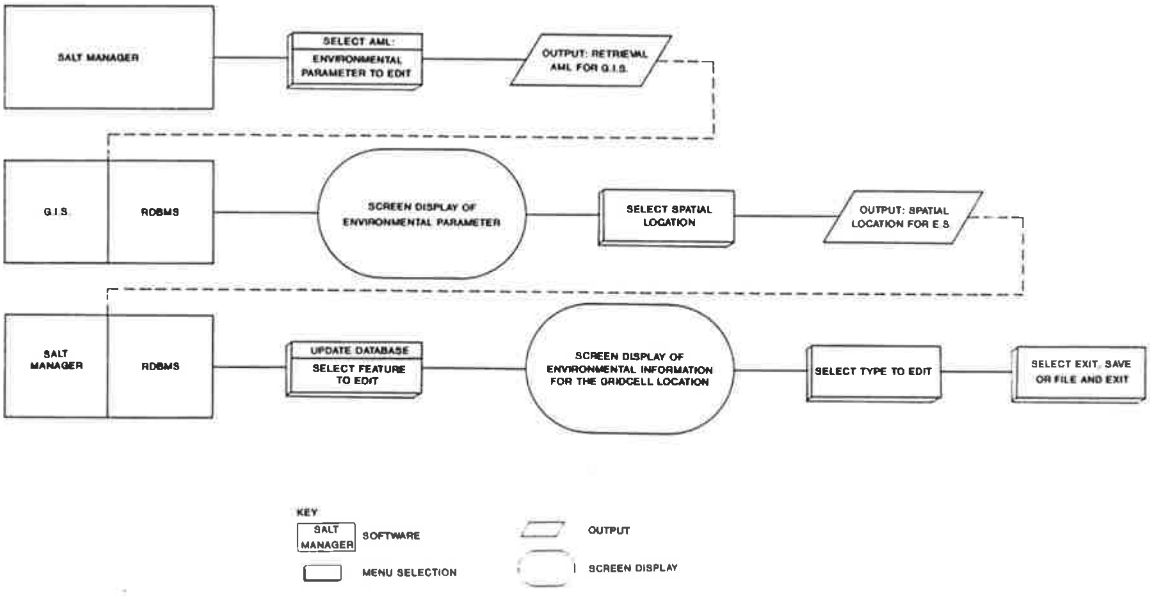
8.2.3 Update Database

Many landholders have valuable knowledge of environmental parameters, such as depth to groundwater, and it is important to capture this information. Updating the database may only occur at the scale of individual gridcells. There are three different methods applied to update the environmental data, with each being specific to a particular environmental data type. The first is via a pop-up contact screen from the ES. Using this update method, the parameters of soil, geology, slope and elevation may be updated. As these parameters are static it is not expected that the end user would alter any of the information provided. Yet the opportunity must be made available. The second is by editing an Arc/Info point coverage in the GIS, this is particularly relevant for the dynamic parameter "depth to groundwater". Figures 8.10 and 8.12 are flow diagrams indicating the decision making process for these two methods. The third is by directly updating the data within the RDBMS. This option will not be considered as only experienced Oracle database managers would be able to complete this task.

8.2.3.1 Update Database via the ES

The first steps in this process are the same as previously considered in the "analysis" and "interrogation" functions of the system (Figure 8.10). The end user selects from the *Salt Manager* menu an environmental feature to view, e.g. soil.

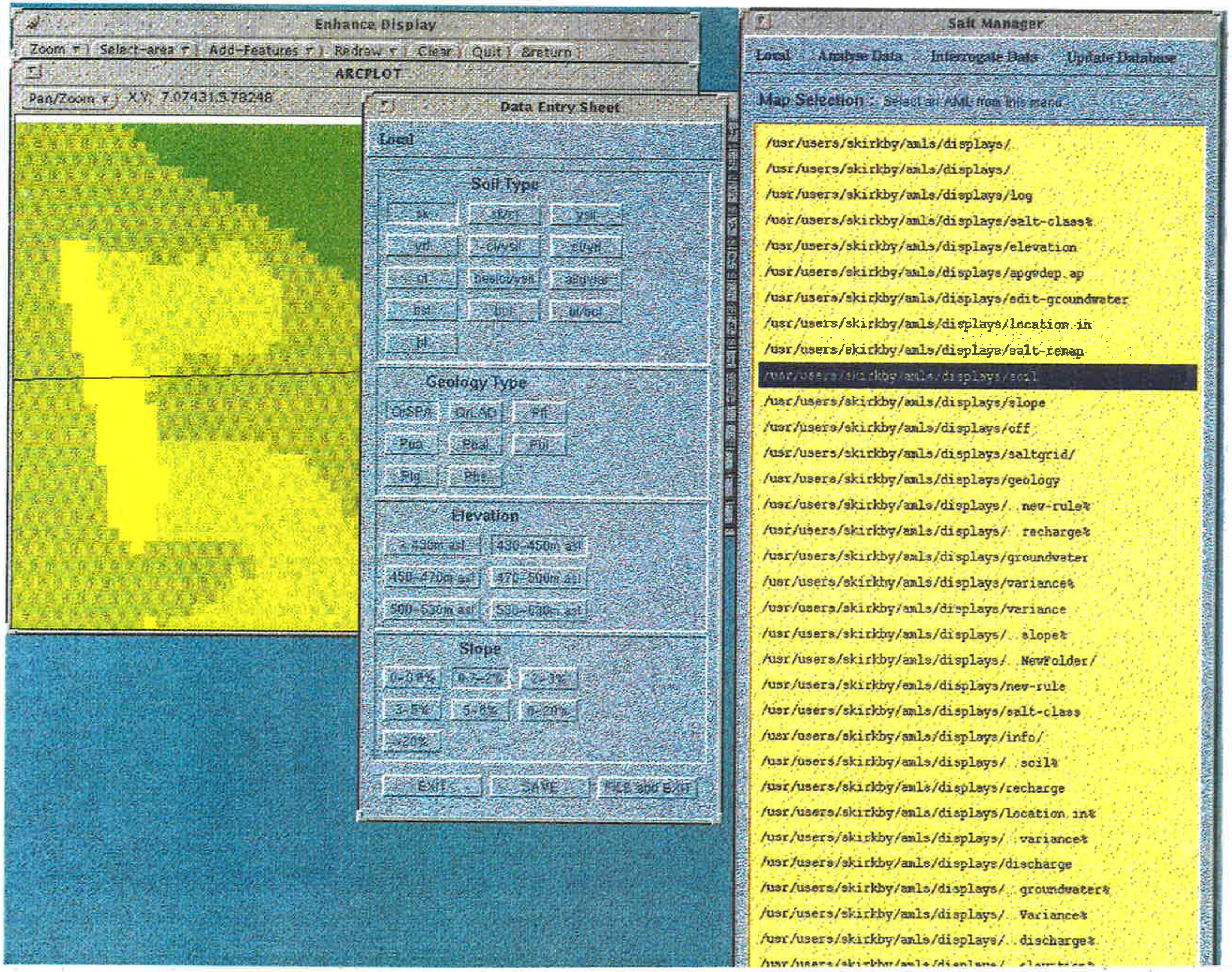
FIGURE 8.10: Updating the Database via the ES



It may be the case that the feature is selected due to a disagreement by the user with an explanation provided by the interrogative component of the *Salt Manager* system, e.g. "the soil in that area does not fit that description". Once, in this case, the soil map is displayed using the functionality of the AML menu, the end user may zoom into an appropriate area and then select the appropriate location. By selecting from *Salt Manager*, the command *Update database*, followed by *select feature to edit*, a data input contact is initiated (Figure 8.11).

As indicated by Figure 8.11, the four types of attribute information, soil, geology, slope and elevation, can be edited by this contact. The cell's current value appears indented when the contact is established.

FIGURE 8.11: **Data Contact Window Enabling Updating of the RDBMS via the ES**



Map Selection : Select an AML from this menu

```
/usr/users/skirkby/amls/displays/
/usr/users/skirkby/amls/displays/
/usr/users/skirkby/amls/displays/log
/usr/users/skirkby/amls/displays/salt-class%
/usr/users/skirkby/amls/displays/elevation
/usr/users/skirkby/amls/displays/apgwdep.ap
/usr/users/skirkby/amls/displays/edit-groundwater
/usr/users/skirkby/amls/displays/location.in
/usr/users/skirkby/amls/displays/salt-remap
/usr/users/skirkby/amls/displays/soil
/usr/users/skirkby/amls/displays/slope
/usr/users/skirkby/amls/displays/off
/usr/users/skirkby/amls/displays/saltgrid/
/usr/users/skirkby/amls/displays/geology
/usr/users/skirkby/amls/displays/. new-rule%
/usr/users/skirkby/amls/displays/. recharge%
/usr/users/skirkby/amls/displays/groundwater
/usr/users/skirkby/amls/displays/variance%
/usr/users/skirkby/amls/displays/variance
/usr/users/skirkby/amls/displays/. slope%
/usr/users/skirkby/amls/displays/. NewFolder/
/usr/users/skirkby/amls/displays/new-rule
/usr/users/skirkby/amls/displays/salt-class
/usr/users/skirkby/amls/displays/info/
/usr/users/skirkby/amls/displays/. soil%
/usr/users/skirkby/amls/displays/recharge
/usr/users/skirkby/amls/displays/location.in%
/usr/users/skirkby/amls/displays/. variance%
/usr/users/skirkby/amls/displays/discharge
/usr/users/skirkby/amls/displays/. groundwater%
/usr/users/skirkby/amls/displays/. Variance%
/usr/users/skirkby/amls/displays/. discharge%
```

With each data input contact, a press of the middle button is an event that signals that "help" information is required. The result of a "help" event manifests itself in the occurrence of a popup window displaying a text file with relevant information. Upon completion of the updating process the user is asked by the *Salt Manager* if they wish to commit all changes to the database. If no, all updated information is cleared; if yes, the changes are committed to the RDBMS. Subsequently, the salinity land classification procedure may be initiated again by selecting the appropriate command from the *Salt Manager* menu.

It results in a new land salinity class map based on the new information being produced and written to the RDBMS. At this stage of development, the run/time for the ES processing the data within the RDBMS is approximately 3 hours. As the system is designed to have interactive capabilities with the user, this needs to be reduced.

8.2.3.2 Update Database via the GIS

By providing a menu from the editing module within the GIS the depth to groundwater data collection points may be updated, altered or deleted by the user. The process is illustrated by a flow diagram (Figure 8.12). Figure 8.12 highlights how the initial decision stages for the operation of the *Salt Manager* system are replicated for each function of the system.

FIGURE 8.12: Updating the Database via the GIS

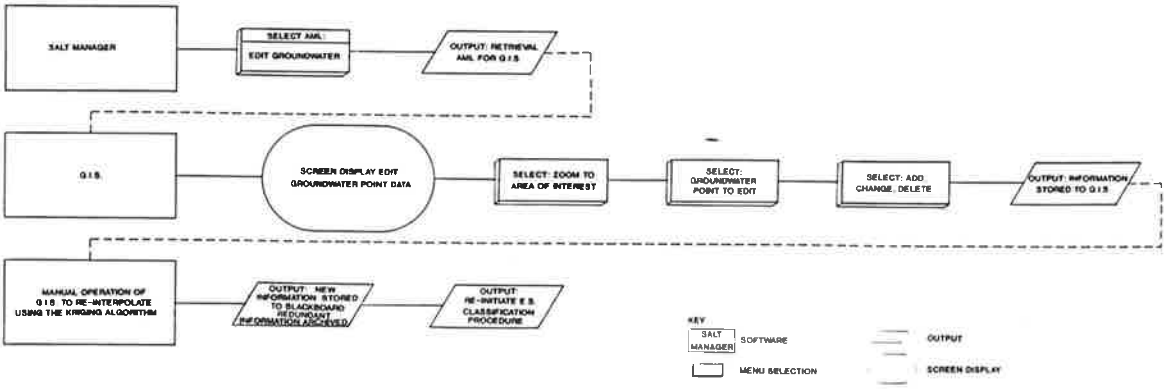
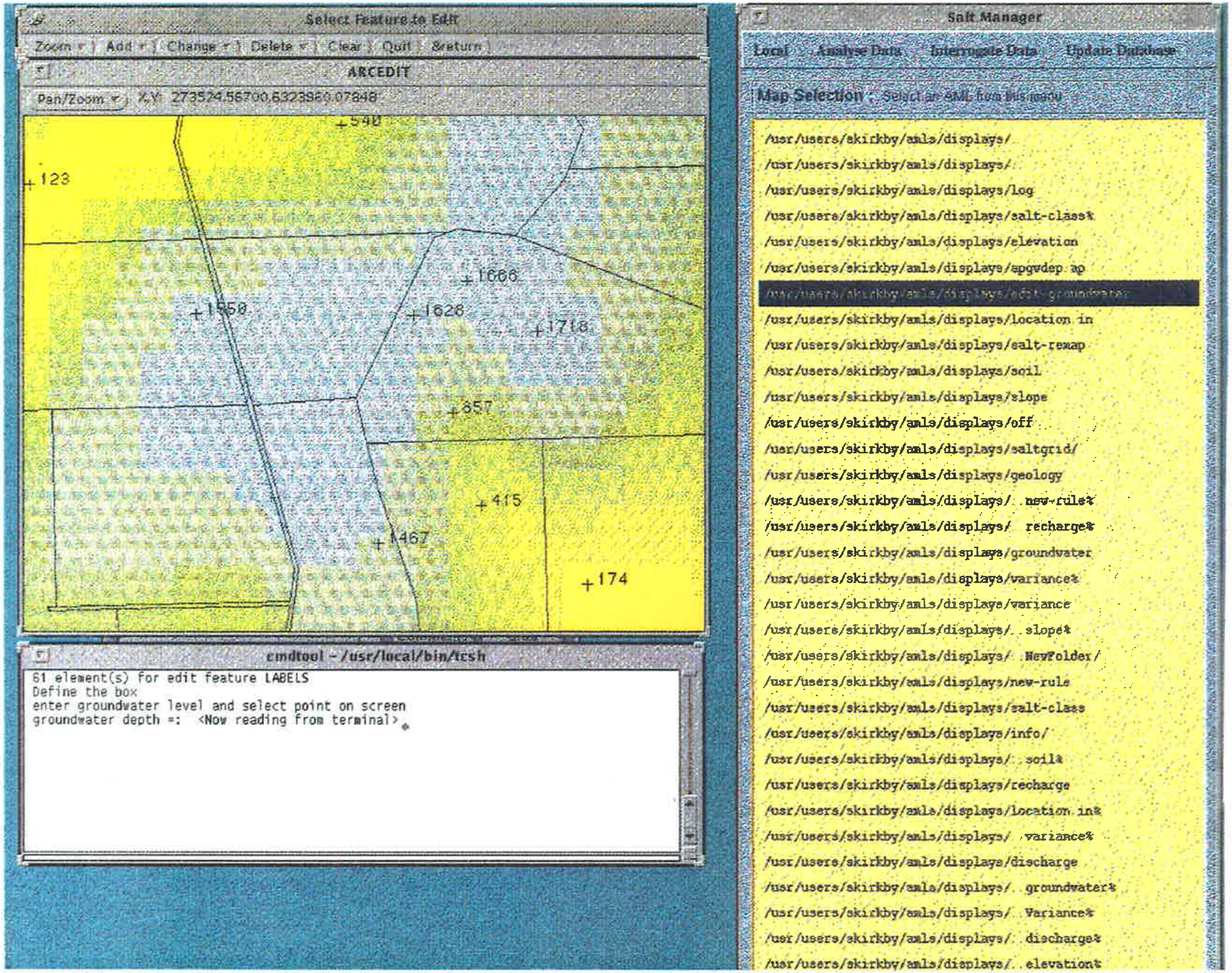


FIGURE 8.13: Editing the Depth to Groundwater Data from the GIS



```
61 element(s) for edit feature LABELS
Define the box
enter groundwater level and select point on screen
groundwater depth =:  <Now reading from terminal>
```

```
/usr/users/skirkby/amls/displays/
/usr/users/skirkby/amls/displays/
/usr/users/skirkby/amls/displays/log
/usr/users/skirkby/amls/displays/salt-class%
/usr/users/skirkby/amls/displays/elevation
/usr/users/skirkby/amls/displays/apgwdep ap
/usr/users/skirkby/amls/displays/edit-groundwater
/usr/users/skirkby/amls/displays/location in
/usr/users/skirkby/amls/displays/salt-remap
/usr/users/skirkby/amls/displays/soil
/usr/users/skirkby/amls/displays/slope
/usr/users/skirkby/amls/displays/off
/usr/users/skirkby/amls/displays/saltgrid/
/usr/users/skirkby/amls/displays/geology
/usr/users/skirkby/amls/displays/. new-rule%
/usr/users/skirkby/amls/displays/. recharge%
/usr/users/skirkby/amls/displays/groundwater
/usr/users/skirkby/amls/displays/variance%
/usr/users/skirkby/amls/displays/variance
/usr/users/skirkby/amls/displays/. slope%
/usr/users/skirkby/amls/displays/. NewFolder/
/usr/users/skirkby/amls/displays/new-rule
/usr/users/skirkby/amls/displays/salt-class
/usr/users/skirkby/amls/displays/info/
/usr/users/skirkby/amls/displays/. soil%
/usr/users/skirkby/amls/displays/recharge
/usr/users/skirkby/amls/displays/location in%
/usr/users/skirkby/amls/displays/. variance%
/usr/users/skirkby/amls/displays/discharge
/usr/users/skirkby/amls/displays/. groundwater%
/usr/users/skirkby/amls/displays/. Variance%
/usr/users/skirkby/amls/displays/. discharge%
/usr/users/skirkby/amls/displays/. elevation%
```

More importantly it also illustrates the editing functionality of the GIS. Each point value may be edited via the AML menu (Figure 8.13). The back coverage being displayed by the GIS (Figure 8.13) is the interpolated depth to groundwater values, where the yellow colours represent the groundwater which is closer to the surface of the earth and the blue/grey colours a deeper depth to the water table. The points represent the original groundwater data collection points. To change the groundwater depth value at each point (displayed in cm), the user selects the point and then enters the value (depth to groundwater) in centimetres. An addition to the point coverage requires that the new point is surveyed using GPS methods into the database. If any of the values are altered, added or selected, the changes are saved and a GIS operator is required to re-interpolate the groundwater coverage. Once interpolated the changes are saved to the related blackboard table in the RDBMS.

A feature of the updating process is the use of the archiving facility within the RDBMS. All changed values for each gridcell are stored in archive files accessible at any later date. This facility allows temporal analysis of environmental change once a history is established.

8.3 MANAGEMENT IMPLICATIONS

Before appropriate management strategies are implemented to inhibit the development of dryland salinisation, the key problem areas must first be identified. This is the first goal of the *Salt Manager* system and it is achieved by analysing the nominated environmental parameters with heuristic knowledge derived from salinity experts. To ensure that land owners understand why certain types of land have been classified as either having recharge/discharge potential the *Salt Manager* system implements a strategy which utilises text and graphic formats to convey information. Part of the information conveyed to the user is an appropriate management strategy for a selected area. It is important to combine the suggested management strategy with the decision making information, as it places the management strategy in context of the overall decision making process. It is envisaged that a better informed farming community will make better land management decisions.

8.4 SUMMARY

The driving force behind the design of this system is the desire to show the decision making process to an end user. It is a system designed to capture the heuristic knowledge of a number of experts and using the system capabilities aims to convey their knowledge to a wider group of people within the community. Uniquely, the *Salt Manager* acknowledges that land owners are themselves a group of experts. In the sense that land owners have expert knowledge of the environmental conditions existing on their properties, i.e. depth to water table. As data collection is time consuming and expensive; it seems better to use the appropriate expert land owner knowledge to collect important environmental data. *Salt Manager* can therefore been seen as implementing a strategy to partition the knowledge and data acquisition stages between the two appropriate "expert groups".

Central to this notion is the assumption that a landowner will be more likely to implement strategies conveyed to them when they have responsibility for the database. It is a "positive re-inforcement method", whereby land owners are firstly fully informed of the decision making process regarding the land classification via the graphic and text facilities of the *Salt Manager*, and secondly encouraged to enter data into the system, resulting eventually in the formulation of another land classification map. At this stage the author is unaware of similar systems which aim to both convey information regarding land classification in this manner and allows data updates from the end user.

Salt Manager is the implementation of a new land classification methodology which will be fully explained in Chapter 10. The following chapter tests the results produced by the *Salt Manager* system.

Chapter 9

VALIDATION, VERIFICATION AND TESTING

This section concentrates on verifying, validating and testing the accuracy of the results produced by the integrated *Salt Manager* system. In informal terms verification can be thought of as putting the question, are we building the product right? And validation as, are we building the right product? (Boehm 1981). In this section we will consider validation as the process of checking that the system actually does what the expert considers correct, while we will apply verification to check that the system meets the requirements of the end user. Testing the accuracy of the results produced is inherent in the validation of the *Salt Manager* system. The chapter will give consideration to three points: the validation of the results; the implementation of another machine learning experiment to quantify the effect of the additional EM data layer; and the verification of the *Salt Manager* system from the user's perspective.

9.1 VALIDATION

To validate the accuracy of the results produced by the *Salt Manager* system both the groundwater recharge and discharge thematic maps must be compared to ground truthed field data obtained from the 1x3km catchment within the study region.

9.1.1 Groundwater Recharge

The test sites to verify groundwater recharge were 8 nested piezometer sites. It should be noted that even though there were 17 point locations where piezometers were located in the test catchment, only 8 had multiple piezometers. The 8 piezometer sites were selected to characterise groundwater systems beneath the various soil-landform associations in the catchment. The 8 piezometer nests enabled vertical gradients at each site to be calculated and the direction of groundwater flow determined thus indicating if the site is acting as a recharge or discharge area. Unfortunately, intensive groundwater data collection over the 27 months, necessary to attain an understanding of the true recharge potential at each site, restricted the number of nested piezometer sites that could be monitored to only eight. But as large amounts of regular data are required to verify recharge potential, the number of sites selected were manageable for the resources allocated to this project.

Monitoring of water levels commenced at the site in February 1991 and were completed in May 1993 (Henschke et al., 1993). Daily rainfall was recorded on-site at the automatic weather station from May 1991. Prior to this, rainfall records for Jamestown were used. The groundwater hydrographs with daily rainfall superimposed are displayed in Appendix 6.

Prior to August 1992, piezometers located on elevated areas in the northern half of the catchment showed a declining trend with small seasonal responses to winter rainfall in 1991. Piezometers located in valley sites at relatively low elevations showed strong annual seasonal response to rainfall with amplitudes ranging from 1.0 to 1.5m. Site 14 located on a mid-upper slope in the southern part of the catchment also displayed a strong seasonal response to rainfall. Significant rises in water level occurred at all sites between August/September 1992 and December 1992/January 1993 (Henschke et al., 1993). This was due to the record rainfall which fell during the spring and early summer. In the period August 1992 to January 1993, 677.8mm of rain fell at Jamestown.

Regression analysis of the hydrographs was used to calculate recharge for the eight points. Recharge (R) was calculated from:

$$R = hS_y$$

where h = effective water level rise due to recharge (being the vertical distance from the zero recharge line to the hydrograph peak), and S_y = specific yield of unconfined aquifer (Henschke et al., 1993). In 1991 recharge rates (Table 9.1) ranged from 6 -

38mm (2-10% of 1991 total rainfall). In 1992 recharge rates ranged from 60 - 130mm (8 - 14% of 1992 total rainfall). Recharge rates are related to EM31 values for each site in Table 9.1, and show a reasonable inverse correlation (Henschke et al., 1993).

TABLE 9.1: Recharge Rates for the Nested Piezometer Sites

Site	EC (mS/m)	Recharge 1991	Recharge 1992
7	34	-	130
2	60	6	60
6	73	28	110
8	80	32	90
14	87	38	92
5	100	6	62
4	105	28	66
1	120	8	70

According to Henschke et al. (1993) the responses indicate that site 7 is located in a high potential recharge zone, 4, 6, 8, 14 in medium potential recharge zones and 1, 2 and 5 in low potential recharge zones.

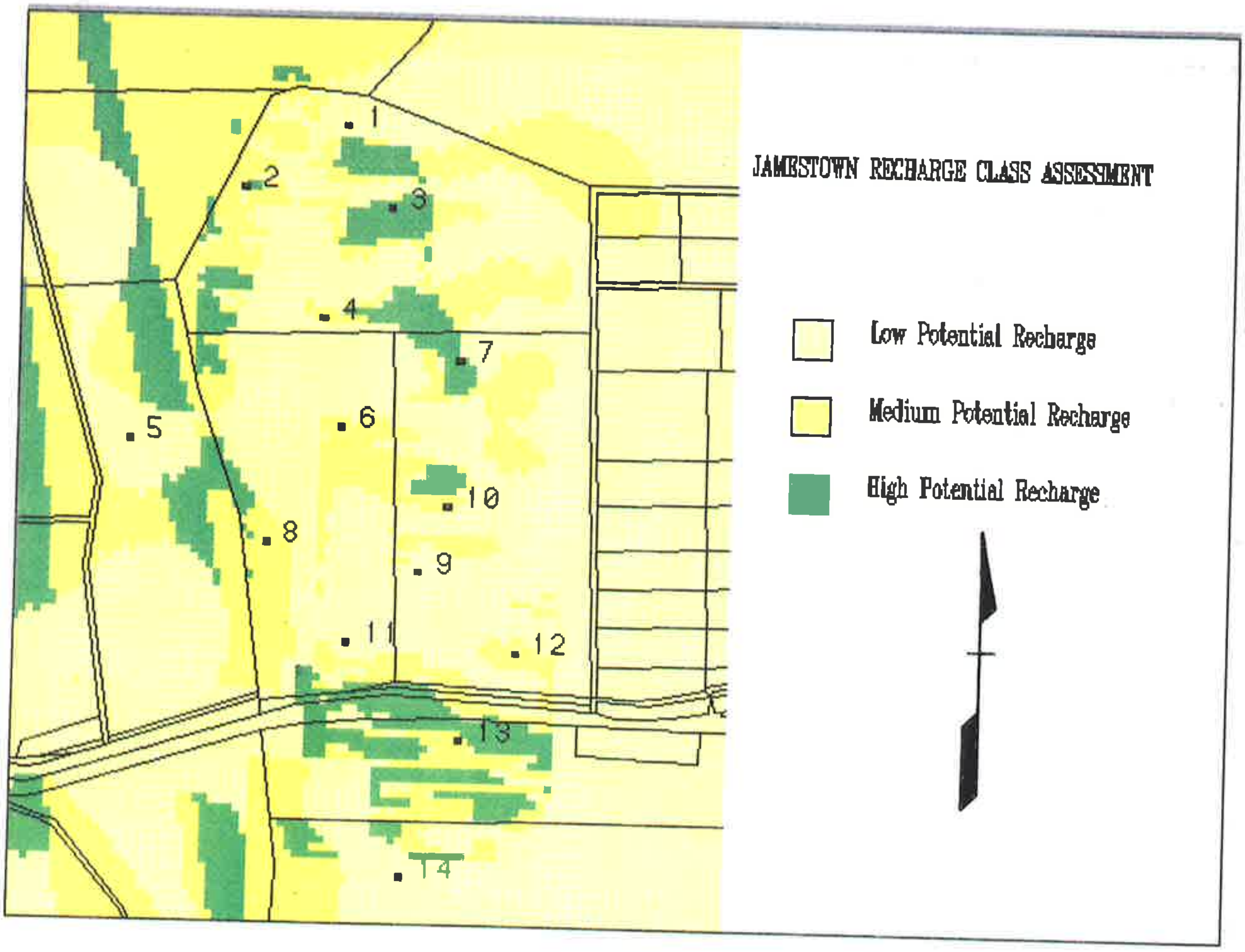
To validate the accuracy of the potential groundwater recharge map generated by *Salt Manager*, the eight nested piezometer points, also registered in the GIS, were intersected. The results are indicated in Matrix 1 (Table 9.2).

The accuracy of the classification is 75% (obtained by dividing the sum of the diagonal by the total number of cases. Misclassified classes were one each for Medium Potential Recharge (MPR) and Low Potential Recharge (LPR) (Figure 9.1). Obviously eight control points is not enough to conclusively state the accuracy of the classification. Nevertheless, it does at least provide an indication of accuracy for the potential groundwater recharge map produced by the *Salt Manager*.

TABLE 9.2: **Field Data Compared to the Salt Manager Groundwater Recharge Land Classification**

		Salt Manager Classification			
		HPR	MPR	LPR	Total
Field Data	HPR	1			1
	MPR		3	1	4
	LPR		1	2	3
Total		1	4	3	6

FIGURE 9.1: **Comparison of Field Sites to the Salt Manager Groundwater Recharge Land Classification.**



9.1.2 Groundwater Discharge

To validate the accuracy of the areas deemed to be salt affected by the *Salt Manager*, a map indicating areas, within the Jamestown test catchment, known to be salt, affected was independently produced by Chris Henschke in October 1991.

To differentiate the saline area, the procedure employed was based on the accepted relationship between halophytic vegetation and increasing rates of soil salinisation (Kirkby et al., 1994a). The six dominant vegetation classes used to identify the saline affected areas were: sea barley grass (Hordeum hystrix), rye grass (Lolium rigidum), wild oats (Arena fatua), spurgalaria (Spurgalaria marginata), samphire (Samphire halasarcia), puccinellia (Puccinellia cilata), and the bare soil patches and seepage areas typically indicative of extreme salt affected soils. Mapping of these indicators involved locating a position on an enlarged 1:2500 1990 colour aerial photograph of the whole catchment area.

Once he was confident of his true location, Chris Henschke proceeded around the perimeter of the halophytic vegetation and marked the apparent boundaries of the halophytic vegetation onto the 1990 aerial photograph. The saline/non-saline boundary was clear but could not be separated into distinct areas indicative of one specific vegetation cover class. After circumnavigating the whole area, he proceeded to walk along three transect lines through the area to determine the internal saline/non-saline boundaries. Particular attention was given to a small hump of high ground which was representative of non-salt affected soils within the discharge zone.

The result derived from this field work was a saline/non-saline boundary map of the discharge area traced onto a 1990 aerial photograph. This boundary was then traced, without ground control points, onto a technical drawing sheet. At a later stage, 17 piezometer points and the catchment boundary were drafted onto the saline map.

There are two methodological weaknesses associated with this procedure. Firstly, the lack of accurate ground control points is a fundamental weakness. A saline discharge map was produced previously by the author and Chris Henscke with 12 referenced ground control points for another study site (Kirkby, 1993b). This saline discharge map produced Root Mean Square (RMS) errors under 0.003 resulting in an accurate representation of the discharge error. This method was to be replicated by this project but due to an oversight the referenced ground control points were not recorded. This error was not conveyed to the author until January 1994, when the author contacted Chris Henschke to obtain his independent results.

The second relates to the use of a colour aerial photograph on which the perceived saline/non-saline boundary was drawn. It is possible that the tones and colours of the aerial photograph may have influenced the marked position of the boundary. The positional errors from each of these issues have not been quantified.

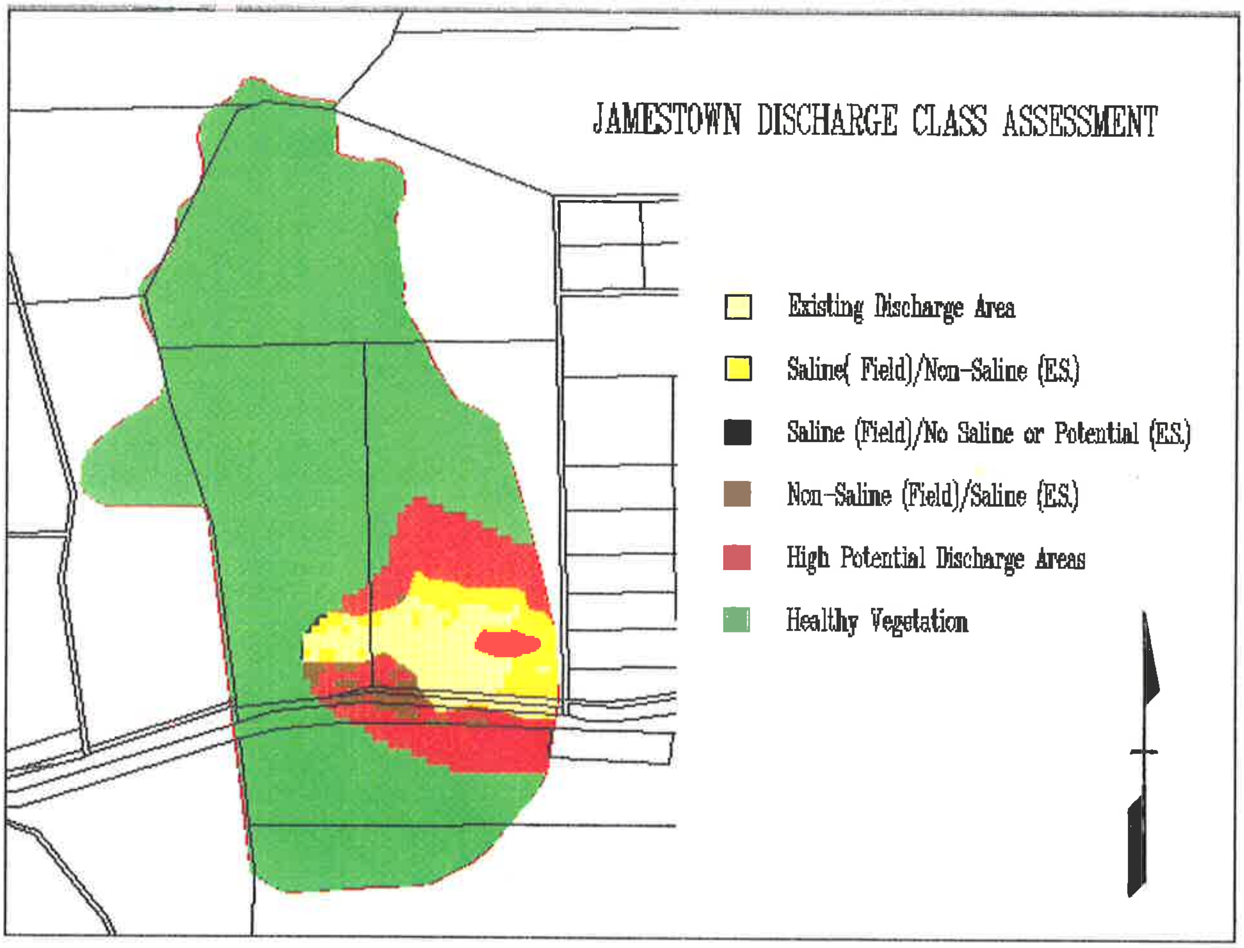
To validate the accuracy of the salt affected areas identified by *Salt Manager*, the saline/non-saline map developed by Chris Henschke was imported into Arc/Info and analysed in relation to the saline discharge land classification produced by *Salt Manager*. The results are displayed by Table 9.3. Table 9.3 indicates the accuracy of the *Salt Manager* classification in relation to the 1991 field data. To ascertain the overall accuracy of the classification the sum of the diagonal values (288.80 hectares) is divided by the sum total (305.99 hectares).

TABLE 9.3: Field Data Compared to the Salt Manager Groundwater Discharge Land Classification

	Salt Manager Classification		
	Existing Discharge	Healthy	Total
1991 Field Data Existing Discharge	20.53	11.71	32.24
1991 Field Data Healthy	5.48	268.27	273.75
Total	26.01	279.98	305.99

Thus the overall accuracy for the *Salt Manager* classification is 94.38%. The value which is of interest is the accuracy of the salinity classification, where 20.53 hectares of the 1991 field work classification has been correctly identified by the *Salt Manager* classification. This represents an accuracy of 64% (20.53/32.24), which is unacceptably low. As previously mentioned the cause of this problem may derive from a weakness in the field data collection methodology. These results are graphically represented by Figure 9.2.

FIGURE 9.2: Comparison of Field Data to the Salt Manager Groundwater Discharge Land Classification



As presented by Figure 9.2 the areas of pale yellow are where the *Salt Manager* classification and the field data correspond (20.53 hectares). The areas of yellow are where the *Salt Manager* has classified areas as healthy vegetation when the field data has classified it as salt affected (11.71 hectares). Similarly, the areas classified as salt affected by *Salt Manager* but classified as healthy vegetation by the field data are represented in brown (5.48 hectares). To further validate the results of the *Salt Manager* classification, areas of high potential discharge (<220 on the Expert scale Table 7.3) were compared to the field data. The areas of black represent the areas where the *Salt Manager* classification has not identified a high potential chance of groundwater discharge occurring, yet the field data has classified it as existing (0.36 hectares). Areas of red represent areas classified as high potential by the *Salt Manager* classification with the field data classifying these areas as healthy vegetation. The areas of green are those classified as healthy by both classifications (268.27 hectares). Interestingly, the areas classified as high potential discharge by the *Salt Manager* classification cover all but 0.36 hectares of the field data classification. Theoretically, the *Salt Manager* classification should have accounted for all of the areas identified as salt affected during the field work. The reasons for why the 0.36 hectares remains unaccounted for will be considered in Chapter 10.

Validation of potential groundwater discharge groups is difficult given the short time frame allowed for this study, i.e. identification of groundwater discharge enchroachment would take quite a number of years. Nevertheless, feedback provided by Henschke et al. (pers. comm., 1994) supports the notion that the groundwater discharge area will expand in a NNW direction.

Similarly the results attained by the *Salt Manager* groundwater recharge classification were also deemed to be very positive by Henschke et al. (pers. comm., 1994). Given that the *Salt Manager* classification for groundwater recharge did not exploit any depth to groundwater information to develop the classification, unlike most other approaches, Henschke et al (pers. comm., 1994) have stated that the *Salt Manager* classification method is a very feasible means by which to ascertain groundwater recharge potential for regional areas. It is a method which addresses problems considered in section 2.3.1.

Ideally, validation of the *Salt Manager* groundwater recharge classification occurring outside of the test catchment area would have confirmed the statement made by Henschke et al. (pers. comm., 1994). This was not the case and a problem evolves from this situation. The catchment is the only area where EM data has been collected, thus the rules which contain EM antecedents are only applicable within the 1x3km catchment zone. As the field data which validates the results comes only from inside the catchment area, extrapolation of the accuracy of the results to the region cannot be directly made.

Comparisons of the independently obtained groundwater recharge field data points to the *Salt Manager* classification, prior to the addition of the EM data, can be made but the accuracy must be considered dubious as the first rule base was designed to include mutual exclusivity for both groundwater recharge and discharge classes. In light of this point, when the initial *Salt Manager* recharge classification is compared to the independently obtained field data the results indicate a 50% accuracy. Four of the eight points (1, 2, 5, 8) were correctly classified whilst the other four (4, 6, 7, 14) were not. To account for this problem the groundwater recharge ES rule base could have been redesigned. This option was rejected as it is clear that the new ES groundwater recharge rule base containing the EM data is definitely more accurate. Rather than precisely document how much more accurate it is, the author assumed it would be more productive to acknowledge the fact that EM data does increase the opportunity of identifying recharge zones (as indicated by other authors in section 3.4.2). Moreover, given "summary point 3" in section 3.4.2, that EM data acquisition is on the increase, a new experiment to verify the effect of the EM data on the final *Salt Manager* classification was initiated. This approach was considered to be a more valid and pertinent option given the increasing effort being given to EM data collection for agricultural regions in Australia. The next subsection considers this experiment.

9.2 VERIFYING THE EFFECT OF THE NEW EM DATA LAYER ON THE DATA COLLECTION REGIME

This component of the results section describes a methodology to assess the addition of remotely sensed EM31 data layers to a rule-based model (Eklund et al., 1994). In particular, it considers the effects of the addition of an EM31 data layer to an existing GIS database containing groundwater depth, soil, geology, slope, elevation and the seven spectral bands of Landsat TM.

To initiate the experiment the field survey groundwater recharge data generated by Heschke et al. (1993) was added, together with the EM31 generated for 1x3km catchment area to the GIS. The resulting database was input to an inductive learning program and produced a decision tree classifier.

The intent of the experiment was to discover: 1) whether the resultant inductive decision tree encapsulated any significant attribute value discriminators in the new EM31 coverage layer; 2) whether the data contained within existing GIS coverages already informationally subsumed the new EM31 coverage; and 3) whether any conclusions were forthcoming about the classification task without prior knowledge of the nature and affect of the EM31 coverage in the domain classification task.

This controlled experiment was able to predict the effect of the EM31 data coverage on groundwater recharge and corresponds to expert domain knowledge on the effect of an EM31 data layer on groundwater recharge.

9.2.1 Results

The principal test of the inductive learning program (C4.5) was to determine whether knowledge of groundwater recharge classes which follow from Table 9.4, could be recovered from the database using the Henschke et al. (1993) field survey recharge data from the 1x3km catchment area. This would have the effect of independently verifying and perhaps recovering additional knowledge concerning the effects of the EM31 layer on the groundwater recharge classes.

The C4.5 program produced a decision tree which contains a number of interesting rules.

Rule A:

```
        groundwater <= 12
        slope >1
        soil in {3}
-->     class 6 (HPR) [89.9%]
```

The occurrence of Yanga Silty Loam (soil type 3) verifies the expert knowledge from Table 9.4 but the non-appearance of EM31 data in Rule A indicates the subsuming role of soil in the high potential recharge class. In contrast, Rule B and Rule C illustrate the importance of EM31 data for other soil types.

TABLE 9.4: Interpreted Groundwater Recharge Classes based only on Soil and EM31 Data from Henschke et al. (1993)

Recharge Potential	Area of Catchment (ha.)	Dominant Soils	EC Zone (mS/m)
High	71.2 (23.7%)	Yangya silty loam Canowie loam Yarcowie loam	0 - 60
Moderate	116.90 (38.8%)	Belalie loam Belalie clay loam	60 - 140
Low	112.9 (37.5%)	Belalie loam	140 - 300

Rule B:

```
        elevation >2
        groundwater <= 6
        slope >3
        emver <4
        soil in {13, 3, 12}
-->     class 6 (HPR) [95.9%]
```

In light of the information produced by Rule A, Rule B substantiates expert opinion on Yanga Silty Loam (soil type 3) in Table 9.4.

Interestingly, Belalie loam (Soil 13) and Belalie clay loam (Soil 12) highlight necessary expansion of the Henschke et al. (1993) recharge classification categories. In other words, a low EM31 response does not preclude Belalie loam and Belalie clay loam soil types as Table 9.4 would argue. A second inference from Rule B is that because EM31 values can vary from low to high across all soil types, an EM31 response can be used to highlight conductivity changes within a single soil type.

```
Rule C
        groundwater <= 8
        soil in {13}
        emhor <= 3
        geology in {8,5}
-->     class 6 (HPR) [89.9%]
```

The idea of discriminating recharge between cells with the same soil type using EM31 is further supported by rule C. This rule (and its twin, rule B) discriminate between high and medium recharge within a single soil type (Belalie loam (13)) using the EM31 response value.

The overall conclusion is that the EM31 coverage is, in general, subsumed by soil type and little additional knowledge is forthcoming from the EM31 coverage for discriminating groundwater recharge (Eklund, et al., 1994). This finding increases confidence in the quality of the soil data coverage.

However, it is precisely the exception to this general rule that is of interest. Additional high potential recharge cells were located using the EM31 coverage in the soil types, Belalie loam and Belalie Clay loam. Variations in the EM31 response indicate that cells previously considered to be low or medium recharge were, in fact, high recharge cells and that the EM31 value was the only discriminating attribute in those cells.

Therefore, for this study area, it is only Belalie loam and Belalie clay loam soils that EM31 can be used for predicting groundwater recharge areas.

This is a significant result because high potential recharge cells can be identified using EM31 and it is exactly these cells that are of interest - identifying high recharge is far more important than discriminating between low and medium potential recharge cells. High groundwater recharge cells are important to identify because they provide 36% of all groundwater recharge within the catchment - a further 28% derives from outside the catchment boundary Henschke et al. (1993).

The results suggest that an EM31 data collection regime should be initiated on the basis of the soil type and in particular restrict itself to the Belalie loam/Belalie Clay loam subset of the soil map for this catchment area. Because 58% of the catchment area has Belalie loam/Belalie Clay loam soil types, EM31 coverage for these soil types will prove useful in identifying groundwater recharge. EM31 data for the remaining 42% of the catchment need not be collected. The savings from this discovery are significant for any future EM31 studies conducted in this area. Especially when the EM31 data is collected in the labour intensive manner documented in section 5.6.

This subsection can be seen as presenting a valuable methodology for incrementally assessing the effects of additional data layer coverages on classification tasks in GIS domains. It demonstrates a technique for determining whether data collection regimes, which may be expensive, have any substantial impact on the classification task. This point is important to further collection of EM data in the regional salinity assessments. The technique also embodies the idea of knowledge discovery since the features of the new coverage which help in the classification may not be well understood.

The following subsection gives consideration to verifying whether the right "product" has been built for the end users.

9.3 VERIFICATION

To verify whether the integrated system developed was suitable for the end user, the "landowner", a meeting was conducted with the Bundaleer Valley Landcare Group in October 1993. At this meeting the *Salt Manager* system was shown via an extended series of slides, each showing the functionality of the system. At the end of the meeting the 9 landowners were asked to complete a questionnaire, during which time the author, left the room. The questionnaire is given in Appendix 7.

Of the 9 landowners 7 (77.78%) responded that they would be interested in using such a system. Furthermore, all of the seven stated that they would be interested in interrogating the thematic classification, updating the database and making use of the information to assist with farm management strategies. Supporting their positive reaction to the system was the setting up of a groundwater observation network in their Landcare area.

In conjunction with the regional Agricultural Officers, the landowners intend to monitor the depth to groundwater levels at nominated bore holes on each of their properties. The nominated boreholes will be surveyed by the Agricultural field Officers to determine the correct height above sea level. This network will provide the opportunity to undertake temporal studies on groundwater fluctuations in the area in relation to groundwater discharge development. It is proposed that the data collected by the landowners will be analysed by the *Salt Manager* system.

9.4 SUMMARY

Validation of the areas identified as groundwater recharge by the *Salt Manager* system is difficult due to the data intensive nature of the field data collection process. Nevertheless, the field data that was collected was considered by the author and Henschke et al. (1993) to be an excellent indicator of the types of groundwater recharge that is occurring within the 1x3km catchment. An accuracy rating of 75% for the potential groundwater recharge classification therefore provides encouragement for further refining of the *Salt Manager* system. Moreover, the second successful machine learning experiment to verify the role of the EM data within the classification task, encourages more efficient data sampling strategies in the refinement process.

The methodological problems associated with the collection of data indicating areas of salt affected land in the catchment resulted in the unacceptably low *Salt Manager* classification result. As previous work in other study areas has successfully identified these areas, this result is of no great concern. If the discharge area was mapped more accurately the results could have substantially improved.

The verification results have provided an interesting platform for further discussion. It is clear that the landowners are supportive of an interactive land classification procedure that requires data input from them. Further discussion on this concept will be considered in the following chapter.

Chapter 10

DISCUSSION AND CONCLUSIONS

This chapter presents a general discussion of the results of this research, considers an alternative land classification approach that derives from the application, identifies areas for further research and lists the final conclusions.

10.1 DISCUSSION

Results from Chapter 9 indicate that the *Salt Manager* analysis of the integrated datasets within the RDBMS and the GIS successfully identifies groundwater recharge and groundwater discharge in the 1x3km catchment in the Jamestown region of South Australia. As the method uses existing datasets, and those which are becoming more readily available, e.g. EM data, the heuristic approach encapsulated by the *Salt Manager* can therefore be applied as a method to ascertain salinity potential in other areas of South Australia.

An advantage of designing this integrated system using "off the shelf" generic software is that it can be adapted to any agricultural region in South Australia. As the only changes to the model that require implementation are the input of the relevant geographic data for the new area into the GIS/RDBMS and the coding of the area specific expert rules into the ES. Facilitating this transfer process will be considered later in this chapter.

Landowner feedback on the viability of the system is, at this stage, very positive. The importance of this point cannot be underestimated as the nature of the salinisation problem requires a coordinated integrated management approach from the landowners. Ideally, the future plan is to provide access to the software from the local Department of Primary Industries Jamestown office (local Agricultural Offices are densely distributed throughout the major agricultural regions of South Australia). This concept will be addressed later in this chapter.

Central to the success of the land classification method developed by this thesis for dryland salinisation problems is a change in how the information is presented to the user.

10.1.1 Conveying Thematic Map Information to the User

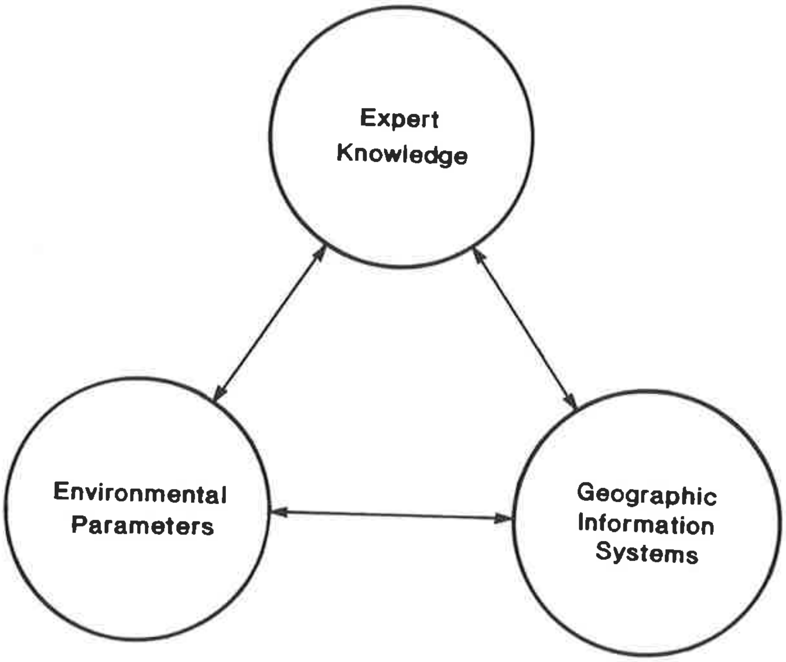
This sub-section describes the development of a land classification method, the *land understanding approach* (McKeown, 1987), which utilises an ES integrated with a GIS (Kirkby et al., 1994b). The *Salt Manager* system provides the basis for the development of this discussion.

10.1.1.1 Land Understanding Approach

The concept of the land understanding approach was first developed by McKeown (1987). He stated that this approach would "utilise knowledge about climate, soil composition, terrain, and general *a priori* knowledge in addition to the spectral information contained in remotely sensed imagery to infer land units suitable for a particular land use" (McKeown, 1987, p. 832). To initiate this approach he noted that an ES and a GIS, that could integrate, store and retrieve data layers was required, but he wrote that "these capabilities did not exist in any of the available GIS organisations" (McKeown, 1987, p. 832). Existing GIS are now able to perform such functions and are therefore able to be integrated with an ES. Research by Newkirk and Wang (1990), Kim and Swain (1989), Hadipriono et al. (1990) Goodenough et al. (1990) and Kirkby and Kurzel (1993a, 1993c) support this claim.

The land understanding approach combines the technical computer emphasis of the parametric/GIS approach and the subjective, qualitative nature of the landscape and genetic land classification approach, as it analyses selected data layers within a GIS with a series of heuristic rules, held within an E.S., to identify specific land units. In simple terms it is a land classification approach that uses new computer techniques to infer relationships between parameters and to justify these inferences in both a textual and graphical format to the end user. The approach is best explained by considering Figure 10.1.

FIGURE 10.1: Representation of Knowledge Components



The environmental parameters represent both the physical occurrence of the selected parameters and, if known, the quantitative knowledge. As a separate entity, prior to land classification, knowledge regarding the parameters is transparent, in the sense that it is documented and can be understood. Similarly, the quantifiable knowledge regarding the GIS is also transparent. Yet the knowledge held by an expert is subjective and not easily understood, especially as experts rely on deeper knowledge or intuition to make decisions regarding land classification. The *Salt Manager* system designed in this project encapsulates the knowledge of experts into a systematic set of heuristic rules which are, in themselves, easily understood and transparent as they can be explained by the reasoning abilities of the ES. This approach therefore bypasses the complicated numeric weighting methods used by the parametric/GIS land classification approach to quantify empirical knowledge. Instead, it substitutes production rules for the numeric weighting process. By doing this the land understanding approach exploits the mimicking capacity of an ES to replicate the initial decision making process made by humans (refer to Gibbons et al. (1968) section 4.0).

A feature of the land understanding approach relates to the reasoning ability of the E.S., i.e. the ability to explain to the user why a specific area of land was classified as a certain type. In this example, the results of the land understanding classification are presented to the user in a "user friendly" software package, utilising both an ES and a GIS, representing a substantial change in the diffusion of knowledge regarding land classifications. The *Salt Manager* system presents to the user the rationale for each decision regarding a specific land class on an interactive computer screen using software designed to be user friendly.

Importantly, (un)certainty regarding the land classification procedure is also conveyed to the end user. The traditional parametric/GIS land classification methods by attributing weighted numeric values to the environmental parameters, inherently convey a degree of certainty to the land classification process, but unfortunately this is not conveyed to the end user. Whereas the approach developed by this thesis aims to convey uncertainties regarding the decision making process and the quality of the data to the end user. It is a method which therefore portrays a more realistic and truthful assessment of the classification result to the end user.

From a systems point of view, the developed *Salt Manager* system is a software program which is external to the GIS database and only accesses the database when spatial information is required. Subsequently, environmental data layers may exist in an unprocessed format within the database, i.e. they have not been weighted or clumped into land system groups. Thus numerous ES programs or deterministic models are able to access the one GIS database.

To ameliorate secondary salinisation, appropriate high water usage management strategies require implementation at both recharge and discharge areas. As recharge areas may vary from metres to tens of kilometres from the relevant discharge areas, it is essential that land owners in the entire region are informed of the implications of their management on land owners in other areas; particularly land owners in the elevated areas who utilise ridges with shallow soils on fractured bedrock (high recharge areas).

To make all land owners in a region more accountable for their management practices an interactive land classification method, *Salt Manager*, has been developed. As an agricultural extension tool, it is a software package designed to be used in the regional agricultural field offices by the regions land owners. The prototype system solves three inter-related problems: (i) it identifies priority areas of potential groundwater recharge and discharge; (ii) it acts as an educational management tool as it interactively provides suitable management information to land owners; (iii) it provides the user an opportunity directly update the information within the database, i.e. alter depth to groundwater levels at surveyed points.

The land understanding approach is a pragmatic method aimed at developing a land classification, based on expert opinion, in a format accessible to any user. As a new approach, it highlights how attitudes have changed to the concept of land classification. The genetic and landscape land classification approaches were criticised for not being able to incorporate enough environmental parameters and for being too qualitative, i.e. in the sense that they were difficult to replicate. With the advent of the parametric approach and GIS, the storage and manipulation of empirical data became possible, overcoming the major problem with the other two land classification approaches. Thus, the focus shifted to the quantification of the environmental parameters in order to develop a replicable method to identify relationships between spatially occurring parameters.

By integrating an ES with a GIS the replication of the human decision making process regarding land classification, as documented by Gibbons et al., (1968), becomes more complete. The data layers recorded within the GIS, are systematically interrogated by the production rules held within an ES but the method of analysis now becomes transparent to the user allowing them to attain some understanding of the reasons decisions were implemented.

The next subsection gives consideration to the results produced during the knowledge elicitation process.

10.1.2 The Knowledge Elicitation Process

The two stage development of the rule base for the *Salt Manager* system provides a number of interesting results. To discuss these results each rule base recharge/discharge will be considered separately.

The first groundwater discharge rule base identified three discrete classes of high/low potential groundwater discharge and existing groundwater discharge and was mutually exclusive to groundwater recharge. To increase the efficiency of the rule base it was subjected to an experiment using machine learning techniques. The results led to a restructuring of the rule base to increase the ES efficiency and the discovery of new knowledge (refer section 7.2.5). After conveying the results of this experiment to the experts they decided to reformulate the ES rules. The knowledge elicitation process was therefore ongoing, as the experts became more familiar with the knowledge elicitation process.

Following from this, the second rule base developed to identify groundwater discharge proposed as its output a groundwater continuum, ranging from existing groundwater discharge to high and then low potential groundwater discharge. As "depth to groundwater" was considered to be the most important parameter to identify potential groundwater discharge, the model required this information to be provided by the users. At this stage it must be noted that the dependence on groundwater data for land classification of potential groundwater discharge is conceptually very different from criticisms applied to traditional identification methods of groundwater recharge using networks of observation bores. As discussed in section 2.2.3, groundwater discharge predominantly occurs in low lying valley areas, thus groundwater point data need only be obtained from valley floors. In these circumstances it is those landowners who have the highest probability of losing land to increasing salinity levels who are required to provide valuable groundwater information. This contrasts to collecting groundwater information for groundwater recharge assessment, as groundwater point data for all topographic areas within a region is required from some landowners who, depending on the locality of their land, may not be affected by increasing salinisation.

The facilities developed within the *Salt Manager* system for data entry update, support the "interactive" concept being developed for the entire land classification procedure. For instance, it is assumed that a user who enters data into the system is more likely to have confidence in the results being produced and are therefore more likely to initiate management information conveyed to them.

The rule base identifying groundwater recharge zones was also formulated in two different stages. The first did not account for the influence of EM31 data and also considered recharge as being mutually exclusive to discharge. After the first machine learning experiment, as previously mentioned, the experts now familiar with the knowledge elicitation process decided to create a separate groundwater recharge rule base. Coincidentally, EM31 data had been collected for the 1x3km catchment and included in the database, thus a new groundwater recharge rule base was formulated to account for this new information.

No direct comparison between the two different groundwater recharge rule bases was conducted. Instead they were independently compared to the results produced for the test points within 1x3km catchment. The second of the two, which included EM31 data, correctly identified 75% of the field study points (compared to 50% for the first rule base). This result supports existing literature on the usefulness of EM31 data as a tool for identifying groundwater recharge.

With the realisation that the results for the regional groundwater recharge analysis, without the EM data, outside of the catchment were poor and with the increase in EM31 data collection, it was decided to conduct a study on the effects of EM31 data on the classification task. Another machine learning experiment was conducted to ascertain this result. The result from the second machine learning experiment concluded that the EM31 data was a useful discriminator for two soil types (Belalie Loam and Belalie Clay Loam), which covered 52% of the catchment, but for the other soil types it was not a discriminating factor in groundwater recharge classification, as it was subsumed in the decision making process by the soil data. This result highlights that the EM31 sampling procedure should have confined itself to the two soil types of Belalie loam and Belalie Clay Loam. The results therefore suggest a methodology for ascertaining relationships between datasets in GIS databases with a classification task.

The next subsection will consider a technical discussion of issues deriving from this thesis.

10.2 TECHNICAL DISCUSSION

The aim of the thesis was to develop a method using the tools of ES, GIS and RDBMS to identify and monitor existing and potential groundwater recharge and discharge zones. In the process of developing this method, the generalisation and synthesis of data was likely to cause errors. Two types of error will be considered in the chapter, namely spatial location errors and errors relating to the estimation of uncertainty in both the GIS data layers and the heuristic decision making process.

10.2.1 Spatial Locational Errors

Spatial locational errors are those generated during: a) digitising the geo-based data including various types of thematic maps and depth to groundwater sample points and b) registering both the data layers and the remotely sensed imagery. Although great care was taken in the digitising process two types of error are likely to occur. The first are physiological errors. These errors occurred when the operator's hand became tired and experienced involuntary movement when digitising large polygon boundaries. The errors produced by this problem were manually edited. The second type of digitising errors, more difficult to edit, were the psychomotor errors. These errors develop when the boundaries representing polygons were not followed precisely by the operator when digitising. Another error factor, relating to the data conversion process, includes the raster/vector dichotomy. As noted in Chapter 6, the digitising process creates data layers in a vector (point, line, polygon) format. In this thesis, for analysis to commence, the spatial data was first converted to raster format using the centroid algorithm within the Arc/Info GIS. This conversion from vector to raster results in errors, e.g. where a line of a polygon dissects a grid cell, leaving one side of the dissected cell as a potential error value. This error factor was not accurately quantified.

The geometric accuracy of the converted data layers tended to be very good as the ground control points were all verified utilising GPS technology. RMS errors for the environmental data coverages: soil; geology; slope; elevation; EM31 and depth to groundwater were all recorded as being less than 0.03. Geometric registration of the remotely sensed data, on the other hand, was difficult and a more likely source of error. Although the statistical estimation of geometric correction of the images was controlled to within 15 metres (i.e. half cell size), cell by cell analysis of a grid containing $30m^2$ pixels ensures that a degree of error is incorporated into the model. This problem relates to the use of raster data structures to represent real world phenomena. The reduction of the landscape into elementary grid cell units results in data analysis being confined to the actual gridcell unit and not to the geographical object. The raster structure therefore enforces a certain type of reality upon us rather than allowing the actual reality to suggest a more appropriate structure for analysis (Kemp, 1993). It is hoped further research into the representation of continuous data will provide solutions for this problem.

Analysis conducted at the gridcell scale is a potential source of error in the model produced by this thesis. The different data types analysed at the gridcell level derive from different scales and therefore have different levels of accuracy and precision, for instance: EM data was accurately collected for 894 points within the 1x3km catchment, whereas Landsat TM data $30m^2$ pixels are known to be displaced by approximately 15m. Without attributing error estimates to each of the data parameters within each gridcell the accuracy of the result cannot be truly attained. Additionally, the cell by cell analysis does not take into consideration neighbourhood relationships, thereby precluding the identification of causal spatial relationships. To compensate for the static nature of the *Salt Manager* model the update facilities provided to add, change and delete the environmental data were developed.

10.2.2 Modelling Uncertainty

This subsection concerns conveying (un)certainty regarding both the accuracy and precision of the GIS data and the decision making process to the end user. In this thesis, as similar to Hunter and Goodchild (1993), the term (un)certainty will be used to denote the fact that so little formal knowledge exists about spatial database error (in the form of models describing its manifestation and subsequent propagation through spatial processing). Thus, until a more appropriate term is developed it will be used in the interim period until this situation is clarified (Hunter and Goodchild, 1993, p. 522).

10.2.2.1 Spatial Data (un)Certainty

There are two main methods by which to convey uncertainty in digital data to the user. The first is by statistical methods. Three commonly used techniques are: the calculation of the mean (an estimate of the truth) and standard deviation (a measure of the uncertainty) of a set of observations; the misclassification matrix used to assess the likelihood of remotely sensed pixels being correctly classified; and the variogram derived from kriging (Hunter and Goodchild, 1993). This thesis has utilised the variogram generated by the kriging process to quantify the zones of uncertainty associated with the depth to groundwater GIS coverage.

Hunter and Goodchild (1993) and Goodchild (1991) state that there are five weaknesses associated with statistical methods. Three are considered here in order to justify the approach taken to represent uncertainty in this thesis. The first criticism relates to the actual error statistic, while useful, it conveys no information regarding the factors contributing to its formulation. Secondly, the error statistic, although clear in meaning to a domain expert, may not be intuitively understood by any individual.

The third is that the uncertainty varies spatially across the map and that its physical distribution may be of more importance than the actual average value for the map. For example, "the calculation of slope aspect is known to be especially sensitive to elevation changes in flatter terrain and thus needs to be shown as it varies throughout the data set" (Hunter and Goodchild, 1993, p. 524).

To compensate for the first two criticisms visual techniques, the other method of representing uncertainty, were deployed. Visualisation offers the ability to organise abstract concepts into meaningful displays, transform numerical displays into visual images and permit manipulation of colour, motion and geometry (DeFranti et al., 1989). Visualisation of data is in fact a key advantage of GIS systems, as it offers users an opportunity to view the spatial distribution of a selected parameter. In this thesis the variogram generated by the kriging algorithm has been processed into a coverage representing spatial variance within the study region for the environmental parameter, depth to groundwater (refer Figure 6.3). To ascertain the effect of this spatial variance on the final *Salt Manager* thematic map classification, the decision making process involving all data parameters must be considered and the results presented to the user in an understandable format.

10.2.2.2 (un)Certainty Associated with the Decision Making Process

This thesis has developed a qualitative, indicative method to combine uncertainties associated with the decision making process and the GIS data layers in order to present visual information to the user. The method is described in section 7.4. There are a number of problems associated with the approach that has been developed. Firstly, the method used to allocate Certainty Factors (CF) to the variance generated by the kriging algorithm was purely subjective. Secondly, as considered in section 7.4 CFs belong to the so called quasi-probabilistic models for uncertainty reasoning which are not mathematically well-founded. Compounding this weakness is the actual allocation of CFs to ES production rules. The validity of this process is difficult to establish, although similar methodologies have succeeded in satisfying a limited form of the Turing test (refer MYCIN, Buchanan and Shortliffe, 1984).

Nevertheless, the method is useful as it enables quantitative uncertainty values generated by kriging to be analysed in relation to other data parameters used in the decision making process, thereby satisfying the third criticism of statistical methods by Hunter and Goodchild (1993). Moreover, this method puts the responsibility of accepting the accuracy of the data and the decision making process on the user, thus forcing them into the position where they either accept or reject the result. This process is deemed to be positive by this thesis as it provides a mechanism through which end users may attain at least some understanding of the rationale behind the decision making process and make a judgement on whether they agree or disagree with that process.

10.3 FURTHER RESEARCH

A report was commissioned by the Australian Land and Water Development Research Development Council and conducted by PDP (1992) to identify key areas for funding in dryland salinisation research. Of the six areas finally nominated, DSS and ES systems were one. It is contended that these tools should not be applied to the problem in isolation but rather, as the results indicate in this thesis, that they should be combined with GIS systems to enable successful identification of areas as well as provide a mechanism by which to inform landowners of problems occurring in their region. In light of this, the Spatial Information Group within the South Australian Department of Primary Industries have applied under the above designated funding area for a substantial sum of research money to continue the work developed in this thesis. The proposal is that the techniques that have been developed should first be applied to another region within South Australia. The initial goal will be to first verify the accuracy of the model in the second area and to then develop a method by which the system can be applied to the agricultural regions within the whole state. To achieve this second goal the concept of meta-knowledge and a renewed approach to knowledge elicitation must be addressed. With meta-knowledge implying a knowledge of descriptions of dominant background themes; forms representation and location; simple relational aspect; and statements of lineage and significance.

Another research area of importance is the assessment of uncertainty in both ES and GIS data models. A qualitative attempt has been implemented within this thesis and, although it conveys an understanding of the uncertainties associated with the land classification results, more accurate quantitative methods require implementation. A library of uncertainty techniques is being developed to address this problem, i.e. fuzzy sets (Zadeh,1965), possibility theory (Dubois and Prade, 1988) and the mathematical theory of evidence (Shafer, 1976). In respect to this point a fuzzy set theory approach is currently being constructed by the author and Dr. Peter Eklund from the Department of Computer Science at the University of Adelaide. Membership functions for the parameters have been developed for each of the parameters but the project is yet to be implemented.

The third area where further research is required is the development of generic ES/GIS systems. Currently, as mentioned in section 4.3, there are a number of generic systems under development which combine different software systems, e.g. combine the functionality of different ES and GIS system. In contrast, commercial GIS systems are now including in their analytical toolkits decision support tools, e.g. IDRISI version 4.1 utilises Bayesian and Fuzzy Set tools to provide decision support. Conceptually, these two approaches are very different as the first seeks to synthesize information from different software systems, whilst the second wants to provide maximum functionality in the one system. The approach taken by this thesis has been to implement the first method of integrating systems. With current changes planned for the storage and management of the aspatial data in the Arc/Info GIS, e.g. ArcSTORM, it is expected that the method of integration implemented by this thesis will have to be changed. Obviously, with ongoing software development being conducted by the separate software packages, methods of integration will also require constant upgrading.

10.4 CONCLUSION

Kemp (1993) writes that a shift is taking place in the environmental sciences, "where the models are moving gradually away from direct description of physical laws to complex models with uncertain spatial parameters, effects, inputs and ill-defined relationships making full system description impossible (Kemp, 1993, p. 4). The methodology adopted by this thesis directly complements Kemp's (1993) observations.

The motivation for the approach developed for this thesis stemmed from the lack of successful methods provided to identify causal factors associated with dryland salinisation. It was contended via the initial objectives that the use of a heuristic approach to analyse environmental information would be an appropriate approach. The results considered in this thesis give support to the notion that the initial objectives proposed in section 1.1 have been achieved.

To conclude four substantive results have been reported by this thesis:

1) A new interactive land classification procedure has been developed to allow the end user to be better informed of the decision making process. This is achieved by providing access to the environmental and heuristic information used in the land classification process. This notion has been previously ignored by previous land classification techniques.
2) The *Salt Manager* system has been developed to implement the new land classification methodology. It integrates GIS/ES/RDBMS technology into a system that has interrogative and data update facilities. Moreover it is a system which both conveys to, and captures information from, the end user.
3) The application of machine learning techniques to the land classification results and the GIS database is unique in this field. [illegible] Two major results have been derived: identifying new knowledge and ascertaining redundant information in the decision making process.
4) The *Salt Manager* system has successfully identified and provided relevant management strategies, for areas of groundwater recharge and discharge within the 1x3km catchment in the Jamestown region of South Australia.

Appendix 1

SALT MANAGER USER MANUAL

SALT MANAGER USER MANUAL

Version 1.0

S.D. Kirkby

Department of Geography
University of Adelaide

skirkby@geog.adelaide.edu.au

INTRODUCTION

Salt Manager is a system which interfaces the Harlequin Expert System (ES) environment with the Arc/Info GIS and the Oracle RDBMS. This manual is intended to outline the details for using the software. Further discussions of the design details may be found in the thesis, Managing Dryland Salinisation with an Integrated ES/GIS, written by the author and located in the library at the Department of Geography, University of Adelaide.

The manual is divided into two sections. The first provides a general description of the three software packages utilised by the *Salt Manager* system. The second gives consideration to the functionality of the *Salt Manager* system.

COMPUTER MODELLING ENVIRONMENT

Three commercial software systems are utilised by this project to develop a seamlessly integrated software system which will allow flexible manipulation of the data. This user manual will provide a brief overview of the structure and theoretical basis of each system. This section is essential to understanding the development of the integrated ES/GIS/RDBMS model.

THE ARC/INFO GIS

Arc/Info, is a commercially available, vector/raster based GIS. It is developed and marketed by the Environmental Systems Research Institute (ESRI) of Redlands, California USA. Without doubt, Arc/Info is the most popular GIS in use. The user manuals consist of 12 volumes and the object library in three additional volumes. The manuals are extensive, thus no attempt will be made to replicate the information. Refer to Morehouse (1989) for an abbreviated discussion on Arc/Info data structures.

THE HARLEQUIN EXPERT SYSTEM SHELL

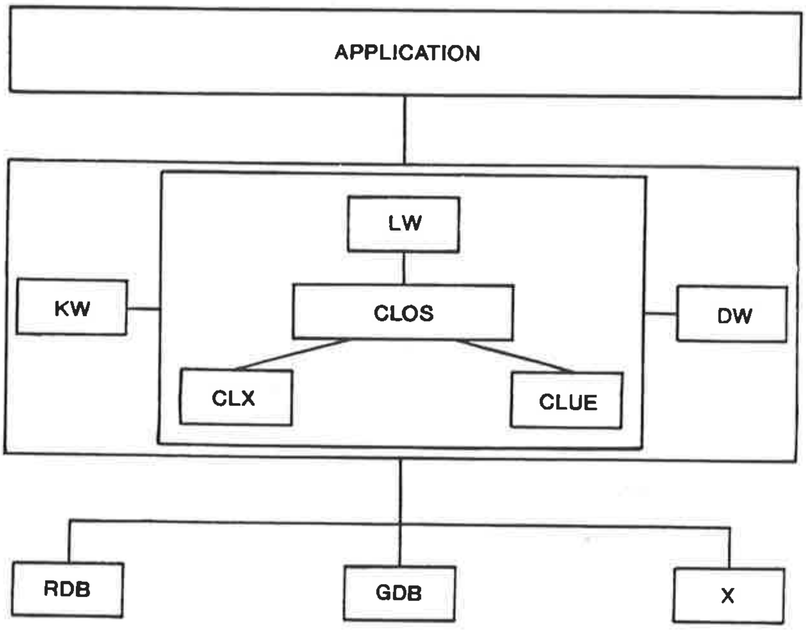
As Harlequin is a relatively new player on the software circuit, information regarding the environment is generally not well known. A detailed explanation of the software is provided in this section.

The Expert system shell used by this project was based on the Common Lisp Object System (CLOS). It is accessed through LispWorks, a module that provides a uniform syntax to all the parts of the system. Accessed via LispWorks is Knowledgeworks, an in-built inference engine that can be used to establish rule based modules within any application (Figure 1).

The data storage provided by Oracle, uses relational database technology and provides a transparent interface for the objects that can be used in the system. Structured Query Language (SQL) queries can be performed if required to select data from the databases.

Notionally supporting the language components are the relational and graphic databases. The CLUE package (Kimbrough and LaMott, 1989) provides the user interface facilities (known as contacts), and CLX (Common Lisp X Windows Library) provides access to X Windows protocols. The application therefore is built upon this hierarchy of components, calling upon facilities as required.

FIGURE 1: The Harlequin Expert System Shell Environment



THE COMMON LISP OBJECT SYSTEM

The Common Lisp Object System (CLOS) is an object orientated programming language embedded in Common Lisp, a language in which all data items are in the form of objects have identity. CLOS aims to be powerful enough to support large, production-quality application programs in both the development and delivery phases - a goal requiring uncompromised efficiency" (Moon, 1987).

Within CLOS there is a distinction between the structure of an entity, and its behaviour. Concepts like multiple inheritance and generic functions are fundamental, as is the notion of extensibility. It is layered in such a way as to provide the programmer with the ability to either use it as an everyday programming language, or access the internal structures and implement new features.

An object has a structure and associated passive properties, and active properties that distinguish its behaviour. Every object has an identity that distinguishes it from others; the Lisp function *eql* determines whether two references to an object are in fact to the same object. Objects have slots which contain other objects; the LISP form *(SLOT-VALUE object symbol)* returns the contents of slot *symbol.*

Classes provide a classification mechanism for objects and every object is an instance of a class. Instances of the same class have the same structure having the same number of slots and the same slot names. The behaviour of the objects refer to the set of operations they can perform. Generic functions perform operations on different types of objects, and have formalised the concept of a method. A generic function can have any number of methods defined for the different possible objects; methods specialise a generic function.

The concept of a class is fundamental to the object-orientated programming paradigm. Classes inherit all the characteristics of their superclasses whilst still allowing extra individual characteristics to be specified. David Moon's succinct description that "a subclass can be regarded as a specialised or customised version of a superclass" (Moon, 1989, pp. 60), highlights the building block approach to the object orientated methodology.

The obvious problem of class conflicts in CLOS are resolved with a declared class precedence list which includes each superclass in the order in which these conflicts are to be resolved. Direct superclasses are resolved using specificity; a subclass is more specific than a superclass. Each class defines some slots and inherits all the slot definitions of superclasses. Similarly, method inheritance allows for situations where a fundamental behaviour can be defined and modified if required.

CLOS also provides mechanisms for combining methods to form one implementation of a generic function; primary methods perform the main function whilst before and after methods initialise, and tidy up after it. The side effects of these latter *daemon* methods are utilised; no results are returned by these. The final type of method is that of around methods; "the term is based on the idea that an around method is wrapped around the other methods, controlling whether and when they are called" (Moon, 1989, pp. 69). Around methods typically pass control via the *CALL-NEXT-METHOD* function.

THE COMMON LISP USER INTERFACE ENVIRONMENT

CLUE is a high level programming interface to the X window system based on the object orientated programming system CLOS, and extending the Common Lisp X Window interface, CLX. Fundamental to the development of interactive applications in this environment, is the fact that the application's control structure is that of an event loop; that is, the application is event driven.

The event loop consists of waiting for an event, typically mouse movements and button presses, working out what to do with the event, processing that event and then going back to wait for the next one. These events, then constitute the user's interaction with the application and make up what is commonly referred to as the user interface.

The user interface which is generally composed of a number of user interface objects like menus, messages, dialogue boxes etc., becomes the front end of the application. Each of these user interface objects is an agent for some portion of the underlying application.

These objects lend themselves to the object-orientated paradigm through the fact that many of the objects have common core functionality, whilst still exhibiting some variance of behaviour. This is well matched by the methodology that allows specific characteristics to be inherited and built upon.

The user interface objects are referred to as contacts. These are specific kinds of windows which inherit basic window functionality but add additional features. For example, contacts can have names, parents, a state, event translation and callbacks. Callbacks provide the connection between the application and the user interface. An individual callback will consist of a callback name and an associated callback function which is invoked with a predefined argument list.

```
(let*
         ((tl (make-contact 'top-level
                            :parent *display*
                            :title "Data Entry
Sheet"))
          (col (make-contact 'constraint-column
                            :parent tl
                            :justification :centre))
          (Soil (make-contact 'choice-item
                            :selected (initialize
soil)
                            :row-margin 5
                            :column-margin 5
                            :columns 3
                            :parent col
                            :name 'soil
                            :item-data-fn 'identity
                            :items '(("sk" . 1)
("sk/cl" . 2) ("ysil" . 3) ("yrl" . 4) ("cl/ysil" .
5) ("cl/yrl" . 6) ("cl" . 7) ("besicl/ysil" . 8)
("alluvial" . 9) ("bsl" . 10) ("bcl" . 11) ("bl/bcl"
. 12) ("bl" . 13))

                           :title "Soil Type"))

  (add-callback Soil :notify 'store-data 'Soil)
   (add-event  Soil  '(:button-press  :button-2)  '(Help
Soil))
```

In the example above, a top level window is created to house a menu contact of class check-list, titled "*Data entry sheet*". The items of the menu are given as "two item" lists; the first item being the string displayed in the menu and the second the associated data item. The add-callbacks command links a menu selection with the function *store-data*.

A contact which is a parent of another contact is known as a composite and is an instance of the composite subclass of contact objects. A parent then, maybe a parent of another composite, thus leading to a hierarchy of contacts. Parent contacts are subsequently responsible for the management of the geometry for the childs contacts.

So too are they responsible for the handling of events that may be of significance to the application. In the previous example, a press of a :button-2 results in the invocation of the help function with its associated argument, specified as a list.

Resource management refers to the specification of certain characteristics of contacts that have been defined. The system maintains a resource database which the user can modify.

If both the background colour and the font type were to be changed for a particular contact, the define-resources macro could be used in the following way to add a binding reflecting a users preference to the database:

```
(Define-resources
  (Salt-Manager * button background) white
   (Salt-Manager  *  pattern-choice  *  font)  helvetica-
12)
```

The last line in the above example is interpreted as specifying the font of everything (the wildcard character *) in the pattern-choice contact, irrespective of which composite it is found in the Salt Manager program, as helvetica-12. Resources are specified in a contact definition with the *:resources* option.

THE LISPWORKS TOOLKIT

A toolkit is provided in the system to simplify the creation of standard contacts like menus, button panels, display panes, popups, prompters, dialogues and graphers. Class definitions for the above are accessible via the toolkit package. A range of generic functions allow the contacts to be modified via the *set-contents* function, interrogated via the *contact-value* function, or simply return useful information for screen management purposes e.g. *contact-best-height*.

Constraint-row and constraint-column classes allow the contacts to be grouped together in specified configurations. Resources like *fixed-sizes* and *size-ratios* can specify and control relative size ratios for the children within the main top level window.

Contact classes for both horizontal and vertical scrolling are provided, along with the generic protocol for attaching and controlling the windows with the bars. The result is the provision of the scrolling user-interface giving visual feedback as to the size and position of the visible portion of the window, and the ability to move to a particular position or backward or forward by one window.

Popups are contacts that appear temporarily on the screen to either report some information, or prompt for some input. In the latter case, they disappear when the information is entered, the former requires some form of *done* button. Windows disappear when they are no longer managed by a higher level window:

```
(setf    (contact-state    (contact-parent    menu))
:withdrawn))
```

This line of code unmaps the window that is the parent of the menu contact.

There are a number of dialogue boxes which are based on the simple-popup behaviour that are provided. The following functions provide commonly used facilities:

confirm-yes-or-no message
send-a-message message
popup-prompt-for-string title
pop-up-prompt-for-form title

Where existing facilities need to be extended, new user-defined popup classes can inherit the popup functionality and extend the feature provided. It is significant in that it demonstrates the binding of *contact creation code* to a slot value.

The graph-pane contact can display graphs using a number of node and edge styles along with a number of generic graph operations such as hiding nodes and collapsing parts of the tree. Button selections of the tree nodes can be captured via the ability to define callbacks, and callback functions for the graph.

Graphics ports implement a set of drawing functions that are interned in the graphics-port package. Each graphics port instance has a gaphics-state object associated with it that defines characteristics of the port. The *with-graphics-state* macro allows the constituents of the object to be accessed and altered. Commonly used graphics functions include:

draw-image port x y image
draw-string port string x y
get-image port x y width height
put-X-image port x y

Further functions provide for the copying, caching, rotation and transformation of graphics states.

SYSTEM DEVELOPMENT SUPPORT

Built in tools for examining, modifying and controlling the current status of the system constitute the system development support. Tools such as browsers, editors, graphers and listeners to name a few, are provided to ensure access to the many classes of objects, both in-built within the toolkit, or defined by the user.

SYSTEM TOOLS

LispWorks provides system tools for managing large applications that may constitute a number of different files. A system is basically a collection of files that together constitute a program, plus rules expressing the interdependencies between these files.

The defsystem macro is used to define a system; the syntax of this facility is:

```
Defsystem system-name options and key members rules
```

Various options may be specified which affect the behaviour of the system; for example, the *:package* and *:default-pathname* provide default values when they are not specified.

Each member of a system corresponds to either a physical file or another system; keyword arguments can specify behaviour. For example, in the following definition for the Salt-Manager system,

```
(defsystem salt
  (:default-pathname  *load-pathname* )
  :members (
            ("defpkg" :source-only t)
             "functions"
             "objects"
             "popmess1"
             "contacts"
             "uncertainty"
             "salt-manager"
            ("salt-rules" :type :kb-system)
   )

  :rules ((:in-order-to :compile :all
            (:requires (:load :serial))
          ))
)
```

the *:source-only* keyword ensures that only the source file is ever loaded for the package definition. Note also that members of a system can themselves be another system. The *:rules* keyword ensures that the dependency for the system are maintained.

The System Browser

As a result of a system definition, the system can be treated as an entity; for example the "*compile* and *load*" option provided by the system browser, could be selected to activate the system compilation. Further still, the member menu allows each member file to be treated individually for editing, or compilation purposes. Message areas ensure that the current state of the system action can be observed; examining the background display allows the developer to view a background trace of the compilation.

Application specific classes and objects can have a visual representation defined for them by the visual object assembler. Their subsequent instantiation and management is then controlled by the system. This includes their on screen representation, the sensing of events like button presses and mouse movements etc., and the corresponding actions (callbacks) to be performed in response.

The Harlequin environment with its associated tools, provides an application development strategy that enable the programmer to specify a complex system composed of many components, and their interrelationships, in a modular manner. Any rules that govern module interdependencies between individual members of the system, can be specified.

KNOWLEDGEWORKS

Knowledgeworks is a LispWorks toolkit for building knowledge based systems in a multi-paradigm programming environment which allows developers to express problems in terms of objects, rules and procedures. It includes OPS5 (refer to Brownston et al., 1986) forward chaining compatibility, along with Prolog compatible backward chaining. CLOS provides the object base for the Knowledgeworks system whilst procedural features are available through Common Lisp foundation. Knowledgeworks objects are standard CLOS object with a *mixin class* providing the Knowledgeworks functionality.

Knowledgeworks rules perform pattern matching over the object base; both forward and backward chaining rules utilise this pattern matching to perform actions in the case of forward chaining rules, or deduce goals in the case of backward chaining.

The Object Base

A Lispworks CLOS relational database class can also be given the Knowledgeworks mixin class, enabling rules to access this data as if there was no database present; that is, transparently. The *def-view-class* macro can be used to specify an object definition where the slots are filled from an existing SQL table.

Forward Chaining

Knowledgworks forward chaining control mechanisms are different from those of other languages in that it uses data sensitive, unordered rules rather than sequenced instructions as the basic unit of computation. Production system architectures typically contain 3 sections; working memory which holds the instances of objects, the rule base that constitutes the program and the inference engine which executes the rules.

The inference engine cycles through its match, select and execute sequence, firing rules from the rule base, changing the contents of working memory. A different set of rules may fire on the next cycle. "In a rule-based system,control is passed on frequent re-evaluation of the data states, not on any static control structure of the program. Thus we say computation in a production-system model data-driven rather than instuction-driven" (Brownston et al., 1986, p. 7).

The process of selecting and firing a rule from a set of possible rules, referred to as an agenda, is repeated until either the forward chaining engine runs out of rules, or a rule instructs it to stop. Each rule can have a priority number associated with it to help in the process of conflict resolution. They are normally grouped in what are referred to as *contexts.*

Contexts, first used in MYCIN, form a partitioning mechanism that allow rules to be modularised. As well as this partitioning, they allow instances of objects in the working memory to be subject to rules that have hierarchical relationships between them.

The forward chaining rule interpreter can be invoked with the Lisp function:

```
(infer [:context '(context list)])
```

Contexts are defined in the knowledge base using the defcontext macro:

```
(defcontext assessment
          :strategy (priority)
          :auto-return t)
```

Every context specifies its own conflict resolution strategy; they include:

priority - highest priorities are preferred first.
recency - the most recently created instantiations are preferred,
order - earliest loaded rules are preferred,
specificity - most specific rules are preferred.

If auto-return is set to true within the context definition, control is passed onto the next context in the agenda list when there are no more rules that can fire. The forward chainer will then proceed down the agenda list, context by context until there are no more rules that can fire; this terminates the execution of the rule interpreter.

Backward Chaining

Backward chaining involves trying to prove a given goal by using rules to generate sub-goals and recursively trying to satisfy those (Harlequin handbook, 4-7). The backward chainer matches directly over Knowledgeworks objects in a similar manner to forward chaining except that goals are satisfied. The action part of a forward chaining rule is syntactically and semantically identical to the antecedent part of the backward chaining rule.

The backward chaining interpreter can be invoked with the following Lisp functions:

```
(any <exp-to-instantiate> <expression-to-prove>)
(findall <exp-to-instantiate><expression-to-prove>)
```

KNOWLEDGEWORKS TOOLS

Knowledgeworks has a number of built-in tools to allow the programmer to examine the system. Firstly, a class browser is available to examine the definition of a class in terms of the slot definitions. The superclasses from where some functionality is inherited, can be examined, along with subclasses.

Both the generic functions and methods defined on a class either directly through inheritance, can be examined. Objects of as class in the working memory can also be examined via the object browser.

A rule browser allows the user to examine the rules, both backward and forward, grouped by contexts. The *forward chaining history* window displays the rules that have fired in each successive cycle along with those that were in the conflict set and could have fired. Spy windows are available to follow the actions of rules as they fire, whilst monitor windows allow the preconditions of forward chaining rules to be monitored.

THE PROGRAMMING SUMMARY

The programming in this environment involves each of the following areas:

1) working with forward and backward chaining rules from a rule-base;
2) employing object-oriented methodologies via generic function methods and CLOS objects;
3) storing and extracting data from relational tables with SQL; and
4) using the procedural language interface of COMMON LISP;

Objects are used to model the real world situations where object slots can store both data values and procedural information. Rules from the knowledge domain can then fire over the objects producing results that can manifest themselves in some relevant action e.g. the changing of a particular value in an object.

Prototypes can be developed and progressively refined in a modular manner. The graphical user interface can be established with hooks that the application can connect into. Callbacks can be used to link the front-end of an application with the applications code.

THE RELATIONAL DATABASE MANAGEMENT SYSTEM (RDBMS)

The relational database management system (RDBMS) utilised by this project is Oracle, a popular RDBMS internationally recognised and utilised within Australia. Oracle is developed and supported by the Oracle Corporation, Belmont, California (Oracle,1990). As Oracle provides 15 manuals to operate version 6.0 no detailed information will be replicated here, rather a general description will be provided.

Oracle RDBMS is designed for a multi-user environment and runs on several platforms ranging from IBM mainframes and DEC and SUN workstations to Personal computers. There is considerable portability, as all versions of Oracle are identical and include the full implementation of SQL. In addition the networking facilities permits sharing of databases.

The Oracle environment consists of the following components: a relational database management system, an active data dictionary, SQL query language, application generator and report writer. The application development tools for Oracle are; SQL*PLUS, SQL*FORMS, SQL*REPORT, SQL*MENU, SQL*GRAPH, EASY *SQL, SQL*QMX, DICTIONARY, SQL*STAR, ORACLE/NET, ORACLE/LINK. The tool utilised frequently by this project is SQL*PLUS.

SQL*PLUS

SQL*PLUS provides a direct interface to Oracle and contains a full implementation of SQL (structured query language). The SQL language is comprised of; a data definition language (DDL) , used to define the tables and make up the database; a data manipulation language (DML), concerned with retrieval and update of the database; and a data control language (DCL), used to specify who can access data in the database and what operations they can perform (McFadyen and Kanabar, 1991).

Within Oracle SQL is initiated to create tables, store and retrieve information from tables, modify information in the tables, manage the database, assist with data sharing and to provide security. Easy*SQL provides an alternative, it consists of menus and panels, but it has reduced functionality, i.e. only three tables can be joined and you cannot grant access to any user.

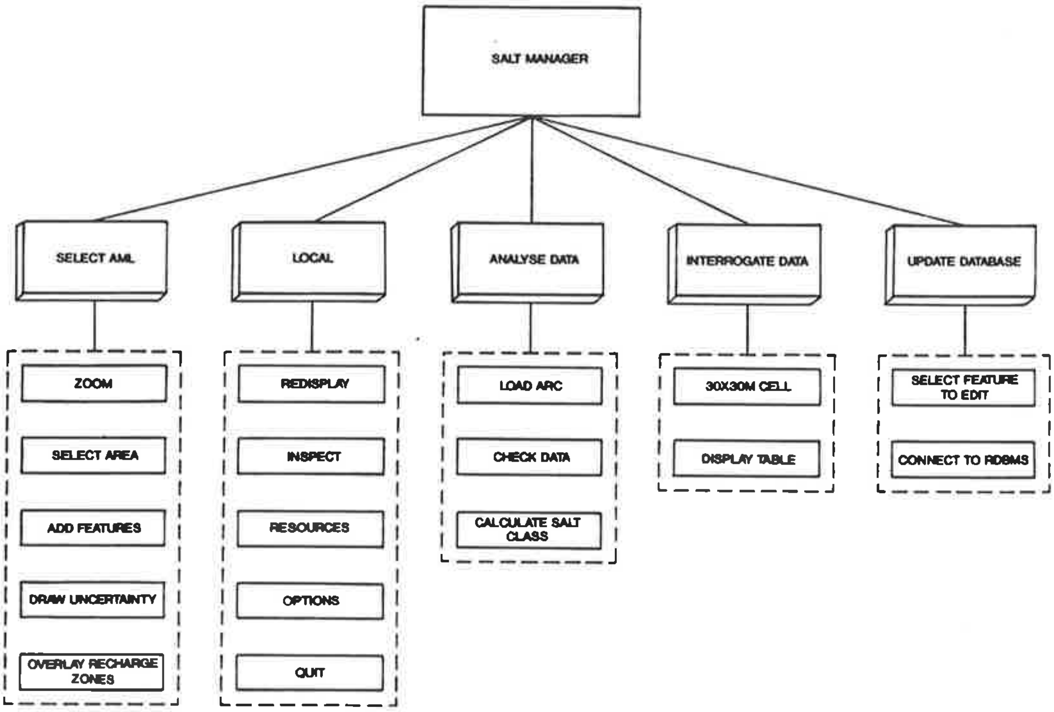
SUMMARY

The approach adopted by this system is the notion that "off the shelf" commercially available software should be integrated into a seamless system where the benefits of each package may be utilised transparently. At no stage does the *Salt Manager* aim to develop software which will replicate any functionality of the above mentioned software programs.

USING THE *SALT MANGER* SYSTEM

Operation of the *Salt Manager* system is via menu bars. A stylised diagram indicating the functionality of the system is displayed in Figure 2. There are five basic functions provided by the system. These are *Analyse Data, Interrogate Data, Update Database, Local* and *Select AML,* with each initiating, when selected by a user, a drop down menu with further selection choices.

FIGURE 2: ***Salt Manager* Functions**



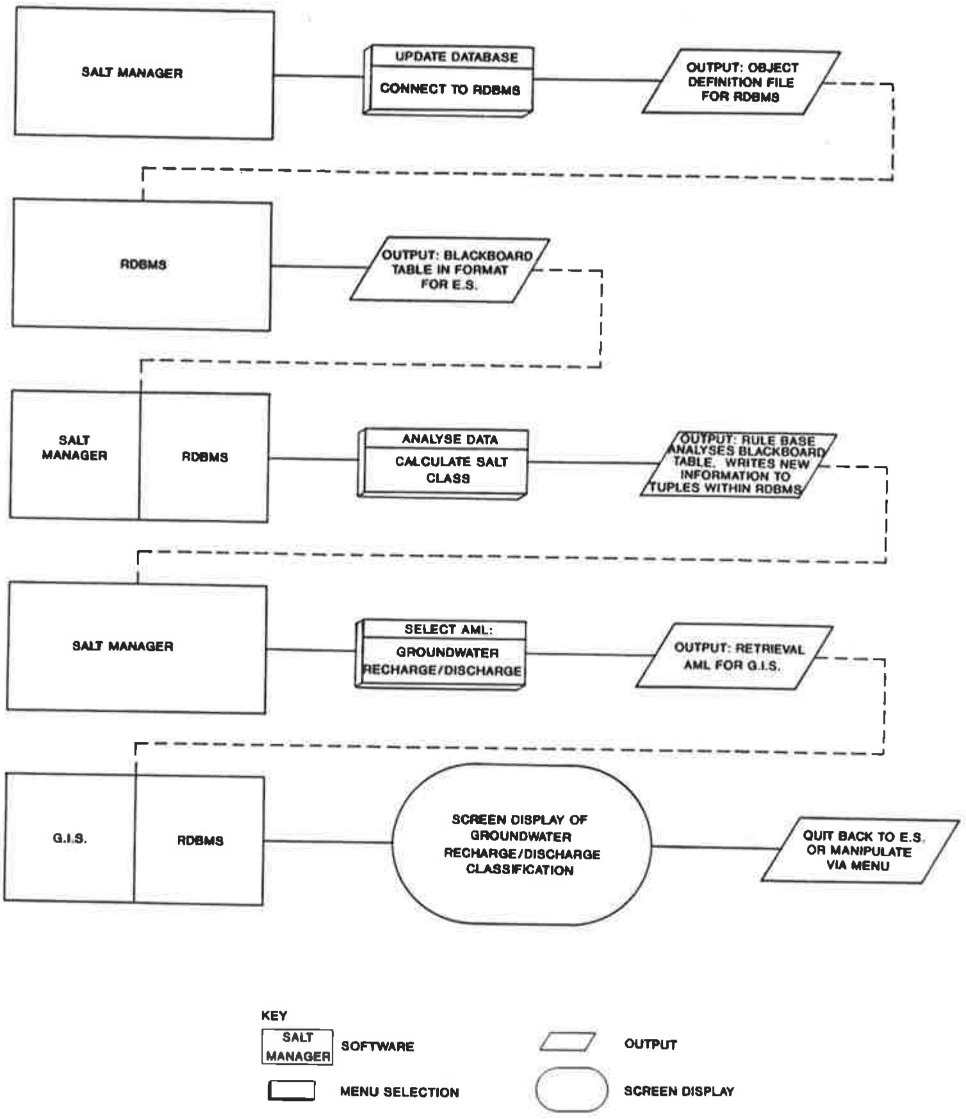
There are three main tasks which the *Salt Manager* system performs, they are: (i) classification of the environmental data (*Analyse Data*); (ii) an interrogation procedure to ascertain the decision making process (*Interrogate Data*); and (iii) updating of the RDBMS (*Update Database*). The fourth menu selection *Local* is the default option provided for each window generated within the Harlequin ES environment. The *Local* function provides the user with the opportunity to ascertain actions being initiated within the Harlequin ES. As this is the default option no further information will be provided in this thesis. The role of *Select AML* is central to the interrogation and updating components of the system and will be considered in relation to these functions.

ANALYSE DATA

The *Salt Manager* system contains two separate rule bases which analyse environmental information to produce two types of land classification maps; an existing and potential groundwater discharge map; and a potential groundwater recharge map.

To initiate the land classification process the user first connects the *Salt Manager* system to the environmental data held within the database. This task is achieved by the user selecting, *Update database*, then *Connect to RDBMS*, from the menu bar. A flow diagram summarising the methodology is displayed in Figure 3.

FIGURE 3: The Classification Process



To activate the inference engine the command *Analyse Data*, is then selected followed by *calculate salt-class*. Once the rule base has classified the environmental data the result can then be displayed. To initiate this process the user simply selects the relevant map to view, i.e. recharge/discharge, from the AML list provided within the *Salt Manager* and then initiates the GIS display function by selecting *Load Arc*, from the *Analyse data* sub-menu. The land classification map is then displayed. Using the sub-menus designed for each of the AML's the user can manipulate and add other information to the displayed image. Options provided by this menu are listed in Figure 2.

The land classification, identifying areas of either potential groundwater recharge and discharge or both, can then be interrogated by the user.

INTERROGATE DATA

The user is offered the option of interrogating the decision making process which results in the land classification map produced by the *Salt Manager*. In this example, the groundwater discharge classification, will be interrogated. A flow diagram, Figure 4, has been developed to illustrate the process.

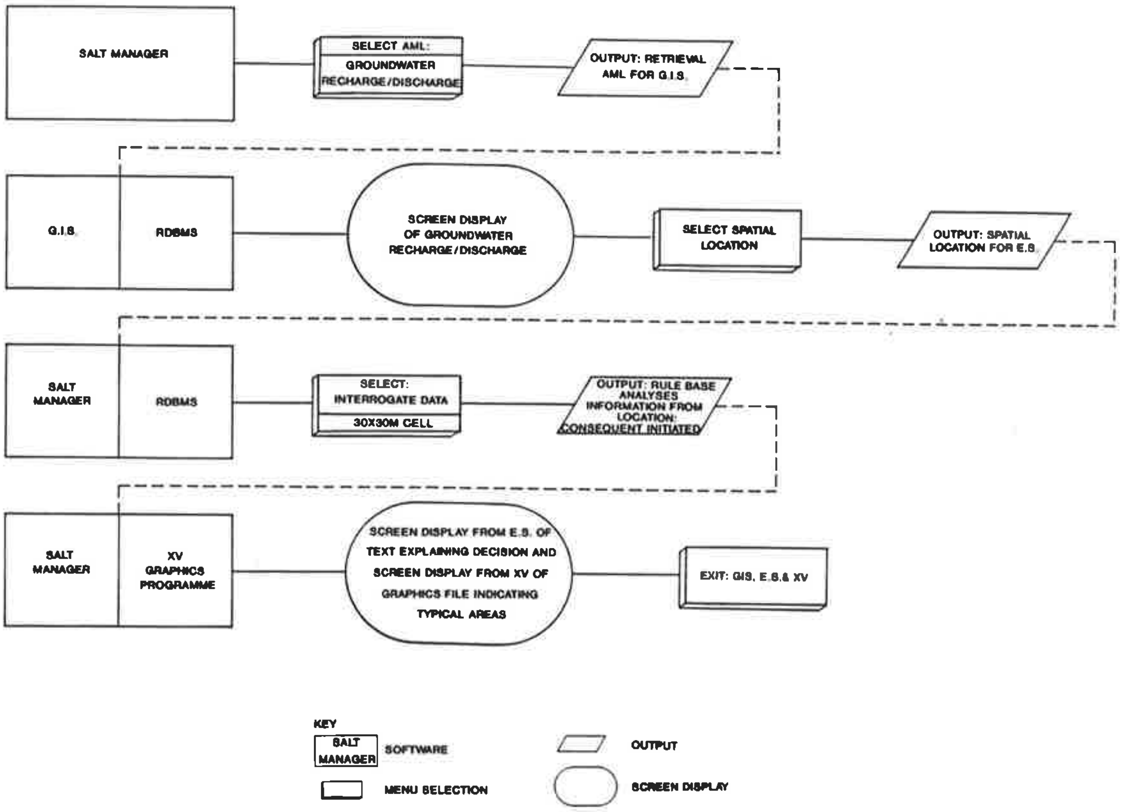
To commence this process the user simply selects the discharge/recharge classification to be displayed from the *Salt Manager*. This process is documented in Figure 4. Once the map is displayed the user may interrogate the map at the gridcell level ($30m^2$), by first selecting from the AML menu displayed by the GIS, the *select area* option. Subsequently, the user then selects the area they wish to interrogate by simply pressing the mouse button. Once selected, the user then presses *interrogate data/30x30m cell* from the *Salt Manager* menu bar.

Automatically, the rule base fires over the environmental information specified at that location.[1]

---

[1] Again, at this stage the type of rule base, either recharge or discharge must be first entered manually. Automation of this process is not difficult and will be initiated in the very near future.

FIGURE 4: Interrogating the Land Classification



Once the antecedent component of the rule has been satisfied the consequent is then initiated. In this case, the different rule base initiates a contact file displaying text information that explains the decision making process and importantly provides a management strategy appropriate for the environmental conditions.[2] Concurrently, it also initiates a graphics file showing an area indicative of the classification.

The Certainty Factor (CF) for the classification at the specific gridcell location is also displayed on the X-window title bar. CFs, are only developed for the groundwater discharge classification. The results of the uncertainty analysis can be graphically presented to the user. The menu options provided by the AML allow the user to display the degree(s) of (un)certainty associated with the classification.

---

[2] It must be noted that there are four rule bases; two each for the groundwater recharge and discharge classifications. There is no difference between the antecedent of each of the two respective rule bases, where they differ is the consequents that they initiate. Obviously the rule bases utilised to interrogate the land classification initiate explanatory acts as their consequents, e.g. initiate graphics and explanatory text.

UPDATE DATABASE

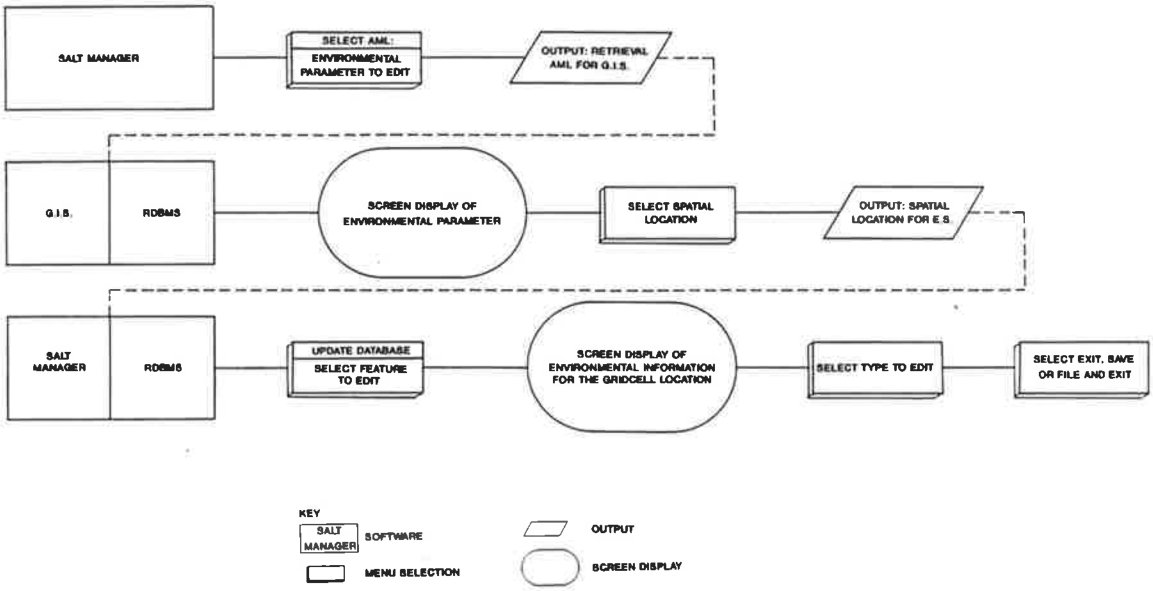
Many landholders have valuable knowledge of environmental parameters, such as depth to groundwater, and it is important to capture this information. Updating the database may only occur at the scale of individual gridcells. There are three different methods applied to update the environmental data, with each being specific to a particular environmental data type. The first is via a pop-up contact screen from the ES. Using this update method, the parameters of soil, geology, slope and elevation may be updated. As these parameters are static it is not expected that the end user would alter any of the information provided. Yet the opportunity must be made available. The second is by editing an Arc/Info point coverage in the GIS, this is particularly relevant for the dynamic parameter depth to groundwater. Figures 5 and 6 are flow diagrams indicating the decision making process for these two methods. The third is by directly updating the data within the RDBMS. This option will not be considered as only experienced Oracle database managers would be able to complete this task.

UPDATE DATABASE VIA THE ES

The first steps in this process are the same as previously considered in the "analysis" and "interrogation" functions of the system (Figure 4). The end user selects from the *Salt Manager* menu an environmental feature to view, e.g. soil. It may be the case that the feature is selected due to a disagreement by the user with an explanation provided by the interrogative component of the *Salt Manager* system, e.g. "the soil in that area does not fit that description". Once, in this case, the soil map is displayed using the functionality of the AML menu, the end user may zoom into an appropriate area and then select the appropriate location.

By selecting from *Salt Manager*, the command *Update database*, followed by *select feature to edit*, a data input contact is initiated. The four types of attribute information, soil, geology, slope and elevation, can be edited by this contact. The cell's current value appears indented when the contact is established. With each data input contact, a press of the middle button is an event that signals that "help" information is required. The result of a "help" event manifests itself in the occurrence of a popup window displaying a text file with relevant information.

FIGURE 5: Updating the Database via the ES



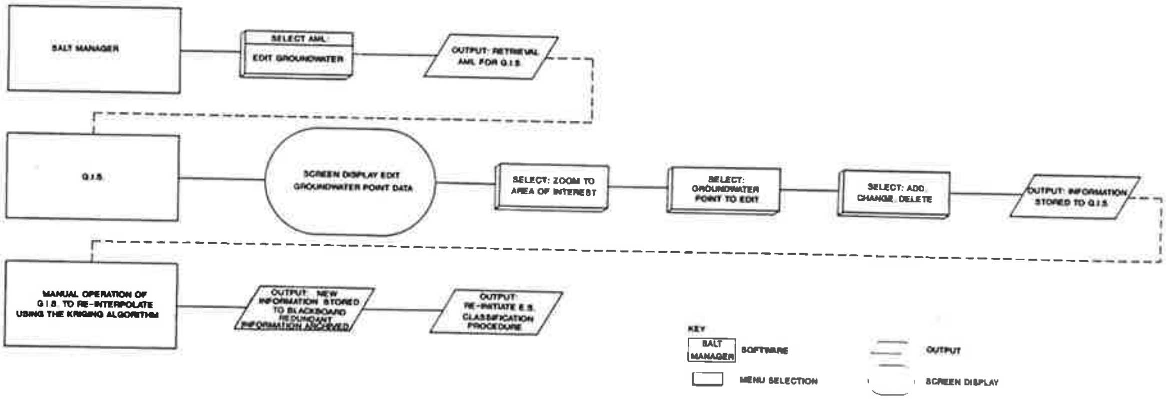
Upon completion of the updating process the user is asked by the *Salt Manager* if they wish to commit all changes to the database. If no, all updated information is cleared; if yes, the changes are committed to the RDBMS. Subsequently, the salinity land classification procedure may be initiated again by selecting the appropriate command from the *Salt Manager* menu. It results in a new land salinity class map based on the new information being produced and written to the RDBMS. At this stage of development, the run/time for the ES processing the data within the RDBMS is approximately 3 hours. As the system is designed to have interactive capabilities with the user, this needs to be reduced.

UPDATE DATABASE VIA THE GIS

By providing a menu from the editing module within the GIS the depth to groundwater data collection points may be updated, altered or deleted by the user. The process is illustrated by a flow diagram (Figure 6).

Figure 6 highlights how the initial decision stages for the operation of the *Salt Manager* system are replicated for each function of the system. More importantly it also illustrates the editing functionality of the GIS. Each point value may be edited via the AML menu.

FIGURE 6: Updating the Database via the GIS



To change the groundwater depth value at each point (displayed in cm), the user selects the point and then enters the value (depth to groundwater) in centimetres. An addition to the point coverage requires that the new point is surveyed using GPS methods into the database. If any of the values are altered, added or selected, the changes are saved and a GIS operator is required to re-interpolate the groundwater coverage. Once interpolated the changes are saved to the related blackboard table in the RDBMS.

A feature of the updating process is the use of the archiving facility within the RDBMS. All changed values for each gridcell are stored in archive files accessible at any later date. This facility allows temporal analysis of environmental change once a history is established.

REFERENCES

Brownston, L., R. Farrell, E. Kant, N. Martin (1986), Programming Expert Systems in OPS5, Addison-Wesley, New York.

Harlequin Ltd. (1992), Knowledgeworks and Lispworks Users Guide, Edition 2, Cambridge, England.

Kimbrough, K. and O. LaMott (1989), Common Lisp User Interface Environment, Texas Instruments Inc.

McFadyen, D. and A. Kanabar (1991), Structured Query Language, W.McBrown Publishers, Dubuque, USA.

Moon, D.A. (1989), "The Common Lisp Object-Oriented Programming Language Standard", in Object-Oriented Concepts, Databases and Applications, W.Kim and F. Lochovsky (Eds.), ACM Press.

Morehouse, S. (1989), The Architecture of Arc/Info, Autocarto, No. 9, pp. 266-277.

Oracle Ltd. (1990), Oracle Users Manuals, Version 6.0, Oracle Corporation, USA.

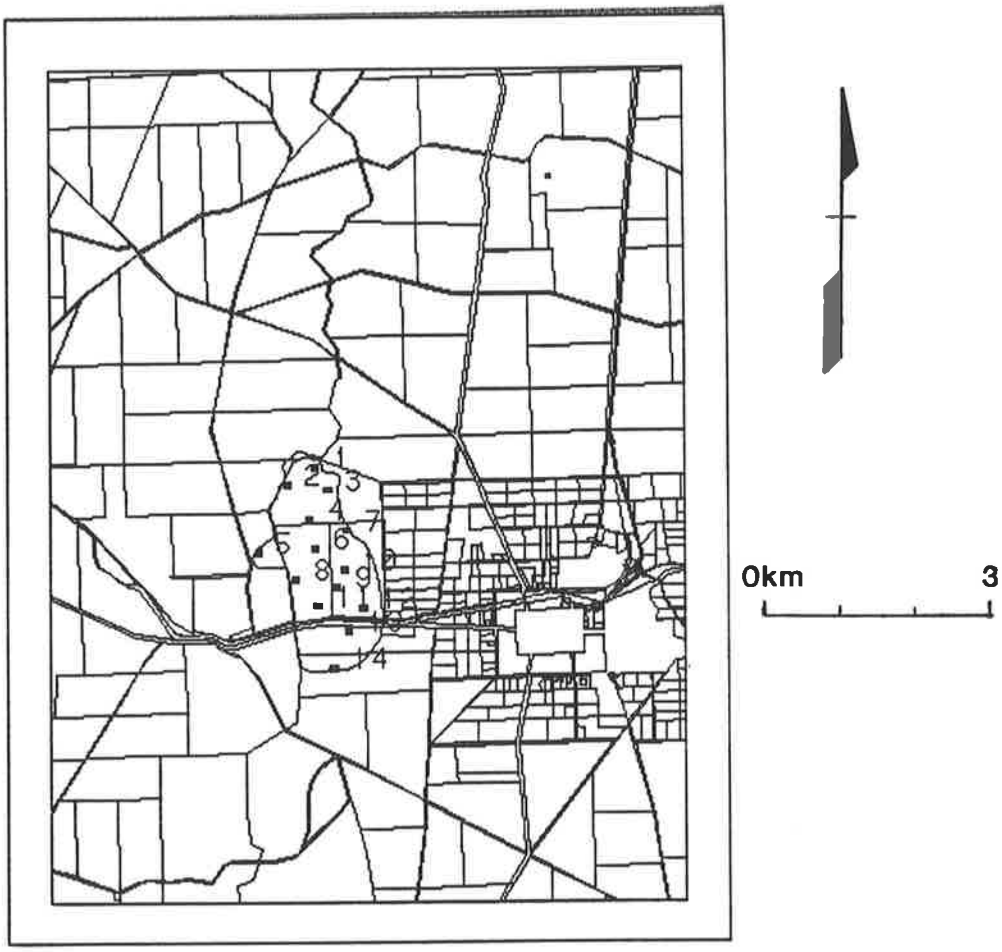
Appendix 2

DRILLING AND SAMPLING

2. Drilling and Sampling

The catchment was drilled at 14 sites with a rotary air blast rig in February 1991, with an additional three sites being drilled in April 1992. Location of the sites is shown in Figure 2.1.

FIGURE 2.1: Catchment and Drill Location sites



Drilling spoil was sampled within the unsaturated zone at 14 sites in the catchment. Samples were initially collected at 0.5m and 1.0m below ground level, and then later at 1m depth intervals until groundwater intersected, or recirculation of drilling fluid commenced. Samples were stored in sealed plastic bags or jars. Drilling logs were obtained by relating drill rig behaviour (i.e. speed and ease of drilling) to drilling spoil collection. Field descriptions included colour, texture, mineralogy, grain size and moisture content. Laboratory analysis included gravimetric moisture content, pH, EC1:5, cations (Na, Mg, K, Ca) and anions (Cl, SO4, NO3, CO3, and HCO3). Matric suction was measured at one site using the filter paper technique (Hamblin, 1981).

Appendix 3

FIELD DESCRIPTION OF SOIL UNITS

Profile: 1

Landform Unit: 1 (steeper slopes/ridges)

Soil Name: "Yangya Silty Loam"

Depth (cm)	Soil Profile Descriptions
0 - 5	Brown silty loam with limestone pebbles, siltstones and quartz stones.
5 - 10	Red brown silty loam with weathered siltstones and minor soft limestones.
10 - 15	Dark red brown clay with weathered siltstones.
15 - 20	Red brown to greenish grey silt loam.
20 - 40	Greenish grey rubbly silty loam with red brown silty loam mottles.
40 - 50	Greenish grey rubbly silty loam with dark brown light clay mottles.

Profile: 2

Landform Unit: 1 (steep slopes/ridges)

Soil Name: "Canowie Loam"

Depth (cm)	Soil Profile Descriptions
0 - 5	Dark red brown sandy loam soil is well structured (self mulching). Stones of quartzite, ironstone and limestone.
5 - 15	Red brown sand clay loam, stony.
15 - 40	Red brown clay - well structured, stony.
40 - 50	Red brown crumbly clay with soft white lime patches.
50 - 60	Brown crumbly clay with soft line and limestone pebbles.

Profile: 3

Landform Unit: 3 (alluvial plain)

Soil Name: "Belalie Loam"

Depth (cm)	Soil Profile Descriptions
0 - 5	Red brown loam.
5 - 20	Light red brown loam to clay loam - weak structure.
20 - 40	Orange brown loam to clay loam - slightly limy.
40 - 50	Red brown loam.
50 - 55	Red brown loam, quartzite pebbles.
55 - 80	Dark red brown crumbly clay with small ferruginous pebbles.
80 - 95	Dark red brown crumbly clay with yellow silty mottles and white limy mottles.
95 - 100	As above but becoming more pebbly.

Profile: 4

Landform Unit: 2 (gentle slopes)

Soil Name: "Belalie Clay Loam"

Depth (cm)	Soil Profile Descriptions
0 - 5	Dark red brown light clay. Self mulching cracking clay.
5 - 10	Red brown light clay with soft lime.
15 - 30	Red brown tight clay with quartz and limestone pebbles.
30 - 35	Brown well structured clay with soft lime, limestone, quartz and ironstone pebbles.
35 - 50	Brown pedal tight clay with increasing lime content.

Profile: 5

Landform Unit: 1 (alluvial plain)

Soil Name: "Belalie Loam"

Depth (cm)	Soil Profile Descriptions
0 - 5	Red brown loam.
5 - 25	Red brown loam to clay loam, weak structure.
25 - 45	Red brown loam.
45 - 50	Red brown clay loam.
50 - 60	Dark red brown clay.

Profile: 6

Landform Unit: 2 (gentle slopes)

Soil Name: "Yarcowie Loam"

Depth (cm)	Soil Profile Descriptions
0 - 10	Red brown clay loam, well structured.
10 - 25	Red brown clay loam with a few yellow and white mottles.
25	Calcrete.

Appendix 4

PIEZOMETER INSTALLATION

4. Piezometer Installation

TABLE 4.1: Piezometer Location Records

BORE ID	AHD[1](m) GROUND	TUBE AGL[2] (m)	CASED DEPTH[3] (m)	SCREEN LENGTH (m)	MATERIALS SCREENED
J1A	456.08	0.38	1.24	0.70	talcy clay
J1C	465.11	0.66	32.60	1.00	siltstone/slate
J2A	463.87	0.32	1.20	0.70	talcy clay
J2C	463.86	0.68	22.07	1.00	weathered rock
J3A	464.67	0.32	1.16	0.70	clay
J3C	464.65	0.50	17.25	1.00	fractured rock
J4A	455.97	0.41	1.22	0.70	clay
J4B	455.93	0.50	11.33	0.50	weathered rock
J4C	455.95	0.52	17.26	1.00	fractured rock
J5A	458.27	0.34	1.48	1.00	clay
J5C	458.28	0.63	19.76	1.00	siltstone
J6A	450.27	0.25	1.26	1.00	clay
J6B	450.28	0.51	8.30	0.50	talcy clay
J6C	450.28	0.60	28.52	1.00	fractured rock
J7A	456.16	0.24	1.25	0.70	clay
J7B	456.17	0.47	10.61	1.00	weathered slate
J7C	456.17	0.49	14.12	1.00	fractured rock
J8A	449.30	0.33	1.36	0.70	clay
J8B	449.30	0.50	8.29	0.50	talcy rock
J8C	449.30	0.60	25.82	1.00	sandstone
J9A	444.42	0.35	2.03		clay
J9B	444.41	0.50	8.50	0.50	talcy clay
J9C	444.41	0.54	19.23	1.00	talcy shale
J10A	444.62	0.25	1.20	0.70	clay
J10B	444.63	0.51	2.61	0.50	mottled clay
J10C	444.64	0.75	26.13	0.50	weathered rock
J11A	445.19	0.32	1.38		clay
J11B	445.18	0.51	2.70	0.50	sandy clay
J11C	445.19	0.64	34.05	0.50	siltstone/slate
J12A	441.06	0.39	2.02	1.00	clay
J12B	441.04	0.53	5.06	1.00	clay
J12C	441.04	0.59	28.37	0.50	weathered shale
J13A	442.24	0.51	1.51	1.00	calcrete
J13C	442.24	0.80	34.59	0.50	siltstone/shale
J14A	449.59	0.29	1.20	0.70	talcy clay
J14C	449.63	0.53	18.30	1.00	gritty talcy clay
J15B	442.04	0.12	3.88	1.00	
J16B	440.58	0.52	4.01	1.00	
J17C	447.18	0.50	14.16	0.50	sandy clay

[1] AHD = Australian Height Datum

[2] AGL = Above ground level

[3] Cased depth is below ground level

Appendix 5

GENERATION OF A DECISION TREE AND GENERALISING FROM DATAPOINTS

GENERATION OF A DECISION TREE AND GENERALISING FROM DATAPOINTS

This appendix considers two symbolic learning tasks that were completed during the knowledge acquisition phase of the thesis development. The first section describes decision tree generation and pruning strategies. The second gives consideration to another experiment conducted to verify which AI technique was more suitable for generalising between datapoints.

5.1 DECISION TREE GENERATION

A decision tree is a tree where nodes stand for attributes, branches for attribute values and leaves stand for classes. A decision tree is read from top to bottom. Every branch from the root of the tree to the leaf represents one decision rule. The skill in constructing a tree is to select an attribute for the current node so that the final decision tree is in some way minimal.

ID3 uses a criterion called *gain*, an information theoretic metric, for minimising the expected number of tests to classify an example (Eklund and Salim, 1993). The information theory that underpins this criterion is: the information conveyed by a message depends on its probability and can be measured in bits as minus the logarithm to base 2 of that probability (Salim, 1993). Therefore, selecting one case at random from a set S of cases and announcing that it belongs to some class Cj will have probability:

$$\frac{freq(C_j, S)}{|S|}$$

where *freq*(Cj,S) stands for the number of cases in S that belong to the class Cj and |S| denotes the number of cases in set S (Salim, 1993). The information it conveys is:

$$-\log_2\left(\frac{freq(C_j, S)}{|S|}\right) bits.$$

The expected information from such a message pertaining to a class membership is found by summing over the classes in proportional to their frequencies in S, giving (Salim, 1993):

$$info(S) = -\sum_{j=1}^{k}\left[\left(\frac{freq(C_j,S)}{|S|}\right) \times \log_2\left(\frac{freq(C_j,S)}{|S|}\right)\right] bi$$

When applied to the set of training cases, info(T) measures the average amount of information needed to identify the class of a case in T. This quantity is also known as entropy of the set S. Now consider similar measurement after T has been partitioned in accordance with n outcomes of an attribute X. The information requirement can be found as the weighted sum over the subsets (Salim,1993), as:

$$info_X(T) = \sum_{i=1}^{n}\frac{|T_i|}{T} \times info(T_i).$$

The quantity,

$$gain(X) = info(T) - info_X(T),$$

measures the information that is gained by partitioning T in accordance with the attribute X (Salim, 1993). In summary, it uses this criterion to select an attribute as a candidate node in the decision tree (Salim, 1993). The gain criterion selects an attribute which maximises this information gain.

Although the gain criterion gives good results, it has a serious deficiency of having strong bias in favour of tests with many outcomes. This is problematic because the examples given to the inductive learning program may not be representative. The problem associated with gain can be seen if one of the attributes in the task contains grid identification which identifies a unique grid point in a spatial addressed database.

Since every identification is intended to be unique, partitioning any set of training cases on values of this attribute will lead to a large number of subsets, each containing just one case. Since all of these one-case subsets, each contain cases of a single class, *info*x(T) = 0, so information gain from using this attribute to partition the set of training cases is maximal (Salim, 1993). From the point of view of prediction, however, such a division is useless.

The bias inherent in the gain criterion was rectified by a kind of normalisation in which the apparent gain attributable to tests with many outcomes is adjusted. Consider the information content of a message pertaining to a case that indicates not the class to which the case belongs, but the outcome of the case. By analogy with the definition of *info*(S), we have (Salim, 1993):

$$split\ info_X(T) = \sum_{i=1}^{n} \frac{|T_i|}{T} \times \log_2 \left(\frac{|T_i|}{T} \right).$$

This represents the potential information generated by dividing T into n subsets, whereas the information gain measures the information relevant to classification that arises from the same division (Salim, 1993). Then,

$$gain\ ratio(X) = gain(X)/split\ info(X),$$

expresses the proportion of information generated by the split that is useful, i.e., that appears helpful for classification (Salim, 1993).

At every decision level C4.5 selects the most informative attribute, an attribute which contributes the most information for classification of the learning examples in the current node. When the best attribute is selected, the set of attribute values is split into subsets of values, each for one branch. Then, the set of learning examples is split according to the values of the selected attribute into subsets. Each subset of the learning examples is then used to construct the subtrees. When a termination condition is satisfied in the current subset of examples, a leaf node is generated. The termination conditions are:

- all examples from a subset belong to a single class;
- no more attributes are available; and
- the pruning factor is satisfied (Salim, 1993).

5.2 DECISION TREE PRUNING

In real-world classification tasks it is frequently the case that the class of objects in the training set cannot be expressed as a function of the attribute values. This can arise either because the attribute values contain errors, or because the attributes collectively provide insufficient information to classify an object.

In these circumstances, continuing to divide the training set until all subsets contain members of a single class may be impossible. Even when such perfect division is possible, it may be inadvisable, as the lower levels of the tree may model idiosyncrasies of the training set rather than structure that is useful for classifying an unseen object; it may be attempting to fit the noise in the training data (Quinlan, 1990).

One way to deal with this problem is to let the tree grow and then remove unimportant or unsubstantiated portions by *pruning* it (Salim, 1993). Pruning involves replacing a subtree by one of its branches or, more commonly by a leaf. It usually results in a smaller subtree that has the additional advantage of increased accuracy when classifying unseen objects.

The following section considers an experiment conducted to compare efficiency of the C4.5 programme to an Instance Based Learning (IBL) programme and a Back Propagation (BP) algorithm.

5.3 GENERALISING FROM DATAPOINTS

In the course of the formulation of the methodology to restructure the rule base, the author became aware of other methods by which to classify a data set. Rather than just accept that the C4.5. programme was most suitable for the classification task it was decided to test which of the methods was more efficient. With efficiency, in this case, being the minimum number of data points required to replicate the classification task. In this sub-section the following software systems are compared: the C4.5 programme, four IBL systems and a BP algorithm (Eklund, 1993)[1]. As the C4.5 programme has already been considered in section 7.2, only the other systems will be defined here.

---

[1] A journal paper describing this topic is currently being formulated. The authors are P.W. Eklund, S.D. Kirkby and A. Salim.

5.3.1 Instance Based Learning

One way to classify a case is to recall a similar case whose class is known and to predict that the new case will have the same class. This philosophy underlies an IBL system which can classify unseen cases by referring to similar remembered cases.

All IBL algorithms are derivable from the well known nearest neighbour pattern classifier and define the similarity of two instances as their Euclidean distance in the instance space (Cover and Hart, 1967). The nearest neighbour classifies and instance as being a member of the same concept as its most similar instance. All IBL algorithms consist of the following three components (Aha and Kibler, 1989; Aha et al., 1991):

1. Similarity function: This computes the similarity between a training instance *i* and the instances in the concept description. Similarities are numeric-valued.
2. Classification function: This receives the similarity function's results and the classification performance records of instances in the concept description. It yields a classification for *i*.
3. Concept description updater: Maintains records on classification performance and decides which instances too include in the concept description. Inputs include *i*, the similarity results, the classification results, and a current concept description. It yields the modified concept description.

The similarity and classification functions determine how the set of saved instances in the concept description are used to predict values for the category attribute. Therefore, IBL concept descriptions not only contain a set of instances, but also includes theses two functions.

IBL algorithms assume that similar instances have similar classifications. This leads to their local bias for classifying novel instances according to their most similar neighbours classification. IBL algorithms also assume that, without a prior knowledge, attributes will have equal relevance for classification decisions. Each instance is normalised to ensure that attributes are assigned equal importance by the similarity function. Assuming that attributes have equal importance is not necessarily correct, but it is a fair approach (Aha, et al., 1991).

Several IBL algorithms are provided by Aha and Kibler (1989), Kibler et al. (1989), and Aha et al. (1991). Brief descriptions of the IB1 and IB2 algorithms and a multi-pass instance learning (MPIL) algorithm are given below (Romaniuk, 1993). Detailed analysis of these algorithms is provided by the articles which have been referred to.

5.3.1.1 The IB1 Algorithm

The IB1 algorithm is the simplest instance-based learning algorithm. It is essentially equivalent to the nearest neighbour algorithm, except it normalises the attributes' ranges, and has the capability to deal with missing values.

5.3.1.2 The IB2 Algorithm

The IB2 algorithm is identical to IB1 except that it saves only misclassified instances. It displays a high degree of similarity with the condensed nearest neighbour algorithm, since it saves and uses only selected instances for making its predictions (Hart, 1968). By being selective in the instances it saves, its storage requirements can be significantly smaller than those of IB1 especially when instances are adequately separated in feature space. On the other hand, IB2 appears to be more sensitive to noise than IB1.

5.3.1.3 Multi-Pass Instance Learning

The multi-pass instance learning allows multiple passes over the complete set of instances as opposed to a single pass as performed by the IB1 and IB2 algorithms. An instance here is represented by features; where a feature represents a continuous attribute or a discrete attribute value. Furthermore, associated with every instance is a neighbourhood (spanned by the instance), which can be either a sphere (MPIL1 algorithm) or a k-dimensional (MPIL2 algorithm). The details of these two algorithms are given in Romaniuk (1993).

The purpose of "neighbourhood" is to define a degree of similarity between instances defining the same concept. Storage reduction can be achieved by firstly: appropriately defining the neighbourhood for every instance; and secondly by removing instances that fall within another instance's neighbourhood (assuming only instances describing the same concept). Instances associated with different concepts are used to define valid neighbourhoods. A neighbourhood is considered valid, if the instances enclosed within it describe the same concept as the instance that spans the neighbourhood (Romanuik, 1993).

5.3.2 Back-Propagation (BP)

The back-propagation algorithm is central to much current work on learning in neural networks. Its discovery in relation to neural networks is generally accredited to Rumelhart et al. (1986a,b). However, in different contexts it was invented independently several times (Bryson and Ho, 1969). The algorithm gives a prescription of changing the weights w_{ij} in any feed-forward network to learn a training set of input-output pairs.

Input units record observed features from the training dataset and pass "activation" forward through an intermediate layer of "hidden" units to output layer. Each node is assumed to be linked to every unit at the next layer via weighted interconnections, which encode the network knowledge. The activation levels of the units in the output layer determine the output of the network. The existence of hidden units allows the network to develop complex feature detectors, or internal representations.

There are two conventions in use for counting the number of layers in the network; some authors count the input units as a layer, some do not. In this thesis the latter convention is used. It is becoming more frequently adopted and seems more logical since the input units play no significant role (Hertz, et al., 1991).

The application of the back propagation rule involves two phrases: forward and backward pass. During the first phase the input is presented and propagated forward through the network to compute the output value O_i for each unit. This output is then compared with the target, resulting in an error propagation term, S, for each output unit. The second phase involves a backward pass through the network during which the S term is computed for each hidden unit in the network. This second, backward pass allows the recursive computation of S. Once these two phases are complete, the actual weight changes can be computed by minimising an error measure, or cost function.

The back-propagation is implemented in the C programming language. The discrete attributes are encoded using a set of input units, one dedicated to each value. For example, Geology is represented with 8 input units. If a case has a Geology value '5', then the fifth input unit of Geology is set to 1.0 while the other seven units to 0.0. Each continuous attribute is represented with one input unit, and their values are scaled between 0.0 and 1.0; 0.0 represents the minimum value, and 1.0 represents a maximum value of that attribute. The output units correspond to classes; one unit for each class. The network was tested with several hidden units, and produced good results with 10 units. Therefore, for the salinity data there were 31 input, 10 hidden and 6 output units.

The network was tested with various learning rates (0.05, 0.1, 0.5, 1) and momentum (0.0, 0.1, 0.5, 0.9). Of these, the 0.05 learning rate and 0.9 momentum value, as initial parameters of adaptive technique, gave best results. The initial weights w_{ij} were set in the range,

where k_i is the number of inputs to unit i, and incrementally updated (one pattern at a time chosen in random order from the training set). This approach of setting and updating weights avoids local minima.

The quadratic cost function is not only the acceptable cost function. Any other differential function that is minimised when its arguments are equal, and derive a corresponding update rule can be used. Therefore, another network was implemented replacing the quadratic cost function with the entropy measure. This is for a +/- 1 range units and, using $g(x) = \tanh(x)$. The results follow in the next sub-section.

5.3.3 Description of Experiment

The three classifier's (C4.5, IBL and BP) were tested on the salinity E.S. classification. The classification was partitioned into two sets: training and test. Of the 121,404 salinity datapoints, 1300 were drawn randomly. To reduce statistical fluctuations, results are averaged over five different training and testing sets produced by randomly placing two-thirds of the examples for each category in the training set and the others in a corresponding test set. Again to reduce statistical fluctuations, each run of BP uses a different random seed number which determines the initial network weights.

5.3.3.1 Training and Testing

All supervised learning classifier's involve two phases: training and testing. Since IB1 stores all training data, it does not involve any training phase. For IB2, MPIL1 and MPIL2 training involves storing prototypical cases that summarises train data. All four IBL algorithms have the same testing phase. Each test case is scanned through the stored data and picks the most similar case to it. If the classes of the two cases are the same then the classification is considered to be correct, otherwise it is incorrect.

In BP training consists of consecutive forward and backward passes. This process is repeated until a satisfactory generalisation performance is obtained and the weights saved. Testing consists of forward passing an input pattern and assigned to the class whose output unit has the highest value. The training phase for inductive learning involves, as discussed in section 7.2, the generation of decision tree and classify the test data using the tree.

5.3.4 Experimental Results

The first set of experiments attempts to compare the performance accuracy of the four IBL algorithms. The performance accuracy is split in three parts: first, the performance on training set only; second, performance on the unseen test set to determine the generalisation capabilities of the algorithms. Finally, the storage requirements of the different approaches are listed. IB1 is used as a base case, since it requires storage of all training instances.

TABLE 5.1: Performance Accuracy of IBL Algorithms

Algorithm	Performance on Salinity Data		
	Train	Test	Storage
IB1	100.0%	89.0%	100.0%
IB2	89.2%	85.6%	13.88%
MPIL1	100.0%	89.0%	12.5%
MPIL2	100.0%	95.8%	9.9%

Table 5.1 gives insight into the performances of the four learning algorithms on the salinity domain. The train set size is 800, and the test size is 500. When comparing the classifier's, the most striking result is the good performance of MPIL1 and MPIL2 over IB1. Both MPIL1 and MPIL2 show the most significant improvement in performance for the data compared to the IB1 and IB2. They achieved perfect results on the train set, and a very high degree of accuracy on the test set, but only required a small fraction of the storage. MPIL2 outperformed the others for the salinity classification.

Table 5.2 lists the results achieved by back-propagation, for the two activation functions: sigmoid and hyperbolic tangent. The number of hidden units was varied from none (one-layer only) to 20. The results are given for the salinity data; for 0, 5, 10 and 15 hidden units. The network gives high results with 10 hidden units and using sigmoid as its activation function. No improvement was visible in the results on increasing the number of hidden units 10 to 20.

TABLE 5.2: Performance Accuracy of Back-Propagation

Hidden Units	Accuracy of Activation Functions: Salinity Area	
	Sigmoid Function	Hyperbolic Tangent
0 5 10 15	92.6% 96.9% 97.6% 95.3%	91.0% 94.7% 97.1% 94.8%

Inductive learning achieved 99.8% and 96% (after pruning the decision tree) accuracies on the salinity data (Table 5.3).

TABLE 5.3: Performance Accuracy of C4.5

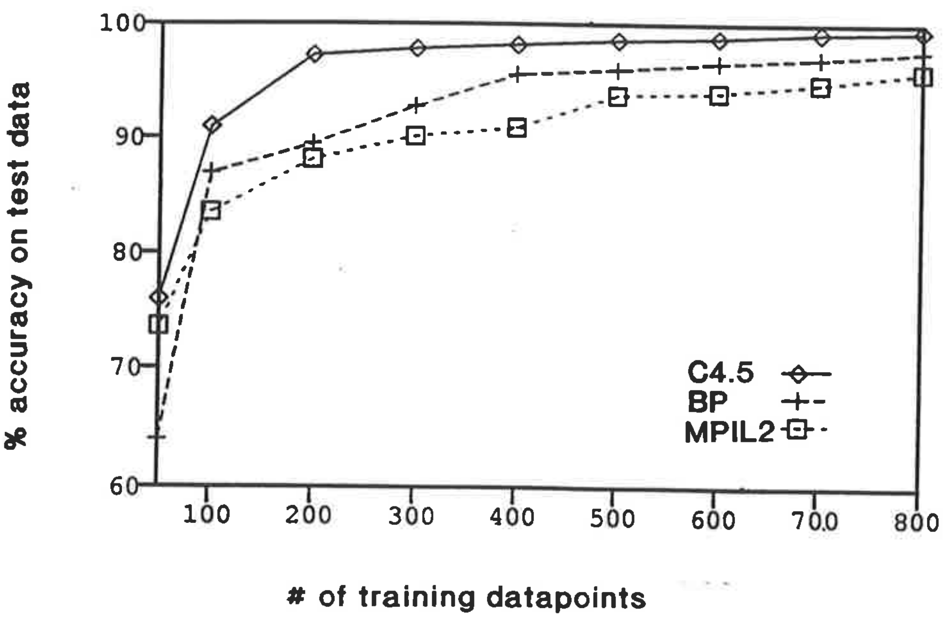
Domain	Accuracy of C4.5		
	Decision Tree	Pruned Tree	Rules Generated
Salinity	99.4%	99.8%	99.8%

5.3.5 Comparing the Results

The graph of Figure 5.1 show the "learning curves" of the three techniques for salinity data. The graph shows the speed at which accuracy increases during training. Given that the main expense of land management strategies is in data collection, a crucial concern of land management and conservation authorities is the number of datapoints which need to be collected in order for generalisations to be drawn. Therefore, the learning curve is a more useful indicator of the performance of the classifier than is accuracy for finite-sized training sets.

From the graphs of Figure 5.1, it can be observed that C4.5 produces a classification accuracy of better than 97% even for a small training set of 200 instances. To achieve the same accuracy using back-propagation more than 500 datapoints are needed, and more than 700 using MPIL2. BP and MPIL2 perform reasonably well for large training sets but their performance degrades rapidly for small training sets.

FIGURE 5.1: Accuracy as a Function of Datapoints



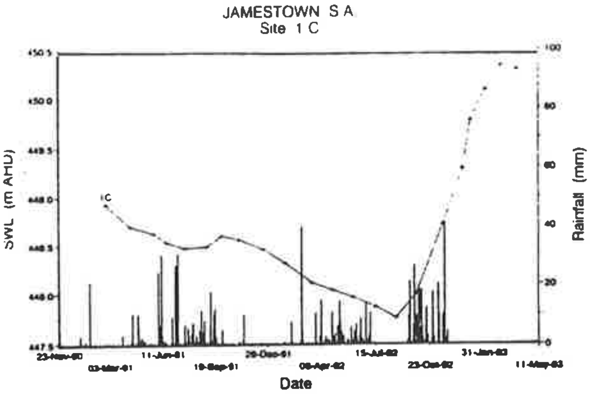
It is fair to observe that the C4.5 significantly outperforms both BP and MPIL2. This is due the fact that there are certain salinity class discriminators which are more significant than others, e.g. depth to groundwater is a prime discriminator of high potential/existing groundwater discharge. In domains where attributes have a more uniform influence the performance of BP and MPIL2 should be much improved.

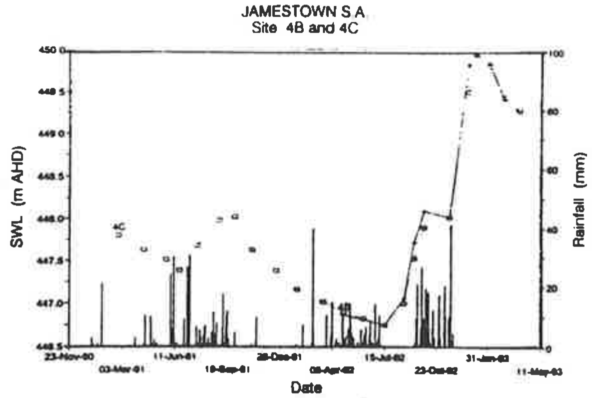
Appendix 6

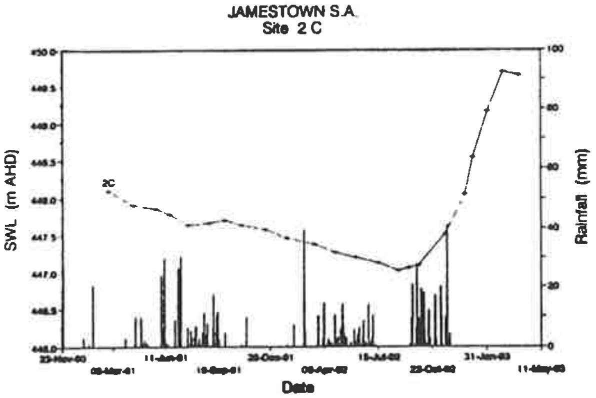
GROUNDWATER HYDROGRAPHS :

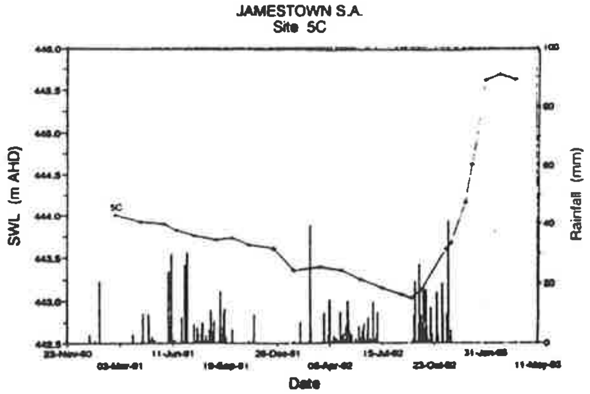
JAMESTOWN, S.A.

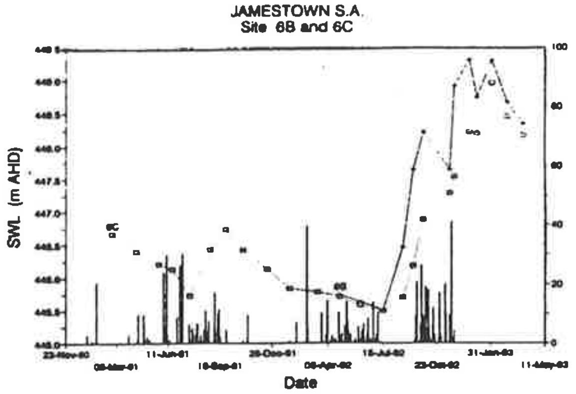
Source: Henschke _et al_. (1993)

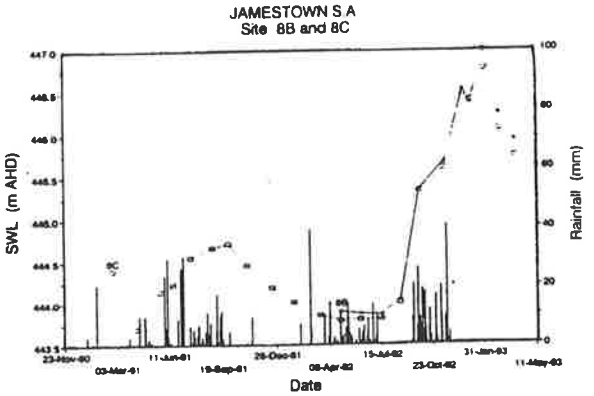

JAMESTOWN S.A.
Site 1 C

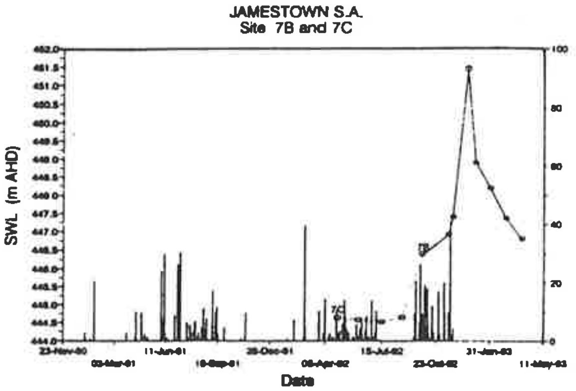

JAMESTOWN S.A.
Site 4B and 4C

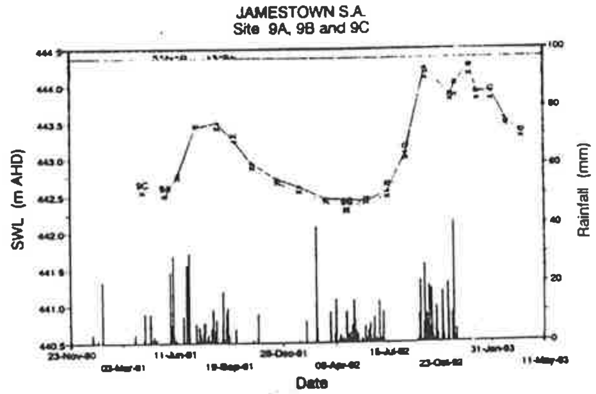

JAMESTOWN S.A.
Site 2 C


JAMESTOWN S.A.
Site 5C

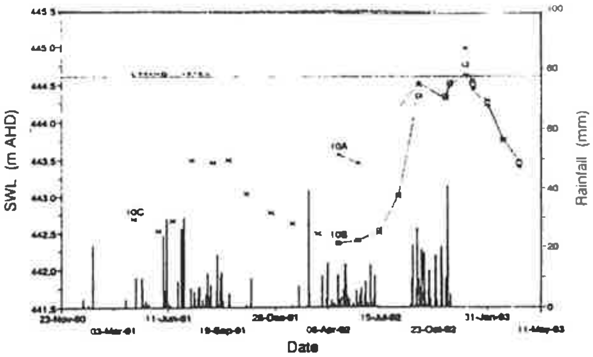

JAMESTOWN S.A.
Site 6B and 6C


JAMESTOWN S.A.
Site 8B and 8C

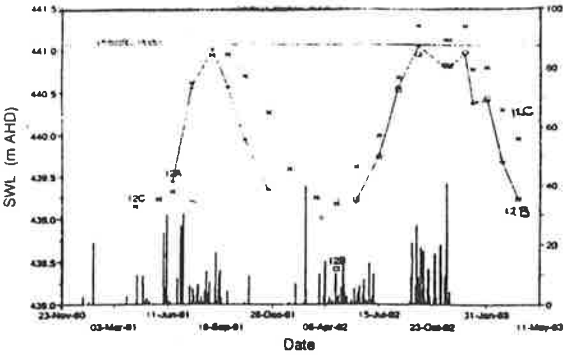

JAMESTOWN S.A.
Site 7B and 7C


JAMESTOWN S.A.
Site 9A, 9B and 9C

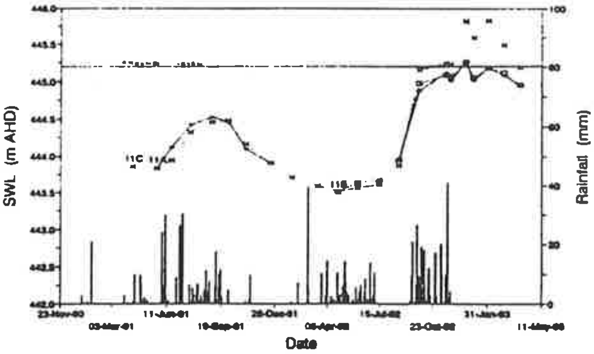
JAMESTOWN S.A.
Site 10A, 10B and 10C



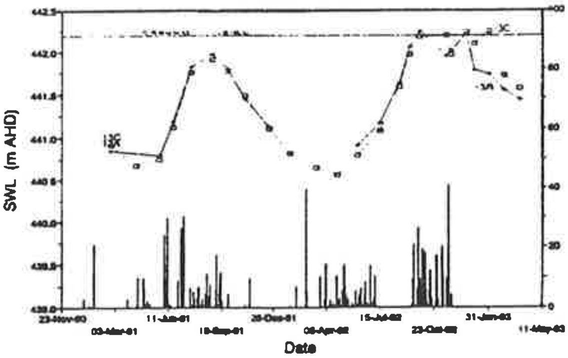
JAMESTOWN S.A.
Site 12A, 12B and 12C



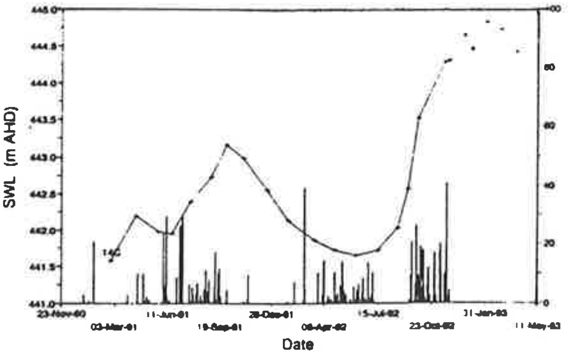
JAMESTOWN S.A.
Site 11A, 11B and 11C



JAMESTOWN S.A.
Site 13A and 13C



JAMESTOWN S.A.
Site 14C



Appendix 7

VERIFICATION QUESTIONNAIRE : BUNDALEER VALLEY LANDCARE GROUP

VERIFICATION QUESTIONNAIRE: BUNDALEER VALLEY LANDCARE GROUP

Questionnaire

1. Do you have a computer at home? Yes / No

2. What do you use the computer for?

   a. Games
   b. Word Processing
   c. Accounting
   d. Crop Management
   e. Stock Management

3. Have you ever used information derived from a computer to assist with your cropping or stocking strategies?

   Yes / No

4. Do you see yourself using a computer to assist with your farm management strategies?

   Yes / No

5. If such a software system, as demonstrated, existed in the local Agricultural Office would you be interested in using such a system?

   Yes / No

6. Do you think you would interrogate the data, i.e. would you be interested in knowing the explanation for a parcel of land?

Yes / No / Undecided

7. Would you update the database, i.e. if you know the groundwater levels for certain points, would you update the database?

Yes / No / Undecided

8. Would you utilise the information provided by the system to assist with your farm management strategies?

Yes / No / Undecided

REFERENCES

Aha, D.W. and D. Kibler (1989), "Noise-tolerant instance-based learning algorithms", Proceedings of the Eleventh International Joint Conference on Artificial Intelligence, Detroit Morgan Kaufmann.

Aha, D.W., D. Kibler and M.K. Albert (1991), "Instance-based learning algorithms", Machine Learning, V.6, pp. 37-66.

Barnett, S. (1990), "Identifying and Monitoring the Cause of Dryland Salinity and Predicting its Spread", in Proceedings of a Technical Forum on Dryland Salinity in South Australia, D. McCarthy (ed.), Department of Agriculture, S.A., Technical Report No. 163, April 1990.

Bell, A. (1989), "Trees Water and Salt - A Fine Balance", Ecos, No. 58, pp. 2-8.

Benn, B.O. and W.E. Grabau (1968), "Terrain Evaluation as a Function of User Requirements", in G.A. Stewart (1968), Land Evaluation, Macmillan, Sydney.

Bettenay, E., A.V. Blackmore and A.J. Hingston (1964), "Aspects of the Hydrologic Cycle and Related Salinity in the Belka Valley, Western Australia, Australia", Journal of Soil Resources, Vol. 2, pp. 187-210.

Bettenay, E. and M.J. Mulcahy (1972), "Soil and Landscape Studies in Western Australia 2. Valley Form and Surface Features in the South-West Drainage Division", Journal Geological Society of Australia, Vol. 18, pp. 359-369.

Boehm, B.W. (1981), Software Engineering Economics, Prentice Hall, Englewood Cliffs, NJ.

Bonham-Carter, G.F., F.P. Agterberg and D.F. Wright (1990), "Integration of geological datasets for gold exploration in Nova Scotia", in D.J. Peuquet and D.F. Marble, Introductory Readings in Geographic Information Systems, Taylor and Francis, London.

Borg, H., R.W. Bell and I.C. Loh (1988), "Streamflow and Stream Salinity in a small Water Supply Catchment in Southwest Western Australia after Reforestation", in Journal of Hydrology, Vol. 103, pp. 323-333.

Bowerman, R.G. and D.E. Glover (1988), Putting Expert Systems into Practice, Van Nostrand Reinhold Co., New York.

Bresler E., B.L. McNeal and D.L. Carter (1982), Saline and Sodic Soils, Springer-Verlag, Berlin, 1982.

Breuker, J.A. and B.J. Wielinga (1984), Techniques for knowledge elicitation and analysis, Report 1.5, Esprit Project 12 (Memorandum 28 of the research project The Acquisition of Expertise, Amsterdam, University of Amsterdam).

Breuker, J.A. and B.J. Wielinga (1987) "Use of Models in the Interpretation of Verbal Data", in Knowledge Acquisition for Expert Systems: A Practical Handbook, A.L. Kidd (ed.), New York, Plennum. pp. 17-44

Brown, P.L., E.C. Cleary and M.R. Miller (1976), "Water Use and Rooting Depths of Crops for Saline Seep Control" in Regional Saline Seep Control Symposium, Bull. No. 1132, Montana State University, pp. 137-144.

Brownston, L., R. Farrell, E. Kant, N. Martin (1986), Programming Expert Systems in OPS5, Addison-Wesley, New York.

Bryant, N.A. and A.L. Zobrist (1981), "Some technical considerations on the evolution of the IBIS systems", Proceedings, 7th Annual William T Pecora Memorial Symposium, South Dakota, pp. 465-75.

Bryson, A.E. and Y.C. Ho (1969), Applied Optimal Control, New York, Blaisdell.

Buchanan, B.G. and E.H. Shortliffe (1984), Rule-Based Expert Programs: The MYCIN Experiments of the Stanford Heuristic Programming Project, Addison-Wesley, Reading, MA.

Bullock, P.R. (1991), "Seepage Scald in Southeastern Australia", Applied Geography, No. 11, pp. 59-81.

Burrough, P.A. (1986), Principles of Geographical Information Systems for Land Resource Assessment, Oxford, Oxford University Press.

Burrough, P.A. (1993), Environmental Assessment and G.I.S, Workshop Conducted at the University of New South Wales, June 7, 1993.

Campbell, N.A., and W.R. Atchley (1981), "The Geometry on Canonical Variate Analysis", Systematic Zoology, V. 30 pp. 268-280.

Cestnik, B., I. Kononenko, and I. Bratko (1987), "Assistant 86: A Knowledge-elicitation tool for sophisticated users", Progress in Machine Learning, Wilmslow, UK: Sigma Press.

Cestnik, B., I. Kononenko, and I. Bratko (1988), Assistant Professional Systems Users Manual, Edward Kardelj University.

Chalmers, L. and M. MacLennan (1990), Expert Systems in geography and Environmental Studies, Department of Geography Publication Series, Occasional Paper 10, University of Waterloo.

Chapman, V.J. (1975), "The Salinity Problem in General, its Importance and Distribution with Special Reference to Natural Halophytes", in A. Poljakoff-Mayber and J.Gale (Eds), Plants in Saline Environments, Springer-Verlag, Berlin.

Chard D.L., and A. Schreiner (1990), "Expert Systems in Obstetrics and Gynaecology", Ballieres' Clinical Obstetrics and Gynaecology, Vol. 4, No. 4, pp. 815-841.

Chaturvedi, L., K.R. Carver, J,C, Harlan, G.D. Hancock, F.V. Small, K.J. Dalstead (1983), "Multispectral Remote Sensing of Saline Seeps", IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-21, No. 3, pp. 239-51.

Clarke, K.C. (1990), Analytical and Computer Cartography, Prentice Hall, Englewood Cliffs, NJ. pp. 290.

Clarke, K.C. (1986), "Recent Trends in Geographic Information System Research", Geo-Processing, Vol. 3, pp. 1-15.

Compagoni, P (1986), "An Environmental Planning Method for Australian Jurisdictions. Comments on the SIRO-PLAN/LUPLAN Scheme", Environment and Planning B: Planning and Design, Vol. 13, pp. 335-344.

Conacher, A.J. and I.D. Murray (1973), "Implications and Causes of salinity Problems in the Western Australian Wheatbelt: the York-Mawson Area", Australian Geographical Studies, Vol. 11, pp. 40-61.

Conacher, A.J. (1975), "Throughflow as a Mechanism responsible for Excessive Soil Salinisation in Non-Irrigated, Previously Arable Lands in the Western Australian Wheatbelt: A Field Study", Catena, Vol. 2, pp. 31-67.

Cook, P.G., G.R. Walker and I.D. Jolly (1989), "Spatial Variability of Groundwater Recharge in a Semi-Arid Region", Journal of Hydrology, Vol. 111, pp. 195-212.

Cook, P.G., G.Walker, S.Kilty, I. Potts, J.Buselli and D.Dodds (1992), "Using air-borne fixed frequency electromagnetic induction devices to map spatial variability of recharge in the South Australian Mallee", Third Murray-Darling

Basin Groundwater Workshop, Extended Abstract. Murray-Darling Basin Commission.

Cook, P.G. and S. Kilty (1994), "Use of a helicopter electromagnetic survey to map groundwater recharge under dryland agriculture", Water Resources Research (in press).

Cooke, J.W. and S.T. Willatt (1983), "Land Management, Water Use and Salinity Prevention", Proceedings Royal Society Victoria, Vol. 95, No. 3, pp. 117-121.

Cooper, L. (1978), Jamestown - A Photographic Study, Corporation of the City of Jamestown.

Cope, F. (1958), Catchment Salting in Victoria, Soil Conservation Authority of Victoria, Melbourne.

Cover, T.M. and P.E. Hart (1967), "Nearest neighbour pattern classification", IEEE Transactions on Information Theory, V.13,pp.21-27.

Cowen, D.J. (1988), "G.I.S. versus CAD versus DBMS: what are the differences?", Photogrammetric Engineering and Remote Sensing, Vol. 54., No. 11, pp. 1551-1555

Curran, P.J. (1985), Principles of Remote Sensing, Longman, New York.

Davis, J.R., J.R.L. Hoare and P.M. Nanninga (1986), "Developing a Fire Management Expert System for Kakadu National Park, Australia", Journal of Environmental Management, Vol. 22, pp. 215-227.

Davis, J.R., P.T. Compagnoni and P.M. Nanninga (1987a), "Roles for Knowledge-based Systems in Environmental Planning", Environment and Planning B: Planning and Design, Vol. 14, pp. 239-254.

Davis, J.R., P.M. Nanninga and D.Cocks (1987b), "The usefulness of computer aids that capture expert knowledge about land management", in Demands on Rural Lands: Planning for Resource Use, C.Cocklin, B.Smit, T.Johnston (Eds)., Westview Press, Colorado.

Davis, J.R., P.M. Nanninga, J.R.L. Hoare and A.J. Press (1989), "Transferring Scientific Knowledge to Natural Resource Managers using Artificial Intelligence Concepts", Ecological Modelling, Vol. 46, pp. 73-89.

DeFranti, T.A., M.D. Brown, B.H. McCormick (1989), "Visualisation: Expanding the Scientific and Engineering Research Opportunities", Computer, Vol. 22, No. 6,pp. 27-38.

Densham, P.J. and M. Goodchild (1989), "Spatial decision Support Systems: A research agenda", Proceedings, GIS/LIS '89, Orlando, Florida, November 26-30 Vol. 2, pp. 707.

Dietterich, T.G. and R.S. Michalski (1983), "A Comparative review of selected data methods for learning from examples", Machine Learning: An Artificial Intelligence Approach, Springer-Verlag, pp. 41-81.

Djokic, D. and D.R. Maidment (1993), Creating an Expert Geographic Information System: the Arc/Info-Nexpert Object interface, Department of Civil Engineering, The University of Texas at Austin, Austin TX 78712 (unpublished draft report).

Dowie, R. (1986), "Relationships between soil salinity, vegetative cover and tree species performance on saline seeps, and the use of such information in providing on farm tree planting guide-lines" in a Final Year Natural Resources Project, Roseworthy Agricultural College, S.A. (unpublished).

Dubois, D., and H. Prade (1988), Possibility Theory, Plenum Press.

Dueker, K.J. (1979), "Land Resource Information Systems: A Review of Fifteen Years Experience", Geo-Processing, No. 1, pp. 105-128.

Dutton, G. (1984), "Geodesic Modelling of Planetary Relief", Cartographica, V.21, pp. 188-207.

Dyson, P.R. (1990), "Dryland Salinisation in South Eastern Australia. An Overview of Salinity Processes", in Management of Soil Salinity in South East Australia, Australian Society of Soil Science Inc.

Dyson, P.R. (1983), "Dryland Salting and Groundwater Discharge in the Victorian Uplands", Proceedings Royal Society of Victoria, Vol. 95, No. 3, pp. 113-116.

Eklund, P.W. (1993), "How many datapoints can we generalise from? Prospects for artificial intelligence in GIS", Proceedings of the North Australian Conference on Remote Sensing and GIS Forum, Darwin.

Eklund, P.W. and A. Salim (1993), "An experiment with automated acquisition of classification heuristics from G.I.S", Proceedings Land Information Management Geographic Information Systems, E.G. Masters and J.R.Pollard (Eds.), Vol. 1, Sydney, pp. 83-94.

Eklund, P.W., S.D. Kirkby and A.Salim (1994), "A Framework for Incremental Knowledge Base Update from Additional Data Coverages", Proceedings 7th Australian Remote Sensing Conference, Melbourne, Vol. 1.,pp. 367-374.

Engelmore, R. and T. Morgan (1989), Blackboard Systems, Addison-Wesley Publishers Ltd., Wokingham, England.

Environmental Systems Research Institute (ESRI) (1991), Arc/Info User's Guide, Version 6.0, Redlands CA.

Estes, J. and R.K. Holz (1985), "Remote Sensing: Coming of Age in the 80's", in R.K. Holz (ed.), The Surveillant Science, Remote Sensing of the Environment, J.Wiley and Sons, New York.

Fayyad, U.M. and K.B. Irani (1990), "What should be Minimised in a decision tree", AAAI, pp. 749-754.

Fedra, K. (1993), "Models, GIS and Expert Systems in Environmental Management", in AURISA 93, The Annual Conference of the Australasian Urban and Regional Information Systems Association Inc., Adelaide, South Australia, pp 223-240.

Feigenbaum, E.A. and P. McCorduck (1984), The Fifth Generation, London, Pan Books

Fenneman, M.N. (1916), "Physiographic Divisions of the United States", Annals Association American Geographers, Vol. 6, pp. 19-98.

Freeze, R.A. and J.A. Cherry (1979), Groundwater, Prentice-Hall Inc, 604 pp.

Gammack, J.G. and R.M. Young (1985), "Psychological Techniques for Eliciting Expert Knowledge", in Research and Development in Expert Systems, (ed.) M.A. Bramer, Cambridge, Cambridge University Press, pp. 105-112.

Gibbons, F.R., J.N. Rowan and R.G. Downes (1968),"The Role of Humans in Land Evaluation", in G.A. Stewart (ed), Land Evaluation, MacMillan, Sydney.

Goodall, A. (1985), The Guide to Expert Systems, Oxford, Learned Information.

Goodchild, M.F. (1991), "Issues of Quality and Uncertainty", Advances in Cartography, J.C. Muller (ed.), Elsevier, New York.

Goodchild, M.F. (1993), "Future Directions: Geographic Information and Analysis", in AURISA 93, The Annual Conference of the Australasian Urban and Regional Information Systems Association Inc., Adelaide, South Australia, pp 458-465.

Goodenough, D.G., M. Goldberg, G. Plunkett and J. Zelek (1987), "An Expert System for Remote Sensing", IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-25, No. 3, pp. 349-359.

Goodenough, D.G., J.C. Deguise and M.A. Robson (1990), "Multiple Expert Systems for Using Digital Terrain Models", IGARSS '90 10th Annual International Geoscience and Remote Sensing Symposium, Vol. 1, Washington.

Grabau, W.E. and W.N. Rushing (1968), "A Computer Compatible System for Quantitatively Describing the Physiognomy of Vegetation Assemblages", in G.A. Stewart (ed), Land Evaluation, MacMillan, Sydney.

Graetz, R.D. and M.R. Gentle (1982), "The relationships between reflectance in the Landsat wavebands and the composition of an Australian semi-arid shrub rangeland", Photogrammetric Engineering and Remote Sensing, Vol. 48, No. 11, pp. 1721-30.

Greenwood, E.A.N. and J. Beresford (1979), "Evaporation from Vegetation in Landscapes developing secondary salinity using the Ventilated-Chamber technique. 1. Comparative transpiration from juvenile Eucalypts above saline groundwater seeps", Journal of Hydrology, Vol. 42., pp. 369-82.

Greenwood, E.A.N., L.Klein, J.D. Beresford and G.D. Watson (1985), "Differences in annual evaporation between grazed pasture and Eucalyptus species in plantations on a saline farm catchment", Journal of Hydrology, No. 78, pp. 261-78.

Guillet, D. (1989), "A Knowledge-based Systems Model of Native Soil Management", Anthropological Quarterly, Vol. 62, No. 2, pp. 59-67.

Hadipriono, F.C., J.G. Lyon, T.W.H. Li, and D.P. Argialas (1990), "The Development of a Knowledge-based Expert System for Analysis of Drainage Patterns", Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 6, pp. 905-909.

Hamblin A.P. (1981), "Filter-Paper method for routine measurement of field water potential", Journal of Hydrology, No.53, pp. 355-360.

Harlequin Ltd (1992), Knowledgeworks and Lispworks Users Guide, Edition 2, Cambridge, England.

Harris, R. (1987), Satellite Remote Sensing: An Introduction, Routledge and Kegan Paul Ltd., London.

Hart, P.E. (1968), "The condensed nearest neighbour rule", IEEE Transactions on Information Theory, V.14, pp.515-516.

Hartley, R.E.R. (1984), "Basic Principles of Salinisation", in W.E. Matheson (ed.), The Recognition and Treatment of Dryland Salinity, South Australian Department of Agriculture, Technical Report No. 51, pp. 6-8.

Hartshorne, R (1939), The Nature of Geography, Association of American Geographers, Lancaster.

Hatton, T. and W. Dawes (1991), The Impact of Tree Planting in the Murray-Darling basin: the use of the Topog-IRM Hydroecological model in Targeting Tree Planting Sites in Catchments, CSIRO Technical Memorandum, 91/14, Division of Water Resources

Henschke, C.J., D.G.McCarthy, S.B. Richardson, T.D. Evans (1993), Dryland Salinity Catchment Investigation: Jamestown Catchment Mid North South Australia, Primary Industries Department, South Australia, Technical Report No. 193.

Herbertson, A.J. (1905), "The Major Natural Regions: An Essay in Systematic Geography", Geographical Journal, Vol. 20, pp. 300-312.

Hertz, J.A., R.G. Palmer and A.S. Krogh (1991), Introduction to the Theory of Neural Computation, CA, Addison-Wesley.

Heuvelink, G.B.M. and P.A. Burrough (1989), "Propogation of errors in spatial modelling with GIS", International Journal of Geographical Information Systems, Vol. 3, No. 4, pp. 303-322.

Hick, P.T. and W.G.R. Russell (1988), Remote Sensing of Agricultural Salinity, CSIRO Report: Presented to the Department of Primary Industries, Canberra, National Soil Conservation Programme.

Hill, S.R. (1988), Dryland Salinity in North Central Victoria: A case study in Detection, Management and Prediction, Monash Publications in Geography No. 33, Melbourne, 1988.

Hill, S.R. (1990), Land Classification using remote sensing and G.I.S.: Salinity a Case Study, Keith Turnbull Research Institute, Department of Conservation, Forests and Lands (unpublished).

Hillel, D. (1980), Applications of Soil Physics, Academic Press, New York.

Honey, F.R., P.T. Hick, R.A. Steckis and L.M. McCullock (1984), "Mapping Dryland Salinity Using Remotely Sensed Data", Third Australian Remote Sensing Conference, Queensland, pp. 44-47.

Hookey, G.R. and I.C. Loh (1985), Groundwater Simulation of the Effect of Catchment Clearing and Partial Reforestation at Maringee Farms, Report No. WRB122, Water Resources Branch, Public Works Department, W.A.

Hookey, G.R. (1987), "Prediction of Delays in Groundwater response to Catchment Clearing", Journal of Hydrology, Vol. 94 No. 1, pp. 181-198.

Hughes, K.K. (1985), "Identification of Lands Susceptible to Seepage Salting", Proceedings of the 4th Australian Soil Conservation Conference, Maroochydore, Qld, 1985.

Hunter, G. and M.F. Goodchild (1993), "Communication of Uncertainty in Spatial Databases", AURISA 93, The Annual Conference of the Australasian Urban and Regional Information Systems, Adelaide, pp. 522-536.

Huntley, D. (1978) "On the detection of shallow aquifers using thermal infrared imagery", Water Resources Research, Vol. 14, No. 6, pp. 1075-1083.

Hushon, J.M. (1990), "Overview of Environmental Expert Systems", in J.M. Hushon, Expert Systems for Environmental Applications, American Chemical Society, Washington D.C.

Hvorslev, M.J. (1951) Time Lag and Soil Permeability in Groundwater Observations, US Army Corps. Eng. Waterways Expt. Station Bull. 36

Jackson, T.J., T.J. Schmugge and P.E. O'Neill (1983), "Remote sensing of soil moisture from an aircraft platform using passive microwave sensors", Hydrological Applications of Remote Sensing and Remote Data Transmission (Proceedings of the Hamburg Symposium), B.E. Goodison (ed.), IAHS Publication, No. 145. pp. 529-539.

Jackson, T.J. and P.E. O'Neill (1987), "Salinity Effects on the microwave emission of soils", IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-25, No. 2, pp. 214-20.

Janssen, L.L.F. and H. Middelkoop (1991), Knowledge based image classification: the use of knowledge about crop rotations in a Thematic Mapper image classification procedure, Winand Staring Centre, Report 40, Wageningen, The Netherlands.

Jenkin, J.J. (1979), Dryland Salting in Victoria, Water Research Foundation and Soil Conservation Authority, Victoria.

Jensen, J.R. (1986) Introductory Digital Image Processing, Englewood Cliffs, NJ, Prentice Hall.

Jolly, I.D. (1988), Non-Irrigated Salinity in South Australia: Options for Treatment, South Australian Department of Agriculture, Technical Report No. 124.

Jolly, I.D. (1989), Investigation into the Potential for Increased Stream Salinisation in the Darling Basin, Report No. 10A, Centre for Research in Groundwater Processes, CSIRO Division of Water Resources, Adelaide, Australia.

Jupp, D.L.B., T.R. McVicar, J.Walker, J.D. Kalma, B.A. Harrison and P.G. Hutton (1990a), "Monitoring land degradation and salinisation in the Murray-Darling Basin of Australia using remote sensing and data integration", Proceedings 5th Australasian Remote Sensing Conference, Perth, W.A.

Jupp, D.L.B., J.Walker, J.D. Kalma, T.McVicar (1990b), "Using thermal remote sensing to monitor land degradation and salinisation in the Murray-Darling basin of Australia", 23rd International Symposium on Remote Sensing of the Environment, Bangkok, Thailand (draft copy of submission).

Jupp, D.L.B., J. Walker, J. Kalma, R.Smith (1990c), "Remote sensing of change in components of the regional water balance of the Murray-Darling basin using satellite imaged and spatially registered environmental data", in Australian Journal of Soil Resources, Vol. 28, pp. 409-15.

Jupp, D.L.B., T.R.McVicar, A.A. Held, T.Guoliang, Q.Yi, L. Fuquin (1994), "Monitoring soil moisture at a regional scale using AVHRR data", 7th Australasian Remote Sensing Conference, Melbourne, Australia, Vol. 2, pp. 961-969.

Kemp, K.K. (1993), Environmental Modeling with GIS: A Strategy for Dealing with Spatial Continuity, Technical Report 93-3, NCGIA, University of California, Santa Barbara.

Kibler, D., D.W. Aha and M. Albert (1989), "Instance-based prediction of real valued attributes", Computational Intelligence, V.5, pp.51-57.

Kim, H. and P.H. Swain (1989), "Multisource data analysis in Remote Sensing and Geographic Information Systems based on Schafer's Theory of Evidence", IGARSS '89 12th Canadian Symposium on Remote Sensing, Vol. 2, Vancouver.

Kimbrough, K. and O.LaMott (1989), Common Lisp User Interface Environment, Texas Instruments Inc.

Kirkby, S.D., M.E. Longmore and M.A. Maslen (1992), "The Integrated Use of SPOT 1 Imagery and GIS to Quantify spatial changes in Saline Discharge Areas", Proceedings of the 6th Australasian Remote Sensing Conference, Vol. 2., Wellington, pp. 384-390.

Kirkby, S.D., and F.Kurzel (1993a), "Interfacing Expert Systems, GIS and Remotely Sensed Information to Identify and Predict Dryland Salinisation", AURISA 93, The Annual Conference of the Australasian Urban and Regional Information Systems, Adelaide, pp. 298-305.

Kirkby, S.D. (1993b), "Monitoring saline discharge areas: Farrell Flat a case study", Australian Geographical Studies, Vol. 31, No. 1, pp. 62-69.

Kirkby S.D. and F. Kurzel (1993c), "Managing Dryland Salinisation with an Integrated GIS/ES system", Proceedings Land Information Management Geographic Information Systems, E.G. Masters and J.R.Pollard (Eds.), Vol. 1, Sydney, pp. 143-156.

Kirkby, S.D., M.E. Longmore and M.A. Maslen (1994a), "Monitoring Salinisation using Remote Sensing Techniques and Geographic Information Systems", in Global Environmental Change, Monitoring, Modelling and Forecasting, R.B. Singh (ed.), Sage Publications, India (in press).

Kirkby, S.D., E.J. Bamford and M.E. Longmore (1994b), "Land Classification A New Understanding", Australian Geographical Studies (in press).

Kolarkar, A.S., R.P. Dhir and N.Singh (1980), "Salt Affected lands in South Eastern Arid Rajasthan - their distribution and genesis as studied from aerial photographs", Proceedings Seminar on Application of Photo-interpretation and Remote Sensing techniques for Nayural Resources Survey and Environmental Analysis, Dehra Dun.

Laut, P., P.C. Heyligers, G. Keig, E. Loffler, C. Margules, and R.M. Scott (1982), Environments of South Australia Handbook, CSIRO Division of Land Use Research, Canberra.

Laut, P. and R. Davis (1989), "Cross-Country Trafficability Ratings for Cape York Peninsula: A Method Based on Simple Landscape Data and Expert Systems", in D. Ball and R. Babbage (1989), Geographic Information Systems: Defence Applications, Pergamon Press, Sydney.

Luard, E.J. and M.E. El-Lakany (1984), "Effects on Casuarina and Allocasuarina species of increasing sodium chloride concentrations in solution culture", in Australian Journal of Plant Physiology, No. 11, pp. 471-481.

Luckman, P.G., M.R. Jessen and R.G. Gibb (1994a), "Use of Expert Systems and GIS in Land Evaluation", New Zealand Geographer Special Issue on GIS (in press).

Luckman, P.G., R.G. Gibb and M.R. Jessen (1994b), "Integration of Expert Systems and GIS for Land Classification and Assessment", First National GIS Conference (in press).

Luger, G.F. and W.A. Stubblefield (1989), Artificial Intelligence and the Design of Expert Systems, Benjamin/Cummings Publishing Co., Redwood City Ca.

Lyle, J. and F.P. Stutz (1983), "Computerised Land Use Suitability Mapping", The Cartographic Journal, Vol. 20, No. 1, pp. 39-49.

Mabbutt, J.A. (1968), "Review of Concepts of Land Classification", in G.A. Stewart (1968), Land Evaluation, MacMillan, Sydney.

Mackenzie, M., G. Bellamy, D. Fraser and G.Ellis (1990), "Mapping dryland salinity recharge areas using EM conductivity measurements and airborne scanner imagery", 5th Australian Soil Conservation Conference: Remote Sensing and G.I.S. workshop, Perth, W.A. pp. 35-52.

Macumber, P.G. (1969), "Interrelationships between Physiography, Hydrology, Sedimentation and Salinisation of the Loddon River Plains, Australia", Journal of Hydrology, Vol. 7, pp. 39-57.

Maidment, D.R. and D. Djokic (1990), Creating an expert GIS: the Arc-Nexpert Interface, Department of Civil Engineering, University of Texas, Austin Texas,78712 (unpublished).

Marble, D.F. and D.J. Peuquet (1983), "Geographic Information Systems and Remote Sensing", in R.N. Colwell (1983), Manual of Remote Sensing Vol 1: Theory, Instruments and Techniques, Sheridan Press, New York.

Marshall, J.K. (1976), "Salt Tolerance of South Western Australian Trees", in Australian Salinity Newsletter, No. 4.

Matheson, W.E. (1984), The Recognition and Treatment of Dryland Salinity, South Australian Department of Agriculture, Technical Report No. 51.

Mazikov, V.M. (1978), "Salinisation of the soils of the new zone of irrigation of the Golodnaya Steppe, according to aerial surveys", Soviet Soil Science, Vol. 10, pp. 604.

McCarthy, D. (1990), Proceedings of a Technical Forum on Dryland Salinity in South Australia, South Australia Department of Agriculture, Technical Report No. 163.

McFadyen, D. and A. Kanabar (1991), Structured Query Language, W.McBrown Publishers, Dubuque, USA.

McHarg, I. (1969), Design with Nature, Natural History Press, New York.

McKeown, D.M. (1987), "The Role of Artificial Intelligence in the Integration of Remotely Sensed Data with Geographic Information Systems", IEEE Transactions on Geoscience and Remote Sensing, Vol GE-25, No. 3, pp. 330-348.

McNeill, J.D. (1980), Electromagnetic terrain conductivity measurement at low induction numbers, Technical Note TN-6 Geonics Ltd. Ontario, Canada.

Michalski, R.S. and R.L. Chilausky (1980), "Learning by being told and learning from examples: An experimental comparison of the two methods of knowledge acquisition in the context of developing an expert system for soybean disease diagnosis", International Journal of Policy Analysis and Information Systems, Vol. 4, No. 2, pp. 125-161.

Miller C. and R.A. LaFlamme (1958), "The Digital Terrain Model - Theory and Applications", Photogrammetric Engineering and Remote Sensing, Vol 24., No. 3, pp. 433-442.

Moffat, I. (1990), "The Potentialities and Problems Associated with Applying Information Technology to Environmental Management", Journal of Environmental Management, Vol. 30, pp. 209-220.

Moon, D.A. (1989), "The Common Lisp Object-Oriented Programming Language Standard", in Object-Oriented Concepts, Databases and Applications, W.Kim and F. Lochovsky (Eds.), ACM Press.

Morehouse, S. (1989), "The Architecture of Arc/Info", Autocarto, No. 9, pp. 266-277.

Morris, J.D. and L.A.J. Thomson (1983), "The Role of Trees in Dryland Salinity Control", Proceedings Royal Society of Victoria, Vol. 95, No. 3, pp. 123-131.

Newkirk, R.T. and F. Wang (1990), "A Common Knowledge Database for Remote Sensing and Geographic Information in a Change Detection Expert System", Environment and Planning B: Planning and Design, Vol. 17, pp. 395-404.

Northcote, K.H., and J.K.M. Skene (1972), Australian Soils with Saline and Sodic Properties, CSIRO Soil Publication. No. 27.

Nulsen, R.A. (1978), Saltland Hydrology; the altered water balance after clearing", in Salinity Problems of Non-Irrigated Lands in Western Australia, Department of Agriculture, W.A., Seminar and Workshop Papers April 11, 12.

Nulsen, R.A. (1980), "Water Movement through Soil" in Journal of Agriculture Western Australia, 4th Series, V.4, No. 19, pp. 106-107.

Nulsen, R.A. and C.J. Henschke, (1981), "Groundwater Systems Associated with Secondary Salinity in Western Australia", in Land and Stream Salinity, J.W. Holmes and T. Talsma (1981) (ed.), Elsevier Publishing Co., Amsterdam.

Oracle Ltd (1990), Oracle Users Manuals, Version 6.0, Oracle Corporation, USA.

Pankova, Y.E.I., V.M. Mazikov, V.A. Isayev and I.A. Yamnova (1978), "Use of Aerial Photographs to Describe the Saline Soils of Non-Irrigated Areas in the Sierozem Zone", Soviet Soil Science, Vol. 10 pp. 235.

Peck, A.J. and D.H. Hurle (1973), "Chloride Balance of some Farmed and Forrested Catchments in South-Western Australia", in Water Resources Research, V. 9, pp. 648-657.

Peck, A.J. (1977), "Development and Reclamation of Secondary Salinity", in Soil Factors and Crop Production in a Semi-Arid Environment, J.S. Russell and E.L. Greacon (Eds), University of Queensland Press, Brisbane.

Peck, A.J. (1978), "Salinisation of Non-Irrigated Soils and Associated Streams: A Review", Australian Journal of Soil Resources, Vol. 16, pp. 157-68.

Peck, A.J., (1983), "Flow in Aquifers" in Groundwater Research, A Report from the CSIRO Division of Groundwater Research.

Perkins, A.J (1927), "Development of Salt Areas on Northern Yorke Peninsula Farms", Journal of the Department of Agriculture of South Australia, Vol. 30, pp. 1194-1195.

Perkins, A.J. (1930), "Treatment of Salt Patches of Land", Journal of the Department of Agriculture of South Australia, Vol. 34, pp. 216-218.

Peuker, T.K. (1977), "Data structures for digital terrain models: discussion and comparison", In Proceedings Advanced Study Symposium on Topological Data Structures for Geographic Information Systems, Harvard University, Cambridge, Massachusetts.

Peuquet, D.J. (1979), "Raster Processing: an Alternative Approach to Automated Cartographic Data Handling", The American Cartographer, Vol. 6, pp. 129-139.

Policy Development Planning (PDP), (1992), Dryland Salinity R&D. A National Review, PDP Australia Pty Ltd.

Poljakoff-Mayber, A. (1975), "Morphological and Anatomical Changes in Plants as a Response to Salinity Stress", in A. Poljakoff-Mayber and J, Gale (Eds), Plants in Saline Environments, Springer-Verlag, Berlin.

Preiss, W.V (1983), The Adelaide Geosyncline, Government Printer, Adelaide.

Price, J.C. (1980), "The potential of remotely sensed thermal infrared data to infer surface soil moisture and evaporation", Water Resources Research, Vol. 16, No. 4, pp. 787-795.

Quinlan,J.R. (1979), "Discovering rules by induction from large collection of examples", Expert Systems in the Microelectronic Age, Edinburgh, University Press.

Quinlan, J.R. (1983), "Learning efficient classification procedures and their application to a chess game", Machine Learning: An Artificial Intelligence Approach, pp. 463-482, Springer-Verlag

Quinlan, J.R. (1986), "Induction of decision trees", Machine Learning, Vol. 1, No. 1, pp. 81-106.

Quinlan, J.R. (1987), "Simplifying decision trees." International Journal of Man-Machine, Vol. 27, pp. 221-234.

Quinlan, J.R., and R.L. Rivest (1989), "Inferring decision trees using the minimum description length principle", Information and Computation, Vol. 80, No. 3, pp. 227-248.

Quinlan, J.R. (1990), "Decision trees and decisionmaking", IEEE Transactions on Systems, Man and Cybernetics, Vol. 20, No. 2, pp. 339-346.

Quinlan, J.R. (1993), C4.5: Programs for Machine Learning. San Mateo, CA: Morgen Kaufmann Series.

Rao, B.R.M. and L. Venkataratnam (1991), "Monitoring of salt affected soils - A case study using aerial photographs, Salyut-7 space photographs and Landsat TM", Geocarto International, Vol. 6, No. 1, pp. 5-11.

Report on the Interdepartmental Committee on Vegetation Clearance (1976), Vegetation Clearance in South Australia, Department of Environment and Planning, Adelaide.

Reutov, E.A. (1989), "Microwave radiometry of saline soils", IGARSS '89, 12th Canadian Symposium on Remote Sensing, Vol. 5, pp. 2747.

Richards, J.A. (1986), Remote Sensing Digital Image Analysis, An Introduction, Springer-Verlag, Berlin.

Richardson, A.J., A.H. Gebermann, H.W. Gausman and J.A. Cueller (1976), "Detection of Saline Soils with Skylab Multispectral Scanner Data", Photogrammetric Engineering and Remote Sensing, Vol. 42, No. 5., pp. 679-84.

Robbins, C.W., W.S. Meyer, S.A. Prathapar, R.J.G. White, (1991), Understanding Salt and Sodium in Soils, Irrigation Water and Shallow Groundwaters, Water Resources Series No. 4, CSIRO, Div of Water Resources, Canberra.

Robinson, V.B. and A.U. Frank (1987), "Expert Systems for Geographic Information Systems", Photogrammetric Engineering and Remote Sensing, Vol. 53, No. 10, pp. 1435-1441.

Romaniuk, S.G. (1993), Multi-Pass Instance Based Learning, Technical Report TRH3/93, Department of Information System and Computer Science, National University of Singapore, Singapore.

Rumelhart, D.E., G.E. Hinton and R.J. Williams (1986a), "Learning representations by back-propagating errors", Nature, V.323, pp.533-536.

Rumelhart, D.E., G.E. Hinton and R.J. Williams (1986b), "Learning representations by back-propagating errors", Parallel Distributed Processing, V.1, Chap8.

Russell, E.W. (1973), Soil Conditions and Plant Growth, Longman, Sydney.

Sabins, F.F. (1978), Remote Sensing, Principles and Interpretation, San Francisco: Freeman

Salim, A.S. (1993), Experiments with Automated Acquisition of Classification Heuristics from Geographic Information Systems, Masters Project Report, Department of Computer Science, University of Adelaide, South Australia.

Salim, A.S. (1994), Experiments with Automated Acquisition of Classification Heuristics from Geographic Information Systems, Technical Report TR94-02, Department of Computer Science, University of Adelaide.

Sands, R. and J.A. Rishworth (1978), "Salt tolerance and avoidance in Eucalyptus Camaldulensis, from different seed sources", Australian Salinity Newsletter, No. 6.

Schofield, N.J. (1988), Methods for Estimating reforestation Area and Distribution for Salinity Control, Report No. WS31, Water Authority of W.A.

Schofield, N.J., I.C. Loh, P.R. Scott, J.R. Bartle, P. Ritson, R.W. Bell, H. Borg, B.Anson and R. Moore (1989), Vegetation Strategies to Reduce Stream Salinities of Water Resource Catchments in South-West Western Australia, Report No. WS33. Water Authority of Western Australia, Water Resources Directorate.

Scofield, C.S., (1942), Pecos River Joint Investigation: Reports of Participating Agencies, U.S. National Resources Planning Board, Washington DC, pp 263-334.

Shafer, G. (1976), A Mathematical Theory of Evidence, Princeton University Press.

Sharma, M.L., R.J.W. Barron, A.B. Craig (1991), Land Use Effects on Groundwater Recharge to an Unconfined Aquifer, CSIRO Division of Water Resources, Divisional Report 91/1.

Sharma, R.C. and G.P. Bhargava (1988), "Landsat imagery for mapping saline soils and wetlands in North-west India", International Journal of Remote Sensing, Vol 9., No. 1, pp. 39-44.

Shortliffe, E.H. (1976), Computer-based Medical Consultations: MYCIN, American Elsevier, New York.

Simonett, D.S. (1983), "The Development and Principles of Remote Sensing", in Manual of Remote Sensing: Volume1, Theory Instruments and Techniques, R.N. Colwell (ed.), Seridan Press.

Skidmore, A.K. (1989), "An expert system classifies Eucalypt forest types using thematic mapper data and a digital terrain model", Photogrammetric Engineering and Remote Sensing, Vol. 55, No. 10, pp. 1449-1464.

Skidmore, A.K. (1990), "Integrating remote sensing and geographical information systems using an expert system", Proceedings 5th Australasian Remote Sensing Conference, Perth, W.A., pp. 840-47.

Slessar G.C., N.J. Murray and R.E. Hammer (1991), "Potential Application of Digital Terrain Mapping to Salinity Evaluation in the Darling Range", Unpublished draft report to S.A. Department of Agriculture, Northfield Research Laboratory.

Starfield, A.M. and A.L. Bleloch (1983), "Expert Systems: An Approach to Problems in Ecological Management that are Difficult to Quantify", Journal of Environmental Management, Vol. 16, pp. 261-268.

State Dryland Salinity Committee (SDSC) (January 1990), Technical Strategy to Address Dryland Salinity in South Australia.

Stephens, C.G., R.I. Herriot, R.G. Downes, T. Langford-Smith and A.M. Acock (1945), A Soil, Land Use and Erosion Survey of part of County Victoria, South Australia, CSIRO, Bulletin No. 188.

Stephens, P.R., J.L. Daigle, J. Cihlar (1982), "Use of sequential aerial photographs to detect and monitor soil management changes affecting cropland erosion" in Journal of Soil and Water Conservation, pp. 101-105.

Stewart, G.A. (1968), "Land Evaluation", in G.A. Stewart (ed), Land Evaluation, MacMillan, Sydney.

Szilagyi, A. and M.F. Baumgardner (1991), "Salinity and Spectral Reflectance of Soils", in ACSM-ASPRS Annual Convention Technical Papers, Baltimore 1991, Vol. 3, pp. 430-437.

Thomson, L.A.J. (1981), "Salt Tolerance in Eucalypts", Australian Salinity Newsletter, No. 9.

Townend, R. (1965), The petrology of selected specimens from the Clare-Riverton 1-Mile sheets, Amdel report MP 426-65, South Australian Department of Mines and Energy open file env. 6713 (unpublished).

Tulloh, N.M. (1978), "Salinity tolerance in agricultural species and Eucalypts", Australian Salinity Newsletter, No. 6.

Van Heusden, W. (1983), "Monitoring changes in heathland vegetation using sequential aerial photographs", International Institute for Aerial Survey and Earth Sciences, No. 2, pp. 160-165.

Walker, G.R. (1990), "CSIRO and CGS Research on Factors Causing Dryland Salinity and Affecting Remediation", in D.McCarthy (ed.), Proceedings of a Technical Forum on Dryland Salinity in South Australia, Department of Agriculture, Technical Report 163.

Webster, A. (1983), "Intercepting Channel Seepage in the Wimmera-Mallee: the use of Trees and Tall Wheat grass", in Water Talk 1983

Weichselberger K., and S. Pohlman (1990), A Methodology for Uncertainty in Knowledge-Based Systems, Springer-Verlag, Munich.

Whitehead, S. (1978), Investigation of Mintaro Slate, Amdel progress report 1, South Australian Department of Mines and Energy open file env, 3333 (unpublished).

Williams, B.G. and S. Arunin (1990), Inferring recharge/dischage areas from multifrequency electromagnetic induction measurements, CSIRO Technical Memorandum 90/11, Division of Water Resources.

Williams, R., P.Crowther and K.Michael (1994), "Interpretation of remotely sensed sea-ice images for climatological studies using an expert system" in 7th Australasian Remote Sensing Conference, Melbourne, Vol. 1, pp. 433-39.

Williamson, D.R. (1978), "Water is the Villain in Dryland Salinity", in Victoria's Resources, December, pp. 9-14.

Williamson, D.R. and E. Bettenay (1980), "Agricultural land Use and its Effect on Catchment Output of Salt and Water - Evidence from Southern Australia" in Progress in Water Technology, Vol. 2., No. 6, Pergammon, London.

Williamson, D.R. (1990a), Salinity - An Old Environmental Problem, CSIRO Technical Memorandum, 90/7, Devision of Water Resources.

Williamson, D.R. (1990b), "Principles and Practice of Management of Dryland Salinity - The Western Australian Experience" in Proceedings of a Technical Forum on Dryland Salinity in South Australia, D. McCarthy (ed.), South Australian, Department of Agriculture, Technical Report 163.

Winston, P.H. (1992), Artificial Intelligence, Addison-Wesley Publishing Co., Reading MA.

Wood, W.E., (1924), "Increase of Salt in Soil and Streams following the Destruction of the Native Vegetation", Journal Royal Society of Western Australia, Vol. 10., No. 7, pp 35-47.

Working Party on Dryland Salting in Australia (1982), Salting of Non-Irrigated Land in Australia, Soil Conservation Authority, Victoria.

Wright, J.R. and K.A. Buehler (1993), "Probabilistic Inferencing and Spatial Decision Support Systems", in Expert Systems in Environmental Planning, J.R. Wright, L.L. Wiggins, R.K. Jain, T.J.Kim (Eds.), Springer-Verlag, Berlin.

Young, R.M., (1988), "Role of intermediate representations in knowledge elicitation" in Research and Development in Expert Systems IV, D.S. Moralee (ed.), Cambridge, Cambridge University Press, pp. 287-288.

Zadeh, L.A. (1965), "Fuzzy Sets", Information Control, V.8, pp. 338-353.

Zadeh, L.A. (1983), "Commonsense knowledge representation based on Fuzzy Logic", Computer, V.16, pp. 61-65.

Zhou, Q. (1989), The Integration of Remote Sensing and Geographical Information Systems for Land Resource Management in the Australian Arid Zone, Ph.D Thesis, School of Geography, University of New South Wales, Australia (unpublished).