The development of a knowledge-based system for the preliminary investigation of contaminated land.

Martin, John Charles

The Development of a Knowledge-Based System for the Preliminary Investigation of Contaminated Land

A thesis submitted to the School of Engineering,
University of Durham for the degree of
Doctor of Philosophy

by John Charles Martin

September 2001

DECLARATION

I hereby declare that the work reported in this thesis has not been previously submitted for any degree. All material in this thesis is original except where indicated by reference to other work.

STATEMENT OF COPYRIGHT

TABLE OF CONTENTS

PAGE

PAGE

PAGE

PAGE

LIST OF TABLES

PAGE

PAGE

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr David Toll, for his continuous support and assistance throughout this research. His enthusiasm and knowledge have been essential in bringing this thesis to its final form.

Special thanks are due to Dr Sue White for her guidance and encouragement.

Also I would like to thank all the members of staff of the School of Engineering and in particular Bernard McEleavey and Steve Richardson for their help during the laboratory demonstration. Trevor Nancarrow for his help and advice relating to numerous computer problems and Olive for her supply of chocolate and banana cake.

Thanks to Trevelyan College rowing club, especially Richard, Sam, Andy, Rachel and Alison, whose friendship made those cold early mornings on the river enjoyable.

I would like to pay tribute to Kate Veal and Dyfrig Williams who sadly past away during this research project. *Kate for a wonderful childhood friendship and Dyfig for introducing me to the great sport of rowing !*

Victoria, what can I say, your love has pulled me through the darkest days of this research. Your patience, thoughtfulness and generally being you has been an inspiration. *I look followed to spending a long and happy life with you !*

Finally, I must thank my parents and brother, whose support and love have allowed me to follow my dreams.

Oh, and thanks to Dave Arscott, for has constant banter during the writing of this thesis. *Dave it looks like you finished the patio first !!*

ABSTRACT

Large areas of the UK have witnessed intense industrialisation since the industrial revolution in the latter part of the 18th Century. Increased environmental awareness and pressure to redevelop brown field sites, have resulted in the majority of civil engineering projects undertaken within the UK encountering some form of contamination.

In order to collect the vast amount of information required to assess a potentially contaminated site, a multi-stage site investigation (preliminary investigation, exploratory and detailed investigation) is usually undertaken. The information collected during the investigation allows the three components of the risk assessment process to be identified. These components are the source of contamination, possible pathways for the movement of contaminants and vulnerable targets on and off site.

A prototype knowledge-based system (ATTIC Assessment Tool for The Investigation of Contaminated Land) has been developed to demonstrate that knowledge-based technology can be applied to the preliminary stage of the investigation of contaminated land. ATTIC assesses information collected during the preliminary stage of an investigation (past use, geological map, hydrological maps etc.) and assists with the risk assessment process, with the prediction of potential contaminants, hazards and risk to neighbouring areas.

The system has been developed, using CLIPS software. It consists of four knowledge-bases (source, pathway, target and health and safety knowledge-base), containing 1600 rules.

The knowledge within the knowledge-bases was obtained from two main sources. The initial and main source was the technical literature. Obtaining knowledge from technical literature involved reviewing published material, extracting relevant information and converting information into rules suitable for the knowledge-base system. The second source of knowledge was domain experts via a knowledge elicitation exercise. The

exercise took the form of a questionnaire relating to the rules and parameters within the system.

A Visual Basic® interface was also developed in conjunction with the knowledge-based system, in order to allow data entry to the system. The interface uses a series of forms relating to different components within the risk assessment process.

On completion of compiling the prototype, the system was validated against a number of case studies. The system predicted the likely contaminants with a reasonable match to those observed, even though the input data for the case studies was limited. The assessment of risks to neighbouring target areas was generally in agreement with the case study reports, matching similar risk values and directions.

In addition to the development of the prototype system, a database modelled on the Association of Geotechnical Specialists electronic format for the transfer of ground investigation data was also developed to store preliminary investigation information. The data structures were implemented using Microsoft Access relational database management system software. This allowed the database to be developed within a Microsoft Windows environment.

CHAPTER 1

INTRODUCTION

1.0 General Introduction

Large areas of the UK have witnessed intense industrialisation since the industrial revolution in the latter part of the 18th Century. This type of land use has resulted in a large percentage of this land becoming contaminated. This means that the majority of civil engineering projects undertaken within the UK are likely to encounter some form of contamination. This usually results in an increase in the development costs of a site and an extended period of design and site works. It is therefore essential that the correct information required for the development of such a site is collected and used in the most cost effective manner.

In order to collect the vast amount of information required to fully assess a potentially contaminated site, a multi-stage site investigation (preliminary investigation, exploratory and detailed investigation) is usually undertaken. The information collected during the investigation allows the three components of the risk assessment process to be identified. These components are the source of contamination, possible pathways for the movement of contaminants and vulnerable targets on and off site.

The scope of this research work is to present a methodology that allows knowledge-based system technology to be applied to the preliminary stage of the site investigation process. The objective of the methodology is to use the data collected within the preliminary stage of the investigation to assist with identification of the three components within the risk assessment process and to produce a risk assessment for the area under investigation.

Such technology can assist with the risk assessment process, as it allows domain knowledge to be structured and represented in a manner necessary for the prediction of the components required within the risk assessment process and for assessment of their overall risk.

Domain knowledge is that knowledge which concerns a particular subject area. In the case of the investigation of contaminated land it involves a number of subject areas, ranging from chemistry to geology. Therefore, the development of the methodology for the system used in this study involved identifying and collecting the required domain knowledge, as well as representing the knowledge in a format that could be implemented within a series of rules.

A prototype of the system has been implemented using CLIPS & Visual Basic© software on a personal computer.

Within the scope of this research a method of storing preliminary investigation data was also required. This led to the design and implementation of a relational database. The database was modelled on the Association of Geotechnical Specialists (AGS) electronic transfer format. However, the AGS format is concerned with the storage and transfer of geotechnical data from ground investigations, and does not include preliminary investigation data. Therefore the development of a format for the transfer of preliminary investigation to other software packages was seen as a major contribution to the area of data transfer, as such a format does not currently exist.

1.1 Overview of the Thesis

A current overview of contaminated land is presented in Chapter 2. The definition, history and amount of contaminated land within the UK are briefly described. Then, the types and sources of contamination are reviewed. This is followed by a full description of the structure, aims and procedures of contaminated land investigation.

An introduction to the concept of database and knowledge-based system technology is outlined in Chapter 3. The chapter starts with a review of the various data models and highlights the benefits of database storage over other methods. This is followed by a review of geotechnical databases describing the development from early systems through to current tends of using a standard format (for example, AGS). This leads on to a discussion on software selection for the purposed database system.

The second half of the chapter reviews knowledge-based system technology, initially detailing the components of such systems. This is followed by a description of the various development tools available for knowledge-based systems, and a discussion on software selection for the purposed knowledge-based system. This leads on to a discussion of knowledge acquisition, highlighting the sources and methods for collecting suitable knowledge. The chapter concludes with a review of the use of information technology within the subject area of contaminated land investigation and details potential areas for development.

Chapter 4 is concerned with the storage of preliminary site investigation data. The chapter starts with a description of the purpose of the database and design considerations. This is followed with a discussion of the implementation of the database. Then, the data structure derived from the technical literature is described in full. The chapter concludes with an outline of the design of the user interface.

The development of knowledge-based system is detailed in Chapter 5. Sources of knowledge for the system and methods of collecting and analysing the knowledge are described at the start of the chapter in the knowledge acquisition section. The chapter continues with the representation of the knowledge required and an overview of the system. This includes a description of rules for representing source, pathway and target information, as well as rules for health and safety issues. This is followed by a description of the implementation of the knowledge using CLIPS software. Then, the process involved in designing the user interface is discussed.

Chapter 6 discusses the evaluation of the system, highlighting the methods used for verification and validation of the system, and detailing how the system performed against four separate case studies. This is followed with a general discussion of the work presented in this thesis. The main features of the system are briefly reviewed, and the knowledge acquisition and evaluation processes are discussed.

Finally, the conclusions reached from the development of the methodology, prototype system and the database are presented in Chapter 7. An overview of possible future work is also identified within the chapter.

CHAPTER 2

CONTAMINATED LAND OVERVIEW

2.0 Introduction

The subject of contaminated land has always been considered a controversial issue, due to financial, social and political interests. Petts et al (1997) suggest that the lack of definitional clarity throughout the environmental and engineering literature serves to illustrate the diversity of awareness, concerns and priorities in relation to contaminated land as an environmental problem.

However, in the 1998 the government set up a contaminated land research programme, investing around £1.3 million per annum. One of the outputs of the programme was the publication of a series of guidelines, which were aimed at rectifying the concerns outlined by Petts et al (1997). The key point from the literature is that it is important to have a clear understanding of all the processes involved, ranging from the type of contaminants likely to be present on a site and their behaviour, to methods of investigation and sampling. This chapter gives a general overview of the key concepts involved within the subject of contaminated land.

2.1 Definition of Contaminated Land

Due to the range of disciplines interested in the study of contaminated land, the simple question "What is contaminated land ?"; has many and varied answers. These varying definitions are related to the different approaches taken by workers in the subject. For example, an area containing high natural levels of elements and compounds may be regarded as contaminated in a general view, but the majority of definitions concern contamination as a result of human activity and so therefore these "natural" areas may not be considered to be contaminated.

The British Standards Institution, in its Draft Code of Practice (DD175, 1988) on the identification and investigation of contaminated land, offered the following definition:

"Land that contains any substance that when present in sufficient concentration or amount presents a hazard. The hazard may:

a) be associated with the present status of the land
b) limit the future use of the land and
c) require the land to be specially treated before use".

In the Department of the Environment's view, no standard definition exists in respect of contaminated land, although the Environment Act 1995 (HMSO, 1995) introduced a legal definition for the first time. Before this definition the Department of the Environment (1990) stated that, "at present it is impossible to define contaminated land unambiguously and that contamination should be regarded as a general concept rather than something capable of exact definition or measurement". The Department of the Environment (1990) did propose a loose definition:

" Land which represents an actual or potential hazard to health or the environment as a result of current or previous use".

The Department of the Environment also adopted the view of the NATO Committee on the challenges of modern society, when defining contaminated land in its sustainable development strategy (Anon, 1994), which defines contaminated land as:

"Land which contains substances which, when present in sufficient quantities or concentrations, are likely to cause harm, directly or indirectly, to man, the environment, or on occasion to other targets."

Smith (1990) criticised the Department of the Environment's definition of contaminated land when giving evidence to the Environment Committee stating that "It is not surprising that the Department would wish to limit the definition of contaminated land because the acceptance of the broader definition would mean that substantial parts of some urban areas would have to be classified as contaminated - as indeed they are."

The Environment Act 1995 legislation defined contaminated land as; "Any land which appears to the local authority in whose area it is situated to be in such a condition, by reason of substances in, on or under the land, that:

- Significant harm is being caused or there is a significant possibility of such harm being caused, or
- Pollution of controlled wastes is being, or is likely to be, caused".

Petts et al (1997) suggested that this should be considered a limited definition for defined legal purposes rather than a general definition of contaminated land. This is because it is only in respect to specific powers of the local authorities and the Environment Agency to enforce remediation of sites.

Another term often used, usually in American literature, to describe land that has been previously used, is "brownfield sites". Syms (1997) states that this term is often regarded as being synonymous with contaminated land but this may not necessarily be the case.

Officially the United Kingdom has not defined the term brownfield site although an attempt was made by Syms (1994) to define such sites as being "any areas of land which have previously been the subject of man-made or non-agricultural use of any type. This would include industrial uses such as chemical works, heavy engineering, ship-building and textile processing together with unfit housing clearance and dock lands, both inland and coastal as well as mineral extraction and those used for landfill purposes".

As regards an official definition, the Department of the Environment define the term "derelict land" as being "land so damaged by industrial or other development that it is incapable of beneficial use without treatment" (Department of the Environment, 1991b). The term treatment refers to ground improvements such as removal of old foundations and the consolidation of fill material, rather than any decontamination works. Therefore in the United Kingdom, "brownfield" is closely related to the term "derelict land".

As mentioned already the majority of definitions refer to contamination as a result of human activity. Syms (1997) states "It is also implicit from the definitions that both contamination and dereliction are seen as having a direct relationship with land use, previous, current or future."

2.2 Historical Overview

Since the start of the Industrial Revolution in the latter part of the 18th Century, large areas of the UK have been intensely industrialised. This type of land use has resulted in a large percentage of this land becoming contaminated.

Although the majority of this contamination has arisen since the Industrial Revolution, some dates back from 2000 years. This includes the sites of copper and lead workings dating back to Roman times. In some places, for example around Shipham in Somerset, there was a continuous history of metal mining and processing for many centuries (Beckett, 1993). It is hard to find evidence of direct impact of these older problems, as their effects tend to become subsumed into the general environmental changes that take place over time in any area.

The impact of more recent industrial activity has caused larger problems. This first became apparent in the 1970's, with the most infamous incident being the discovery of the Love Canal, Niagara Fall, New York (Attewell, 1992). The Love Canal was a 3000 m trench that had been abandoned in 1896. The canal was an attempt by William Love to link the upper and lower sections of the Niagara River, above and below the Falls. In 1942 the land containing the trench was purchased by a chemical company and used for the dumping of chemicals between 1947 and 1953. In total 22,000 tonnes of solid and liquid chemical waste was deposited and buried. On completion of dumping the site was capped and purchased by the Niagara Falls Board of Education for a price of only $1 with the understanding that the site would not be disturbed by building works. This caveat was ignored and several hundred houses and a school were constructed on the site. By 1977 it had become apparent that the chemicals were migrating across the site. The chemicals began to seep into basements and residents started to complain of unexplained illnesses (asthma, urinary tract problems, hyperactivity, eye irritations, skin rashes, intestinal problems, incontinence, strictures, renal failures, central nervous system problems, miscarriages, still-births, birth defects, seizures and learning problems), (Attewell, 1992).

Two hundred and forty eight different chemicals were identified within the canal, including benzene, carbon tetrachloride, vinyl chloride, dichloroethane, hexachlorobenzene, hexachlorocyclohexane, lindine, polychlorinated biphenyl's, trichlorophenols, tetrachlorodibenzene-p-dioxin, toluene and xylene. Altogether, there were 34 neurotoxins, 4 pulmonary toxins, 20 hepatoxins, 15 renal toxins, 34 carcinogens, 18 teratogens, and 30 foetotoxins or embryotoxins (Bridges, 1991).

In the U.K. the recognition of the problem of contaminated land also became apparent around the same time, although in the years following the Aberfan Disaster of 1966, a very significant increase in Government support for land reclamation took place (Beckett, 1993). This support from the Government addressed the issues of dereliction, rather than contamination.

It was not until the 1970's when certain Local Authorities were faced with redeveloping sites that had been contaminated by their former use, that the problems of contamination started to be addressed. In one case, that of the Greater London Council's new-town development at Thamesmead on the site of the former Woolwich Arsenal, development on part of the site had already begun when severe contamination, associated with former munitions manufacture, town gas generation and the dumping of waste materials, was encountered (Lowe, 1984). This led to a large scale clean up of oils and tars, organic compounds and "heavy metals". Many other sites around the country were also found to have similar problems, and this prompted the Department of the Environment and the Department of Health and Social Security (DHSS) to undertake an inquiry into the problem of contaminated land. The result of the inquiry led to the establishment of a government committee to co-ordinate advice from various departments and make this advice available to Local Authorities. The committee known as the Interdepartmental Committee on the

Redevelopment of Contaminated Land (ICRCL), was eventually set up in 1976 and includes representatives from the Department of the Environment, Department of Health and Social Security, The Welsh Office, The Health and Safety Executive and The Ministry of Agriculture, Fisheries and Food, with the Scottish Development Department joining in the 1980's. The main output of the committee was a series of guidance notes covering different types of industrial land, and the publication of the document "Guidance on the Assessment and Redevelopment of Contaminated Land" (ICRCL, 1987).

Throughout the late 1980's the emphasis was placed on identifying the likelihood of contaminants well in advance. Eventually, in 1988, the first British Standard code of practice (DD175, 1988) was published, although only in draft form.

In 1990 the Environmental Protection Act (EPA) reached the statute book. The legislation was widely regarded as one of the most comprehensive pieces of environmental protection legislation ever to have been introduced in the UK (Denner & Harris, 1997).

The EPA 1990 (Department of the Environment, 1992) introduced a new regime for the regulation of industrial facilities, waste management and two sections (S143 & S61) with implications for the management of contaminated land. Section 143 was concerned with setting up a register of contaminated land by local authorities. However, during the public consultation process concerns were raised by property owners and funding institutions on the grounds of the effect on property values.

In March 1993 the government withdrew the proposed contaminated land register. A number of other developments (publication of consultation papers) took place throughout the early to mid 1990's; these are discussed by Denner & Harris (1997).

One of the major developments was the introduction of Section 57 of the Environment Act 1995. The act inserts new sections 78A to 78YC into the Environment Protection Act 1990, placing a duty on local authorities to inspect their region and determine using a new statutory definition of contaminated land, whether land is contaminated.

Nancarrow (1998) outlines the key statutory duties placed on local authorities under section 57, highlighting how local authorities can fulfil their duties in the context of limited resources.

In 2001, the draft British Standard code of practice DD175 (1988), was published in full as BS10175:2001 "Investigation of Potentially Contaminated Land Site - Code of Practice".

As regards a world view to the development of contaminated land, Meyer et al (1995) cover US and European Union policies in detail.

2.3 Amount of Contaminated Land

A comprehensive survey into the extent of contaminated land in the UK has never been undertaken, unlike the majority of other European countries. Since the early 1980's a number of isolated studies have been undertaken, but most of these have attempted to assess certain types of contaminated land, e.g. old landfill sites. Haines & Harris (1987), suggest that, " any estimate of the size of the problem would be highly dependent on the choice of definition" This becomes very apparent, when past surveys are examined.

A survey conducted by the Welsh Office in 1988 recorded 746 sites in total covering some 4,000 hectares (10,000 acres). The survey excluded all sites that were currently in use, and those sites of 0.5 hectare or less, so this obviously gave a limited view of the problem.

From a survey of Derelict Land in 1988, the Department of the Environment extrapolated a possible maximum figure of 27,000 hectares (67,000 acres) of derelict land in England which could be classed as potentially contaminated (Department of the Environment, 1991b). This figure amounts to 65% of the total derelict land, and 0.2% of the total land area. Although the estimated figure is considered to be a maximum, the estimate excluded land which was both in use and contaminated.

A survey by Environmental Resources Ltd (ERL) in 1987 estimated that there were some 50,000 to 100,000 potentially contaminated sites in the U.K. This assessment included land which was currently in use, and was estimated from a pilot survey carried out in Cheshire. The survey pointed out that it is likely that only a small number of these areas would present an immediate threat to public health or the environment.

These figures were backed up by European figures. The Dutch inventory of contaminated land now stands at over 110,000 sites, Germany's at 100,000, and Denmark's and Finland's at 20,000 each. Britain's land use history would indicate that there are potentially 50,000 to 100,000 sites which could be expected to be identified if a national register had been collated (ENDS Report 193, 1991).

2.4 Types of Contaminants

A range of heavy industrial activities have developed since the Industrial Revolution. This has led to a diversity in the materials and processes used, and this has inherently produced a wide suite of contaminants. These contaminants may be present in three forms; gases, liquids and solids. Table 2.1 highlights some of these significant contaminants.

Each of the contaminants in Table 2.1 has a varying effect on the redevelopment of a site. When in solution, some of the contaminants, particularly sulphate, may have an aggressive and corrosive action on contact with building materials. Hazards to human and animal health may also occur due to inhalation, ingestion or direct contact with contaminants. For example, Phenols are readily absorbed through the skin on direct contact, causing white and blistered skin, or burning on prolonged contact. Severe exposure may result in digestive disorders and central nervous system (CNS) effects such as fainting (Haines & Harris, 1987).

Metals and their Compounds:- arsenic, barium (soluble), beryllium, boron *, cadmium, chromium, copper *, iron *, lead, manganese *, mercury, molybdenum *, nickel, selenium, thallium, zinc *
Non-metals :- chlorides, sulphides, sulphates, sulphur
Acids :- hydrochloric, phosphoric, sulphuric
Alkalis :- caustic solutions, ammoniacal liquors
Organic Substances :- phenols, cyanides (free and complex), thiocyanates, hydrocarbons, oils, tarry wastes, polychlorinated biphenyls (PCBs), pesticides, herbicides and other chlorinated hydrocarbons, polynuclear aromatic hydrocarbons (PAHs)
Putrescible, biodegradable matter :- domestic waste, food and vegetable residues, paper, packaging
Miscellaneous materials :- asbestos, radioactive substances, glass, rubble, coal wastes, pyrite shales, methane

* In trace amounts, essential to plant and animal health.

Table 2.1 : Significant Contaminants. (Source : Leach & Goodger, 1991)

As regards the metals in Table 2.1, the majority of them do not present a risk to site workers unless in the form of dust. For example, dust containing arsenic can behave as a skin irritant causing inflammation and ulceration (Haines & Harris, 1987). Metals in the soil are usually a greater hazard to subsequent site occupiers as they may receive prolonged exposure to them. The reverse is true for oils and tars, where site workers are likely to have greater contact with the substances, than the later occupants. Problems with electrolytic reactions between metallic contaminants and metallic building materials can also occur. Dissolved salts may also cause similar problems.

The main hazards have been summarised by Crowcroft & Young (1992):

(1) Health Effects

Exposures to contaminants can give rise to health problems through ingestion (e.g. contaminated food or water), inhalation (e.g. toxic gases or dust, including asbestos), or direct skin contact with irritants or harmful chemicals.

(2) Pollution of Water

Drinking water may become contaminated if water pipes pass through soils containing organic compounds (such as phenols) or soluble metal compounds. Contamination of groundwater or surface waters can arise from leaching of rainwater through contaminated land, and breaches in buried tanks and pipework may release contaminated water and liquids into the ground. Several surface water pollution incidents have resulted from the removal of hard standings over contaminated ground and subsequent leaching or washing of contaminants whilst rehabilitation work is in progress.

(3) Phytotoxicity

Substances which are harmful to plant growth are termed phytotoxic. Phytotoxicity is particularly associated with certain metals (copper, nickel, zinc), but other substances including boron, oils, coal tars, phenols and sulphate can also exhibit phytotoxic effects, even when they occur at concentrations which are not toxic to humans. Carbon dioxide is directly phytotoxic and, together with methane, can be indirectly toxic to plants through the depletion of oxygen levels in the soil.

(4) Chemical Attack

Conditions which may lead to chemical attack on buildings and service materials include sulphate attack on concrete and the attack of plastic materials by phenolic compounds. Chloride or extremes of pH may also present corrosion problems, compounds. Chloride or extremes of pH may also present corrosion problems, particularly with metal (such as pile reinforcement) in the ground. Corrosion of underground pipework can lead to the ingress of potentially toxic fluids into water supplies, and sulphate attack on concrete can lead to structural failure in buildings.

(5) Fires

Fires may be propagated underground if material of suitable calorific value is present together with an ignition source and a sufficient supply of oxygen; examples include ground containing coal or coal dust, oil shales, oils and domestic waste. Underground fires are difficult to extinguish and principal hazards include the emission of toxic gases and subsidence into void spaces caused by the fire.

(6) Explosions

Flammable gases, for example methane produced by the degradation of organic material in the absence of oxygen, may form explosive mixtures if they accumulate in confined spaces under buildings, in service ducts, or other enclosed spaces. Any process which generates a spark can trigger an explosion if an explosive gas mixture is present.

(7) Asphyxiation

Where degradation of organic materials occurs, for example at landfill sites, concentrations of carbon dioxide and other gases may exclude oxygen from confined spaces - resulting in an asphyxiating atmosphere.

(8) Odour Problems

These may be associated with landfills, and may also be a problem in land contaminated with organic substances such as coal tars. Although odour problems do not necessarily represent a particularly hazardous situation, the nuisance imposed can often be difficult to control.

(9) Radioactivity

The disposal of low-level radioactive wastes has occurred legally at older landfills, and certain industrial processes, such as the manufacture of luminous dials and gas mantles can give rise to radioactivity in the soil. Exposure for significant periods of time to radioactivity is potentially carcinogenic.

2.5 Sources of Contamination

The source of land contamination is usually the result of human activity, with certain contaminants related to specific industries or activities. It can, however, also be of natural origin, for example emissions of methane and radon or enhanced concentrations of metals in rocks or soils. The industries and activities, which later create contaminated land problems, may be broadly split into four categories; industrial sites, commercial sites, municipal sites and mineral extraction sites.

Industrial sites (heavy industry) such as gas works, iron & steel works and chemical works, may create numerous problems. As well as hosting a suite of contaminants, they are often on "fill" or "made" ground and consequently badly compacted. They often contain massive foundations and underground pipework, tanks and other structures, and there may be abandoned and derelict, unsafe contaminated buildings still standing (Smith, 1985).

The commercial site category includes light industry such as printing works, abattoirs and scrap yards. For example, contaminants on scrap yards can be present in a variety of forms including liquid and solid waste and sludges. It is usually impossible to generalise the distribution of contaminants on the site, since the ground surface will frequently be covered with metal dust, waste oils and other organic contaminants. Common contaminants are lead, copper, zinc, cadmium and nickel, as well as cyanides, sulphates, acids and alkalis (ICRCL 42/80, 1983).

The municipal site category contains a range of uses including residential sites, hospital sites and landfill sites all with varying problems. Residential sites typically consist of densely packed late-Victorian dwellings which have been demolished for modern building works. Often small cottage industries such as dental mechanics would be running in individual houses, this resulting in high levels of mercury being found in the back gardens of these properties.

Urban garden soils are also, in fact, typically contaminated with lead, zinc, mercury, etc. because of aerial deposition, dumping of coal ash and soot, burning of refuse on bonfires, flaking lead paint, and the breakdown of galvanising (Leach & Goodger, 1991).

Hospital sites can present very different problems, as it is common to find areas contaminated by pathogens or pharmacological waste; asbestos is also commonly found lagging old pipe works. Deep basements, old foundations and drying out of clays beneath furnace floors also hamper redevelopment.

Landfill sites can contain a full range of contaminants ranging from heavy metals, such as lead, zinc, copper and nickel depending upon the nature of the waste deposited, to gases and leachates produced from the biological breakdown of the deposited waste. Gases include methane, carbon dioxide and hydrogen sulphide. These sites may also contain combustible material that may spontaneously combust when exposed to the air during redevelopment, or ignited or heated from an external source (Leach & Goodger, 1991).

The mineral extraction sites category covers a range of mining activities and methods; these include quarries, gravel pits, clay pits, coal mining (deep & open cast) and metalliferous mining. The relics of these activities exist across the United Kingdom from the ore fields of Devon and the North & South Pennines to the china clay pits of Cornwall and the limestone quarries of the Mendips.

Open cast workings usually cover a vast number of hectares but are often relatively free of biodegradable material and contamination (excluding open cast coal mining). The infilling after the completion of works usually causes problems, as fill material can range from domestic waste to industrial chemical waste.

By contrast metal mining has resulted in major contamination. The contaminants are normally the metals that were actually mined, or the associated metals that occur within the mineralised deposit, for example, cadmium associated with lead and zinc deposits. The older sites are usually heavily contaminated by metals due to crude extraction techniques which produced tailings often containing up to 10% metal, in comparison to modern flotation techniques that reduce tailings to 0.1 % metal (Johnson & Bradshaw, 1977). Besides tailings, mine water containing contaminants may cause problems by entering ground or surface water systems.

A full series of forty-eight industrial profiles have been drawn up by the Department of the Environment (1995). These identify the range of chemical pollutants associated with each industry, and are designed to assist with treatment of resulting problems.
As regards the severity of the problem caused by these sources, Myers et al (1994) have sub divided the categories by severity of contamination caused by an industry. Table 2.2 outlines four categories; highly contaminative; moderately contaminative; slight contamination and low contamination.

Category 1: *High contamination*	**Principal type of contamination ‡**
Hazardous waste treatment	O/I
Bulk organic chemical manufacture	O
Fine chemical manufacture	I
Coal gasification/carbonisation	O
Landfill and other waste treatment/disposal	O/I
Steelworks	O/I
Lead metal ore processing and refining	I
Oil refining and petrochemical production	O/I
Pesticide manufacture	I
Asbestos & asbestos products manufacture	I
Scrap yards	O/I
Pharmaceutical manufacture	O
Category 2 *Moderate contamination*	**Principal type of contamination ‡**
Drum and tank cleaning/recycling	O
Fertiliser manufacture	I
Non-ferrous metal ore mining	I
Wood preservatives production & timber treatment	O
Docks	I
Electric/electrical equipment manufacture	O

Mechanical engineering	I
Garages/filling stations	O
Mineral processing (bricks, cement, tarmac, *etc*.)	I
Power stations	I
Sewage treatment works	I
Shipbuilding/ship breaking	O/I
Textile production and dying	O
Tyre manufacture and other rubber processing	O/I
Metal (other than iron or lead) processing/refining	I
Pulp and paper manufacture	O
Paint and ink manufacture	O/I
Electroplating and other metal finishing	O/I
Precious metals recovery	I
Foundries	I
Tanneries	O/I
Airports & airfields	O
Category 3 ***Slight contamination***	**Principal type of contamination ‡**
Timber products manufacture	O
Animal processing works	O
Glass manufacture	I
Road haulage yards	
Building trades products manufacture	O
Printing works	I
Research laboratories	O
Vehicle manufacture	O
Railway yards/sidings	O
Toiletries, detergents, disinfectants, *etc*, manufacture	O
Electricity sub-stations	O
Dry cleaners	O
Category 4: ***Low contamination***	**Principal type of contamination ‡**
Food preparation/processing (inc. brewing)	I
Distilleries	I
Railway tracks	I
Agriculture	None
Oil shale & coal mining	I

Table 2.2: Categorisation of Major Industrial Land Uses. † (Source : Myers et al, 1994)

Notes:

"O" signifies organic contamination; I signifies inorganic contamination.

‡ The categorisation is for illustrative purposes. It will give a broad indication only of whether the business concerned involves a contaminative use, and the likely nature of the contamination arising from a contaminative use.

† Whether a business falls within a particular category will depend on a number of further factors, such as: (a) whether the previous business use of the site has given rise to contamination; (b) the period of time for which the site has been used for the business purpose; (c) the overall sensitivity of the site with respect to its broader environmental setting; (d) assessment of the extent to which the business follows good environmental practices and management controls; (e) the extent of the manufacturing or processing activity which is carried on by the business and the site; and (f) assessment of the influence of the underlying geology and its hydrogeological characteristics.

2.6 Investigation of Contaminated Land

The investigation of a contaminated site requires both the collection of qualitative information (site use, past use, etc.) and quantitative data (ground conditions, contaminant concentrations, etc.), which then need to be evaluated and assessed in terms of the effect on the environment, human health, construction materials and other sensitive targets. Therefore it is important that an integrated approach is undertaken, combining the site investigation findings with the remediation requirements. This should then allow the site to be developed in a satisfactory, safe and economic manner.

Herbert (1995) suggested that, up until recently, many investigation practices and remediation techniques used in the UK relied heavily on standard civil engineering methods and procedures, which lacked guidance in terms of risk management. This has, however, begun to be rectified to a certain extent, with the publication of a series of CIRIA Special Reports (Harris et al, 1995a) giving full guidance on all aspects of contaminated land investigation, assessment and remediation in terms of risk management.

2.6.1 Risk Management

When investigating a potential contaminated site it is generally acknowledged that a risk management approach should be undertaken. The term risk management is best described by the Royal Society (1992) definition as " The process whereby decisions are made to accept a known or assessed risk and/or the implementation of actions to reduce the consequences or probabilities of occurrence".

This process covers a range of activities from the initial site visits to post-treatment monitoring. This relationship of the risk management process and the main stages of a work programme of site investigation, assessment and remediation are shown in Figure 2.1.

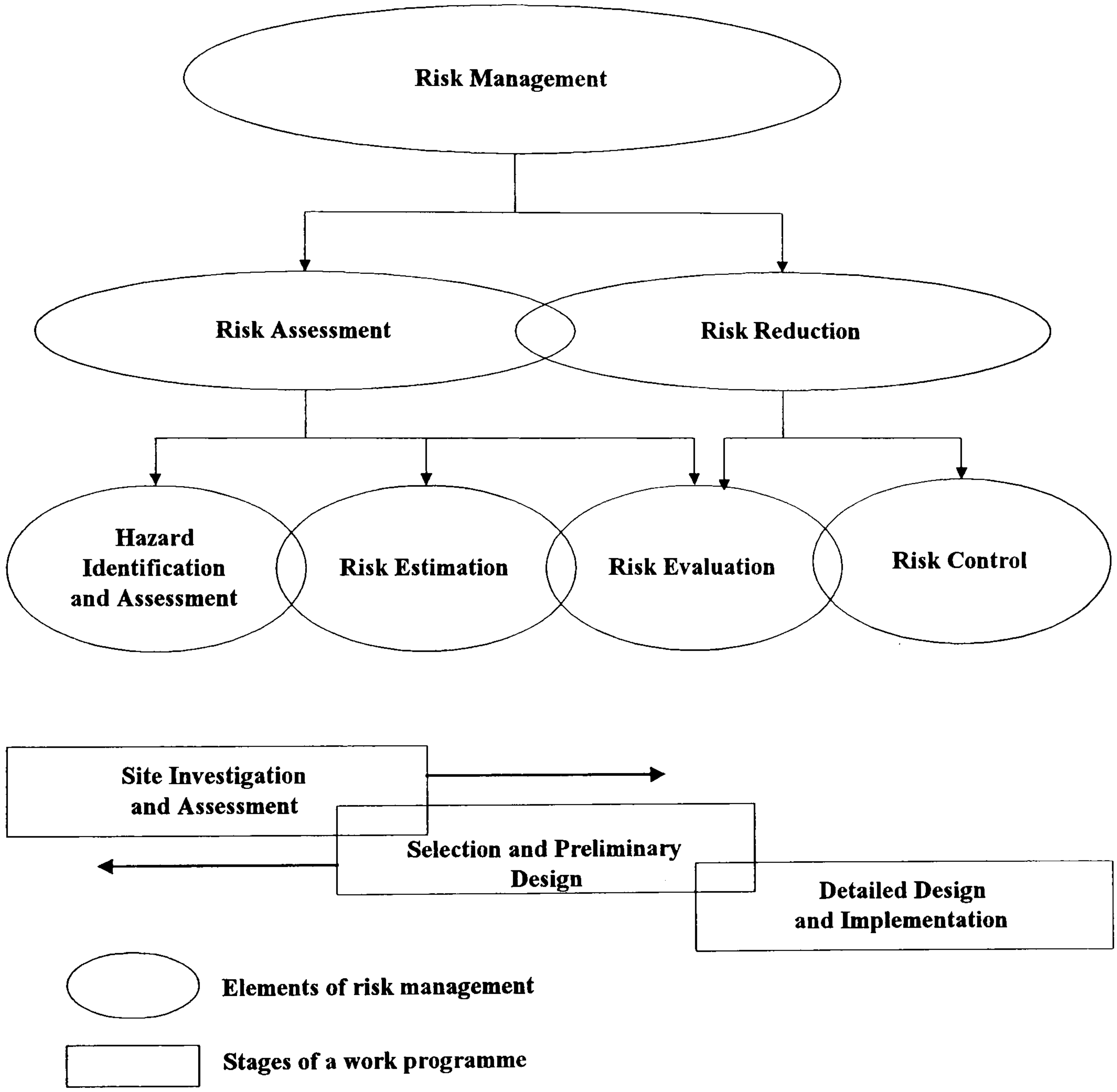


Figure 2.1 : Relationship Between Risk Management and Main Stages of a Work Programme.
(Source : Harris & Herbert, 1994)

Petts (1993) highlights the advantages of a risk management approach to contaminated land as:

- Structured;
- Objective;

- Comprehensive;
- Explicitly considers uncertainties;
- Provides a rational, transparent and defensible basis for discussion of a proposed course of action with e.g. regulators, funders, insurers, the local community.

As shown in Figure 2.1 the risk management process splits into two distinct categories; risk assessment and risk reduction. The risk assessment category is normally dealt with during the site investigation stage with the risk reduction category involved with the remediation and implementation. Figure 2.1 also indicates (highlighted with arrows) that the stages of work programme overlap.

2.6.1.1 Risk Assessment

A risk assessment strategy involves three separate components; hazard identification and assessment, risk estimation and risk evaluation (Figure 2.1). The main objectives of the risk assessment strategy and these three components are outlined by Harris et al (1995b) as :

- To determine systematically any risk arising from any contamination present on the site and whether these are 'unacceptable';
- To provide, at least, a qualitative statement about the magnitude and nature of the risks where they exist;
- To determine the effects of foreseeable events, such as weather extremes, rising water-table, flooding, increase in neighbouring populations etc. on the nature and magnitude of the risks;
- To determine the consequences (e.g. potential impacts on the environment, groundwater resources, public health) of a change of use, development, redevelopment or other works on the site;

- To identify the critical contaminants and associated factors (e.g. pathways) relevant to the site so that the steps necessary to reduce risks to 'acceptable' levels, both currently and in the foreseeable future, can be determined;

- To help to set objectives and priorities for reducing risks;

- To make judgement about the significance and acceptability of identified risks;

- To provide a rational and defensible basis for discussion about a proposed course of action with third parties (e.g. regulators, insurers, local community etc.).

The process is normally undertaken on a site-specific basis. Ellis & Rees (1995) highlight the reasoning behind this, as the fact that the accuracy of risk assessment is highly dependable upon a thorough understanding of the fate and effects of contaminants under site-specific conditions and use.

2.6.1.2 Hazard Identification & Assessment

The hazard identification and assessment component involves collecting enough reliable and accurate information about the site (geotechnical/hydrological properties), possible contaminants and the neighbouring environment to identify possible hazards and plausible scenarios that may cause problems. This information is normally collected though the site investigation process. A plausible scenario consists of three main elements: a source of contamination (hazard), a pathway for movement of contaminants and a sensitive receptor or target. Young et al (1997) outline these basic data blocks that make up the process;

- Definition of source of contamination
    - Location of contamination
    - Nature of contamination
    - Concentration
    - Total loading

- Identification of Pathways
  - Site topography
  - Soil/rock permeability
  - Joint/bedding systems
  - Man-made pathways (mine shafts, pipe backfill, etc.)
  - Surface drainage channels

- Location of Sensitive Receptors
  - Depth to groundwater
  - Proximity of surface water.

Other possible targets (sensitive receptors) not covered by Young et al (1997) may include site works, future occupiers, neighbouring users, soil & air quality, flora and fauna and building services.

The Department of the Environment (1994a & b) report sets out a framework for the assessment of impacts of contamination on ground and surface waters and discusses techniques available for quantitative predictions.

2.6.1.3 Risk Estimation

The risk estimation process normally involves constructing a model to estimate the amount of a contaminant that may travel from a source to a possible target and the effect on the target. This usually involves two different procedures, an exposure assessment and toxicity assessment. The aim of the exposure assessment is to define the environmental transport and fate of contamination. Harris & Herbert (1994) suggest the factors that need to be considered are;

- Chemical form and physical properties;

- Characteristics of the host medium (soils, rock, groundwater etc.) and effect on contaminant concentrations along travel pathways;

- Concentration of contaminants at the source, at points along the travel pathway and at the point of exposure (e.g. ingested by the target);

- Rate of movement along the pathway;

- Amount, frequency and duration of exposure;

- Characteristics of exposure route (e.g. ingestion, inhalation, direct contact) that determine how much of the contaminant is taken in by the target;

- Data limitations.

Ferguson (1996) reviews recent research involved with the assessment of human health risk from exposure to contaminated land. The toxicity assessment involves determining the effect of the hazard on the target under the conditions defined in the exposure assessment. The effects may range from examining the effect on human health, therefore undertaking a toxicity assessment, to focusing on the impact on building materials (corrosion assessment, BRE (1994) & Cairney (1995) outlines in detail the risk of attack on construction materials). On completion of the risk estimation exercise the resulting output may be presented in qualitative terms (a statement that a risk of defined level of harm is high, medium or low), or quantitative terms (e.g. the risk of excess cancer over the lifetime of the individual is less than 1 in 10^6).

Harris et al (1995b) suggests that, an important feature of risk estimation is that it enables action values (i.e. the point at which further assessment or remedial action should be taken) and remedial values (e.g. the residual contaminant concentrations which any remedial action must achieve) to be determined on a site-specific basis.

2.6.1.4 Risk Evaluation

The risk evaluation process consists of making valid judgements concerning the acceptability of the risk estimates. This involves taking into account the uncertainties associated with the risk estimates as well as using available guidance. This acceptability can obviously differ between different parties involved in the process. Also, the acceptability of a risk may change as more information about the cost and feasibility of the remedial action is evaluated. Harris & Herbert (1994) suggest, for example, that a risk may be considered unacceptable (even when judged to be low) if there are serious consequences (e.g. an explosion leading to human fatalities). A high risk (e.g. death of a proportion of young landscape plants) may be tolerated if the cost and practical problems of removing the source of the risk (moderately high concentrations of phytotoxic metals) are more onerous than those associated with rectifying the damage (e.g. periodic replacement of stock) should it occur. The essential aim of the evaluation is to assess how changes in assumptions made during the assessment may affect the outcome of the project. In marginal cases a small adjustment in assumptions can have a major effect on the risk estimate, which in turn may have implications which may lead to costly problems with remediation.

2.6.1.5 Risk Reduction

The risk reduction process consists of two main elements; risk evaluation and risk control, with risk evaluation overlapping with the risk assessment process. Each element contributes to the decision about the level of contamination that is taken to be unacceptable for the defined targets, out-lining the type of response required to reduce or control risks to defined levels and finally undertaking a remedial strategy and monitoring procedures that achieve the remedial action objectives in both the short and long term.

Therefore the most important aim of the process is to select the correct remedial strategy that offers the best risk reduction that is feasible in terms of the available skills, plant, time, and engineering properties (including environmental impacts), as well as being cost effective and acceptable to other relevant parties.

It is general practice in the UK that the risk reduction process takes the form of agreed remedial standards; these are normally known as Contamination Related Objectives (CROs), and take into account the proposed end use of the site. The CROs are usually expressed as residual concentrations of contaminants in affected media (i.e. soils, groundwater, surface water). Harris and Herbert (1994) highlight typical examples of CROs;

- The concentration of specified contaminants that should not be exceeded in soils remaining in place following excavation;

- The concentration of specified contaminants that should not be exceeded in recycled/imported replacement materials;

- The concentration of specified contaminants present in the coarse clean fraction of a soils washing plant;

- The concentration of contaminants present in groundwater following a pump-to-treat operation;

- The concentration of contaminants in groundwater on the "clean" side of a cement-bentonite cut-off wall.

If after the evaluation period it becomes apparent that complying with the CROs is likely to be impossible, then one or a combination of alternatives may be utilized. These may include: using the site for a less sensitive propose and hence redefining the CROs, extending the completion time for the remedial action and increasing the resources available to overcome short- and long-term constraints.

2.6.2 Site Investigation

The term site investigation is widely used, and is often taken to mean physical exploration on site, such as the excavation of trial pits or the sinking of boreholes. For the investigation of contaminated land a more intensive investigation is usually required, which unfortunately in the past has not always been delivered.

Crowcroft (1994) confirms this, stating that "throughout the 1980's, this was the case. The investigation of contaminated land became a bolt on part to geotechnical investigations and not an investigation in its own right".

Harris & Herbert (1994) highlight this problem; indicating that poorly informed and executed site investigation works may expose investigation personnel, and the general public, to unacceptable health risks and could lead to more extensive or intractable contamination problems than those which previously existed on the site.

The British Standards Draft Code of Practice DD175 defines site investigation of contaminated land as follows :

"The planned and managed sequence of activities carried out to determine the nature and distribution of contaminants on and below the surface of a site that has been identified as being potentially contaminated. These activities comprise identification of the principal hazards; design of sampling and analysis programmes; collection and analysis of samples; and reporting of results for further assessment".

Therefore, it is important that the investigation is more than just the excavation of trial pits and boreholes, and that clear plans and objectives are set from the outset of the investigation. Table 2.3 illustrates examples of the investigation objectives. It can be clearly concluded that a full investigation of contaminated land crosses a range of disciplines, including: geology, chemistry, ecology, hydrogeology and geotechnics/civil engineering.

Due to the fact that the investigation covers a range of fields, it is important that a multi-disciplinary approach is taken when an investigation of contaminated land is undertaken.

Aspect	Objectives/information needs
Contamination	To determine the : • Nature, extent, source and distribution of contaminants (on and off-site) in a range of media - soil/fill/wastes, ground/surface water, air, biota, containers (drums etc.) • Form of contamination or contaminated media - gaseous, liquid, semi-solid, solid • Ground temperatures • Level of microbial activity • Health of ecosystems (soil, water, land area)
Water environment	To determine where appropriate the : • Groundwater levels/pressures and their variation with time • Direction and volume of flow of ground and surface water • Abstraction and recharge activities having an influence on the site • Chemical and mineralogical quality of ground and surface water • Background chemical composition of surface and groundwater in the area • Geological strata composition and structure • Primary and secondary permeability/porosity • Propensity of site to flood • Rainfall and evaporation characteristics • Tidal fluctuations
Geotechnics	To determine, where appropriate, the : • Physical characteristics of the ground e.g. presence of in-ground obstacles, services etc. • Physical characteristics of contaminated matrices e.g. mineralogy, moisture content, permeability, chemical composition, particle size distribution • Geotechnical characteristics e.g. strength, compressibility, stability of slopes, existing structures, potential subsidence etc.

	• Presence of old mine workings
Actual and potential targets and pathways	• Potential exposure pathways identified from detailed analysis of all above information • existing or proposed use of site and surrounding land • Potential human targets including site workers (investigation/remediation/construction/maintenance) , occupants, users, neighbours and trespassers • Proximity to sensitive ecosystems • Proximity of water bodies • Proximity to economically valuable natural resources (e.g. mineral deposits)

Table: 2.3 Examples of Investigation Objectives. (Source : Harris et al, 1995b)

2.6.2.1 Phases of Investigation

It is clear that a vast amount of information is required to fully assess a potentially contaminated site. Therefore in order to assess the information fully and prioritise the needs of the investigation, undertaking the process in a series of phases enables the investigation to be refined as more information is identified.

Harris & Herbert (1994) state that "the investigation of contaminated sites should involve at least three phases (preliminary, detailed and compliance/performance investigations) and may involve up to five:-

♦ Preliminary investigation (comprising desk study and site reconnaissance);

♦ Exploratory investigation (e.g. preliminary sampling, monitoring);

♦ Detailed investigation (involving detailed on-site exploratory work);

- Supplementary investigation (the collection of additional site investigation data for specified purposes);

- Investigation for compliance and performance (comprising on-going monitoring and validation of remedial action, and post-treatment management)".

Each phase has different objectives, as highlighted by Table 2.4, and involves different types of investigation methods; with the results from each phase assisting in the design of the next phase.

Phase of Investigation	Typical Objectives	Typical Activities
Preliminary Investigation	To provide background information on past and current uses, hazards, geology and hydrology, possible scale of contamination etc. To inform design of on-site work (including sampling and analysis, health and safety, environmental protection) Can be used to rank a number of sites based on hazard potential. May provide initial indication of remedial needs.	Literature review (Desk Study) Consultation (e.g. site owners, neighbours, regulatory authorities) Site Visits (walkover survey)
Exploratory Investigation	To confirm initial hypotheses about contamination and site characteristics To refine design of detailed investigation	Preliminary sampling (e.g. surface deposits, vegetation) Preliminary monitoring (e.g. gas composition and groundwater quality, flora and fauna)

Detailed Investigation	To characterise fully contaminants, geology, hydrology of site and associated pathways and targets To inform risk assessment and selection of remedial methods.	Comprehensive investigation of ground (e.g. using trial pits/trenches, boreholes) Monitoring (e.g. gas composition and water quality, flora and fauna)
Supplementary Investigation	To obtain additional information in support of risk assessment and/or selection of remedial strategies.	Further ground investigation and monitoring Treatability testing
Investigation for Compliance and Performance	To confirm effectiveness of remedial action	Post-treatment validation and monitoring as appropriate

Table 2.4 : Examples of Objectives and Activities Associated with Site Investigation. (Source : Harris & Herbert, 1994)

2.6.2.2 Preliminary Investigation

The preliminary investigation is split into two stages. The desk study and the site reconnaissance. Each stage plays an important role in achieving the required objectives and the procedures for the exploratory phase of the investigation. It also assists with the health and safety and environmental protection requirements for on-site work. Harris et al (1996) suggest that the preliminary investigation should also reduce the risk of;

- An investigation design which requires the comprehensive measurement of contaminants and other hazards which, in reality, are unlikely to be present or relevant to the objectives of the investigation;

- An inadequate investigation design which fails to provide the data needed either to assess the hazards and risks or to select appropriate remedial measures where necessary.

2.6.2.3 Desk Study

McEntree (1991) suggests that the importance of undertaking a thorough desk study prior to the site-work stage of the investigation cannot be over stressed. This is also reinforced by many other authors, including Forde et al (1992), Young et al (1997), Harris et al (1995b) and Jewell et al (1993). Government publications such as BS5930 (1999) and DD175 (1988) also highlight the merits of the desk study process.

The results from such studies provide important information for designing the ground investigation stage, as well as assessing the hazards likely to be encountered by site investigation/construction personnel and end users of the site. McEntree (1991) reports that there have been instances where site investigation personnel have worked on site with no knowledge whatsoever of the chemical hazards affecting the site, and consequently have taken no precautions for their own safety.

The process is often quite time consuming as it usually involves searching old archives and records, but often results in gathering a great deal of relevant information about the site. Such information often reduces the time and money spent on later stages of the investigation.

Steed et al (1996) highlight typical information that may be gathered :

- History of the site, details of its owners, occupiers and users, as far back as possible;

- Processes used, including their locations, raw materials, product waste residues and methods of disposal;

- Chemical and physical properties of potential contaminants on site;

- Layout of the site above and below ground at each stage of the development including roadways, storage areas and other hard-cover areas, and the presence of any buildings, pits, and services, i.e. gas, sewer, electricity, drains, water, telecommunications;

- Presence of waste disposal tips, abandoned pits and quarries, without standing water;

- Mining history including shafts and roadways (worked seams);

- Previous survey data, e.g. borehole and trial pit logs, sample analysis results;

- Information on geology and hydrogeology, including presence of groundwater and surface water;

- Presence of nearby contaminated sites from which contaminants could spread via air and/or groundwater to site in question;

- Populations at risk, e.g. proximity of local population centres.

2.6.2.4 Site Reconnaissance

The site reconnaissance, if undertaken by an experienced investigator, can identify abiotic and biotic indicators which can confirm findings from the desk study stage and assist in planning sampling patterns and frequency of the exploratory phase of the investigation. Hobson (1993) suggests that the reconnaissance should, wherever possible, be conducted on foot and it is usually best to walk around the perimeter of the site first, before inspecting the central area and points of detail. This gives an understanding of the overall scale of the site and allows landmarks to be easily located.

Department of the Environment (1994c) define abiotic indicators as; "debris and structures on site; anomalies in topography and soil between the site and adjacent land or within the site; the presence of characteristic colours and odours."

The abiotic indicators of past or current activities may sometimes be the only evidence of the presence of contaminants. Although the abiotic indicators are rarely used by themselves, they are typically combined with other factors to point to a particular contaminant. Besides the obvious features such as characteristic buildings, infrastructure and machinery which indicate past or current land use, surface deposits and soil colouration can also be characteristic of contamination. For example, Sury & Slinsby (1991) suggest that white surface deposits can be one of a number of chemicals including Calcium Sulphate. Hidding (1986) and Forth & Beaumont (1996) also suggest that another reliable indicator is “blue billy” (a complex of spent oxides containing iron and cyanide compounds) due to its characteristic colour and smell, which is indicative of waste from gasworks.

Other features such as bare patches have many possible causes, including toxicity, made ground, or mechanical wear compaction by vehicles, as well as natural stresses such as drought and nutrient deficiency (IEHO, 1988).

Odours can also be associated with different types of contamination. James et al (1985) have published a number of descriptions, for example; Carbon tetrachloride is described as being strongly odorous; pungent; ether like, or Chlorobenzene; chlorinated moth balls; aromatic; faint; pleasant. It should be borne in mind that certain odours can be produced naturally from decomposing vegetation, anaerobic mud (hydrogen sulphide), and other organic sources.

Surface water and drainage patterns can also provide evidence of potential contaminated areas of a site. For example, surface water draining from the site, should be inspected both upstream and downstream of the site, to ensure that the stream is not already contaminated before entering the site. The Department of the Environment (1994d) highlight the most obvious signs of possible contamination to be.

- Turbidity of the water (other than after heavy rainfall);

- Discoloration of water and sediments - e.g. dark or reddish ochre staining;

- Odours associated with the water;

- Presence of sewage fungus;

- Foaming;

- Presence of oily deposits or film on water surface (natural processes can also produce a film from decaying organic matter, but man made oils can often be distinguished from this by smell);

- Gases bubbling continuously through the water;

- Lack of, or abnormal, aquatic vegetation and fauna.

There are also usually very simple abiotic indicators that may be found in the vicinity of the potentially contaminated site. Street/house names or public house names which can give clues to particular past industrial uses, for example, Coal Tar Lane, Brickmakers Arms and Gas Works Alley, are often found.

BS5930 (1999) & Richard *et al.* (1996) also suggest interviewing neighbours of the site and other parties. Although of variable reliability, the information may give a lead to past use of the site.

The abiotic indicators obviously play a vital role in helping to identify types and locations of possible contaminants. It is also important to remember that information regarding the types of soil and geological features can also be examined during the site reconnaissance. This information plays an important part in identifying contaminant flow pathways in the risk assessment process.

DD175 (1988) & BS5930 (1999) recommend the examination of nearby railway cuttings, road cuttings or old excavations, as these can often reveal local soil and rock types and their characteristics. Similarly the examination of embankments, buildings and other structures may indicate a history of settlement, and maybe the presence of compressible or unstable soil.

As with abiotic indicators, biotic indicators can play an important part in helping to identify possible contaminants. The Department of the Environment (1994d) state that; "Biotic indicators are related to biological features of the site and include : the type of animal or plant species present; symptoms of effects of contamination in any species; the conditions of the soil".

The Department of the Environment (1994c) suggest that biotic indicators are rarely of use unless considered in the context of abiotic indicators and information on site history. Biotic indicators are only useful on sites where concentrations of contaminants are sufficient to affect biota.

The use of plants as indicators for ground conditions has been known for at least half a century. Cannon (1971) describes how toxicity symptoms and physiological and morphological changes in plants caused by varying soil conditions, such as unusual amounts of metals in the soil, have been used in mineral prospecting, geological mapping and groundwater surveys. The use of biological monitoring of fish has been used by the National Rivers Authority (now a part of Environment Agency) (NRA, 1994) for classifying water quality.

2.6.3 Investigation Methods

Both the exploratory and detailed phases of the site investigation require a comprehensive collection of data regarding the ground conditions, contaminant concentration, etc. In order to undertake this task, there are a number of techniques normally used. The techniques can be split into two separate classes; non-intrusive activities and intrusive methods that require physical sampling on the site. DD175 (1988), Crowcroft (1994) and Young et al (1997) summarise the techniques available;

- Non-Intrusive Techniques
  - Surface gas emission testing;
  - Geophysical testing;
  - False colour infrared photography;
  - Thermography;
  - Tracer gas testing;

- Intrusive Techniques
  - Boreholes;
  - Trial pits & trenches;
  - Probing techniques;
  - Window sampling;
  - Gas & water monitoring wells.

The actual techniques used will vary depending upon the needs of the investigation, which may not always suit the nature of all the contaminants on site. For example, trial pits and trenches provide an excellent method for visible inspection of the contaminants present and the media within which they are contained. Syms (1997) suggests that this method is unsuitable for volatile contaminants, due to problems with sample collection. In this case a borehole investigation would provide better results.

Therefore it is important to assess the likely contaminants that may be present and compare the advantages and disadvantages of the techniques available. Another major consideration is to anticipate the sample testing programme required, and select an appropriate sampling pattern and the correct number and size of samples to be collected.

2.6.3.1 Non-Intrusive Techniques

There are a range of non-intrusive techniques as previously outlined, that may be used to identify anomalies in ground or vegetation patterns that are indicative of contamination.

One of the most popular range of techniques used involves geophysics. These techniques range from resistivity and seismic methods to ground penetrating radar techniques. Such techniques, in particular seismic refraction, have been widely used and have a proven record in mineral exploration, hydrogeology and geotechnical engineering. In terms of investigating contaminated land there are a range of applications that the techniques may be used for, which include; locating buried storage tanks, drums and pipes, estimating the general composition of landfill, mapping leachate/contaminant plumes, detecting cavities and investigating hydrogeological and soil/bedrock conditions both laterally and vertically.

Jewell et al (1993) suggested that for the best results it is essential to use a combination of techniques, for example, seismic refraction and resistivity sounding, or transient electromagnetic techniques and ground penetrating radar.

The geophysics techniques also require other ground investigation methods, such as drilling and geochemical testing to be undertaken in conjunction with them, in order to confirm their findings. Leach & Goodger (1991), Jewell et al (1993) and Crowcroft (1994) highlighted the available range of geophysics methods available in more detail.

Other non-intrusive techniques involve collecting data from aerial views of the site using a balloon or an aircraft (real or model). These include, False Colour Infrared Photography which gives an overall view of a site and highlights areas of vegetation distress. Problems with interpretation can arise with this method as waterlogged ground can cause vegetation distress and produce the same results as those seen for contaminated distressed vegetation.

Another method undertaken from the air is the thermography technique, which involves detecting small variations in surface temperature. Elevation in temperature can be indicative of underground fires. Such temperature changes may also relate to human activities or installations such as manholes, therefore caution is required when interpreting the results.

Surface gas emission testing equipment can be another useful screening technique. Sensitive flame ionization detectors can give indications of areas of greater gas generation. These areas may be leaks from underground storage tanks or former landfill sites. Emissions may also reflect the quality of capping of gas releasing wastes. Monitoring off-site surfaces will also indicate the detection of gas migration. Samples may also be collected in gas-bags and subsequently laboratory tested.

Volatile organic compounds favour the soil vapour phases and the dissolved and adsorbed phases, therefore testing soil vapour for such material is also an excellent indicator of the presence and location of such contaminants. There are a range of techniques available for testing, monitoring and collecting gas and soil vapours; these are outlined in detail by Smith (1993), Figg et al 1980 and Farias et al (1993).

2.6.3.2 Intrusive Techniques

There are a number of intrusive techniques available in order to collect information regarding the sub-surface. The drilling of boreholes remains one of the most commonly used tools for contaminated land investigation, even with advances in less invasive techniques. The use of boreholes has three main purposes; collecting soil samples for testing, retrieving stratigraphical and lithological data and installing monitoring wells for groundwater sampling for both short and long term investigation.

The collection of such data also allows information from non-intrusive techniques to be confirmed. There are a range of drilling techniques available depending upon ground conditions, depth of hole required and the type and nature of sampling and monitoring required. These include light cable percussion drilling, rotary drilling (open-hole drilling, coring using double or triple tube core barrels) and auguring (hollow and solid-stem). Hobson (1993), Jewell et al (1993) and Crowcroft (1994) explain the techniques available and the advantages and disadvantages in detail. Jewell et al (1993) also highlight the importance of decontaminating equipment between uses to ensure that cross contamination of samples and uncontaminated ground does not occur. This is obviously important for all intrusive techniques.

On completion of drilling, some form of monitoring well (standpipe, piezometer) is usually installed within the borehole. This allows in-situ monitoring of both gas and water on site. There is an array of monitoring installations available as well as a number of sampling methods. These are described in detail by Chilton (1996) and Bell (1993) for groundwater and Department of the Environment (1991a), Smith (1993) and Crowcroft (1994) for gas monitoring and sampling.

As well as the drilling of boreholes, the excavation of trial pits or trenches is widely used during the investigation of contaminated land. These provide the only method of examining a relatively large cross-sectional area of the sub-surface. The technique also allows the collection of large disturbed samples, although undisturbed samples may also be obtained by driving sampling tubes into the side of the pit. The pits are usually excavated using mechanical excavators and are normally between one and one and half metres wide and up to seven metres in depth, although support is required below 1.2m to allow the investigator to enter the pit. Such pits may be easily extended into a trench if required.

The main advantage of the trial pit method over other intrusive techniques is that it is relatively cheap and reasonably quick to excavate, Hobson (1993) suggests up to twenty holes per day can be excavated. The disadvantage is the disturbance that they create (loosely backfilled holes can obviously cause problems), as well as the cost of reinstating the damage caused at the surface of an existing development.

Other intrusive methods include probing techniques, and there are two common forms of probe available; the Standard Penetration Test (SPT) and the Static Cone (CPT). Both have their advantages and disadvantages. The Static Cone probe allows continuous measurements over the depth of penetration (end resistance, sleeve friction, pore pressure and conductivity) although small hard obstructions can prevent progress. The SPT can overcome obstructions but is less sophisticated (number of blows to penetrate a certain distance relates to ground strength) and can only measure a limited range of parameters. The SPT also has an additional cone attachment that is used within coarse material, such as gravel. Any measurement method used has the problem that it does not hold well in heterogeneous made ground or landfill.

The advantage of the Cone probe technique is that it may be used to insert gas/water monitoring tubes as well as being modified to obtain high quality samples from the ground. This is useful where sites have limited access and where relatively shallow sampling is required.

2.6.4 Sampling Strategies

The primary purpose of the exploratory phase of the investigation is to collect samples that are representative of the bulk medium both chemically and geotechnically. Therefore choosing an appropriate sampling strategy is extremely important. DD175 (1988) highlights the factors that need to be considered when designing an optimum sampling programme;

- The number of stages of sampling;
- The number of sampling points;
- The choice of sampling pattern;
- The size of sample required by the analyst;
- The need to define the position of each sampling accurately.

It also states that the strategy should be designed to suit the particular needs of the site and the methods of collection and analysis.

Contaminants are often contained within isolated areas across the site, rather than evenly distributed in the ground across the site. These areas are commonly known as "Hot-Spots". Ferguson (1993) suggests that the sampling should be designed to answer three key questions;

- Which hazardous substances, if any, are present in the soil?

- Do contaminant hot-spots exist on the site, and if so where?

- What size and shape are the hot-spots, if they exist?

Due to the nature of the hot-spots, it is obvious that locating them all is difficult, although the data collected during the preliminary investigation can assist in locating such areas with a reasonable degree of accuracy. The Department of the Environment (1994e) suggest that a decision needs to be made on the largest hot-spot that could be accepted or dealt with economically if it were missed in sampling. This critical hot-spot size is an important design parameter.

As regards to sampling patterns, there are three main sampling patterns usually discussed in the contaminated land literature. These include; square grid, simple random and stratified random. The mathematical theories associated with these patterns are outlined by Ferguson & Abbachi (1993) and the Department of the Environment (1994e). The square grid is the most popular due to its obvious practical advantages. Ferguson (1992) suggests that an efficient sampling pattern should satisfy four conditions;

- It should be stratified (that is, the area to be sampled should be partitioned into regular sub-areas);

- Each stratum (sub-area) should carry only one sampling point;

- It should be systematic;

- Sampling points should not be aligned.

Unfortunately, the three most commonly used sampling patterns do not satisfy all of the above conditions. A fourth sampling pattern, the herringbone sampling pattern has been devised by Ferguson (1992). This pattern overcomes the disadvantages of the other methods and satisfies all four design conditions. It is also relatively easy to set out on site.

In general the exploratory phase of the investigation may therefore comprise a mixture of specifically targeted trial pits or boreholes with others conforming to a grid or herringbone pattern (Syms, 1997).

Another consideration is the number of sampling points required. DD175 (1988) highlights the minimum number of sampling points required, according to different site areas (Table 2.5). Besides the number of sampling points DD175 (1988) also suggests that, at least three samples should be taken at each sampling location.

Area of Site (in hectares)	Minimum number of sampling points
0.5	15
1.0	25
5.0	85

Table 2.5 : Minimum Number of Sampling Points. (Source : DD175, 1988)

A full review of research developments within sampling methodologies is given by Smith (1996).

2.7 Conclusion

The subject of contaminated land has always been considered a controversial issue, due to financial, social and political interests. Even the simple question "What is contaminated land ?" has many and varied answers due to the range of disciplines involved in the subject area. For example, an area containing high natural levels of elements and compounds may be regarded as contaminated in a general view, but the majority of definitions relate to contamination as a result of human activity and so these areas may not be considered to be contaminated.

Due to the confusion of a consistent definition for contaminated land, data on the extent of contaminated land in the UK are sparse but it is usually estimated that 100,000 ha of land is contaminated comprising of some 50,000 to 100,000 sites; this represents around 0.4% of total land area in the UK. The majority of these sites have arisen since the start of the industrial revolution in the later part of the 18th century, although there are examples of areas contaminated by Roman copper & lead workings from some 2000 years ago.

The figures quoted indicate that the majority of civil engineering projects undertaken are likely to encounter contamination, and therefore the need for effective contaminated land investigations is increasing. In the past, the investigation of contaminated land became a bolt on part to geotechnical investigations and not an investigation in its own right.

Since the recognition of the problems associated with the redevelopment of contaminated land, the government have produced guidance to ensure that contaminated land is identified well in advance of any redevelopment. It is now widely recognised that due to the hazardous nature of the redevelopment of contaminated land, it is essential that the investigation is undertaken in a number of stages (preliminary, exploratory, detailed investigation and an investigation for compliance and performance if required) and that each stage of the investigation is revised as more information becomes available. The data collected during the investigation provides information for the risk assessment process. The risk assessment process highlights factors such as hazards to end users, site workers and the local environment. It also gives an indication to the size and cost of reclamation programme required. The basic data blocks for the risk assessment process consist of: definition of contaminants on site, identification of possible pathways for the movement of contaminants and location of vulnerable targets on and off site. Therefore as more information regarding the site becomes available the risks posed by the site can be minimised. Addressing the complex parameters involved in the risk assessment process comprehensively and successfully requires expertise and knowledge from a number of disciplines, ranging from geotechnical engineers to chemists.

The preliminary stage of the investigation can often yield a wealth of useful information, but it is often not used to its full potential, or unfortunately, neglected completely. This stage of the investigation is generally split into two stages; the desk study and the site reconnaissance. Each stage plays an important role in achieving the required objectives and the procedures for the exploratory phase of the investigation.

The desk study stage usually involves searching old archives and records in order to gather relevant information about the site under investigation. The results from such studies provide important information for designing the exploratory phase as well as assessing the hazards likely to be encountered during site work. The site reconnaissance stage is designed to identify abiotic and biotic indicators, which can confirm findings from the desk stage. This stage also assists in planning sampling patterns and frequency of the exploratory phase of the investigation.

The exploratory phase of the investigation requires a comprehensive collection of data regarding the ground conditions, contaminant concentration, etc. In order to undertake this task, there are a number of techniques normally used. The techniques are generally split into two separate classes; non-intrusive activities and intrusive methods that require physical sampling on the site.

CHAPTER 3

REVIEW OF DEVELOPMENT TOOLS

3.0 Introduction

In order to develop an effective system using the most suitable tools, a review of database management systems and knowledge-based system technology was undertaken.

This chapter starts with the description of the various data models and the benefits of using a database over other storage methods (section 3.2). This is followed by a review of the development of geotechnical database systems and the implications of a national standard for the storage and transfer of geotechnical data. The selection of suitable database software is then discussed within section 3.4.

The second half of this chapter introduces the concept of knowledge-based system technology, initially reviewing the definitions offered by various authors (section 3.5). This is followed by a brief summary of knowledge-based system architecture, discussing the three main components of such systems and how they relate to each other.

There are a number of tools available to a developer when constructing a knowledge-based system. Such tools have evolved in order to simplify the development process; section 3.6 reviews the development tools available. This is followed with a discussion of the selection of a suitable knowledge-based system development tool (section 3.7). Section 3.8 details knowledge acquisition during the development of a knowledge-based system, highlighting sources of knowledge and methods of collection.

The chapter concludes with a review of the use of information technology within contaminated land and also highlights areas for knowledge-based system development in contaminated land investigation (section 3.9).

3.1 Databases

A database is usually described as a collection of data that is stored in such a manner that the data contained within it can be accessed in a range of different ways and formats, and used in an effective way.

A number of authors (Beynon-Davis, 1996; Date, 1995) refer to the term as analogous to a filing cabinet, or more accurately to a series of filing cabinets. Hence the database is a structured repository for data. The overall purpose of such a repository is to maintain data for some set of organisational objectives. Bamford and Curran (1987) take the approach that the term database relates to the combination of physically stored data and the software required to allow that data to be stored.

The database approach offers a number of potential advantages compared to traditional file approaches. Date (1995) outlines the benefits as follows;

- Redundancy can be reduced
- Inconsistency can be avoided
- The data can be shared
- Standards can be enforced
- Security restrictions can be applied
- Integrity can be maintained

3.2 Database Models

Database management systems are generally based on one of four different data architectures. The data model specifies the way in which data are structured and manipulated within the database. The structural component of the model defines how data are represented (e.g. tree, tables etc.). The manipulative component of the model outlines the standard operations such as print, search, add and so on. The models have been developed for a range of environments from large mainframe computers to personal computers. The four approaches are :

a) The hierarchical model
b) The network model
c) The relational model
d) The object-oriented model

3.2.1 Hierarchical Data Model

The hierarchical model arranges the data into tree-like hierarchies. The structure of the tree is designed to represent the sequence in which the data will be accessed. The tree consists of one or more levels, the top level known as the root level and the lowest level known as the leaf level (database trees are turned upside down). Each level of the structure, except the root level, contains a number of record types. The record type consists of one or more fields in a specified order. Records are linked by branches; this is often described as a parent-child link and can consist of a one-to-many relationship between two record types. The hierarchic approach has a number of constraints, as summarised by Beynon-Davis (1996) :

1) No record occurrence, except a root record, can exist without being linked to a parent record occurrence. This means that :
    a) A child record cannot be inserted unless it is linked to a parent record.

b) Deletion of a parent record causes automatic deletion of all linked child records.

2) If a child record type has two or more parent record types, then a child record must be duplicated once for each parent record.

These problems mean that it is extremely important that the database designer understands the structure and number of levels required, as at the planning stage, unforeseen connections can cause problems in retrieving data.

3.2.2 Network Data Model

The network data model, sometimes referred to as the CODASYL model, uses a two-level tree as its basic data structure (Bontempo & Saracco, 1995).

The model consists of two data structures : record types and set types. The record type is the same as in the hierarchical model, although the fields may be used to store multiple values or to represent a composite of values which repeat. The set type is a description of a one to many relationship between two record types.

The manipulation of data within the model is similar to the hierarchical model. The database-specific functions are embedded using the host language. The functions can usually be split into three distinct groups : data navigation commands; retrieval commands; update commands. The host programming language and database system are usually connected together by a common interface.

Again as with the hierarchical model problems with maintaining the consistency of the database can be difficult. Hussain & Hussain (1991) also suggest that confusion often occurs amongst users due to variations in core concepts of the model.

3.2.3 Relational Data Model

The relational data model is based on set theory and logic. It relies on appropriately defined tables as basic objects of retrieval and update operations. Mayne & Wood (1983) summarise the basic concepts of the relational system as follows:

- within a relational system the table must contain only one type of record. Each record has a fixed number of fields, all of which are explicitly named. The database will usually contain numerous tables, so that different kinds of records are held in different tables;

- within a table the files are distinct, and repeating groups are not allowed;

- each record within a table is unique; there are no duplicate records;

- the order of the records within the table is indeterminate. The records may come in any order, and there is no predetermined sequence;

- the fields within any column take their values from a domain of possible field values. The same domain can be used for many different field types, perhaps in several tables;

- new tables can be produced on the basis of a match of field values from the same domain in two existing tables. The formation of new tables from existing tables is the essence of relational processing.

3.2.4 Object-Oriented Data Model

The object-oriented approach is the most recent development in the database management field, and is closely linked to object-oriented programming languages and concepts.

Generally the model can be described as consisting of a number of records that are represented by entries called objects. The objects store data and provide methods or procedures to perform specific tasks. The objects of similar type are stored together as a class. The class is a template that describes the common characteristics of a set of objects. The structure of the object database model is usually unlike other earlier data models as the object database does not rely on specific ways to structure data. A class or object type can have a range of structures ranging from a linked list, a set, an array and so on, depending on the programmer's preference.

3.3 Review of Geotechnical Databases

The use of information technology in geotechnical engineering developed slowly in the past, although recently there has been more rapid progress. The process of site investigation, by its nature, produces a vast amount of data, which often causes problems with managing the data efficiently. The introduction of information technology within the discipline has assisted in solving this problem, in particular through the development of geotechnical databases. These provide an economic way of storing the large amount of data acquired from a site investigation.

Buller (1964) is widely recognised as introducing the use of computers for the storage and retrieval of geological data. Buller's work involved the development of such a system to store well records for the Department of Mineral Resources in Canada. The system although operational, had a number of problems due to its cumbersome nature and the lengthy time taken for a search. This is due to the fact that a search could involve multiple passes through various sources of data. Even with such problems the system was still an advancement on manual storage and searching of local geological data. It also started a trend for oil and mining companies to turn their attention to developing methods of storing geological data.

However the problem that arose from the early developments was that they used punch card systems, which meant that they could never be used to their full potential. Rhind & Sissons (1971) developed a database for the storage of drift borehole records in

Edinburgh, using a mixture of numerical and free form text storage. The main output from this development was that it allowed layer descriptions and their associated depths to be stored in an accessible manner.

During the late 70's and early 80's further developments came from Berner (1975), Cripps (1978), De Beer & Biggs (1978) and Day et al (1983). Ibrahim (1993) gives a detailed review of these developments. The main problem of these early systems is that they were specially designed for the requirements of a particular user, which made them difficult to use by other interested parties.

The late 1980's saw the introduction of Database Management Systems (DBMS) and procedural languages leading to implementation of more sophisticated databases. Rapier & Wainwright (1987) developed such a system known as Geoshare, which was implemented using the CODASYL Database Management System (DBMS) running on a ICL 2988 mainframe computer. The Geoshare system proved a successful prototype, highlighting the benefits of centralised data storage to the geotechnical community. The system concentrated on efficient data manipulation, retrieval and searching, and also gave the user the opportunity to use free form English within the data fields. This highlighted the need for the system to be accessible to both computer skilled and non-skilled personnel.

Other examples of such systems include; Strata 3 (Greenshaw et al, 1987) using Oracle; Greenwood's (1988) geotechnical database implemented on an IBM PC; gINT (Staten & Caroona, 1992) using the Betrieve data file structure running on a personal computer; SID/GDMS (MZ Associates, 1994); TechBASE (MINEsoft Ltd, Denver, Colorado, USA). Oliver (1994) outlines such systems in detail, also reviewing the development of applications for the production of borehole logs. Oliver (1994) suggests that it is important to note that, whilst such systems are not strictly databases, they do store geotechnical data in data files and hence have led to the foundation of sophisticated geotechnical databases. Such examples include systems produced by Howland & Polanski (1985), Chaplow (1986) and Finn & Eldred (1987).

Commercial software packages have developed even further since the review by Oliver (1994). This has led to a range of systems that usually operate in a Windows environment on personal computers. The systems have sophisticated reporting facilities for the output of borehole logs, graphical displays for laboratory test results, cross sections and contouring diagrams as well as producing costed fieldwork summaries. Appendix 1 reviews a number of such systems.

Another major development within the field of database systems has been the introduction of products such as Dbase, FoxPro, Access and SuperBase which are all PC based database systems. Such systems allow large organisations to produce a range of "in house" systems for storing geotechnical data. Unix platforms have also been utilised by large organisations. Malenke (1991) describes the development of a large management system by the Bureau of Reclamation, Denver, USA. The system supported both the administrative needs and engineering functions (borehole data, soil samples storage) of the Bureau using the Ingres relational database management system. Threadgold (1992) argued that such systems restrict the transfer of data between other systems due to their own specialisation. Work on Geoshare at Queen Mary College, University of London (Day et al, (1983); Rapier & Wainwright, (1987)) tried to solve this problem of sharing data, with the aim of setting up a national database. The British Geological Survey (BGS) have also developed a national data system (Forster & Culshaw, (1990)) which consists of a national borehole index, containing borehole information logged with BGS. The problem with such systems is the commercial value of the data stored. Rodger (1992) reinforces this point, arguing that, whilst a national borehole database would be beneficial to all, data security and commercial implications of such a system can be highlighted as possible areas of complication.

These problems led to the Association of Geotechnical Specialists (AGS, 1994) suggesting that a standard format for exchanging factual ground investigation information was a more realistic approach than that of setting up a national database. The AGS format is now widely used and has been incorporated into the majority of commercial geotechnical data management packages, allowing contractors and consultants to freely exchange data.

Oliver (1994) highlights the problem that, although the geotechnical database systems available are able to store soil descriptions, they are stored as text fields. This makes access to individual parts of the text description difficult and inefficient, requiring processing of the text string in order to abstract any part of the description. To overcome this Oliver (1994) developed GeoTec, a geotechnical database developed as part of a knowledge-based system for interpreting site investigation information. GeoTec was developed within the framework of the AGS format and was implemented using the Ingres relational database management system. To deal with the soil and rock description problem GeoTec contains five extra tables not included within the AGS format; Layer, Structure, Stratum, Constituent, Stratum Structure and Colour (Toll & Oliver, 1995). This type of format allows for the possibility of multiple strata within a layer. This is required because descriptions of layers may often contain more than one stratum, for example SANDSTONE interbedded with SILTSTONE. This example contains two distinct strata within a layer, yet they cannot be distinguished as separate layers (a layer being defined by depth and thickness).

An additional feature of GeoTec was the ability to store structured geotechnical test information into levels. The top level information stored are the interpreted geotechnical parameters from the test. Subsequent levels contain more detailed information on the derivation of the parameters. The system also has the ability to store raw data in an unstructured form, which can include pictures, formatted document files or simple ASCII files.

Although GeoTec could contain a wealth of site investigation data, there were limited facilities for preliminary site investigation data (desk study & walkover information). The Geology table was the only such inclusion. It allowed storage of details of stratigraphic information (such as geological horizons) that have been obtained from the desk study.

3.4 Database Software Selection

On consideration of the various data models (section 3.2) it was concluded that a relational data model would be the best technical solution to implement the database system. This is primarily because such a model can meet the requirements of dealing with potentially large quantities of data and the required multi-user capability. The relational model also has greater flexibility over other models, due to the fact that it is very easy to alter. New tables and records may be added at any time without disrupting the database schema.

Before selecting an appropriate software development tool, an important factor to consider was the hardware platform on which the software was to be developed. On examination of the hardware platforms available within the market place, it became apparent that there were two main options available, either a Unix Workstation or a personal computer. In order to assist in selecting the most suitable platform a list of requirements was compiled. These included :

(1) Able to support a range of commercial software packages, ranging from database development software to knowledge-based system shell software.
(2) To have PC multi-user capability via networking
(3) To have ample storage facilities
(4) To be common place in the majority of engineering design offices
(5) To lie within the financial constraints of the project both in purchasing and maintaining the hardware.

Taking into account the requirements highlighted previously, it was concluded that a personal computer platform would be the best tool to use for the development of the proposed software. Advances in personal computer technology have resulted in such computers offering facilities that had only been previously available on workstation platforms. The cost of personal computers has also fallen rapidly in recent years making workstation technology expensive in comparison. This has in turn resulted in most geotechnical design offices using personal computers and therefore the end user (geotechnical engineers) of the proposed system will be familiar with this platform.

Using a personal computer platform also required selection of an operating system. This was a relatively straight forward choice, as the majority of database software packages available run within a Windows environment. It is usually only pre 1992 software that ran within the MS-DOS environment. Therefore the Windows environment was chosen, in particular the Windows 95 operating system.

Once the fundamental decisions regarding hardware platform and database model had been decided, it was necessary to select a suitable relational database package to use. At the time of choosing the software, there were a number of relational database packages available. These included Paradox, FoxPro and Microsoft Access, any of which were suitable for this application. After considering the advantages and disadvantages of the packages available it was decided to use Microsoft Access.

Microsoft Access version 2.0 is a relational database management system for creating Windows 3.1 (and higher) desktop and client-server database applications (Jenning & Person, 1994). It provides the developer with a range of tools that can be used to create a powerful relational database system. The point-and-click and drag-and-drop capabilities make creating user interface forms very easy. It also has the facility to create modules and macros using the Access Basic programming language. It also allows data tables, indices, queries, forms, reports, macros and Access Basic code modules to be stored within a single database file. The facility to import data from and export data to other applications such as FoxPro, Paradox etc., is another useful feature.

In addition to the tools provided, Access is generally supplied as part of the Microsoft Office package and the majority of commercial organisations tend to use this suite of packages. Therefore even if the end user is not familiar with Access, they are still likely to be at ease with the Microsoft Windows environment. This was seen as a positive factor as it is likely to encourage end users (geotechnical engineers) to use the database. The popularity of this software is also reflected in the fact that Access is the recommended database package for Durham University use. This meant that full technical support was available, which was seen as a major advantage during development.

3.5 Knowledge-Based System Components

The field of artificial intelligence within computer science is involved with the development of computer reasoning, in particular pursuing the possibility of computer reasoning in a similar manner to humans. Artificial intelligence includes a number of areas; robotics, cognitive modelling, artificial neural systems, speech and knowledge-based systems.

The field of knowledge-based systems within artificial intelligence makes use of specialised knowledge to solve problems at a similar level to that of human experts. The terms expert systems, intelligent assistants or knowledge-based expert systems are also often used when referring to this type of application. However, Adeli (1988b) suggests that due to the fact that very few true "expert systems" exist, it is more appropriate to use the term knowledge-based system. Within this thesis the term knowledge-based system is therefore used.

Various definitions regarding knowledge-based systems exist within the technical literature, ranging from lengthy descriptions to simple statements. Ibrahim (1993) presents a detailed description of the definitions. One of the earliest developers, Feigenbaum (1981), defines an expert system as " An intelligent computer program that uses knowledge and inference procedures to solve problems, that are difficult enough to require significant human expertise for their solution". Giarratano & Riley (1988) suggest that a more meaningful way to define knowledge-based system technology is to examine where it differs from conventional programming (Table 3.1).

Characteristic	Conventional Program	Expert System
Control by	Statement order	Inference engine
Control and data	Implicit integration	Explicit separation
Control strength	Strong	Weak
Solution by	Algorithm	Rules and inference
Solution search	Small or none	Large
Problem solving	Algorithm is correct	Rules
Input	Assumed correct	Incomplete, incorrect
Unexpected input	Difficult to deal with	Very responsive
Output	Always correct	Varies with problem
Explanation	None	Usually
Applications	Numeric, file, and text	Symbolic reasoning
Execution	Generally sequential	Opportunistic rules
Program design	Structured design	Little or no structure
Modifiability	Difficult	Reasonable
Expansion	Done in major jumps	Incremental

Table 3.1: Typical Differences Between Conventional Programs and Expert Systems. (Source: Giarratano & Riley, 1998)

3.5.1 Architecture of Knowledge-Based Systems

The architecture of a knowledge-based system consists of three main components; knowledge-base, working memory and inference engine.

These three components of a knowledge-based system try to mimic the human expert. Durkin (1994) illustrates how the components relate to each other and to the human expert (Figures 3.1, 3.2).

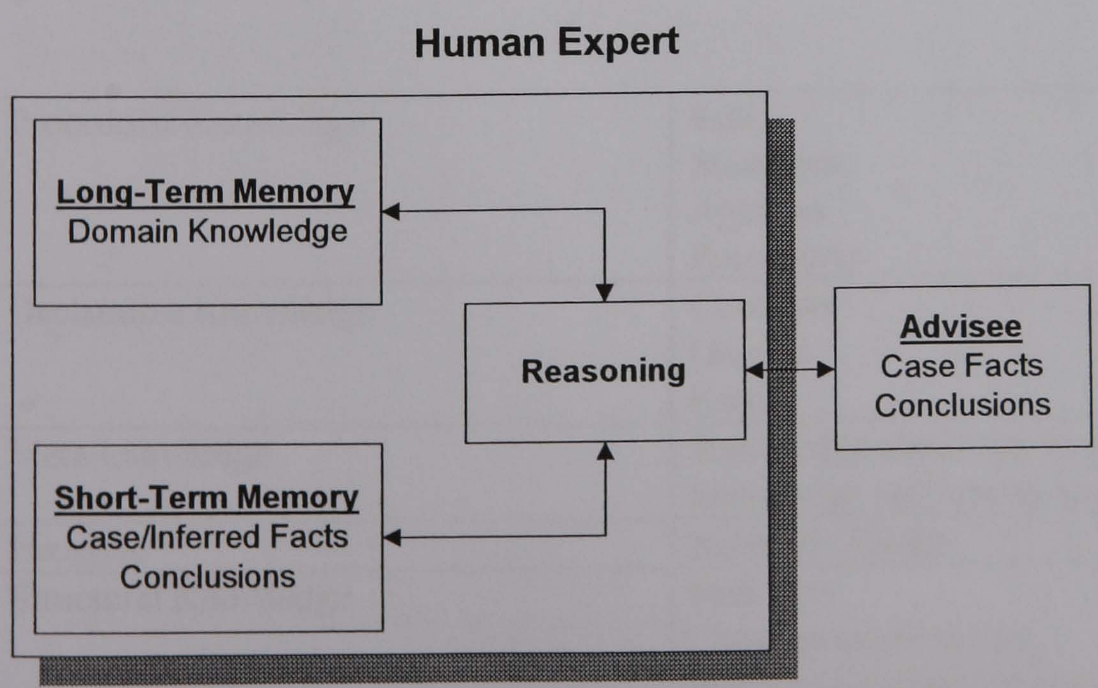


Figure 3.1 Human Expert. (Source: Durkin, 1994)

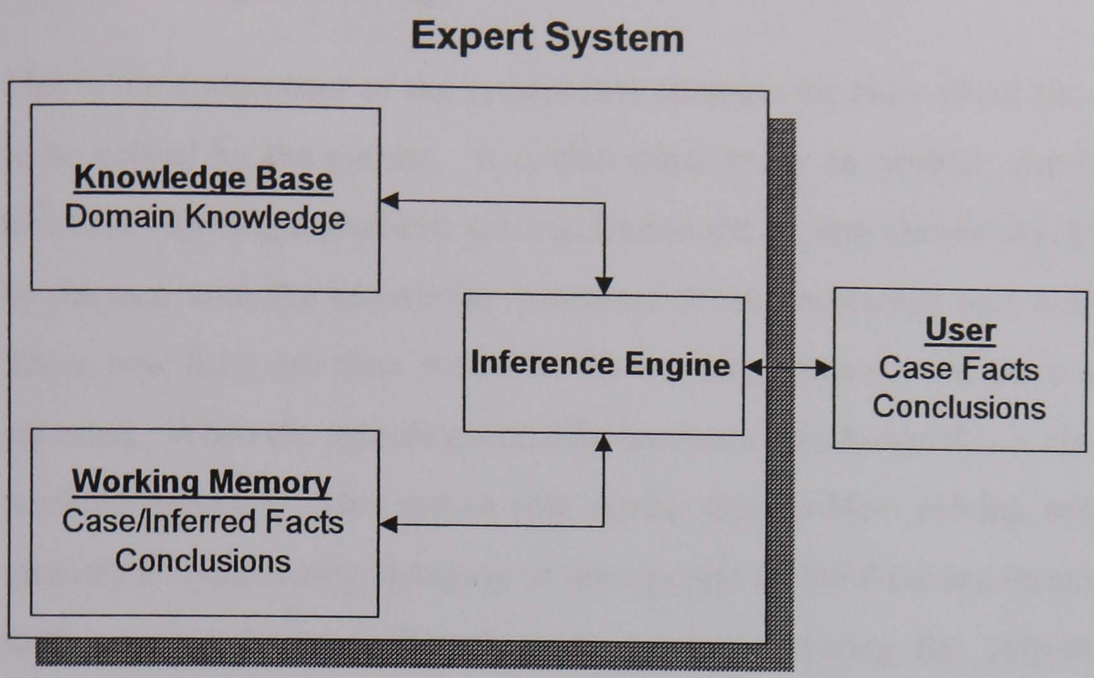


Figure 3.2: Expert System. (Source: Durkin, 1994)

3.5.2 Knowledge Base

The knowledge base is the component that contains information about the particular field under consideration by the system. This is usually known as domain knowledge. The knowledge is most commonly represented using sets of rules (If-then statements). However, knowledge may also be represented by documented definitions, facts, heuristics and concept relationships. A summary of different types of knowledge is given in Table 3.2 (from Durkin, 1994).

Procedural Knowledge	Rules Strategies Agendas Procedures
Declarative Knowledge	Concepts Objects Facts
Meta-knowledge	Knowledge about the other types of knowledge and how to use them.
Heuristic	Rules of Thumb
Structural Knowledge	Rule sets Concept relationships Concept to object relationships

Table 3.2: Types of Knowledge. (Source: Durkin, 1994)

3.5.3 Working Knowledge

This is the component of the system that contains the facts about the problem which is to be solved by the system. It is also often know as context, short-term memory or fact base. During a problem solving session the system checks the information entered by the user with the knowledge contained in the knowledge base and infers new facts. These new facts are then stored in the working memory and the matching process is repeated. When the system eventually reaches a conclusion this is also entered into the working memory. This means that during any problem solving session the working memory is dynamically changing to incorporate all the facts and intermediate results, as well as the solution. Therefore, at any point during the session, the amount of information stored within the working memory reflects the state of the problem currently being solved by the system. The working memory may also load information in at the beginning of a session from an external source (e.g. a database or spreadsheet).

3.5.4 Inference Engine

The inference engine (sometimes known as the reasoning mechanism or control mechanism) is the component of the knowledge-based system that specifies the reasoning process of the system. It examines known facts and beliefs and, if possible, derives new facts and beliefs in order to solve the problem the system is dealing with. There are two main inference strategies for rule-based system. These are "Forward Chaining" (data driven) and "Backward Chaining" (goal driven).

3.5.5 User Interface

Generally, the interaction between a knowledge-based system and the user is conducted through a natural language style interface. This, requires the interface to be well designed, in order for reliable information to be obtained from the user. This may be the only access the knowledge-based system has to the information required. The user interface therefore allows the user to communicate with the system. It is important that the interface expresses clearly to the user what information is required by the system and is simple to use, as well as being aesthetically pleasing.

Understanding how humans interact with such interfaces has been widely researched in the past, and areas that should be avoided or encouraged have been outlined in these studies. Card et al (1983) suggests several principles that should be considered when designing an interface;

Consistency - similarity of patterns and in presentation of information - consistency reduces human learning load and increases recognition by presenting familiar patterns. The human mind is excellent at pattern matching.

Compatibility - new designs should be compatible with, and therefore based upon, the user's previous experience.

Economy - interface designs should reduce the number of operations required by the user to a minimum and lessen the work of the user whenever possible.

Adaptability - interfaces should be able to adapt to different levels of user, from speed of operation through to the skill level of particular users. When this is not possible the interface should be clear and concise, not laborious for the experienced user yet clear for the novice.

Guidance not control - interfaces should guide the user through a set of terms and inform and instruct in the process. The interface should function at the user's pace according to the user's command and should not attempt to control the user.

Structure - interfaces should be designed to reduce the complexity of a given framework. Information should be presented and organised so that only relevant information is passed to the user in a simple manner.

3.5.6 Other Components

In addition to the three main components, there are a number of other components that are usually associated with knowledge-based systems. These include: explanation facility and a knowledge acquisition component.

3.6 Development Tools for Knowledge-Based Systems

The development of a knowledge-based system involves capturing knowledge from a particular domain and creating an inference procedure. Selecting the correct tool is an important first step in the development of such an application.

Since the early developments within the field of knowledge-based systems, a number of tools have evolved in order to simplify the task of construction. When developing a system a developer has three options; General Purpose Programming Language (GPPR); General Purpose Representational Languages (GPRL), and Knowledge-Based System shells (KBSs).

3.6.1 General Purpose Programming Languages

The first option, general purpose programming languages (GPPL), includes conventional procedural languages such as Pascal, Fortran and C. Such languages are designed for numerical algorithmic computation, and so are more applicable to solving mathematical, engineering and scientific problems. Therefore, they do not provide the most appropriate environment for the development of knowledge-based systems. However, Adeli (1987) suggests that a number of knowledge-based systems have been developed in procedural languages since they offer easy portability between different types of computer and compatibility with numerous pieces of software available in these languages. Procedural languages may also be suitable for producing rule-based systems.

3.6.2 General Purpose Representational Languages

The second option, General Purpose Representational Languages (GPRL), includes symbol manipulation languages which have been developed for use in building knowledge-based systems. The information within these languages is normally represented in a descriptive form rather than a numerical system like other languages. These are often known as AI languages; popular examples of such languages include LISP (LISt Programming), and PROLOG (PROgramming in LOGic).

The descriptive nature of these languages provides a greater flexibility in implementing knowledge-based systems. Vamos (1998) suggests that such languages remain the first choice of designers, although recently object-oriented languages, such as C++, have become popular among system developers. The basic idea behind such languages, is to program with objects. Each object is defined by data specific to it (its characteristics) as well as by the operations and computations that it is able to execute when a message is sent to it. The inheritance property of object-oriented languages makes them suitable for knowledge-based systems using semantic networks or frames.

3.6.3 Knowledge-Based System Development Shells

Early developments within the field of knowledge-based systems usually meant creating systems from scratch, using some form of programming language. As the development of such systems increased it soon became apparent that the systems often had a lot in common. Generally the system consisted of a set of declarative representations (rules) combined with an interpreter for the representations. This meant that it was possible to separate the interpreter from the domain specific knowledge. This allowed new systems to be created by simply changing the knowledge held within the system and replacing it with knowledge that corresponded to the new problem domain.

These interpreters are generally known as shells. Rich & Knight (1991) suggest that one of the most influential examples of such a shell is EMYCIN (for Empty MYCIN) (Buchanan & Shortliffe, 1984), which was derived from MYCIN, the rule-based expert system developed at Stanford University to aid physicians in diagnosing and treating patients with infectious blood diseases caused by bactereia and meningitis. A number of commercial shells are now available, providing an established environment for creating systems. The shells usually provide mechanisms for knowledge representation, reasoning and explanation. These also provide knowledge acquisition and user interface development facilities. Another important feature is the ability to integrate knowledge-based systems with other kinds of programs, as operating such systems within a vacuum limits their capabilities.

Fully integrated systems allow access to commercial database systems and also enable the systems to be embedded within larger application programs that use primarily conventional programming techniques. Both of these features greatly enhance the efficiency and data storage of the system, as well as providing an easy-to-use interface between the larger program and the shell.

3.7 Knowledge-Based Systems Software Selection

There are a number of commercial knowledge-based system development tool "shells" available, all of which are suitable to run on the chosen hardware platform (personal computer). The selection of suitable software can therefore be an extremely difficult task, as selecting the wrong shell can result in the production of an expensive unuseable end product. Vedder (1989) outlines the factors that should be considered; these include:

(1) Flexibility of knowledge representation
(2) Variety of interface mechanisms and their control
(3) Ease of use
(4) Editing, tracing and debugging aids
(5) Explanation facilities
(6) Interface to other applications
(7) Uncertainty management
(8) Support and consultancy services

(9) Hardware requirements

(10) Price (runtime and development environment).

Due to the large number of factors that require consideration it becomes apparent that it is not possible to meet all the requirements with one development shell. Citrenbaum et al (1990) therefore suggest that a shell should be compatible and portable in knowledge representation with other products, so knowledge-based systems can be transported to the most appropriate tools for each user in each task.

CLIPS version 6.10 was selected for this research project, and installed on a P100 personal computer. The selection was made for a number of fundamental reasons, principally the fact that CLIPS is not a commercial product and not copyrighted. It can, therefore, be freely distributed, which is an important feature if the final package is to be used by other organisations. Other software development tools, that would require the end user to purchase a copy of the development tool, could alienate end-users due to the cost which can reach thousands of pounds in licence fees etc. The development tool usually acts as the underlying software which is required to run the main package.

However, one major disadvantage of CLIPS not being a commercial product, is that the support service is not as good as that available with commercial software products.

The other main reason for the selection of CLIPS is that it is a general purpose, development environment. It comes as source code, which allows it to be expanded to deliver additional capabilities, and has the ability to execute external programs written in any language. This, therefore, satisfies the criteria of Citrenbaum et al (1990), that the development tool should be portable and compatible with other products.

Although CLIPS is a useful development tool, it did not provide facilities to develop a user friendly interface. Therefore, Visual Basic© version 4 was chosen as the tool for developing an interface in order to acquire the information required by the system from the user in an effective and efficient manner. Visual Basic© allowed a Microsoft Windows interface to be produced quite simply. This was seen as a positive factor as most end-users are likely to be familiar with such an environment, which conforms with the majority of commercial software package.

3.7.1 Overview of CLIPS

CLIPS (C Language Integrated Production System) is an expert system tool that provides support for rule-based, object-oriented, and procedural programming. It was developed at NASA/Johnson Space Centre using the C programming language, and was designed with the specific purpose of providing high portability, low cost, and easy integration with external systems.

Giarratario & Riley (1998) suggest that the inferencing and representation capabilities provided by the rule-based programming language of CLIPS are similar to, but more powerful than, those of OPS 5. It only supports forward chaining rules, but can emulate backward chaining. The procedural language within CLIPS is syntactically similar to LISP and has similar features to languages such as C, Pascal and Ada. CLIPS operates on many platforms including IBM-PC compatibles, HP, Sun and Macintosh.

3.8 Knowledge Acquisition

When developing a knowledge-based system it is important that the type and source of information is of the highest possible standard. There are generally two main sources of knowledge that may be drawn upon during the knowledge acquisition phase. These are technical literature and domain experts.

Technical literature generally includes published literature such as technical reports, conference proceedings, journals and textbooks. Obtaining suitable knowledge from such sources is relatively straight forward although usually involving an extensive literature study.

The type of knowledge collected from domain experts is often known as private knowledge, as experts usually acquire their expertise through experience of working on similar problems. This allows the human expert to make educated decisions, and deal with incomplete data. This can be particularly useful for ranking rules and applying certainty to information within rules. Therefore, it is possible to derive rules from technical literature and use domain experts to validate such rules. This suggests that it is critical that the appropriate domain experts are identified and involved in the knowledge acquisition stage of the system development.

Extracting knowledge from experts may be problematic, due to the way experts access their problem-solving knowledge in order for them to solve problems efficiently. Durkin (1994) suggests that this becomes apparent when experts are asked to describe their problem-solving methods, as they will often make mental leaps over important issues and have difficulty in explaining the knowledge used in detail. Water (1986) labels this dilemma as the knowledge engineering paradox, "the more competent domain experts become, the less able they are to describe the knowledge they used to solve problems." Other problems include experts providing incorrect knowledge. This may be either because the expert is uninformed or due to a simple mistake. Also experts often provide irrelevant knowledge when questioned.

In order to extract knowledge from domain experts the developer has a range of methods available. One method used generally involves compiling a questionnaire and mailing it to suitable domain experts. Miller (1991) suggests, that such a method has a number of advantages and disadvantages, as highlighted.

Advantages of mail questionnaire

(1) Permits wide coverage for minimum expense, both in money and in effort

(2) Affords wider geographic contact

(3) Reaches people who are difficult to locate and interview

(4) Greater coverage may yield greater validity through larger and more representative samples

(5) Permits more considered answers

(6) More adequate in situations in which the respondent has to check information

(7) More uniformity in the manner in which questions are posed

(8) Gives respondent a sense of privacy

(9) Affords a simple means of continual reporting over time

(10) Lessens interview effect

Disadvantages of mail questionnaire

(1) The problem of non-returns must be addressed
    (a) Response rates to mail questionnaires usually do not exceed 50% when conducted by private and relatively unskilled persons
    (b) Intensive follow-up efforts are required to increase returns

(2) Those who answer the questionnaire may differ significantly from non-respondents, thereby biasing the sample.

If the questionnaire method is used during the knowledge acquisition phase, it is essential that the format of the questionnaire is suitable to collected the required knowledge, and allows the developer to analysis the knowledge easily. Sekaran (1992) suggests three principal areas that require consideration when designing a questionnaire; (1) wording of the questions, (2) how the variables will be categorised, scaled or coded and (3) the general appearance of the questionnaire.

Questions may be worded in two formats, giving varying results. An open format allows the respondent to answer questions freely without any restrictions on the answer they may choose. Jankowicz (1992), recognises the resulting data from this format as being rich data but also "disorganised data". This type of question requires the researcher to categorise, process and analyse the replies. The process of analysing data gathered from open questions can therefore be very time consuming and difficult. The other format, the closed question format, forces the respondent to select from answers provided in the questionnaire, which have been determined in advance. This allows the questionnaire to be designed in a standardised format, which allows much simpler and less time consuming analysis of results.

3.9 Use of Information Technology within Contaminated Land

Information technology is widely used within the field of civil and geotechnical engineering for a range of tasks, from design packages that assist with foundation design to data management systems that produce high quality graphical outputs such as borehole logs and laboratory results. Such data management software also plays a vital role during site investigations, as it allows the vast amount of data collected to be manipulated, validated and analysed. Bond (1995) reviews geotechnical design software and examines the factors affecting the quality and validity of the software. The introduction of a standard format for exchange of ground investigation data (AGS, 1994) has encouraged the use of information technology within geotechnical engineering, by eliminating the problems caused when data are transferred between parties involved with an investigation (contractors and consultants).

In many cases the same software can be used for the investigation of contaminated land. The Department of the Environment (1994f) gives guidance to the use of information systems for land contamination. The guidance covers:

- ♦ Consideration of the types of organisation which may need to compile or make use of information on land contamination, and the ways in which they may use it;

- The types of information which may be useful for such organisations to hold; and

- The management of information, including recommended actions on quality control.

The computerisation of investigation data concerned with contaminated land offers a number of advantages relating to manipulation and presentation of data as well as improving the efficiency, flexibility and accessibility of stored data.

One branch of information technology that is slowly developing in the field of geotechnical engineering is the application of knowledge-based systems, although Durkin (1994) states that, in engineering in general, the development of such systems is rather limited compared with other disciplines. Toll (1990) suggests the geotechnical specialists could operate more widely if they could make use of other people's expertise, and that knowledge-based systems can be an effective means of disseminating this knowledge.

Moula et al (1995) reviewed the knowledge-based systems available within the field of geotechnical engineering, concentrating mainly on soil engineering applications developed up until 1993. They suggest that more systems are likely to be developed to a commercial stage over the next decade. Toll (1996) updated the earlier review including more rock engineering applications, concluding that many systems are still simple prototypes although progression beyond this point was starting. As regards the development of tools for assisting with the investigation of contaminated land, it has been extremely limited. Some of the systems highlighted by Toll (1996) can aid the investigation, although they are not designed specifically for contaminated land. In addition to the systems highlighted by Toll (1996), Law et al (1986), Heynisch et al (1994), Tucker et al (1997) and Kelly and Lunn (1998) have also contributed to the development of knowledge-based system software.

Adeli (1988a) described the development of one of the earliest systems developed by Law et al (1986). The aim of this system was to classify inactive hazardous waste sites in terms of the level of groundwater contamination, surface-water pollution and air pollution. The knowledge base within the system included rules and facts documented in handbooks as well as rules of thumb obtained from the experts in the field. The US Environmental Protection Agency (EPA) hazardous rankings system was used to score and rank the sites for their potential to cause health, ecological or environmental problems. The ranking system was represented within the system as a series of production rules. Although the hazard ranking system divided the assessment of groundwater migration into four groups (route characteristics, containment, waste characteristic and the site environment) Law et al (1986) concentrate on the first group which included permeability, groundwater flow direction and gradient.

Heynisch et al (1994) developed a knowledge-based system (HYDRISK) that evaluated hydrogeological properties and chemical criteria relevant to contaminant transport. The system focused on highlighting which groundwater pathways must be protected and recommended possible future land uses for the area under investigation. A risk classification was worked out for factors that have an influence on the climatic water balance and hence the groundwater (surface flow, water balance in soil, effective precipitation etc). Due to the fact that the site and its geological and hydrological properties are spatially related, the system defined them in terms of a geometric model. The model was composed of horizontal layers. The site was also gridded and thickness and further depth related criteria were represented as parameters belonging to grid cells. During evaluation, each spatial unit was assessed for its associated attributes and was ranked. The various attributes were then replaced by risk values. The system was written in the computer language C, and was supported by a geographic information system (GIS) which enabled mapping of the spatially distributed properties and results of evaluations.

Tucker et al (1997) developed an expert support system Site ASSESS (Assessment of Sampling Strategies Expert Support System), for assisting site assessors when compiling preliminary investigation information and developing an initial hypothesis on the likely locations of hot-spots on a contaminated site. The system was based upon expert knowledge that was distilled and structured as a series of numerical coefficients.

Tucker et al (1997) suggested that the coefficients were a "snap-shot" of current knowledge, and as knowledge increased through a better understanding of contamination indicators, it was expected that these coefficients would be improved. The knowledge base converted desk study information into a score of indicators in order to produce a prior probability map of hot-spot locations. The total number of sample locations was then computed and distributed over the site to reflect the prior information and hot-spot specification.

Kelly & Lunn (1998) developed a prototype Contaminated Land Assessment System (CLASS) within the framework of the geographic information system ARC/INFO, in conjunction with Newcastle City Council. The system assisted in predicting pollution migration using a source-pathway-target approach to rank past and present industrial land. The system comprised two main components; first a database for identification and characterisation of contaminant sources, pathways and targets within the Newcastle area; and the second, a hazard modelling system to classify each site in terms of its pollution potential. A hazard index was determined based upon an estimation of the distribution of chemical travel times to near-by surface water targets. The index was derived using five physical and chemical attributes; water travel time; contaminated site area; sorption; persistence and toxicity.

3.9.1 Areas for Knowledge-Based System Development in Contaminated Land Investigation

As highlighted previously very few have tackled the development of knowledge-based systems within the subject area of contaminated land investigation. This illustrates the potential for development in this area.

On examination of the processes involved within the investigation of contaminated land, (outlined in section 2.6.2 - 2.6.3), key elements can be identified. These are, (1) the investigation requires knowledge from a number of disciplines and (2) a structured multi-stage approach is required during the investigation, in order to reassess the needs of the investigation as more information is gathered. Both of these elements lend themselves to the knowledge-based system concept. A knowledge-based system has the ability to hold information from a number of subject areas, as well as structuring the entry of data, and allowing reassessment after input of new data.

This obviously makes knowledge-based system technology an ideal tool for use within such an investigation. The stage of an investigation that would benefit most from such technology is the preliminary stage, as it plays an essential role in identifying potential problem areas of the site and likely contaminants before the exploratory stage of the investigation starts. This is important as, due to the expense involved in an investigation, it is not economic or feasible to examine all areas of the site in detail. So, having prior knowledge about the site reduces the risk of encountering unforeseen hazards. Unfortunately investigators often overlook the collection of such information or do not use it to its full potential.

3.10 Conclusion

Database management systems are generally based on one of four different architectures. These are, the hierarchical model, the network model, the relational model and the object-oriented model. The object-oriented model is the most recently developed, although generally the relational model is the most commonly used model within database management systems.

Along with the development of data models, personal computer technology has also developed tremendously. This has resulted in database development tools for personal computers becoming extremely sophisticated, and hence increasing the use of database systems within a number of subject areas.

One such area that has benefited from this development is the area of geotechnical engineering. Database systems provide an extremely useful tool, to store and manipulate the vast amount of data that is collected during a site investigation. The development of the AGS Data Exchange Format has also encouraged the use of such technology, as it allows ground investigation to be transferable between parties. This had been difficult in the past, as the early geotechnical database systems had been restricted to particular users.

The Microsoft Access Relational Database Management System was selected to implement a database, as it allowed a user interface to be developed within a Windows environment. This was seen as a positive factor as it is likely to encourage end users to use the database, due to the fact that most design offices use Windows based software.

Knowledge-based systems are computer programs that contain domain knowledge stored within their knowledge base as sets of rules (IF-then statements) or as documented definitions, facts, heuristics or concept relationships. A separate inference procedure (inference engine) is usually employed to manipulate knowledge in order to solve the defined problem. The manipulation of data usually consists of checking the information input by the user against knowledge within the knowledge base and inferring new facts. The new facts are stored within the working memory of the system and the matching process repeated until the problem has been solved. The user generally communicates with the knowledge-based system via a natural language style interface. It is important that the interface expresses clearly to the user what information is required by the system and is simple to use, as well as being aesthetically pleasing.

The tools that are available for developing a knowledge-based system can be divided into three main categories; General Purpose Programming Languages (GPPR), General Purpose Representational Languages (GPRL), and Knowledge-Based System Development Shells.

During the development of a knowledge-based systems, there are generally two sources of knowledge to draw upon, these being technical literature and domain experts. The most effective and popular method of obtaining knowledge from domain experts is via a questionnaire format.

The development of such systems within the field of contaminated land has been very limited. CLIPS development software was selected as a tool to develop a knowledge-based system to assist with the investigation of contaminated land and in particular the preliminary stage of the investigation, making full use of desk study and site reconnaissance data.

CHAPTER 4

DATABASE DEVELOPMENT

4.0 Introduction

The need to design a database to store preliminary site investigation data was an important part of this research project, as this would allow the end user to store such data independent of the knowledge-based system. It was also seen as a major development in the area of geotechnical databases, as the storage of preliminary investigation data is not addressed by existing geotechnical database systems.

The design history of this database system is discussed in this chapter, starting with the introduction of the design process (section 4.1). This is followed with a description of the purpose of the database in section 4.2. Section 4.3 then highlights the implementation of the designed data structure. The chapter concludes with a discussion of the design of the user interface.

4.1 Database Design

As discussed in Chapter 3 a database provides a useful tool for the manipulation of large volumes of data. This therefore fits well with the needs of storing preliminary investigation data. However, to achieve this, it was essential that a clear plan was developed from the onset of the design process. To develop this plan involved examining a number of factors; these include;

- Understanding the purpose of the database
- Assessing the type and volume of data to be stored
- Selecting an appropriate database model and hence designing a data structure which fits the requirement of the data to be stored

- Choosing development software that allows the data structure to be implemented in a logical manner
- Considering the best procedures for data input, including the design of a suitable user interface

Failure to consider such points is likely to result in poor storage and therefore creating problems with processing and data access, which in turn reduces the processing speed and counteracts the main advantage of using modern database technology.

As regards the geotechnical database design, Greenwood (1988) suggests that computer systems used for storing and retrieving such data should incorporate the following features:

(1) Data input once only
(2) Data not constrained within any particular database or spreadsheet format
(3) No restrictions on the extent of data storage
(4) Data should be readily transferable between different computer systems
(5) Accessible by contractors for the preparation of reports, by engineers for analysis and cross-referencing of gathered data.

4.2 The Purpose of the Database

In order to achieve a well structured database it was important that all the requirements of the database system were identified at the start of database design. The main requirements of the proposed system were outlined as;

- Preliminary Investigation data storage
  Data to be stored will essentially be preliminary investigation information, detailed in sections 2.6.2.2 to 2.6.2.4. Enforcing a structured approach to data input encourages the user to undertake a full and structured preliminary investigation (this includes both desk study and site reconnaissance).

- AGS format

  A similar structure and format to that used for the AGS (1994) data transfer format. This allows data to be placed into groups within the structure and uses key fields to identify each group. It allows links to be forged between existing AGS tables and newly designed preliminary investigation data tables. This also makes data available for transfer between parties involved in projects relating to the redevelopment of contaminated land.

- Data Input

  Data input via a user friendly interface within a Windows environment, thus allowing data to be input by non-computer experts. The data entry should also be either using a network system (multi-user platform) or via a single stand alone personal computer.

- Allow links to knowledge-based system

  The data within the database needs to be accessible for the knowledge-based system to use. Therefore it is important to allow the user to input data either by direct entry into the knowledge-based system or via the database for the knowledge-based system to use at a later date.

- Data manipulation

  Data should be available for use with other packages that the user may require. Other packages may include GIS packages (e.g. ArcView) or other commercial geotechnical packages.

4.3 Implementation

A "top-down" approach (Malenke, 1991) to design and implementation was adopted. The "top-down" approach starts with more general requirements for the database and gets progressively more detailed as the final design is reached. By contrast the "bottom-up" approach starts detailed and develops towards a more general concept or design. For the purpose of this design process a "top-down" approach was decided to be the best solution, as it seemed logical to highlight the main areas that needed to be considered for storage, and to work down through the subject areas, pin-pointing the detailed areas that required consideration. In contrast, the “bottom-up” approach may result in some subject areas being overlooked, especially if the detailed level has not been fully completed.

The first stages in the implementation involved defining the data entities to be stored in the final data structure. The entities were selected using technical literature (detailed in section 2.6.2.2). The entities consist of general types of data that need to be dealt with in the database, such as geology, topography etc.

On completion of identification of the relevant entities their inter-relationships and attributes were identified. An Entity-Relationships, E-R, diagram was constructed, allowing the relationships between entities to be clearly defined. The entities typically have "one-to-many" relationships that is, one record in an entity could possibly relate or join with many records in another entity. For example, entities known as project and site were identified. The site entity contained information relating to the area under investigation, such information included site address, owner etc. The project entity contained data regarding project name, project client etc., therefore it was concluded that within a large project there may be a number of sites. Hence project can have a one-to-many relationship with site. This process also helped to eliminate undesirable relationships that may occur. These included many to many relationships; in this case one of the entities was decomposed into two entities. This resulted in two of the entities showing a "one to many" relationship with the new entity.

After the entity relationships had been established the attributes (properties possessed by an entity) were identified. A normalisation process was undertaken in order to identify individual tables from the entities and reduce the level of duplication and redundancy to a minimum. This led to each entity being translated into an individual table, although, in certain cases, entities were broken into two or more tables, depending on how general the entities were. For example, an entity known as hydrology was initially identified. On examination it became apparent that this was too general and therefore was split into tables “groundwater” and “surface water”.

Once individual tables had been derived from the entities, fields within the tables were identified along with referential keys required to link the tables. The referential key consists of a single or multiple field that uniquely identifies that table.

4.3.1 Data Structure

From the technical literature (Department of the Environment 1994 (c & d) and British Standard DD175 (1988)) regarding site investigation of contaminated land, it was clear that there were a number of subject areas that are usually taken into consideration (detailed in section 2.6.2.3). These areas are: topography, geology, hydrology, services, geography, history, fauna, meteorology and vegetation. The identification of these areas made it possible to split the subject areas further and form a relational data structure, with such a data structure allowing the storage of preliminary information in an electronic format.

An outline scheme for the database is shown in Figure 4.1. The boxes represent tables within the relational database structure. The tables within the structure are data groups that represent the parameters required for the preliminary investigation. The structure allows potentially large volumes of data to be retrieved, searched and handled in an effective manner. The names of the tables have been adopted to be compatible with the AGS format. A full list of the database tables is given in Table 4.1. The details of all the database tables are outlined in Appendix 2.

Legend	Table Name	Legend	Table Name
PROJ	Project	GRDW	Groundwater
SITE	Site	TOPO	Topography
ZONE	Zone	GEOG	Geography
PREL	Preliminary Investigation	METE	Meteorological
VEGE	Vegetation - General	VEDT	Vegetation - Detail
FAUA	Fauna – General	FADT	Fauna - Detail
SERV	Services – General	SEDT	Services - Detail
SURW	Surface Water – General	SUDR	Surface Water Drainage
GEOL	Geology – General	SUST	Surface Water Storage
STFT	Geology Structural Features	GEDT	Geology - Detail
HIST	History – General	HIDT	History - Detail

Table 4.1: Legend for Database Structure.

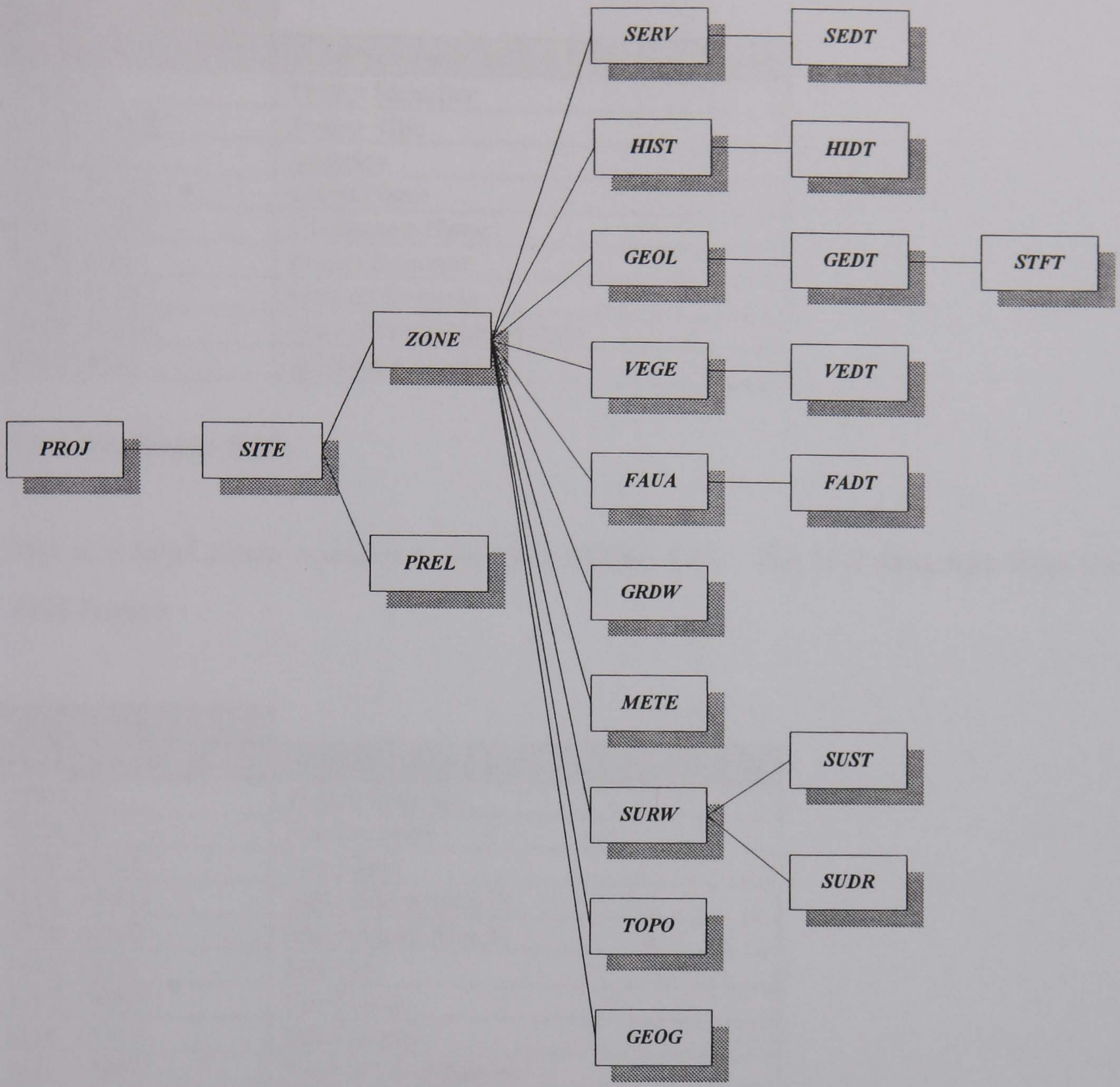


Figure 4.1: Schema for Database. (see Table 4.1 for Legend)

There are six identifiable levels within the structure, with the top-level table being the project table. This is one of the original AGS format tables; it contains information on the location and date of the project and the parties involved. Each project will therefore have its own table, which is identified by a *Project ID* key (Table 4.2). Using this table also allows the preliminary investigation information to be linked to the ground investigation data, as both types of data could be assigned to the same project table.

PROJ	
Field Name	**Field Description**
PROJ_ID	Project Identifier
PROJ_NAME	Project Title
PROJ_LOC	Location
PROJ_CLNT	Client Name
PROJ_CONT	Contractors Name
PROJ_ENG	Project Engineer
PROJ_REM	General Remarks
PROJ_DATE	Date of Production of Data
PROJ_AGS	AGS Issue Number

Table 4.2: Project Table.

The next level down contains a site table (Table 4.3). This is a departure from the AGS format.

SITE	
Field Name	**Field Description**
PROJ_ID	Project Identifier
SITE_ID	Site Identifier
SITE_NAM	Site Name
SITE_ADD1	Site Address (line 1)
SITE_ADD2	Site Address (line 2)
SITE_CITY	Site City
SITE_CONT	Site County
SITE_COTR	Site Country
SITE_CORT	Type of Co-ordinates
SITE_XCOR	X-Co-ordinates
SITE_YCOR	Y-Co-ordinates
SITE_AREA	Area Site Covers
SITE_CUOW	Current Owner
SITE_ADAU	Administration Authority
SITE_PLRS	Planning Restrictions
SITE_ACBY	Accessibility
SITE_ACPT	Access Points to Site
SITE_REM	Remarks

Table 4.3: Site Table.

The site table allows storage of the location of the site, including a full postal address and co-ordinates of the site, along with general information such as current ownership, accessibility and planning restrictions. Having information regarding accessibility is extremely useful at the preliminary investigation stage, as it allows the investigator to gain an understanding of how equipment (drilling rigs etc.) may be bought on to site.

Planning restrictions can often play an important role in deciding investigation and construction methods, therefore having such knowledge early in the project is vital. For example, Regional Important Geological Sites (RIGS), often prevent shotcrete being used on rock slopes. Therefore on such sites, an appropriate alternative must be decided early in the project. This type of table becomes particularly useful on major development projects where there are a number of sites within a project. An example of this may be a major road development project. One site may be involved in the construction of an underpass and another involved in the construction of a bridge. Both constructions are part of the same project but on different sites, therefore highlighting the need for the site table.

The level below the site table splits into two further tables, preliminary investigation table and a zone table. The preliminary investigation table is linked directly to the site table, with the *SITE_ID* key. This table contains information regarding the details of the desk study and site reconnaissance. The data stored includes the date the preliminary investigation was undertaken, the engineer responsible and any remarks relating to the investigation.

At the same level as the preliminary investigation table, a zone table (Table 4.4) is linked to the site table. This allows a site to be divided into various sub-areas (zones), with the principle that zones are selected to reflect changes within the site. For example, a zone may be identified due to a change in land use (historical or current), which may give rise to distinct ground contamination changes. A change in the subsurface ground conditions may also warrant identification of another zone. Besides physical conditions, zones may also be identified to represent different components of a redevelopment project. For example, one zone may be used to represent the construction of an embankment, another for the foundations of a building. Therefore, the zoning system plays a useful role in the investigation process. If varying zones have been identified during the preliminary investigation the investigator may select an appropriate ground investigation technique to suit the zone.

ZONE	
Field Name	**Field Description**
PROJ_ID	Project Identifier
SITE_ID	Site Identifier
ZONE_ID	Zone Identifier
ZONE_RESN	Reason for Zone
ZONE_AREA	Area of Zone
ZONE_COR1	Co-ordinate of Zone
ZONE_COR2	Co-ordinate of Zone
ZONE_COR3	Co-ordinate of Zone
ZONE_COR4	Co-ordinate of Zone
ZONE_COR5	Co-ordinate of Zone
ZONE_REM	Remarks

Table 4.4: Zone Table.

In order to represent this concept, the zone table contains co-ordinates of the zone and the reason for its selection. The co-ordinates of the zone relate to a polygon shape made up of a number of nodes, representing the geographical area of the zone. It is also important to note that the number of zones within a site is unlimited, as each one has its own identifier which can be linked back to the site table. The zoning system also allows zones to inherit properties from the zone it is within.

The fourth level consists of the ten main parameters, derived from the nine subject areas highlighted earlier, the parameters are namely: geology, topography, geography, groundwater, surface water, history, services information, vegetation, fauna, and meteorological data. These tables are linked to the zone table by the *ZONE_ID* key, and each table has its own unique identifier. Each zone can also have as many general tables linked to it as required.

The information contained within these tables is likely to be general information. For example the history table (Table 4.5) contains information such as archaeological interest, evidence of subsidence or evidence of seismic activity. The more detailed information regarding the history is stored in the next level down.

HIST	
Field Name	**Field Description**
PROJ_ID	Project Identifier
ZONE_ID	Zone Identifier
HIST_ID	History General Identifier
HIST_ARCH	Archaeological Interest
HIST_REMA	Archaeological Interest Remarks
HIST_SUBS	Evidence of Subsidence
HIST_REMS	Remarks Regarding Subsidence
HIST_EVSA	Evidence of Seismic Activity
HIST_REMSA	Remarks Regarding Seismic Activity
HIST_SOIF	Source of Information
HIST_REM	Remarks

Table 4.5: History Table.

The fifth level of the data structure contains detailed information regarding five main subject tables in the above level (vegetation, fauna, services, geology and history), and for this reason are known as detail tables. Again each detail table has been assigned a unique identifier. In the case of the history detail table *HIDT_ID* (Table 4.6), this allows the general table to have connections to as many entries within the detail table as required. Therefore, in the case of the history, most areas (zone) being investigated are likely to have a number of past uses, which can be represented by assigning a history detail table entry to each use.

HIDT	
Field Name	**Field Description**
PROJ_ID	Project Identifier
ZONE_ID	Zone Identifier
HIST_ID	History General Identifier
HIDT_ID	History Detail Identifier
HIDT_NAME	Name of Owner
HIDT_USE	Previous Use
HIDT_FEAT	Features Associated with Use
HIDT_STAT	Start Date of Use
HIDT_FINS	Finish Date of Use
HIDT_DURT	Duration of Use
HIDT_LVOD	Level Above Ordnance Datum
HIDT_SOIF	Source of Information
HIDT_REM	Remarks

Table 4.6: History Detail Table.

For example, one history detail table entry may contain information regarding a gas works that relates to the early history of the area. Another detail table entry relating to the same area, may contain information about a steelworks that relates to the later history of the area. Therefore the type of information stored within the history detail table includes; past use (this plays an important role in identifying contaminants associated with site use), start and finish for this site use (this allows a judgement to be made on how long contaminants may have been on site) and features associated with past uses, (this allows the system to identify hazards associated with such features), and level above ordnance datum (which can indicate whether the area has been infilled or excavated since the land use described).

Also among the detail tables at this level is the geology detail table (Table 4.7). This table contains any data regarding subsurface material, including information concerned with made ground, superficial geology or bedrock geology. This again allows the zones to have as many geology types as required. For example, within a zone where there are three distinct layers of material, e.g. layer one: made ground, layer two: Coal measures and layer: three Sandstone, each layer would be assigned an entry in the detail geology table. Within each table, details of the type and age of material, depth to top of layer, main characteristics and source of information are stored. The information stored within this table allows the investigator to have an understanding of the geology located within the area under investigation. This in turn allows permeabilities to be assigned to different types of geology and also allows judgements to be made regarding possible movement of contamination through the different types of geology.

It is also important to note that the stratum descriptions table within the AGS format is given the group name GEOL. This is obviously the same name as has been assigned to the geology general table described in this chapter. However, this is not seen as a problem, due to the fact that the IDs for each table do not conflict. The stratum descriptions table sits below the hole table within the AGS format, which relates to individual boreholes from exploration investigations. This means that key fields within this table are *HOLE_ID*, *GEOL_TOP* and *GEOL_BASE*. In the preliminary investigation data structure the key fields for the geology general table are; *PROJ_ID*, *ZONE_ID* and *GEOL_ID*. This means that if the AGS format is used in conjunction with the format detailed within this chapter, the two tables can sit within different levels of the data structure and not conflict.

GEDT	
Field Name	**Field Description**
PROJ_ID	Project Identifier
ZONE_ID	Zone Identifier
GEOL_ID	Geology General Identifier
GEDT_ID	Geology Detail Identifier
GEDT_DESC	Stratum Description
GEDT_LYNO	Layer Number
GEDT_LYDT	Depth to Top of Layer
GEDT_LYTH	Layer Thickness
GEDT_CHAR	Characteristics
GEDT_FEPT	Features Present
GEDT_SOIF	Source of Information
GEDT_REM	Remarks

Table 4.7: Geology Detail Table.

Other detail tables include the services detail table (Table 4.8). This table includes information regarding the type of service, responsible authority, elevation, trend of service and co-ordinates of service. The information here is vital for identifying possible pathways for contaminant movement as well as ensuring that boreholes and trial pits are not excavated at the location of services.

SEDT	
Field Name	**Field Description**
PROJ_ID	Project Identifier
ZONE_ID	Zone Identifier
SERV_ID	Services General Identifier
SEDT_ID	Services Detail Identifier
SEDT_TYPE	Type of Service
SEDT_RSAT	Responsible Authority
SEDT_ELEV	Elevation
SEDT_TRED	Trend of Service Across Zone
SEDT_STCX	Start X-Co-ordinate of Service
SEDT_STCY	Start Y-Co-ordinate of Service
SEDT_FNCX	Finish X-Co-ordinate of Service
SEDT_FNCY	Finish Y-Co-ordinate of Service
SEDT_SOIF	Source of Information
SEDT_REM	Remarks

Table 4.8: Services Detail Table.

The other two tables at this level include the vegetation detail table (Table 4.9) and the fauna detail table (Table 4.10).

VEDT	
Field Name	**Field Description**
PROJ_ID	Project Identifier
ZONE_ID	Zone Identifier
VEGE_ID	Vegetation General Identifier
VEDT_ID	Vegetation Detail Identifier
VEDT_TYPE	Type of Vegetation
VEDT_HATH	General Health of Vegetation Type
VEDT_REMH	Remarks
VEDT_LEVH	Health of Leaves
VEDT_REML	Remarks
VEDT_ROTH	Health of Roots
VEDT_REMR	Remarks
VEDT_YSRG	Young Seedling Regeneration
VEDT_REMY	Remarks
VEDT_VGDB	Vegetation Die Back
VEDT_REMD	Remarks

Table 4.9: Vegetation Detail Table.

Both tables have been compiled in order to store information that assists in identifying likely contaminants. For example, the health of certain types of vegetation, seedling regeneration and vegetation die back can be used to identify contamination within the ground. The same is also true for fauna health as well as the abundance and diversity of certain species.

FADT	
Field Name	**Field Description**
PROJ_ID	Project Identifier
ZONE_ID	Zone Identifier
FAUA_ID	Fauna General Identifier
FADT_ID	Fauna Detail Identifier
FADT_SPCE	Fauna Species
FADT_HLTH	Health of Species
FADT_HEDT	Details of Health
FADT_ABNC	Abundance of Species
FADT_DIVS	Diversity of Species
FADT_REM	Remarks

Table 4.10: Fauna Detail Table.

The final level within the data structure is designed to store detailed information relating to the fifth level, and these are again linked by their unique identifier. The reason for this extra level is outlined in the following example; a zone may have a number of types of geology within it, one of which may be a sandstone containing a number of faults and folds. The general geological information about the zone will be contained within the fourth level of the data structure, the detailed information about the sandstone will be within the fifth level. However, a problem of storing data regarding the structure features (folds and faults) within the layer arises at this point. It is impossible to store these features within this fifth level, as the number of structural features is variable. To overcome the problem another level has been added below the fifth level with the aim of storing such features.

STFT	
Field Name	**Field Description**
PROJ_ID	Project Identifier
ZONE_ID	Zone Identifier
GEDT_ID	Geology Detail Identifier
STFT_ID	Structural Feature Identifier
STFT_TYPE	Structural Feature Type
STFT_FTDD	Dip & Direction of Feature
STFT_FTLC	Location of Feature
STFT_FTSZ	Size of Feature
STFT_SOIF	Source of Information
STFT_REM	Remarks

Table 4.11: Geology Structural Features Table.

Within this level sits a structural features table (Table 4.11), which allows the storage of data regarding the type of feature and its dip and direction. Each feature is assigned an entry in the table. Therefore, within the example outlined above, one entry may detail information about a fault within the sandstone and another about a fold. This allows the geology detail table to have as many structural feature table entries linked to it as required. The information stored at this level plays an important role in identifying possible pathways for contaminant movement.

4.4 User Interface Design

The user interface plays an important role in the effective running of the database system. There are a number of requirements that the interface needs to meet; these include; allowing the user to input data in a systematic and clear manner; allowing the user to retrieve and change data in a logical and easy manner; prompting the user for data required by the application to continue; allowing the user to navigate around the system independently without losing track of where they are within the system; displaying stored data in a professional and clear manner on screen and allowing hard copies of data to be output.

To achieve these requirements Access offers a number of features. These include; forms, reports, macros and modules. On opening the database, an Autoexec macro has been used to initiate the user interface and set up the required tool bars. The macro opens a main screen that allows users to select the type of data they wish to input. This includes preliminary information, laboratory test data and ground investigation information.

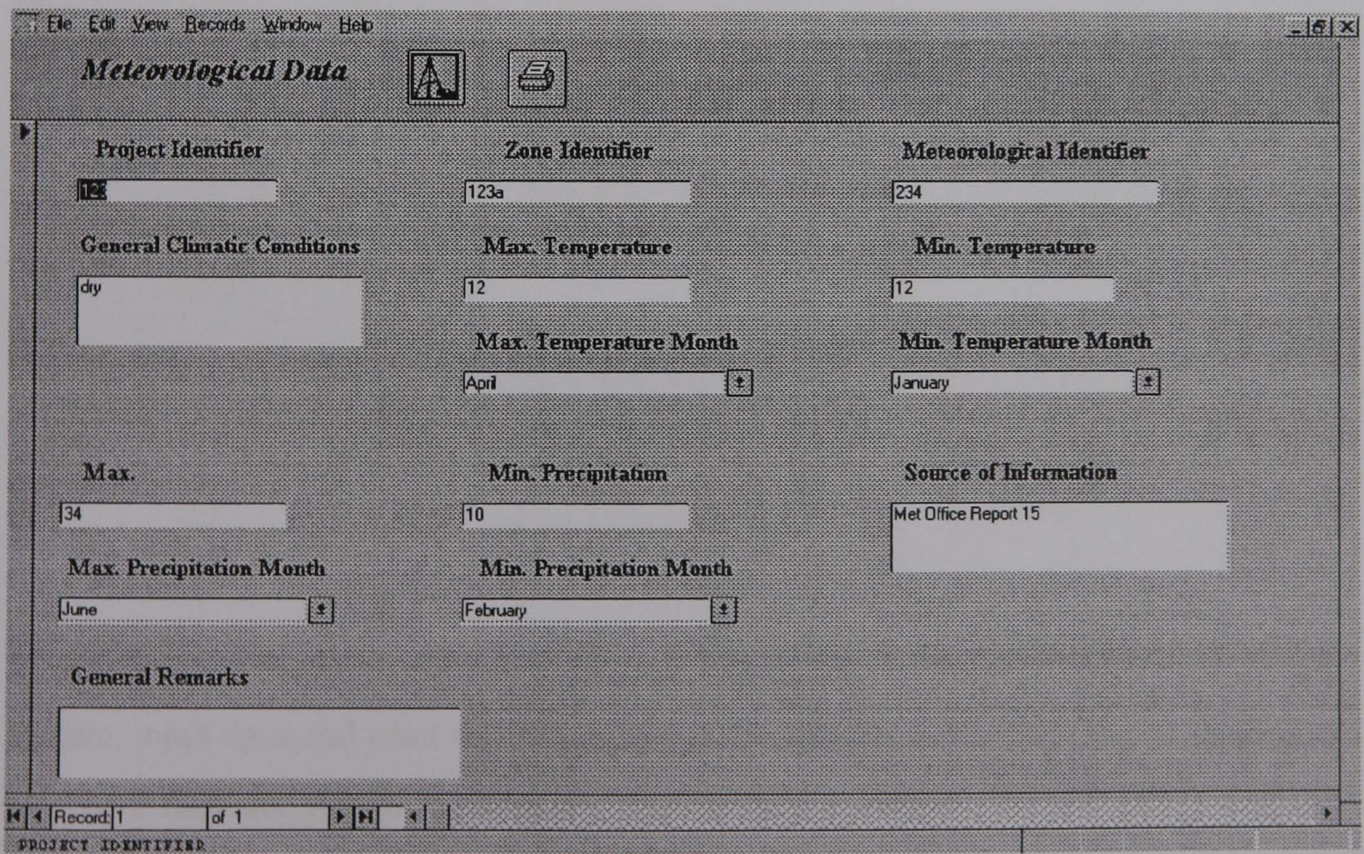


Figure 4.2: Non-Scrolling Areas and Command Buttons.

As displayed in Figures 4.2 and 4.3 the form feature within Access offers the most convenient layout for allowing the user to enter, change and view records within the database. Access also contains a number of functions to assists with these tasks.

Functions such as command buttons, menu commands are highlighted in Figure 4.2. Command buttons are generally used within the form to allow the user to navigate around the system i.e. move from one data entry screen to another, undertake queries and produce hard copies of data (button containing printer). The menu commands are displayed to the user when the menu name is highlighted with the cursor. Again such commands act in a similar manner to the command buttons, although have the advantage of occupying less room on the form. However the drawback is that the user must open the menu to see the commands.

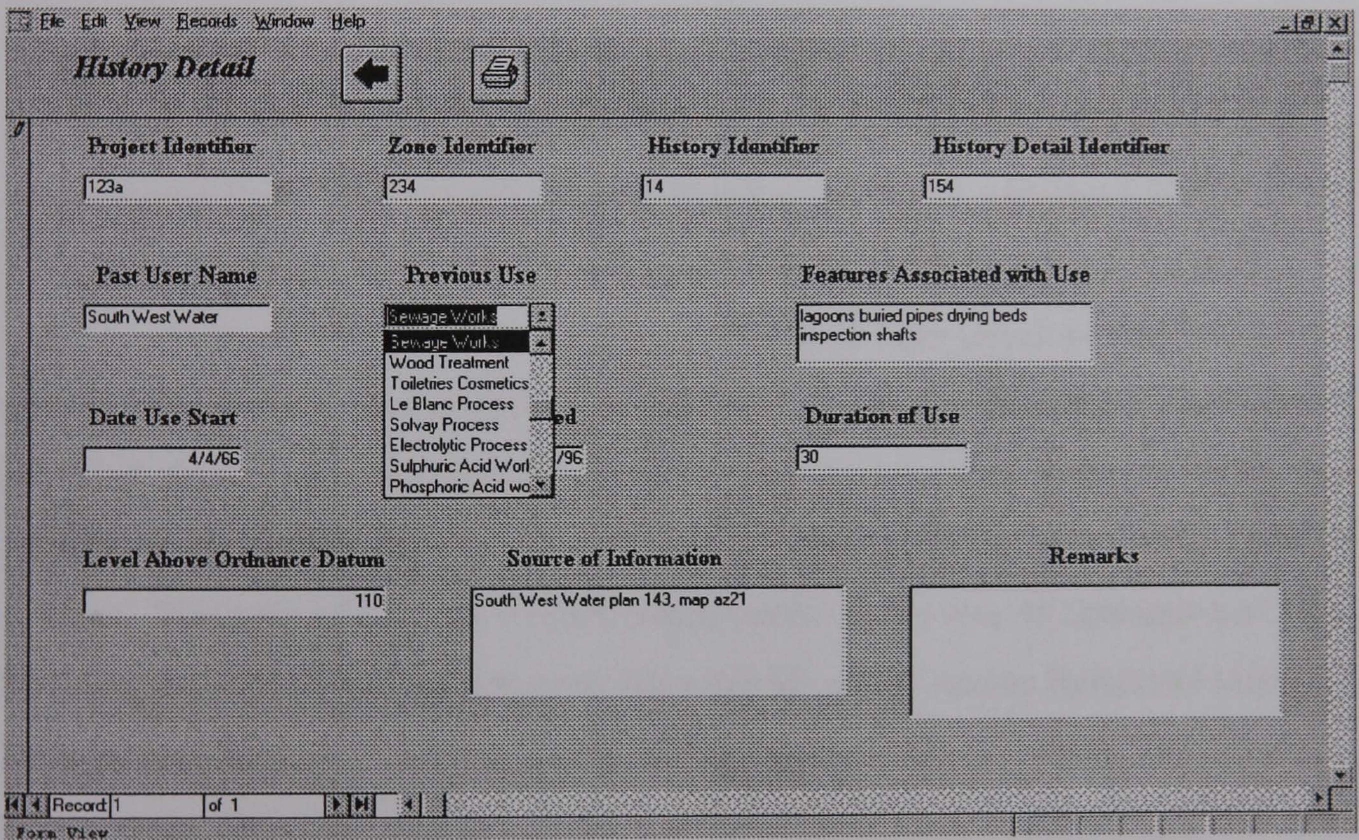


Figure 4.3: Drop Down List Box and Navigation Buttons.

In addition to displaying command buttons that allow the user to navigate around the system, input data and print records, it is also essential that the user does not lose track of where they are within the system. This problem was solved with the use of non-scrolling headers and footers. The buttons situated within the footer allow the user to move either back to a previous record or forward to the next or new record, as shown in Figure 4.3. The command buttons within the header play a similar role, but allow the user to move from one table to another, i.e. from the history detail table form to the history general table form. This ability to allow users to move back and forwards between records and tables, prevents the user being lost within the system.

Other features used within the user interface included; dialog boxes, which display important messages and collect information from the user, and drop down list boxes also shown in Figure 4.3. The drop down list box controls the input from the user by limiting the data that may be entered, although if required the user may add to the list display. The report feature was also added to the user interface to allow the user to generate search reports.

4.5 Conclusion

The storage of preliminary investigation data is extremely important within the area of contaminated land. The storage of such data has yet to be addressed in full by the AGS Data Exchange Format or by any geotechnical database systems available. In order to overcome this problem, data structures have been developed for storing all aspects of preliminary site investigation information, ranging from geological data to historical data. The data structure designed also contains the ability to store data relating to vegetation and fauna, which is a major advance over other database systems. This type of data can be particularly useful in the area of contaminated land. The data structures were implemented using the Microsoft Access Relational Database Management System. A user interface was also developed within the Windows environment. This was seen as a positive factor as it is likely to encourage end users (geotechnical engineers) to use the database, as they will generally be familiar with Windows based software.

With the preliminary investigation data stored in such a manner it also allows data to be passed to other software, therefore making data available for the knowledge-based system, to use within its rules.

CHAPTER 5

KNOWLEDGE-BASED SYSTEM DEVELOPMENT

5.0 Introduction

This chapter is concerned with the development of a knowledge-based system, known as ATTIC (Assessment Tool for The Investigation of Contaminated Land). The aim of the system is to assist with the preliminary stage of investigation of a contaminated site.

To achieve the desired structure and appropriate outcomes, described in section 5.1 and 5.1.1, it was essential from the outset that a clear and structured approach to system development was undertaken. This involved the collection of domain knowledge from a wide range of sources via a structured knowledge acquisition process, detailed in section 5.2.

As a result of the knowledge acquisition exercise, it was necessary to process the large volume of information collected into a suitable format for use in a knowledge-based system. This involved segmenting the information into rules that allow the system to compare information input by the user with knowledge derived from technical literature and domain experts. The knowledge representation process not only involved compiling suitable rules but grouping the rules into complementary sets known as knowledge-bases, that allowed the results from the rules to be passed from one set to another and therefore maximising the use of the available information. Section 5.3 describes the knowledge representation process in detail, highlighting the rules and the division of rules into sets (knowledge-bases).

This is followed by a description of the implementation of the knowledge using CLIPS software, in section 5.4. The construction of the user interface for the system is outlined in section 5.5. The final section (5.6) describes the conclusions relating to this chapter.

5.1 Areas and Objectives for System Development

The development of a prototype knowledge-based system is intended to address the following:

(1) Store information collected in the preliminary stages of the investigation in a standardised form, making it accessible to the knowledge-based system involved in the interpretation, as well as to other packages the user may require (e.g. GIS packages). This also makes data available for transfer between all parties involved in a redevelopment project.

(2) Consist of a user-friendly package that is accessible to individuals with varying degrees of computer experience that can be used on a standard stand-alone personal computer.

(3) Offer advice to all levels of staff involved in the investigation ranging from junior engineers to senior consultants. This should be seen more as a tool to assist professionals rather than a replacement of experts.

(4) Through the input and collection of data, enforce a structured approach to the investigation.

(5) Use a range of abiotic and biotic indicators (e.g. tolerant plant species, soil staining etc.) to assist with location of the source and types of contaminants likely to be present on site, together with any related hazards (e.g. buried tanks etc.)

(6) From the list of predicted contaminants give advice regarding health and safety requirements for site workers.

(7) Use information from published literature (e.g. geological maps etc.) to locate possible migration pathways for contaminants both on and off site.

(8) From published data and site reconnaissance assist with location of vulnerable targets both on and off site.

5.1.1 System Overview

The development of a knowledge-based system to assist with the investigation of contaminated land requires the representation of knowledge regarding the source of contaminants, likely pathways and possible targets. It was decided that it was possible to consider these three distinct components separately and compile a knowledge-base for each component. The production of the knowledge-bases involved collecting relevant data from technical literature as outlined in knowledge acquisition (Section 5.2), and compiling realistic rules that represent the data collected. Each knowledge-base has its own series of rules with its own related data. Four distinct knowledge-bases were identified. These included the source knowledge-base, pathway knowledge-base, target knowledge-base and the health & safety knowledge-base. Figure 5.1 highlights the relationship between the four knowledge-bases and the final output to the user. As shown in Figure 5.1 each knowledge-base is closely related, with facts being passed from one knowledge-base to another, in order to achieve the final output to the user.

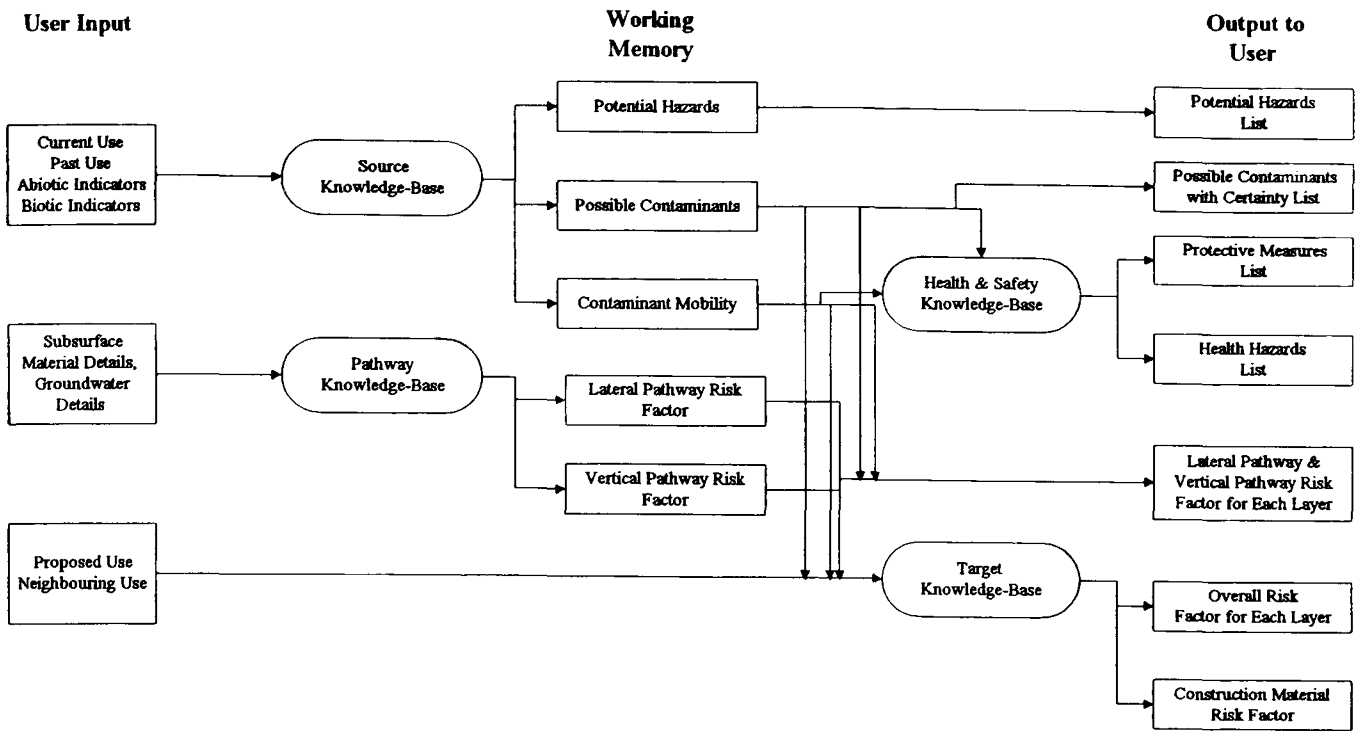


Figure 5.1: Overview of System.

The source knowledge-base contains rules that utilize the facts related to current use, past use and abiotic & biotic indicators input by the user. As a result of these rules the user is presented with a list of potential hazards and contaminants. In addition, the rules output facts relating to contaminant mobility.

The pathway knowledge-base relies on facts relating to subsurface material and groundwater details input by the user. The rules within this knowledge-base convert such data into a pathway risk value for both vertical and lateral movement of contaminants. This risk value relates to the likelihood of contaminants moving from their source in a vertical or lateral direction. The rules assess a range of factors from permeability of subsurface material through to the attitude of structural features. The derived risk values are presented to the user in order to give the user an understanding of possible contaminant movement within the subsurface.

The target knowledge-base draws upon facts from two sources. The initial source is from the information input by the user relating to proposed and neighbouring use. The second source is facts derived from the source (list of contaminants, mobility of contaminants) and pathway (lateral & vertical risk values) knowledge-bases. The rules within this knowledge-base combine the facts and present the user with an overall risk factor which indicates the direction of risk, for every layer under investigation. The target knowledge-base also contains rules concerned with the risk that possible contaminants pose to the building materials on site. The rules simply check the type of contaminants available and assess the risk they pose to main construction material types (concrete, steel, plastic and rubber). This process returns a result in terms of high, medium or low risk to the material under investigation.

The final knowledge-base within the system is related to health and safety issues. The rules contained within this component of the system require the results derived from the source knowledge-base, in particular, the likely contaminants that may be present within the zone being investigated and the mobility of contaminants. The result is a list of potential health hazards and the required protective measures. The results relate to worst case scenarios, as no quantitative data are usually available during the preliminary investigation stage. Therefore it was decided to advise based on the highest values.

Once the system has completed a run through of the four knowledge-bases, the results stored within the working memory are presented to the user in the form of a list. This list includes; possible contaminants and a certainty value, a list of potential hazards (relating to past use), lateral and vertical pathway risk factor for each layer, a risk value for construction materials, a list of health hazards and protection measures required and finally an overall risk relating to neighbouring land use.

5.2 Knowledge Acquisition

When developing a knowledge-based system one of the most important and difficult tasks faced by the developer is the capture of knowledge for the knowledge-base. This process is usually known as knowledge acquisition, and, as suggested by Rich & Knight (1991), remains a major bottleneck in applying knowledge-based system technology to new domains.

Liou (1998) defines knowledge acquisition as the process of extracting, structuring, and organising knowledge from several knowledge sources, usually human experts so that the problem solving expertise can be captured and transformed into a computer readable form.

5.2.1 Source of Knowledge

To develop a successful knowledge-based system it is essential that the quality of the knowledge collected during the acquisition phase of the development is of the highest possible standard. Therefore, it is important that the most relevant sources of information are reviewed and the most appropriate domain experts are consulted. The knowledge collected from such sources may include; rules of thumb, case studies, definitions, formulae, facts, rules, definitions and hypotheses. Therefore, defining knowledge precisely is a difficult task.

It is also essential for the developer to have a clear understanding of the various knowledge types and how they may be used within a system. Within the domain of the investigation of contaminated land, it can be concluded that two main sources of knowledge may be utilised during the knowledge acquisition phase of the knowledge-based system development. These two sources are, technical literature and domain experts.

5.2.2 Technical Literature

This is often known as public knowledge. It consists of such sources as codes of practice (e.g. BS5930, DD175), geo-environmental engineering textbooks, journals (e.g. Land Contamination & Reclamation, Journal of the Institution of Water and Environmental Management), conference proceedings and technical reports (e.g. D.o.E. contaminated land research reports). The literature review of this data (section 2.0 - 2.6) outlines the parameters that need to be considered when investigating a potentially contaminated site. This includes data regarding potential contaminants, characteristics of different subsurface materials and vulnerable land use information.

The high quality and amount of information available made it possible to construct all the rules required for the system from this source. The technical literature, however, did have a number of drawbacks, including problems with certainty and ranking. Technical literature rarely contains information relating to the certainty of the data. Therefore, it is difficult to define how necessary a condition is, for a particular conclusion to be drawn.

5.2.3 Domain Experts

The domain of contaminated land covers a range of disciplines covering such areas as geology, chemistry, hydrology and geotechnics. Therefore it is possible to have a number of domain experts relating to this range of disciplines. However, in reality, often one expert will normally be trained within one subject area, but have enough knowledge of other areas to make appropriate decisions. This domain expert is usually a geo-environmental engineer or a geotechnical engineer who specialises in contaminated land investigation. These experts usually work within geo-environmental or geotechnical consultancies.

For the purpose of this research it was decided that such domain experts would be engaged to validate the system. On taking into account the issues discussed in Chapter 3 (section 3.8 knowledge acquisition) it was decided that a questionnaire would be compiled relating to the rules and parameters within the system.

The questionnaire format (Appendix 3) was selected as a knowledge acquisition tool as it was hoped that this would allow a large population of domain experts to be consulted. It also meant that experts from a range of disciplines could be consulted.

On consideration of the type of information required by the system, it was possible to identify three main subject areas. The first area related to biotic and abiotic indicators and their usefulness in identifying potential contaminants. It was decided that respondents would be asked to rate fourteen such indicators (detailed in Chapter 2) using a scale of one to five, with one relating to an indicator of very little use when used as a sole indicator, to five indicating an extremely useful indicator. This was viewed as a good way to gather the views of domain experts on such indicators, as such detail is not available within the technical literature. The values from this exercise could then be used as certainty values within rules. This pre-set scaled answer (closed question format) to the question (i.e. rating answers from 1 - 5) was selected after considering the views detailed within Chapter 3, knowledge acquisition section 3.8. This closed question format fully controls the answer from the respondent, therefore making the results easier to use. The alternative, would be to ask this question in an open format, which could result in a number of terms describing indicators being used, for example "a good indicator", "an ok indicator" etc. Such results would make analysis of the questionnaire very difficult.

The second area identified concentrated on the movement of groundwater. As discussed in Chapter 2 groundwater plays an important role in the movement of contaminants. Knowing the velocity of groundwater movement within a potentially contaminated site is extremely useful in the risk assessment process. Respondents were therefore asked to provide appropriate ranges of velocity (m/day) for the following terms; Fast, Medium and Slow. Such terms would then be incorporated into the pathway rules within the system. This type of question is therefore an open format (section 3.8) and varies considerably to the closed format question used for the indicators question. It was felt that the open format eliminated any bias that may occur if predefined velocities were presented to the respondents. It was also extremely difficult to find suitable values for such terms within the technical literature. Although this type of question may cause problems of analysing ambiguous responses, it was thought to be the best option available.

The final area identified related to vulnerable targets. It was felt necessary to obtain a judgement from domain experts relating to the different land uses. Initially an open format was considered, to ask respondents to list different land uses and assign a vulnerability to contamination. However it was thought that this may result in a number of land uses and terms describing their vulnerability being identified. This would obviously make analysis of results difficult. On consideration of such factors it was decided to use a closed question format, therefore respondents were asked to use three terms (high, medium and low) to classify ten land uses (targets). The land uses included; school, public open space, agricultural area, general commercial, low density residential area, medium density residential area, high density residential area, light industrial area, heavy industrial area and permanently paved area.

In addition to the three main areas outlined, respondents were also asked in an open question format about their occupation, number of years of experience within their field and the type of organisation they are employed by. This was to overcome potential problems, relating to the expertise of the respondent. It was hoped that the result from such questions would be a way of judging the expertise of the respondent and hence how valid their replies were, although this clearly relies on the respondent answering the questions honestly.

A covering letter was also attached to each questionnaire, detailing the research study and how to return the questionnaire. Each section within the questionnaire also contained a short expansion outlining the background to the question being asked. The final stage of the questionnaire development involved checking, to ensure that the questions were clear and short.

On completion of the development of the questionnaire a method of distributing it had to be considered in order to achieve a high return rate and useable results. After reviewing the advantages and disadvantages of distributing questionnaires as suggested by Miller (1991), (section 3.8) it was decided that the questionnaire would be mailed out. However, rather than using the standard postal service, it was decided that with increased use of computer technology communications, for example Internet web sites and e-mail, this form of medium would provide an ideal way of distributing the questionnaire. It was, therefore, decided to e-mail the questionnaire to a range of mailing lists whose members are likely to be involved in the field of contaminated land. In total the questionnaire was mailed to six mailing lists which included; engineering-geotech (subject area: geotechnical engineering), geo-env (subject area: environment and geology), bsss-soil (subject area: soil science based topics), bioregional (subject area: natural regions and human habitation), water-env-info-systems (subject area: develop of water and environmental information systems) and env-chem (subject area: environmental chemistry). It was estimated that around 560 members subscribe to such lists.

Although this method was thought to be a good way of reaching a relative large number of experts in the chosen domain, it was important to note the disadvantages of such a method. Generally there are two main disadvantages, the first being that there is limited control over the population being questioned. This is due to the fact that the questionnaire may be posted to a mailing list that contains members who are involved or interested in the field of contaminated land but who may not have a great deal of experience within the domain. The question asking the respondent to state their experience and occupation was hoped to overcome this problem. The second and minor disadvantage is that it may be difficult to work out the exact size of the population to which the questionnaire has been posted to. This is due to the fact that some members may subscribe to several lists, or may not use their e-mail account, although their address is on the list, therefore mailing list numbers may not be a true reflection of actual members.

Forty six responses to the questionnaire were received. The response was considered to be good and some extremely useful information was identified as a result of the questionnaire. The results were analysed relating to experience and occupation (Table 5.1 & 5.2). Two categories were identified for experience; 5 - 10 years experience and >10 years experience, with a split of approximately 52% and 48% respectively between the two categories.

Experience Category	Responses
5 - 10 Years Experience	24
>10 Years Experience	22
Total	46

Table 5.1: Questionnaire Responses by Experience Category.

The open question format relating to the occupation of the respondents identified five categories of occupation. These included Chemist, Geo-environmental engineer, Geologist, Hydrogeologist, and other. As expected the greatest response came from the Geo-environmental engineer category. It is assumed that this is due to the fact that this category is the one generally involved in the investigation of contaminated land and therefore the category with the most experience.

Occupation Category	Responses
Chemist	7
Geo-Environmental Engineer	15
Geologist	11
Hydrogeologist	8
Other	5
Total	46

Table 5.2: Questionnaire Responses by Occupation Category.

On completion of categorising the results by occupation and experience, each category was analysed in terms of their responses to the subject areas identified earlier; biotic and abiotic indicators, groundwater velocities and vulnerable targets.

Bar charts were plotted for the results from each subject area of the questionnaire (excluding groundwater velocities). For example, with the biotic and abiotic indicators the fourteen indicators were plotted in terms of certainty value and number of respondents, with certainty plotted on the x axis and number of respondents on the y axis. This process was performed for each occupation category identified and each experience category, hence resulting in seven sets of results. A similar process was also undertaken with the results relating to vulnerable targets, with target values (high, medium and low) plotted against number of respondents. Analysing the results in such a manner allowed a comparison to be made between the different occupations and different scales of experience. Full sets of plotted results are presented in Appendix 4.

On completion of plotting the results, the most popular certainty value in the case of indicators and target value in the case of vulnerable targets was identified and summarised in Tables 5.3 & 5.4. It is clear from the results within the two tables that generally all seven categories were in agreement with values assigned to the subject areas. For example, the indicator ground surface staining colour was assigned a certainty value of three by five of the categories and a value of two by the remaining two categories. Therefore it was reasonable to accept the mode value, i.e. the value that occurs most frequently in the set of data, for each indicator as a reflection of the domain experts opinion. However, within some of the categories the domain experts identified two values, in such cases the worst case (lowest value) was considered when identifying the mode value. For example, the indicator terrestrial invertebrates visible health symptoms was assigned 1 - 2 value by the Hydrogeologist category. The worst case value in this situation is value 1 so this was the value used when identifying the mode value.

On completion of identifying the mode value for each biotic and abiotic indicator and vulnerable target, the mode value was incorporated into the relevant rules as a certainty value, as will be described later in the chapter.

In the case of groundwater velocities it became apparent from the results that identifying such values was a difficult task. Nineteen of the respondents did not answer the question and of those who did, very few agreed on a common answer. It therefore became necessary to select an appropriate value from the limited results that were generally similar. The final values selected are displayed within Table 5.5, with the full set of results tabulated in Appendix 4.

Indicators \ Category	5 - 10 Years Experience	>10 Years Experience	Chemist	Geo-Environmental Engineer	Geologist	Hydrogeologist	Other	Mode Value
Current use of site	5	5	5	5	4	3	4	5
Historical use of site	5	5	4-5	5	5	5	4	5
Presence of odours	3	3	3	3	3	3	3-4	3
Ground surface deposit type	3	3	3	3	3-4	4	5	3
Ground surface deposit colour	3	3	3	3	3	2	2-3	3
Ground surface staining colour	3	3	3	3	3	2	2	3
Terrestrial vegetation tolerant species	2	2	4	2	2	2	3	2
Terrestrial vegetation visible health symptoms	2	2	2-4	2	2	2	2-3	2
Visible symptoms concerned with soil microbiology	2	2	2	2	2	2	2	2
Terrestrial invertebrates tolerant species	1	2	2	1-2	1	1	1-2	1
Terrestrial invertebrates visible health symptoms	1	1-2	2	1	1	1-2	1	1
Aquatic invertebrates tolerant species	2	2	2	2	1-2	2	1	2
Aquatic invertebrates visible health symptoms	2	2	2	1-2	2	2	1-2	2
Visible health symptoms relating to mammals	2	2	2	1	2	2	1-2	2

Table 5.3: Questionnaire Responses for Biotic and Abiotic Indicator Values from all Categories.

Category \ Target	5 - 10 Years Experience	>10 Years Experience	Chemist	Geo-Environmental Engineer	Geologist	Hydrogeologist	Other	Mode Value
School	High	High	High	High	High	High	High	High
Public Open Space	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium
Agricultural Area	High	High	High	High	High	High	High	High
General Commercial	Low	Low	Low	Low	Low	Low	Low	Low
Low Density Residential Area	High	High	High	High	Medium	Medium	High	High
Medium Density Residential Area	High	High	High	High	High/Medium	High	High	High
High Density Residential Area	High	High	High	High	High/Medium	High	High	High
Light Industrial Area	Low	Low	Low	Low	Low	Low	Low	Low
Heavy Industrial Area	Low	Low	Low	Low	Low	Low	Low	Low
Permanently Paved Area	Low	Low	Low	Low	Low	Low	Low	Low

Table 5.4: Questionnaire Responses for Vulnerable Target Values from all Categories.

Velocity of Groundwater m/day	Term
< 0.01	Slow
0.01 – 0.1	Medium
>0.1	Fast

Table 5.5: Selected Values for Groundwater Velocities from Questionnaire Responses.

5.3 Knowledge Representation

To represent the four distinct knowledge-bases (source knowledge-base, pathway knowledge-base, target knowledge-base and health & safety knowledge-base) identified in section 5.1.1, it was vital that appropriate knowledge was compiled into a suitable format.

The following sections describe the methods and rules used to represent the knowledge required in each knowledge-base. Between the four knowledge-bases there are approximately 1600 rules contained within the system. Due to this large number of rules, a limited number of examples have been presented to give an understanding of the purpose of the rules and how they interact between the knowledge-bases and the desired outputs to the user.

5.3.1 The Source Knowledge-Base

When investigating a potential contaminated site, one of the initial objectives of the investigator is to assess the possible contaminants that may be present on site. This is undertaken in a number of phases as outlined in section 2.6.2.1, with an initial phase of desk study and site reconnaissance survey. The aim of this system is to capture such data within the source knowledge-base and assist the investigator in deciding upon the likely contaminants present, their mobility and related hazards.

Within the source knowledge-base there are four main categories of rules producing outputs. The four categories of rules rely on the user inputting data relating to the past and current uses of the area under investigation and evidence from abiotic and biotic indicators. The initial category identifies a list of possible contaminants, with the other categories producing a list of possible hazards, a hazard ranking for the area and a mobility of contaminants ranking. The mobility ranking also draws upon data from the pathway knowledge-base.

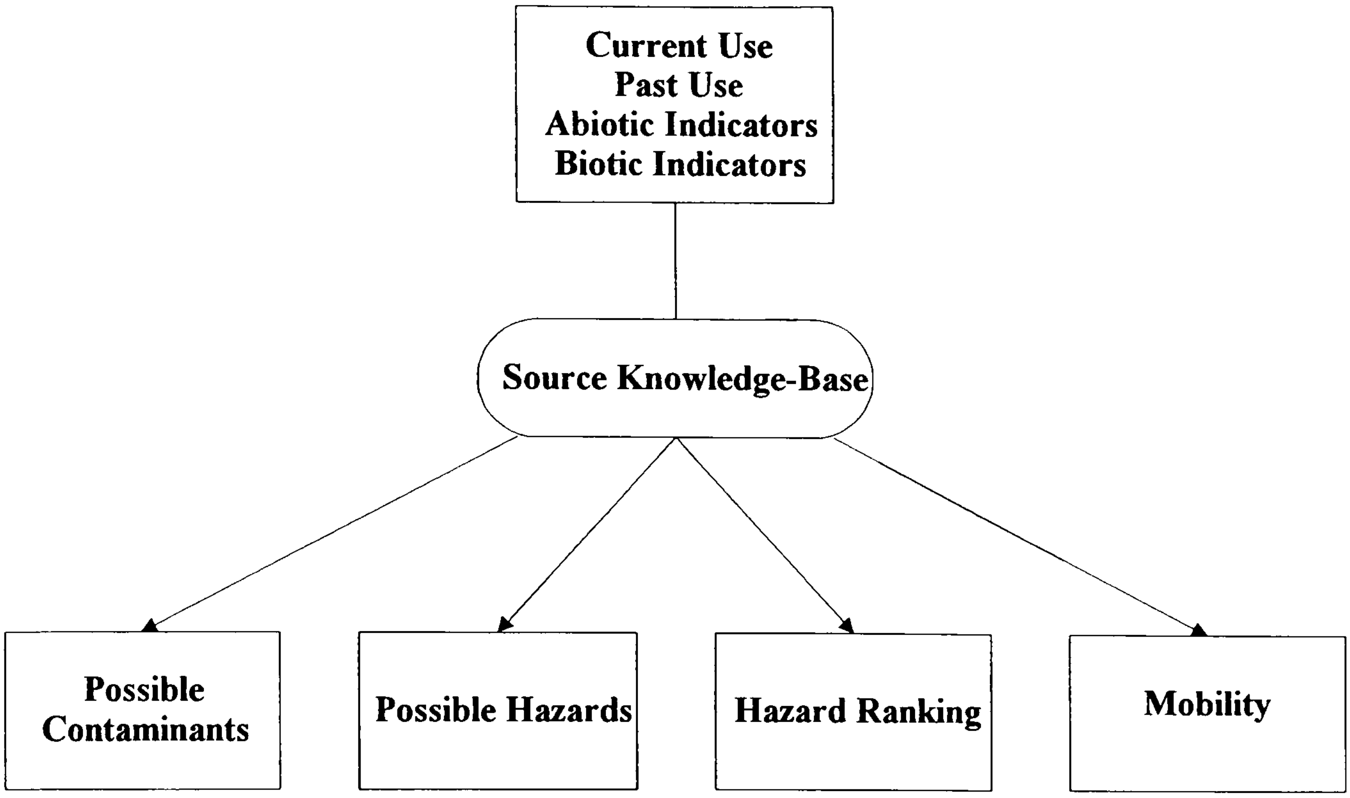


Figure 5.2 Results from Source Knowledge-Base.

As previously mentioned the initial category produces a list of possible contaminants. A review of the technical literature revealed that there are a number of basic sources of information that are commonly used to identify potential contaminants (highlighted below). Such factors range from historical maps of the area to odours that relate to different contaminants.

Source Components

- Current use
- History
- Fauna *(type & visible symptoms)*
  - (a) Mammals
  - (b) Terrestrial Invertebrates
  - (c) Aquatic Invertebrates
- Surface Deposits *(type & colour)*
- Odours
- Surface Stains *(colour)*
- Flora *(type & visible symptoms)*
  - (a) Trees & Shrubs
  - (b) Grasses

(c) Forbes

(d) Mosses

(e) Microbiology

- Others *(road & pub names, oil drums, waste materials)*

There is a considerable amount of knowledge available regarding such factors. During the literature search such knowledge was tabulated and a Microsoft® Help file was compiled. This provided a useful reference source when converting such knowledge into sets of rules that predict likely contaminants. It was also envisaged that the Help file would be attached to the final knowledge-based system interface and made available as a reference point for the user. Obviously certain types of data are more useful than others. For example, information gathered from historical maps can give an insight to likely past uses of the site, which in turn relates to the types of contaminants associated with such past uses. This type of data is reasonably reliable, due to the fact that records relating to manufacturing processes are well documented. In comparison, observing soil staining or tolerant/susceptible plant species is less reliable, due to the fact that other environmental factors can affect such features. This is confirmed by Department of the Environment (1994c), which suggests that such features are of little use, without considering other factors.

When constructing the knowledge-base for these factors, it was essential that this reliability was taken into account. This was achieved by assigning certainty values to the sets of rules, with more reliable knowledge having a higher certainty than less reliable knowledge. The reliability and in turn the certainty of the indicators was assessed by domain experts during the questionnaire exercise. A mode value for each indicator was derived from the results of the questionnaire as discussed within section 5.2.3. This mode value then became the certainty value for each rule. Table 5.6 outlines the certainty values assigned to each indicator type, which in turn relates to a set of rules.

Indicator Type	Certainty Value	Rule Sets within Indicator Type
Current use of site	5	current-related-contaminant
Historical use of site	5	history-related-contaminant
Presence of odours	3	odours-related-contaminant
Ground surface deposit type	3	deptype-related-contaminant
Ground surface deposit colour	3	deposit-related-contaminant
Ground surface staining colour	3	stain-related-contaminant
Terrestrial vegetation tolerant species	2 2 2 2	grass-vege-related-contaminant forb-vege-related-contaminant trees-shrubs-related-contaminant mosses-liverworts-related-contaminant
Terrestrial vegetation visible health symptoms	2 2 2 2	grass-symptoms-related-contaminant forb-symptoms-related-contaminant trees-shrubs-symptoms-related-contaminant mosses-liverworts- symptoms-related-contaminant
Visible symptoms concerned with soil microbiology	2	soil-microbiology-related-contaminant
Terrestrial invertebrates tolerant species	1	terrestrial-invertebrates-related-contaminant
Terrestrial invertebrates visible health symptoms	1	terrestrial-invertebrates-symptoms-related-contaminant
Aquatic invertebrates tolerant species	2	aqu-invbra-related-contaminant
Aquatic invertebrates visible health symptoms	2	aqu-invbra- symptoms -related-contaminant
Visible health symptoms relating to mammals	2	mammal-symptoms-related-contaminant

Table 5.6: Certainty Values & Rule Sets.

Table 5.6 also details the sets of rules relating to each indicator, for example in its simplest form, *the Historical Use of Site* indicator relates to the *history-related-contaminant* rules. Each rule predicts a list of contaminants on the basis of the historic use of the site. In a more complex situation the *Terrestrial Vegetation Tolerant Species* indicator relates to four sets of rules (*grass-vege-related-contaminant, forb-vege-related-contaminant, trees-shrubs-related-contaminant, mosses-liverworts-related-contaminant*), with each set predicting likely contaminants on the basis of presence of indicator species of vegetation on the area under investigation. The four sets relate to broad classes of different vegetation types.

To illustrate the way in which the sets of rules predict the contaminants, a number of examples are shown below. Each rule produces a list of contaminants with a certainty rating attached to each contaminant.

Example One High Certainty Rule

IF History is Wood Treatment
THEN History-related-contaminant is zinc with certainty 5
AND History-related-contaminant is copper with certainty 5
AND History-related-contaminant is mercury with certainty 5
AND History-related-contaminant is chromium with certainty 5
AND History-related-contaminant is arsenic with certainty 5
AND History-related-contaminant is tar with certainty 5
AND History-related-contaminant is phenols with certainty 5
AND History-related-contaminant is boron with certainty 5
AND History-related-contaminant is fungicides with certainty 5
AND History-related-contaminant is organic solvents with certainty 5
AND History-related-contaminant is lime with certainty 5

Example Two Low Certainty Rule

IF Grass vegetation is Creeping bent grass
THEN Grass-vege-related-contaminant is copper with certainty 2
AND Grass-vege-related-contaminant is arsenic with certainty 2
AND Grass-vege-related-contaminant is lead with certainty 2

AND Grass-vege-related-contaminant is fluoride with certainty 2

Example Three Low Certainty Rule

IF Grass-symptoms is stunted plant growth

THEN Grass-symptoms-related-contaminant is zinc with certainty 1

AND Grass-vege-related-contaminant is PCB's with certainty 1

AND Grass-vege-related-contaminant is cyanide with certainty 1

AND Grass-vege-related-contaminant is vanadium with certainty 1

AND Grass-vege-related-contaminant is manganese with certainty 1

Example Four Low Certainty Rule

IF Odour is Antiseptic

THEN Odours-related-contaminant is phenols with certainty 1

AND Odours-related-contaminant is arsenic with certainty 1

AND Odours-related-contaminant is copper with certainty 1

Example one is a high certainty rule, since the domain experts assigned a value of five to the 'historical use of the site' indicator. Therefore any contaminant produced as a result of the 'historical use of the site' indicator rule is given a certainty value of five. In comparison a contaminant produced as a result of a 'tolerant plant species' indicator or an 'odour' indicator is given a low certainty value and as result is classed as a low certainty rule, as shown in example 2 and example 4. The 'visible health symptoms' indicators (terrestrial vegetation, soil microbiology, terrestrial/aquatic invertebrates and mammals) are also classed as a low certainty rule sets, because although certain contaminants do produce visible health symptoms on certain plants other environmental factors may also cause similar symptoms, as shown in example 3, 'terrestrial vegetation visible health symptoms' example.

The terrestrial vegetation visible health symptoms indicator acts in an opposite way to the vegetation tolerant species indicator. As the health symptoms indicator identifies contaminants on the basis of poor health symptoms that appear on vegetation as a result of certain contaminants. In contrast the tolerant species indicator relies on species that have evolved to grow in extreme conditions, and therefore do not show poor health symptoms. This principle is also true for terrestrial and aquatic invertebrates.

On completion of a run through the sets of rules, a final list of possible contaminants is produced. The list is a cumulative result from all the rules fired. Figure 5.3 highlights the process of checking biotic indicators. The process may range from the result from just one rule or from all the rules, depending upon the facts available. A cumulative certainty is also produced. This again relates to the number and type of rules that produce the results. For example, the contaminant *arsenic* is produced from three of the example rules and the cumulative certainty is eight. In comparison the contaminant *lead* is only produced from one rule and the cumulative certainty is two. This therefore implies that the likelihood of the investigator identifying *arsenic* on site is greater than *lead* from the information available.

The maximum cumulative certainty that any one contaminant may achieve is forty-eight, which would relate to a contaminant being identified by all twenty sets of rules. If a contaminant is not identified by any of the rules then it does not appear on the final list of contaminants presented to the user. This eliminates the problem of having a long list of contaminants with certainty values of zero.

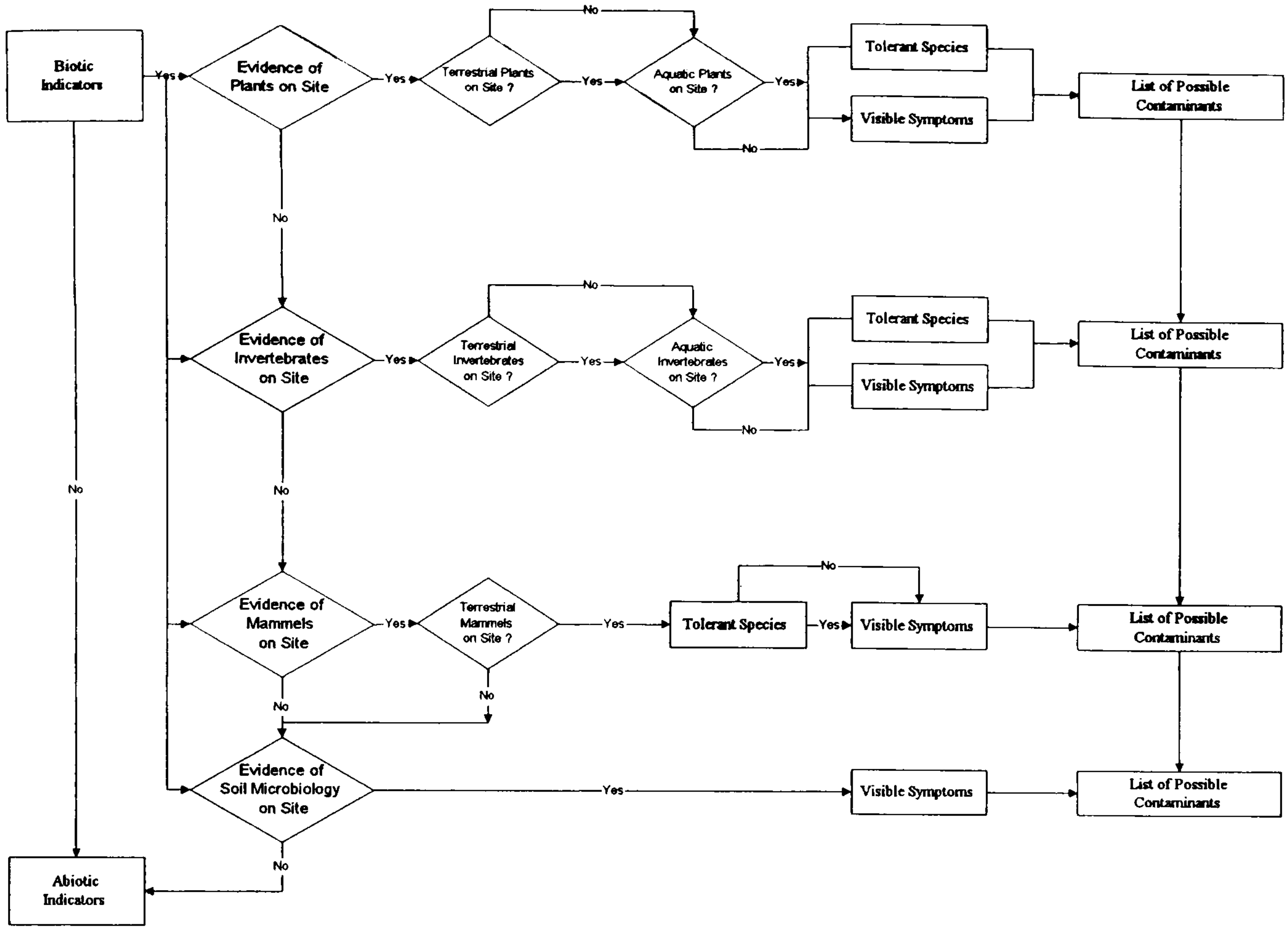


Figure 5.3: Process for Checking Biotic Indicators.

It is acknowledged that all the source biotic indicator rules within the source knowledge-base require contaminants to be present. In some situations, if a non tolerant species is observed on the site and does not show any poor health symptoms, then it may be inferred that there is no contamination present. To test such a situation would require the knowledge-based system to have knowledge of a large number of species (plants, invertebrates etc.), thus resulting in a large volume of rules being produced. It was therefore concluded that due to the fact that the certainty of such rules would be low, and that generally investigators recognise that biotic indicators only play a minor role in the investigation of contaminated land, such issues would not be addressed by the knowledge-based system.

Within the knowledge-base for source information, there also reside rules concerned with potential hazards that are likely to be present within the area being investigated. These rules have again been derived from technical literature, and in particular from information relating to past use. The identification of such hazards is relatively straight forward, as manufacturing processes of past industries are well documented. Such hazards include buried tanks, deep foundations, effluent lagoons etc. An example of a hazard related rule is given in example 5:

Example Five Hazard Rule

IF History-use is sewage-works
THEN related-hazard is drying-beds-with-highly-unstable-ground
AND related-hazard is buried-pipelines-and-tanks
AND related-hazard is overgrown-lagoons
AND related-hazard is inspection-shafts
AND related-hazard is raw-and-treated-sludges

On completion of the rules being checked, the user is presented with a list of potential hazards. This allows the user (investigator) to have an insight into the type of problems facing the investigation.

The identification of potential contaminants and hazards obviously plays a vital role in assessing a contaminated site, although such factors do not cover all the needs of the risk assessment process. Information regarding the mobility of contaminants is also extremely important. Therefore this factor had to be taken into consideration when constructing the knowledge-base for source data. Again a series of rules were developed from technical literature. The first step was to split the types of contaminants and rate them according to their properties. The initial split was to classify the contaminants into organic, inorganic or other groups of compounds. The organic group was assigned the highest risk factor, as it is widely recognised (Department of the Environment 1994b, Farias et al 1993) within the literature that organic compounds generally have a high mobility. For example, phenols are very soluble and can migrate considerable distances from their source. The risk factors regarding the other groups were ranked in descending order with 'inorganic' assigned as medium risk and 'other' as low risk.

This classification is relatively simple, whereas in reality the factors involved in contaminant migration are much more complex. For example, it is widely acknowledged (Department of the Environment 1994b, Barry 1991) that the solubility of some metals (copper, lead and zinc) may increase under acidic conditions, whereas other compounds such as arsenic may become more soluble at higher pHs. Ground composition may also affect the movement of metals, for example the presence of clay minerals and organic matter significantly retards the movement of metals. Therefore it was essential to capture such information within the knowledge-base. This meant constructing a series of rules that considered the type of contaminant and any environmental conditions (pH, ground composition etc.) that may affect the mobility.

On completion of a run through of the rules the user is presented with a high, medium or low ranking relating to the mobility of the contaminants. The drawback with such rules is that the user requires some knowledge of the pH of the ground. This type of information may not always be directly available during the preliminary investigation stage, although pH may be interpreted from other factors available, such as vegetation type, soil type or past use. Therefore rules were compiled to take these factors into consideration, example six;

Example Six Inferred Data Rule

IF trees-shrubs is bracken
THEN pH-condition is low

The final factor that had to be incorporated into the source knowledge-base was a past use or current use age factor. This was to take into consideration the industrial processes at the site over time, as industrial working practices have changed considerably over the past fifty years. The initial changes came after the Second World War, when some control over development was introduced. Until then practices were relatively unrestrained, especially during the war years. However, Myers et al (1994) suggest that the biggest change in working practices has occurred since the 1970s. This is due to increased legislation that has forced industries to develop good housekeeping and health and safety arrangements (outlined in section 2.2).

Therefore it was decided that the age of the most recent industrial process undertaken on the area under investigation would be ranked. For example, an area that was last used for wood treatment before 1970, is likely to be a higher risk than its more modern equivalent. A high risk factor was assigned to pre-1970 industry, medium to the period 1970 to 1990 and 1990 to date has a low risk ranking. This was implemented using a series of simple rules relating age of industry to ranking.

On completion of the four categories the derived results (outputs) and facts are stored within the working memory of the system. These include, type of contaminant, potential hazard, potential mobility of contaminants and an industrial process ranking related to age. The results are stored ready to be presented to the user once the system has completely finished the risk assessment process. On completion of this process the user is presented with a list of contaminants and hazards, see Figure 5.4. and Figure 5.5. The remaining outputs (including list of contaminants) are stored, so that they are available for the other knowledge-bases to access.

SELECTED CONTAMINANTS

CONTAMINANTS	CERTAINTY
arsenic	12
copper	5
zinc	5
cadmin	10
nickel	10
iron	6
lead	7
fungicides	8
toluene Ex. Matter	10
phenols	5
oils	10
cyanide	10
chromium	8
mercury	8

Figure 5.4: Example of System Output of Predicted Contaminants.

SELECTED HAZARDS

HAZARDS
drying beds with highly unstable ground
buried pipelines and tanks
overgrown lagoons
inspection shafts
raw and treated sludges

Figure 5.5: Example of System Output of Predicted Hazards.

5.3.2 Pathway Knowledge-Base

The assessment of pathways within the investigation of contaminated land involves locating possible routes for migration of contaminants within the site or off site (outlined in section 2.6.1.2). Creating a knowledge-base for such a task involves capturing knowledge from a range of subject areas, with the aim of presenting the user with a list of possible pathways.

Possible movement along pathways may occur both vertically and laterally, and this was an extremely important consideration to take into account when compiling the knowledge-base. On consultation of the technical literature there were eight major parameters that were identified: soil/rock permeability, soil/rock thickness, structural features, direction and dip of features, presence of groundwater, groundwater flow direction and groundwater velocity. Not all the parameters are relevant to both vertical and lateral movement. Therefore it was decided that two sets of rules would be compiled, one for lateral movement and another for vertical movement. Each parameter was split into categories and risk values assigned to each one of these. The risk values were derived by assessing the parameters and using technical literature to judge which parameters are likely to have the greatest effect on the outcome of the rules. The parameters were then combined to construct the lateral and vertical movement rules. The assigned risk values were summed and used to derive the results of the rules. This process was undertaken for each layer present within the zone being investigated, up to a maximum of five layers. This means that a pathway risk factor is produced for each layer both in terms of vertical and lateral movement. The vertical movement factor gives an indication of likelihood of contaminants moving down through the layer under investigation. This indication is presented to the user in terms of low, medium or high rating. A similar rating is produced for the lateral pathway factor, although an indication of the direction of movement is also given in this case. The direction is presented in simple terms of north, south, east or west.

As previously mentioned the system assesses up to five layers for both vertical and lateral movement. This was decided to be a suitable number, as it is unlikely that the investigator (user) is able to identify useful data for more than five layers from preliminary investigation data (maps etc.).

The permeability parameter is considered to be one of the most important factors within the pathway process. It was split into categories as shown in Table 5.7.

Permeability	m/s	Risk Value
impermeable	10^{-12} - 10^{-9}	0
very low	10^{-9} - 10^{-7}	1
low	10^{-7} - 10^{-5}	2
medium	10^{-5} - 10^{-3}	4
high	10^{-3} - 1	6

Table 5.7: Permeability Terms.

These categories relate to corresponding permeability values derived from technical literature (Carter,1983). Risk values were then assigned to these categories (Table 5.7), with the lowest risk corresponding to the 'impermeable' condition and the highest to the 'high permeability' situations. This reflects the fact that contaminant movement is more likely in a highly permeable layer than an impermeable layer. A non-linear scale of risk values was assigned to the permeabilities to emphasise its importance.

These categories obviously rely on the user knowing the permeability of the ground conditions within the layer being investigated. During the preliminary stage of the investigation (which this system is aimed to assist) such values may not be available. It is likely that the only information regarding geology or soil types for the investigator to use at this stage will be available from maps. Therefore it was important for the pathway knowledge-base to have the facility to relate ground type to permeability values. This meant capturing such values from the technical literature and compiling them into a series of rules. This is shown as example seven;

Example Seven

IF layer-one-material	is Upper-Greensand
THEN permeability-layer-one	is High.

In total one hundred and forty-five different material types were collected and permeability values assigned to them. These ranged from Bituminous concrete to geological units such as the Green Ammonite Beds. In a number of cases the geological unit type has a different permeability depending upon its location. In this situation a location factor was incorporated into the rule, as demonstrated in example eight. The permeability values for each material type were sourced from NRA Groundwater Vulnerability Maps of England and Wales (NRA 1995). This proved to be the most effective way of collecting the vast amount of data required.

Example Eight

IF layer-one-material	is Great-Oolite-Limestone
AND location	is South-Oxford
THEN permeability-layer-one	is High.

The thickness parameter also plays a vital role when assessing possible pathway movement in the vertical direction, as the thicker the layer the longer it is likely to take a contaminant to move through it, thus reducing the risk of contaminant movement. Again suitable categories were obtained from the technical literature, to take into account the range of thicknesses available. The limits for the thicknesses were derived from the BS5930 (1999) description of bedding plane spacing. Once the thickness categories had been compiled, risk values were assigned to them as shown in Table 5.8. These values take into account the effects that the thickness parameter has on pathway assessment. A very thin layer has the highest value, due to the fact that such thickness would have a very limited effect in slowing contaminant movement.

Description of Thickness	Limits for Thickness	Risk Value
Very Thick	over 2 m	0
Thick	0.6 – 2 m	1
Medium	0.2 – 0.6 m	2
Thin	60 mm - 0.2 m	3
Very Thin	20 – 60 mm	4

Table 5.8: Description of Thickness.

To obtain the thickness parameter from the user, the user is simply asked which category the layer they are describing fits into.

Another important factor within the pathway assessment process is identification of the presence of structural features. These features range from man-made structures, such as sewage pipes, electricity cables, drainage channels etc., to geological features that include folds, faults and joints. Such features create possible movement pathways. Therefore if a feature is present within a layer the risk of contaminant movement within the layer increases. A risk value was assigned to this parameter, as highlighted in Table 5.9.

Structural Feature	Risk Value
Yes	1
No	0

Table 5.9: Risk Value for Structural Features.

The information regarding structural features is input by the user. A simple Yes/No question, asks the user to highlight the major structural feature within the area under investigation. This is then used later in the pathway assessment process. The user is expected to obtain information concerning the features from geological maps, guides etc. for geological features, and site plans, ordnance survey maps etc. for man-made structures. Although the presence of structural features is important, the direction and the dip of the feature is even more important within the pathway assessment process. Therefore this was another parameter that required consideration within the pathway knowledge-base. Geological literature was consulted, in order to obtain likely parameters that relate to structural features.

In the case of bedding, it was decided to adopt terms that are normally used to describe the attitude of the axial surface (plane) of a fold as shown in Table 5.10.

Dip of Bedding Plane	Terms	Vertical Risk Value	Lateral Risk Value
0°	Horizontal	0	3
1 - 10°	Sub-Horizontal		
10 - 30°	Gentle	1	2
30 - 60°	Moderate		
60 - 80°	Steep	2	1
80 - 89°	Sub-Vertical	3	0
90°	Vertical		

Table 5.10: Attitude of Bedding Plane and Terms.

Bedding planes are usually regarded as pathways for contaminant movement when joints between the planes have developed which allow groundwater to move along the planes, and therefore transport contaminants.

It was decided that this approach of classification could also be used for man-made features, for example, inspection shafts, electricity cable and drainage channels. Such features are generally vertical (e.g. inspection shafts) or horizontal (e.g. electricity cables), therefore it is likely that only these two categories are required. If there is any variation, other options are available for describing them.

As outlined in Table 5.10 two sets of risk values were assigned to the terms. This was to take into account lateral and vertical movement of contaminants. A lateral feature will have very little effect on vertical movement and vice versa, therefore the risk values were ranked accordingly. A number of terms within Table 5.10 were assigned the same risk values, for example Sub-Vertical and Vertical are both given the vertical risk value of 3. This was due to the fact that the difference between the two terms is only small (10°) and therefore it was concluded that this did not pose a significant difference in terms of risk.

Another geological feature that needed to be described in terms of dip is a fault. Faults have their own classification system, which means they are described according to the attitude of the fault plane. The dip of the fault plane is the angle between the fault plane and the horizontal. Table 5.11 outlines the fault classification system and the risk values assigned to the terms. Again different values were assigned to vertical movement and horizontal movement for reasons mentioned previously.

Due to the classification system for bedding plane dip and fault plane dip being different, there is a difference between the risk values assigned to the angles of dip. For example, using the bedding classification; 10° - 60° is assigned a vertical risk value of 1, whereas, in comparison a vertical risk value of 1 is assigned to angles up to 45° using the fault system. This reflects the fact that a fault is likely to act as a better pathway than a bedding plane and hence has a higher risk value for a lower angle.

Dip of Fault Plane	Terms	Vertical Risk Value	Horizontal Risk Value
< 45°	low-angle	1	3
> 45°	high-angle	2	2
90°	Vertical	3	0

Table 5.11: Fault Classification.

The direction (orientation) in which the feature runs through the area being investigated was also required by the system. Therefore this meant asking the user to input the direction (orientation) of the feature in a simplified form, for example 'North to South' or 'North-East to South-West'.

The final parameter that needed to be represented within the pathway knowledge-base was the presence of groundwater. Groundwater plays an important role in the movement of contaminants, within the sub-surface layers, and therefore contributes to the pathway assessment process. The initial assessment for the knowledge-base was to confirm if water was present within the layer being assessed. This was achieved by simply asking the user, for which a Yes or No reply was obtained, for each layer. Risk values were than assigned to these answers, as highlighted in Table 5.12. The presence of groundwater assists the movement of contaminants, therefore increasing the risk, which is reflected in its risk value.

Groundwater	Risk Value
Yes	**1**
No	**0**

Table 5.12: Groundwater Risk Value.

The second parameter related to groundwater, is concerned with the direction of groundwater movement. This is obviously only relevant within the assessment of lateral movement of contaminants. The user is required to input the direction of groundwater flow using a simple notation, such as ‘north to south’, or ‘north-east to south-west’ etc., in a similar manner to that used within the structural features section. Such information is extremely useful for assessing if contaminants are likely to move towards a target, although this is enhanced if combined with other information, such as the direction in which structural features run across the area being investigated. It was therefore decided to combine these data when compiling the risk values, and these are outlined in Table 5.13.

		Groundwater Direction			
		N-S	E-W	NE-SW	NW-SE
Direction (orientation) of structural feature	N-S	2	0	1	1
	E-W	0	2	1	1
	NE-SW	1	1	2	0
	NW-SE	1	1	0	2

Table 5.13: Combined Groundwater Direction and Structural Features Parameters.

As highlighted within Table 5.13, the worst case scenario occurs when the groundwater is flowing in a similar direction to that of the structural feature. This is due to the fact that the groundwater is likely to flow along the line of the feature and transport contaminants. For this reason the highest risk value was assigned to this scenario. The reverse is true, when the structural feature runs perpendicular to the direction of groundwater flow. For this reason, such a scenario was given the lowest risk value.

Another important sub-parameter of groundwater, is the velocity of the groundwater as this plays a role in increasing the movement of contaminants. The velocity of groundwater was split into three main categories, as highlighted in Table 5.14. This was achieved by consulting domain experts using the questionnaire, described in selection 5.2.3 (knowledge acquisition). The results of the questionnaire assisted in deriving typical values for slow moving groundwater through to fast moving groundwater. Risk values were assigned to each one, with the highest value relating to the highest risk.

The user is expected to input which category the groundwater they are describing fits into. This parameter is then used later in the lateral movement pathway assessment process. The information required for the user to make such decisions may be collected from NRA groundwater maps, old site investigation reports etc.

Velocity of Groundwater m/day	Term	Risk Value
< 0.01	Slow	1
0.01 – 0.1	Medium	2
>0.1	Fast	3

Table 5.14: Groundwater Velocity Terms and Risk Values.

Within the pathway knowledge-base, once each one of the eight parameters have been assigned a value, they are combined within a series of rules for vertical and lateral movement. The rules were compiled in order to take into consideration all the possible scenarios, ranging from the worst case scenario, when contaminants are likely to move easily to the best case, when contaminant movement is hampered.

Example Nine

Best Case Scenario for Vertical Movement

		Risk Value
IF	Permeability is impermeable	0
AND	Thickness is very thick	0
AND	Structural features is No	0
AND	Groundwater is NO	0
THEN	Pathway Risk is Very Low	0

Example Ten

Worst Case Scenario for Vertical Movement

		Risk Value
IF	Permeability is High	6
AND	Thickness is Very thin	4
AND	Structural features is Yes	1
AND	Dip of feature is Vertical	3
AND	Groundwater is Yes	1
THEN	Pathway Risk is Very High	15

The result from the rule is a pathway risk (Very Low, Low, Medium, High or Very High), as highlighted in examples nine and ten. The pathway risk was derived as a result of a simple summation of the individual risk values assigned to each parameter within the rule. This was viewed as an acceptable approach due to the fact that a rule that contains a large number of parameters with high risk values represents a higher risk than a rule that contains parameters with low risk values assigned to them. Table 5.15 defines categories relating to the summation of the risk values.

Pathway Risk	Σ risk value
Very Low	0 – 3
Low	4 – 5
Medium	6 – 8
High	9 – 12
Very High	12 – 15

Table 5.15: Pathway Risk Values.

For lateral movement assessment not all the parameters previously outlined were used. For example, the thickness of a layer will not have a great deal of effect on the lateral movement of contaminants. Therefore this parameter was excluded from the lateral movement assessment process. However, the direction of structural features and groundwater movement play a vital role within this process, and so were included along with the parameter relating to the velocity of the groundwater. Therefore, the rules relating to lateral movement are slightly different to the rules compiled for vertical movement.

Example Eleven

Worst Case Scenario for Lateral Movement

		Risk Value
IF	Permeability is High	6
AND	Structural Feature is Yes	1
AND	Dip of feature is Horizontal	3
AND	Direction of feature is North-South	(1)
AND	Groundwater is Yes	1
AND	Groundwater Direction is North-South	(1)
AND	Groundwater velocity is Fast	3
THEN	Pathway Risk is Very High to South	16

The result from the rules is a lateral pathway risk value as highlighted in example eleven. The pathway risk is derived from the sum of the risk values, in a similar fashion to that used within the vertical movement process. The brackets around the two risk values indicate a combined value, derived from Table 5.13. Due to the increased number of parameters and hence the possibility of higher totals, the risk categories are derived slightly differently to those used within the vertical movement process. Risk values are outlined in Table 5.16.

Pathway Risk	Σ Risk Value
Very Low	0-2
Low	3-5
Medium	6-10
High	11-13
Very High	14-16

Table 5.16: Lateral Movement Risk Values.

On completion of both the vertical movement and lateral movement assessment process, the pathway knowledge-base produces a pathway risk for both movement types. This result is produced for every layer described by the user. This pathway risk is then combined with data from the source knowledge-base and target knowledge-base, in order to produce an overall risk profile, see Figure 5.6.

```
Possible PathwayS

PATHWAYS Layer 1
-------------------------------------
low



lateral pathway Layer 1
-------------------------------------

low-north
low-south
low-east
low-west

PATHWAYS layer 2
-------------------------------------
very-low

lateral pathway Layer 2
-------------------------------------
very-low-north
very-low-east
very-low-west
very-low-south
```

Figure 5.6: Example of System Output of Pathway Risk Factors.

5.3.3 Target Knowledge-Base

The identification of sensitive targets is vital in order to complete the source-pathway-target assessment process. A risk cannot exist unless there is a plausible scenario, therefore if a sensitive target is not identified, the other components have no worth.

A range of possible targets that are usually associated with the investigation of a contaminated site was complied from technical literature (DD175 1988, Harris et al 1995, Barry 1991 and Young et al 1997). These are listed below:

- Site workers
- Future occupiers or users
- Neighbouring occupiers and users
- Soil quality
- Surface and groundwater quality
- Ambient air quality
- Flora and fauna
- Buildings and services
- Mineral resources

On examination of potential targets, it became apparent that certain targets require quantitative data, which may not be available from the preliminary investigation. This meant investigating and eliminating targets that the system is able to assist with. It was concluded that the following targets; site workers, soil quality, surface and groundwater quality, ambient air quality, flora and fauna, buildings & services and mineral resources were unsuitable for the system. Such targets generally require quantitative data relating to concentration of contaminants, volume of contaminants and soil/groundwater chemical composition. Although the remaining targets (future occupiers or users, neighbouring occupiers and users) also generally require the type of data previously mentioned, it is still possible to predict a useful risk assessment for such targets using preliminary investigation data.

It was also decided that although quantitative data was required for buildings and services targets, a simple risk assessment could be carried out by the system. In addition the risk to site workers was also addressed in broad terms, however this element was contained within its own knowledge-base (health & safety knowledge-base).

The initial aspect of a target that is relatively straight forward to assess is the risk to future users of the site, as certain end uses of the site are more sensitive to contaminants than others. Therefore a series of rules were compiled for this factor. The result from the rules is a risk factor, high, medium, low. The risk factor is based on exposure (pathway) to end users via ingestion, dermal contact and inhalation. Therefore contaminated areas that are likely to be permanently covered are a lower risk than exposed areas. This is illustrated by examples twelve and thirteen;

Example Twelve

Low Risk Example

IF future end use is permanent paved area

THEN Target value is Low

Example Thirteen

High Risk Example

IF future end use is public open space

THEN Target value is High

If the contaminated zone and the target are one and the same it is not necessary to include a pathway risk analysis. However, when assessing neighbouring occupiers and users these parameters are required. A similar set of rules to those outlined above were compiled, for neighbouring use, but with an additional check for possible pathways to the target area. In order to assess whether potentially vulnerable targets exist, the user is required to input the land use for the four possible neighbouring zones (north, east, south and west) by simply selecting an appropriate land use from the land use classification suggested by Jewell et al (1993), shown in Table 5.17.

The land use classification outlined in Table 5.17 was used to cover a range of land uses that may be found neighbouring a potential contaminated area. The term "neighbouring use", does not indicate that this area has to be completely off-site. It can simply be a zone within the site.

To assess the risk to such neighbouring use, the system draws upon facts derived from both the source and pathway knowledge-bases. From the source knowledge-base facts regarding mobility of contaminants are accessed and from the pathway knowledge-base lateral and vertical pathway risk facts are utilised. The facts are combined along with the target value for the neighbouring use and an overall risk for the layer being tested is derived. The values assigned to the neighbouring uses (target value) outlined in Table 5.17, were derived from the results of the questionnaire, as described in section 5.2.3. The values range from low through to high, with low indicating a use that is unlikely to be vulnerable to contaminants and high relating to a very vulnerable use.

Neighbouring Land Use	Target Value
School	High
Public Open Space	Medium
Agricultural	High
General Commercial	Low
Low Density Residential	High
Medium Density Residential	High
High Density Residential	High
Light Industrial	Low
High Industrial	Low
Permanently Paved	Low

Table 5.17 Target Values Relating to Land Use.

This process is undertaken for every layer within the zone being investigated, up to a maximum of five layers, resulting in an overall risk being derived for each layer. The system searches for a neighbouring use that relates to the direction of pathway lateral risk. For example, if the direction of pathway lateral risk is south, the system will not fire rules relating to the north, east and west neighbouring uses (see example fourteen, layer one overall risk). This is due to the fact that the potential contaminants are unlikely to be transported towards these targets. Therefore the pathway risk lateral value is compared with the neighbouring use to the south. In the case of example twelve the neighbouring use to the south has a high target value and the pathway risk lateral value is high to the south. Therefore the overall risk of contaminants migrating to the south and causing problems to the neighbouring area in the south is high.

Example Fourteen

IF contaminant mobility is High

AND pathway risk-one lateral is South-High

AND Neighbouring use South is High

THEN Overall risk one is High-South

As previously mentioned, an overall risk is calculated for every layer up to five layers, within the zone under investigation. This process is relatively straightforward for layer one as shown in the example fourteen. However, when investigating a layer that has another layer overlying it, the situation becomes more complex. This is due to the fact that the vertical pathway properties of the layer above are required. This involved constructing rules, that combined these facts, as shown in the example below.

Example Fifteen (using vertical pathway facts from layer above)

IF contaminant mobility is High

AND pathway risk-one vertical is high

AND pathway risk-two lateral is south-high

AND Neighbouring use-south is High

THEN overall risk two is High South

This process of using vertical pathway risk values continues, as more layers overlie the layer under investigation. During this process, it becomes extremely important that the rules are written efficiently. Therefore to facilitate this, the pathway vertical facts were combined using a sub-set of rules before the overall risk rules were fired. This eliminates having to compile a rule for every situation (see example fifteen).

On completion of firing all the relative rules, within this section of the knowledge-base, the user is presented with an overall risk factor which indicates the direction of risk, for every layer under investigation.

The final section of the target knowledge-base is the assessment of buildings and services and although, as previously mentioned, quantitative data is usually required for this, it was decided that a simple assessment would be presented to the user. Even a simple assessment may be useful for the investigator at the preliminary stage of an investigation.

As highlighted by Smith (1991), buildings and services are also classed as sensitive targets that require assessment. Cairney (1995) suggests that the main components of buildings and services that cause most concern are concrete, steel, iron, plastic or rubber. Cairney (1995) also outline the contaminant conditions which may pose risk to construction materials, including risk factors. These conditions have been reproduced and combined with data from the source and pathway knowledge-base to produce a set of rules that highlight the risk to construction materials, as shown in example sixteen.

Example Sixteen

IF pH is Low
AND Groundwater is Yes
AND Contaminant is Sulphate
THEN Concrete-construction-material is High Risk.

The potential risk to construction materials is thus presented to the user. It ranges from low to high, with risk depending upon the number of parameters available to make the decision. The example highlights a high risk example, although if only one parameter was available for example, pH low, then the risk would be low.

5.3.4 Health and Safety Knowledge-Base

The final knowledge-base within the system is concerned with health and safety issues. When investigation personnel enter a potentially contaminated site it is essential that adequate health and safety precautions are taken. This component of the system is aimed at presenting the user with a list of required precautions and health hazards, see Figures 5.7 & 5.8. It consists of a series of simple rules compiled from the technical literature (Stead et al 1996).

The rules utilise data derived from the source knowledge-base. These data consists of the likely contaminants that may be present, as shown in examples seventeen and eighteen;

```
Protective Measures
-------------------------------------------
cadmium-protection:gloves
cadmium-protection:overalls
cadmium-protection:suitable-respiratory-protective-equipment
lead-protection:gloves
lead-protection:overalls
lead-protection:suitable-respiratory-protective-equipment
zinc-protection:gloves
zinc-protection:overalls
zinc-protection:suitable-respiratory-protective-equipment
copper-protection:gloves
copper-protection:overalls
copper-protection:suitable-respiratory-protective-equipment
phenols-protection:gloves
phenols-protection:overalls
methane-protection:infrared-gas-analyser
methane-protection:ventilation-control
methane-protection:suitable-respiratory-protective-equipment
methane-protection:ban-ignition-sources
```

Figure 5.7: Example of System Output of Protective Measures.

Example Seventeen

Health Hazard Rule

IF Contaminant is Mercury

THEN Health-Hazard is highly-toxic-by-ingestion

AND Health-Hazard is highly-toxic-by-skin-absorption

AND Health-Hazard is inhalation of dust.

Example Eighteen

Safety Precautions Rule

IF Contaminant is Mercury

THEN Protective-measures is Protective gloves

AND Protective-measures is Overalls

AND Protective-measures is Respiratory-protective-equipment.

```
Health-Hazard
-----------------------------------------
zinc-hazard:toxic-dust-inhalation
zinc-hazard:corrosive_to_skin
zinc-hazard:carcinogenic
copper-hazard:dust-inhalation_damages_respiratory_system
mercury-hazard:highly-toxic-ingestion,dust-inhalation,and_skin-absorption
chromium-hazard:toxic-dust-inhalation
chromium-hazard:suspected_carcinogen
chromium-hazard:corrosive_to_eyes,skin,nasal-passage
arsenic-hazard:skin-contact_linked_to_cancer
phenols-hazard:acute-exposure_can_be_lethal
phenols-hazard:chronic-exposure_can_result_to_vomiting,diarrhoea,liver_and_kidney_damage
phenols-hazard:toxicity-via-ingestion,skin-absorption,inhalation
boron-hazard:compounds-corrosive_to_skin,eyes
cadmium-hazard:toxic-ingestion,dust-inhalation,and_skin-absorption
lead-hazard:toxic-ingestion,dust-inhalation,and_skin-absorption
lead-hazard:cumulative-poison
nickel-hazard:toxic-dust-inhalation,and_skin-absorption
nickel-hazard:suspected-carcinogen
cyanide-hazard:acute-toxicity-via-ingestion,skin-absorption,eyes
cyanide-hazard:dust,gas-inhalation_results_in_collapse_and_death
methane-hazard:asphyxiation
methane-hazard:explosive_in_confined_spaces
methane-hazard:flammable-limits-5-15%
```

Figure 5.8: Example of System Output of Health Hazards.

The results from the rules are only the precautions required in a worst case scenario. Therefore such precautions may not be required in all cases. This means it is important that the user is aware of this, and it is up to the investigator to take the necessary precautions. The system only provides a simplified result and the issues are likely to be more complex.

5.4 Implementation of Knowledge

On completion of the representation of knowledge into a suitable format, the implementation of such rules was required using the CLIPS software. CLIPS allows knowledge-bases to be developed and executed in a modular environment. Modules are defined using the *defmodule* construct, which allows sets of constructs (*deftemplate*, *defrule*, *deffacts*) to be defined within a module in such a manner that explicit control can be maintained over the constructs. Thus, it is possible to interchange data (code, rules, facts etc.) from one module to another. This allows modules to be defined for the definition of rules, generation of results and displaying results to the user.

The rule module (*defmodule RULE*) example shown has been compiled to define how rules are structured. The module contains two constructs, *deftemplate* and *defrule*.

The *deftemplate* construct is used to create a template which can then be used to access fields by name. The *deftemplate* construct informs CLIPS of the list of valid slots for a given name (keyword). In the rule template the keywords *if* and *then* are defined as *multislots*. The *multislot* term allows more than one value to be assigned to the keyword. This type of template may be used for a range of tasks where definitions of keywords (task) is required. The use of a template is also demonstrated in *the Rule Selection* and *Rule Generate Module* examples.

The *defrule* construct allows rules to be defined within CLIPS. Rules within CLIPS consist of a collection of conditions and actions to be taken if the conditions are met. The construct contains a Left-Hand Side (LHS) and Right-Hand Side (RHS). The LHS is made up of a series of conditional elements which consist patterns to be matched against. The arrow (=>) separates the LHS from the RHS. The RHS contains a list of actions to be performed when the LHS of the rule is satisfied.

Example of Rule Module

```
;;******************
;; The RULES module
;;******************

(defmodule RULES (import MAIN ?ALL) (export ?ALL))

(deftemplate RULES::rule
 (slot certainty (default 100.0))
 (multislot if)
 (multislot then))

(defrule RULES::throw-away-ands-in-antecedent
 ?f <- (rule (if and $?rest))                 ; Left-Hand Side (LHS)
 =>
 (modify ?f (if ?rest)))                      ; Right-Hand Side (RHS)

(defrule RULES::throw-away-ands-in-consequent
 ?f <- (rule (then and $?rest))
 =>
 (modify ?f (then ?rest)))
```

The defined *Rule Module* allows the rule format to be used by any of the knowledge-bases, from the source knowledge-base to the health & safety knowledge-base. The two examples shown, pathway rule example and health & safety rule example demonstrates the modules in use.

Example of Pathway Rule

```
;;*********************************
;; *layer three impermeable groundwater slow*
;;*********************************

(defmodule CHOOSE-PATHWAY)

(Import RULE ?ALL)
(Import Main ?ALL)
(defrule CHOOSE-PATHWAY)

(rule (if permeability-layer-three is impermeable
          and structural-features-three is yes
          and dip-three is horizontal
          and direction-three is NW-SE
          and groundwater-three is yes
          and groundwater-direction is S-N
          and groundwater-velocity is slow)
          (then pathway-risk-three-north  is medium))

(rule (if permeability-layer-three is impermeable
          and structural-features-three is yes
          and dip-three is sub-horizontal
          and direction-three is NW-SE
```

and groundwater-three is yes
and groundwater-direction is S-N
and groundwater-velocity is slow)
(then pathway-risk-three-north is medium))

Example of Health & Safety Rule

(defmodule CHOOSE-HEALTH_SAFETY)

(Import RULE ?ALL)
(Import Main ?ALL)
(defrule CHOOSE-HEALTH_SAFETY)

(rule (if contaminant is sulphur)
(then protective-measures is sulphur-protection:gloves
and protective-measures is sulphur-protection:overalls
and protective-measures is sulphur-protection:suitable-respiratory-protective-equipment"if-necessary"
and protective-measures is sulphur-protection:face-shields_if_splashing))

(rule (if contaminant is phenols)
(then protective-measures is phenols-protection:gloves
and protective-measures is phenols-protection:overalls
and protective-measures is phenols-protection:suitable-respiratory-protective-equipment"if-necessary"
and protective-measures is phenols-protection:face-shields_if_splashing))

Both examples start by defining the module name; in the pathway example the module is assigned the name *CHOOSE-PATHWAY*. The next line of code imports the *RULE Module*. This allows data to be structured into a *if-then* rule format. This means that when the rule is fired the system matches facts against the *if* statements and a *then* statement is returned if a match is found.

In addition to the definition of the rules, modules are compiled to define valid facts. This process is carried out using the *deffacts* constructs. This enables the facts to be assigned a name and values that are accepted by the rules. This allows facts to pass from the user interface to the knowledge-base via a text format as displayed in the facts list example below.

Examples of Facts List

(attribute (name has-history) (value timber_yard))
(attribute (name current-use) (value timber_product_manufacturing))
(attribute (name structural-features-one) (value no))
(attribute (name structural-features-two) (value no))
(attribute (name structural-features-three) (value no))
(attribute (name structural-features-four) (value no))

```
(attribute (name thickness-two) (value v.thick))
(attribute (name layer-two-material) (value till))
(attribute (name neighbouring-use-east) (value Light-Industrial))
(attribute (name neighbouring-use-south) (value Low-Density-Residential))
(attribute (name neighbouring-use-west) (value Low-Density-Residential))
(attribute (name neighbouring-use-north) (value Heavy-Industrial))
```

Once rules have been defined and facts have been received the system matches the facts with appropriate rules and generates facts that may be output to the user via the *Print Selected Rules Module* or passed to another *Rule Selection Module* and used to generate further facts. For example facts from the *Generate Contaminant Module* are passed to the *Hazards Selection Rule Module* in order to generate facts relating to hazards from past usage of the area under investigation.

The pathway selection rule example demonstrates the process of the selection of appropriate rules and the generation of facts. It can been seen that within the process an *attribute* and *names* are defined. These relate to the facts list input by the user via the text file. The text file is a product of the user interface described in section 5.5 with an example shown in Figure 5.14.

Example of Rule Selection and Rule Generate Modules

```
;;**********************************
;;* PATHWAYS layer3 SELECTION RULES *
;;**********************************

(defmodule PATHWAYS3 (import MAIN ?ALL))

(deffacts any-attributes
  (attribute (name pathway-risk-three) (value any)))


(deftemplate PATHWAYS3::pathways3
  (slot name (default ?NONE))
  (multislot path3 (default any)))

(deffacts PATHWAYS3::the-pathways3-list

;;****************
;;*pathway3  only*
;;****************

      (pathways3 (name very-high)    (path3 very-high))

      (pathways3 (name high)       (path3 high))

      (pathways3 (name medium)      (path3 medium))
```

```
    (pathways3 (name low)        (path3 low))

    (pathways3 (name very-low)    (path3 very-low)))


;;********************
;;*GENERATE PATHWAYS3
;;********************

(defrule PATHWAYS3::generate-pathways3
 (pathways3 (name ?name)

    (path3 $? ?pa3 $?))

 (attribute (name pathway-risk-three) (value ?pa3))

 =>

 (assert (attribute (name pathways3) (value ?name))))
```

On completion of the generation of facts/results a module known as *Print Selected Rules* presents the user with the results. This type of module has been created for each result from each knowledge-base. An example of the module relating to the pathway knowledge-base is displayed below.

Example of Print Selected Rules Module

```
;;*****************************************
;;* PRINT SELECTED PATHWAYS Layer 3 RULES *
;;*****************************************

(defmodule PRINT-RESULTS5 (import MAIN ?ALL))

(defrule PRINT-RESULTS5::header ""
  (declare (salience 10))
  =>
  (printout t t)
  (printout t "      SELECTED PathwayS" t t)
  (printout t " PATHWAYS layer 3                 " t)
  (printout t " ----------------------------------------" t)
  (assert (phase print-pathways3)))

(defrule PRINT-RESULTS5::print-pathways3 ""
 ?rem <- (attribute (name pathways3) (value ?name))
 (not (attribute (name pathways3))))
 =>
 (retract ?rem)
 (format t " %-24s %2d%%%n" ?name))

(defrule PRINT-RESULTS5::remove-poor-hazards-choices ""
 ?rem <- (attribute (name pathways3)))
```

```
  =>
 (retract ?rem))

(defrule PRINT-RESULTS5::end-spaces ""
   (not (attribute (name pathways3)))
  =>
  (printout t ))
```

5.5 User Interface

A systematic and logical approach to data entry was considered vital in order for the knowledge-based system to receive suitable data from the user. To achieve this, it is important that the user interface was developed in a suitable manner and using an appropriate development tool. The user interface described in this section is concerned with data entry into the knowledge-based system and is independent of the database data entry interface described earlier. This means that data entered via the knowledge-based system interface is not saved within the database but as a separate text file.

The design of the interface was undertaken using the principles outlined in section 3.6.5 as guidelines. A Visual Basic© development tool was selected to build the interface, as this allows the developer to present the user with several options for entering data and navigating around the system. This met the two principles of consistency and compatibility, due to the fact that information presented to the user is in a windows environment, and ensured consistency with the majority of commercial software packages used in design offices, thus presenting the user with a familiar work environment.

The principle of guiding the user through the system was considered vital, in order to obtain the required data from the user. Therefore the user is guided from the initial display within the interface through to the final screen display. For example, the initial display checks to ensure the user has collected suitable data (preliminary investigation data) before continuing onto the first set of options, regarding contaminant source data.

Evidence is provided using a simple dialog box, asking users with limited experience in the subject area to select the help button. This presents the user with a help file. This allows the user to browse through an overview of the issues involved in the investigation of contaminated land, and therefore allows them to understand the type of information that should be collected during the preliminary investigation stage. This type of help underlines the principle of adaptability, as it allows a novice within the subject area to learn more about the issues involved before continuing, but also allows a user of a higher skill level to bypass the help and continue with the data entry.

Once into the main body of the system the user has three main sections to navigate through. Each section collects data regarding the different components within the risk assessment process. Thus, the sections relate to the source, pathway and target components described within section 2.6.1.1.

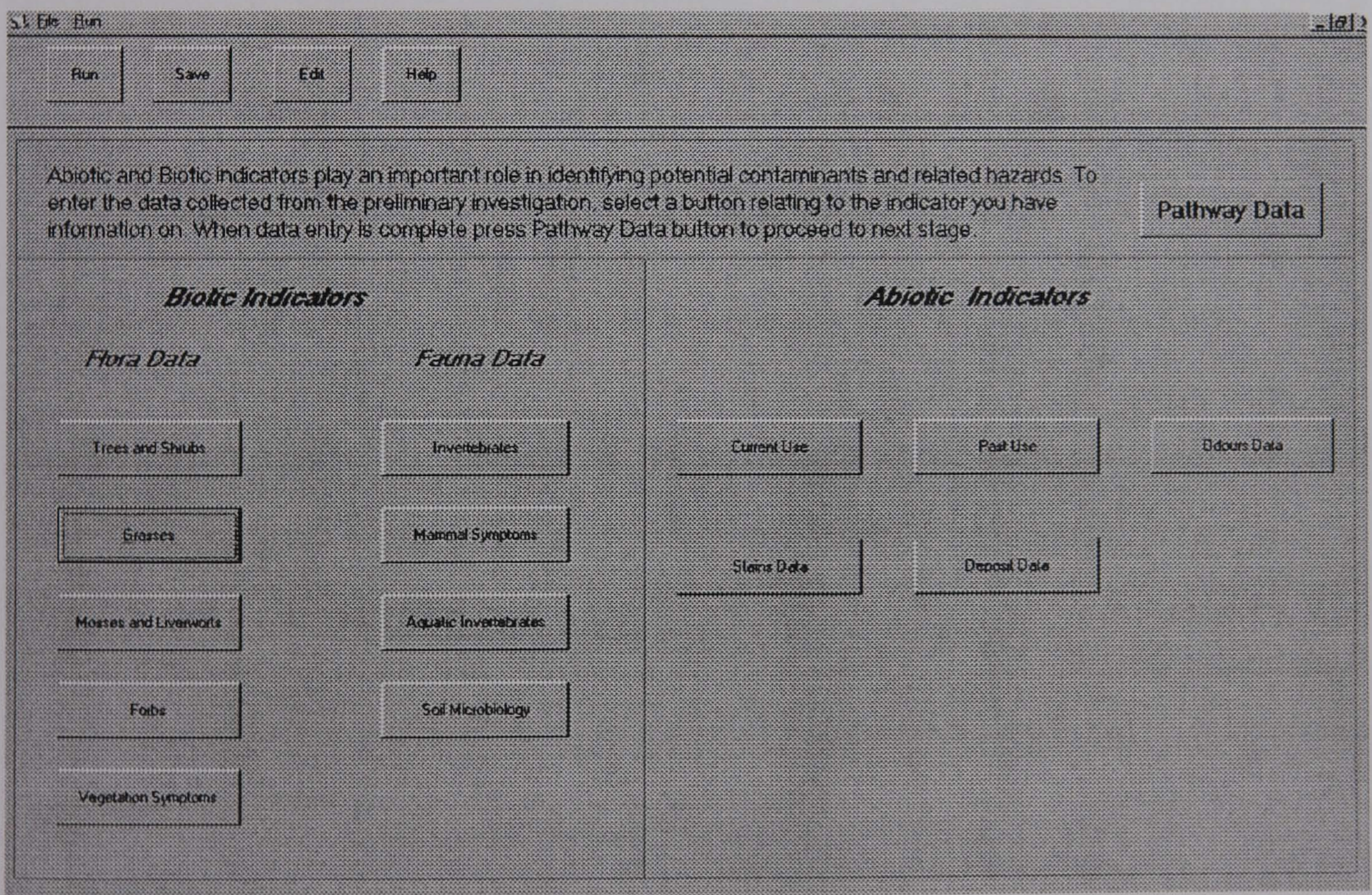


Figure 5.9: Source Section Introductory Interface Display Example.

Within each section the user is presented with an introductory screen display. This gives a background to the type of information required within the section, and presents the user with a series of command buttons relating to information required, as shown in Figure 5.9. Thus within the source section, the user is asked to enter information relating to both biotic (trees & shrubs, grasses, mosses & liverworts, forbs, vegetation symptoms, invertebrates, mammal symptoms, aquatic invertebrates and soil microbiology) and abiotic (current use, past use, odours, stains and deposit data) indicators.

On activating one of the command buttons, a data entry form is displayed to the user. To reduce the number of operations required by the user, data entry involves the user highlighting check boxes as shown in Figures 5.10. This avoids the user having to type the required information, which in turn eliminates misspelling or mis-typing during data entry.

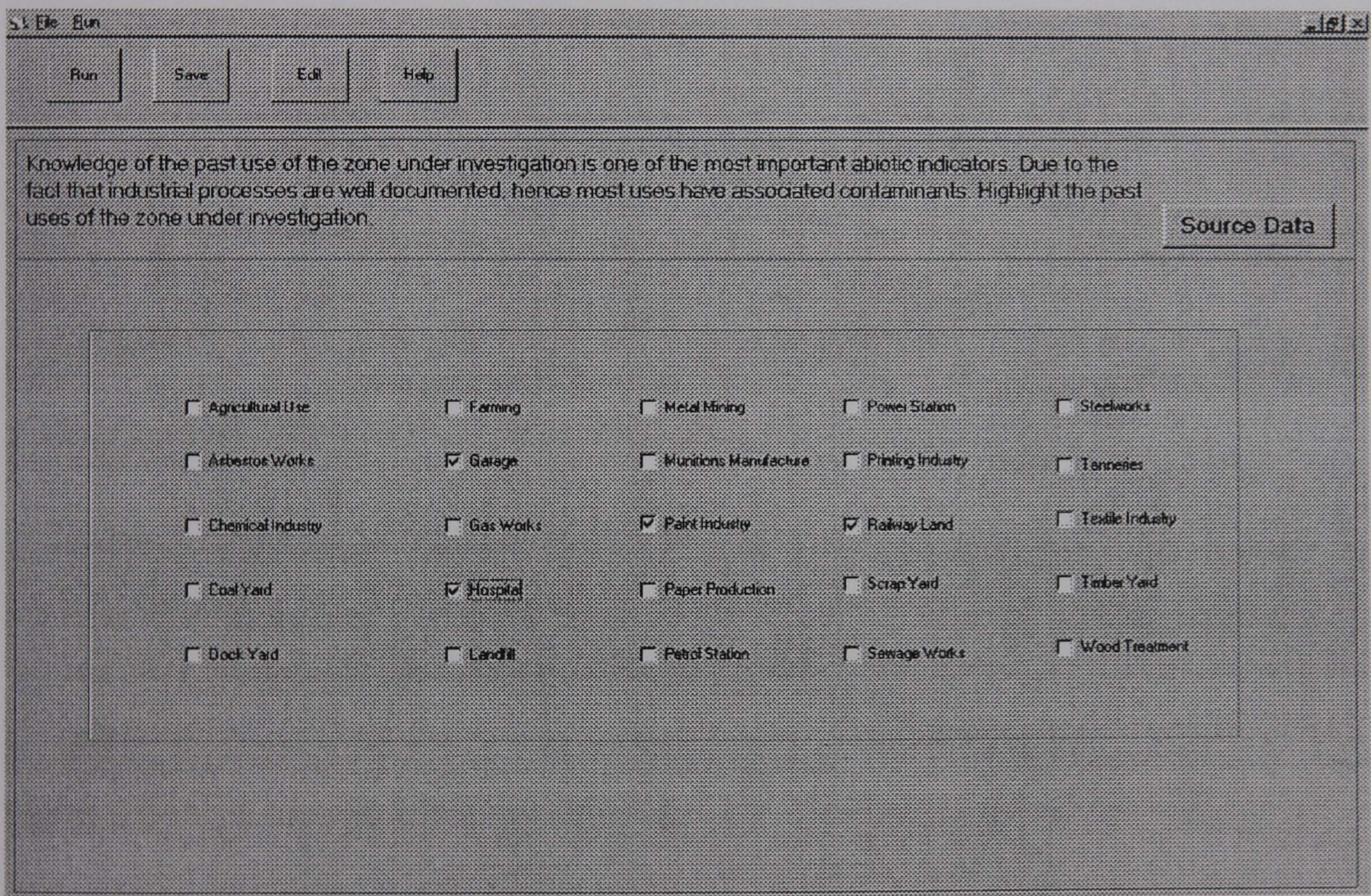


Figure 5.10: Check Box Interface Display Example (Past Use).

The check box system example from the source section shown in Figure 5.10 allows the user to select past uses of the area under investigation. Using a check box system enables the user to enter more than one entry under a specific biotic or abiotic indicator. Therefore, within the history example the user can enter more than one past use.

On completion of the data entry form the user returns to the introductory screen for that section and from this screen the user may advance to the next section. For example, as displayed in Figure 5.9, the user clicks the "Pathway Data" button to move to the pathway related section.

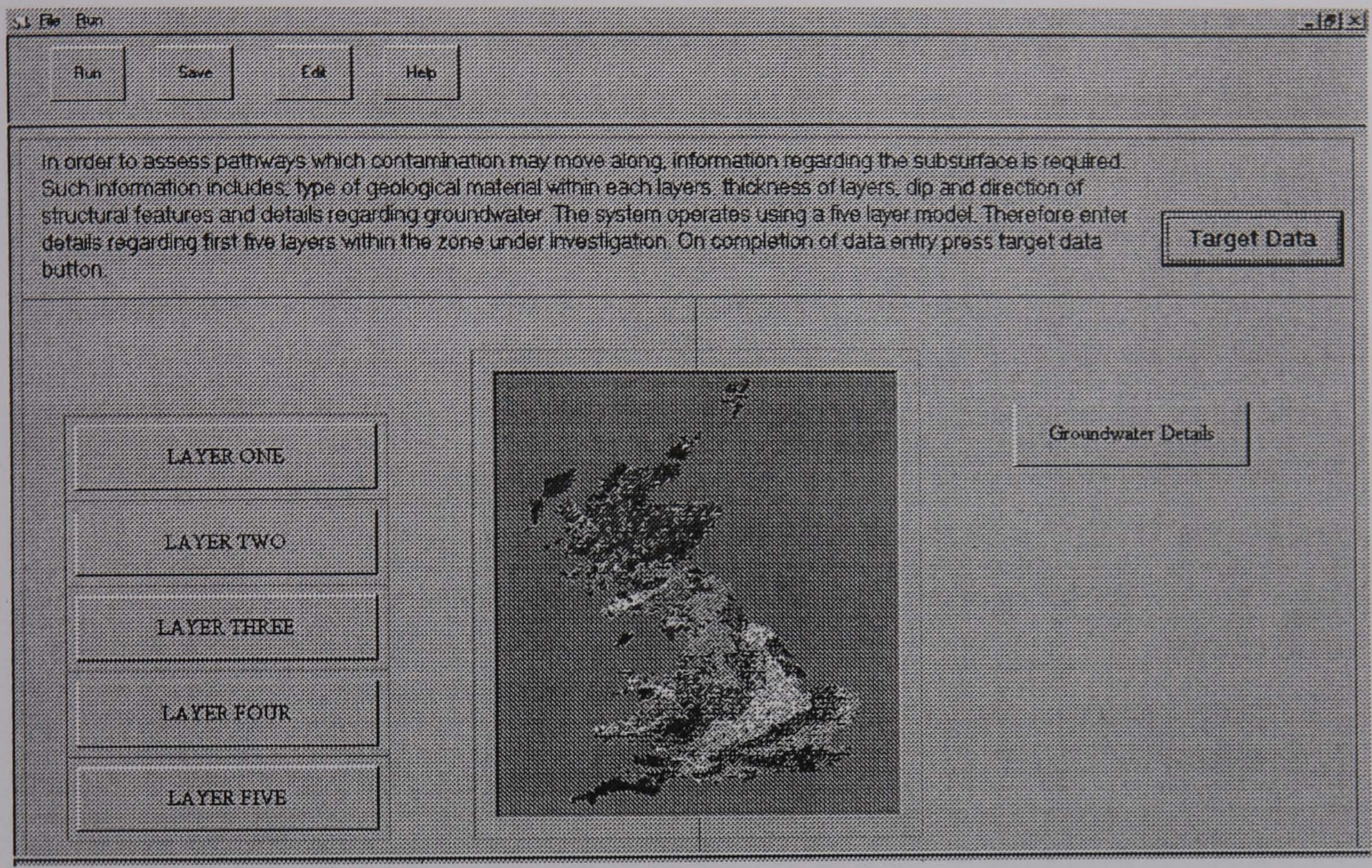


Figure 5.11: Pathway Introductory Display Example.

The pathway introductory screen shown in Figure 5.11, again presents the user with a list of command buttons relating to the data required. As within the source section, on activating one of the command button, a data entry form is displayed.

An option box system is also used within the pathway interface, shown in Figure 5.12. The option box system allows only one option to be selected, hence avoiding conflicting data being entered. However, the user is permitted to enter multiple layers under the location of groundwater as it is possible that groundwater may be in any of the layers. A check box system is utilised in this part of the form (shown in Figure 5.12, under the heading of "Location of Groundwater").

However both the check and option box data entry methods are not without their disadvantages. Both methods completely control the user's input and limit the number of values the user may select. To limit the effect of such features a number of special options were constructed within certain sections. For example, the pathway section contains two additional options; the first allows the user to enter a permeability term (high, medium, low, very low and impermeable) rather than selecting a material type, when the material type required is not displayed, and the second enables the user to select the 'permeability unknown' option (this passes a worst case scenario to pathway knowledge-base), when required material type and permeability is unavailable or unknown.

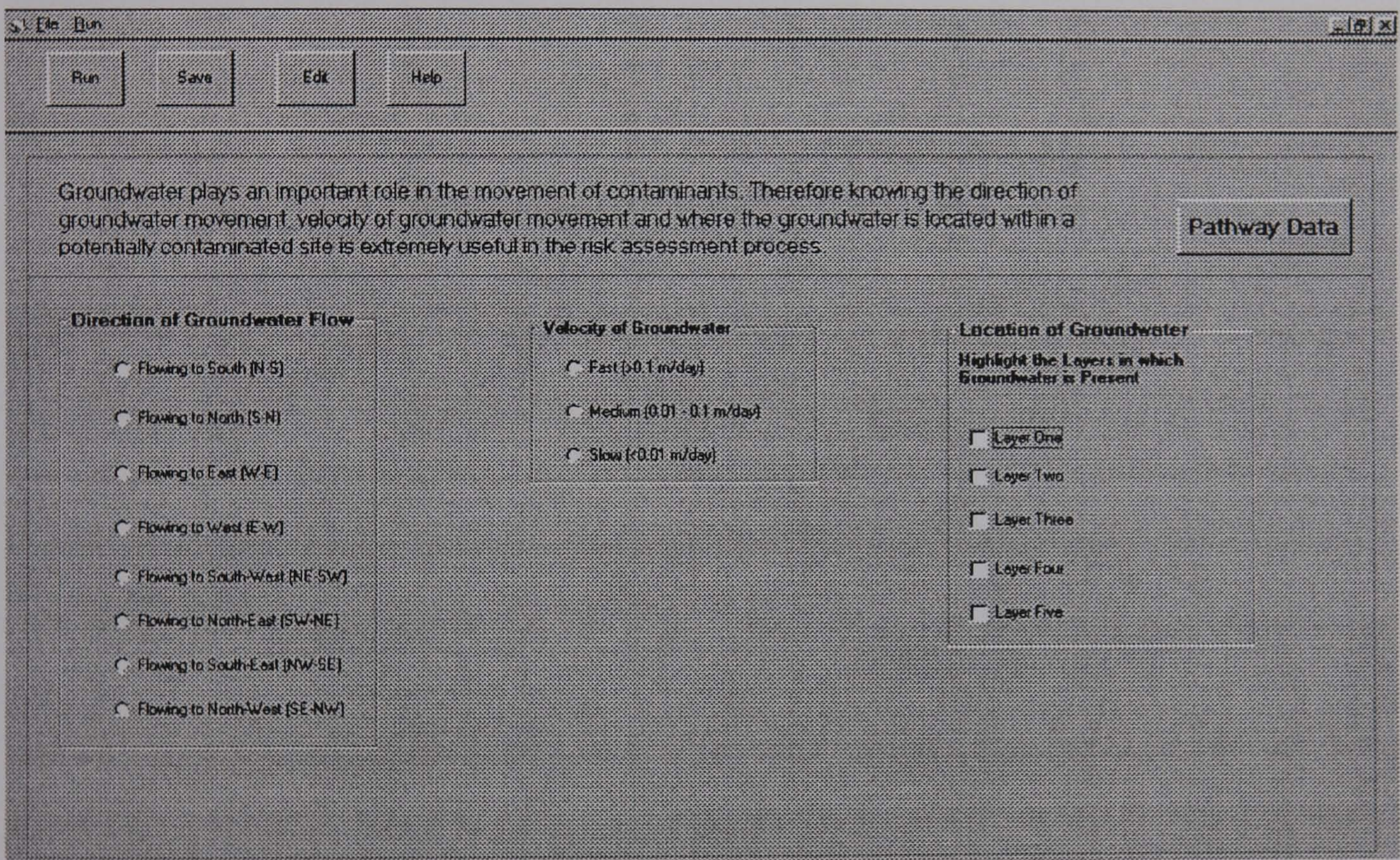


Figure 5.12: Option Box Interface Display Example (Groundwater).

On completion of data entry the user activates the command button which allows movement to the next section (target section), and a simple warning dialog box appears. The box prompts the user to ensure that the minimum amount of data has been provided. For example, between the pathway and target section, the dialog box reads "Please ensure that the system has details for at least one layer including details regarding groundwater".

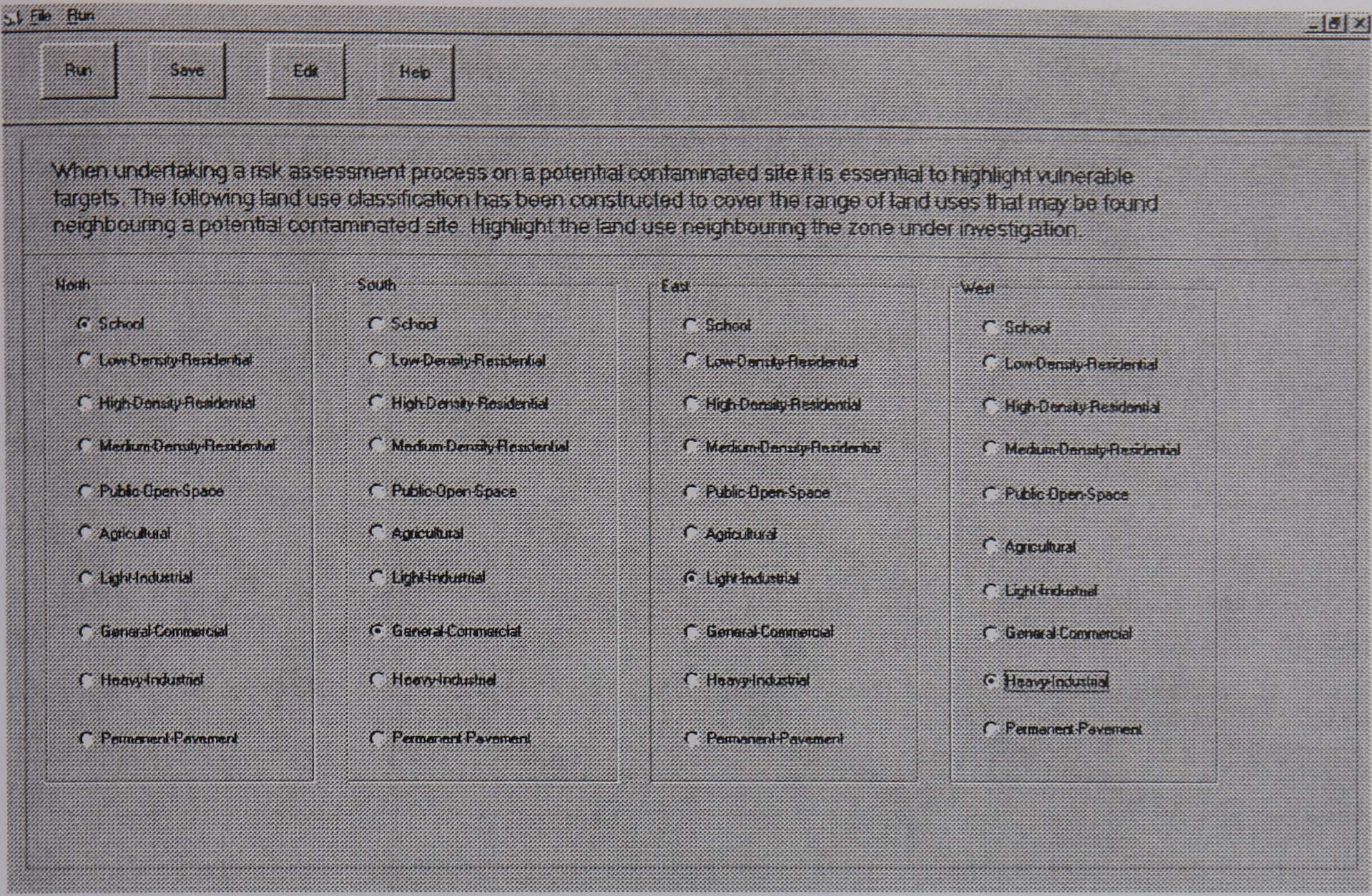


Figure 5.13: Target Section Interface Display Example.

The target section again contains option boxes, to enable the system to collect information relating to neighbouring land use. This allows the user to select from a pre-set classification, which has been compiled to cover a range of land uses, shown in Figure 5.13. Again, as previously mentioned this method of data entry prevents invalid data being entered, i.e. unclassified land use.

Once the user has been guided through the three sections, the data input by the user is saved into a suitable format for the knowledge-bases to use. Each entry made by the user is converted into a line of text, that includes an attribute name and the related value input by the user (shown in Figure 5.14). This process is automatically completed when the user activates the "save" command button. This command starts a routine that searches each interface form for check boxes or option boxes that have been selected as true, and then produces the relevant lines of text. The attribute names within the lines of text relate to the rules within the knowledge-bases. Therefore, on loading the contents of the text file into the knowledge-bases, facts are declared to the system which in turn triggers the relevant rules to fire.

```
File  Edit  Search  View
(attribute (name has-history) (value paper_production))
(attribute (name has-history) (value power_station))
(attribute (name has-history) (value railway_land))
(attribute (name has-history) (value tanneries))
(attribute (name grass-vegetation) (value brown_bent_grass))
(attribute (name has-stains) (value red))
(attribute (name has-deposits-type) (value ash))
(attribute (name has-deposits) (value green))
(attribute (name has-odours) (value musty))
(attribute (name structural-features-one) (value yes))
(attribute (name structural-features-two) (value no))
(attribute (name structural-features-three) (value no))
(attribute (name structural-features-four) (value no))
(attribute (name structural-features-five) (value no))
(attribute (name thickness-three) (value v.thin))
(attribute (name thickness-two) (value medium))
(attribute (name neighbouring-use-north) (value Light-Industrial))
(attribute (name neighbouring-use-south) (value Low-Density-Residential))
(attribute (name neighbouring-use-east) (value Low-Density-Residential))
(attribute (name neighbouring-use-west) (value Heavy-Industrial))
(attribute (name terrestrial-invertebrates) (value Lumbricus_rubellus))
(attribute (name layer-one-material) (value made_ground))
(attribute (name thickness-one) (value thick))
(attribute (name dip-one) (value horizontal))
(attribute (name direction-one) (value N-S))
(attribute (name groundwater-velocity) (value medium))
(attribute (name groundwater-direction) (value S-N))
(attribute (name groundwater-one) (value yes))
(attribute (name groundwater-two) (value yes))
(attribute (name groundwater-three) (value yes))
(attribute (name groundwater-four) (value yes))
(attribute (name groundwater-five) (value no))
(attribute (name layer-four-material) (value Eden_Shales))
(attribute (name layer-two-material) (value upper_greensand))
```

Figure 5.14: Text List Display Example.

5.6 Conclusion

ATTIC has been developed, using CLIPS software, as a tool to assist with the investigation of contaminated land and in particular the preliminary stage of an investigation, making full use of desk study and site reconnaissance data.

ATTIC consists of four knowledge-bases that contain approximately 1600 rules between them. The rules represent knowledge required for predicting possible contaminants, likely pathways for contaminant movement and sensitive targets, as well as knowledge concerning health and safety issues involved in the investigation of contaminated land.

The knowledge within the system has been drawn from two major sources; technical literature and domain experts. The knowledge gained from the domain expert was collected via a questionnaire and allowed certainty values to be applied to the knowledge derived from the technical literature.

A user interface containing questions has been constructed to allow the user to input data into the knowledge-based system.

CHAPTER 6

EVALUATION AND DISCUSSION

6.0 Introduction

A vital component within the development of the knowledge-based system is the evaluation of the system. It is often thought that this should be undertaken as the last stage within development. However, this is a process that must be considered throughout the entire design and development process. Leaving such a process to the end of the development may result in practical difficulties when trying to make changes to the system.

Green and Keyes (1990), suggest that there are two formal methods of testing any computer code, these being verification and validation. However, Ayel and Laurent (1991) suggests that validation and verification of knowledge-based systems is different from that for other types of computer systems. These differences include; a focus on symbolic knowledge rather than numeric data, an investigation of previously uninstructed problems, inclusion of both symbolic and numerical information in the same program (i.e. rules of the form "IF...THEN" and uncertainty factors on the weights), and the general lack of a means by which to determine the quality of a solution other than by human validation.

Verification can be defined simply as the process to ensure that the computer code is written without bugs, logical flaws and any other mistakes made by the knowledge engineer when translating expert knowledge into rules. This process was pursued on a regular basis as the system was developed and revised, and is discussed in Section 6.1.

Validation, in contrast, involves ensuring whether the meaning and context of the rules within the system are correct, supplying the user with valid results, which meet the design criteria and assist the users' performance. The key to a successful validation process requires the design criteria to be clearly defined. Kirk and Murray (1988) refer to this as the external correctness which is expressed in correct or desired output when the system is operating in a realistic environment. This process is addressed in section 6.2.

Section 6.3 follows with a discussion relating to the general development of the system and database, highlighting the differences between the system developed during this research and existing systems. Section 6.3 also discusses the knowledge acquisition process and reviews the evaluation process.

6.1 System Verification

Checking the syntax of any software is an important task, and is usually undertaken during the building of the system, as this avoids causing serious problems on execution of the system. The checking of the syntax within the knowledge-based system was undertaken using the tools provided by CLIPS. The syntax checker clearly shows the errors and their locations. Again with the Visual Basic© development environment there is a similar syntax checking facility.

Checking the syntax within both environments was therefore relatively straight forward. It consisted of loading the code into the respective compiler and checking for errors. This process was carried out with the addition of every new set of rules or lines of code, and was therefore performed countless times throughout development. Typical syntax errors included; unbalanced brackets, missing keywords or misspelt keywords and missing quotes. Besides, the simple errors, the check also highlighted more complex errors, such as the use of undeclared attributes.

Another component within the system verification process involved evaluating the consistency and completeness of the system. Evaluating such parameters, within a knowledge-based system can be difficult compared with other software systems. This is due to the fact that, if a routine in a standard software package fails, it will usually be obvious whereas if a rule within the knowledge-based system fails, it may be more difficult to locate the problem. Often a rule may not fire because it requires information that it has never received. Adelman (1992) highlights common anomalies associated with knowledge-based systems that need to be detected within the evaluation process.

(1) Redundant rules: individual rules or groups of rules that essentially have the same conditions and conclusions.

(2) Subsumed rules: when one rule's meaning is already expressed in another's that reaches the same conclusion from similar but less restrictive conditions.

(3) Conflicting rules: rules that use the same or very similar conditions, but result in different conclusions.

(4) Circular rules: rules that lead one back to an initial (or intermediate) condition(s) instead of a conclusion.

(5) Unnecessary If conditions: the value on a condition does not affect the conclusion of any rule.

(6) Unreferenced attribute values: values on a condition that are not defined; consequently, their occurrence cannot result in a conclusion.

(7) Illegal attribute values: values of a condition that are outside the acceptable set of values for that condition.

(8) Unreachable conclusion: rules that do not connect input conditions with output conclusions.

The "watch tool" within CLIPS permits the developer to observe facts being asserted and retracted, and also watch rules being executed. This provides an extremely useful facility, when checking for such anomalies highlighted by Adelman (1992). The checking process was undertaken in a logical and systematic way, starting with rules from the source knowledge-base and working through to the knowledge-base concerned with the health and safety issues. For each knowledge-base, facts relating to the rules were input and the output results monitored, to check whether the rules were firing correctly.

The Visual Basic© interface also required checking for incompleteness. This involved testing each form within the interface by producing a text output file and reading it within a text editor. This was then checked to make sure the correct output had been produced for the interface form tested. Common errors included, misspelt facts, missed fact lines or facts relating to the wrong rules.

On completion of the checking of both the CLIPS and Visual Basic© components of the system, an overall evaluation was required. This involved inputting a set of facts that contained data relating to the source, pathway and target components, to produce a set of results. To check whether the system was working consistently one of the components was altered and the other two kept the same. For example, the direction of the groundwater within the pathway component was altered, which in turn should result in the change of risk direction. Another example involved changing the target component, selecting a target where a low risk result would be expected (i.e. a permanently paved area) and then selecting the opposite in which a high risk result would be expected (i.e. a school). This process was undertaken for a number of combinations until the knowledge-base was seen to be running efficiently.

6.2 System Validation

The validation of the system was applied once the prototype knowledge-based system had been fully developed. The common method for validating a system usually involves running examples with known results, and comparing the performance of the program against the correct answers.

Therefore, to carry out this process successfully the collection of a number of case studies was required. This involved contacting a civil engineering consulting company and obtaining suitable data from a number of projects that the company had been involved in.

In total four case studies were collected for the validation process. The information for each case study included; data relating to preliminary investigation information, physical exploration results consisting of contamination test results and interpretative reports. Table 6.1 details the preliminary information available within the case studies.

Case Study One	
Site History:	1920 - 1945 - Timber storage, 1945 - 1963 Timber treatment plant 1963 - 1987 Car breaking yard 1987 Waste land
Geology:	Layer 1: Made Ground, thickness 1.5m Layer 2: Glacial clay, thickness 2m Layer 3: Coal Measures Shales, thickness 20m
Groundwater movement direction:	Unknown
Groundwater:	Layer 1: No Layer 2: No Layer 3: Yes
Neighbouring Use North: Neighbouring Use East: Neighbouring Use South: Neighbouring Use West:	Derelict (heavy industrial) Light industrial use Housing Housing
Case Study Two	
Site History:	1880 - 1964 Iron foundry 1964 - 1986 Site derelict
Geology:	Layer 1: Made Ground, thickness 3m Layer 2: Sands & Gravels, thickness 0.6 m Layer 3: Coal Measures Shales, thickness 5m Layer 4: Carboniferous Limestone, thickness : >30m
Groundwater movement direction:	East
Groundwater:	Layer 1: Yes Layer 2: Yes Layer 3: Yes Layer 4: Yes

Neighbouring Use North: Neighbouring Use East: Neighbouring Use South: Neighbouring Use West:	Light industrial use Light industrial use Housing Housing
Case Study Three	
Site History:	1880 - 1914 Gas works 1914 - 1950 Royal Navy fuel depot. 1950 - 1981 Private oil storage depot. 1981 - site demolished
Geology:	Layer 1: Concrete/fill, thickness 1.5m Layer 2: Sands, thickness 0.5m Layer 3: Dawlish Formation Sandstone, thickness 20m
Groundwater movement direction:	South
Groundwater:	Layer 1 : No Layer 2 : Yes Layer 3 : Yes
Neighbouring Use North: Neighbouring Use East: Neighbouring Use South: Neighbouring Use West:	Housing Light Industrial use Heavy Industrial use Heavy Industrial use
Case Study Four	
Site History:	1923 Farmland 1938 - 1947 Sand & Gravel pit 1947 - 1973 Sand & Gravel pit, concrete product manufacture and timber yard 1973 - 1992 Concrete product manufacture, Sand & Gravel pit infilled.
Geology:	Layer 1 : Madeground, thickness 0.5m Layer 2 : Alluvial deposits, thickness 3m Layer 3 : Lower Lias, thickness 182m
Groundwater movement direction:	East
Groundwater:	Layer 1 : No Layer 2 : Yes Layer 3 : Yes
Neighbouring Use North: Neighbouring Use East: Neighbouring Use South: Neighbouring Use West:	Housing Housing Industrial use Industrial use

Table 6.1: Case Study Desk Study Information

The available information for each case study was separately input into the system. The outputs from the prototype system were then correlated and compared with the physical exploration results and interpretative reports. Table 6.2 details the source knowledge-base outputs and the contaminant test results from the physical exploration. The level of contamination recorded from testing is not shown in Table 6.2, as it was viewed to be more important that the contaminants were present rather than the level of contamination.

It is clear from Table 6.2 that the prototype system results generally agree with the results produced from the physical exploration. However, there were a number of anomalies within the results. The anomalies can be split into two: either the contaminant appears in the predicted and not the proven results, or the reverse, the contaminant appears in the proven results and not the predicted. For example, within case study one contaminant mercury was identified by the prototype system but was not reported in the physical exploration results. A number of reasons were identified for this difference between the two. The first reason is that mercury may not have been tested for during the physical exploration and therefore not reported. The second reason is that the prototype is over-predicting the number and type of contaminants because it is simply relying on past history data. Case study 2 also identified the contaminants; phosphates, vanadium, manganese, aluminium and PAH, which were not identified by the physical exploration. This type of anomaly is also seen in case study 3, with the contaminants cadmium, chromium, mercury, oils and vanadium. Again case study 4 also predicted contaminants such as PAH and oils, which were not reported by the physical exploration. Although this form of anomaly is over predicting contaminants, it is not seen as being detrimental to the system, as it is generally better to over predict than to under predict. However, this could lead to increased investigation costs, if contaminant testing is recommended for contaminants that may not exist.

As previously mentioned the second type of anomaly seen in the results involved, contaminants being identified by the physical exploration but not predicted by the system. This type of anomaly is seen in case studies one, two and four. In case study one the system failed to identify ammonium, sulphate and sulphide; case study two cadmium, copper and mercury were not identified and the results of case study four were very poor. Cyanide, cadmium, nickel, phenol, selenium, sulphate and sulphide were all unidentified by the system. This is obviously a greater problem than the first type of anomaly identified, as the system is failing to identify contaminants. In defence of the system, the available desk study information in each case study was very limited. In the situation of case study four the only information available for the system to use included past history of the site (timber yard & car park). The geology also indicates that layer 1 contains made ground. Such limited information makes predicting contaminants extremely difficult, especially as the made ground may contain a large range of contaminants. To overcome this problem would require the system to provide a blanket recommendation to the types of contaminants expected. This is generally not seen as being very useful for the end user. Thus, if the system is only presenting a minimum list of results because of the lack of desk study information, it is important that the user is aware of this fact.

As previously mentioned the overall prediction of contaminants proved to be reasonable successful. However, Table 6.1 highlights that the preliminary investigation information available was extremely limited, consisting generally of one abiotic indicator (history of site) and information from geological maps and groundwater details. This made testing all the options within the system impossible, for example the use of biotic indicators. However the available data was deemed acceptable for the initial validation process. In order to overcome this problem would require the collection of desk studies that contained more extensive information. It became apparent from contacting a number of companies that generally due to financial and time constraints such extensive desk study information is rarely collected.

In order to fully test all the options within the system would require the system to be used by a company on a day to day basis and encourage the collection of information that is not normally collected. This would highlight areas of the system that require further work and areas that could be omitted from the system. A rigorous validation would also allow components (bedding thickness, groundwater direction, dip bedding etc.) that have risk values assigned to them to be fully evaluated and amended if required. Finding companies willing to invest the time required for the trial of the system is likely to be difficult, unless the company can see tangible benefits from using the system.

An alternative to using company data would be the use of published case histories. However, at the time of this research project, no suitable detailed studies were identified.

Contaminant	Proven Case Study One Results	KBS Case Study One Predicted Results	Proven Case Study Two Results	KBS Case Study Two Predicted Results	Proven Case Study Three Results	KBS Case Study Three Predicted Results	Proven Case Study Four Results	KBS Case Study Four Predicted Results
Arsenic	X	X	X	X	X	X	X	X
Ammonium	X							
Cadmium	X	X	X			X	X	
Chromium	X	X	X	X		X	X	X
Cyanide	X	X	X	X	X	X	X	
Copper	X	X	X		X	X	X	X
Lead	X	X	X	X	X	X	X	
Mercury		X	X			X	X	X
Nickel	X	X	X	X			X	
Phenol	X	X	X	X	X	X	X	
Selenium							X	
Sulphate	X		X	X	X	X	X	
Sulphide	X		X	X	X	X	X	
Toluene Ex. Matter	X	X	X	X	X	X		
Zinc	X	X	X	X	X	X	X	X
PAH				X	X	X		X
Oils		X	X	X		X		X
Iron		X		X				
Aluminium				X				
Manganese				X				
Vanadium				X		X		
Methane			X	X				

Table 6.2: Source Predicted Results Compared with Proven Results.

The testing of the source (likely contaminants) knowledge-base proved to be relatively straight forward, as previously demonstrated. This is due to the fact that generally the results of contaminant testing are presented in a standard format, with no interpretation from the investigator. This therefore makes a comparison of results from the system with test results extremely easy. However, the comparison of results from the other knowledge-bases proved not to be as straight forward. It became apparent from studying the interpretative/factual reports relating to the four case studies, that information regarding possible pathways and targets is not presented in a similar format to that of the results produced by the system. For example, the system presents a risk factor for each layer and for every direction away from the source. In contrast, within a consultant's interpretative report, the investigator generally will only make comment to the direction in which contaminants may migrate and the most likely type of ground conditions that are likely to make this possible.

It is unusual to see a break down of each layer, although the exercise of identifying the risk for each layer will probably be carried out by the expert, but not presented in the report. This exercise may have been carried out in a formal process or simply within the experts head, utilizing his own experience.

In order to compare the results from both sources (system/interpretative reports) Table 6.3 was compiled. Table 6.3 presents the results from the system and relevant information from the interpretative reports. The system generally seemed to perform well, identifying similar risks and direction for contaminant movement.

The results from the system and the interpretative report for case study one both highlighted a low to medium risk of contaminant movement, although the system presented a very low risk for layer 2. Due to the fact that no groundwater direction data was available from the preliminary investigation, the system produces the same risk for all directions. The system also failed to identify perched water tables and therefore possible localised contaminant movement. To overcome this problem it may be necessary to alert the user to the possibility of such features, when preliminary investigation data highlights Made Ground. This would obviously be the worst case scenario, however at the preliminary stage it is useful to have an understanding of all the possibilities.

Predicted Pathway	Proven Pathway
Case Study One	
Predicted Pathway	**Proven Pathway**
Layer 1 Vertical Risk: Medium Lateral Risk: Medium North, East, South, West	The Made Ground has variable permeability, a number of perched water tables were discovered within the Made Ground. The risk of contaminant movement is low - medium due to the absence of groundwater flow. However, localised movement may occur in the location of the perched water tables.
Layer 2 Vertical Risk: Very Low Lateral Risk: Very Low North, East, South, West	
Layer 3 Vertical Risk: Low Lateral Risk: Low North, East, South, West	
Case Study Two	
Predicted Pathway	**Proven Pathway**
Layer 1 Vertical Risk: Medium Lateral Risk: Medium East	The granular nature of the underlying strata and the presence of groundwater results in a high risk of contaminant movement. The Coal Measures strata below the granular layer is likely to inhibit contaminant movement to a degree. It is therefore likely that any contaminant movement will take place in the top two layers. The groundwater flow indicates the possible movement of contaminants to the east. Further investigation and consultation with the EA will confirm if this is the case. Remediation options will assess the risk of contaminant disturbance.
Layer 2 Vertical Risk: High Lateral Risk: Medium East	
Layer 3 Vertical Risk: Low Lateral Risk: Low East	
Layer 4 Vertical Risk: Medium Lateral Risk: Medium East	
Case Study Three	
Predicted Pathway	**Proven Pathway**
Layer 1 Vertical Risk: Medium Lateral Risk: Medium South	The underlying Sands were generally found to have high permeability. Contaminant leaching is likely to have occurred between the Made Ground and the sand layer. The presence of groundwater within the sands layer will increase the risk of contaminant movement in southerly direction. The underlying strata is classified to have a pathway risk value of High.
Layer 2 Vertical Risk: High Lateral Risk: Medium South	
Layer 3 Vertical Risk: Medium	

Lateral Risk: Medium South	
Case Study Four	
Predicted Pathway	**Proven Pathway**
Layer 1 Vertical Risk: Medium Lateral Risk: Medium East	The permeability of the underlying strata was found to be variable with the made Ground recording a range of permeability values.
Layer 2 Vertical Risk: Low Lateral Risk: Low East	The Alluvial deposits are composed of interbedded sands and gravels with the occasional clay layer, recording medium to high permeability values.
Layer 3 Vertical Risk: Very Low Lateral Risk: Low East	The presence of groundwater will also enhance the movement of contaminants. The general groundwater flow indicates the possibility of contaminant movement to the east. The overall risk of contaminant movement within the underlying strata is considered to be medium - high risk.

Table 6.3: Pathway Predicted Results Compared with Proven Results.

As regards case study two, the system presented some good results, highlighting the high risk in layer 2 and agreeing with the interpretative report regarding layer 3 (Coal Measures strata). The only concern from this case study was the fact that a medium risk factor was identified for layer 2 lateral risk and a high risk factor was identified for the vertical risk. This problem is related to the difference between lateral and vertical movement values in Table 5.16 and 5.17. This may need further investigation to assess if this is a major problem.

The results for case study three again were good, although the problem of a difference between the vertical and lateral risk values was also identified again. On reflection it is understandable to expect a difference between the values if structural features are present. Such features would increase the risk of contaminant movement. However, this is not the case in either of the case studies.

Problems were also identified within case study four. The system produced a pathway risk value of very low to medium, whereas the interpretative report suggests a medium to high risk value for the underlying strata. It is believed that this error is related to the permeability value the system used for the Alluvial deposits. The physical exploration of the site found interbedded Sands and Gravels, hence producing medium to high permeability values. The Alluvial in the system relates to strata with a high content of finer material and hence lower permeability.

Although the results are particularly encouraging it is important to note that the case studies are relatively straight forward and to fully test the system would require an investigator to use the system on a day to day basis. This would highlight gaps or problems within the system.

It is also apparent from the case studies, that the pathway risks presented within the interpretative reports, often rely on permeability values obtained from a combination of technical literature and limited test results. This relates to the fact that often the ground conditions may not be suitable for in-situ testing or laboratory testing of permeability due to ground conditions being extremely variable. This, therefore, often results in the investigator having more confidence in published values, unless a large number of samples have been collected and their permeability tested. This fact gives the system some credibility as the system is likely to be using the same published permeability values as the experts.

The prediction of target risks also proved to be successful, Table 6.4 details the outputs from the system and relevant information from interpretative reports.

<table>
<tr><th>Case Study One</th><th></th></tr>
<tr><th>Predicted Targets</th><th>Report Targets</th></tr>
<tr><td>Layer 1
Overall Risk: Low North and East; Medium South and West</td><td rowspan="2">The risk to the groundwater target is low, due the presents of the thick Clay and Coal Measures strata. Contaminant movement via groundwater is likely to be limited, therefore resulting in a low risk to neighbouring sites. All remediation options</td></tr>
<tr><td>Layer 2
Overall Risk: Low North and East; Medium South and West</td></tr>
<tr><td>Layer 3
Overall Risk: Very Low North and East; Low South and West</td><td>should consider the neighbouring housing estates to the south and west of the site and ensure airborne contaminants are eliminated. The risk to site workers is low providing suitable Personal Protective Equipment (PPE) is worn.</td></tr>
</table>

Case Study Two	
Predicted Targets	**Report Targets**
Layer 1 Overall Risk: Low East, West and South, Very Low North,	The risk to the groundwater target is high. The granular nature of the underlying strata will assist with the migration of contaminants in the direction of groundwater flow (easterly direction).
Layer 2 Overall Risk: Medium East, Low North, West and South	The risk of Methane migration off site is high. The neighbouring site to the East currently houses a series of light industrial units. Due to the nature (permanently paved) of the neighbouring site to the East the target risk value from groundwater transported contaminants is medium. However, methane migration results in a target risk of High for all neighbouring sites.
Layer 3 Overall Risk: Medium East, Low West and South, North	Further investigation and consultation with the EA will establish the current status of groundwater contamination and methane contamination on neighbouring sites.
Layer 4 Overall Risk: Low East, West and South, North	The risk to site workers is low providing suitable Personal Protective Equipment (PPE) is worn.
Case Study Three	
Predicted Targets	**Report Targets**
Layer 1 Overall Risk: Medium South, Low North, West and East	The risk to the groundwater target is high. Contaminant movement is likely to occur in a southerly direction, the target risk to the neighbouring site to the south is medium.
Layer 2 Overall Risk: Medium South, Low West, East, Very Low North	Remediation options will need to consider the disturbance of contaminants and the risk to housing estate to the north of the site. The housing estate is considered a high risk target from airborne contaminants under certain conditions (dry conditions and winds from the south).
Layer 3 Overall Risk: Medium South, Low West, East, Very Low North	The risk to site workers is low providing suitable Personal Protective Equipment (PPE) is worn.
Case Study Four	
Predicted Targets	**Report Targets**
Layer 1 Overall Risk: High East, Medium North, South, Low West	The risk to the groundwater target is high. Contaminant movement is likely to occur in an easterly direction, the target risk to the neighbouring site to the east is high, due to the direction of groundwater flow (easterly direction).

Layer 2 Overall Risk: Medium East, Low North, South and West	Further investigation and consultation with the Statutory Authorises will assess the current status of contaminant movement to neighbouring sites. Remediation options should prevent migration of contaminants offsite via groundwater and airborne dust.
Layer 3 Overall Risk: Low East, Very Low North, South and West.	The risk to site workers is low providing suitable Personal Protective Equipment (PPE) is provided.

Table 6.4: Target Predicted Results Compared with Report Targets.

As previously mentioned the results from the system were generally in agreement with the case study reports, matching similar risk values and directions. However a number of discrepancies were identified. The initial problem discovered, related to the system not presenting a risk value for site workers, although the health and safety knowledge-base presents Personal Protective Equipment (PPE) requirements.

The issue regarding methane migration in case study two was also not identified by the system. It is therefore clear that the system needs to be amended to account for methane migration. A similar issue was also identified with airborne contamination with all the case studies. This issue will also need to be addressed, by adding knowledge to the target knowledge-base or pathway knowledge-base.

The system also identified a risk for all the neighbouring areas unlike the case studies which only identified limited neighbouring areas. It is acknowledged that an area will not be mentioned within the reports if the neighbouring area is not connected to the investigation and there is no risk of contaminant movement.

In terms of health and safety issues, generally the case study reports highlight the measures that exceed trigger levels (level likely to cause harm). For example, the general statement used in all the reports states "General Personal Protective Equipment (PPE) should include disposable overalls and gloves. If conditions are dry suitable dust masks or suitable dust suppression methods should be employed. Good personal hygiene practice must be observed on site. This should include the removal of overalls and gloves prior to entering welfare facilities and the washing of hands before eating and drinking". In addition to this general statement, sites where workers may encounter oils and leachates the recommendations include the use of suitable face shielding equipment and the banning of ignition sources. As reference is made to sites where methane may be encountered, protection measures include; gas testing equipment and the banning of ignition sources. All the PPE recommendations within the reports also have the caveat that all PPE is dependent on the remediation option selected for the site.

In contrast, the system produces a list of all the measures for all the contaminants derived from the source knowledge-base as shown in Figure 6.1. This is due to the fact that the system cannot derive contaminant concentration levels. As a result the system can only provide health and safety measures to meet all contaminants. This is obviously an over estimation. However, at the desk study stage, this is extremely useful information.

```
Protective Measures
------------------------------------------
cadmium-protection:gloves
cadmium-protection:overalls
cadmium-protection:suitable-respiratory-protective-equipment
lead-protection:gloves
lead-protection:overalls
lead-protection:suitable-respiratory-protective-equipment
zinc-protection:gloves
zinc-protection:overalls
zinc-protection:suitable-respiratory-protective-equipment
copper-protection:gloves
copper-protection:overalls
copper-protection:suitable-respiratory-protective-equipment
phenols-protection:gloves
phenols-protection:overalls
methane-protection:infrared-gas-analyser
methane-protection:ventilation-control
methane-protection:suitable-respiratory-protective-equipment
methane-protection:ban-ignition-sources
```

Figure 6.1: Example of Contaminant PPE Measures Produced by System.

The measures that are detailed within the reports generally match the recommendations presented by the system. This is only to be expected, because the domain experts are likely to be using the same technical literature as the system to base their recommendations on. This therefore highlights that the correct knowledge had been used within the health and safety knowledge-base. However the system has overlooked one protective measure and that is the statement referring to good personal hygiene practice on site. This measure is not identified by the health and safety knowledge-base and therefore this problem needs to be rectified.

6.3 Discussion

The knowledge-based system and database system developed as part of this research have been designed to assist with the preliminary investigation of potentially contaminated land.

The development of such systems has been extremely limited within the domain of contaminated land. Law et al (1986), Heynisch et al (1994), Tucker et al (1997) and Kelly & Lunn (1998) have all used knowledge-based system technology as a tool within contaminated land investigation, described in detail within section 3.9.

Law et al (1986) and Heynisch et al (1994) both concentrate on evaluating hydrogeological properties to assess possible contaminant movement and potential problems relating to environmental problems, future land use etc. Both systems make full use of quantitative investigation data (data obtained from exploratory investigation). The system described within this thesis has acknowledged the importance of such parameters used within Law et al (1986) and Heynisch et al (1994) and incorporates them into the system. However, rather than using input data from the exploratory investigation, the system utilises data collected from the preliminary investigation. This is seen as an extremely important step forward as it allows the investigator to have an understanding of hydrogeological properties before exploratory investigation is undertaken. Therefore, planning of the exploratory investigation can be improved, which in turn should result in a higher quality and more applicable data being collected during the investigation. The input of higher quality data into systems such as Law et al (1986) and Heynisch et al (1994) should also result in better outputs from such tools.

In comparison, Tucker et al (1997) use preliminary investigation data to highlight possible hot-spots of contaminants. Tucker et al (1997) highlight the fact that outputs such as likely contaminants, groundwater assessment etc. would be useful but their system did not achieve this. In contrast, the system described in this thesis can successful assess likely contaminants, pathways, targets and also draws in other issues such as health and safety.

Again Kelly & Lunn (1998) make use of preliminary investigation data (generally historical data) to assists in predicting pollution migration using a source-pathway-target approach. As with the other systems the detail and extent of the knowledge incorporated within the system does not compare with the depth of knowledge drawn together in the system described within this thesis. As a result the inputs and outputs from the Kelly & Lunn (1998) system are limited. However, it could be argued that there are too many options available for entering data into the system compiled during this research and the additional options are unlikely to be utilized by an investigator.

However, generally, it can be stated that the system is a great improvement over existing systems in terms of type and extent of knowledge used within the system. In addition, the independent database constructed during the research is also seen as a major step forward over other systems and in the use and exchange of preliminary investigation data.

The prototype system constructed during this research project has addressed the problem of not making full use of preliminary investigation data within the risk assessment process. This has been accomplished by assessing the type of information available within the technical literature and converting it into suitable rules. The results of these rules present the user with an overall risk assessment based entirely on preliminary investigation data.

The knowledge acquisition process, which involved converting technical literature into 1600 rules, was found to be one of the most time consuming tasks within the development of the system, as it involved reviewing a range of subject areas from geological material types to the properties of contaminants. This demanded an understanding of a number of disciplines such as chemistry, geology, hydrogeology etc.

The process also required the rules from certain knowledge-bases to be inter-related with the results from other knowledge-bases. For example, the source knowledge-base that produces a list of likely contaminants is inter-related to the health and safety knowledge-base, in which, health hazards relating to the contaminants is produced.

Although identifying suitable information from the technical literature was relatively straight forward, it was more difficult to assess the value of such information, in terms of its usefulness in the risk assessment process. For example, certain indicators selected from the technical literature are used to identify contaminants. However, the literature does not give value judgements on how successful these indicators are, nor does it detail which indicators are most commonly used. To overcome this problem a questionnaire exercise, described within Chapter 5, was undertaken. It was decided that the questionnaire would be conducted through the e-mail system, as it was hoped that it would receive better responses than using a standard mailing questionnaire, as a low return rate is often the biggest problem when undertaking such a survey. The questionnaire was e-mailed to six mailing lists, with an estimated subscription of five hundred and sixty members. Forty six responses to the questionnaire were received. The response was considered to be good and some extremely useful information was identified as a result of the questionnaire. The results were analysed relating to experience and occupation. Although the results from the questionnaire were useful in setting values within the system, the responses of forty six is still relatively low. Therefore it may be questionable how such values are viewed by a larger population of experts working in the field of contaminated land. One way of amending this problem would be to send copies of the prototype of the system to consulting firms and ask them to evaluate the system over a number of months. The problem of this solution is that it is extremely difficult to get companies to agree to such a task, as often they have their own procedure to follow during an investigation. Therefore, they are unlikely to want to devote time to evaluating a system unless they are likely to benefit from it. Such an evaluation period also adds a considerable amount of time to the research project, but can be an extremely valuable addition to it.

In addition to the evaluation of the knowledge within the system via the questionnaire, a verification and validation of the system was also undertaken (described in sections 6.1 and 6.2). The verification process was carried out at regular intervals during the development of the system. It involved evaluating the consistency and completeness of the system, by detecting redundant rules, syntax errors etc. As errors and bugs were identified the appropriate measures were taken to correct such anomalies.

After the initial development of the prototype system a validation process was undertaken. The validation process ensures that the meaning and context of the rules within the system are accurate and the system meets the design criteria and assists the users performance. The process consisted of entering a series of four case studies into the system and comparing system outputs with proven results. The information available from the case studies was very limited, consisting generally of one abiotic indicator and information from geological maps and groundwater details. This meant that testing all the options within the system was impossible. However, for an initial validation it was considered acceptable. Obtaining case studies with more extensive details proved to be difficult. It became apparent that a high level of detail from preliminary investigations is uncommon.

The results from the validation process proved to be encouraging, although direct comparison of results with case studies was sometimes difficult. This was due to the risk assessment format used by the company from which the data were obtained. However, the problems identified by the validation process need to be addressed before the process is considered a success.

On completion of the development and testing of the prototype system, it was found that the system produced relevant information. However, a number of short falls within the system were discovered, with the main problem being the speed of data processing. This is due to the system being very code heavy and therefore this problem would need to be solved before the system could be used within industry.

Using such a system within industry is also likely to be viewed with caution by many domain experts at first. This initial caution may be due to a misunderstanding of the concepts of knowledge-based system technology or expert system technology as it is often referred to, as the expression "expert" often leads domain experts to believe the system is over-rated. In reality such systems, and particularly the system described, are intended to be used as tools to assist the engineer and not to produce definitive solutions.

During the development of the prototype system it became apparent that the storage of preliminary investigation data had yet to be addressed in full by the AGS Data Exchange Format or by any geotechnical database system available. In order to overcome this problem, data structures were developed for storing all aspects of preliminary investigation information, ranging from geological data to historical data. The ability to store data relating to vegetation and fauna was also included within the data structures, and this is seen as a major advance over other database systems. The data structures were implemented using Microsoft Access.

The development of such a standalone database system allows the user to store preliminary investigation data in a standard format similar to the AGS Data Exchange Format, and use the data either within the prototype system or with other suitable software.

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER WORK

7.0 Conclusions

The investigation of potentially contaminated land requires a multi-disciplinary and multi-stage approach, in order to collect the vast amount of information required to make a full risk assessment of the site.

This type of problem lends itself to knowledge-based system techniques, due to the fact that such technology can store knowledge from a number of domains and utilise the knowledge to solve problems input by the user.

The potential for the use of knowledge-based system technology has been demonstrated within the field of geotechnical engineering with the development of a number of systems addressing a range of geotechnical engineering problems. However, the development of such systems within the field of contaminated land investigation has been extremely limited. Therefore, scope for the development of such a tool to aid the investigation process of a potentially contaminated area was identified.

As part of this research project a prototype knowledge-based system containing a series of 1600 rules has been compiled. This was done with the aim of demonstrating that such technology may be applied to the preliminary stage of the investigation of contaminated land.

The prototype system assesses information collected during the preliminary stage of the investigation (past use, geological map, hydrological maps etc.) and assists with the risk assessment process, with the prediction of potential contaminants, hazards and risk to neighbouring areas.

The system has been developed, using CLIPS software, consisting of four knowledge-bases (source, pathway, target and health and safety knowledge-base). The results produced from each knowledge-base are stored within the working memory of the system until the final results list is presented to the user. This allows the results produced from each knowledge-base to be used as facts for rules within other knowledge-bases. For example, results from the source knowledge-base can be used as facts within the target knowledge-base.

A Visual Basic® interface has also been developed in conjunction with the knowledge-based system, in order to allow data entry to the system. The interface uses a series of forms relating to different components within the risk assessment process. Data entry to the form involves the user highlighting option boxes or check boxes, this avoiding the user having to type the required information, which in turn eliminates misspelling or mis-typing during data entry. On completion of the data entry, the resultant data is passed to the knowledge-base system in a text format.

The knowledge within the knowledge-bases was obtained from two main sources. The initial and main source being technical literature. Obtaining knowledge from technical literature involved reviewing published material, extracting relevant information and converting information into rules suitable for the knowledge-based system.

The second source of knowledge was domain experts via a knowledge elicitation exercise. The exercise took the form of a questionnaire relating to the rules and parameters within the system. This allowed views of domain experts to be sampled. The increased use of computer technology communications allowed the questionnaire to be delivered to domain experts via e-mail. This form of communication allowed the questionnaire to be accessed and returned by domain experts within a short period of time.

On completion of compiling the prototype, the system was validated against a number of real site investigation data sets. The system predicted the likely contaminants with a reasonable match to those observed, even though the input data for the case studies was limited. The assessment of risks to neighbouring target areas was generally in agreement with the case study reports, matching similar risk values and directions.

In addition to the construction of the prototype knowledge-based system and the user interface, the need to develop a relational database to allow preliminary investigation data to be stored, was identified within the scope of this research. The database system was modelled on the Association of Geotechnical Specialist electronic format for the transfer of ground investigation data. The data structures were implemented using the Microsoft Access relational database management system software. This allowed the database to be developed within a Microsoft Windows environment.

Finally, it may be concluded that the research project undertaken has demonstrated the contribution that knowledge-based system technology can make to the preliminary investigation of potentially contaminated land. The need for a standard format for the exchange of preliminary investigation data has also been highlighted and the construction of a relational database system for the storage of such data is seen as a major contribution in allowing the electronic transfer of preliminary investigation data.

Further development of the prototype system described within section 7.1, including the combination of the database system, would produce an extremely useful support tool for an end user within the field of contaminated land investigation.

7.1 Further Work

The knowledge-based system and database described within this thesis have been produced as prototypes, therefore leaving much scope for further work.

A link between the knowledge-based system and the database is seen as one of the first stages of further development. This would enhance the system, by allowing the user to enter data straight into the knowledge-based system via the database or via a Visual Basic© interface that updates the database while inputting data into the knowledge-based system. The output of results should also be enhanced, in order to present the results in a user friendly format similar to the Visual Basic© input interface.

Since the development of the prototype system, knowledge-based system development tools (shells) have advanced making development of such system within a windows environment easier and more powerful. Therefore, further development of the system should be employed using an advanced updated Windows based development shell. This would have a number of benefits, of which the first would be the presentation of the user interface. Although the current user interface within the prototype system is Microsoft Windows based and user friendly, it has been developed using Visual Basic© and is not fully integrated with the system. Therefore, values input by the user are saved in a text format and passed to the knowledge-based system. However, having an integrated interface would eliminate the need to use the routine highlighted above, hence increasing the efficiency of the system. The use of such software should also allow the database and knowledge-based system to be linked using ODBC link facility.

Another important aspect that could be incorporated using more advanced software may be a knowledge acquisition facility. Such a facility would enable the modification (addition or deletion) of the information within the existing knowledge bases. The modification of information within existing knowledge bases may include both rule and rule rating changes.

On completion of the amendments to the system a further period of validation is required. A more efficient and easy to use system is likely to encourage experts within the field of contaminated land to assist with an intensive validation exercise. This process would enhance the number of suitable components within the system. For example, certain biotic indicators that are rarely used may be eliminated from the system. Knowledge not discovered within the technical literature may be also added to the system in the form of further rules or an additional knowledge-base.

In addition to the work required to improve the efficiency of the system, there is vast scope to expand the system to aid other stages of an investigation. One option of expansion may be to add the facility to assess the level of contamination observed during the exploration stage and suggest possible remediation methods. Another useful addition to the system would be to link the system to a GIS package (e.g. Arcinfo). This would allow plans and 3D models to be constructed of the area under investigation. Combining this feature with the results from the exploration stage of the investigation would allow the system to calculate the volume of contaminated material within an area under investigation. This type of data is particular useful when selecting an appropriate remediation option for the area. Data relating to costs of remediation options could also be combined within the system. This would allow the system to select the most suitable and cost effective remediation option. This type of knowledge could be stored within separate knowledge bases.

REFERENCES

Adeli H. (1987), *Knowledge-Based Systems in Structural Engineering* , The Application of Artificial Intelligence Techniques to Civil and Structural Engineering, pp. 71-78.

Adeli H. (1988a), *An Overview of Expert Systems in Civil Engineering*, Expert Systems in Construction and Structural Engineering (Ed. Adeli H.), Chapman & Hall, London, pp. 45-83.

Adeli H. (1988b), *Artificial Intelligence and Expert Systems*, Expert Systems in Construction and Structural Engineering (Ed. Adeli H.), Chapman and Hall, London, pp. 1-12.

Adelman L. (1992), *Evaluating Decision Support and Expert Systems*, John Wiley & Sons, Chichester.

AGS (1994), *Electronic transfer of geotchnical data from ground investigations*, Publ. AGS/Issue 07/94, Association of Geotechnical Specialists, Camberley, Surrey.

Anon (1994), *Sustainable Development : The UK Strategy*, HMSO, London.

Attewell P. (1992), *Ground Pollution : Environment, Geology, Engineering and Law*, Chapman & Hall, London.

Ayel M. and Laurent J.P. (1991), Validation, Verification and Test of Knowledge-Based Systems, John Wiley & Sons Ltd, Chichester.

Bamford C. and Curran P. (1987), *Data Structures, Files and Databases*, Chapman & Hall Ltd, London.

Barry D. (1991), *Hazards in Land Recycling*, Recycling Derelict Land, (Ed. Fleming G.) Thomas Telford, London, pp. 28-63.

Beckett M.J. (1993), *Land Contamination*, Contaminated land Problems and Solutions, (Ed. Cairney T.), Blackie Academic & Professional, Glasgow, pp 8-28.

Bell F.G. (1993), *Engineering Geology*, Blackwell Scientific Publications, Oxford.

Berner H. (1975), *Geomap - Computer Based System for Geological Field Data*, Canada, Geol. Surv, paper 74-63, pp. 8-18.

Beynon-Davis P. (1996), *Database Systems*, Macmillan, London.

Bond A. (1995), *Hard Facts on Software*, Ground Engineering, October, Vol. 28, No. 8, pp14-16.

Bontempo C.J. & Saracco C.M. (1995), *Database Management Principles and Products*, Prentice Hall PTR, New Jersey, USA.

Bridges E.M. (1991), *Waste Materials in Urban Soils*, Soils in the Urban Environment, (Eds. Bullock P. & Gregory P.J.), Blackwell Scientific Publications, Oxford, pp.28-46.

British Standard 5930 (1981), *Code of Practice for Site Investigations*, British Standard Institution, London.

British Standard 5930 (1999), *Code of Practice for Site Investigations*, British Standard Institution, London.

British Standard DD175 (1988), *Code of Practice for the Identification of Potentially Contaminated Land and it's Investigation,* Draft for Development, British Standard Institution, London.

Buchanan B.G. and Shortliffe E.H. (1984*), Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, M.A.

Building Research Establishment, (1994*), Performance of Building Materials in Contaminated Land,* B.R.E. Report BR 255,

Buller J.V. (1964), *A Computer-Oriented System for the Storage and Retrieval of Well Information,* Bulletin of Canadian Petroleum Geology, Vol. 12, No. 4, pp. 847-891.

Cairney T. (1995), *The Re-use of Contaminated Land: A Handbook of Risk Assessment,* John Wiley & Sons, Chichester.

Cannon H.L. (1971), *The Use of Plant Indicators in Ground Water Surveys, Geologic Mapping and Mineral Prospecting,* Taxon 20 (2/3), pp.227-256.

Card S.K., Moran T.D. and Newell A.A. (1983), *The Psychology of Human Computer Interaction,* Lawrence Erlbaun Associates, Hillsdale, NJ, US.

Carter M. (1983), *Geotechnical Engineering Handbook,* Pentech Press, London.

Chaplow R. (1986) Production of Borehole Logs using a Microcomputer, Quarterly Journal of Engineering Geology, Vol. 19, No. 3, pp 291-300.

Chilton J., (1996), *Groundwater, Water Quality Assessments,* (Eds. Chapman D.), 2nd Edition, E & FN Spon, London, pp. 413-510.

Citrenbaum R., Geissman J.R. & Schultz R., (1990), *Selecting a Shell*, Expert Systems A Software Methodology for Modern Applications, IEEE Computer Society Press Reprint Collection (Ed. Raeth P.G.), IEEE Computer Society Press, California, pp. 52-61, 1990.

Cripps J.C. (1978) *Computer Storage of Geotechnical Data for use during Urban Development*, Bull. Int. Ass. Eng. Geol., 19, pp. 290-299

Crowcroft P. (1994), *Contaminated Ground Investigation: Matching Techniques to Site Conditions*, Journal of the Institution of Water and Environmental Management, No 8, October.

Crowcroft P. and Young P.J. (1992), *Investigating Contaminated Land*, IBC Conference on Contaminated Land, London, November 1992.

Date C.J. (1995), *An Introduction to Data Systems*, sixth edition, Addison-Wesley, Reading Massachusetts.

Day R.B., Tucker E.V. & Wood L.A. (1983), *The Computer as an Interactive Geotechnical Data Bank and Analysis Tool*, Proc. Geol. Assoc., Vol. 94, No. 2, pp. 123-132.

De Beer J.H. and Biggs D.C. (1978), *Urban Geotechnical Data Banking*, Proc. 3rd Int. Cong. Int. Ass. Engineering Geologists, Madrid, Special Session 4, pp. 101-137.

Denner J & Harris M.R. (1997), *UK Government Policy and Controls*, Contaminated Land and its Reclamation, (Eds. Hester R.E. & Harrison R.M.), The Royal Socity of Chemistry, Cambridge, pp. 25-46.

Department of the Environment (1990), *Contaminated Land*, The Government's Response to the First Report from the House of Commons Select Committee on the Environment, Cm1161, HMSO, London.

Department of the Environment (1991a), *Landfill Gas*, Waste Management Paper No. 27, HMSO, London.

Department of the Environment (1991b), *Survey of Derelict Land in England*, Main Report, Vol. 1, HMSO, London.

Department of the Environment (1992), *Environmental Protection Act 1990 (Section 143 Registers Regulations)*, Consultation Paper, DoE, London.

Department of the Environment (1994a), *A Framework for Assessing the impact of Contaminated Land on Groundwater and Surface Water*, Contaminated Land Research Report No. 1, Vol. 1, DoE, London.

Department of the Environment (1994b), *A Framework for Assessing the impact of Contaminated Land on Groundwater and Surface Water*, Contaminated Land Research Report No. 1, Vol. 2, DoE, London.

Department of the Environment (1994c), *Guidance on Preliminary Site Inspection of Contaminated Land*, Contaminated Land Research Report No. 2, Vol. 1, DoE, London.

Department of the Environment (1994d), *Guidance on Preliminary Site Inspection of Contaminated Land*, Contaminated Land Research Report No. 2, Vol. 2 DoE, London.

Department of the Environment (1994e), *Sampling Strategies for Contaminated Land*, Contaminated Land Research Report No. 4, DoE, London.

Department of the Environment (1994f), *Information Systems for Contaminated Land*, Contaminated Land Research Report No. 5, DoE, London.

Department of the Environment, (1995), *Industry Profiles 1 - 48*, DoE, London.

Durkin J. (1994), *Expert Systems Design and Development*, Prentice Hall International, London.

Ellis. B & Rees J.F. (1995), *Contaminated Land Remediation in the UK with Reference to Risk Assessment - 2 Case-Studies*, Journal of the Institution of Water and Environmental Management , Volume 9 No. 1, pp. 27-36

ENDS Report (1991), *Contaminated Land Policy Stutters*, ENDS Report, 193, February 1991.

Farias C.M., Moss D.W. & Jewell C.M. (1993), *A Rapid Screening Technique for Analysis of Volatile Organic Compounds (VOC's) in Groundwater and Soil*, Geotechnical Management of Waste and Contamination, (Eds. Fell R., Phillips T. & Gerrard C.), AA Balkema, Rotterdam, pp. 385-388.

Feigenbaum E. (1981), *Handbook of Artificial Intelligence,* Heuris Tech Press/William, Stanford.

Ferguson C. & Abbachi A. (1993), *Incorporating Expert Judgement into Statistical Sampling Designs for Contaminated Sites*, Land Contamination & Reclamation, Vol.1 No. 3, pp. 135-142.

Ferguson C. (1996), *Assessing Human Health Risks from Exposure to Contaminated Land: A Review of Recent Research,* Land Contamination & Reclamation, Vol. 4, No. 3, pp.159-170.

Ferguson C., (1992), *The Statistical Basis for Spatial Sampling of Contaminated Land*, Ground Engineering, 25 (5), pp.34-38.

Ferguson C., (1993), *Designing Sampling Strategies for Contaminated Sites, Site Investigation for Contaminated Sites*, IBC Conference, 23-24 September 1993, London.

Figg J.W., Souch K.W., Tublin A.P. (1980), *Some Experiences of Testing Soils and Materials from Contaminated Sites, Reclamation of Contaminated Land*, Proceedings of an SCI Conference.

Finn P.S. & Eldred P.J.L (1987) Data Management with Microcomputers in Geotechnical Engineering Practice, Quarterly Journal of Engineering Geology, Vol. 20, No. 2, pp 347-354.

Forde M.C., Batchelor A.J. & Ironside C.D., (1992), *Site Investigations and funding Strategies for Contaminated Land*, Proceedings of the Institution of Civil Engineers, Municipal Engineer, Vol. 93, March, pp. 19-26.

Forster A. and Culshaw M.G. (1990), *The Use of Site Investigation Data for the Preparation of Engineering Geological Maps and Reports by Planners and Civil Engineers*, Engineering Geology, Vol. 29, pp. 347-354.

Forth R.A. & Beaumont D., (1996), *Coal Carbonization in Northeast England : Investigation and Cleaning-up an Historical Legacy*, Engineering Geology of Waste Disposal (Ed. Bentley S.P.), Geological Society Engineering Geology Special Publication, Geological Society, London, No. 11, pp. 103-110.

Giarratano J. and Riley G. (1998), *Expert Systems : Principles and Programming*, PWS Publishing Company, London.

Green C.J.R., & Keyes M.M. (1990), *Verification and Validation of Expert Systems*, Expert Systems A Software Methodology for Modern Applications, IEEE Computer Society Press Reprint Collection (Ed. Raeth P.G.), IEEE Computer Society Press, California, pp. 444-449.

Greenshaw L.M., Bentley S.P. and Rice S.M.M. (1987), *The Management of Site Investigation Data using "Strata 3"*, Civil-Comp 87, Civil-Comp Press, Edinburgh, pp. 223-227.

Greenwood J.R. (1988), *Developments in Computerised Ground Investigation Data, Ground Engineering*, Vol. 21, No. 6, pp. 36-41.

Haines R.C. & Harris M.R. (1987), *Main Types of Contaminants*, Reclaiming Contaminated Land, (Ed. Cairney T.), Blackie, Glasgow & London. pp. 39-61.

Harris J.A. , Birch P. & Palmer J., (1996), *Land Restoration and Reclamation Principles and Practice*, Addison Wesley Longman, Harlow.

Harris M. and Herbert S. (1994), *Contaminated Land, Investigation, Assessment and Remediation,* I.C.E. Design and Practice guide, Thomas Telford, Suffolk, UK.

Harris M.R., Herbert S.M. & Smith M.A., (1995a), *The Remedial Treatment of Contaminated Land*, Special Publication SP 101-112, CIRIA, London.

Harris M.R., Herbert S.M., & Smith M.A., (1995b), *Remedial Treatment for Contaminated Land Volume III : Site Investigation and Assessment*. Special Publication 103, CIRIA, London.

Herbert S.M. (1995), *Design Guidance Provided for Contaminated Land, Proceedings of the Institution of Civil Engineers*, Civil Engineering, February 1995, pp. 43-44.

Heynisch S., Pekdeger A., Richter B., Schafmeister M & Skala W. (1994), *Knowledge-Based Systems for Risk Assessment of the Subsurface of Contaminated Sites*, Mathematical Geology, Vol. 26, No. 4, pp.505-517.

Hidding H. (1986), *Investigation and Remediation of the Site of the Former Gasworks at Tilbury*, Contaminated Soil. (Eds. Assink J.W., Van den Brink W.J., Matinus) Nishoff Publishers.

HMSO (1995), *Environment Act 1995*, HMSO, London.

Hobson D.M.(1993), *Rational Site Investigation*, Contaminated land Problems and Solutions, (Ed. Cairney T.), Blackie Academic & Professional, Glasgow, pp 29-67.

Howland A.F. & Polanski (1985) A Microcomputer Based System to Provide Report Quality Borehole Records, Quarterly Journal of Engineering Geology, Vol. 18, No. 4, pp 357-361.

Hussain D.S. and Hussain K.M. (1991), *Information Systems for Business*, Prentice Hall, London.

Ibrahim J.A.R. (1993), *The Application of Knowledge-Based Technology to Geotechnical Site Investigation*, PhD thesis, Heriot-Watt University, Edinburgh.

ICRCL (1983), *Notes on the Redevelopment of Scrap Yards and Similar Sites*, ICRCL Guidance Note 42/80, 2nd Edition, DoE, London.

ICRCL (1987), *Guidance on the Assessment and Redevelopment of Contaminated Land*, ICRCL Guidance Note 59/83, 2nd Edition, DOE, London.

Institute of Environmental Health Officers (I.E.H.O.) (1988), *Contaminated Land : Development of Contaminated Land* - Professional guidance.

James S.C., Kinman, R.N., Nutini, D.L. (1985), *Toxic and Flammable gases*, Contaminated Land Reclamation and Treatment. (Ed. Smith M.A.), Plenum Press, New York and London, pp. 207-255.

Jankowicz A.D. (1992), *Business Research Projects for Students*, Chapman and Hall, London.

Jennings R. & Person R. (1994), *Using Access 2 for Windows*, Que Corporation, Indianapolis.

Jewell C.M., Hensley P.J., Barry D.A., & Acworth I. (1993), *Site Investigation and Monitoring Techniques for Contaminated Sites and Potential Waste Disposal Sites*, Geotechnical Management of Waste and Contamination (Eds. Fell R., Phillips T. & Gerrard C.) AA Balkema, Rotterdam, pp. 3-37.

Johnson M.S. & Bradshaw A.D.B. (1977), *Prevention of Heavy Metal Pollution from Mine Wastes by Vegetative Stabilisation*, Trans. Inst. Min. and Metal A., No. 86, pp. 47-55.

Kelly C. & Lunn R.J. (1998), *Development of a Prototype Contaminated Land Assessment System*, Unpublished Report.

Kirk D.B. & Murray A.E. (1988), *Verification and Validation of Expert Systems for Nuclear Power Applications*, Science Applications International Corporation, New York.

Law K.H., Zimmie T.F. & Chapman D.R. (1986), *An Expert System for Inactive Hazardous Waste Site Characterization*, Expert Systems in Civil Engineering (Eds. Kostem C.N. & Maher M.L.), American Society of Civil Engineers, New York, pp. 159-173.

Leach B.A. & Goodger H.K. (1991), *Building on Derelict Land*, Special Publication 78, CIRIA, London.

Liou Y.I. (1998), *Expert System Technology : Knowledge Acquisition*, The Handbook of Applied Expert Systems (Ed. Liebowitz J.), CRC Press, London, pp. 2-1- 2-10.

Lowe G.W. (1984), *Investigation of Land at Thamesmead and Assessment of Remedial Measures to bring Contaminated Sites into Beneficial Use*, Proc. Conf. Management of Uncontrolled Hazardous Waste Sites, Washington D.C., HMCRI, Silver Spring, Maryland, pp.560-564.

M Z Associates (1994), SID : *A Geotechnical Data Management System Information Pack*, M Z Associates, Llwynhaf, Camarthen, Wales, UK.

Malenke M. (1991), *Relational Database Design and Technology*, Geotechnical Engineering Congress 1991 (Eds. Mclean F.G., Campbell D.A. & Harris D.W.), Geotechnical Special Publication No. 2, Vol. 1, June 10-12, Boulder, Colorado.

Mayne A. and Wood M.B. (1983), *Introducing Relational Database*, The National Computing Centre Limited, Manchester,England.

McEntee (1991), *Site Investigation*, Recycling Derelict Land, (Ed. Fleming G.) Thomas Telford, London, pp. 64-87.

Meyer P.B., Williams R.H. & Yount K.R. (1995), *Contaminated Land: Reclamation, Redevelopment and Reuse in the United States and European Union*, Edward Elgar Publishing Limited, Aldershot, UK.

Miller D.C. (1991), *Handbook of Research Design and Social Measurement*, Fifth Edition, Sage Publications, London.

Moula M., Toll D.G. and Vaptismas N. (1995), Knowledge-Based Systems in Geotechnical Engineering, Geotechnique, Vol. 45, No. 2, pp 209-221.

Myers K., Vogt T. & Wales J. (1994), *Hazard Ranking Criteria for Contaminated Sites*, Land Contamination & Reclamation, Vol. 2, No. 1, pp. 13-18.

Nancarrow D. (1998), *Section 57 - What Will it Mean for Local Authorities ?*, Land Contamination & Reclamation, Vol. 6, No. 3, pp. 119-122.

NRA (1994), Implementation of the EC Freshwater Fish Directive: Water Quality Requirements for Support of Fish Life, Water Quality Series No. 20, HMSO, London.

Oliver A. (1994), *A Knowledge Based System for the Interpretation of Site Investigation Information*, Ph D Thesis, University of Durham.

Petts J. (1993), *Risk Assessment for contaminated sites*, Proceedings of a Conference on site investigation for contaminated sites. IBC Technical Services (London) paper No. 1.

Petts J., Cairney T. & Smith M. (1997), *Risk-Based Contaminated Land Investigation and Assessment*, John Wiley & Sons, Chichester.

Rapier J. & Wainwright D. (1987), *Use of the Geotechnical Database GEOSHARE for Site Investigation Data Management*, Quarterly Journal of Engineering Geology, Vol. 20, No. 3, pp. 221-230.

Rhind D.W. and Sissons J.B. (1971), *Data Banking of Drift Borehole Records for the Edinburgh Area Report 71/75*, Institue of Geological Science, London.

Rich E. and Knight K. (1991), *Artificial Intelligence*, International Edition, McGraw-Hill, New York.

Richards B.P., Perry T.E. & Boyle J.S. (1996), *Contaminated Land : Legal Implications, Assessment and Impact on the Environment*, GEC Review, Vol. 11 No. 1, pp. 21-36.

Rodger L. (1992), *Borehole Exchange Support*, Geoscientist, Vol. 1, No. 6, pp. 30-31.

Royal Society (1992), *Risk : Analysis, Perception and Management*, Royal Society, London.

Sekaran U. (1992), *Research Methods for Business - A Skill Building Approach*, Secound Edition, John Wiley & Sons, Chichester.

Smith M. (1985), *Contaminated Land - Reclamation and Treatment*, (Ed. Smith M.), Plenum Press, New York.

Smith M. (1990), *Memorandum of Evidence*, Report of the Environment Committee, Vol. 11, HMSO, London.

Smith M. (1991), *Data Analysis and Interpretation*, Recycling Derelict Land, (Ed. Fleming G.), Thomas Telford, London, pp. 88-144.

Smith M. (1993), *Landfill Gases*, Contaminated Land Problems and Solutions, (Ed. Cairney T.), Blackie Academic & Professional, Glasgow, pp. 160-190.

Smith M. (1996), *An Overview of Current Research into Site Ivestigation Technologies and Methodologies*, Land Contamination & Reclamation, Vol. 4, No. 3, pp.149-157.

Staten P.M. & Caronna S. (1992), *Geotechnical Data Management of a Major Highway Scheme in the UK*, Proc. International Conference on Geotechnics and Computers, Paris, Presses de I'Ecole Nationale de Ponts et Chaussees: Paris, pp. 741-748.

Steed J.E., Shepherd E. & Barry D.L. (1996), *A Guide for Safe Working on Contaminated Sites*, Report 132, CIRIA, London.

Sury M. & Slinsby A. (1992), *Reclamation of the Former Laporte Chemical Works*, Ilford, Essex, Proceedings of the 2nd International Conference on Construction on Polluted and Marginal Land, Brunel University, London, pp. 379-384.

Syms P.M. (1994), *The Funding of Developments on Derelict and Contaminated Sites*, Industrial Property Policy and Economic Development, (Eds. Ball R. & Pratt A.C.) Routledge, London, pp. 63-82.

Syms P.M. (1997), *Contaminated Land, The Practice and Economics of Redevelopment*, Blackwell Science Ltd, London.

Threadgold L. (1992), *Borehole Database*, Geoscientist, Vol. 2, No. 2, pp 30-31.

Toll D.G. (1990), *Do Geotechnical Engineers Need Expert Systems ?*, Ground Engineering, Vol. 23, No. 23, pp. 32-36.

Toll D.G. (1996), Artificial Intelligence Application in Geotechnical Engineering, Electronic Journal of Geotechnical Engineering, http://geotech.civen.okstate.edu/ejge/ppr9608/.

Toll D.G. and Oliver A.J. (1995), *Structuring Soil and Rock Descriptions for Storage in Geotechnical Databases*, in Geological Data Management (ed. J.R.A. Giles), Geol. Soc. Special Publ. No. 97, Bath: Geological Society Publishing House, pp 65-71.

Tucker P., Ferguson C., Abbachi A. & Nathanial P. (1997), *A Decision Support System for Designing Sampling Strategies for Potentially Contaminated Sites*, Advances in GIS Research, Proceedings of the Seventh International Symposium on Spatial Data Handling (Ed. Kraak M.J. & Molenar), Taylor & Francis Ltd, London 1997.

Vamos T. (1998), *Knowledge Representation*, The Handbook of Applied Expert Systems (Ed. Liebowitz J.), CRC Press, London, pp. 3-1- 3-22.

Vedder R.G. (1989), *PC Based Expert System Shell : Some Desirable and Less Desirable Characteristics*, Expert Systems, Vol. 1, No. 1, pp. 28-42.

Water (1986), *A Guide to Expert Systems*, Addison-Wesley, Reading, M.A.

Young P.J. , Pollard S. & Crowcroft P. (1997), *Overview : Context, Calculating Risk and Using Consultants*, Contaminated Land and its Reclamation, (Eds. Hester R.E. & Harrison R.M.), The Royal Society of Chemistry, Cambridge, pp. 1-24.

APPENDICES

LIST OF APPENDICES

APPENDIX 1

COMMERCIAL DATABASE MANAGEMENT SYSTEMS

Brief Description of Commercial Database Management Systems

Name : Environmental WorkBench

Publisher : SSESCO

Platform : UNIX, OS/2

Status : Commercial

Description : A suite of integrated applications for processing Environmental Data to analyse and visualise large datasets. Interfaces to many models and monitoring sources. Includes the programs: savi3D - A 3D visualisation program: MeRAF - the internal data format. Binary, Random Access, self describing netCDF based database. Development Library (documented netCDF based C, C++, and FORTRAN library available for DOS, OS/2, and UNIX platforms): ShowME - A text based data file display tool: ToolME - A tool for handling large datasets: DLG_Extract - A command line utility to extract maps containing specified attributes from a USGS Graphics file: Ground Water Process - A program which facilitates input of observational data for soil and groundwater study sites, interpolates these observations onto a grid, runs as a simple, scrolling text based application.

Supplier in USA : SSESCO

Name : EQuIS

Publisher : EarthSoft

Platform : Win3x, Win95, WinNT

Status : Commercial

Description : An environmental data management system written in Visual Basic with a Microsoft Access database engine. Contains a comprehensive, fully relational environmental database that includes the data fields and entity relationships necessary to store and manage all the technical data that is generated during site characterisation, field remediation and data monitoring projects. Graphical Applications - Two kinds of interfaces are available that let EQuIS share data with several low-cost popular PC software products, such as Surfer, Grapher, Crystal Report Writer and StatMost. The Casual interface generates 2D and 3D graphs and contours without knowledge of the underlying software. The Power user interface offers full control of the target graphics system. ArcView 2 and the DoD Groundwater Modelling System will be added in release 2.0. Interfaces to GTGS Boreholes, AutoCAD and others are under development. Reporting - EQuIS Crystal Reports Pro. Canned reports are available, or the user can create customised reports. Data Integration - Historical data from IRPIMS, IRDIMS, GIS/Key, CLP, ITEMS and NEDTS can be loaded into the system. The Lab Data Verification Tool data loader electronically loads and verifies deliverables from LIMS systems. EQuIS can also produce data in different formats for data sharing with other systems. Customisation - The EQuIS system is Open and Customizable. As new capabilities are required on a project, new functionalities can be integrated into the system. With a Source Code contract, Earthsoft will provide source code for user development requirements.

Cost : US Dollars 4000 plus 15% annual support

Supplier in USA : EarthSoft

Supplier in USA : Environmental Systems & Technologies Inc

Name : KeyHOLE

Publisher : Key Systems

Platform : AutoCAD

Status : Commercial

Description : KeyHOLE is an add-on for AutoCAD that provides a relational database for data storage, manipulation, modelling and presentation of geotechnical data. Reads AGS format data. Produces borehole logs, long and cross sections, contamination profiles. Contouring of geological or topographic surfaces. Links to HoleBASE+.

Cost : GB Pounds 1500

Name : LYNX

Publisher : Lynx Geosystems Inc

Platform : UNIX

Status : Commercial

Description : LYNX is a complete software system of 3D application tools for characterisation, analysis and geo-engineering of the subsurface. From 3D integration of all geo-data sources to prediction, risk assessment and visualisation of complex conditions with application across the geosciences. Comprehensive suite of application modules. The base system functionality for full 3D geological characterisation can be extended by surface and underground engineering options and enhanced by the 3D visualisation option. LYNX is available with single, multi-seat or network licensing options for a range of graphic workstations. Integrated 3D functionality; Total information management; Interactive geological interpretation; Geostatistical prediction; Spatial analysis and query tools; Surface & underground geo-engineering; Risk assessment and planning; Presentation quality visuals. Links to CAD, GIS and spreadsheet systems. Borehole logs, maps, samples, plans, surveys, sections, surfaces, volumes and gridded variations are accessible with a range of analytical reporting and visualisation options. Applications: Mineral and energy resource evaluation; Surface and underground mining; Environmental contamination assessment; Remediation planning and design; Geotechnical and tunnelling applications.

Name : PC-XPLOR

Publisher : Gemcom Services Inc

Platform : DOS

Status : Commercial

Description : PC-XPLOR system stores, manipulates, analyses and displays all types of exploration data. Database Features: databases build on one another hierarchically using templates; each database can consist of multiple related tables; one project can support a variety of databases simultaneously; data entry and editing are interactive; import or merge data from tables stored in ASCII files; redefine and restructure databases; customise the screen, files and printed reports; functions for flexible data manipulation, filtering, extraction and sorting; modular design. Graphics and Plotting Features: integrates QuickPlot module to provide WYSIWYG graphics and plotting; overlay data prepared from different databases; prepare detailed drillhole plans and sections annotated with values and histograms; extract subsets of data from any database and display colour coded symbols and text at sample sites; overlay contour plots produced from gridded point data; export graphics to AutoCAD; export drillhole sections or composites to GEO-MODEL for interactive interpretation and polygonal reserves. Statistics and Geostatistics Features: histograms and line graphs to show frequency distributions; scattergrams and interactive regression analysis; down-hole or three-dimensional semi-variograms; interactively fit variogram models. Gridding and Contouring Features: two and three dimensional inverse distance interpolation and kriging; three dimensional surface fitting; contour preparation with full smoothing and labelling. Compositing Features: composites by length, level, cut-off grades and by lithologic intervals; display and analyse composited data in the same way as assay data; export composited data to PC-MINE to interpolate three dimensional block models.

Name : QUEST

Publisher : Environmental Systems & Technologies Inc

Platform : DOS

Status : Commercial

Description : A graphical relational database for integrated data analysis and modelling of environmental data. Uses DXF format base maps to which can be added overlays of sample locations, posted values, contour and gradient plots. The database fields are user definable but can cover soil boring information, chemical data and groundwater readings. Plots can be produced of time series data or x-y graphs of any pairs of database fields. Data can be contoured using kriging. Links to ARMOS, BIOTRANS and SPILLCAD programs to share data, results and maps. QUEST is not available separately, but is included as a part of ARMOS, BIOTRANS, and SPILLCAD.

Name : RECALL

Publisher : Z&S Consultants Ltd

Platform : UNIX

Status : Commercial

Description : An engineering database designed to store all types of borehole data for the oil and gas E&P industry. This includes conventional wireline logs, MWD, LWD, core measurements, core laboratory results, borehole images, dipmeter, waveforms, core photographs, thin section, SEM and borehole seismic data. In addition to data storage, it provides an integrated environment for well data applications. The modular system includes: INCLINE II : dipmeter processing and interpretation system: IMAGE : borehole image processing and interpretation system: PETROS II : petrophysics processing and interpretation system: TRANSCRIPT: modelling language: SPATIAL : plotting system for borehole, well location and well trajectory data: RtBAN : induction, laterolog and electric log modelling system: SONIC : full wave-form acoustic processing system. Input data formats such as LIS, DLIS, BIT, Atlas CLS Field Tape, Geoshare, LAS, SEG-Y, SPWLA, TIFF and ASCII files. Data manipulation tools to edit, calibrate, merge and splice and depth match borehole data. Interfaced to other database products and applications which include OpenWorks, Finder, Tigress, Iris21, Terrastation, Stratlog and IRAP.

Cost : GB Pounds 10000+

Name : SiteGIS

Publisher : GeoTrans

Platform : Win3x, MapInfo

Status : Commercial

Description : An add-on for MapInfo to store, analyse and present environmental data used in subsurface remediation investigations. Maps may be imported and exported as DXF files and in Arc/Info format. Links to Excel for the plotting of time series and other graphs. Links to Surfer for DOS to grid and contour data. Data points and contoured data can be overlaid on base maps and aerial photograph images.

Cost : US Dollars 1500

Name : Spase

Platform : DOS

Status : Commercial

Description : Relational database management system for scientific and engineering data where there is some element of spatial map based data. The additional module EnviroSpase provides analysis of environmental data for site monitoring or remediation. Handles base maps, data locations, samples and chemical data. Data can be mapped. listed and plotted. The program can be adapted to handle different data types and its functionality can be extended by a scripting language.

Cost : US Dollars 2000, US Dollars 3000 with EnviroSpase

Database systems (with log production)

Name : BLDM

Platform : DOS, Intergraph Microstation

Status : Commercial

Description : A PC-based boring log database management system and site characterisation tool. It allows users to create and maintain project boring log data, print summary and detail reports, create data files compatible with Intergraph's INSITU system, and generate boring log design file plates for plotting or display on any Intergraph CAD platform.

Cost : US Dollars 194

Name: GEOBASE

Publisher : Earthware Inc

Platform : DOS

Status : Commercial

Description : A relational database for geological, hydrogeological and environmental data. The geology workstation provides customised borehole log production with lithological symbols, depth plots of parameters and details of borehole installation construction. Up to 8 stratigraphic tags may be added to each borehole to allow automatic section drawing, surface and isopachyte contouring. Cross sections and fence diagrams can be plotted with borehole logs, well construction details, numeric data plots and labels. 2D and 3D contouring. The environmental module includes map and data entry, chemical parameter editor, spreadsheet data import, time plotting, Stiff, Piper, bar, pie, vector, tickel, disk, Durov, Schoeller groundwater plots and salinity hazard diagrams. Time logs for meteorological data. The hydrogeology module has analytical models for pump testing, flow line modelling and slug testing. Data can be imported and exported as ASCII files. The program produces DXF files for import into CAD packages.

Cost : Geology GB Pounds 1747, Environment GB Pounds 1747, Hydrogeology GB Pounds 940

Supplier in USA : Earthware Inc

Supplier in United Kingdom : Natural Systems Software

Name: Geodasy

Publisher : A F Howland Associates

Platform : Win95/98, WinNT

Status : Commercial

Description : A modular program for geotechnical data management. Uses plain English entry screens to allow data entry and amendment. Produces report quality records for percussion boring, rotary drilling, trial pits, dynamic penetrometer, instrumentation, soil and chemical laboratory sheets. Carries out data processing with graphical output including nominal sections, x-y plots, nmc/LL/PL plots, composite grading curves, A-line plots, groundwater and gas readings v time. Project management and administration features include costed fieldwork summaries, sample logging sheet, laboratory test schedule, sample store record sheet, sample transmittal sheets and AGS data export.

Supplier in United Kingdom : A F Howland Associates

Name: GeODin

Publisher : CivilServe

Platform : Win3x, Win95/98, WinNT

Status : Commercial

Description : The GeODin system is a structured data model made up of four modules: The GeODin Base module comprises a combination of editors that provide tools for data collection: Features includes: General borehole information editor; Geological editor; Well design editor; Geotechnical and chemical editor; Unlimited number of user-definable boreholes, sampling points, depths, intervals; Up to 9 piezometers can be shown with any specified well diameter, casing, filter material; Unlimited number of depth profiling parameters (e.g. chemical analysis, borehole tests); German geological dictionary; Import of borehole coordinates (e.g. after digitising) and ASCII-files. The GeODin Graph module combines CAD functionality with graphic elements to enable presentation of environmental and geological data. Features include: Geotechnical information presented either as full-length text or in abbreviations, in tabular format or user defined shapes (circular, rectangular); Main soil/rock types and secondary constituents can be shown separately or as a percentage; Individual profiles can be shown in different formats an unlimited number of times within a single document; Printing support for tiling (multiple sheets) at selected scale or automatic scaling (fit to page); Display in black and white or colour; Printout in accordance with national standards (e.g. DIN); Full-featured CAD program with design tools for drafting and drawing (e.g. worksheet with company logo etc.). The GeODin Analyse module is a combination of an environmental and a geological database contained within the GeODin project structure and controlling the administration and evaluation of the geo-environmental data. Features include: Report preparation and presentation of data as business graphics, including time-series and statistical analysis; Comparison of environmental data with recommended guideline values and user defined reference lists; Tools for time-series analysis, organisation into data types and parameter groups; Import and export filters to external programs (e.g. Arc/Info, Surfer, SPSS) and links to other GeODin modules. The GeODin View module provides compatibility with ArcView so that area and point information can be selectively combined from specified data / data types / projects and viewed at source in the GeODin System. Features include: Presentation and evaluation of all available information at a particular measuring point, including changes in parameters with time or grouping of chemical contaminants as pie charts; Construction of schematic borehole profiles as ArcView themes.

Name: gINT

Publisher : Geotechnical Computer Applications

Platform : Win3x, Win95/98, WinNT

Status : Commercial

Description : A stand-alone database manager and report generator for geotechnical and geo-environmental investigations. Free-form reporting styles include logs, graphs, 2D and 3D fence diagrams, histograms, tables, and data summaries. An unlimited number of report templates for all the report styles can be user-defined. At output time the template type and name, and the data range are specified. Output can be to a printer, AutoCAD DXF file, Windows Bitmap file, or gINT Drawing file. Text tables can be output to a variety of ASCII file formats and to a spreadsheet file. Output data can be filtered to obtain a subset of the full database. Text table reports and histograms also support the printing of data statistics. Supports user-definable libraries for material, sample, well, graph data marker, graph specification curve, logo, and fill patterns. The program supports both export and import of ASCII format and the UK AGS data interchange format files. Each project is stored in one Microsoft ACCESS compatible file. Technical support by telephone, fax, and e-mail at no charge. Program comes with a 30 day money-back guarantee.

Cost : US Dollars 1295 to 3995 depending on options and corporate discounts

Supplier in USA : Geotechnical Computer Applications

Supplier in United Kingdom : SWK Ltd

Name: GIS-Key

Publisher : GIS-Solutions, Inc.

Platform : DOS, Win3x, Win95/98, WinNT

Status : Commercial

Description : An integrated system for managing, reporting and visualising subsurface geology, hydrology and chemical contaminant data. It integrates data import/validation and compliance reporting, with the graphics and mapping functions needed for a complete CERCLA, RCRA, DOD or DOE site assessment or monitoring project. Contains links to other GIS software. Evaluated by the US EPAs Superfund Innovative Technology Evaluation (SITE) program. Features for chemistry: Isopleth maps of soil/water quality in plan or section view; Chemical concentration time series graphs inter/intra-well; Trilinear piper diagrams; Chemical concentration versus distance graphs; User defined alerts; Graphical summary of statistics; Presentation quality reports and data export; IRPIMS export; Direct exporting to leading 3D modelling and analysis tools. Features for geology: User customisable boring logs; Smart geological cross-section diagrams; Isopach maps; Structure maps; Modflow integration. Features for hydrology: Density corrected water level contour maps; Floating product contour maps; Hydraulic conductivity contour maps; Water elevation versus time graphs; Floating product thickness versus time graphs; Extraction well graphs; Modflow integration.

Supplier in USA : GIS-Solutions, Inc.

Name: HoleBASE II

Publisher : Key Systems

Platform : Win95/98, WinNT

Status : Commercial

Description : HoleBASE II is a database application dedicated to the storage, manipulation, and presentation of geotechnical and geo-environmental data from ground investigations. Features include Site Plan and Geological Sections, laboraratory test and contaminant data reporting functions, batch printing of multiple forms and borehole logs, and extensive query and report facilities. It also creates XYZ files of ground surface, geological, chemical and contamination data for proprietary modelling software to create digital ground models. It is a modular application, additional modules may be added by remote licensing. It may be run on stand-alone PC's or installed on a network server. The optional data entry module for networks allows data entry simultaneously from any number of networked PC's. The Base Module includes: Multiple Project Relational Database, Geotechnical and In situ Test Tables: Data Entry Screens with on-line Help and Spell Checking: Editable Libraries for Legends, Descriptions, and Text Snippets: Form Designer for Borehole Logs, Data Charts, and Header Sheets, including many ready-to-use example layouts. Single or Batch Printing. Import/Export of Borehole and In situ Test Data, AGS Format Checker, Worksheet-to-AGS Converter. Other available modules include: Query Manager - Report Designer (QM-RD); Laboratory Test (LT); Penetrometers and Downhole Testing (PDT); Site Plan and Geological Sections (SPG); Digital Modelling Link (DML).

Cost : Base module GB Pounds 1195, Optional modules GB Pounds 145 to 495

Supplier in United Kingdom : Key Systems

Name: SID

Publisher : M Z Associates

Platform : DOS, Win3x

Status : Commercial

Description : Enter, store and output geotechnical data from site investigation fieldwork and laboratory testing. Produce borehole and trial pit logs in predefined or to any user customised format. Plot field and laboratory test results on over 170 predefined graphs or set up your own graph types. Contour geological strata or test data. Data for plotting can be selected by borehole and stratum or by multiple criteria. Data can be imported and exported in AGS format or to and from spreadsheets. Data integrity checking. Site plan and geological section plotting. Calculation of derived parameters from test results. Links to AutoSketch or AutoCAD for log, plan and section drawing, and to Grapher and Surfer for plotting. The Standard version provides the data input, edit and reporting functions, log production and basic graphing and includes AGS input and output. The Professional version provides more comprehensive data manipulation functions including data integrity checks, parameter calculations, QA tracking and over 170 pre-formatted graphical outputs including strata contouring and 3D block diagrams.

Cost : From GB Pounds 595

Supplier in United Kingdom : M Z Associates

Name: TECHBASE

Publisher : MINEsoft Ltd

Platform : DOS, Win3x, Mac, UNIX, MIPS, HP, Sun, DEC, Silicon Graphics, IBM R/S

Status : Commercial

Description : Modular software package for Mining, Engineering, Environmental, Geotechnical industries. Based around a relational database for exploration geology and engineering information provides facilities for database management, statistics, graphics and graphical analysis, 2D and 3D modelling, mining and mineral exploration and development, oil and gas exploration and development, groundwater, slope stability, coordinate conversions, data and graphical transfer to and from most other programs. Accommodates 1D, 2D, 2.5D and 3D data as well as polygonal data for property and geological limits. Data can be imported from dBase, Lotus 123 spreadsheets or ASCII files. Handles AGS format data and carries out format checking. Graphical output can be exported as DXF files. Produces: cross sections with modelled soil layers, user definable format borehole logs, 3D perspective views with x-y lines or contours, maps with contoured data, time series plots. Includes: mathematical modelling options, the calculation of cut and fill volumes, input of digitised data. Mapping capabilities include: multiple layers, contouring of surface, subsurface, isopach, geologic unit and properties, integration of logos, scales and legends. Geochemical capability: store, analyse, and manipulate multi-element laboratory data including duplicate and standard samples, carry out statistical analyses, output with proportional symbols, data posting and data stacking. Downhole and surface geophysical data can be graphically displayed. Borehole data management: collar information, downhole assay information, lithology, alteration, compositing.

Cross sections: posting of vertical, deviated and inclined hole information, graphical display of analytic and geologic data along the hole. Optional modules available for: groundwater flow prediction using the USGS MODFLOW equation, 3D slope stability analysis, water quality analysis in Piper and Stiff plots.

Cost : Modular, approximately US Dollars 2500

APPENDIX 2

STRUCTURE OF DATA TABLES IN THE PRELIMINARY INVESTIGATION DATABASE

Structure of Data Tables in the Preliminary Investigation Database

Legend:
K - Unique Identifier, **T** - Text Field, **N** - Numerical Field

PROJ			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
PROJ_NAME	T - N	Project Title	
PROJ_LOC	T - N	Location	
PROJ_CLNT	T	Client Name	
PROJ_CONT	T	Contractors Name	
PROJ_ENG	T	Project Engineer	
PROJ_REM	T	General Remarks	
PROJ_DATE	dd/mm/yy	Date of Production of Data	
PROJ_AGS	N	AGS Issue Number	

SITE			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
SITE_ID	T - N	Site Identifier	K
SITE_NAM	T	Site Name	
SITE_ADD1	T - N	Site Address (line 1)	
SITE_ADD2	T - N	Site Address (line 2)	
SITE_CITY	T	Site City	
SITE_CONT	T	Site County	
SITE_COTR	T	Site Country	
SITE_CORT	T	Type of Co-ordinates	
SITE_XCOR	N	X-Co-ordinates	
SITE_YCOR	N	Y-Co-ordinates	
SITE_AREA	N (m^2)	Area Site Covers	
SITE_CUOW	T	Current Owner	
SITE_ADAU	T	Administration Authority	
SITE_PLRS	T	Planning Restrictions	
SITE_ACBY	T	Accessibility	
SITE_ACPT	T	Access Points to Site	
SITE_REM	T	Remarks	

PREL			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
SITE_ID	T - N	Site Identifier	K
PREL_ID	T - N	Preliminary Investigation Identifier	K
PREL_DENG	T - N	Desk Study Engineer	
PREL_DSRT	dd/mm/yy	Start of Desk Study	
PREL_DCMP	dd/mm/yy	Completion Date of Desk Study	
PREL_DMAP	T	Maps Relating to Site	
PREL_REM	T	Remarks	
PREL_RENG	T	Site Reconnaissance Engineer	
PREL_RSRT	dd/mm/yy	Start Date of Site Reconnaissance	
PREL_RCMP	dd/mm/yy	Completion Date of Site Reconnaissance	
PREL_RAIM	T	Main Aim of Site Reconnaissance	
PREL_REM	T	Remarks	

ZONE			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
SITE_ID	T - N	Site Identifier	K
ZONE_ID	T - N	Zone Identifier	K
ZONE_RESN	T	Reason for Zone	
ZONE_AREA	N (m^2)	Area of Zone	
ZONE_COR1	N	Co-ordinate of Zone	
ZONE_COR2	N	Co-ordinate of Zone	
ZONE_COR3	N	Co-ordinate of Zone	
ZONE_COR4	N	Co-ordinate of Zone	
ZONE_COR5	N	Co-ordinate of Zone	
ZONE_REM	T	Remarks	

GRDW			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
GRDW_ID	T - N	Groundwater Identifier	K
GRDW_VULB	T	Groundwater Vulnerability	
GRDW_SOFV	T	Source of Information (vulnerability)	
GRDE_HIGH	N (m)	Groundwater Depth High	
GRDE_LOW	N (m)	Groundwater Depth Low	
GRDE_TIRG	N (m)	Tidal Range	
GRDE_FLDT	T	Groundwater Flow Direction	
GRDW_VELC	N (m/s)	Groundwater Velocity	
GRDW_QUTY	T	Groundwater Quality	
GRDW_SOFQ	T	Source of Information (quality)	
GRDW_LWAT	T	Local Water Authority	
GRDW_REM	T	Remarks	

GEOG			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
SITE_ID	T - N	Zone Identifier	K
GEOG_ID	T - N	Geography Identifier	K
GEOG_TERN	T	Type of Terrain	
GEOG_POP	T	Population of Area	
GEOG_VGTP	T	General Vegetation Types	
GEOG_PEDS	T	General Soil Types	
GEOG_SOIF	T	Source of Information	
GEOG_RENA	T	Relevant Pub, Street Names	
GEOG_INLR	T	Information from Local Residents	
GEOG_REM	T	Remarks	

VEGE			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
VEGE_ID	T - N	Vegetation General Identifier	K
VEGE_COVR	T	General Vegetation Cover	
VEGE_REMA	N (m^2)	Details of Areas	
VEGE_TREE	Y/N	Trees Present on Site	
VEGE_REMT	T	Details of Trees	
VEGE_SHBS	Y/N	Shrubs Present on Site	
VEGE_REMS	T	Details of Shrubs	
VEGE_GRSS	Y/N	Grasses Present on Site	
VEGE_REMG	T	Details of Grasses	
VEGE_FOBS	Y/N	Forbs Present on Site	
VEGE_REMF	T	Details of Forbs	
VEGE_MOSS	Y/N	Mosses & Ferns on Site	
VEGE_REMM	T	Details of Mosses & Ferns	
VEGE_GSHE	Y/N	General Signs of Poor Health	
VEGE_SOIF	T	Source of Information	
VEGE_REM	T	Remarks	

VEDT			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
VEGE_ID	T - N	Vegetation General Identifier	K
VEDT_ID	T - N	Vegetation Detail Identifier	K
VEDT_TYPE	T	Type of Vegetation	
VEDT_HATH	T	General Health of Vegetation Type	
VEDT_REMH	T	Remarks	
VEDT_LEVH	T	Health of Leaves	
VEDT_REML	T	Remarks	
VEDT_ROTH	T	Health of Roots	
VEDT_REMR	T	Remarks	
VEDT_YSRG	Y/N	Young Seedling Regeneration	
VEDT_REMY	T	Remarks	
VEDT_VGDB	Y/N	Vegetation Die Back	
VEDT_REMD	T	Remarks	

METE			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
METE_ID	T - N	Meteorological Identifier	K
METE_GNCC	T	General Climatic Conditions	
METE_MXTP	N (deg C)	Maximum Temperature	
METE_XTMH	T	Maximum Temperature Month	
METE_MNTP	N (deg C)	Minimum Temperature	
METE_MTMH	T	Minimum Temperature Month	
METE_MXPC	N (mm/day)	Maximum Precipitation	
METE_MXMH	T	Maximum Precipitation Month	
METE_MNPC	N (mm/day)	Minimum Precipitation	
METE_MNMH	T	Minimum Precipitation Month	
METE_SOIF	T	Source of Information	
METE_REM	T	Remarks	

HIST			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
HIST_ID	T - N	History General Identifier	K
HIST_ARCH	Y/N	Archaeological Interest	
HIST_REMA	T	Archaeological Interest Remarks	
HIST_SUBS	Y/N	Evidence of Subsidence	
HIST_REMS	T	Remarks Regarding Subsidence	
HIST_EVSA	Y/N	Evidence of Seismic Activity	
HIST_REMSA	T	Remarks Regarding Seismic Activity	
HIST_SOIF	T	Source of Information	
HIST_REM	T	Remarks	

HIDT			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
HIST_ID	T - N	History General Identifier	K
HIDT_ID	T - N	History Detail Identifier	K
HIDT_NAME	T	Name of Owner	
HIDT_USE	T	Previous Use	
HIDT_FEAT	T	Features Associated with Use	
HIDT_STAT	dd/mm/yy	Start Date of Use	
HIDT_FINS	dd/mm/yy	Finish Date of Use	
HIDT_DURT	N (Year)	Duration of Use	
HIDT_LVOD	N (m)	Level Above Ordnance Datum	
HIDT_SOIF	T	Source of Information	
HIDT_REM	T	Remarks	

GEOL			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
GEOL_ID	T - N	Geology General Identifier	K
GEOL_NULY	N	Number of Subsurface Layers	
GEOL_TYPE	T	Types of Subsurface Material	
GEOL_FEAT	Y/N	Subsurface Features Within Zone	
GEOL_MAPS	T	Maps Used	
GEOL_REM	T	Remarks	

GEDT			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
GEOL_ID	T - N	Geology General Identifier	K
GEDT_ID	T - N	Geology Detail Identifier	K
GEDT_DESC	T	Stratum Description	
GEDT_LYNO	N	Layer Number	
GEDT_LYDT	N (m)	Depth to Top of Layer	
GEDT_LYTH	N (m)	Layer Thickness	
GEDT_CHAR	T	Characteristics	
GEDT_FEPT	T	Features Present	
GEDT_SOIF	T	Source of Information	
GEDT_REM	T	Remarks	

STFT			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
GEDT_ID	T - N	Geology Detail Identifier	K
STFT_ID	T - N	Structural Feature Identifier	K
STFT_TYPE	T	Structural Feature Type	
STFT_FTDD	N (deg)	Dip & Direction of Feature	
STFT_FTLC	T	Location of Feature	
STFT_FTSZ	N (m)	Size of Feature	
STFT_SOIF	T	Source of Information	
STFT_REM	T	Remarks	

SERV			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
SERV_ID	T - N	Services General Identifier	K
SERV_NUMB	N	Number of Services Within Zone	
SERV_TYPE	T	Type of Services	
SERV_PLAN	T	Details of Plans For Services	
SERV_REM	T	Remarks	

SEDT			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
SERV_ID	T - N	Services General Identifier	K
SEDT_ID	T - N	Services Detail Identifier	K
SEDT_TYPE	T	Type of Service	
SEDT_RSAT	T	Responsible Authority	
SEDT_ELEV	N (m)	Elevation	
SEDT_TRED	T	Trend of Service Across Zone	
SEDT_STCX	N	Start X-Co-ordinate of Service	
SEDT_STCY	N	Start Y-Co-ordinate of Service	
SEDT_FNCX	N	Finish X-Co-ordinate of Service	
SEDT_FNCY	N	Finish Y-Co-ordinate of Service	
SEDT_SOIF	T	Source of Information	
SEDT_REM	T	Remarks	

TOPO			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
TOPO_ID	T - N	Topography Identifier	K
TOPO_FEAT	T	Feature	
TOPO_DRSL	T	Direction of Slope	
TOPO_DPSL	N (deg)	Dip of Slope	
TOPO_ELFT	N (m)	Elevation of Feature	
TOPO_SOIF	T	Source of Information	
TOPO_REM	T	Remarks	

FAUA			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
FAUA_ID	T - N	Fauna General Identifier	K
FAUA_TYPE	T	Types of Fauna	
FAUA_EVDC	T	Evidence of Fauna	
FAUA_HLTH	T	General Health of Fauna	
FAUA_DIVS	T	Diversity of Zone	
FAUA_ABNC	T	Abundance of Zone	
FAUA_OWNR	T	Owner of Fauna (if applicable)	
FAUA_REM	T	Remarks	

FADT			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
FAUA_ID	T - N	Fauna General Identifier	K
FADT_ID	T - N	Fauna Detail Identifier	K
FADT_SPCE	T	Fauna Species	
FADT_HLTH	T	Health of Species	
FADT_HEDT	T	Details of Health	
FADT_ABNC	T	Abundance of Species	
FADT_DIVS	T	Diversity of Species	
FADT_REM	T	Remarks	

SURW			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
SURW_ID	T - N	Surface Water General Identifier	K
SURW_SWZN	Y/N	Surface Water Present Within Zone	
SURW_TYPE	T	Type Surface Water	
SURW_GNQU	T	General Water Quality Within Zone	
SURW_REMQ	T	Remarks Regarding Quality	
SURW_GQEZ	T	Quality of Water Entering Zone	
SURW_REMQE	T	Remarks Water Entering Quality	
SURW_GQLZ	T	Quality of Water Leaving Zone	
SURW_REMQL	T	Remarks Water Leaving Zone	
SURW_SOIF	T	Source of Information	
SURW_REM	T	General Remarks	

SUDR			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
SURW_ID	T - N	Surface Water General Identifier	K
SUDR_ID	T - N	Surface Water Drainage Identifier	K
SUDR_DRTY	T	Type of Drainage	
SUDR_DRDI	T	Direction of Drainage	
SUDR_DRVL	N (m/s)	Velocity of Drainage	
SUDR_APWT	T	Appearance of Water	
SUDR_PHWT	N	pH of Water	
SUDR_ODPR	T	Odours Present	
SUDR_BUPR	Y/N	Bubbles Present	
SUDR_SPRL	T	Speed of Release	
SUDR_SPREM	T	Remarks Regarding Bubbles	
SUDR_GQWT	T	General Quality of Water	
SUDR_SOIF	T	Source of Information	
SUDR_REM	T	Remarks	

SUST			
Field Name	**Field Type**	**Field Description**	**Remarks**
PROJ_ID	T - N	Project Identifier	K
ZONE_ID	T - N	Zone Identifier	K
SURW_ID	T - N	Surface Water General Identifier	K
SUST_ID	T - N	Surface Water Storage Identifier	K
SUST_STTY	T	Type of Storage	
SUST_LOC	N	Location	
SUST_ETVL	N (m^2)	Estimated Volume	
SUST_APWT	T	Appearance of Water	
SUST_PHWT	N	pH of Water	
SUST_ODPR	T	Odours Present	
SUST_BUPR	Y/N	Bubbles Present	
SUST_SPRL	T	Speed of Release	
SUST_SPREM	T	Remark Regarding Bubbles	
SUST_GQWT	T	General Quality of Water	
SUST_SOIF	T	Source of Information	
SUST_REM	T	Remarks	

APPENDIX 3

QUESTIONNAIRE FORMAT

Questionnaire Format

Dear Participant

The questionnaire below is designed to gather views and information from a range of experts involved in the field of contaminated land assessment and reclamation. It is hoped that the data collected can be used to validate a knowledge-based system that has been compiled as part of a Ph.D. project. The project has investigated the use of knowledge-based systems within the investigation process of potentially contaminated sites. If you are interested in knowing more about the study, you can contact me at J.C.Martin@durham.ac.uk.

I would be extremely grateful if you could take a few minutes to complete the questionnaire and return it via e-mail.

Thank you very much for your participation and assistance
Apologies for cross-posting

John Martin
University of Durham

QUESTIONNAIRE

Evaluating the parameters used during the preliminary investigation of a potentially contaminated site.

Section A: Details of Respondent

Occupation :-

Number of years of experience :-

Type of organisation employed with :-
(example; academic institution, consultants, contractors)

Section B: Details of Indicators

During a preliminary investigation of a potentially contaminated site, there are a number of indicators that may provide clues to likely contaminants.
Please input the number which in your opinion best applies to each indicator, using the key below.

KEY: **5 = extremely useful, 4 = very useful, 3 = useful, 2 = little use, 1 = very-little use when used as sole indicator.**

(1) Current use of site [ ]
(example; railway land: oils, coal-dust, asbestos, lead, etc.)

(2) Historical use of site [ ]
(example; wood treatment: zinc, arsenic, tar, phenols, etc.)

(3) Presence of odours on site [ ]
(example; bad-eggs due to sulphur, organic effluents)

(4) Ground surface deposit type [ ]
(example; sewage sludge: mercury, nickel, zinc, etc.)

(5) Ground surface deposit colour [ ]
(example; blue due to copper, sulphur, zinc, etc.)

(6) Ground surface staining colour [ ]
(example; green due to copper, chromium, arsenic, etc.)

(7) Terrestrial vegetation *(grasses, trees, shrubs, etc.)*: tolerant species [ ]
(example; Brown bent grass due to zinc, lead)

(8) Terrestrial vegetation: visible health symptoms [ ]
(example; yellowing or browning of roots due to arsenic, lead)

(9) Visible symptoms concerned with soil microbiology [ ]
(example; reduced decomposition of leaf litter due to arsenic)

(10) Terrestrial invertebrates: tolerant species [ ]
(example; clubionid spiders: zinc)

(11) Terrestrial invertebrates: visible health symptoms [ ]
(example; earthworms, loss of saddle due to mercury)

(12) Aquatic invertebrates tolerant species [ ]
(example; crayfish: cadmium)

(13) Aquatic invertebrates visible health symptoms [ ]
(example; reduced number of mollusc taxa due to: zinc)

(14) Visible health symptoms relating to mammals [ ]
(example; reduced growth rate, bleaching of incisors due to: cadmium)

Section C: Details of Groundwater Movement

Groundwater plays an important role in the movement of contaminants. Therefore knowing the velocity of groundwater movement within a potentially contaminated site is extremely useful in the risk assessment process. Please indicate suitable ball park ranges of velocity for the following terms (m/day).

Fast	............m/day
Medium	............m/day
Slow	............m/day

Section D: Details of Targets

When undertaking a risk assessment process on a potential contaminated site it is essential to highlight vulnerable targets. The following land use classification has been constructed to cover the range of land uses that may be found neighbouring a potential contaminated site. Please indicate the risk that you would assign to the following neighbouring land uses. Using the terms high *(H)*, medium *(M)* and low *(L)*, with high relating to a land use that is most vulnerable to contamination.

(1) School [ ]

(2) Public open space [ ]

(3) Agricultural area [ ]

(4) General commercial [ ]

(5) Low density residential area [ ]

(6) Medium density residential area [ ]

(7) High density residential area [ ]

(8) Light industrial area [ ]

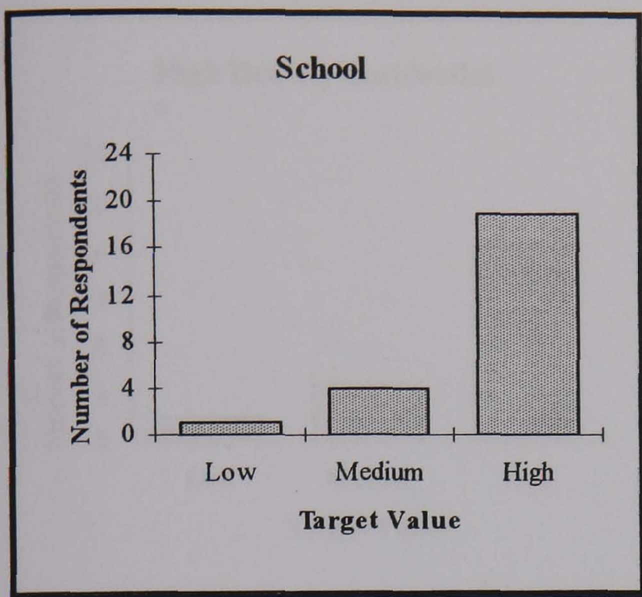
(9) Heavy industrial area [ ]

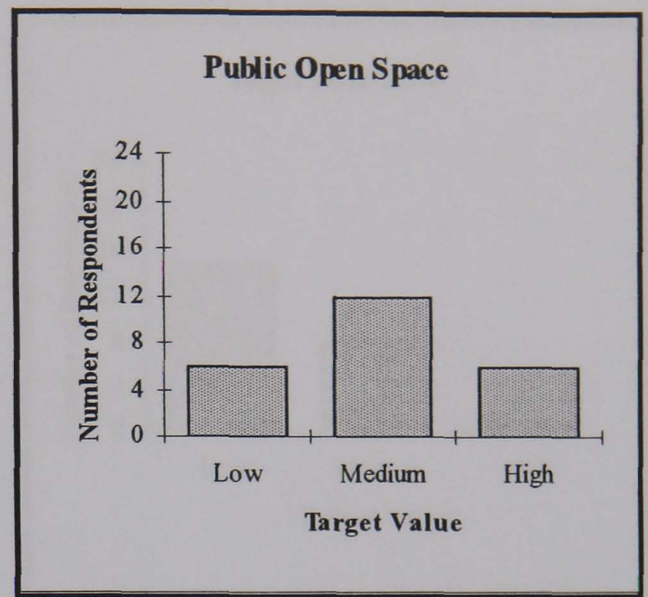
(10) Permanently paved area [ ]

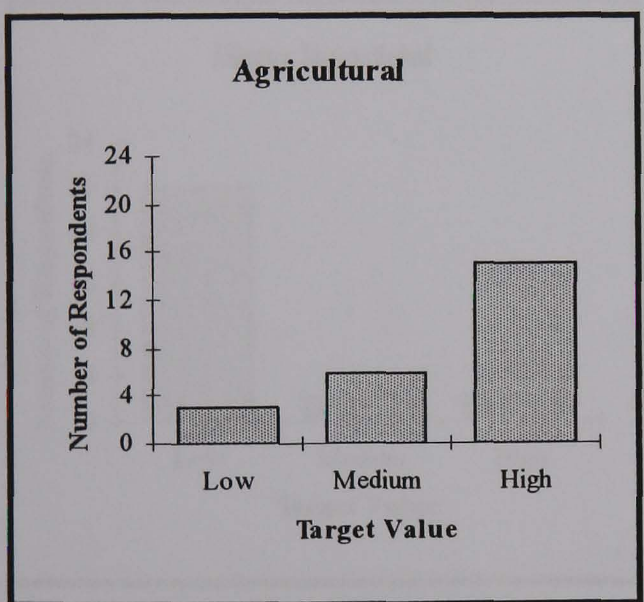
APPENDIX 4

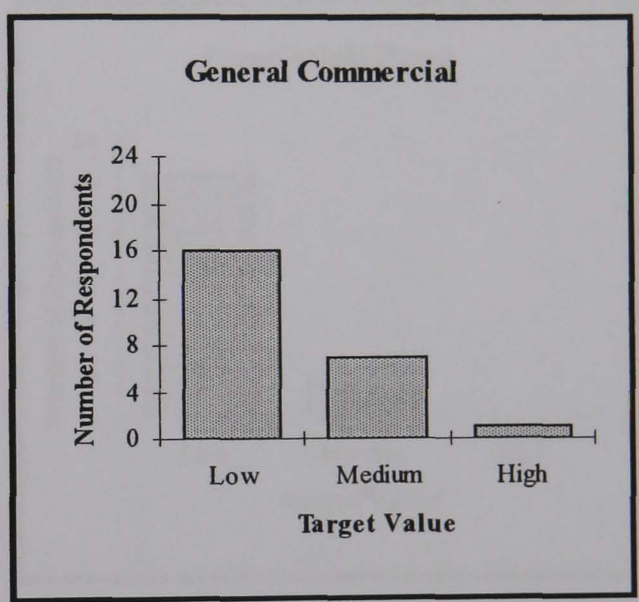
QUESTIONNAIRE RESULTS

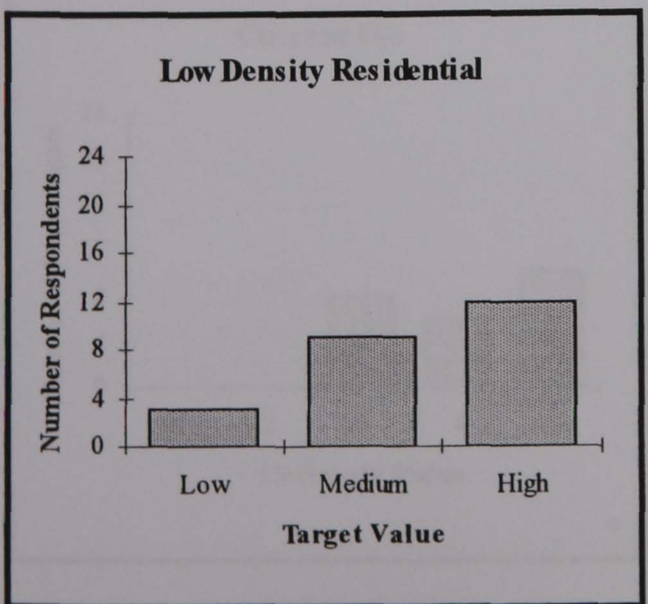
Questionnaire Results from the ">5 Years Experience" Classification

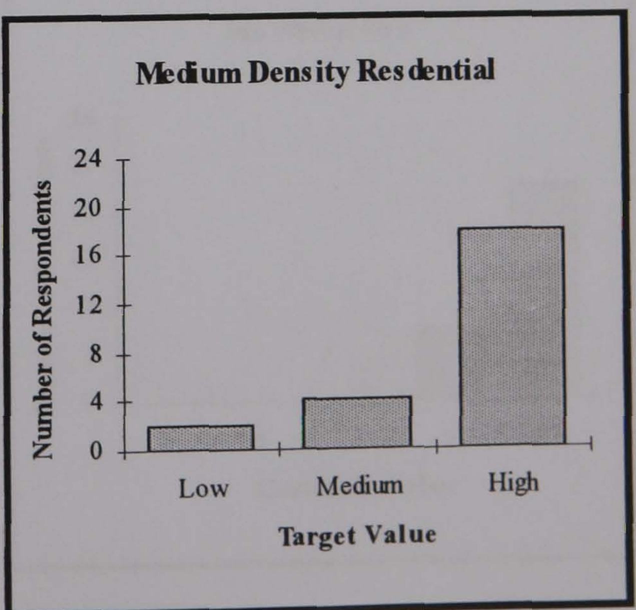




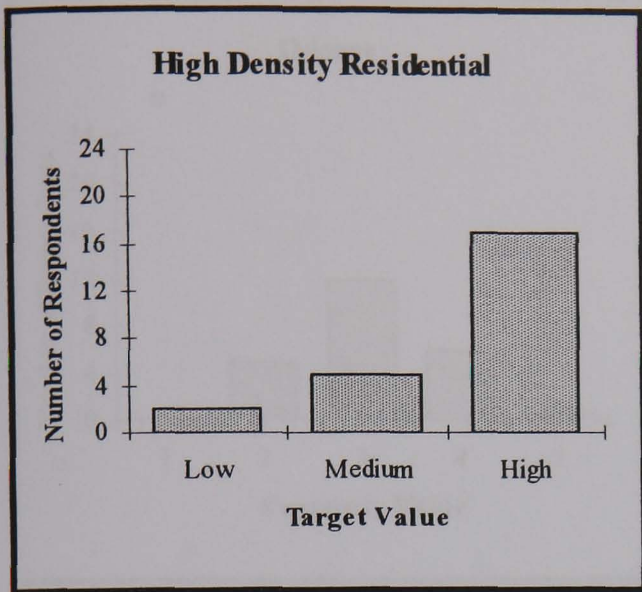


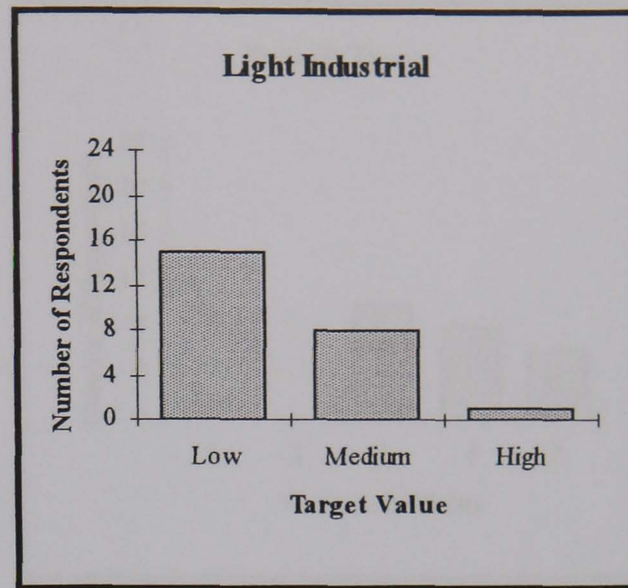


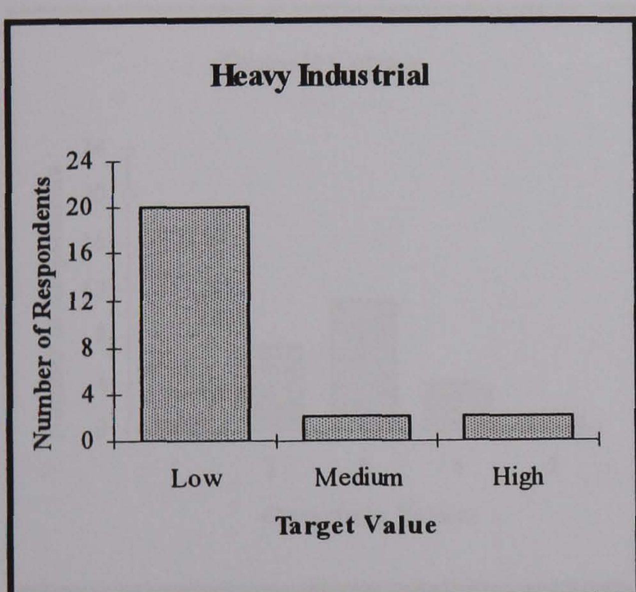


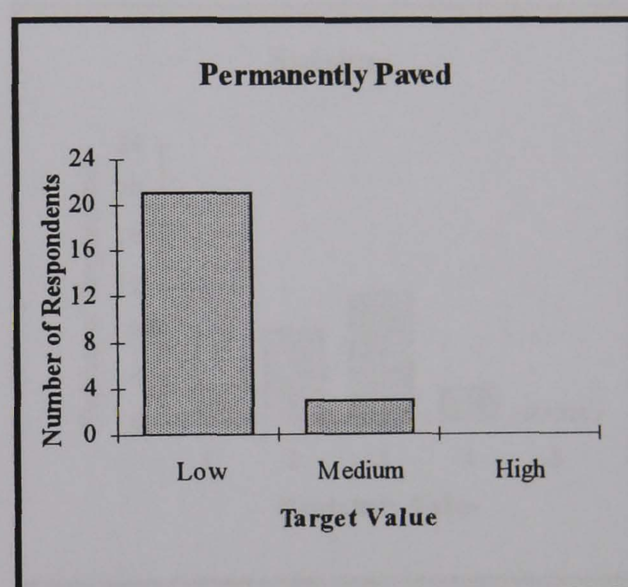


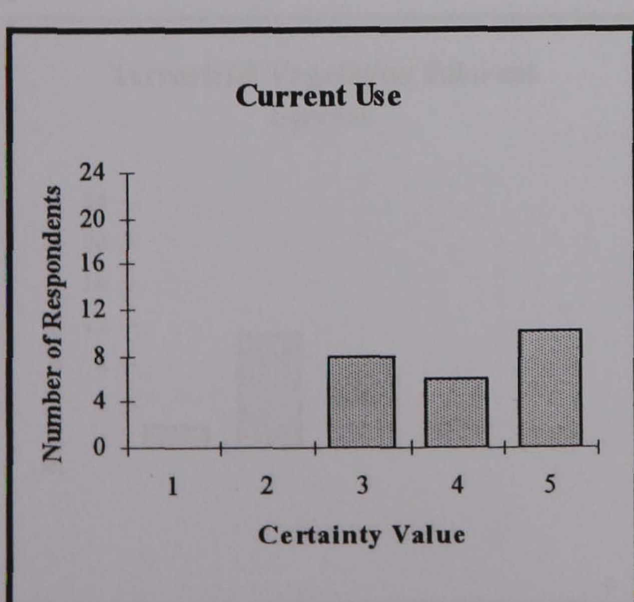
Questionnaire Results from the ">5 Years Experience" Classification

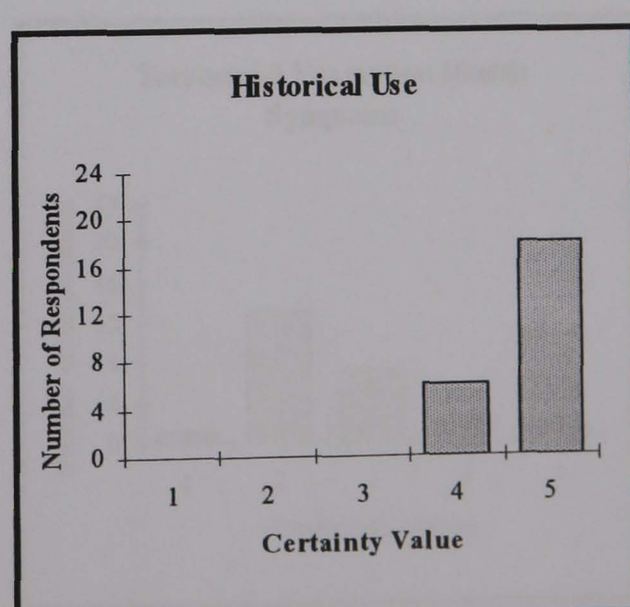




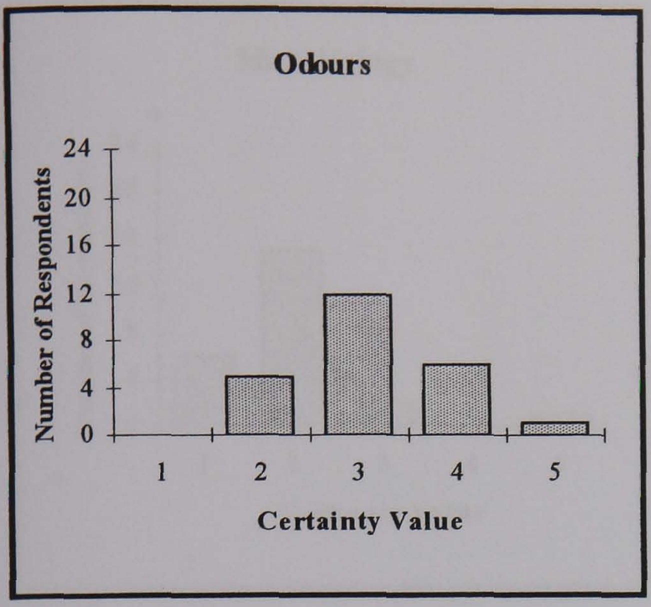


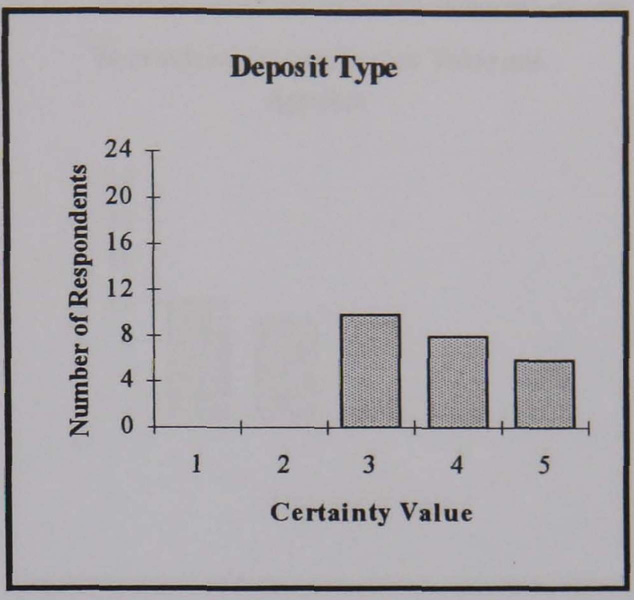


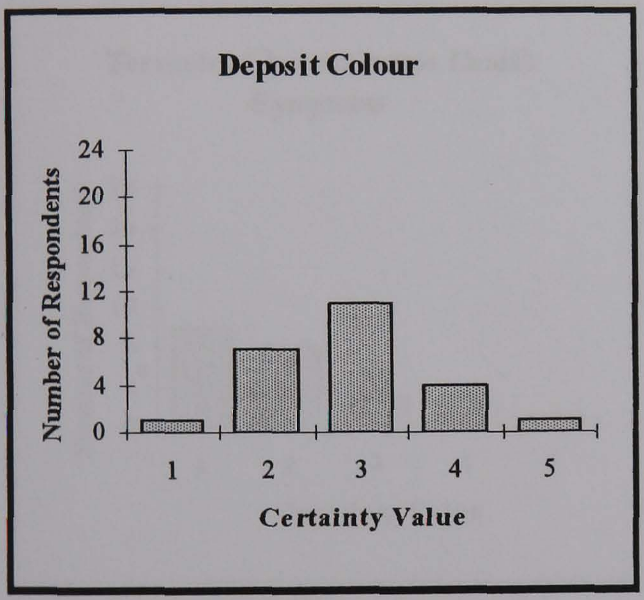


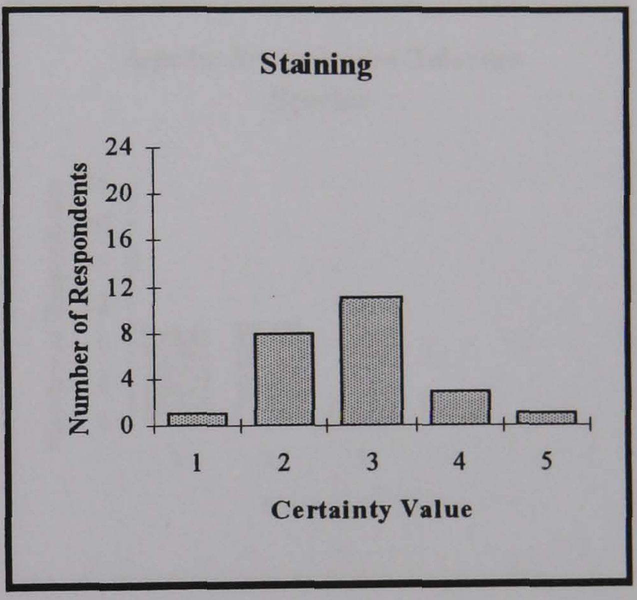


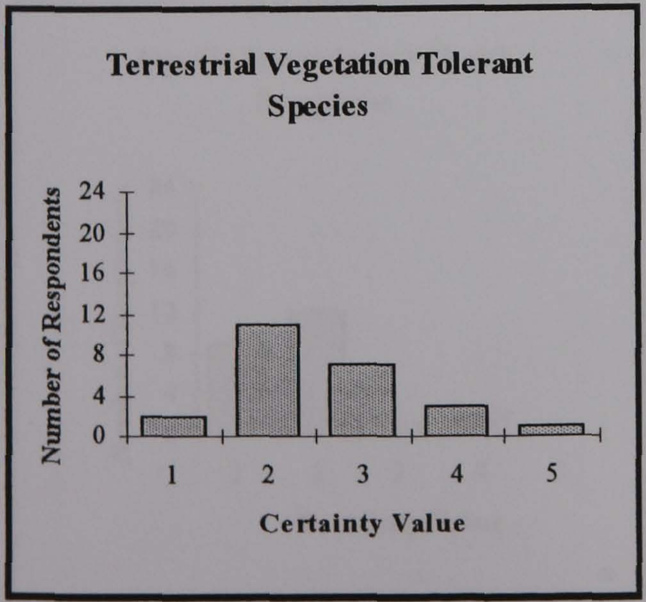
Questionnaire Results from the ">5 Years Experience" Classification

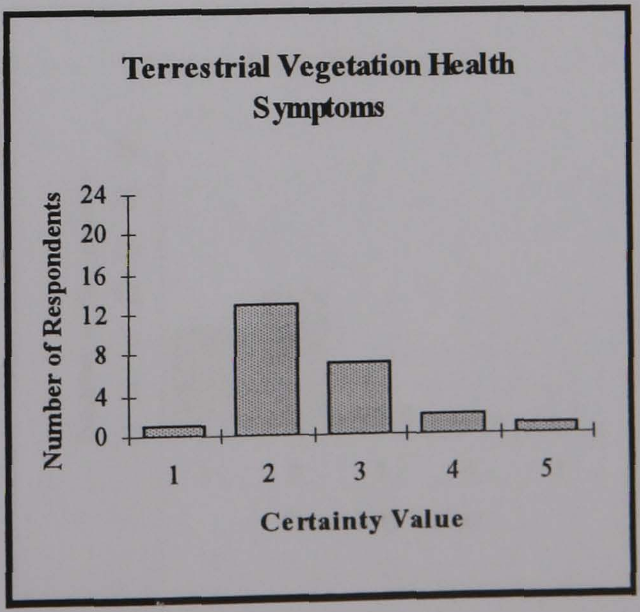




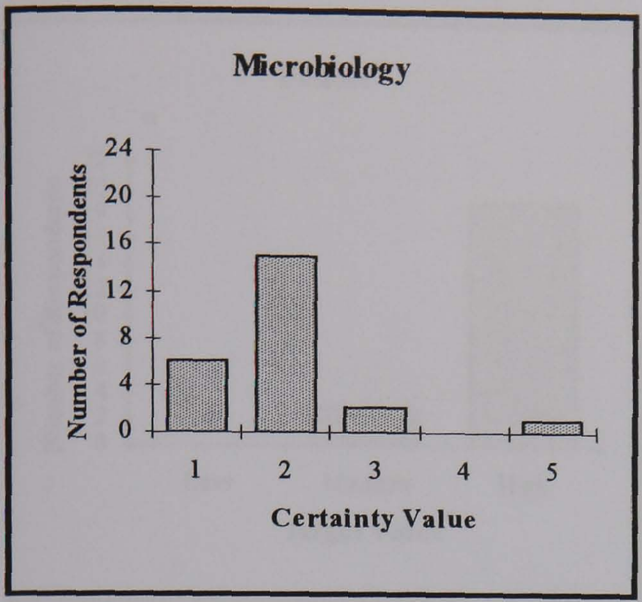


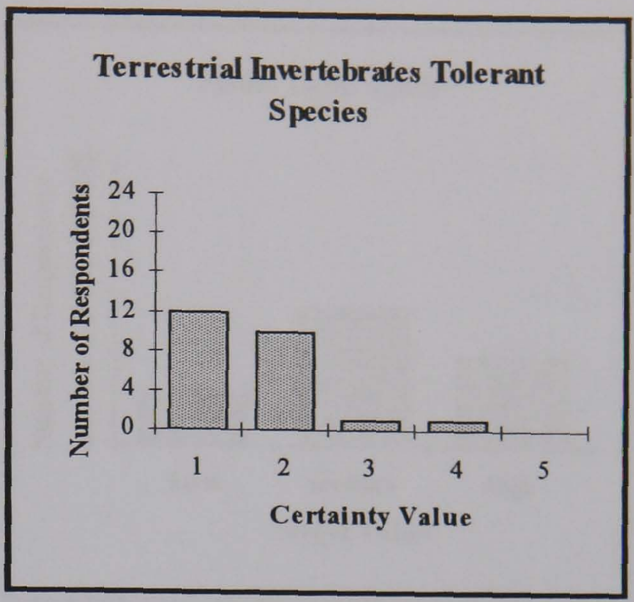


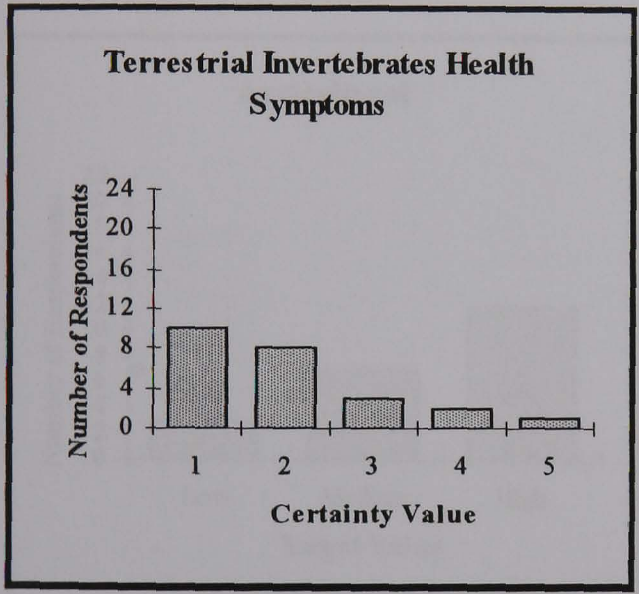


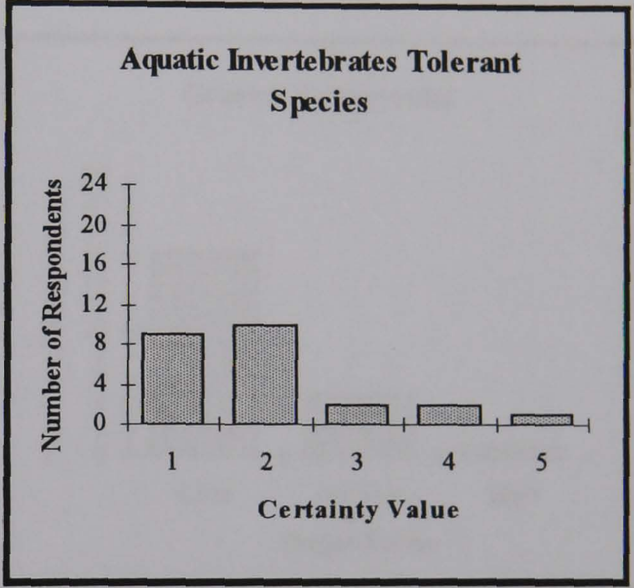


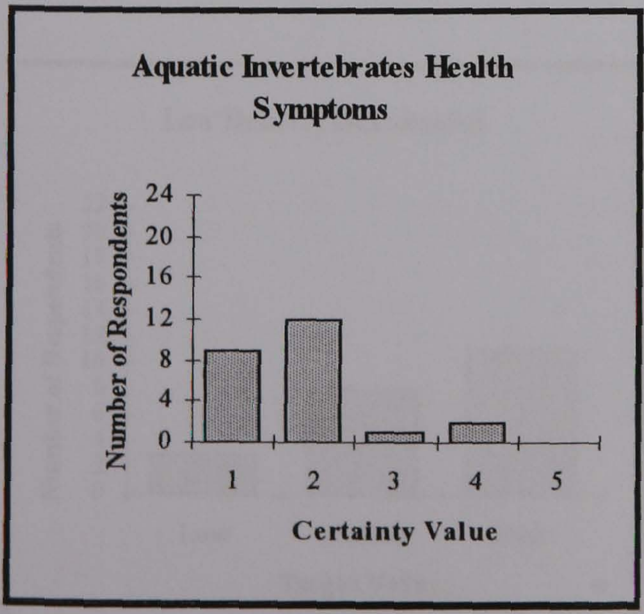
Questionnaire Results from the ">5 Years Experience" Classification

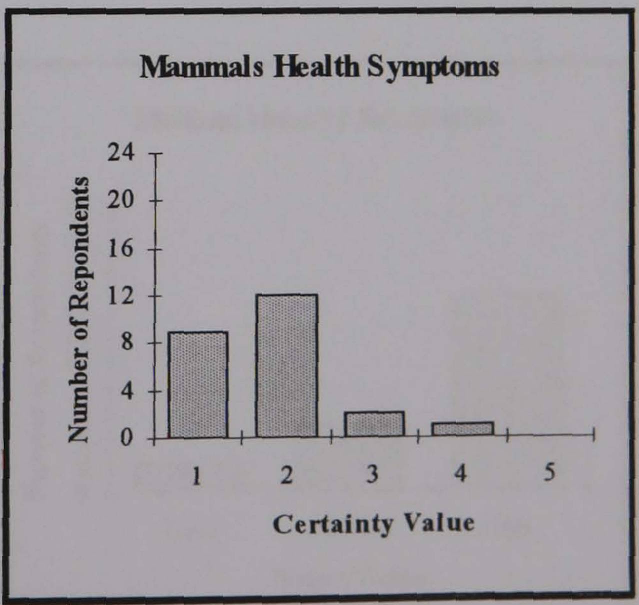




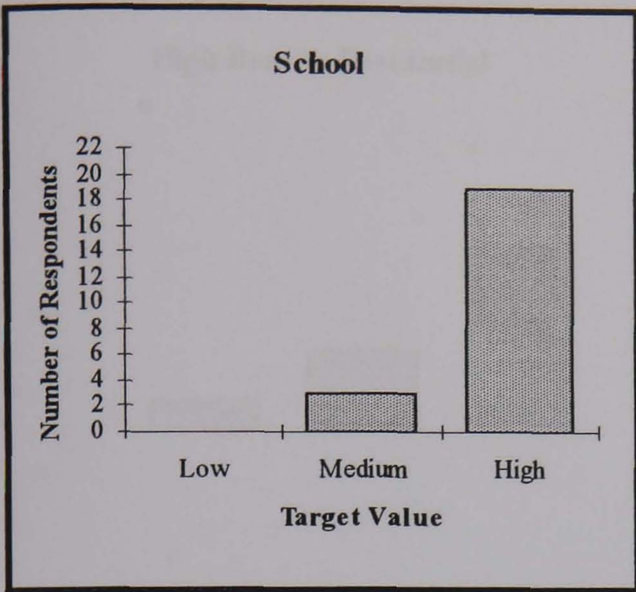


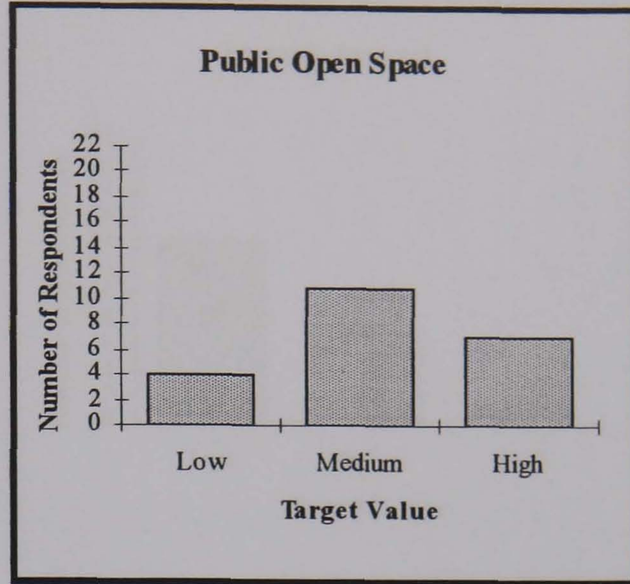


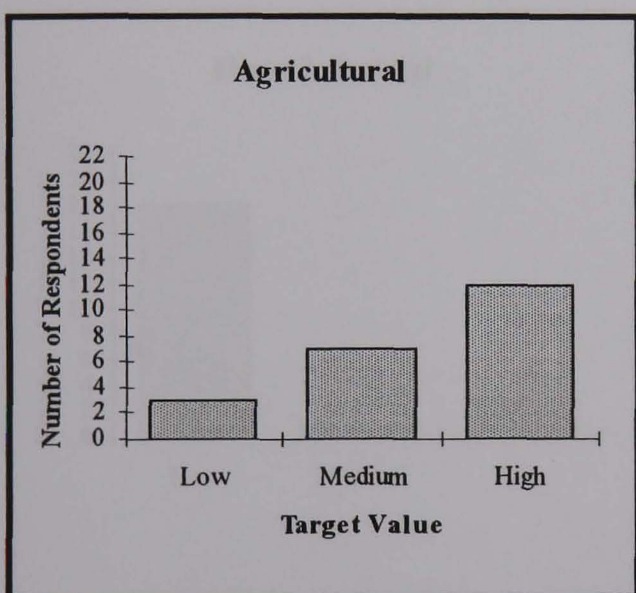


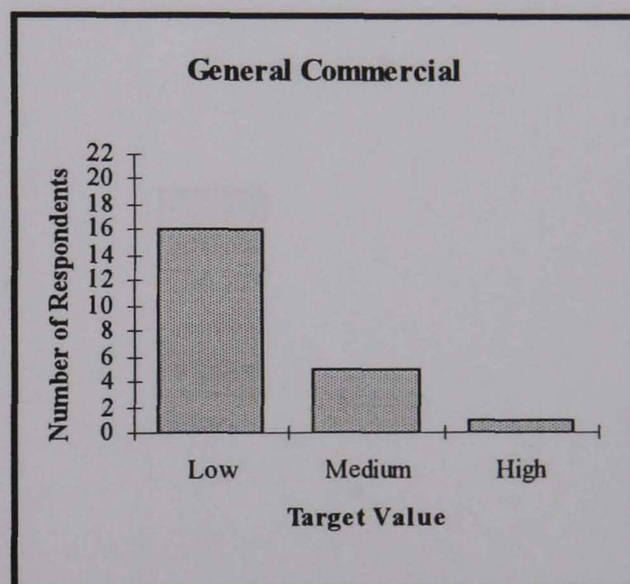


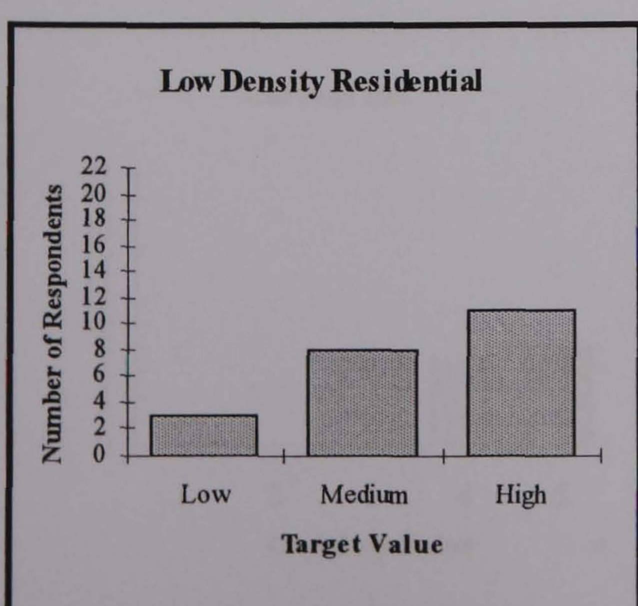
Questionnaire Results from the “>10 Years experience” Classification

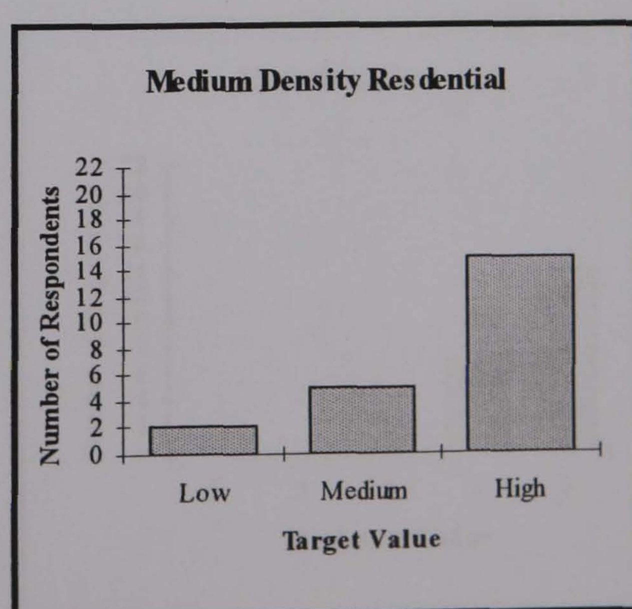




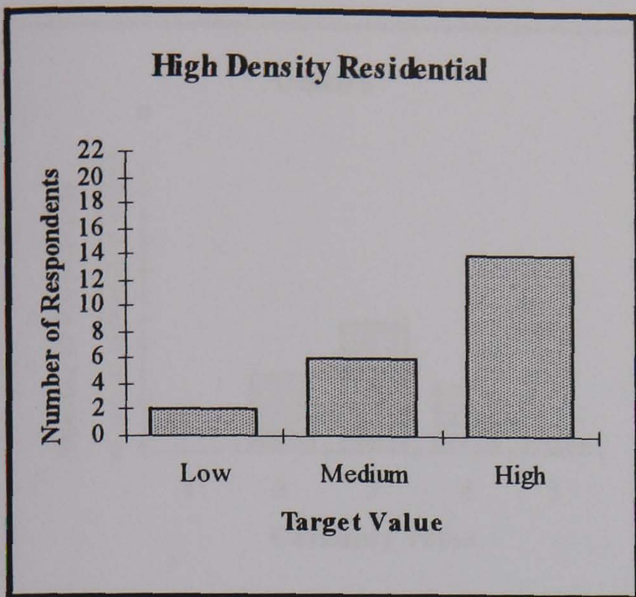


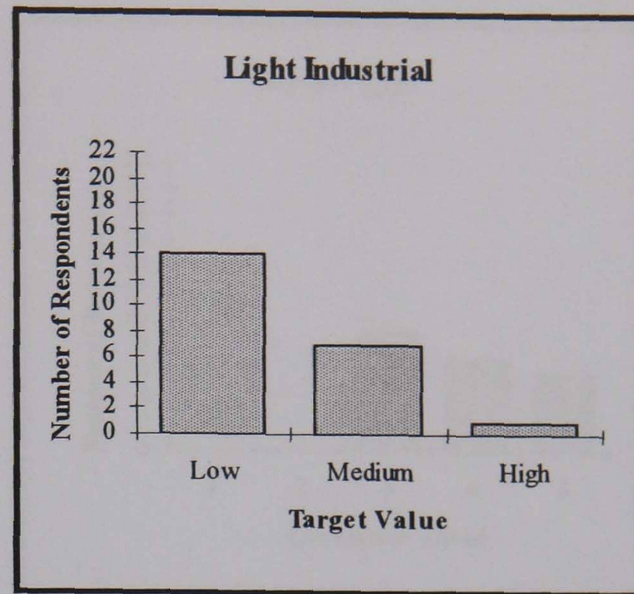


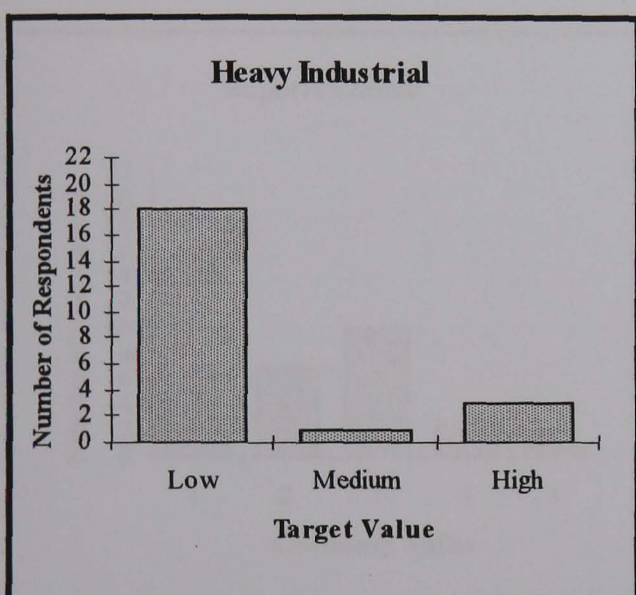


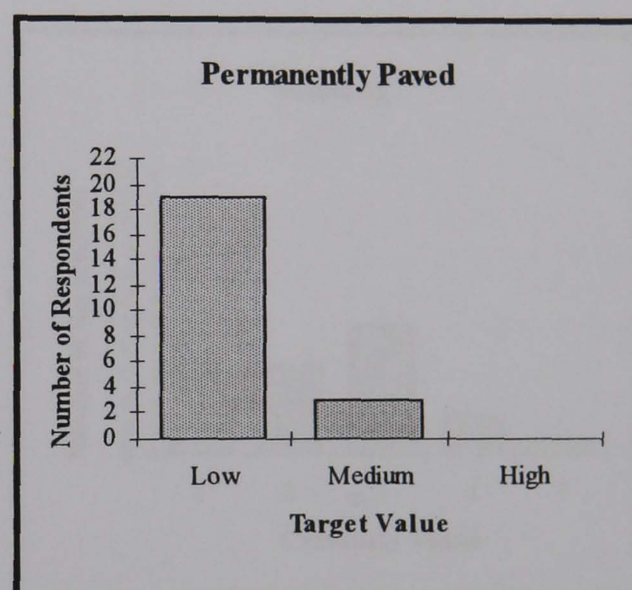


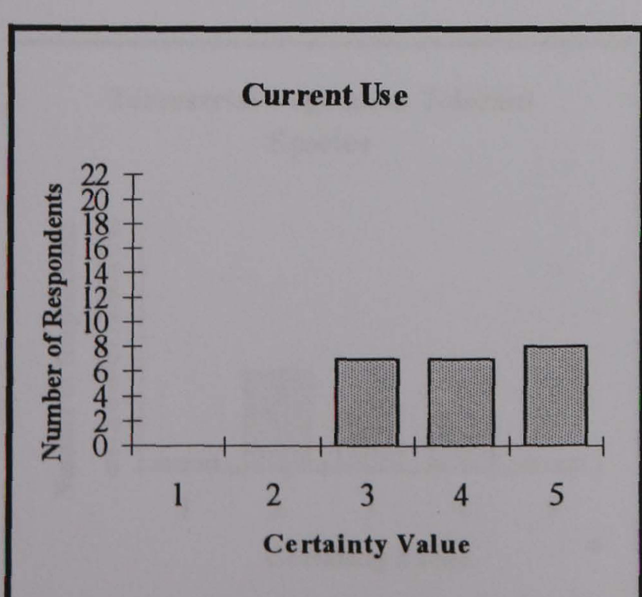
Questionnaire Results from the ">10 Years experience" Classification

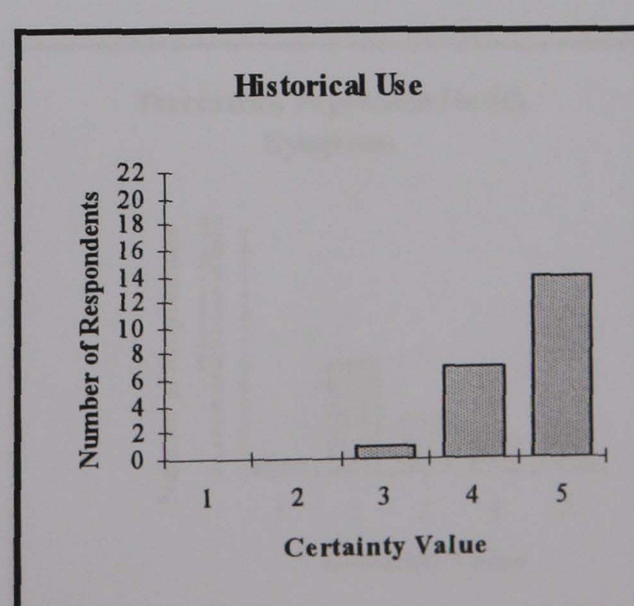




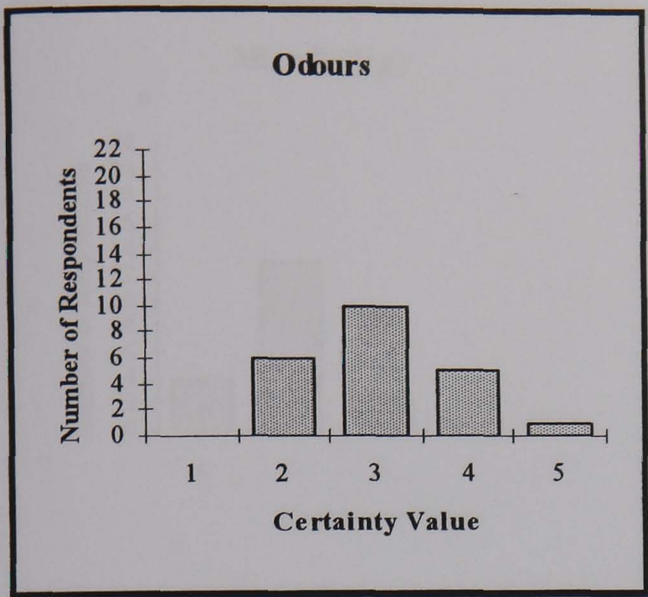


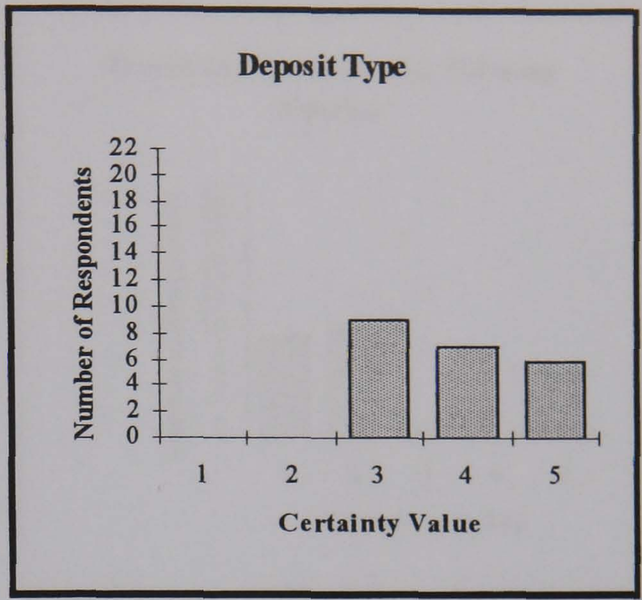


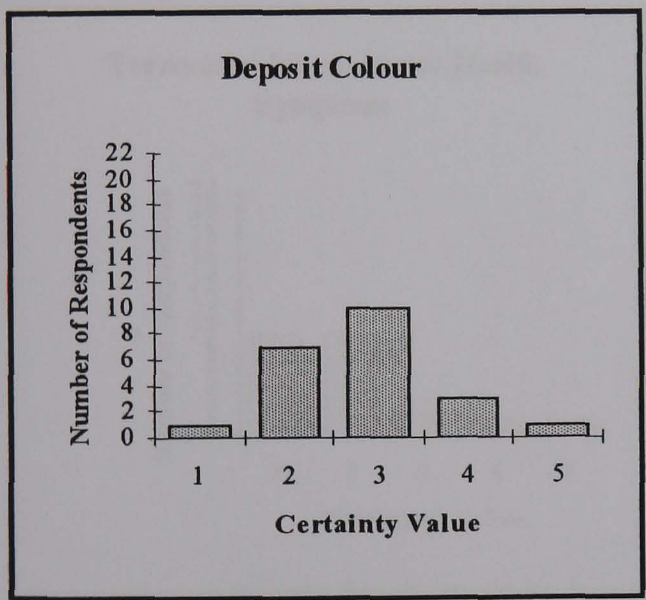


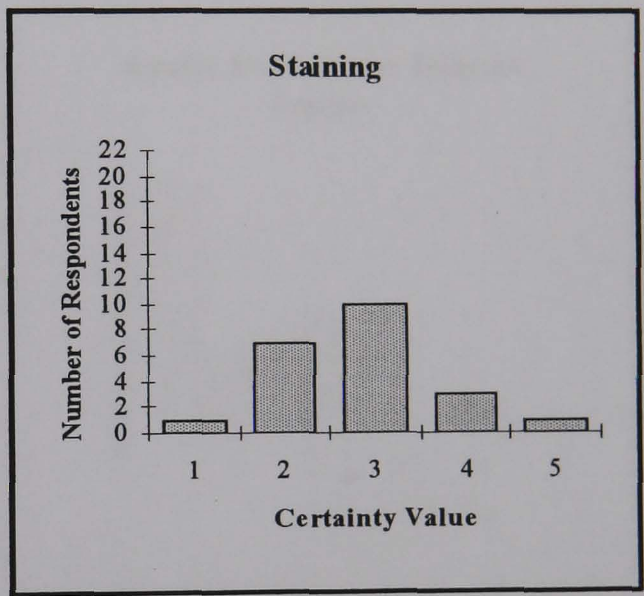


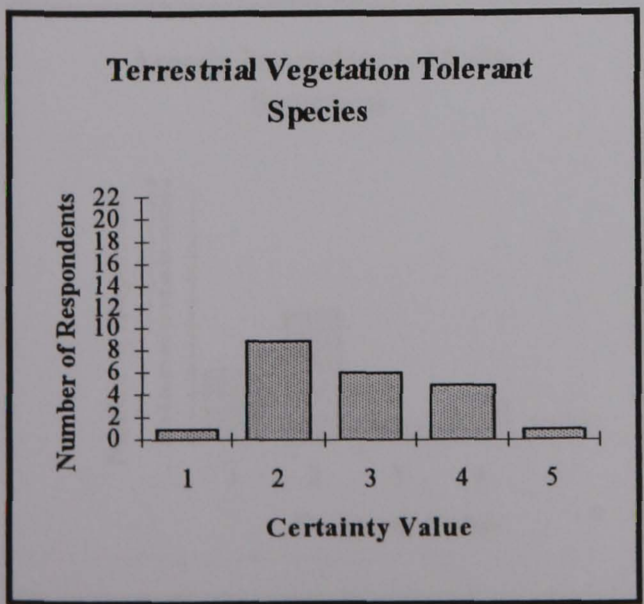
Questionnaire Results from the ">10 Years experience" Classification

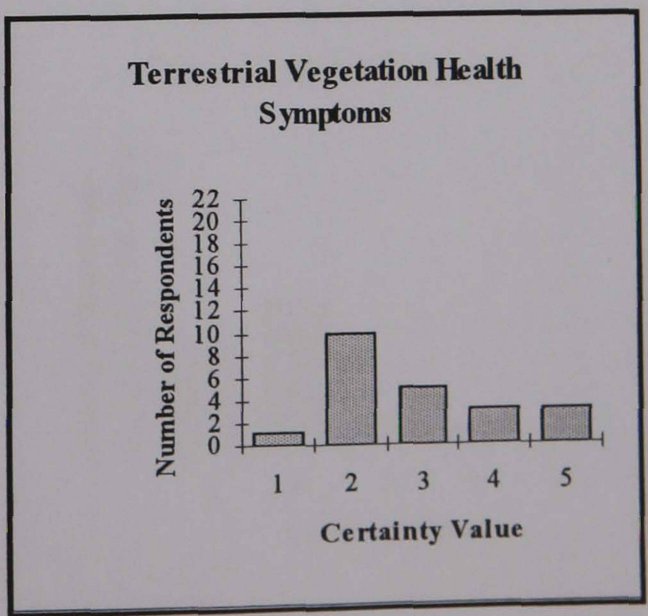




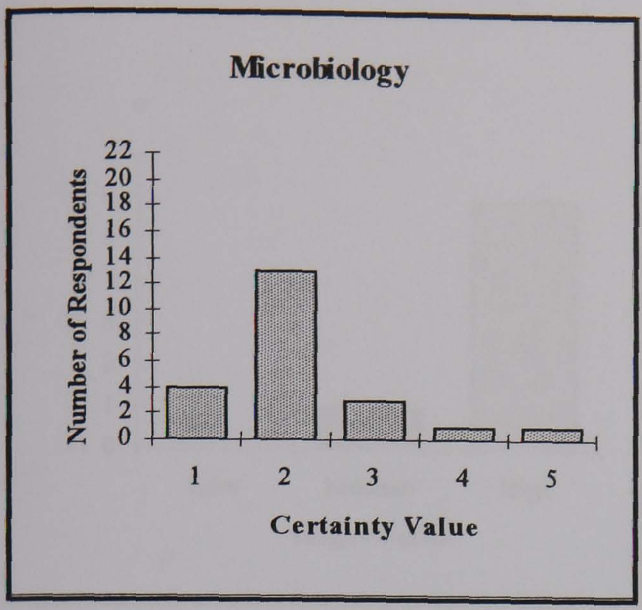


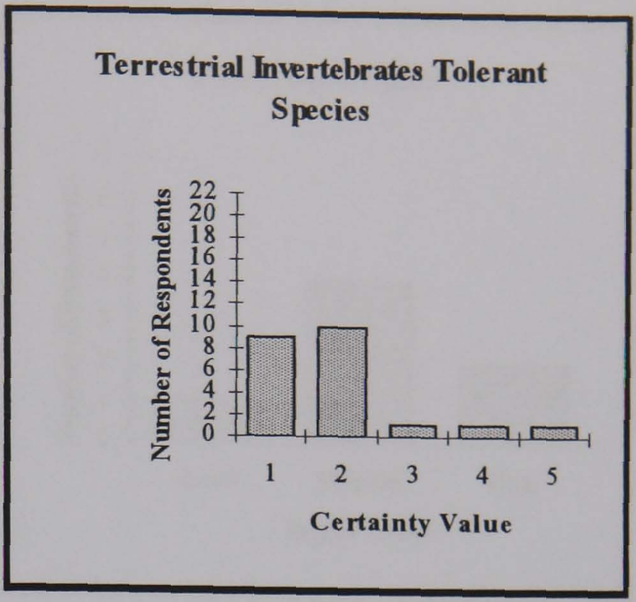


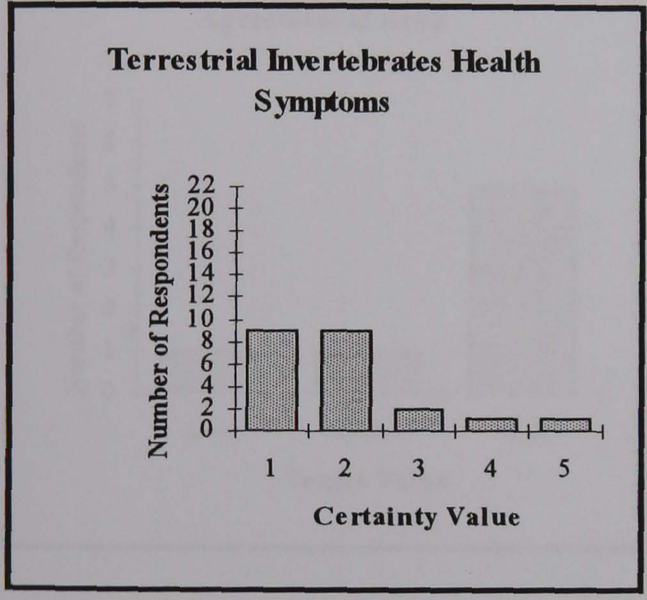


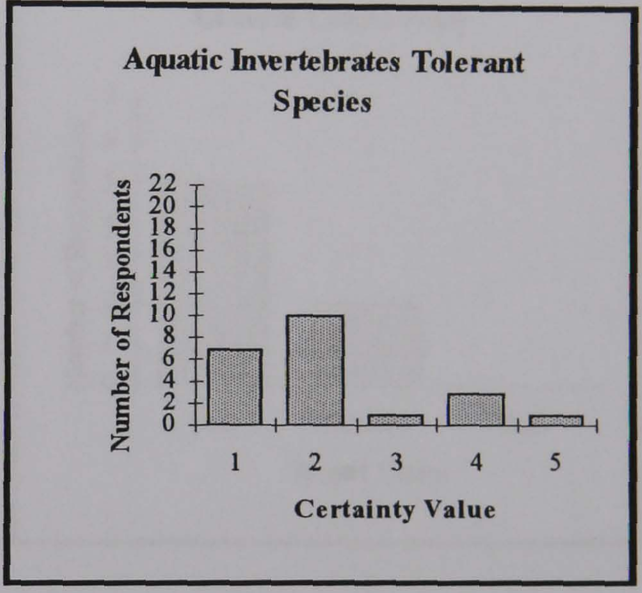


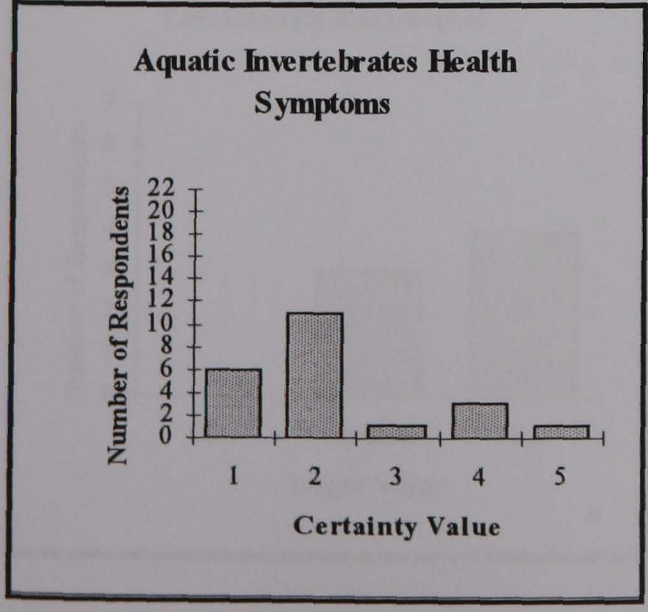
Questionnaire Results from the “>10 Years experience” Classification

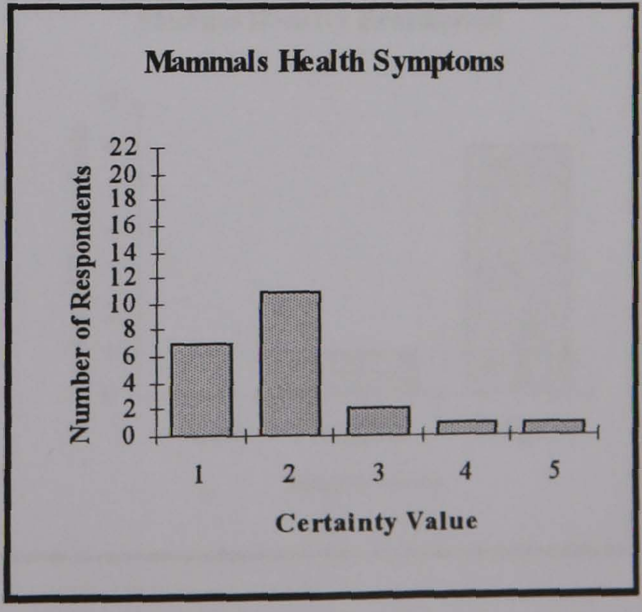




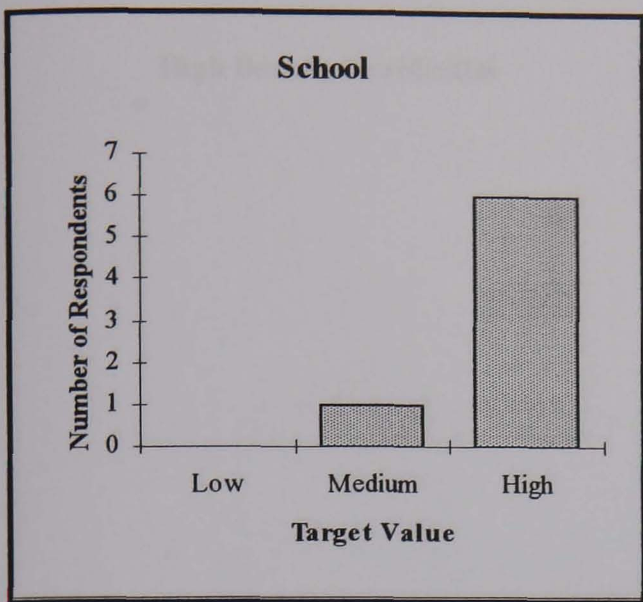


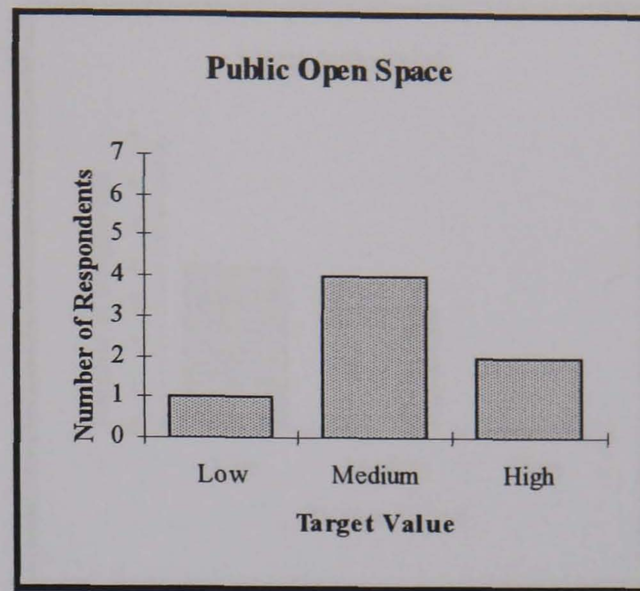


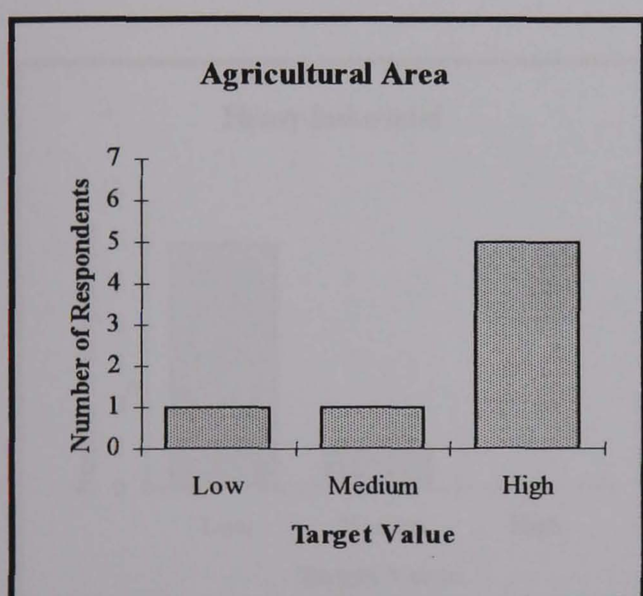


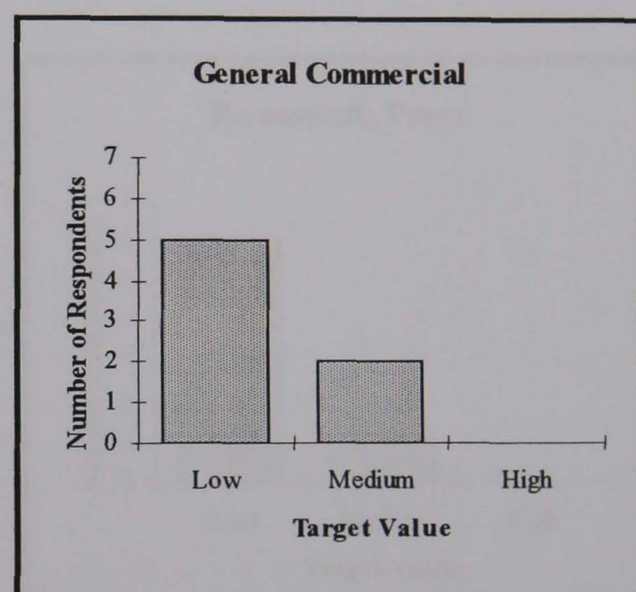


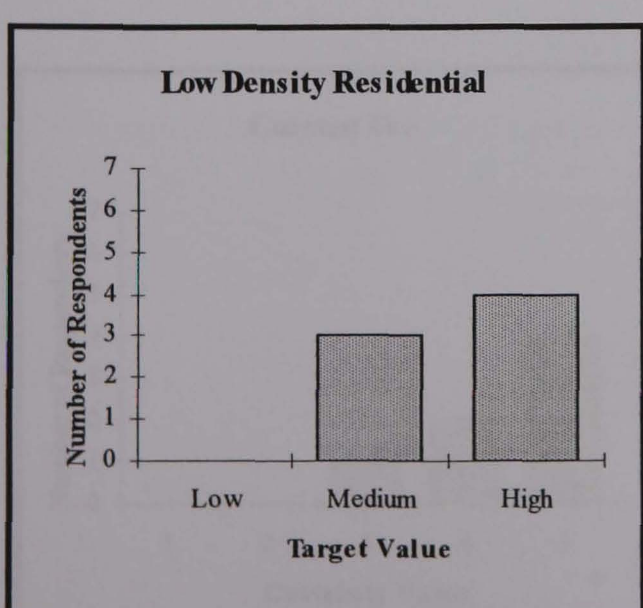
Questionnaire Results from the "Chemist" Classification

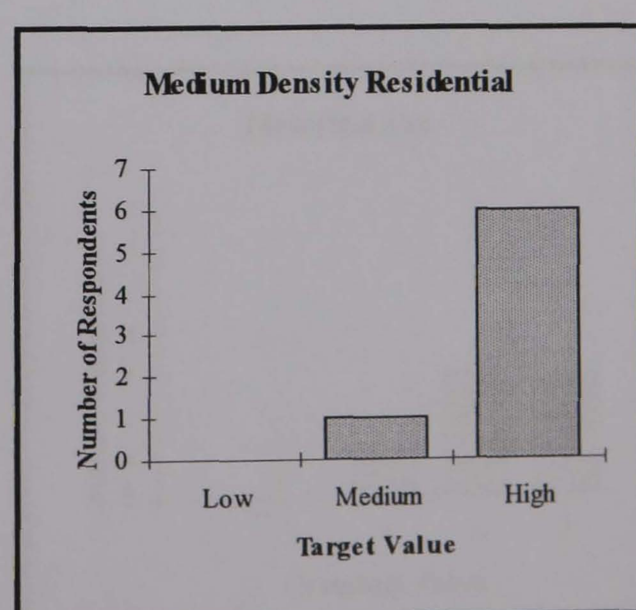




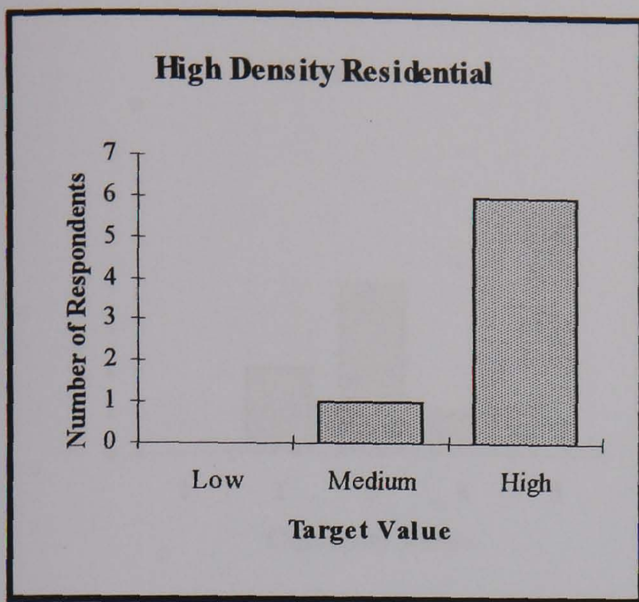


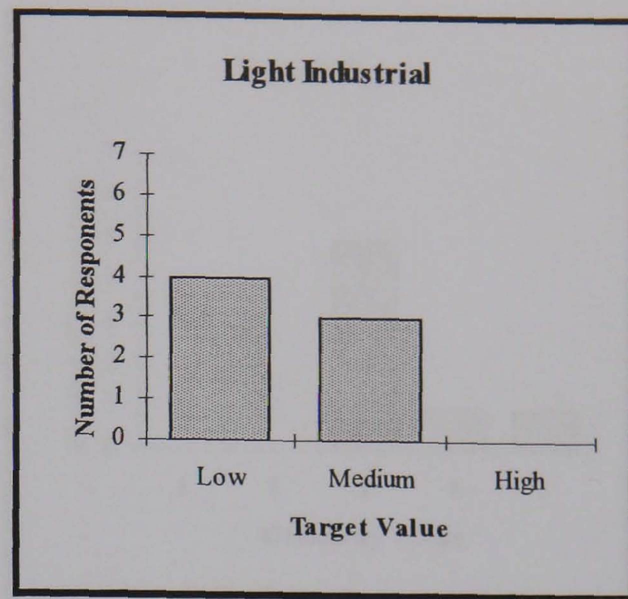


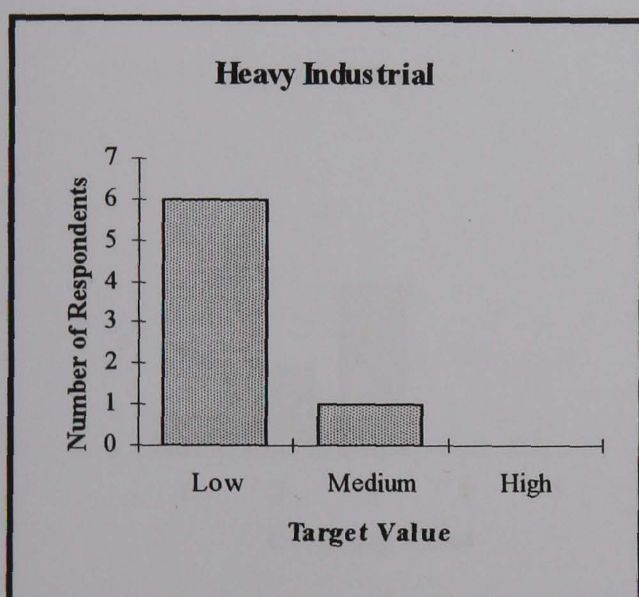


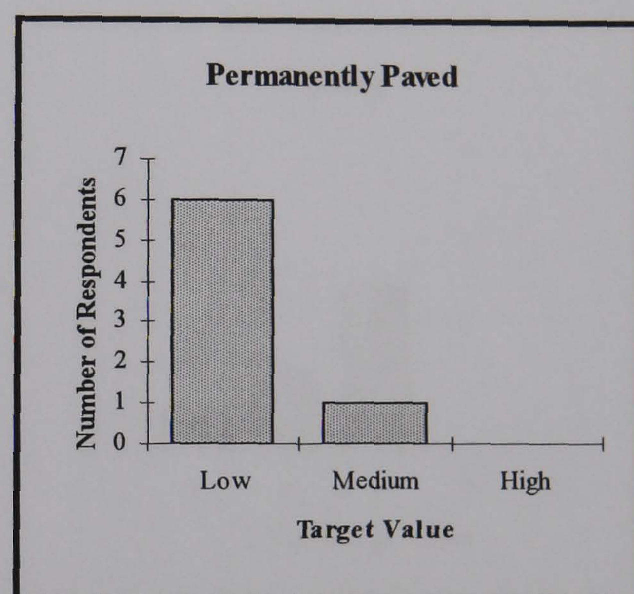


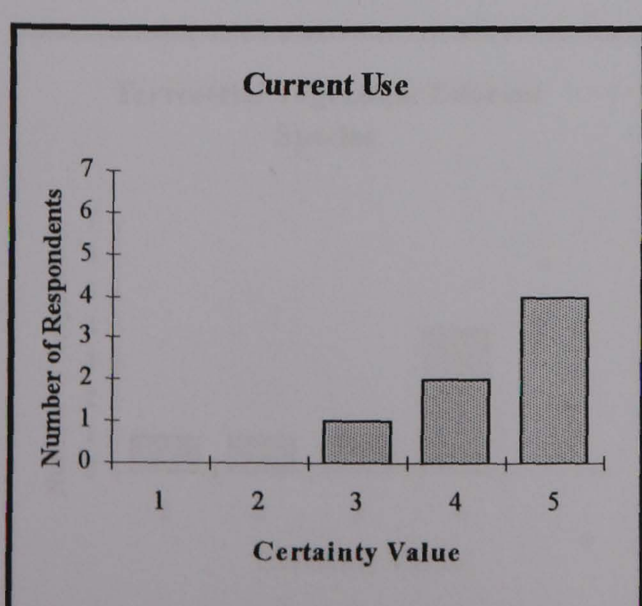
Questionnaire Results from the "Chemist" Classification

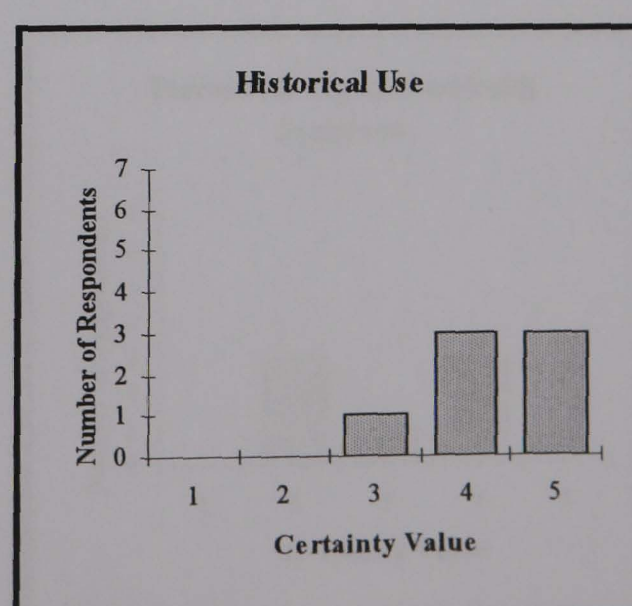




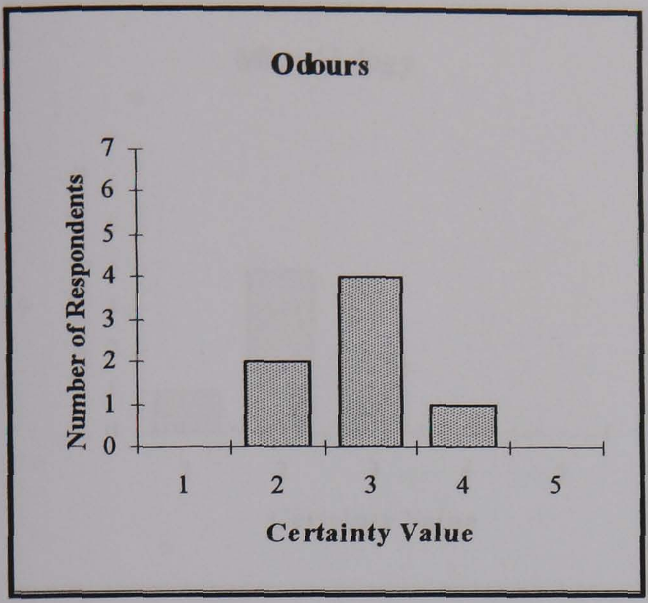


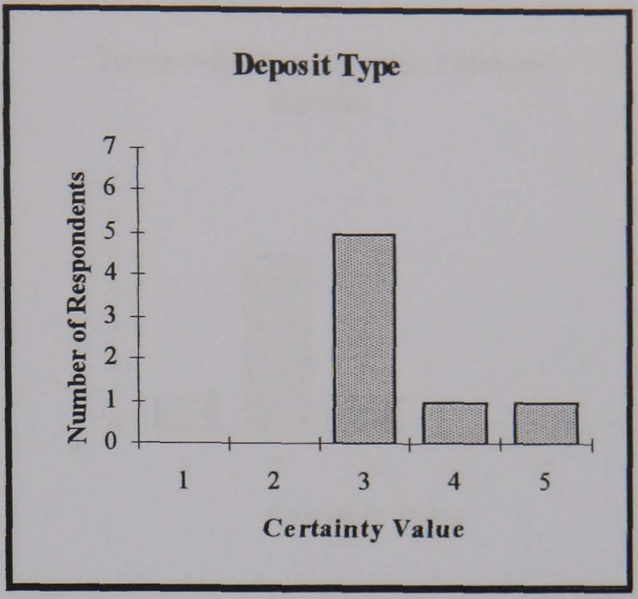


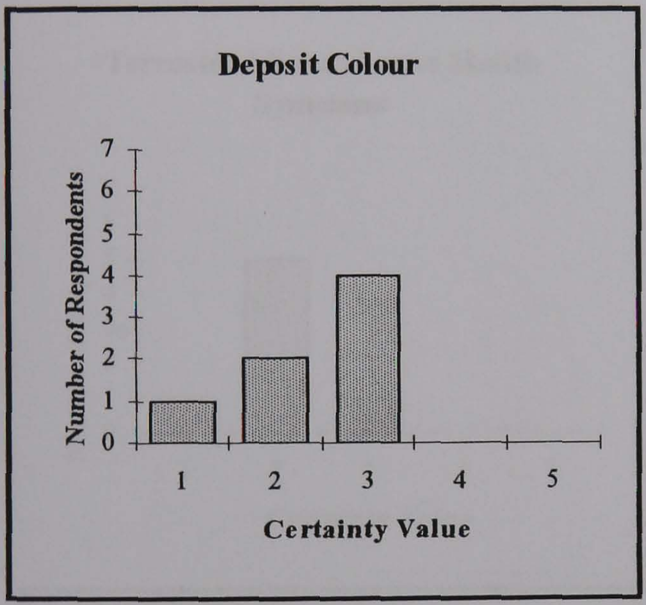


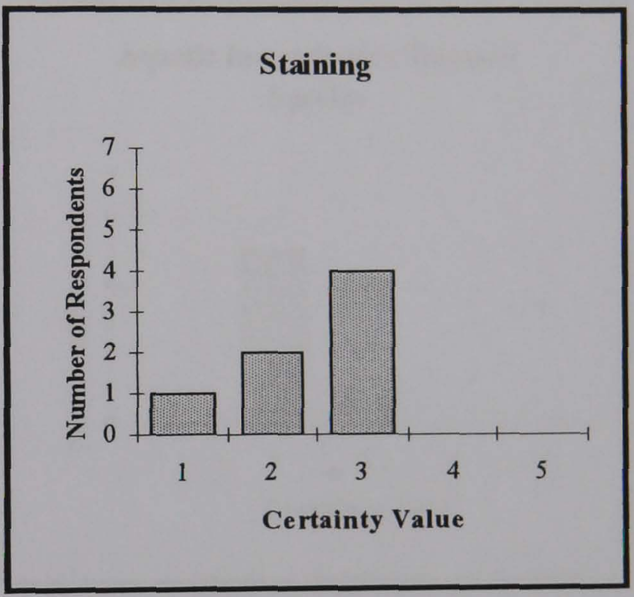


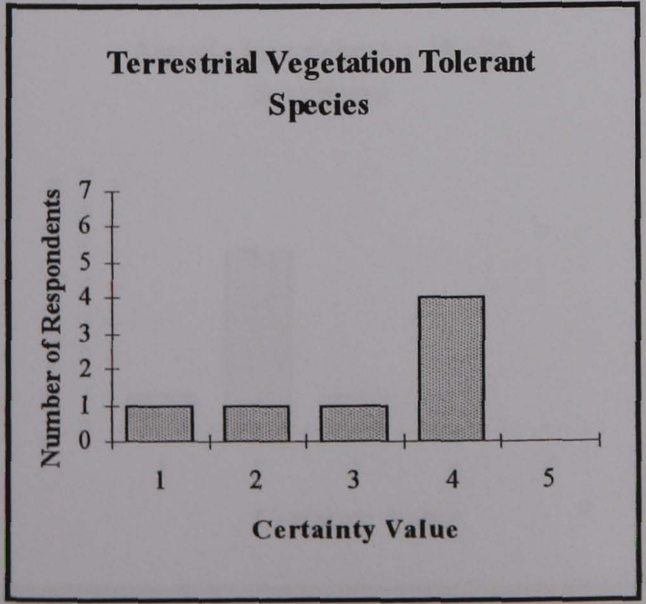
Questionnaire Results from the “Chemist” Classification

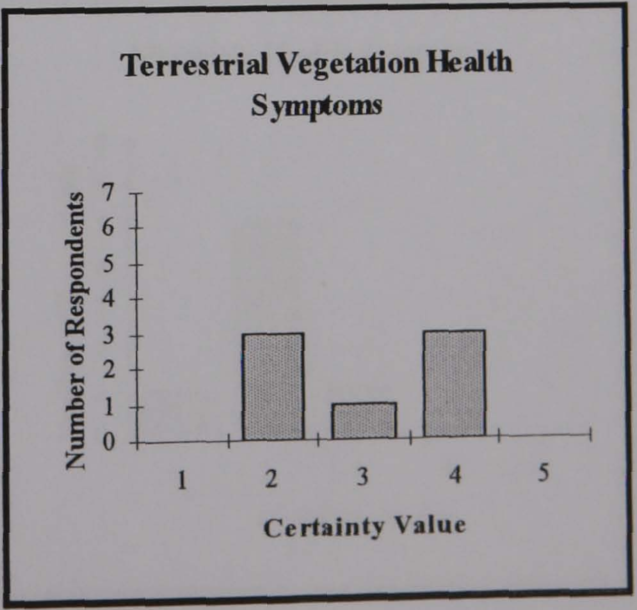




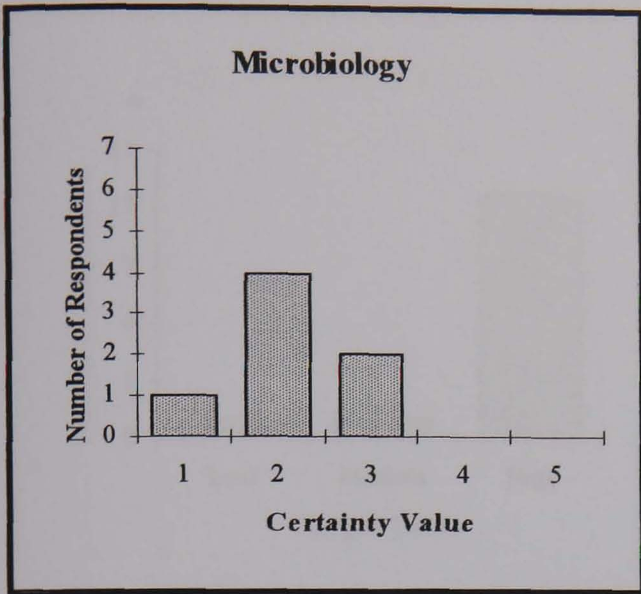


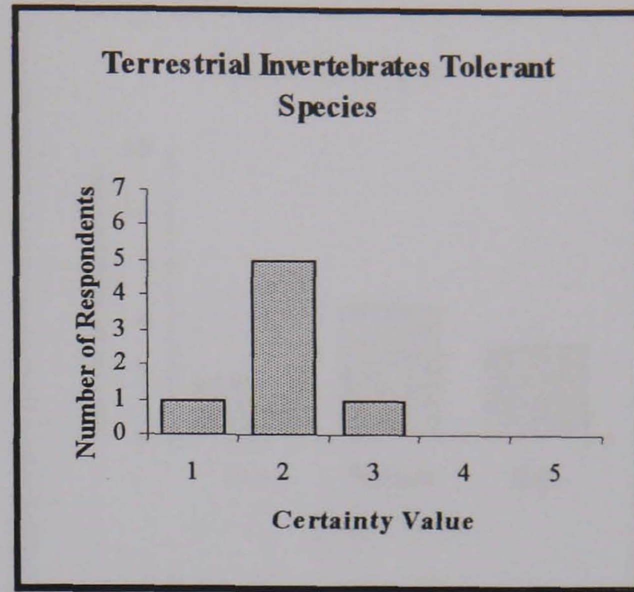


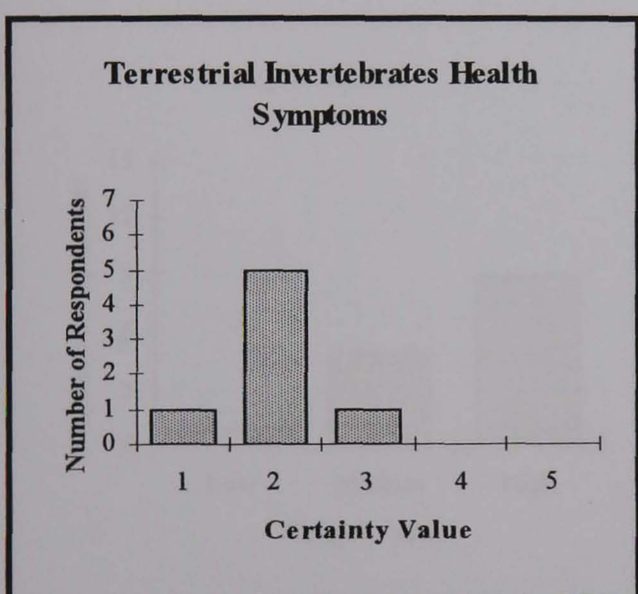


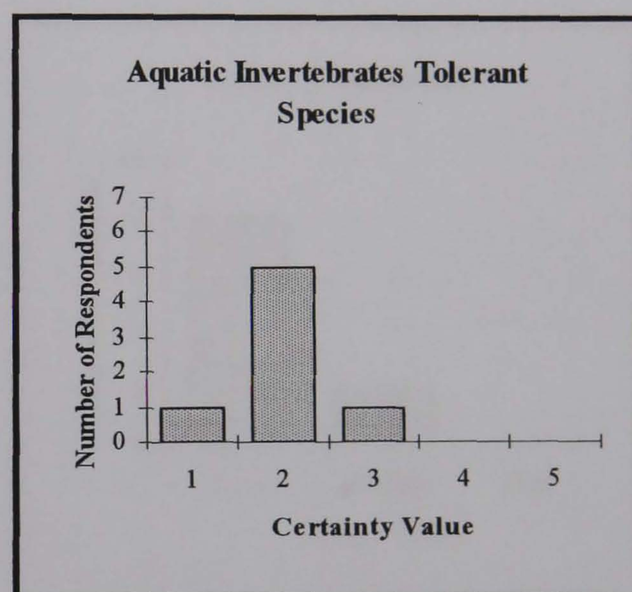


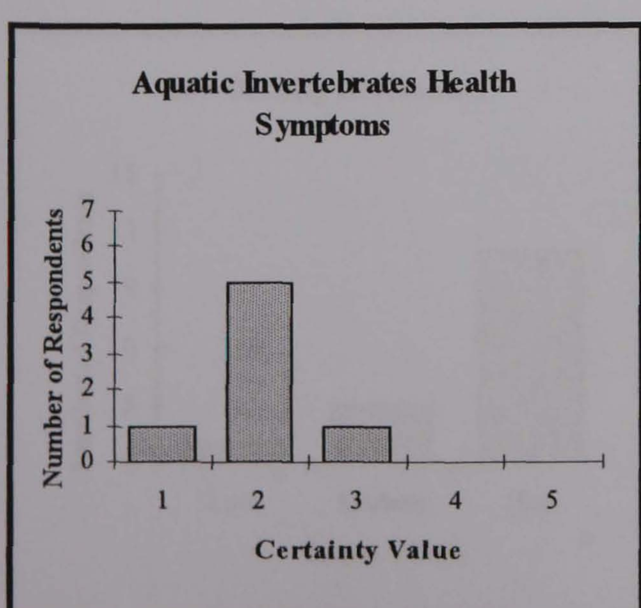
Questionnaire Results from the "Chemist" Classification

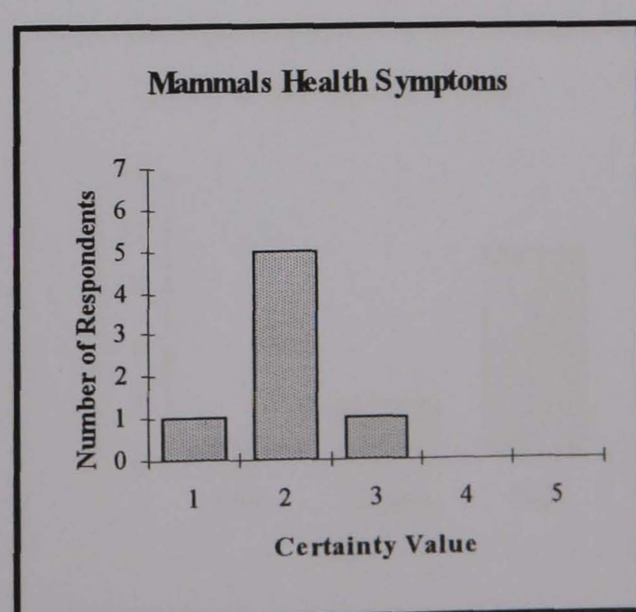




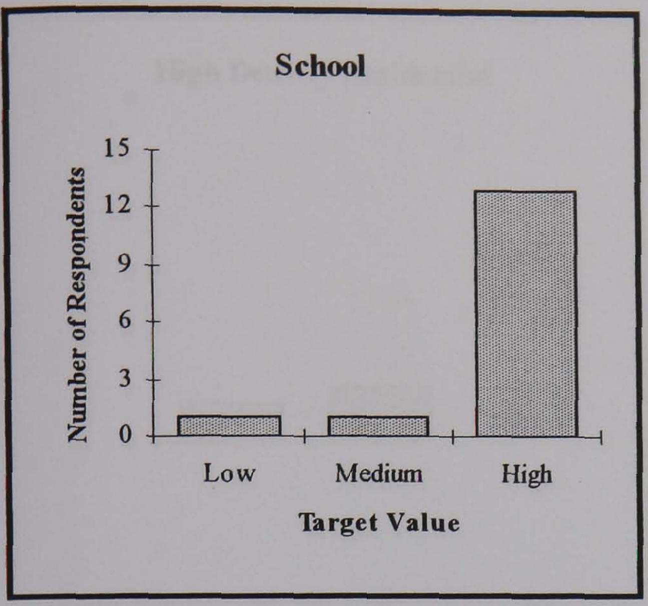


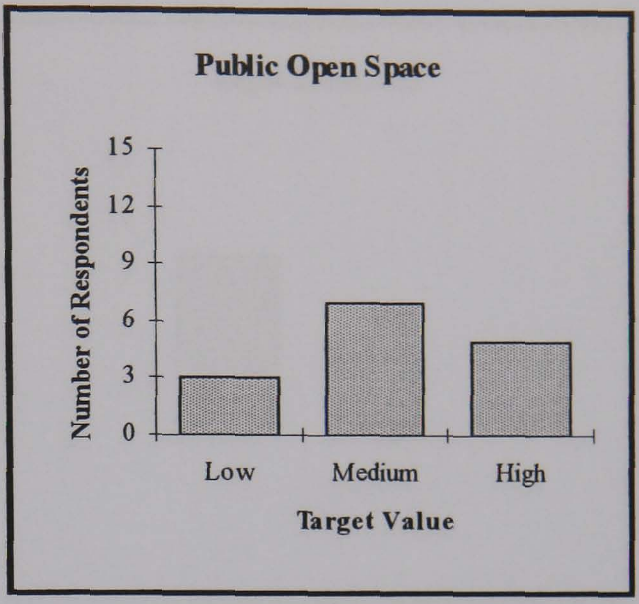


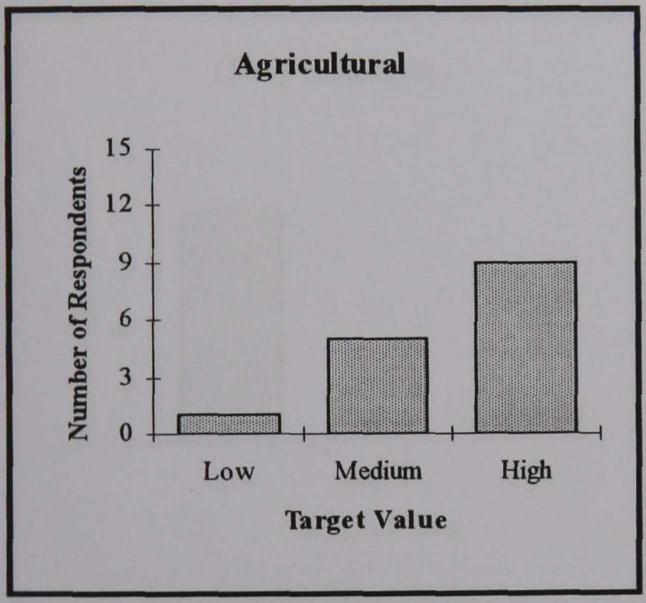


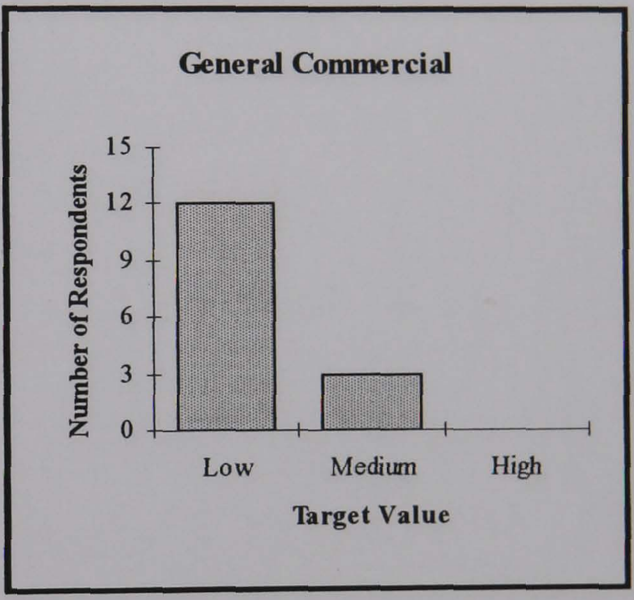


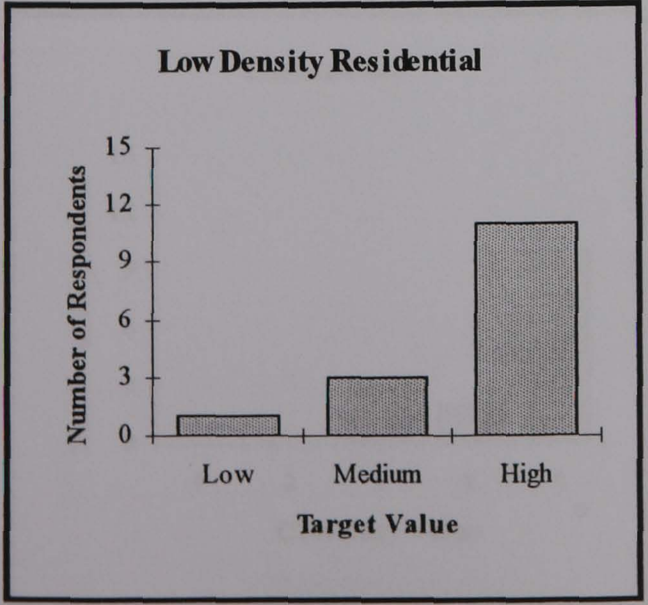
Questionnaire Results from the "Geo-Environmental Engineer" Classification

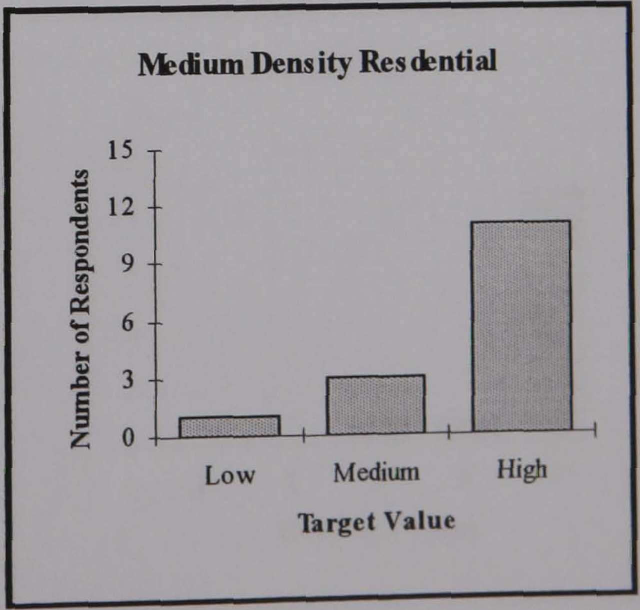




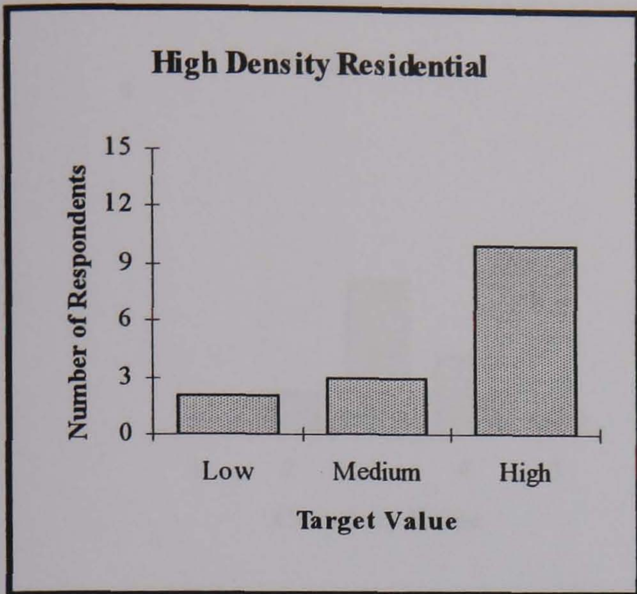


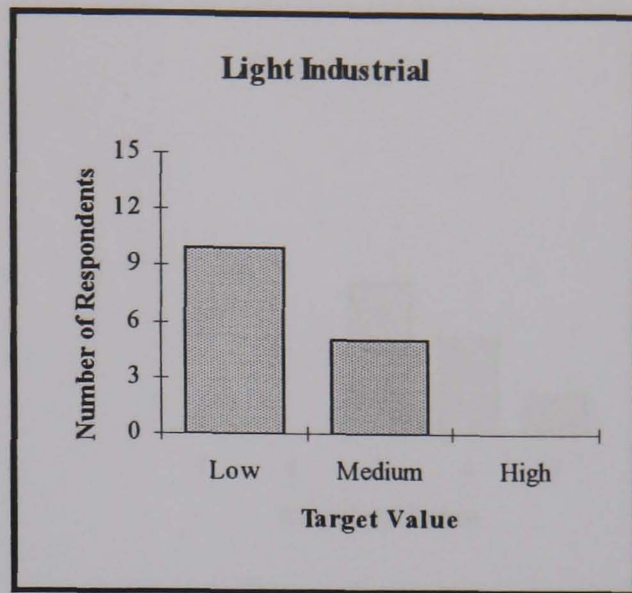


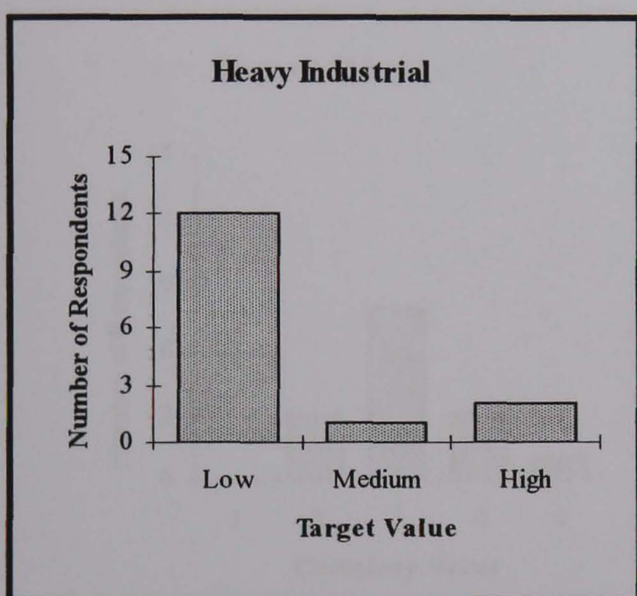


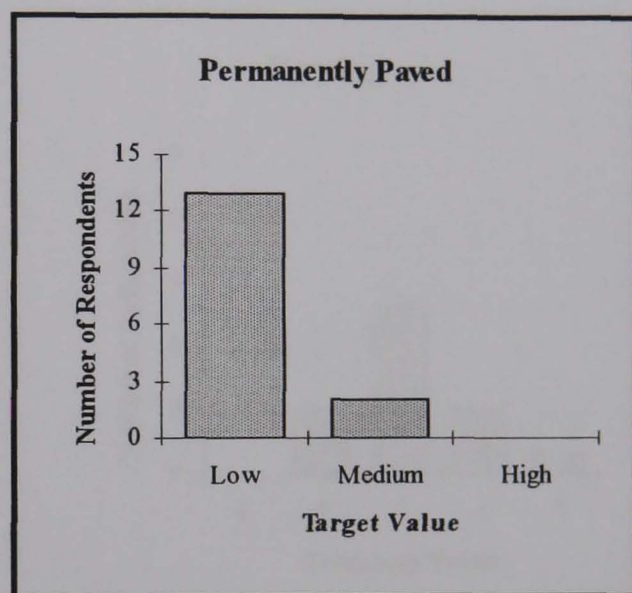


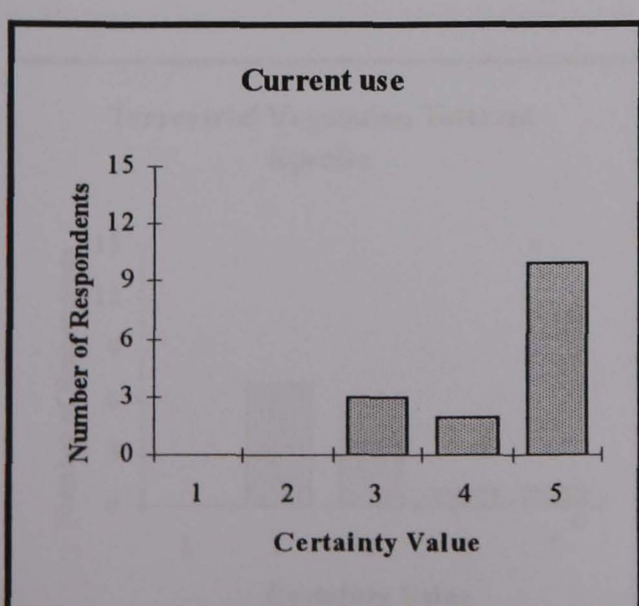
Questionnaire Results from the "Geo-Environmental Engineer" Classification

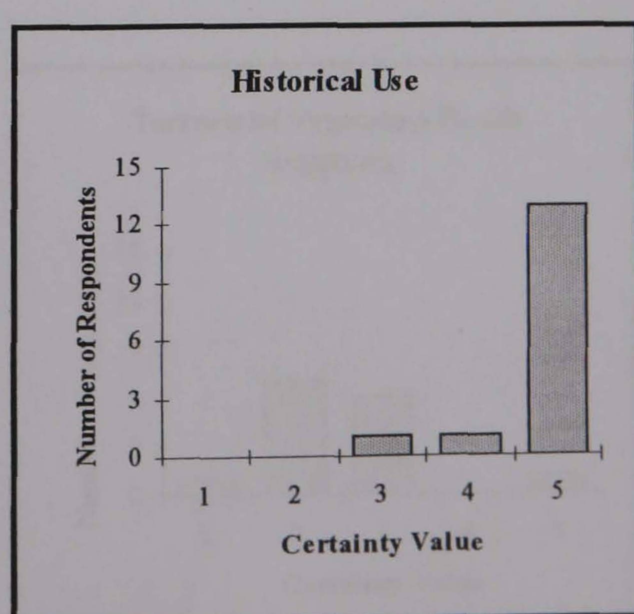




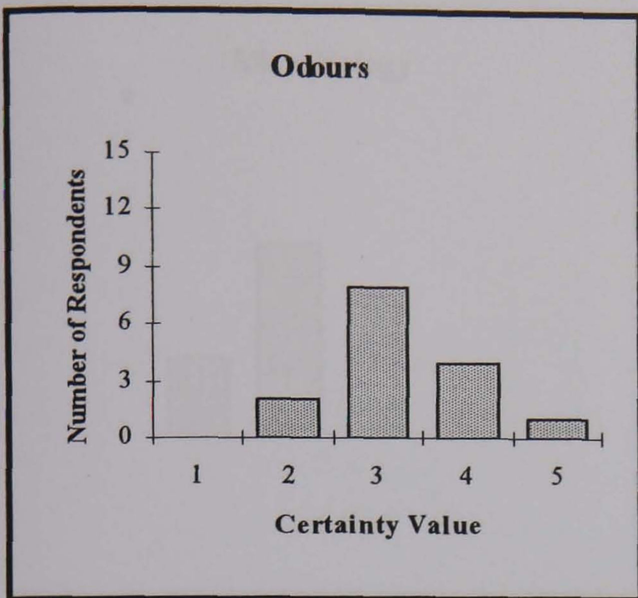


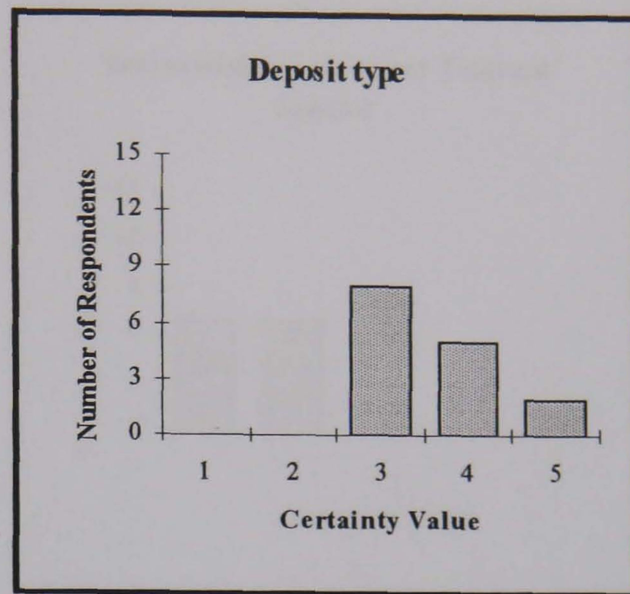


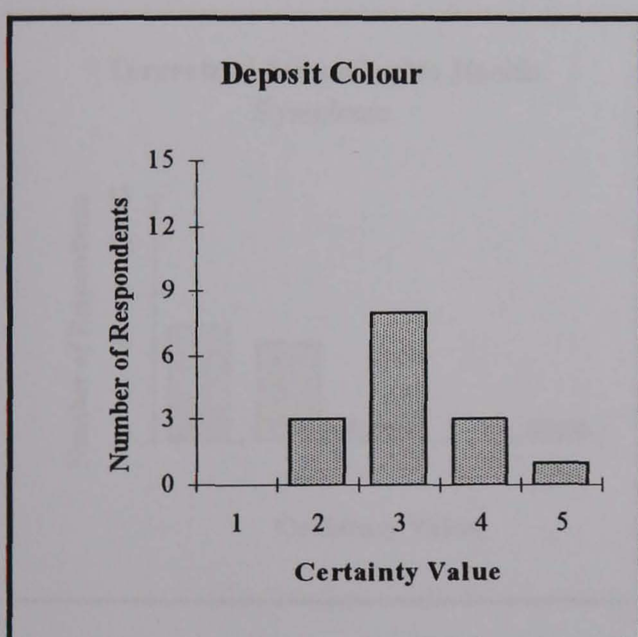


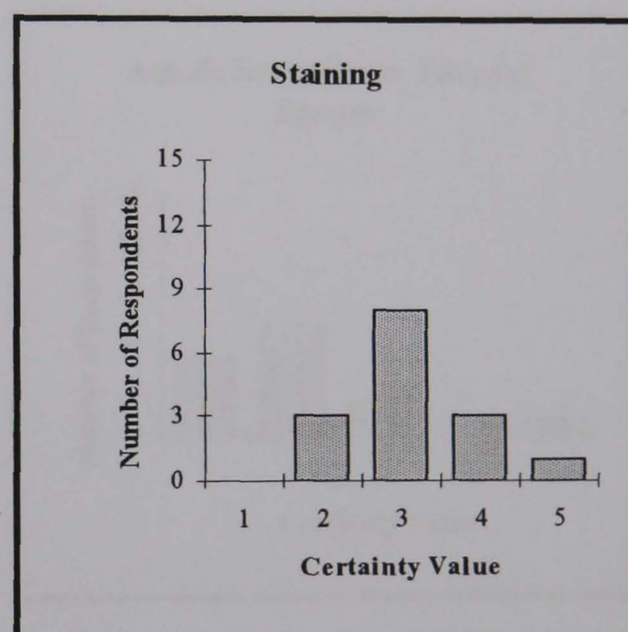


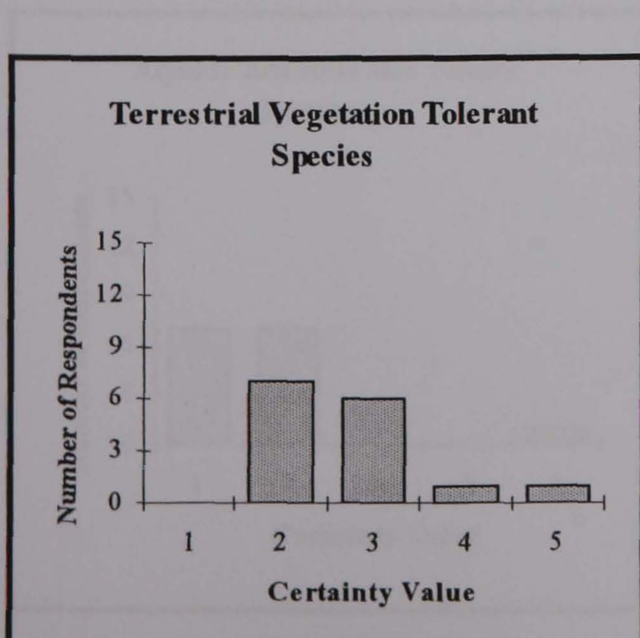
Questionnaire Results from the "Geo-Environmental Engineer" Classification

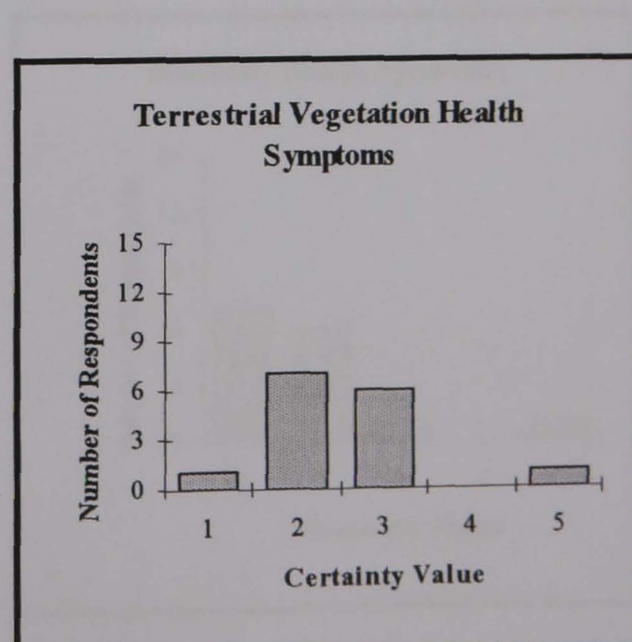




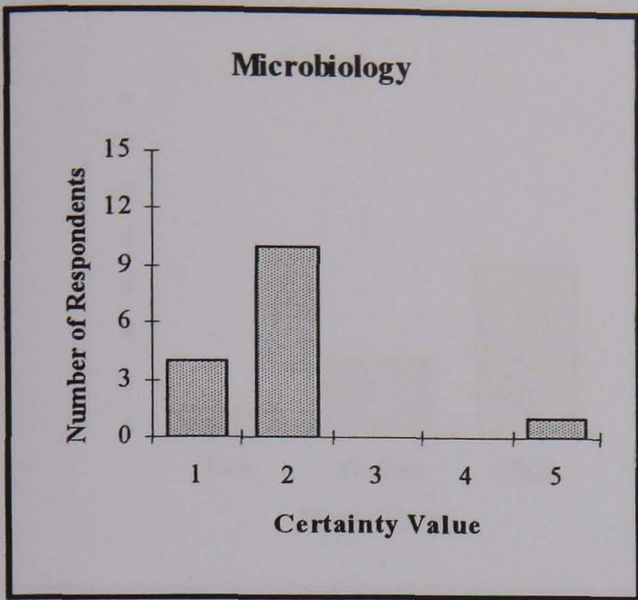


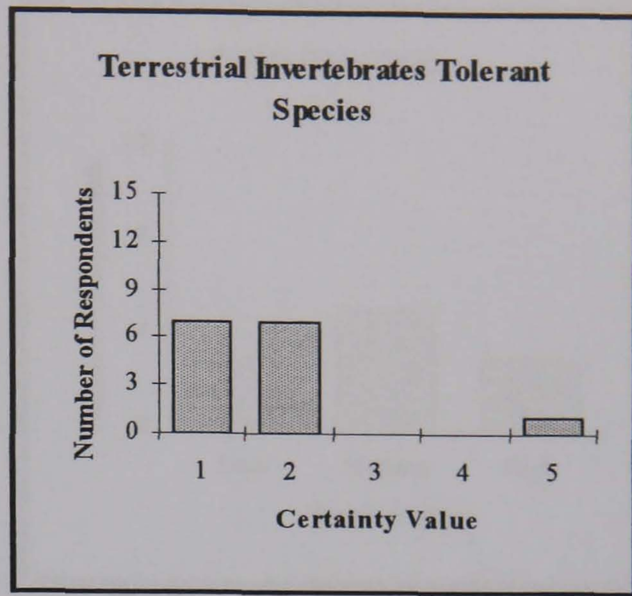


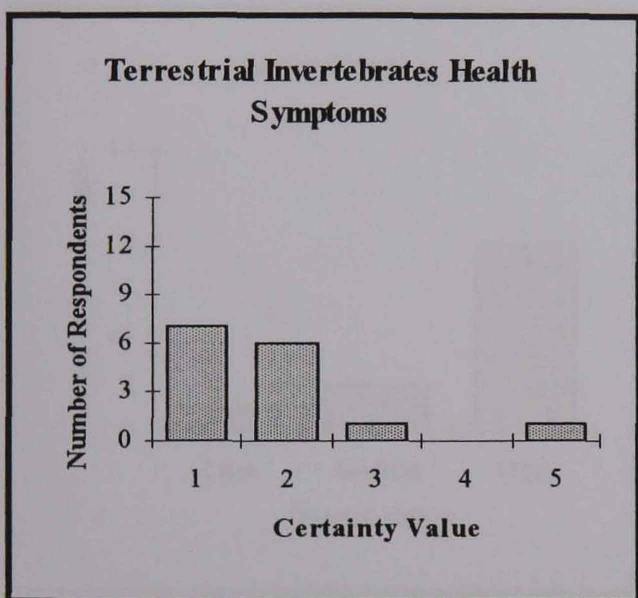


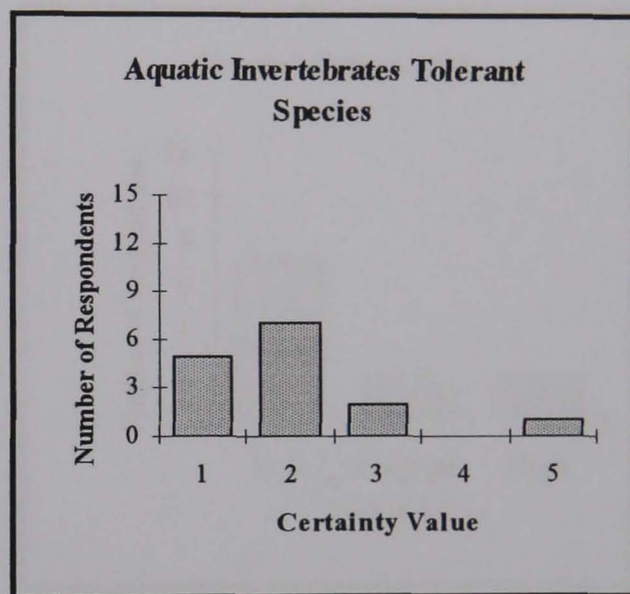


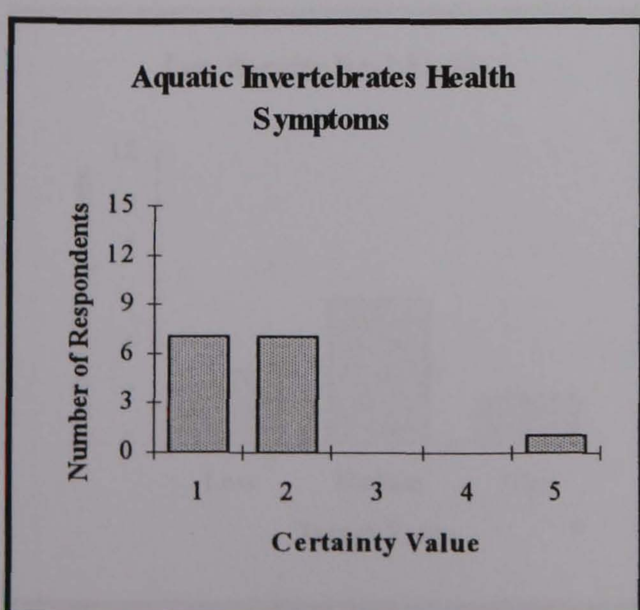
Questionnaire Results from the "Geo-Environmental Engineer" Classification

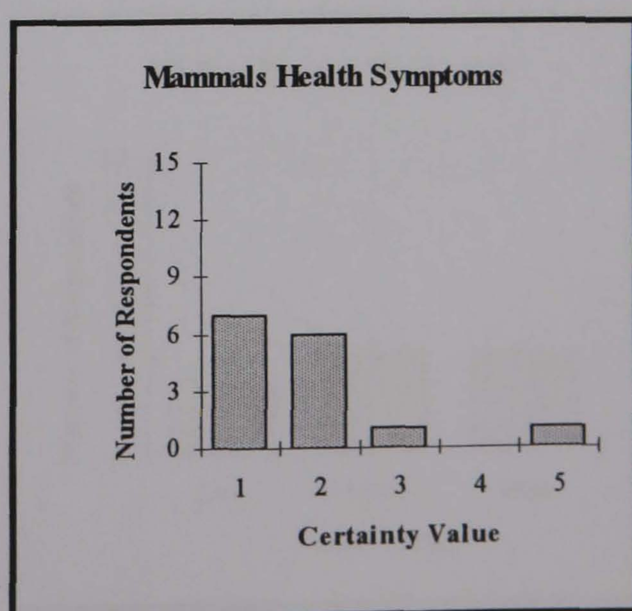




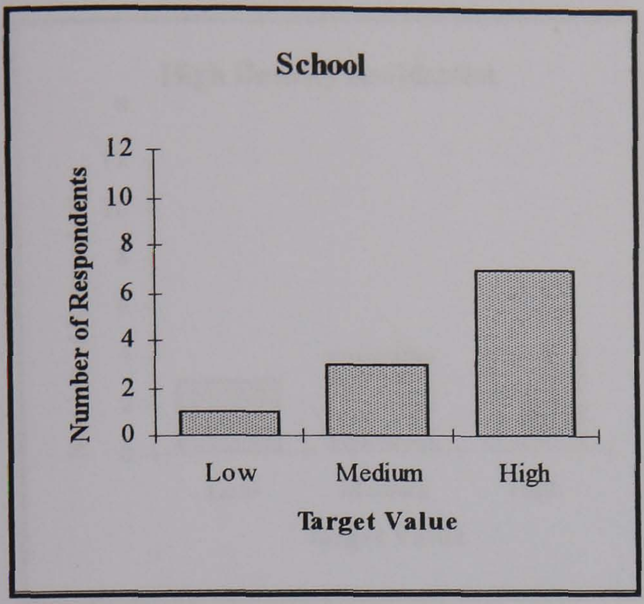


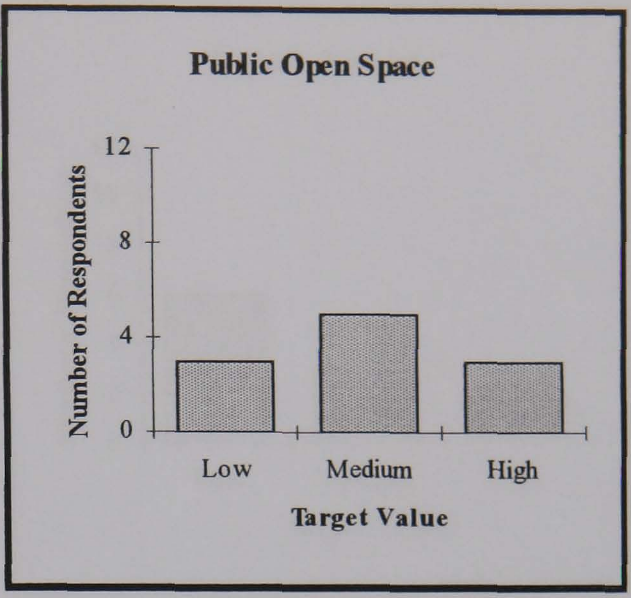


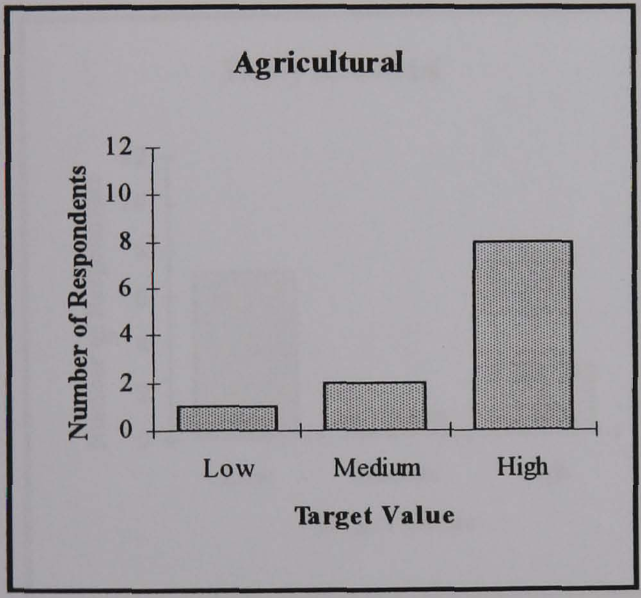


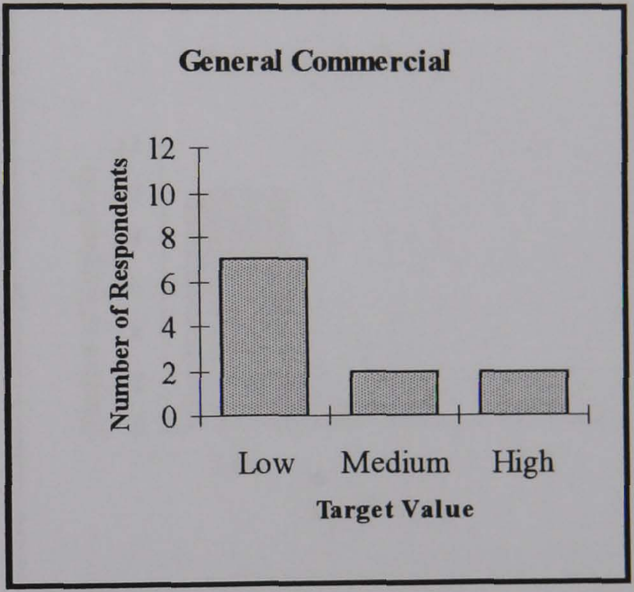


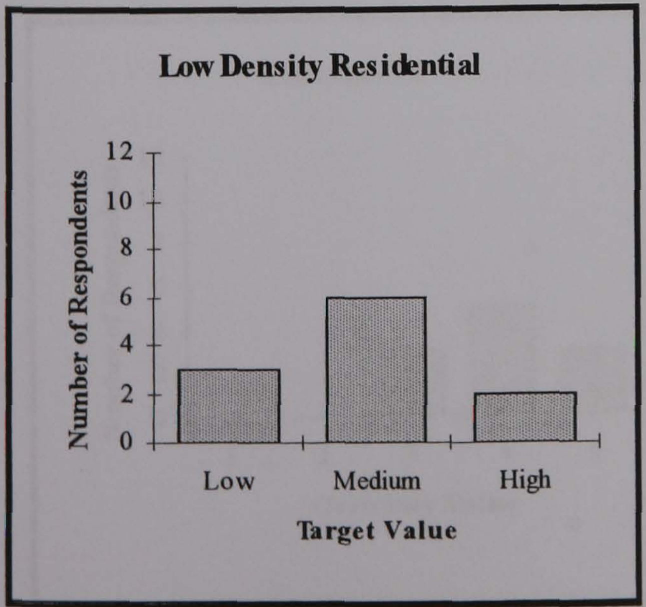
Questionnaire Results from the "Geologist" Classification

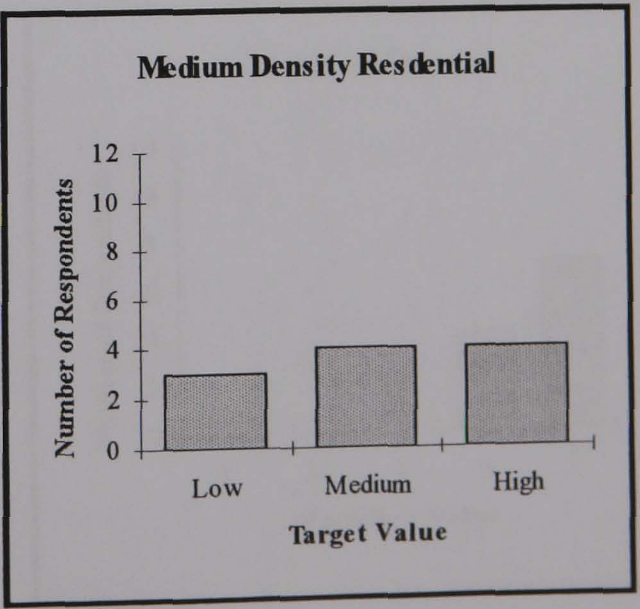




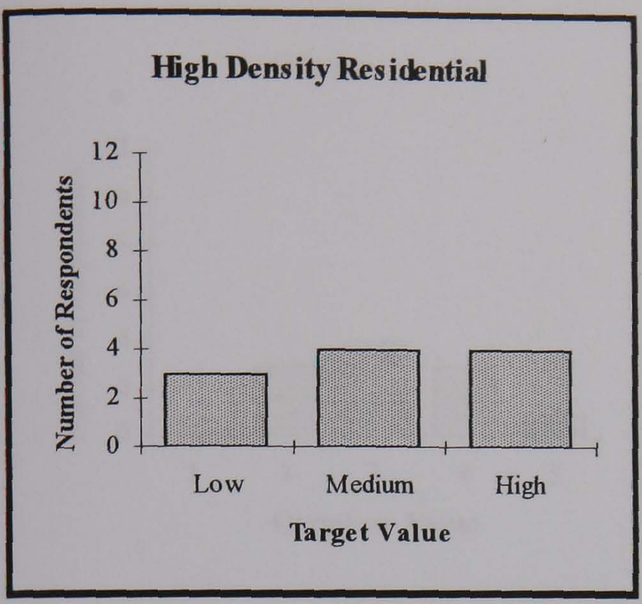


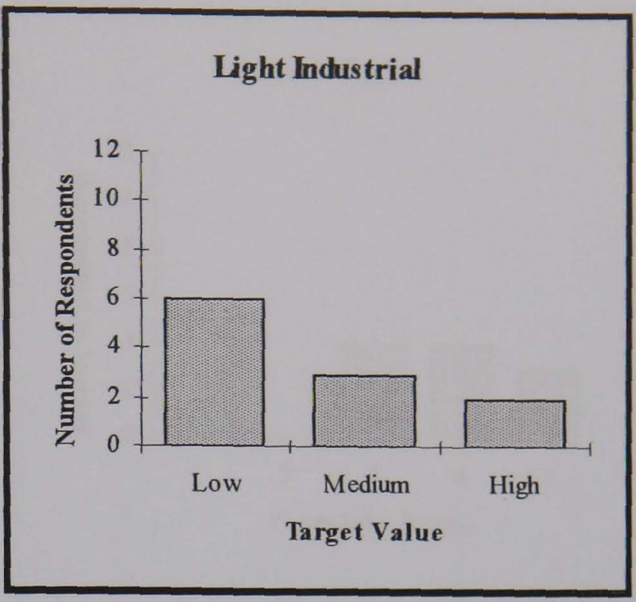


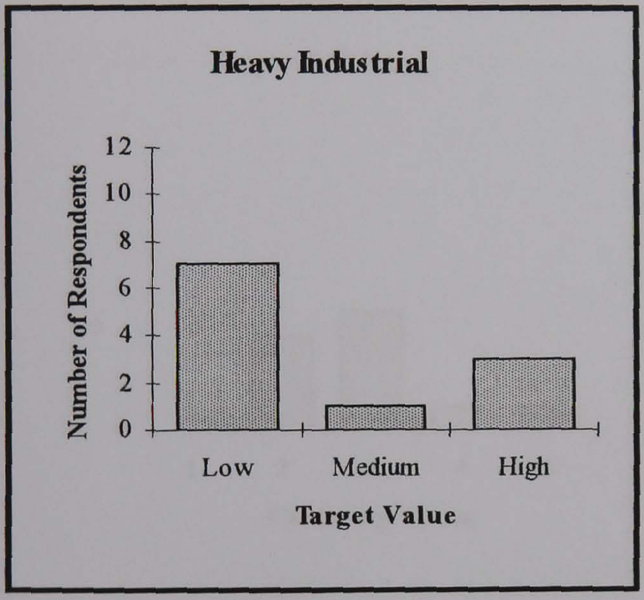


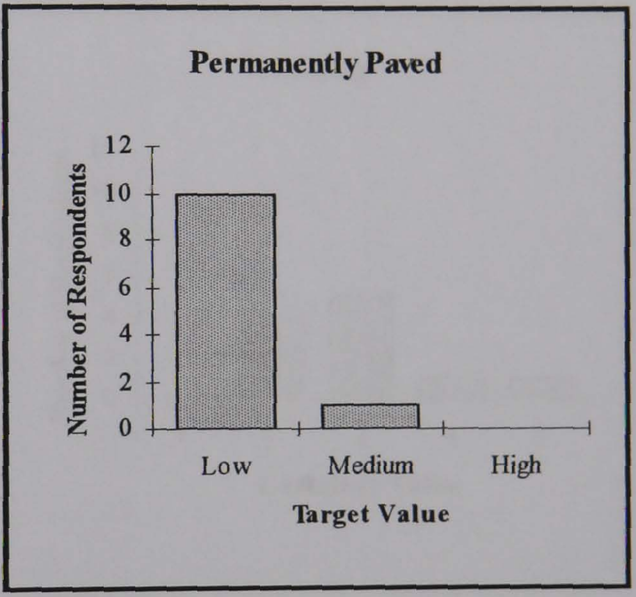


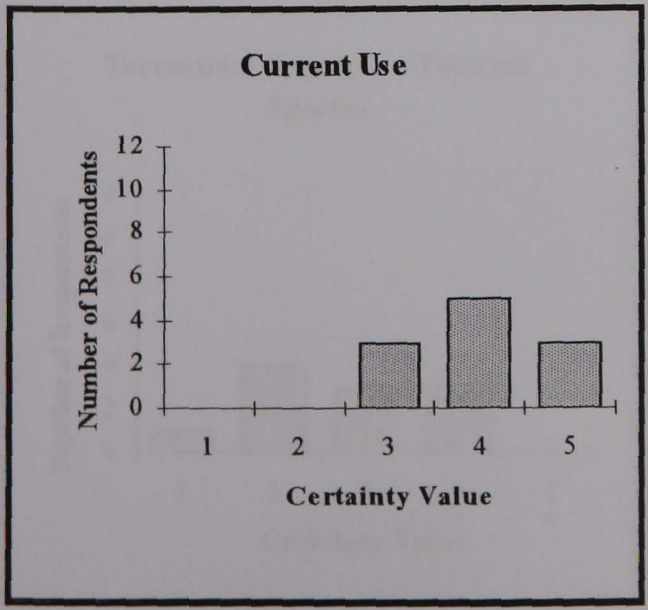
Questionnaire Results from the "Geologist" Classification

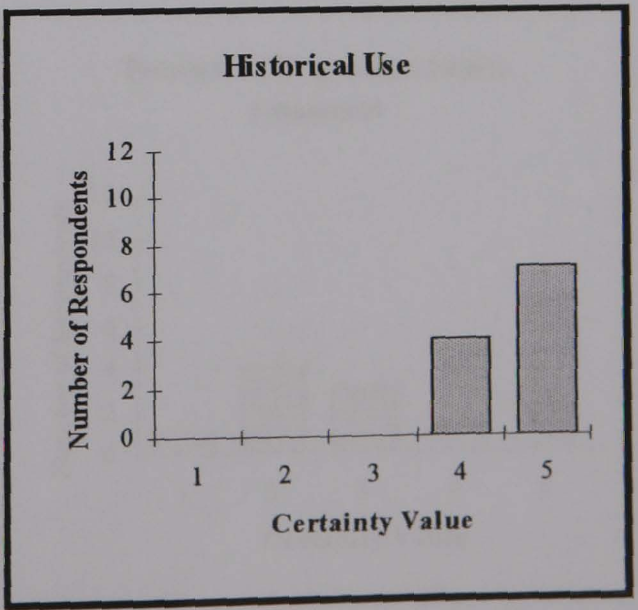




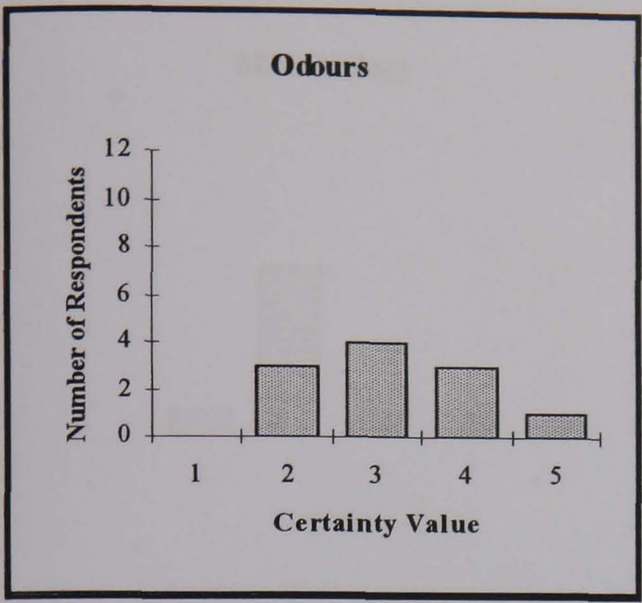


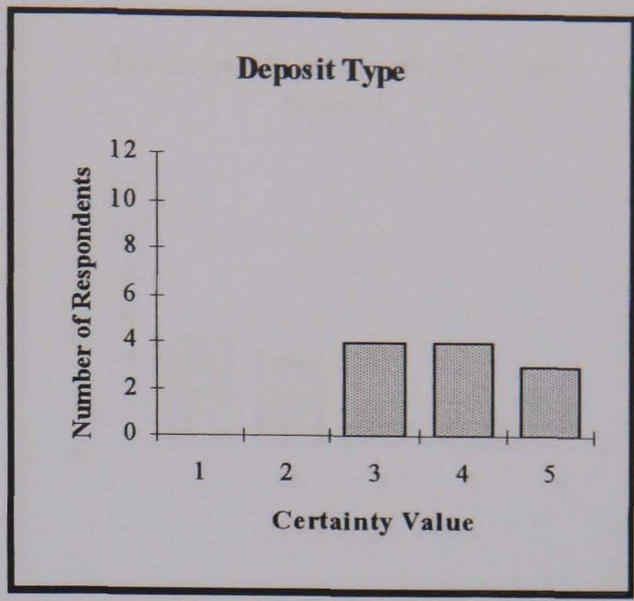


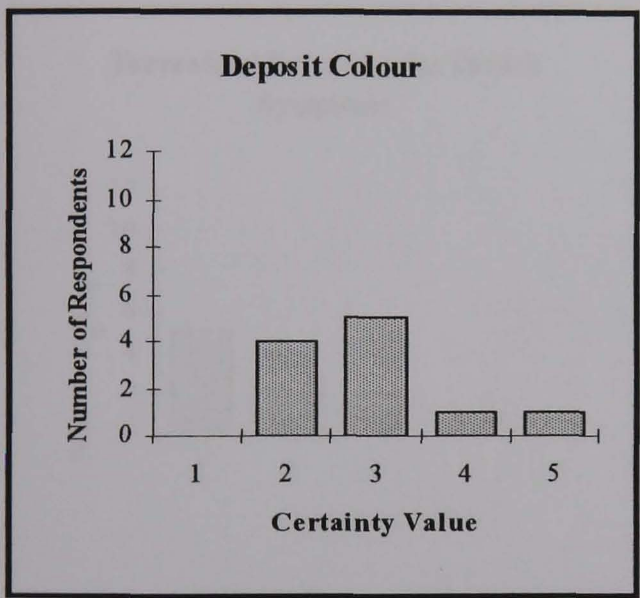


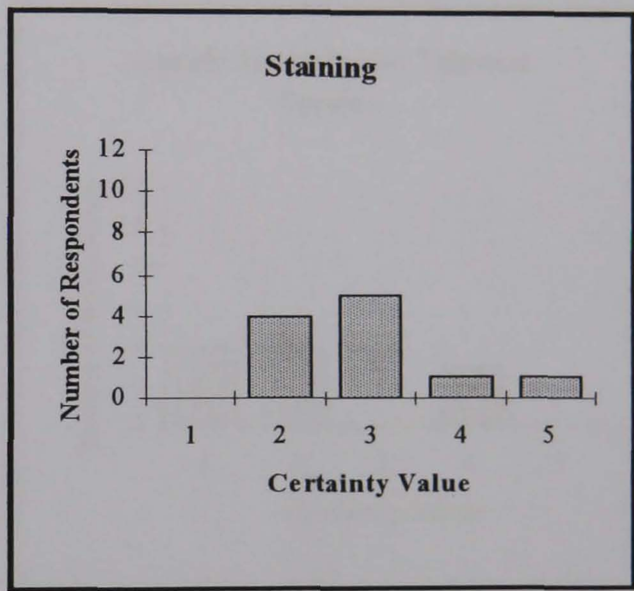


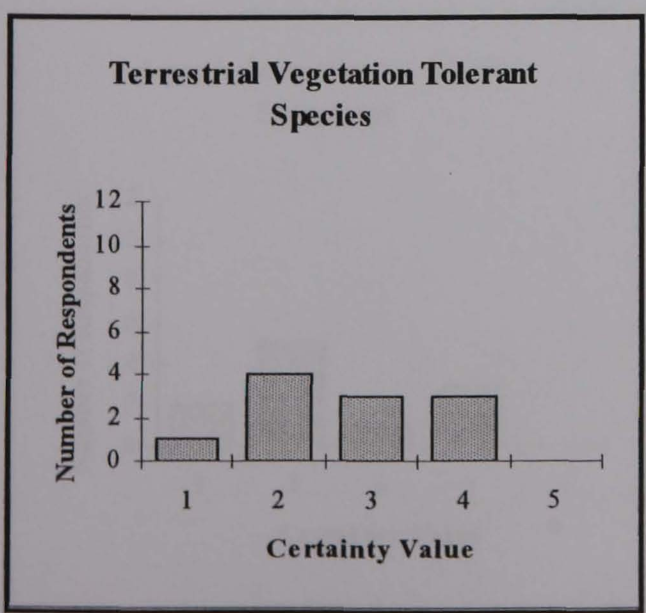
Questionnaire Results from the "Geologist" Classification

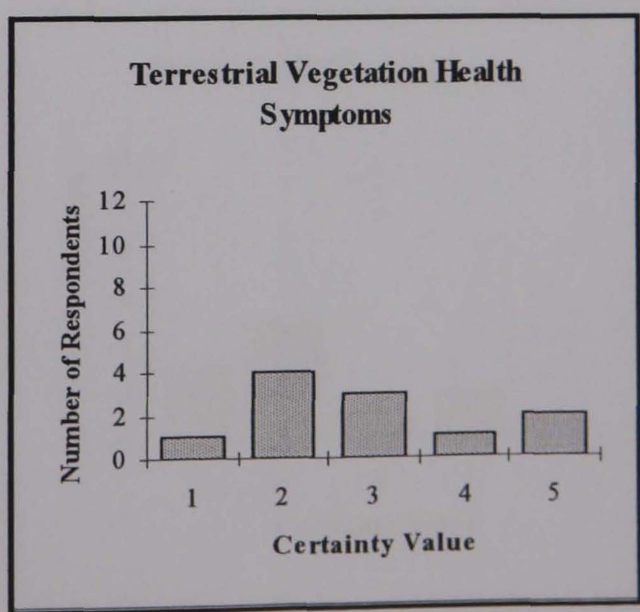




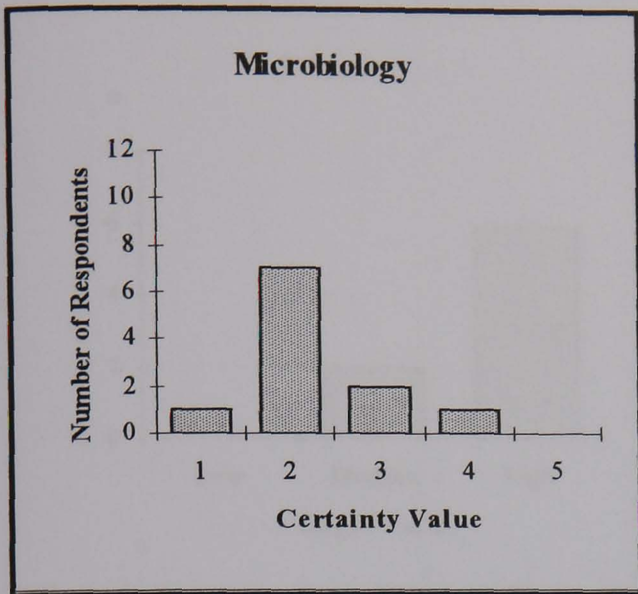


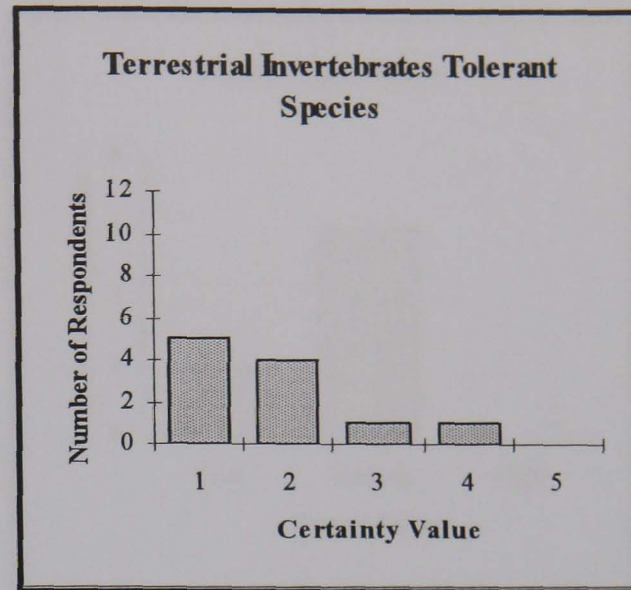


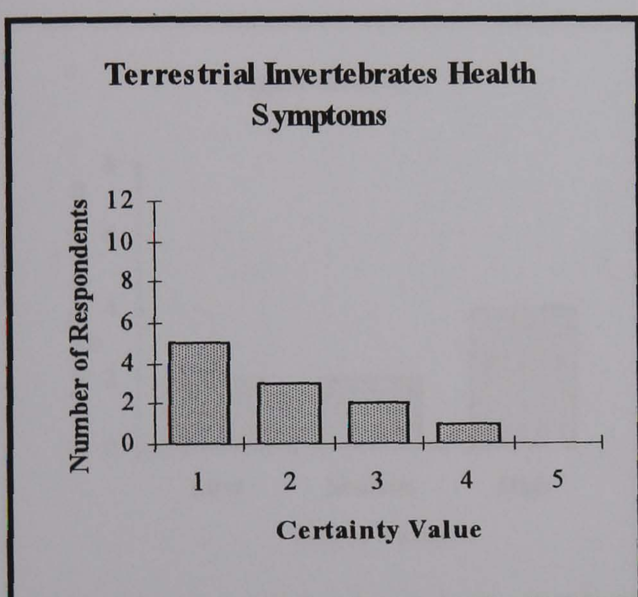


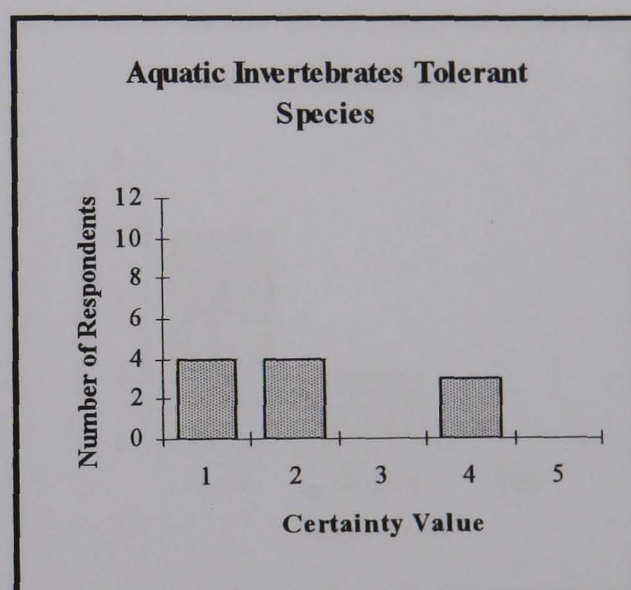


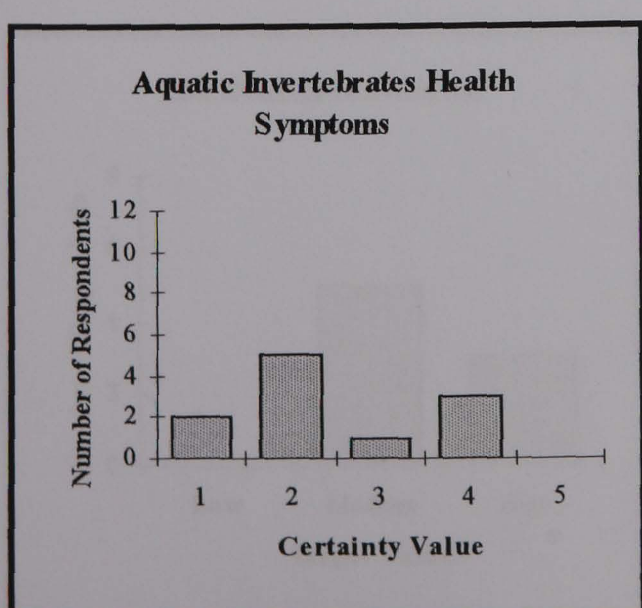
Questionnaire Results from the "Geologist" Classification

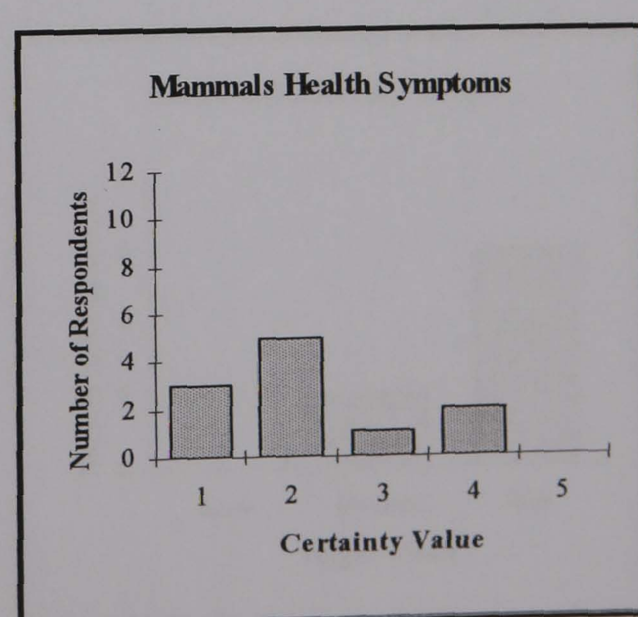




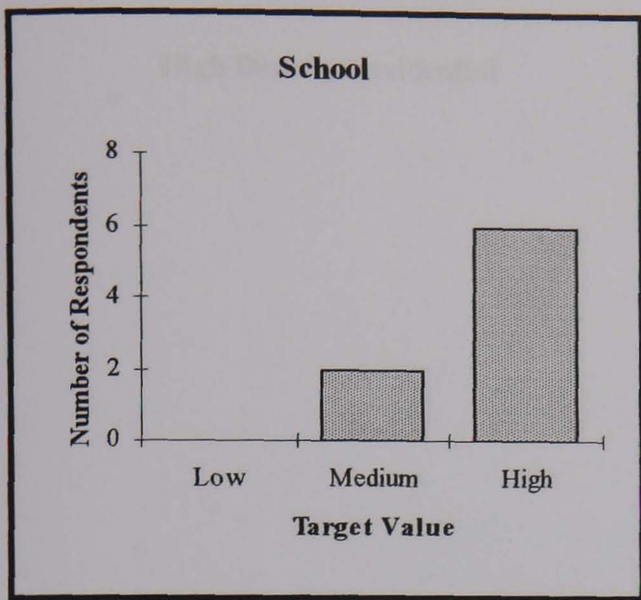


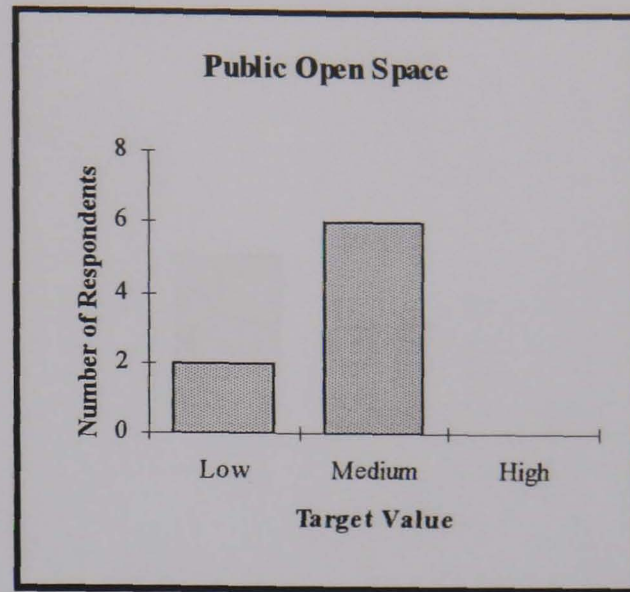


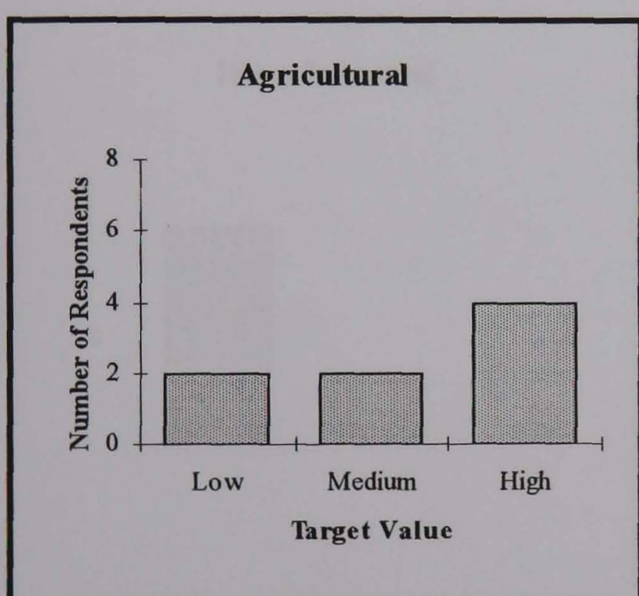


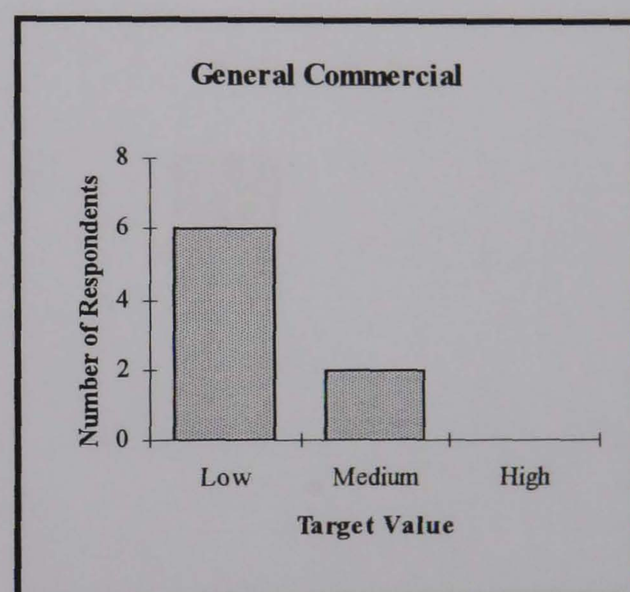


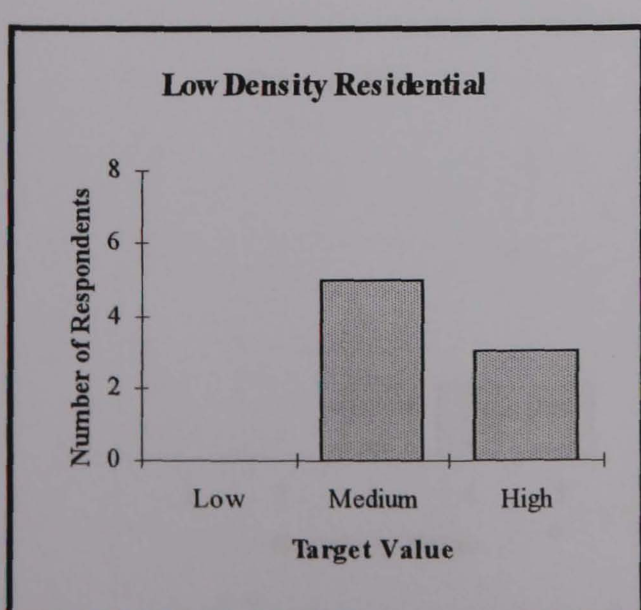
Questionnaire Results from the "Hydrogeologist" Classification

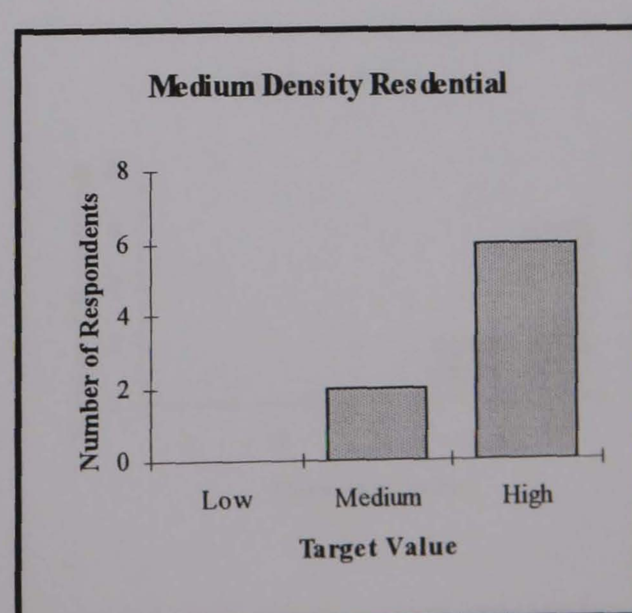




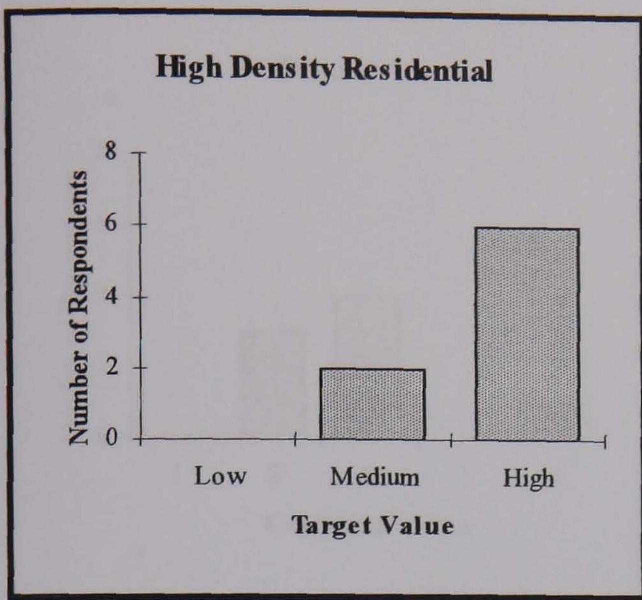


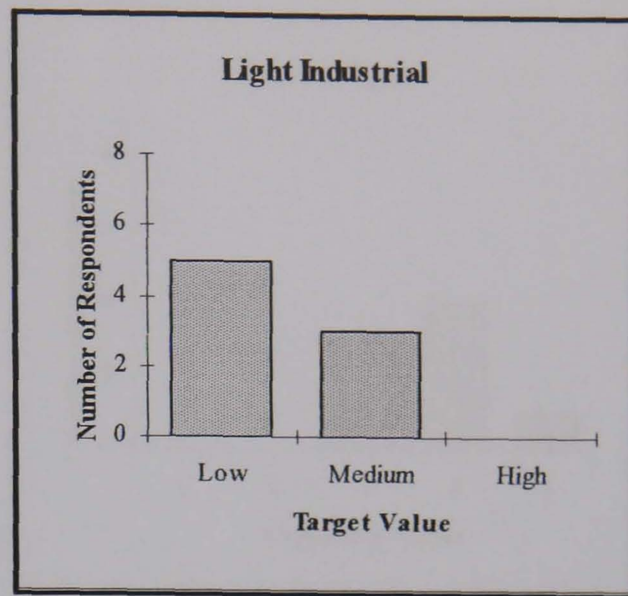


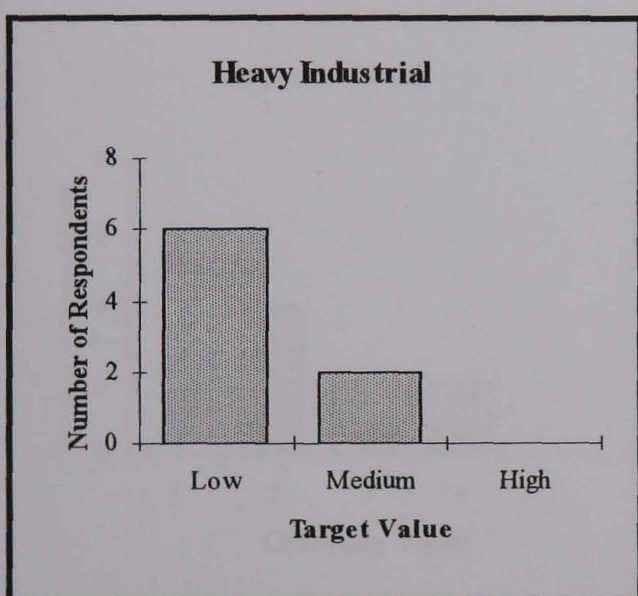


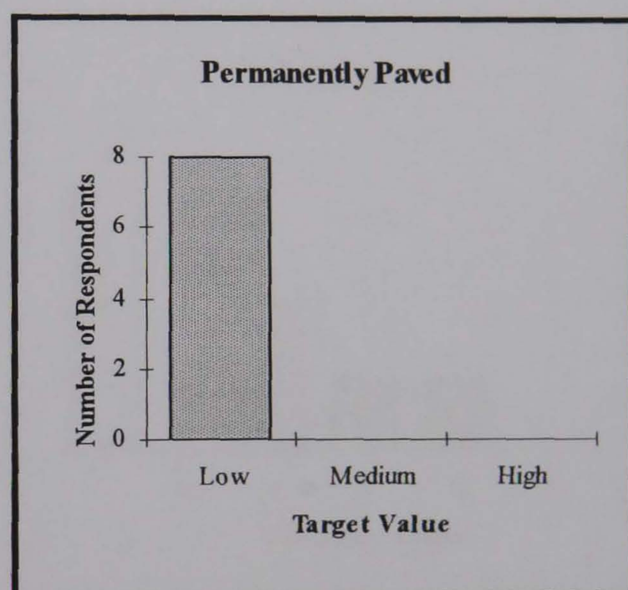


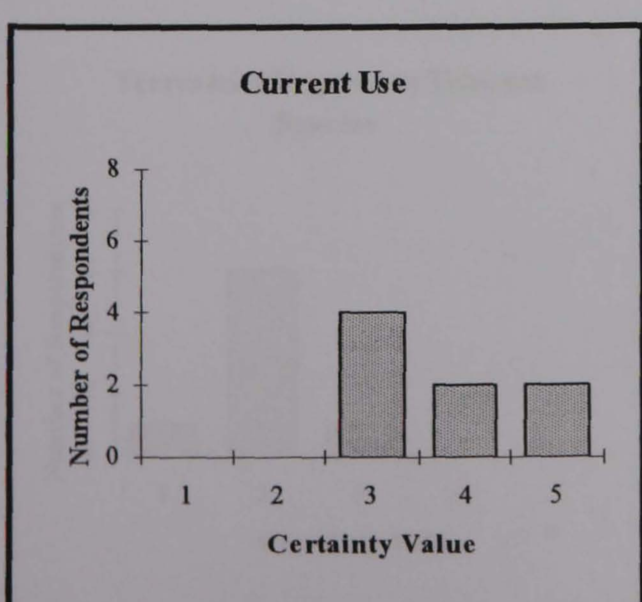
Questionnaire Results from the "Hydrogeologist" Classification

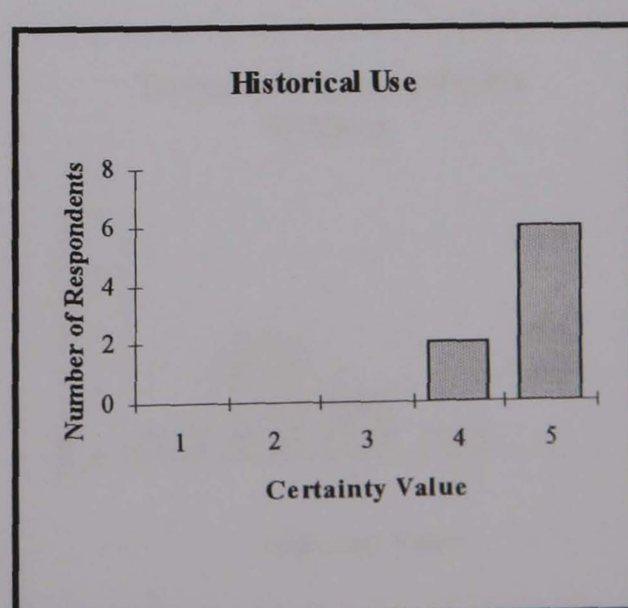




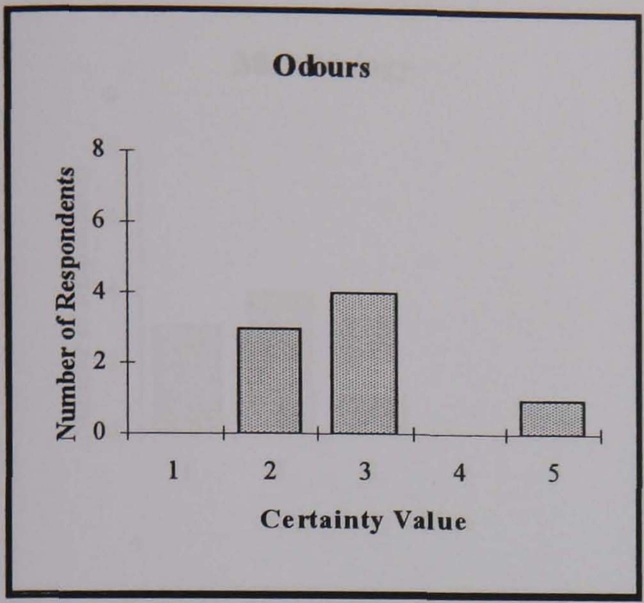


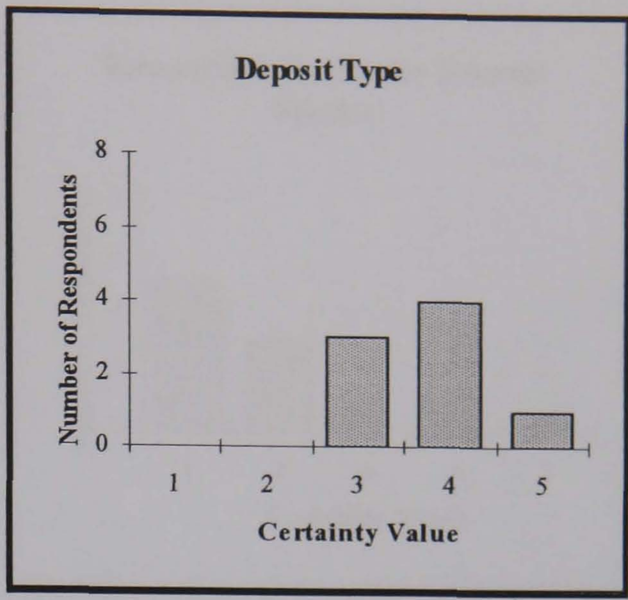


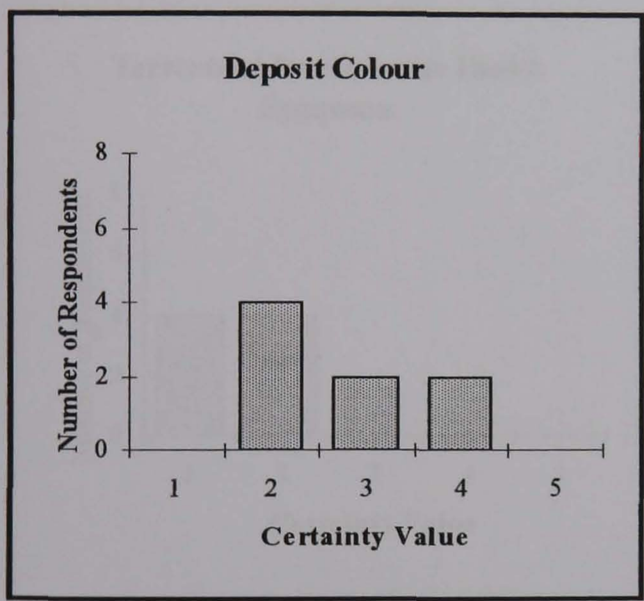


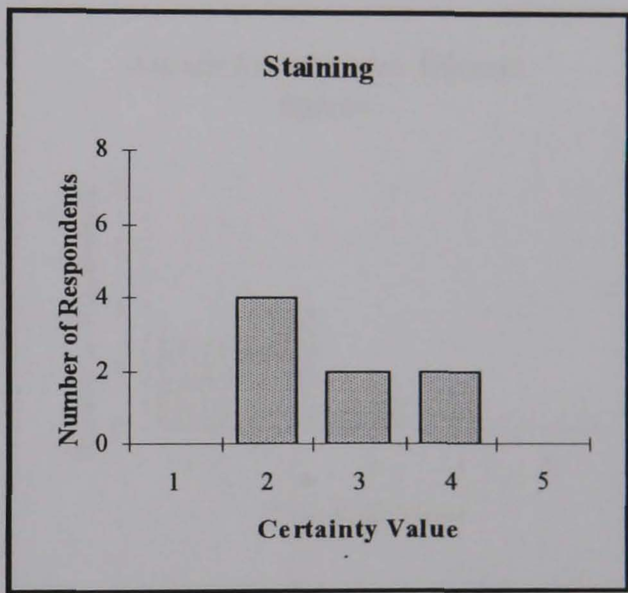


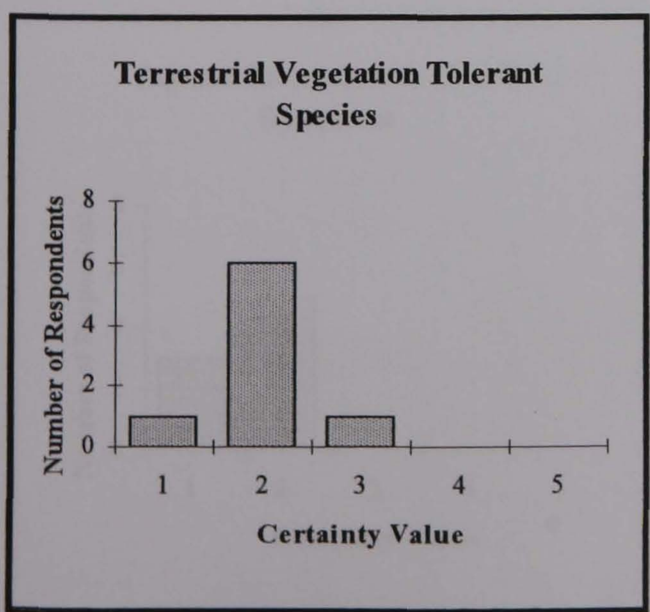
Questionnaire Results from the "Hydrogeologist" Classification

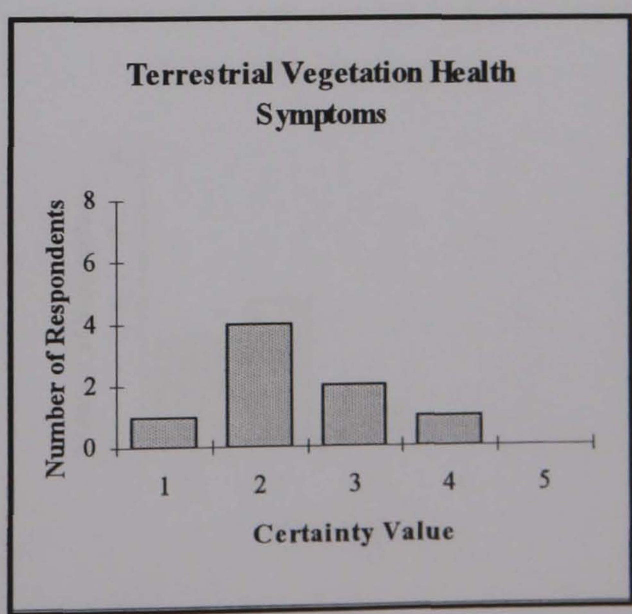




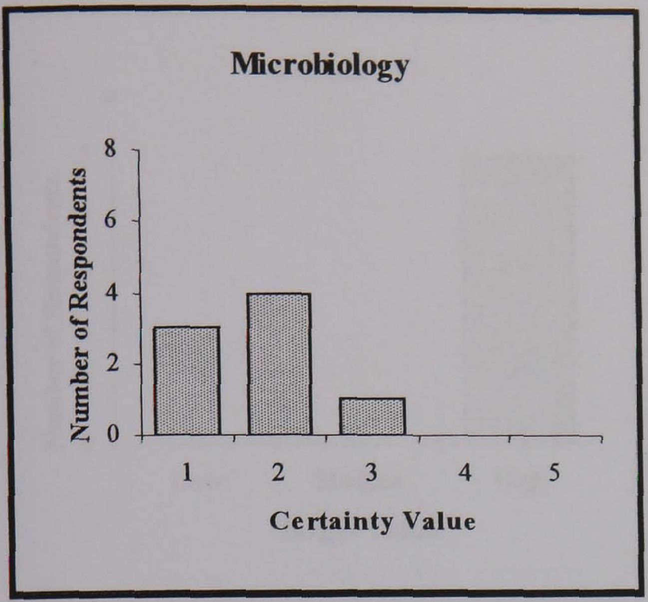


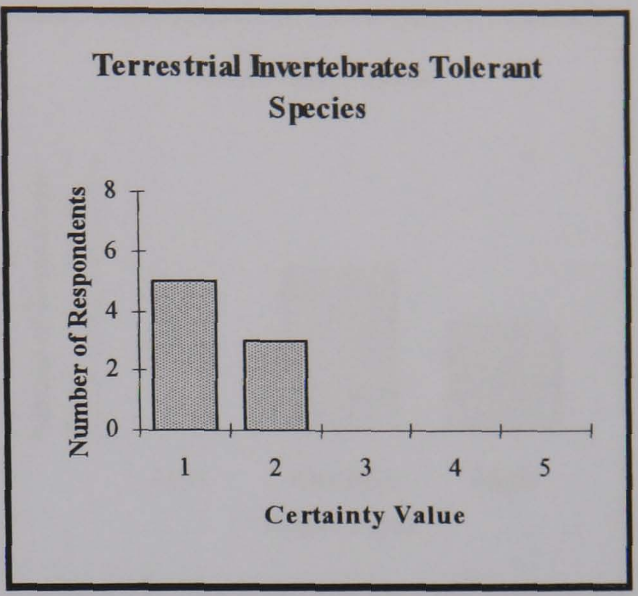


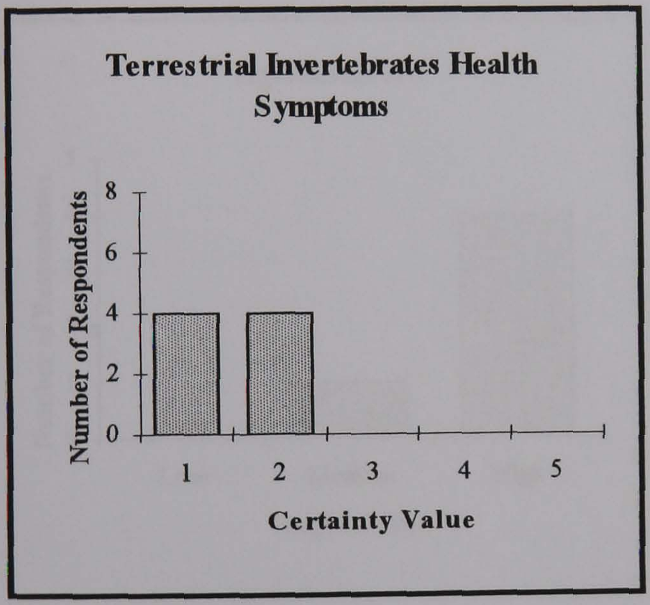


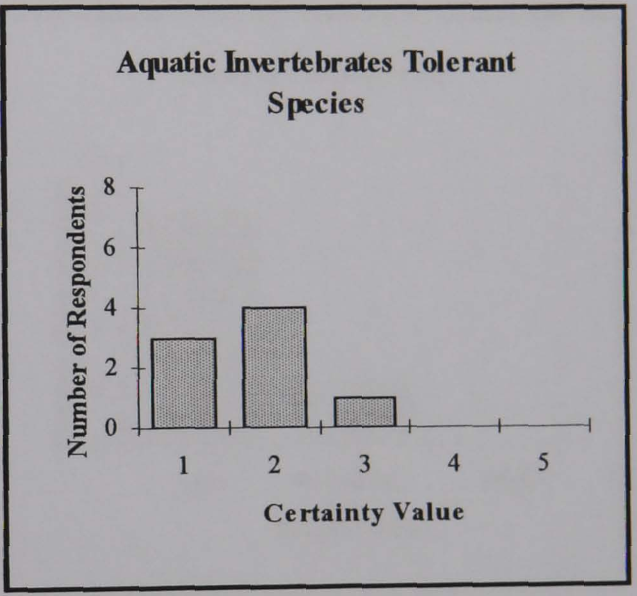


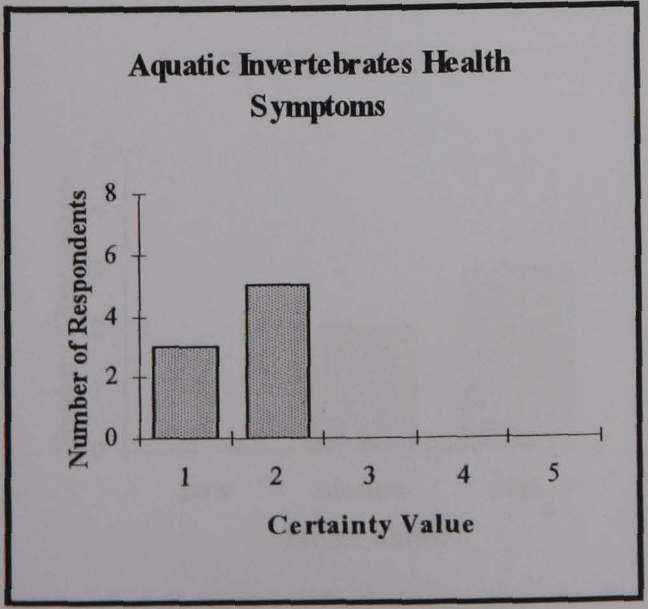
Questionnaire Results from the "Hydrogeologist" Classification

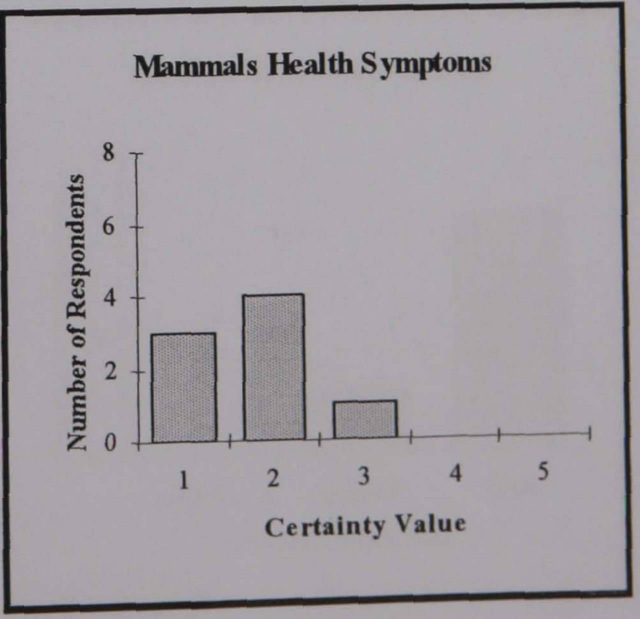




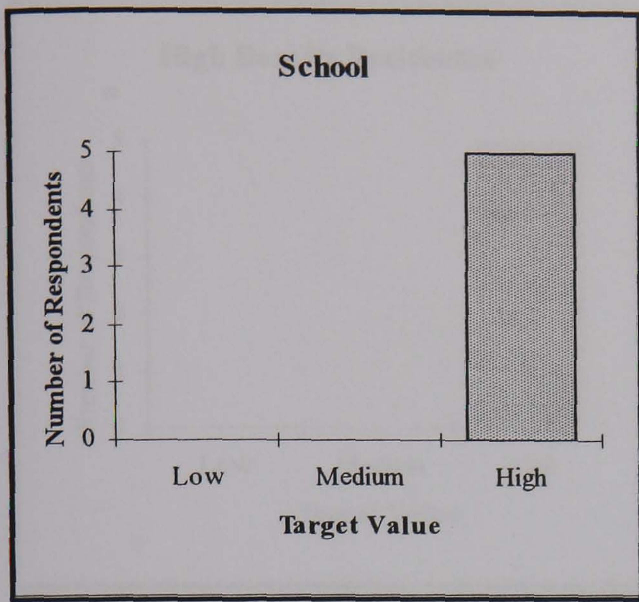


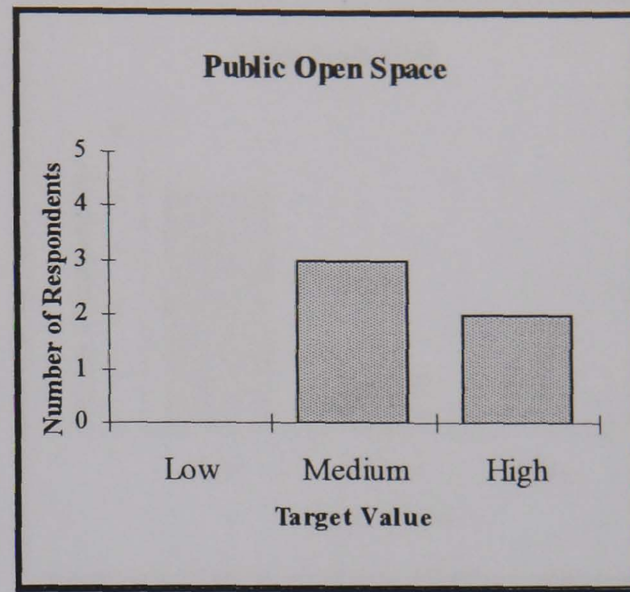


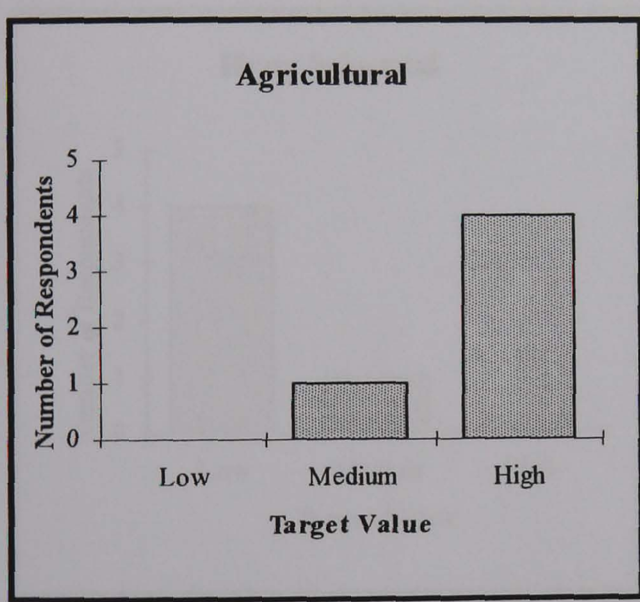


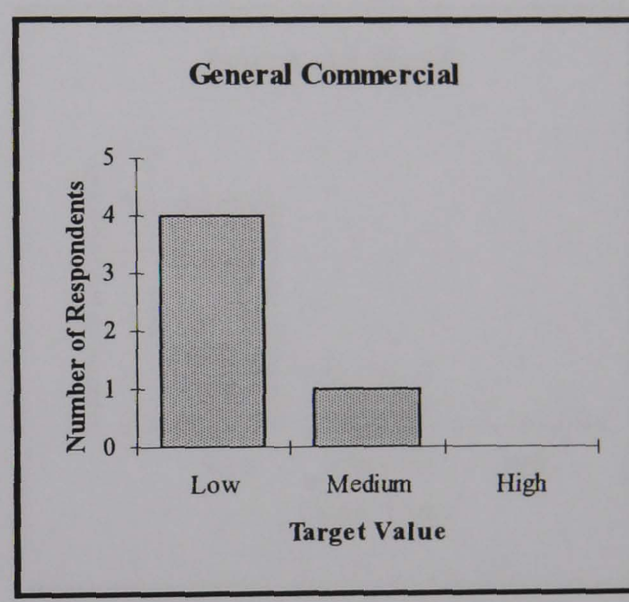


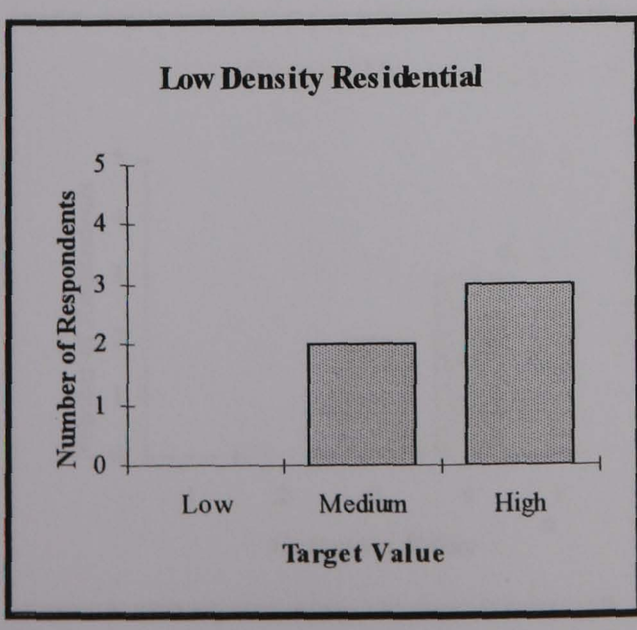
Questionnaire Results from the "Other" Classification

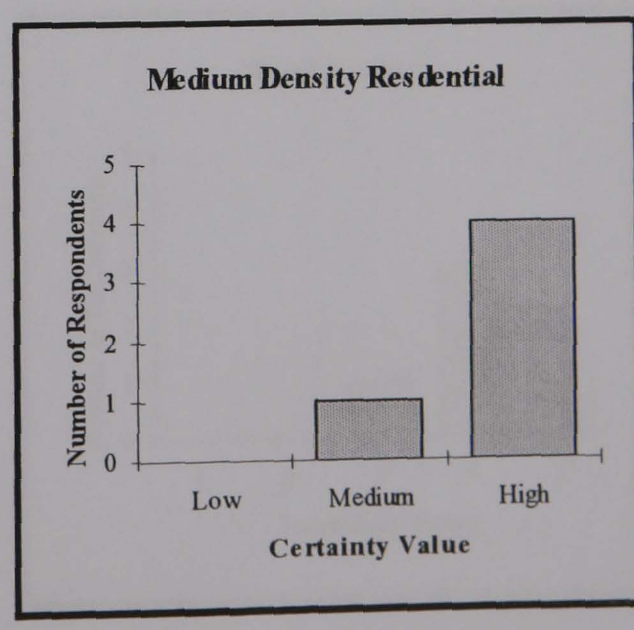




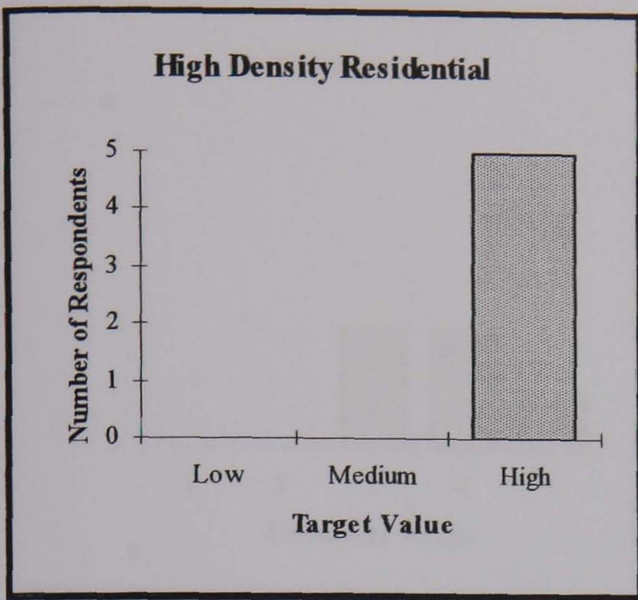


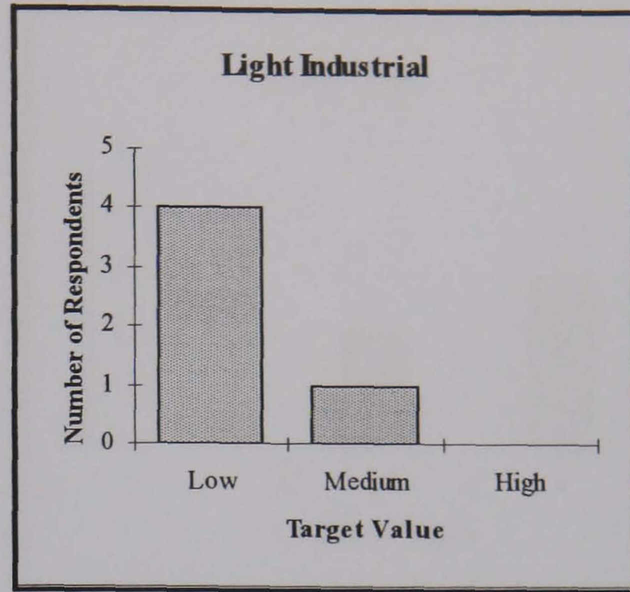


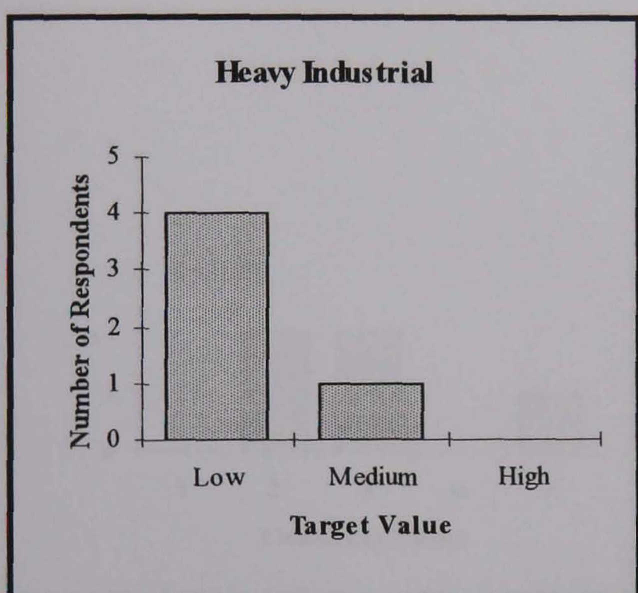


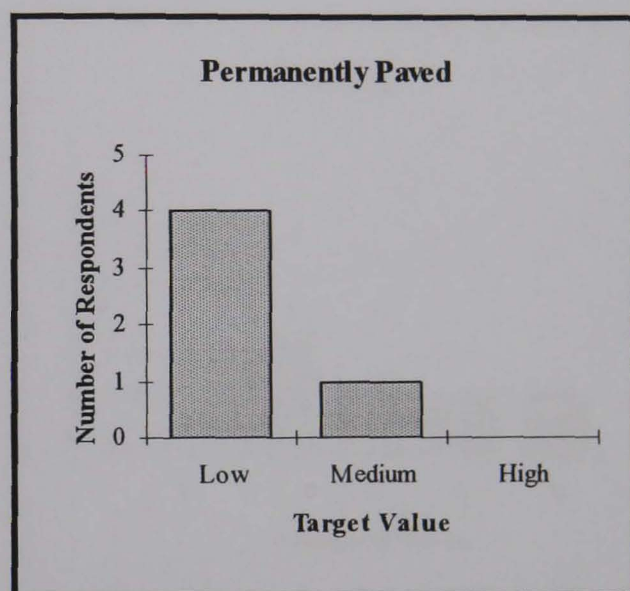


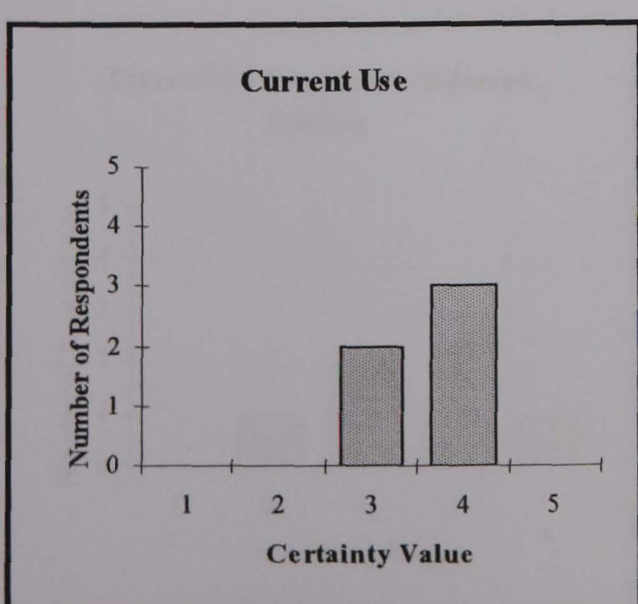
Questionnaire Results from the "Other" Classification

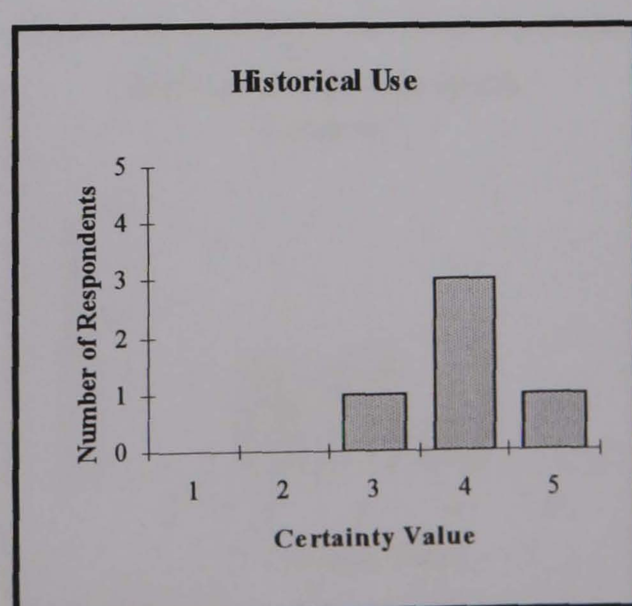




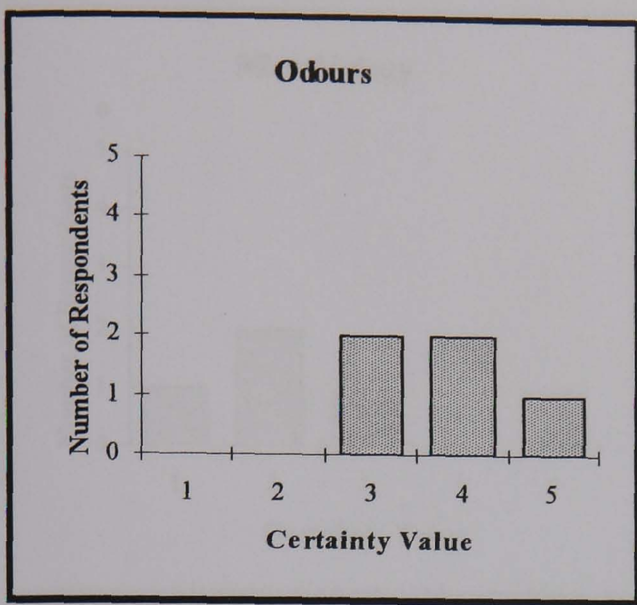


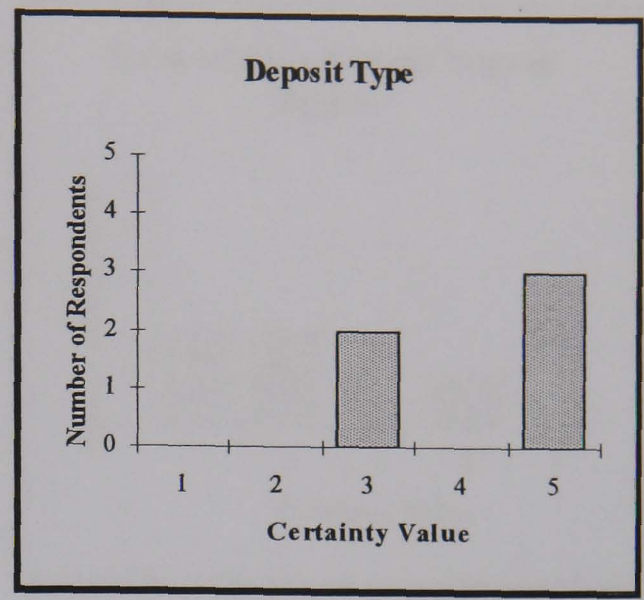


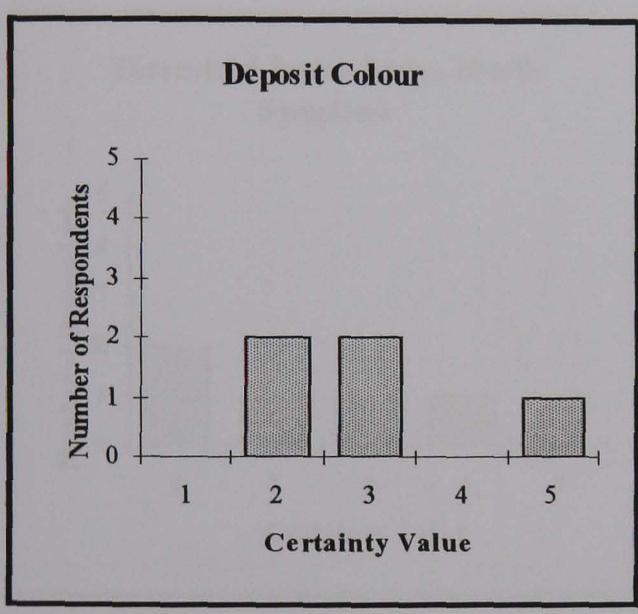


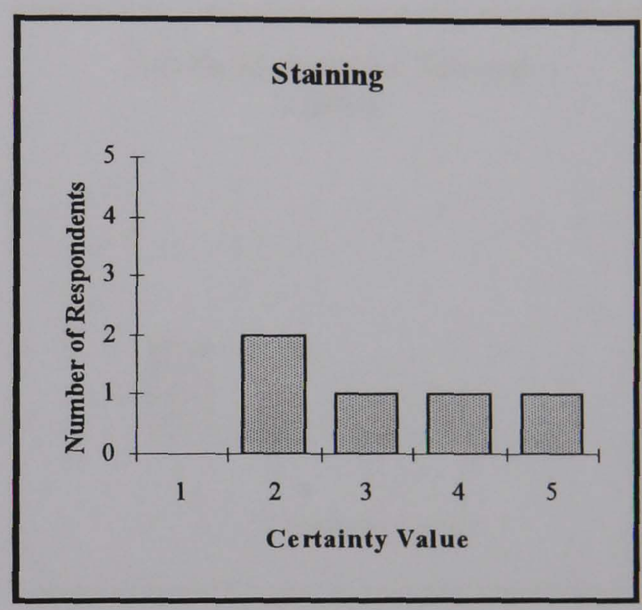


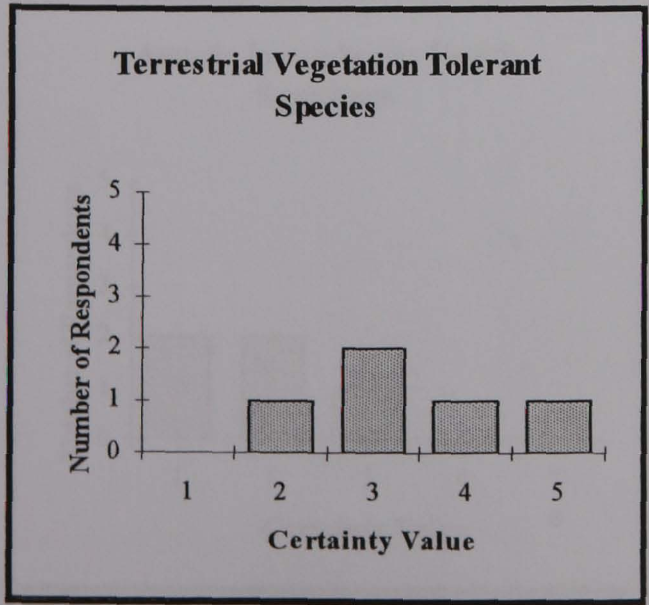
Questionnaire Results from the "Other" Classification

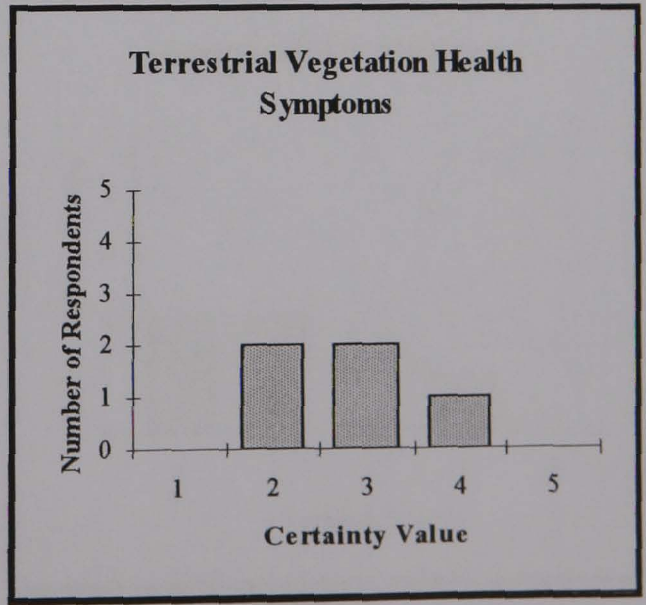




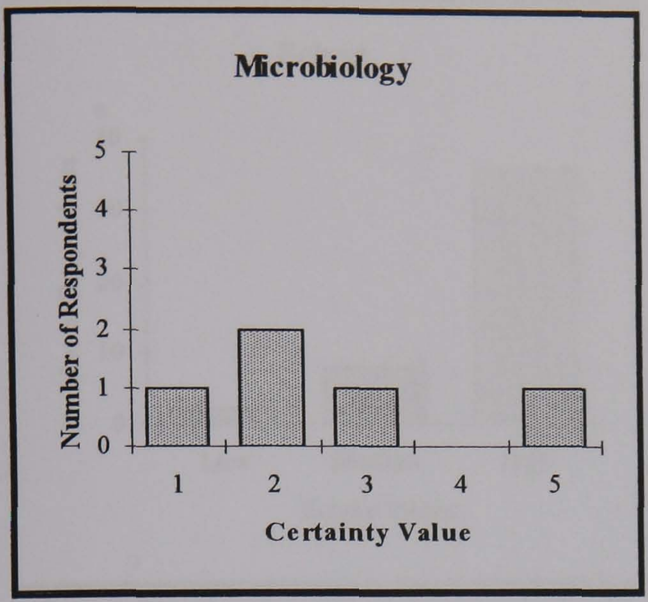


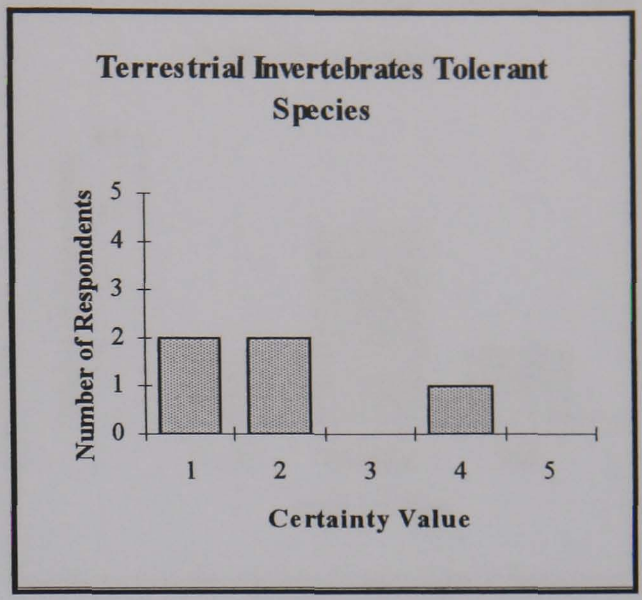


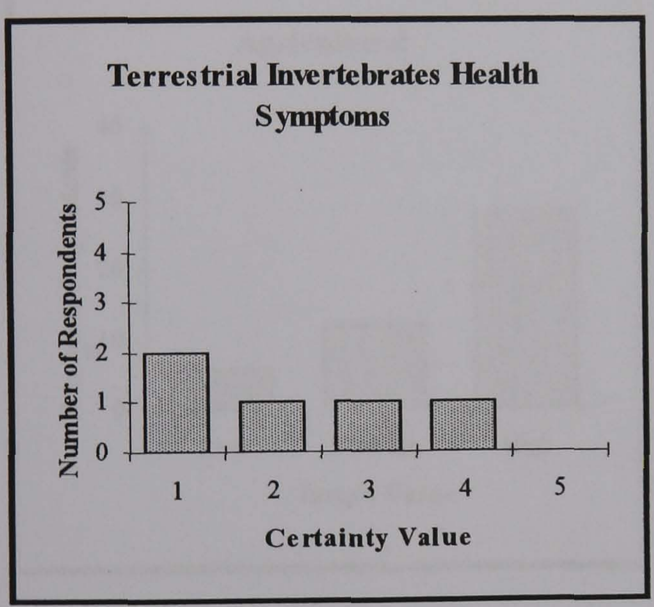


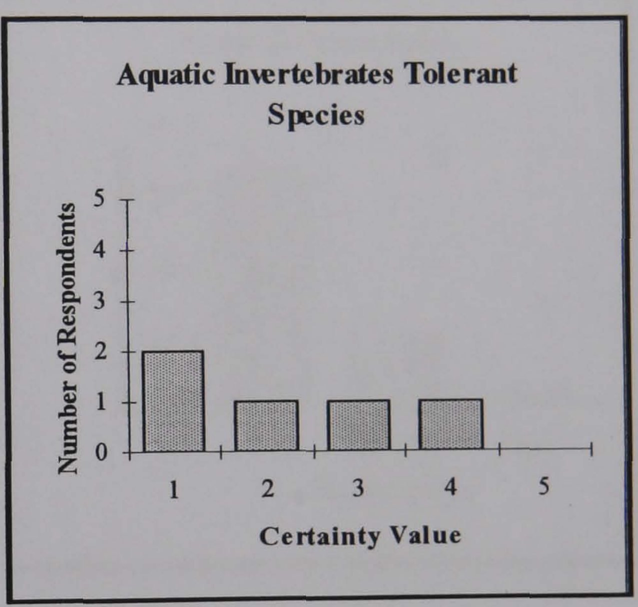


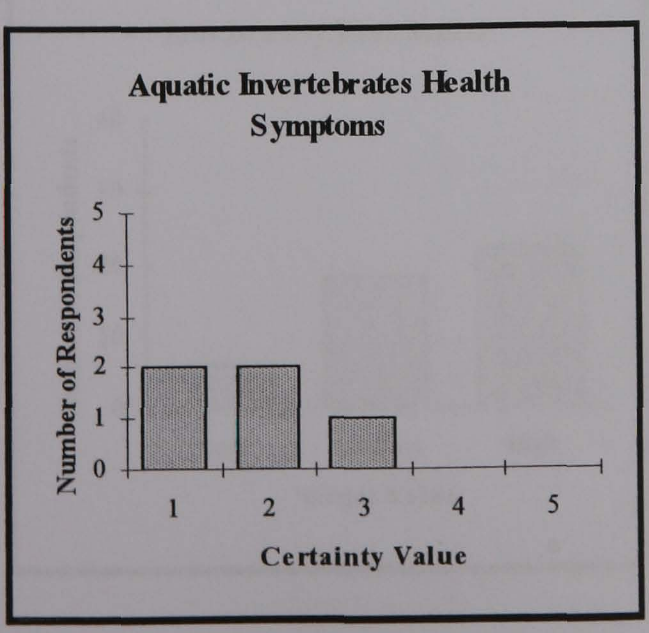
Questionnaire Results from the "Other" Classification

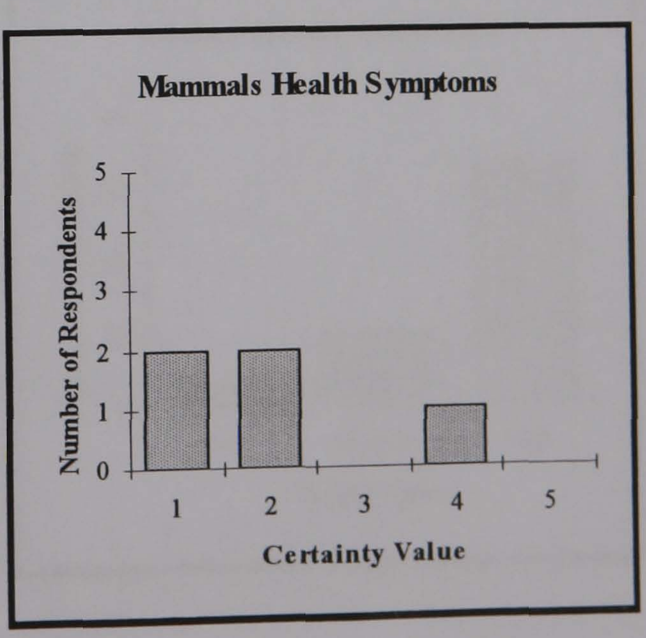




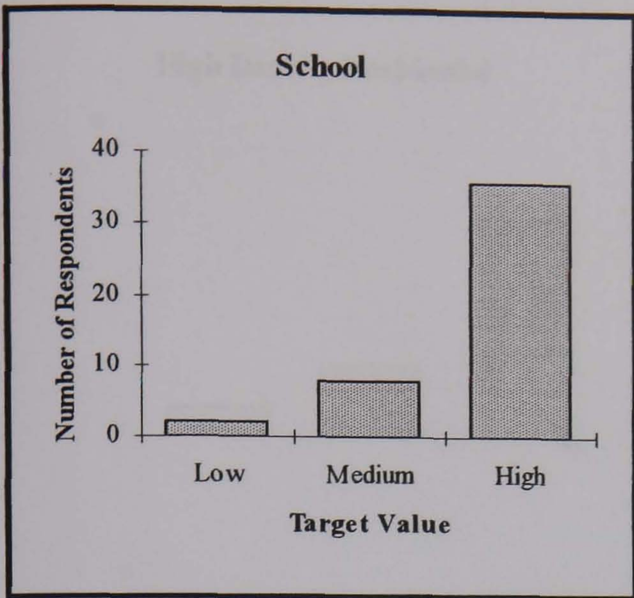


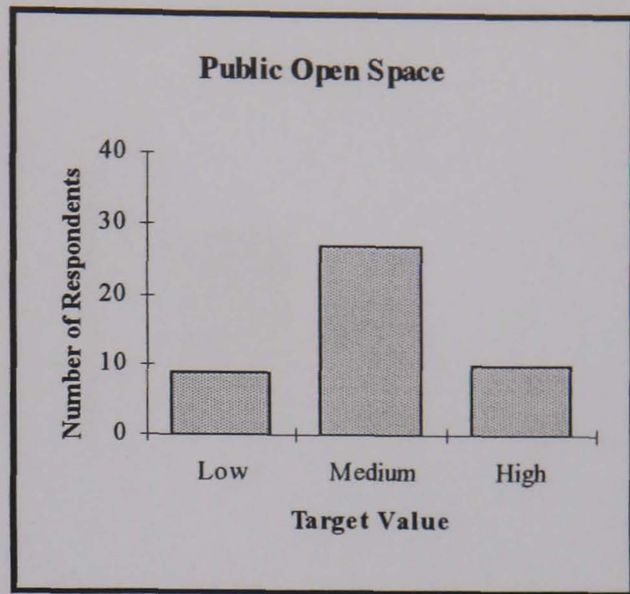


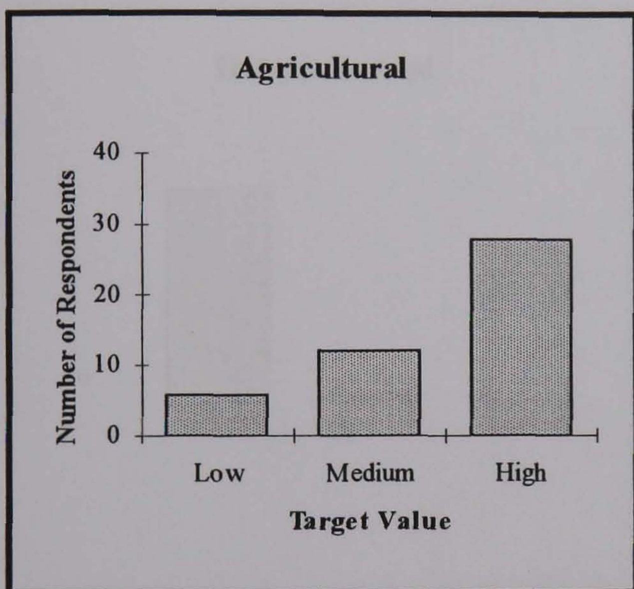


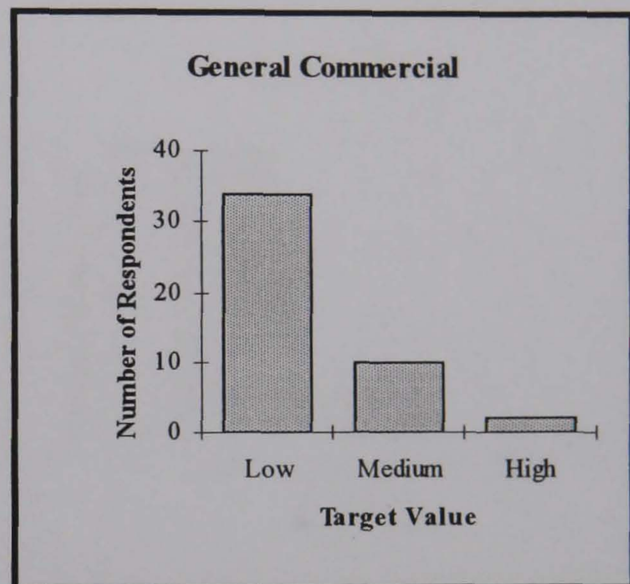


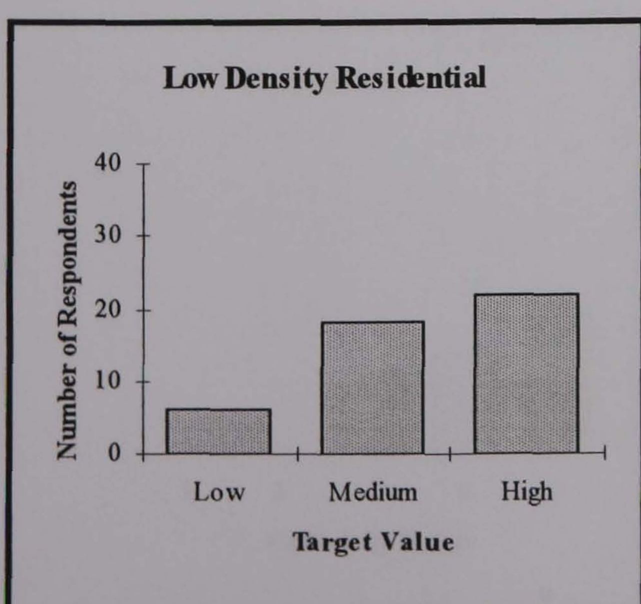
Questionnaire Results from the "Combined" Classification

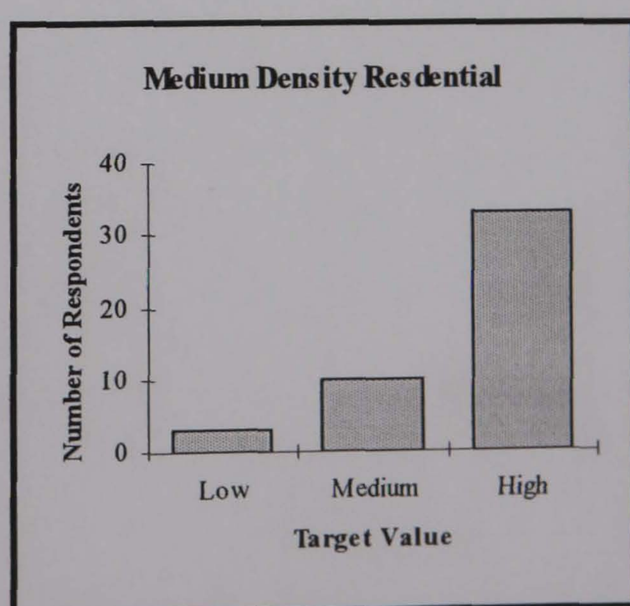




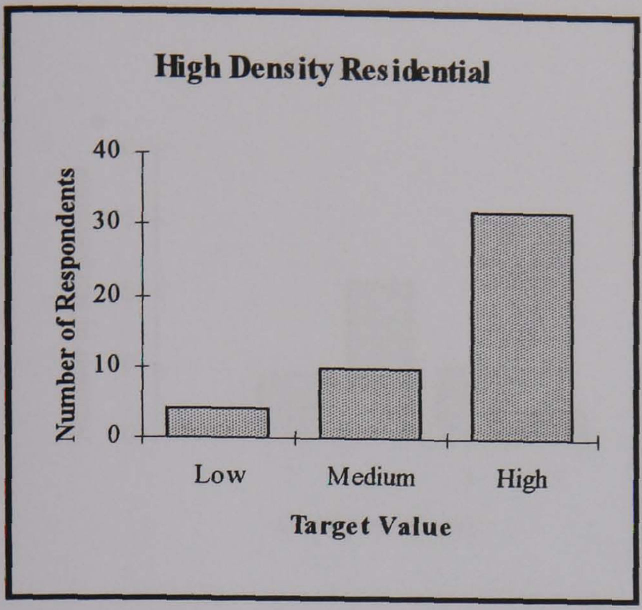


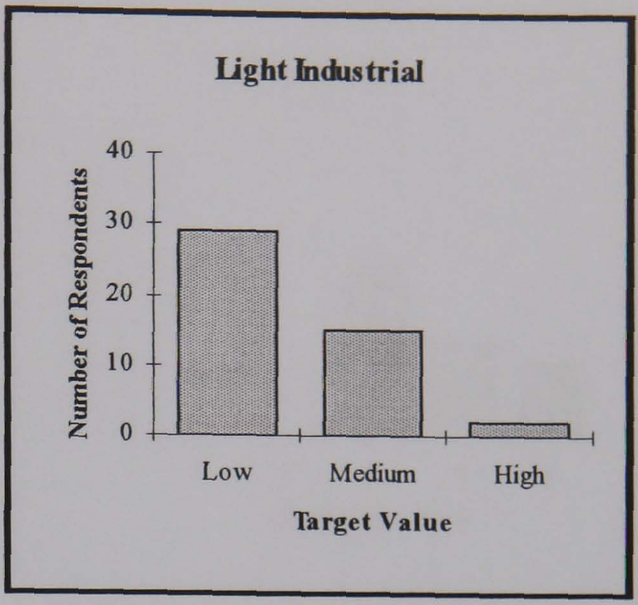


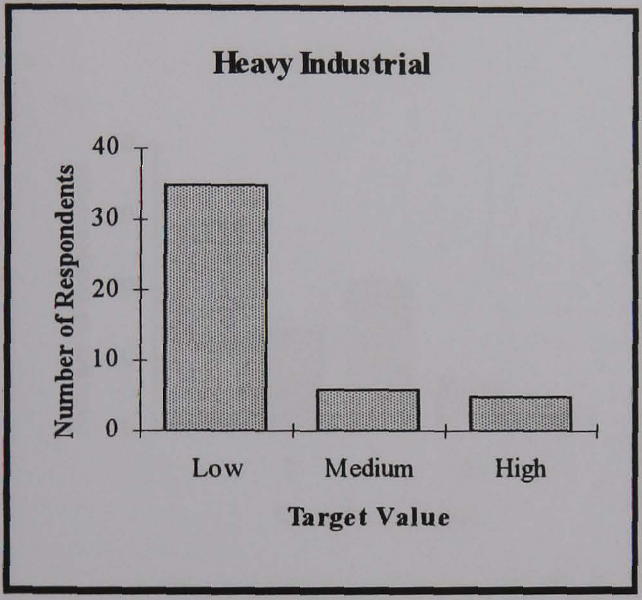


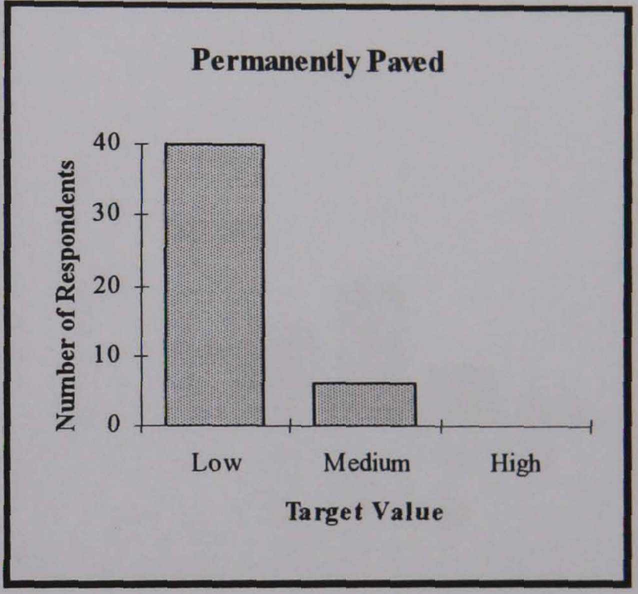


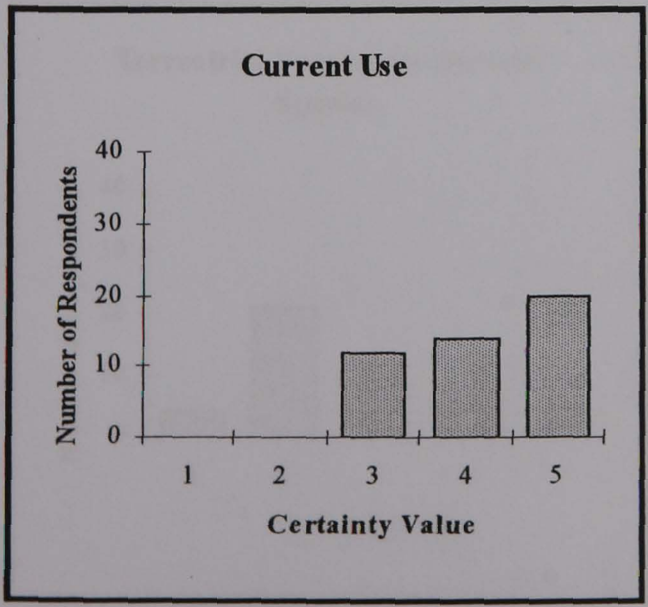
Questionnaire Results from the "Combined" Classification

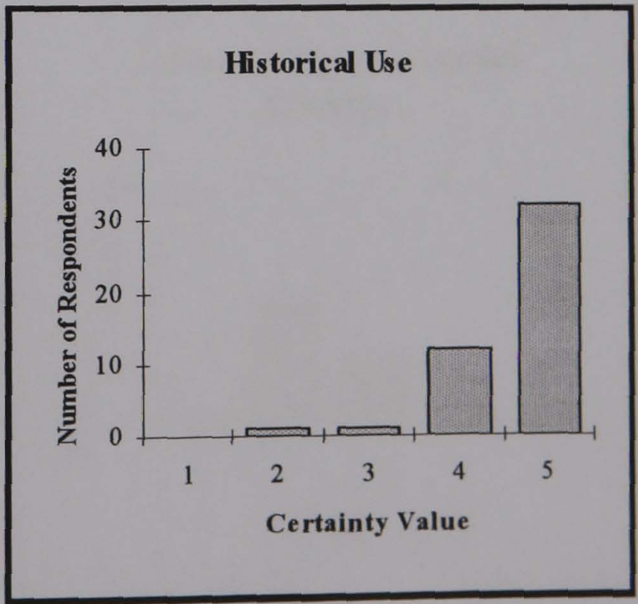




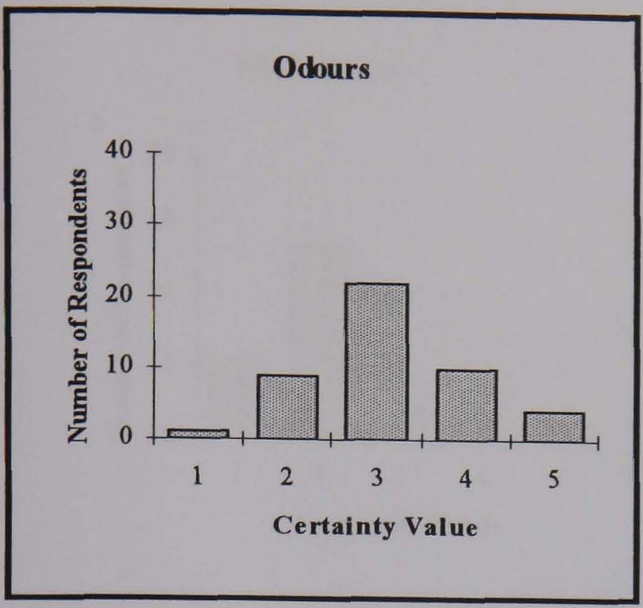


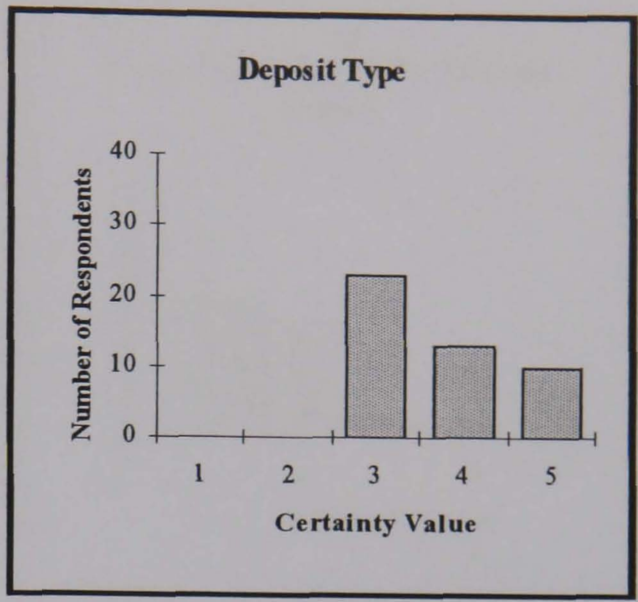


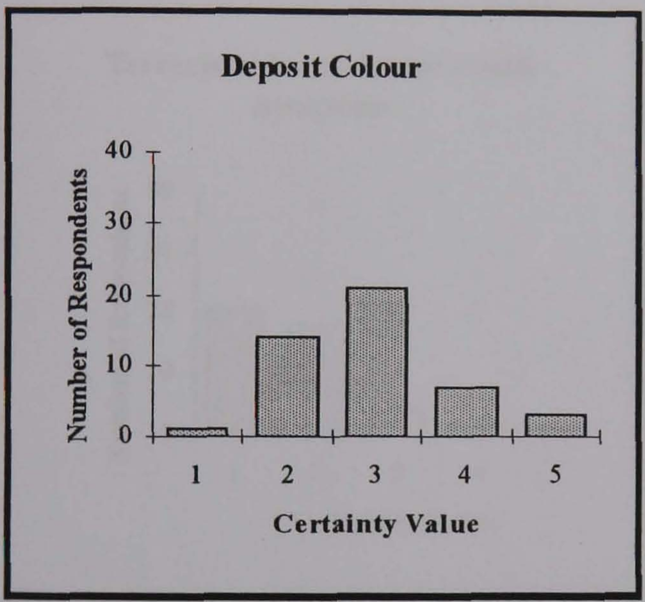


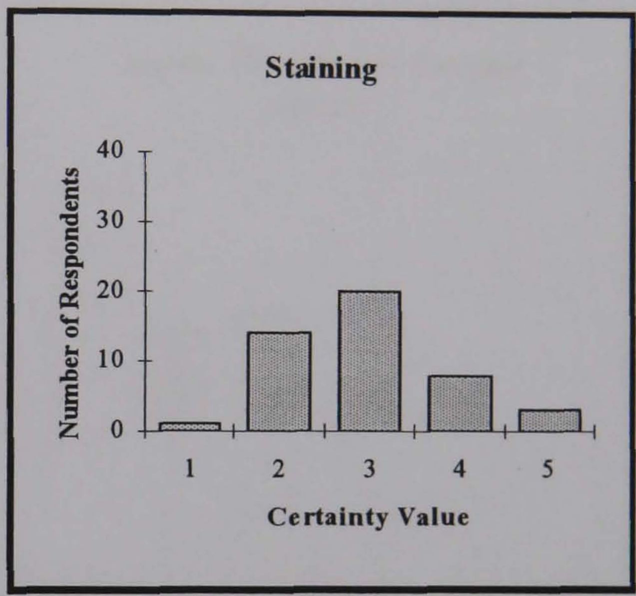


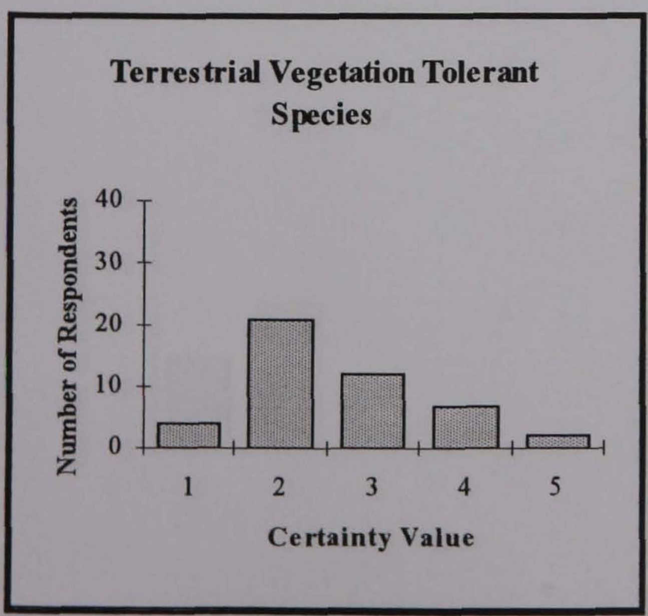
Questionnaire Results from the "Combined" Classification

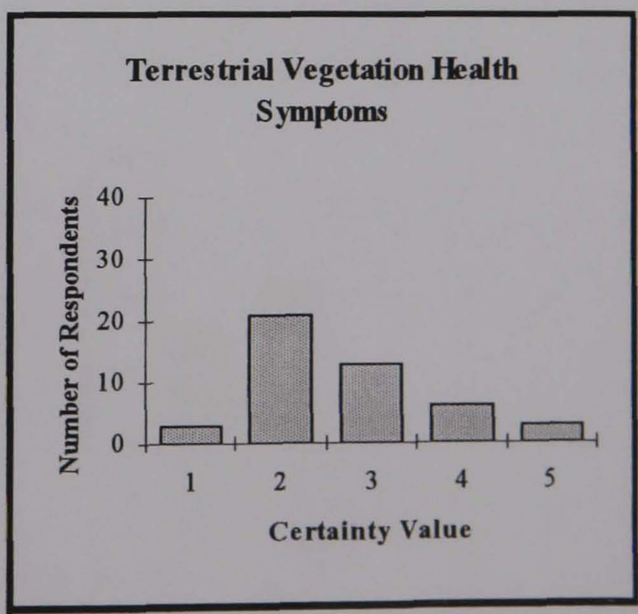




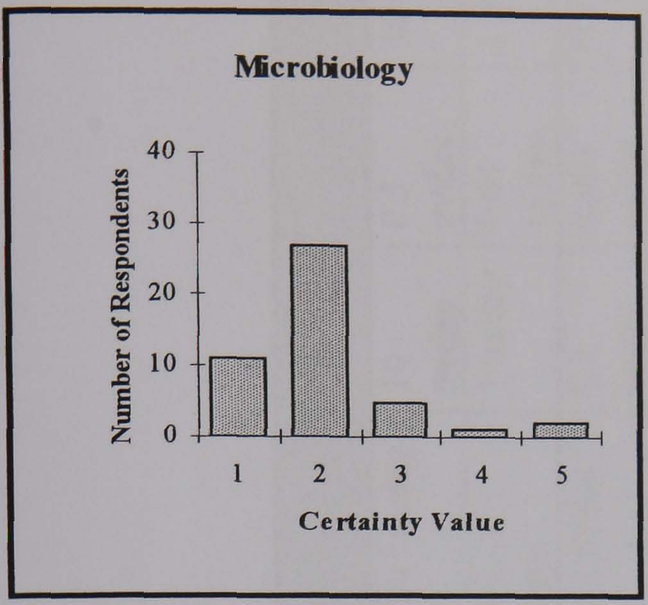


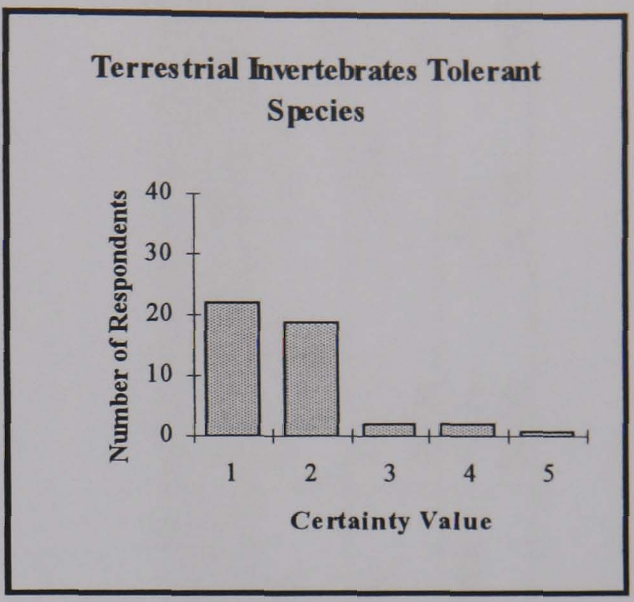


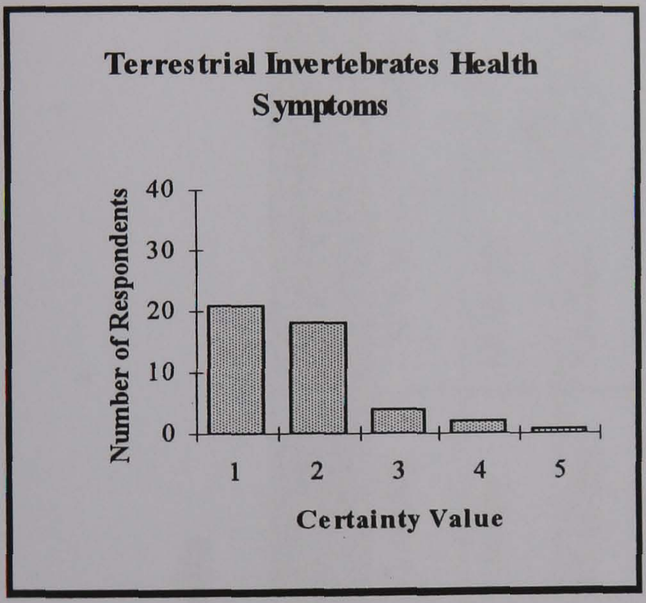


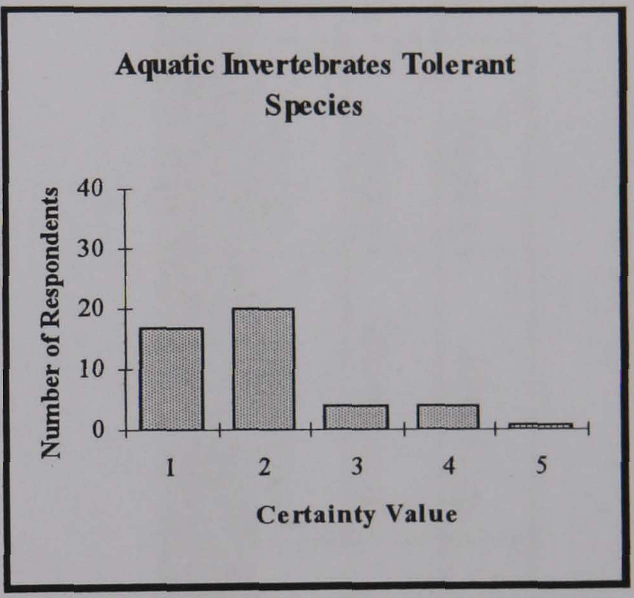


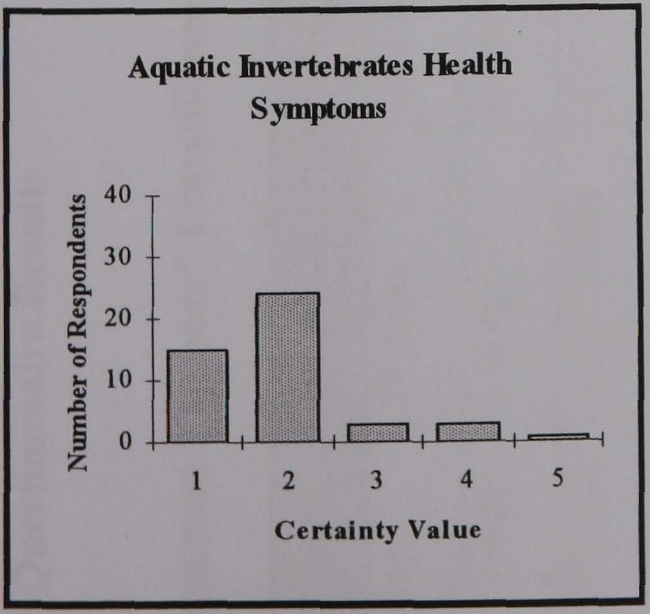
Questionnaire Results from the "Combined" Classification

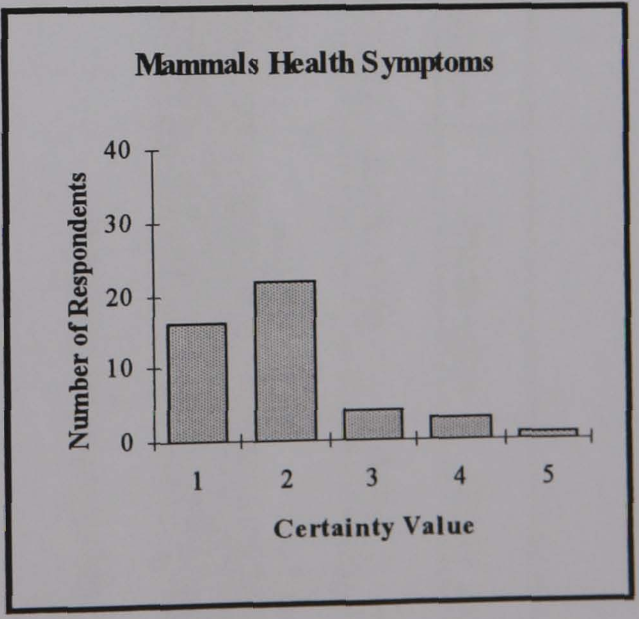












Groundwater Questionnaire Results

"Geo-Environmental Engineer" Classification Results

Term	Velocity of Groundwater m/day									
Fast	5 m/day	10 m/day	2.5 m/day	25 m/day	250 m/day	>1 m/day	<0.01 m/day	10 m/day	0.5 m/day	10 m/day
Medium	0.5 - 1 m/day	1 m/day	0.5 m/day	1 m/day	10 m/day	0.01 - 1 m/day	0.01 - 0.1 m/day	1 m/day	0.01 m/day	1 m/day
Slow	5 x10^{-2} m/day	1 x10^{-5} m/day	0.01 m/day	0.5 m/day	1 m/day	<.01 m/day	>0.1 m/day	0.1 m/day	0.001 m/day	0.1 m/day

"Geologist" Classification Results

Term	Velocity of Groundwater m/day								
Fast	>5 m/day	<.02 m/day	1 m/day	>10 m/day	10 m/day	0.1 m/day	1-10 m/day	10 m/day	1-10 m/day
Medium	0.05 - 5 m/day	.02-.0001 m/day	0.001 m/day	10 - .01m/day	6 m/day	0.01 m/day	0.01 - 1 m/day	1 m/day	0.1-1 m/day
Slow	<0.05 m/day	.0001 m/day	<0.001 m/day	<0.1 m/day	3 m/day	0.001 m/day	<0.01 m/day	0.01 m/day	<0.1 m/day

"Hydrogeologist" Classification Results

Term	Velocity of Groundwater m/day							
Fast	<0.01 m/day	>.3 m/day	>.5 m/day	>0.1 m/day	>0.1 m/day	>100 m/day	>300 m/day	>1 m/day
Medium	0.01 - 1 m/day	.01-0.3 m/day	0.1-0.5 m/day	0.1-0.001 m/day	>0.01 m/day	<100 - 10 m/day	50 - 300 m/day	0.001 m/day
Slow	>0.1 m/day	<.01 m/day	0.1 m/day	<.001 m/day	<0.01 m/day	<10 m/day	0.0001 -50 m/day	<0.001 m/day