SLEMS: A Knowledge Based Approach to Soil Loss Estimation and Modelling.

by

E.G.Mtalo

MSc. (Civil Eng.) — Technical University of Budapest, 1979.

A Thesis Submitted in Partial Fulfillment of

the Requirements for the Degree of

MSc. E

in the

Department of Survey Engineering.

This thesis is accepted.

Dean of Graduate Studies and Research

THE UNIVERSITY OF NEW BRUNSWICK

December, 1990

Abstract

In this thesis the modern information processing and management tools of remote sensing and geographic information systems are investigated as information sources for the knowledge based solution of soil erosion problems. Data and information requirements for the knowledge based modelling and estimation of soil loss from remote sensing and geographic information systems sources have been analyzed and established. The thesis also examines the problem of representation of vague, imprecise information, or fuzzy data, often used by the soil erosion domain specialists, in modelling and estimating soil loss.

The main thrust of the research is the knowledge based management and application of spatial attributive data in soil erosion modelling and estimation. Because of the complex nature of this problem, and the extent of the information requirements for the solution of soil loss related problems, the soil loss estimation and modelling system (SLEMS) has been designed and implemented in the C programming language, as a general purpose knowledge based system consisting of four subsystems.

An indexed relational database management subsystem handles conventional data. Two independent but cooperating knowledge based subsystems implemented as domain independent expert shells facilitate domain knowledge acquisition and intelligent query processing. The fourth subsystem constitutes SLEMS mechanism for representing and manipulating fuzzy data and knowledge.

The utility of the system has been tested on a rule-based expert system prototype for soil loss estimation and modelling. The system accepts both precise and vague data and uses a simple natural language interface to process queries on soil erosion related problems. Currently the system runs on a Sun 4 UNIX Workstation but it has been implemented with a view to porting it to MS DOS based personal computers.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to give my sincere thanks to Prof E. Derenyi for constant guidance and tolerance during the long period I have been working on the thesis research. Most of all I deeply thank him for giving me a free hand in selecting the thesis proposal, in assisting me in the preparation of the ground work, and for his financial support without which I would not have been able to complete writing the thesis.

It would not be appropriate however to start giving thanks to any one else before paying homage to my sponsors the IDRC, for making this dream come true. In particular I am deeply indebted to Prof. Omari who so to say, discovered me in the backwoods of Tanzania and opened my eyes to the possibilities offered by the IDRC.

I also thank my colleagues in the CanLab-INSPIRE whose cool manners allowed me to concentrate on my project without hindrance.

Lastly my sincere gratitude goes to my family especially the younger sons and daughters of my late parents Basil Kisakisa and Antonia Mkatomboka, and my beloved daughter Marcelina who had to put up with the hardships caused by my long absence. I would be amiss if I did not mention my daughter's mother Martha for the courage and perseverance, in raising our daughter alone during these three long years.

Chapter 1

Introduction

1.1 The Soil Erosion Problem Overview.

Technically soil erosion is defined (Morgan, 1986; Hóly, 1980) as the process where-by detachment and transportation of soil from its natural location takes place usually with adverse impact on the environment. Factors and causes of soil erosion are characterised (Wischmeier and Smith, 1957; Hóly, 1980; Goldman et al, 1986; Morgan, 1986) as climatic and hydrological agents (rain, runoff, antecedent moisture, wind, geographical position, and altitude), morphological agents (slope, slope length, surface roughness), geological and soil agents (soil type, parent material, texture), vegetative agents (plant cover type, density, height), technical agents (construction activities, conservation management, tillage systems, farm equipment) and social economic agents.

Six types of soil erosion can be identified according to the causative agent and physical characteristics of erosional features (Hóly, 1980; Goldman et al, 1986; Morgan, 1986). These are splash erosion due to rain drops; sheet erosion of soil by shallow "sheets" of runoff water; rill erosion caused by rapid concentrated runoff flow; gully erosion resulting from the deepening of erosional rills, and channel erosion due to disturbance of bank vegetation and increased stream flow.

Soil loss estimation is the process of determining the amount of soil and plant nutrients transported from agricultural lands, forest lands and rangelands by rain and wind energy. The estimation may be made by extrapolation from field and laboratory measurements or by modelling the effective soil loss caused by each known factor (Smith and Wischmeier, 1957; Meyer, 1984; Foster and Wischmeier, 1974, Foster et al, 1981). Soil loss modelling is therefore a complex process which requires the identification of causative factors, adoption of some empirical or physical models, and the determination of the values of the parameters of the adopted models (Smith and Wischmeier, 1957; Hóly, 1980; Goldman et al, 1986; Wischmeier and Smith, 1957).

Soil erosion is considered by some authorities as a major problem facing humanity (Hóly, 1980; Morgan,1986). As the world population continues to expand rapidly the need to conserve the small percentage of productive and fertile portion of the earth's soils becomes imperative (Hóly, 1980). Not withstanding the current level of understanding of the mechanism by which fertile lands lose their life giving nutrients and turn into deserts, human activities still continue to be the major factor in accelerated soil loss (Hóly, 1980; ESRI, 1984; Morgan, 1986; L'vovich et al, 1990).

Jacks and Lowdermill (Morgan, 1986) postulated that the destruction of past civilizations in North Africa, ancient Mesopotamia, the Bay of Arabia, and North China can be attributed to the devastating effects of accelerated soil loss due to over exploitation of agricultural land and forest resources.

To appreciate the contribution of human activity to accelerated soil loss Morgan (1986) gives the following rough measures. It takes 100,000 years to wash away 12in. of soil if covered by native sod, 12,000 years for marsh covered silt loam, and 29 to 36 years when the same land is cultivated to corn on 8% slope. The acceleration is obviously astronomical.

It is estimated that the USA was losing (in 1987) an annual amount of 4 billion Mg of soil representing an increase of 30% over the last 50 years since the great dust

bowl event (Morgan,1986). This milestone event in the history of soil conservation, occurred in the early thirties as a result of irresponsible agricultural practice and over-exploitation of the US and Canadian prairies (Hóly, 1980; Morgan, 1986). A devastating event of similar magnitude and nature also took place in the so called virgin lands of the Russian Siberian plains (Hóly, 1980; Morgan, 1986).

In both cases serious efforts at soil conservation were initiated after the catastrophes, to study the soil erosion phenomenon and to determine methods for preservation of agricultural lands against soil erosion by rain and wind. In the USA the CSD (Soil Conservation Department) was established under the department of agriculture USDA, and it has continued erosion monitoring to date under the new name SCS (Soil Conservation Service) (Goldman et al, 1986). It was also at that time, in history, that the new mapping technique of aerial photography was put to use in North America, Soviet Union and Europe for monitoring and mapping soil erosion.

Timely delivery of data on the state of soil erosion and the need for monitoring vast territories was the main incentive for the introduction of aerial photo-interpretation methods. For the same reasons remote sensing and geographic information systems (GIS) technology are extensively used nowadays to monitor the environment and to provide critical data for assessing soil erosion hazard potential and land degradation (Ringrose and Matheson, 1987; Stephens et al, 1985; Askari and Faust, 1981; Fansworth and Canterford, 1980).

In spite of the vast financial and scientific resources at the disposal of developed nations, such as the US and USSR, soil erosion still continues to be a major problem in their countries. Soil erosion monitoring and control is a difficult problem because of its complexity. Morgan (1986) considers soil erosion to be a multi-faceted problem involving:

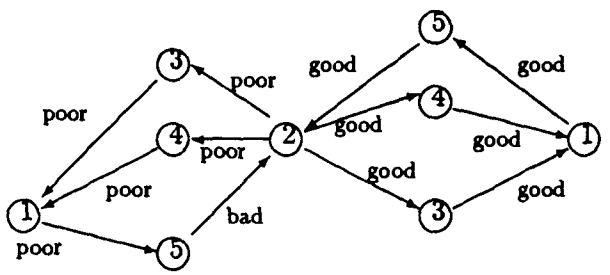
1. Policy.

2. Measurement and inventory (monitoring) of soil erosion extent and severity.
3. Assessment and evaluation of soil loss conservation efforts.
4. Modelling and estimation of soil loss.
5. Social and economic aspects.

The five components in the above macro-scale model of the problem are interwoven (figure 1.1). For example to formulate policy, one requires data and information on the soil erosion causes and current erosion state, assessment of existing soil loss conservation measures, resources available for conducting conservation, social and political impact of any adopted policy. Assessment and evaluation relies on appropriate measurement and inventory practice. Modelling and estimation are dependent on the quality of available data and knowledge of the various factors influencing soil erosion.

The right (or respectively left) part of figure 1.1 attempts to represent the propagation of good (or respectively bad) influences resulting from proper(or respectively improper) consideration and implementation of each facet on its related facets.

Practical solution to the soil loss problem requires a micro-scale analysis of the problem. In this respect the type of erosion, soil characteristics, precipitation and wind factors, terrain geomorphology and topography, engineering and construction activities, agricultural, range management, and forest management practice, ocean wave action, and other causative factors must be studied and modelled (Hóly, 1980; Goldman et al, 1986). Specific solutions must then be designed and initiated for each type (Hóly, 1980; Goldman et al, 1986). At micro-scale planning, issues of equipment, data sources, data processing, and information extraction methods must also be addressed.



KEY:

1 Policy

5 Social and Economic Influences.

2 Measurement and inventory of erosion.

3 Assessment and Evaluation of Conservation Measures.

4 Modelling and Estimation.

poor/good ⟶ Effect on the antecedent if the precedent is poor (or good).

Figure 1.1: Inter-relationships and Propagation of Influence Among the Facets of the Soil Erosion Problem.

1.1.1 Policy and Lack of Policy Impact on the Soil Degradation Problem: The Canadian Experience.

The general model of the soil erosion problem outlined above is in good agreement with situations existing in various countries. In Canada, policy or lack of comprehensive policy was identified by the Standing Senate Committee on Agriculture, Fishery and Forestry, as the most serious factor impeding progress in soil and water conservation efforts(Hon. Sparrow, H.O. 1984).

Specifically inappropriate or conflicting policies instituted by government agencies at various levels were found to be the major culprit. Examples were cited in Sparrow (1984) of certain economic incentives/disincentives which coerce farmers into producing more while ignoring the adverse effects to the soil. Similarly, policies intended to conserve soil moisture during the summer season, such as the summer fallow practice, were identified as a major causes of increased soil erosion and salinization of the soil. The net effect of bad policies and inadequate financing for research and implementation of good soil conservation measures have resulted in ever increasing soil loss by water and wind erosion. In 1984 it was estimated that the loss to Canadian farmers due to soil degradation was one billion dollars annually.

Although Canada is a vast country, only about 9% of its land is cultivatable and of this only 4.5% is farmed. At the same time 40% of Canada's GNP, 10% of all jobs and 10% of its export surplus comes from the agricultural sector (Hon. Sparrow, 1984). Obviously agricultural land is a priceless commodity in Canada.

The Senate report on the state of soil degradation in Canada is a major effort to focus government and public awareness to the dangers of irreversibly losing Canada's agricultural soils. The report hopes that increased awareness will result in better and comprehensive policies on the management of soil. It is also believed, that by placing a dollar value on the losses to farm production due to soil erosion, farmers will be more willing to adopt new solutions such as conservation tillage and zero tillage

methods.

Even with good policies in place, existing bottlenecks in the delivery of conservation technology to farmers have to be removed. The report therefore recommends the training of specialized soil conservation extension officers to facilitate efficient transfer of technology from research to the application points.

Assuming that most Canadian farmers are literate, the introduction of suitable PC-based expert systems to facilitate direct consultation to farmers, may provide a better solution to the technology transfer problem.

1.1.2 The Nature of the Soil Erosion Problem: Third World Perspective

In tropical and mid-latitude developing countries, where the amount of rainfall coincides with the maximum erossivity part of the soil erosion curve (see Hóly, 1980, pp. 9, fig. 7), the magnitude of erosion losses is immeasurable.

The situation was especially aggravated by the extreme exploitation of land resources which followed the colonialization of the African, Asian and American continents. The role played by the western civilisation in the aggravation of the soil erosion hazard in Tanzania for example can be illustrated by the following short story.

In my childhood, I remember humming to a local lyric[1] which the village songstresses had composed to deride the WaChakka's chief, who had decreed that every villager construct "matuta" (swahili for terraces) on coffee grown lands on orders from the English Governor General Hon. R. Turnbull.

Prior to the forced cultivation of coffee[2] villagers had cultivated the land on a continual basis without any apparent damage to the steep slopes of mount Kilimanjaro.

[1] "Wee lele, wee lele, hamba!, Sabasi kahamba le, hamba!, lureme matuta, hamba!, na matuta mali ha serikali, hamba!", which translates to: "oh, oh, speak up!, Sabasi has decreed, speak up!, that we must build terraces, speak up!, and terraces are government's property!"

[2] This was cleverly contrived by a so called head tax which could only be paid for in English money obtained from coffee sales

Coffee cultivation, however, required clearing of relatively large tracts of land (Temple, 1972) thereby exposing the light volcanic soils to the ravages of tropical torrents which caused devastating landslides (personal experience). Judging from the words of the lyric it is obvious that initially no one had bothered to educate the people on the usefulness of the terraces. As the people saw it then, this was yet another insensitive order by a foreign government, just like the coffee they had been forced to grow in the first place!. Evidently the villagers anger was justified! The chief's efforts were however rewarded, when the people soon put the "matuta" implementation to good use by using "masale" a densely rooted tropical plant which, while acting as soil retainer also served as forage for their enclosure animals. These modest soil conservation efforts were rewarded, and the soil erosion hazard in Kilimanjaro was permanently checked.

Unfortunately, the Wachakka's success at soil conservation efforts were not repeated in other areas of Tanzania such as the central regions of Dodoma and Kondoa, and the Uluguru Mountains of Morogoro where similarly mistaken agricultural policies by the British agriculture officers led to unreparable damage to land (Temple, 1972, Morgan, 1986, personal experience).

In North Western Africa the expanding sahelian desert is regarded by most experts to be partly due to the destruction of the ground plant cover by overgrazing and firewood harvesting (Aidoo, 1987, Morgan, 1986).

Due to the weak economies of most of the developing nations, monitoring and conservation activities are inadequate or non-existent. It is therefore not possible to tally with any measure of confidence the total amount of soil loss in those areas (L'vovich et al, 1990). At best soil erosion and agricultural experts in developing nations can only hazard a good guess based on vague information and experience.

Awareness of the seriousness of the erosion problem by some governments has however contributed to limited but critical efforts in erosion monitoring activities. For example in Tanzania, the government has continued soil loss conservation efforts since

independence. Erosion monitoring stations in operation within the country include Shinyanga, Lyamungu in Kilimanjaro, Tengeru in Arusha, Mpwapwa in Dodoma, and Mfumbwe in Morogoro, which have continued for the past 50 or so years (Temple, 1972, Morgan, 1986). However according to Temple (Morgan, 1986), monitoring has been operating intermittently rather than continuously due to inadequate financial resources and technical support.

A number of international organisations have also been actively studying the soil erosion problem in some African countries. These include the UNEP (United Nations Environmental Program) and FAO (Food and Agricultural Organisation) (ESRI, 1984) who have sponsored research in Kenya (Burrough, 1987, Simonet et al, 1987), Uganda (Simonet et al, 1987) and Ghana (Aidoo, 1987). Bilateral agreements between countries also provide poor countries with badly needed research potential.

For example Kenya has benefited from a collaborative association with the Agricultural University of Wagenigen's Tropical Soil Science research branch (Burrough, 1987), which has conducted long term erosion research in the Kisii District. Similarly Tanzania and Lesotho have benefited from research results from Rapp and Temple (Morgan, 1986) who conducted limited but very useful research work on erosion in the sixties and early seventies.

Efforts have also been made by individual researchers (L,vovich, 1990) and international organisations (ESRI, 1984) in the assessment of the potential threat of soil erosion on a global level. Such projects are a useful source of default data and information to researchers in developing countries.

Reasons for the apparent lack of success in controlling soil erosion in developing countries include (Morgan, 1986, personal observations)

1. Lack of adequate data for planning conservation efforts.
2. Lack of facilities and resources to monitor and assess the magnitude of the problem and to judge the effectiveness of conservation efforts.

3. Inadequate human resources or expertise in tackling the soil erosion problem.
4. Increasing cost of living and dwindling arable land which make it difficult for small scale farmers to abide to instituted soil conservation practice.
5. Increasing pressure from international money lenders on poor countries to produce and export more.

It is therefore clear that from the developing country perspective the soil erosion problem is indeed complex and requires financial and human resources, and international political good will to solve.

1.1.3 Role of The Surveying Sciences in the Solution of the Soil Erosion Problem

The role of the modern surveyor in the solution of the complex problem of soil erosion is two-fold. First as a specialist in the art and science of measurement he provides the instruments, methods and measurements necessary to provide the domain scientists with their requisite data. Secondly the modern surveyor assisted by computers is regarded as an information management specialist. In this role the surveyor can provide methods and tools for manipulating geographical and attributive data and presenting it in a form more conducive to the domain scientists needs.

1.1.4 Motivation for Research into the Soil Loss Problem.

Soil loss estimation and modelling is an area where special techniques for information acquisition, processing and management can go a long way towards fulfilling the information needs of domain specialists. During extensive travels over the Tanzanian country-side between 1979 and 1986 I became acutely aware of the physical extent

of the soil erosion problem in the country. It also become evident that current topographic maps do not provide any information on either the areal extent or the type of soil erosion within the country.

Between 1980 and 1987 I tried to get some of my students at the Ardhi Institute to look into the viability of using Landsat imagery for mapping landuse patterns over the Tanzanian and Ugandan territories. We were able to show tentatively that black and white and colour composites of Landsat MSS and Landsat TM imagery could indeed provide updating information at 1:250,000 and 1:100,000 scales (ARI, 1980 - 1987).

Between 1967 and 1969 I also had the benefit of working temporarily with the photogrammetry and map reproduction section at the Survey and Mapping Division of the country. During that time I became aware of the vast amounts of aerial photographic cover which existed then in the country. Since then additional photographic coverage has been added to the stock.

Although according to an investigation done by one of my students in 1982 (ARI, 1980 - 1987) the older film is in bad condition, it may still be useful for photo interpretation purposes. Further more the old photographs could provide a good historical record to detect changes in the landscape caused by erosion, flooding etc. For example Rapp (Morgan, 1986) demonstrates the use of 1960 aerial photographs supplemented with field completion data from 1969 - 1971, in the identification and mapping of erosion features in Dodoma, Tanzania.

The nature and magnitude of the soil erosion problem relating to the Tanzanian experience was therefore the major motive for focusing my attention to the soil loss and estimation problem. In pursuing this research I hope that, I will be able to contribute to the efforts being conducted by the agricultural and soil conservation experts in Tanzania and other developing countries at large by providing a method which will facilitate efficient transfer of technology from the remote sensing and GIS domains into their field of application.

Further more with an eye sight on the special circumstances existing in developing countries at large (Morgan, 1986), I consider it important that local researchers seek out ways of processing and incorporating substantial but, non-conventional and non-precise data and information in the form of qualitative aerial photo interpretation reports, map interpretation, reports, and written advice of past experts in the soil erosion business.

1.1.5 The Scope and Limitations of the Research

Within the scope of the five facets of the complex soil erosion problem (fig. 1.1), the specific area where this research seeks to provide a contribution is in the modelling and estimation aspect. The thesis research is further restricted within this narrow area of application to the design of tools and methods which will aid the modelling and estimation process through the efficient transfer of GIS and Remote sensing technology into the soil erosion problem domain.

Specifically the research has been guided by the desire to provide an efficient tool for the collection and organisation of data, facts, and extraction of knowledge from GIS and remote sensing sources, in a manner which will have a favourable impact on the solution of soil loss modelling and estimation problems.

No attempt is made to comment on the suitability of existing soil loss estimation and modelling methods. Neither is the business of this research to specify new soil loss estimation models as these issues are considered subjects which belong to the soil erosion domain experts.

The primary input of the research is therefore the identification of data sources, information processing and information management techniques to facilitate soil loss estimation and modelling using existing models. In keeping with current trends in information processing and management (Martin, 1984; Robinson, 1987; Stonebraker, 1990) the study looks into the viability of a knowledge based approach to the management and extraction of information from remote sensing and GIS sources and

inter-domain technology transfer in general. In addition the research seeks for solutions which can be implemented in third world countries.

1.1.6 Research Goals

Viewed from the points raised in the previous section the proposed research will pursue the following issues and goals.

- Ways in which the proposed research can directly impact on the soil loss estimation and modelling problem.
- Ways in which modern tools and methods of data acquisition and information management can be placed at the disposal of the soil erosion domain scientists and experts.
- Methods for representing vague non-precise data and information in a meaningful way for computer manipulation.

From the previous discussions it must be clear by now that the main focus of the research is information management and technology transfer from the mapping sciences at large to the soil erosion research domain. It is the intention of this research to address cheap but reasonably efficient means for harnessing GIS and remote sensing information resources for the solution of soil erosion related problems.

To this end the study will undertake a knowledge based approach to information management, motivated by the desire to maximize the information value of uncertain and vague data which may have to be used under the circumstances discussed in section 1.1.2. This approach is indicated by the complex nature of the soil erosion problem and the over-reliance on human experts in its solution. Related indicators for which real world problems require knowledge based solutions are addressed in section 1.2.2.

The fast pace at which new technologies have been evolving in the mapping sciences is another reason for opting for a knowledge based approach. In this respect the research focus will be on the development of tools which will facilitate knowledge transfer from GIS and remote sensing by automating the inter-domain knowledge flow process.

Finally application of GIS and expert system (ES) technology in solving soil erosion problems has been shown to be viable (Morrison et al, 1989). In this research I will therefore have the benefit of drawing from related past solutions.

1.1.7 Thesis Outlay

In the previous sections the scope of the research and the goals pursued by the research were outlined. In this section a brief outline of the thesis contents and organisation is given in order to make it easier to follow the rest of the report. The rest of chapter one discusses in depth the data requirements for soil loss estimation and modelling. Specifically, this chapter focuses on sources of data, methods of data acquisition, and also attempts to discover those critical areas in data acquisition and knowledge extraction best suited to the knowledge base approach. The chapter also gives a third world perspective of the soil erosion problem to lend credence, support, and motive for the choice of research focus.

Chapter two is a short discussion of the SLEMS DBMS utility. No attempt is made to give a detailed discussion of database management principles. However design and management aspects relevant to the CDATA based SLEMS DBMS are briefly outlined.

In chapter three the focus of the discussion shifts towards the knowledge based approach to information management. Basic principles of knowledge base systems are introduced and discussed with respect to SLEMS proposed application. The chapter covers in more detail the semantic network structure and rule based knowledge representation. It also provides a fairly extensive literature review of knowledge structures

and knowledge representation requirements and strategies.

The subject-attribute (OA) tuples and subject-verb-object (SVO) triplet knowledge structures used as the basic elements of the knowledge structures implemented in the SLEMS are introduced and discussed in the first few sections. The last part of chapter three however deals more with the actual design and implementation of the SLEMS LEARN subsystem.

The SLEMS EXPERT and FUZZ subsystems are also given a brief introduction to smoothen the passage into chapter 4 and chapter 5 which are reserved for the FUZZ and EXPERT subsystems respectively. A few examples are also given to illustrate the role of the LEARN subsystem in knowledge acquisition and query processing. The chapter discusses in detail the LEARN subsystems knowledge manipulation and inference modules. Chapter 4 introduces new concepts in the management and handling of fuzzy knowledge. Specifically a fuzzy geometrical partition approach to the representation of fuzzy restrictions of the type *about(x)* is introduced and developed for application in the manipulation of fuzzy objects in the SLEMS knowledge base. Results of test application of the theory in a test database and the SLEMS knowledge base are presented for illustration.

The objective of chapter 5 is to present the SLEMS EXPERT subsystem and its Knowledge Base. This is done through a discussion of general principles and design specifications of expert systems followed by a detailed discussion of the EXPERT's knowledge manipulation modules. Several examples and extracts of the EXPERT's sessions are presented to illustrate its applicability. It also buttresses the theory discussed in chapter 2 and three.

Chapter 6 is the conclusion part and it gives a summary of the thesis research, detailing achievements, unresolved problems and future role of the proposed system.

1.2 Soil Loss Estimation and Modelling Data Requirements

As an outcome of serious and long research of the soil erosion problem Wischmeier (Meyer, 1984; Wischmeier, 1984) introduced USLE, the first universally adopted soil loss model which was adopted by the USDA SCS in 1978 . The model (Eq. 1.1) was derived empirically from data collected from forty seven research stations in twenty four US states and it summarises the effects of various factors observed continuously for 5 to 30 years at individual stations (Wischmeier, 1984).

$$A = R \times K \times S \times L \times C \times P \tag{1.1}$$

These factors include, soil characteristics, rainfall and runoff effects, ground cover, farming and conservation practice and topographic factors such as slope gradient and slope length. The six parameters corresponding to these factors are R, a dimensionless rainfall erossivity factor, K, the soil loss rate per unit area under ideal conditions (continuous fallow, 9% slope, 22.1 m. long slope) due to soil erodibility alone, L, a dimensionless slope length factor, S, a dimensionless slope gradient factor, C, a dimensionless cropping management factor, and P, a dimensionless conservation practice factor.A is the predicted average annual soil loss in tons per acre.

The main sources of data and information required for soil loss model parameter estimation and soil loss estimation comes from the following sources (Wischmeier and Smith, 1957; Bali and Karale, 1977; Morris-Jones and Kiefer, 1978; Hóly, 1980; Logan et al, 1982; Spanner, 1983; Walsh, 1985; Best and Westin, 1984):

1. Conventional topographic and thematic maps.
2. Aerial photographs and digital images.
3. Space photographs and imagery.
4. Geographic information systems.

5. Field measurements of the erosion and soil loss process.

6. Field and laboratory experiments.

Table 1.2 gives a summary of the information requirements for the determination of each factor in Eq. 1.1. A more detailed list of information requirements for erosion and sediment yield measurements in general is given in Kólař (1977) and Fleming (1981).

Over the past decade the mapping sciences and profession at large has been shifting towards computer based digital mapping and spatial databases. The extraction of information from computer based maps is done by spatial data analysis techniques (Blais, 1987; Burrough, 1987; Ripple and Ulshoefer, 1987; Goodenough, 1987, 1988; McKeown, 1989; Schenk and Zilberstein, 1990; Egenhofer and Frank, 1990; Hadiprionо et al, 1990).

Aerial photographs have been used in solving soil erosion related problems for a very long time (Goosen, 1964). They provide both quantitative and qualitative information about ground cover, watershed characteristics, topographic data such as elevation, slope gradient, and slope length, soil texture and class. The relevant methods applied in this case are aerial photointerpretation and photogrammetry (Goosen, 1964; Webster and Wong,1969; Webster and Beckett, 1970; Parry and Beswick, 1973; Speight, 1977).

Space and airborne imagery is increasingly becoming a major data and information source for environmental and natural resources studies including soil erosion monitoring and range degradation assessment (Bondelid et al, 1980, 1981; Stephens, 1985; Ringrose and Matheson, 1987). Data which can be obtained from these sources include ground cover and terrain morphology (Parry, 1973; Cermak et al, 1979; Schneider et al, 1979), soil texture and soil moisture (Anderson, 1979; Schmugge et al, 1979), runoff curve numbers (Bondelid et al, 1980, 1981; Stephens et al, 1985), precipitation (Barret et al,1979; Heilman et al, 1979; Fowler, 1979; Scofield and Oliver, 1979;

Woodley, 1979; Ullaby et al, 1979; Fansworth and Canterford, 1980; Richardson et al, 1981) etc. Techniques for extracting the necessary information are remote sensing image analysis and classification (Lillesand and Kiefer, 1987; Asrar, 1989).

Geographic information systems and spatial databases provide a sophisticated tool for managing and analyzing data and information gathered from various data sources. Using GIS techniques areas of critical soil erosion can be identified (Morris-Jones and Kiefer, 1978; Burrough, 1987). Factors affecting soil loss and the universal soil loss equation parameters can also be determined by photogrammetric, remote sensing and GIS techniques (Dean and Schneider, 1977; Morris-Jones and Kiefer, 1978; Askari, 1981; Spanner, 1983; Stephens, 1985; Gilley et al, 1987)

The level to which each of these sources satisfies the requirements for soil loss modelling, estimation and monitoring depends on the data quality and ease with which the information required by the relevant scientists and researchers can be accessed. For example in developed nations where access to accurate data is not a problem, the emphasis is more on accuracy and completeness of data from the individual sources, and efficient information delivery. In developing nations where according to Temple (Morgan, 1986) lack of data is the norm, any available data is useful for planning and even implementation of soil loss prevention measures. However even in developing nations information delivery bottlenecks can aggravate the problems faced by domain experts.

Similarly computing facilities or services in developed nations guarantee efficient means of processing the data and extracting information essential to the solution of soil loss related problems. The same cannot however be said of developing nations. It is therefore necessary when proposing solutions to take cognisance of the very different situations pertaining to developed and developing areas of the world.

Domain scientists and researchers in the developing countries are more likely to be forced to make decisions using uncertain and unreliable data than their colleagues in the developed nations. Because of this access to knowledge based solutions, which

have in-built facilities to cope with uncertainty in data and information, is just as important or even more so to third world researchers as it is to those in the developed world.

1.2.1 Requirements Analysis.

The solution of soil loss estimation and modelling problems require multi-source information and knowledge. General proposals on data requirements for soil erosion and hydrological modelling in general have been given in Kolář (1977) and Fleming (1981).

Remote sensing and GIS data needed for soil erosion studies includes, soil data, ground cover data, rainfall data, and others (table 1.1 and 1.2) as indicated in the references given in section 1.1.7. Traditionally this kind of information has been derived from existing topographic maps, aerial photo interpretation reports, soil surveys etc.(Smith and Wischmeier, 1957; Foster and Wischmeier, 1974; Foster et al, 1981; Logan et al, 1982; Stephens et al, 1985; Goldman et al, 1986; Mills, 1986; Burrough, 1987; McCool et al, 1987; Wischmeier and Smith and Wischmeier, 1957). The traditional applications of these sources of information in soil erosion are characterised and summarised in table 1.2 based on a modification of Moris-Jones and Kiefer (1978).

In figure 1.2 the human expert is a hidden source because he contributes actively to the process of information extraction by deduction using the available data and his own experience. This role is illustrated by the HUMAN/COMPUTER and HUMAN EXPERT/EXPERT SYSTEM interface boxes in figure 1.2 depicting a model of the processes and flow of information in the soil loss modelling and estimation activity.

Based on figure 1.2 possible improvements to the traditional soil loss estimation approach can be made by:

- Automation of the human expert's contribution into the soil loss estimation and modelling process.

- A knowledge based approach to the incorporation of non-precise or fuzzy information into the soil loss estimation and modelling task.

Broadly stated, the first proposed improvement, requires the storage of the past experience or knowledge of human experts in a form conducive to automated knowledge extraction and decision making. The second recommendation requires an ability to integrate non-precise or vague information with precise data.

The HUMAN/COMPUTER and HUMAN EXPERT/EXPERT SYSTEM interface boxes in figure 1.2 are intended to show those critical areas where either full automation or an open ended man-machine interaction is desirable. A knowledge based solution would enable the critical tasks to be performed by less skilled technicians and would boost productivity.

INFO SOURCE	CONTENT	REQUIRED/USED INFO
Topo maps	terrain morphology, slope and aspect, ground cover type, ground cover area.	qualitative terrain data, slope, length, aspect, qualitative slope info, cover identification, cover type data. qualitative info
Soil maps	soil class, texture, soil distribution.	soil series data, erodibility, soil group, and distribution.
Meteorological reports	rain data, Wind data	rain intensity, distribution and frequency, qualitative rain data, wind speed, direction etc.
Spatial DBMS	same as topo maps, predicted data,	same as topo maps, derived data & parameters, qualitative info.
Space Imagery sensing	same as topo maps, digital image data, classified images,	same as topo maps, qualitative info, metric info,
Aerial photographs	metric info, qualitative photo-interpretation.	same as topo maps and GIS, qualitative info.

Table 1.1: Sources and Uses of Information in Soil Erosion Studies.

Raw data	Modelled Process	Method or Source	Modelled Parameter
Precipitation	Detachment	Precipitation measurement, cloud mapping, Radar sensing.	Kinetic energy, R factor.
Runoff data	Transportation and Detachment	Field measurement.	Runoff Volume, Curve Number (CN).
Wind data	Detachment Transportation	Wind meter, Cloud speed, Dust speed, Deposition distance.	
Vegetative cover.	Rain effect modulation, Conservation practice.	Ground measurements, Aerial photography and Space Imagery.	Canopy density, C factor, Cropping and tillage systems, P Factor.
Morphometric data.	Slope, slope length effects.	Topographic, Photogrammetric, Remote Sensing and Digital mapping.	Slope, Slope length, S factor, L factor.
Soil Data.	Resistance to erosive forces.	Soil survey, Laboratory measurements.	Soil class, K factor.
Anthropogenic factors.	Human and animal impact.	Aerial photo interpretation, Image interpretation, Thematic maps.	Land degradation, Deforestation, C factor, P factor.

Table 1.2: Data requirements for Soil Loss Modelling Parameters

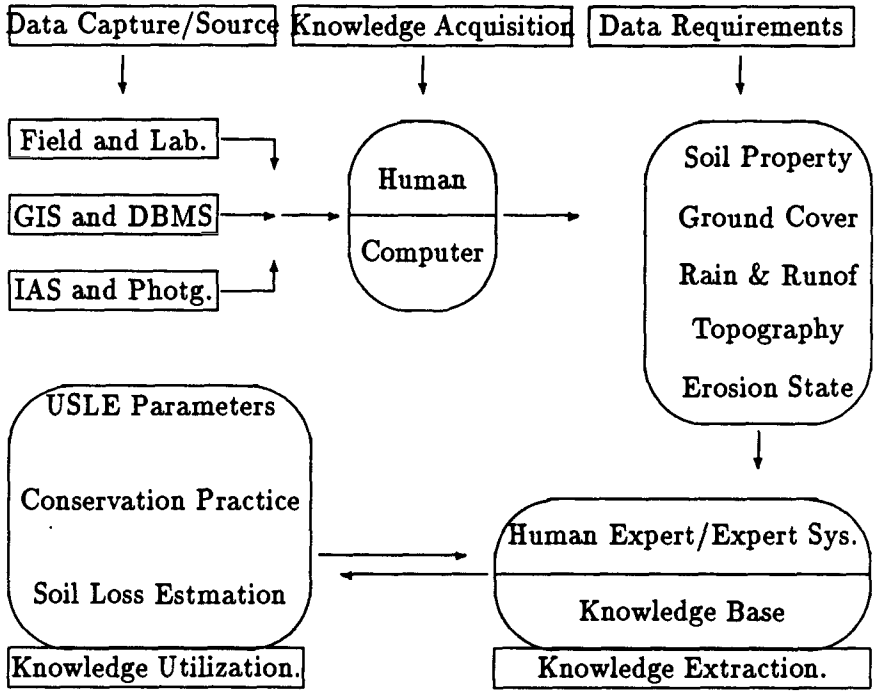


Figure 1.2: A Conceptual Model of the Soil Erosion Estimation and Modelling Process and Knowledge Requirements.

1.2.2 Motivation for the Knowledge Based Approach.

Current research direction in the soil erosion field has been shifting towards the development of expert systems to handle this complex problem (Morrison et al, 1989). In countries where a sufficient number of human experts are available elimination of the physical human expert connection is only needed to increase information throughput. However in developing nations where scarcity of human experts is prevalent, even partial automation of the critical decision making process may mean a lot in terms of increased productivity.

The best defence for introducing a knowledge based solution to the soil loss problem is offered by Rolston's (1988) desirable characteristics of problems for which expert systems technology is reasonable or suggested. Using this criteria the soil erosion problem qualifies for expert systems based solution because:

1. It is a complex problem in which use of fuzzy information is often encountered.
2. It is a problem which requires intensive human expert involvement in its solution, and is normally a lengthy affair.
3. Solutions of specific parts of the problem such as determination of the USLE parameters are well known and documented by erosion domain experts.

Thus the soil loss problem satisfies most of the generally accepted criteria (Rolston, 1988; Luger and Stubblefield, 1989) for expert systems application.

1.3 Review of Existing Knowledge Based Solutions.

At present few knowledge based systems have been designed and implemented for soil erosion estimation and modelling. In general GIS techniques have been successfully applied in characterisation and assessment of land suitability, and soil erosion potential (Morris-Jones and Kiefer 1978; Burrough, 1987; Aidoo, 1987). In these solutions direct human interaction in the analysis process and in drawing final conclusions is critical.

The most recent expert system approach to the soil erosion problem is PLANTING, an expert system developed by Morrison et al (1989) for advising farmers on conservation planting . The expert system incorporates a knowledge acquisition facility for inputting cultivation equipment parameters such as weight, depth of penetration, compaction factor, etc. which have an impact on agricultural soil erosion (Morrison et al, 1989). It uses knowledge about soil characteristics, weather, tillage practice, tolerable soil erosion, soil loss conservation efforts etc. to advise farmers on optimal equipment selection and cultivation method which will produce the least erosion intensification.

The required raw data for the estimation of soil loss by the USLE such as soil characteristics is kept in a consolidated soil data file by the USDA-SCS (US Department of Agriculture and Soil Conservation Service). The rule based system is implemented on the EXSYS expert shell and it consists of the following rules:

- Rules for estimating annual soil loss by the USLE.
- Rules for selecting planting machine components.
- Rules for matching available machines to the selected components.

The system has also an explanation facility to assist farmers in understanding and accepting its advice.

1.3.1 Expert Systems in General.

An expert system is a computer based system that uses knowledge, facts, and reasoning techniques to solve problems that normally require the abilities of human experts (Martin and Oxman, 1988; Luger and Stubblefield, 1989). The process of analyzing, classifying, and designing computer based models of real world knowledge for storage and manipulation by expert systems is referred to as knowledge representation. The implementation of the knowledge representation, knowledge acquisition and compilation in an expert system is called knowledge engineering (Martin and Oxman, 1988; Barr and Feigenbaum, 1982; Barr et al, 1989; Luger and Stubblefield, 1989). The intelligent behaviour of expert systems can be attributed to three important factors:

- Appropriate knowledge representation structures for capturing the semantics or inherent hierarchical organisation of real world knowledge.
- Inference strategies and mechanisms which exploit the hierarchical knowledge structures to provide intelligent answers to queries by a human-like reasoning process.
- An ability to explain the reasoning process of the inference mechanism.

Armed with these capabilities intelligent programs are able to mimic the human reasoning processes and display intelligent behaviour.

Generally expert systems consist of, a user friendly interface, a knowledge acquisition facility, an explanation facility, an inference engine and a knowledge base of facts, rules or other data structures (Barr and Feigenbaum, 1981, 1982; Martin and Oxman, 1988; Barr et al, 1989; Luger and Stubblefield, 1989). On the basis of the knowledge representation format used expert systems are characterised as, logic programming systems (Martin and Oxman, 1988; Luger and Stubblefield, 1989), rule-based systems (Barr et al, 1981, 1982; Martin and Oxman, 1988; Luger and Stubblefield, 1989), frame systems (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Luger et

al, 1989), semantic network systems (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Luger and Stubblefield, 1989), blackboard architectures (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Nii, 1989; Argialas and Harlow, 1990).

Expert systems can also be characterised by the control strategy employed by the inference mechanism. Goal driven strategies make use of backward chaining to control the search path (Barr and Feigenbaum, 1981, 1982; Martin and Oxman, 1988). Data or event driven control systems make use of forward chaining search control algorithms (Barr and Feigenbaum, 1981, 1982; Luger and Stubblefield, 1989). Expert systems based on logic programming languages such as PROLOG use various forms of the unification algorithm to facilitate search control (Barr and Feigenbaum, 1981; Cohen, 1985; Luger and Stubblefield, 1989).

Other search control strategies used in expert systems include breadth-first and depth-first search methods, various forms of heuristic search, pattern matching, problem reduction, and hierarchical control (Barr and Feigenbaum, 1981, 1982; Martin and Oxman, 1988). In practice individual search control strategies may be augmented by one or more of the methods mentioned above (Cohen, 1985; Martin and Oxman, 1988). Alternatively search control strategies may be acquired automatically by the expert system through a learning process (Davis and Lenat, 1982; Minton, 1988). The search control strategies mentioned above are not discussed further in this study but their details may be found in the referenced literature.

Expert systems are usually implemented by means of special expert systems development languages and tools. The most famous among the early expert systems development tools was LISP which was used to implement systems such as MYCIN, E-MYCIN, PROSPECTOR and ELIZA (Barr and Feigenbaum, 1981, 1982; Martin and Oxman, 1988). In recent times this language has not fared well outside the AI community because it requires special hardware (Barr and Feigenbaum, 1981, 1982).

In general expert systems development tools can be divided into four categories. In the first category are low level programming languages such as C, Fortran, and

Pascal (Martin and Oxman, 1988). The second category consists of languages such as LISP, PROLOG, OPS5, C++, and SMALL TALK (Martin and Oxman, 1988) which are usually 4GL programming languages. Expert system shells such as E-MYCIN, ART, KEE, EXSYS, and PC Plus (Martin and Oxman, 1988; Barr et al, 1989) constitute the third category in the hierarchy of expert systems development tools. The fourth level in this hierarchy consists of computer assisted software engineering (CASE) environments (Martin and Oxman, 1988; Barr et al, 1989).

Because 4GL languages such as LISP are interpreted languages and they generally produce slower executable codes, the faster C and C++ programming languages have become important in the implementation of commercial systems. The general trend is therefore to conceive and design and prototype the expert system around some 4GL language and then implement it in C or C++ for delivery.

A list of current (1988) commercially available expert systems and their implementation languages organised by domain of application may be found in Martin and Oxman (1988). A similar list can also be found in Barr et al (1989)

1.3.2 Handling of Uncertainty in Expert Systems in General.

Cohen (1985) identifies several techniques used to manage uncertainty in knowledge based systems. These methods can briefly be summarized as, the engineering approach, diversification or assumptions heuristics, parallel certainty factors, the control approach, support justification and the theory of endorsement (Cohen, 1985).

The engineering approach is a general purpose solution similar in spirit to techniques of "error elimination" in the surveying sciences which can be achieved by adopting ideal models of uncertainty. The diversification approach is also referred to as the assumptions based approach (Cohen, 1985). This approach requires the

uncertainty to be "shared" among the pieces of uncertain evidence when no knowledge is available about the uncertainty of any specific piece of evidence. Solutions employing this approach include the assumptions based truth maintenance system by De Kleer (1987). The control structure approach makes use of knowledge about the nature of the uncertainty in the domain knowledge to develop appropriate search control heuristics which are not affected by the inherent uncertainty. Practical systems based on this approach include HEARSAY-I, HEARSAY-II and generally all black board architecture systems (Barr and Feigenbaum, 1981, 1982; Cohen, 1985; Barr et al, 1989). Support justification is an approach considered to give better performance because unlike the parallel certainty factor approach it separates the reasons for believing from reasons for not believing a piece of evidence (Cohen ,1985). A practical implementation of this approach is the truth maintenance system discussed in Doyle (1987).

The model of endorsements introduced by Cohen (1985) has superficial similarities to the support justification method of Doyle (1987). It however differs fundamentally in the fact that whereas support justification summarises all the reasons for believing (disbelieving) a piece of evidence by a single real number, the endorsement approach keeps a history of all the reasons for believing (disbelieving) each piece of evidence. The pro and con endorsements are then employed by the inference mechanism to facilitate reasoning with uncertainty (Cohen, 1985).

In the classical probability method uncertainty is assumed normative (Cohen, 1985) and propagation is achieved by taking the product of all the certainty factors of the rule premises. As explained in Cohen (1985), Barr and Feigenbaum (1981, 1982), Barr et al (1989) there is little justification for assuming that the uncertainty associated with rule-based expert systems is normative because in general the rule premises are not independent and may not even be exclusive and exhaustive as required by the probability theory assumptions.

1988; Zadeh, 1989):

1. Computing the meaning of linguistic fuzzy variables.
2. Handling of uncertainty associated with fuzzy variables.
3. Evaluating vague natural language query lists.
4. Designing and using rules and meta rules about domain knowledge to facilitate control of the computer reasoning process.

Computation with fuzzy variables is however generally complex, and requires the introduction of the necessary fuzzy arithmetic into the computer before their application (Shmucker, 1984; Yamakawa, 1988). As it will be seen in this research certain assumptions and simplifications can be used to permit direct use of non-precise, fuzzy information in performing database searches. The resulting method, called the fuzzy geometric partitions method of vague data representation, is discussed in chapter 4.

1.4 General Specifications for the SLEMS.

Based on the objectives introduced in section 1.1.6 the proposed Soil Loss Estimation and Modelling System (SLEMS) must address the following specific problems:

1. Handling of non-precise data and information available in report form, from various sources with a view to maximizing its usefulness in areas with inadequate data.
2. Knowledge extraction from non-precise sources for application to problem solving in situations where human expertise is scarce.
3. Integration of precise and non-precise data within an updatable database or knowledge base to enable revision of the knowledge base as more or precise information becomes available.

Using Cohen's argument (Cohen 1985), both the parallel certainty factors approach and the support justification approach suffer from the disadvantage that after a domain expert has assigned the numerical values, the reasons which he used to arrive at the specific values will no longer be available to the system to facilitate reasoning. The advantage of the other methods is however that, they require less complex implementation strategies (Cohen, 1985). The method of parallel certainty factors was therefore very popular in many of the original expert systems and expert shells such as MYCIN, E-MYCIN and PROSPECTOR (Cohen, 1985, Barr and Feigenbaum, 1981, 1982; Barr et al, 1989) and it has been chosen for use in the EXPERT subsystem.

1.3.3 Handling of Fuzzy Qualitative Information in General.

Qualitative and vague information is frequently used in soil loss estimation and modelling problems (Bali, 1977; Hóly, 1980; Morgan, 1986). Qualitative information is defined (Barr et al, 1989) as information for which no precise measures or quantification are made. Usually vague data is specified in terms of fuzzy natural language expressions. Alternatively fuzzy entities may be described in terms of vague numeric ranges and fuzzy probabilities (Zadeh et al, 1975; Baldwin, 1979; Kandel, 1979; Zadeh, 1979; Dubois and Prade, 1980; Kandel, 1986).

Until the introduction of the fuzzy set theory by Zadeh (Zadeh et al, 1975; Dubois and Prade, 1980) computations involving vague information were not possible. Using the new method however computers can be used to manipulate and compute with non-precise qualitative information to facilitate solution of complex problems (Adamo, 1980; Giles, 1980; Baldwin, 1986; Zenner, 1985; Wenstøp, 1979). Fuzzy mathematical techniques facilitate computer based application and use of non-precise information by offering a framework for (Zadeh et al, 1975; Dubois et al, 1980; Klir and Folger,

1988; Zadeh, 1989):

1. Computing the meaning of linguistic fuzzy variables.
2. Handling of uncertainty associated with fuzzy variables.
3. Evaluating vague natural language query lists.
4. Designing and using rules and meta rules about domain knowledge to facilitate control of the computer reasoning process.

Computation with fuzzy variables is however generally complex, and requires the introduction of the necessary fuzzy arithmetic into the computer before their application (Shmucker, 1984; Yamakawa, 1988). As it will be seen in this research certain assumptions and simplifications can be used to permit direct use of non-precise, fuzzy information in performing database searches. The resulting method, called the fuzzy geometric partitions method of vague data representation, is discussed in chapter 4.

1.4 General Specifications for the SLEMS.

Based on the objectives introduced in section 1.1.6 the proposed Soil Loss Estimation and Modelling System (SLEMS) must address the following specific problems:

1. Handling of non-precise data and information available in report form, from various sources with a view to maximizing its usefulness in areas with inadequate data.
2. Knowledge extraction from non-precise sources for application to problem solving in situations where human expertise is scarce.
3. Integration of precise and non-precise data within an updatable database or knowledge base to enable revision of the knowledge base as more or precise information becomes available.

4. Ability to serve as a cheap repository for knowledge gathered from human domain experts and its use as a consulting system (advisor) for less skilled technical personnel.
5. Provision of conventional database management capability for storing and manipulating non vague data required by domain experts.

In the course of pursuing solution to these problems a system with the following subsystems was conceived and implemented:

1. The SLEMS database and database management sub-system or the CDATA based SLEMS DBMS subsystem.
2. The SLEMS knowledge acquisition subsystem called the LEARN subsystem
3. The SLEMS knowledge manipulation subsystem or the EXPERT subsystem.
4. The SLEMS fuzzy comparison operator or the FUZZ subsystem.

The main body of the thesis therefore, consists of the design, implementation, and demonstration of their application to the solution of soil erosion related problems.

1.4.1 The SLEMS DBMS

In developing the SLEMS DBMS subsystem certain general database development principles were applied. These principles as outlined in Whittington (1988), Longstaff (1984), Grundy (1985), Yao et al (1982), Kambayashi et al (1978) and others include:

1. Data analysis
2. Data structure.
3. Data validation.

4. Data security and Concurrency control.
5. Data definition Language.
6. Data manipulation Language.
7. Software development environment.

Data analysis generally involves data normalization which is an iterative process with four elements (Dutka, 1989):

- Identification of initial attributes.
- Specification of dependencies.
- Grouping of attributes.
- Selection of primary keys.

As it will be evident later, the decision to make use of an existing database management package (the CDATA) for the SLEMS DBMS removed the need to perform the details in items 2 to 7 above. The discussion on database management in this thesis is therefore restricted to the data analysis issues with particular attention to the specification and implementation of the database schema.

1.4.2 The SLEMS Knowledge Based Subsystems

The design of knowledge based systems requires familiarisation with expert systems (ES) and artificial intelligence (AI) concepts. General specifications for expert systems are not yet common (Green and Keyes, 1987, Ebrahimi, 1987). However some basic principles regarding basic system components have to be followed. The provision (Rolston, 1988; Luger and Stubblefield, 1989) that expert systems must satisfy users of the validity of their answers requires the ES to be equipped with reasoning and explanation mechanism. Briefly, the important issues for consideration in designing knowledge systems include (Luger and Stubblefield, 1989):

1. Knowledge representation.
2. Truth maintenance (consistency enforcement).
3. Knowledge manipulation language.
4. Inference mechanism.
5. Explanation facility.
6. Output facility.

Although the knowledge based subsystems were all implemented from scratch using the C language, considerable advantage was gained by making use of available source code (Schildt, 1987) for some important modules of the EXPERT and the LEARN subsystems. Essential modifications were made to the existing code and new modules added to facilitate the desired performance. The performance objectives used as guideline for the design and implementation of the knowledge based subsystems are:

1. To facilitate domain independent input of data and knowledge into the SLEMS knowledge base.
2. To provide the means for introducing domain expert knowledge on problem solving.
3. To provide a simple enough user interface which will allow subsequent use and consultation of the knowledge base by non-expert personnel.
4. To facilitate intelligent query of the SLEMS knowledge base using natural language.
5. To enable useful response to incomplete or vague queries.
6. To facilitate user acceptance of the SLEMS query solutions by providing explanation of the solution.

These are general requirements of expert systems (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Luger and Stubblefield, 1989) but they were also dictated by

the observation that in most developing countries where foreign experts are frequently employed, the knowledge brought by the expert usually departs with him at the end of his tenure. Most of the instructions given to local technicians are usually abbreviations of the experts knowledge, which are inadequate to facilitate continued guidance after the experts departure.

One would therefore like to have a facility which can capture and store the experts knowledge for continued reference after his departure. Such a facility should therefore have a simple enough interface to allow the less skilled personnel to consult it. In addition, some developing countries have vast unused amounts of information in the form of aerial photographic coverage. These can be valuable source of photo-interpretation data for the solution of soil loss related problems.

With this short introduction to the nature of the soil erosion problem and the proposed knowledge based approach to its solution the discussion now focuses on the technical issues of system design, development, implementation and testing. Because the knowledge based approach is a recent development (Christofer, 1987; Ebrahimi, 1987; Lowry, 1989) the steps taken in undertaking the various facets of the development of the SLEMS subsystems is necessarily ad-hoc.

Chapter 2

Design and Development of the SLEMS DBMS

2.1 Introduction.

The SLEMS database management subsystem is based on the Cheap Database or CDATA (Stevens, 1987). Decision to use this particular system followed from the evaluation of three alternatives:

1. Use of facilities existing within database management and GIS systems at UNB such as the CARIS, the ARC-INFO, and the PCI EASI-PACE.
2. Development of a database management system from scratch using the Sun 4 UNIX Workstations software development facilities.
3. Acquisition of a simple DBMS package and its modification to suit the desired needs.

After examining and weighing each of these alternatives the CDATA was chosen and purchased because:

1. It is a relatively cheap database which comes with source code and manual for the price of a book.
2. The CDATA is written in the C programming language which was chosen as the programming language for the project because of its versatility, it's availability on the Sun 4 Workstation and availability of published source code on other topics of interest (Schildt, 1987).
3. The package has built-in portability for a number of C compilers, specifically; Ci C86, Datalight C, DesMet C, ECO-C88, High C, Lattice C, Microsoft C, Mix C, QC88, Turbo C, Let's C, Whitesmith's C and Wizard C (Stevens, 1987).
4. The CDATA carries no copy right restrictions, giving the user freedom of use and modification to suit specific needs (Stevens, 1987).

The main disadvantage of the CDATA[1] is that it was developed to run on PC DOS. This was however considered to be a minor detraction to its advantages. The first task was therefore to modify it so that it could run on the UNIX based Sun 4 Work station. Further modifications were then made to tailor it to the specific needs of the thesis research.

[1]The CDATA Database Development book and source code disk is available from,
Management Information Source, Inc.
P.O. Box 5277,
Portland, OR 97208-5277.

The Small Computer Book Club,
279 Humberline Drive,
Rexdale, ON, M9W 6L1

for a total cost of $43.95 at 1989 prices.

2.2 The CDATA Database Management Package.

The CDATA is an indexed relational database management package (Stevens, 1987). It uses the B-tree data structure to manage index files which contain the necessary information for accessing the data stored in the database files. Figure 2.1 shows the general scheme of the index file structure. Details of the B-tree algorithm used in the CDATA are available in Stevens (1987). General topics on B-tree data structures can also be found in Overmars (1983) and Pfaltz (1977).

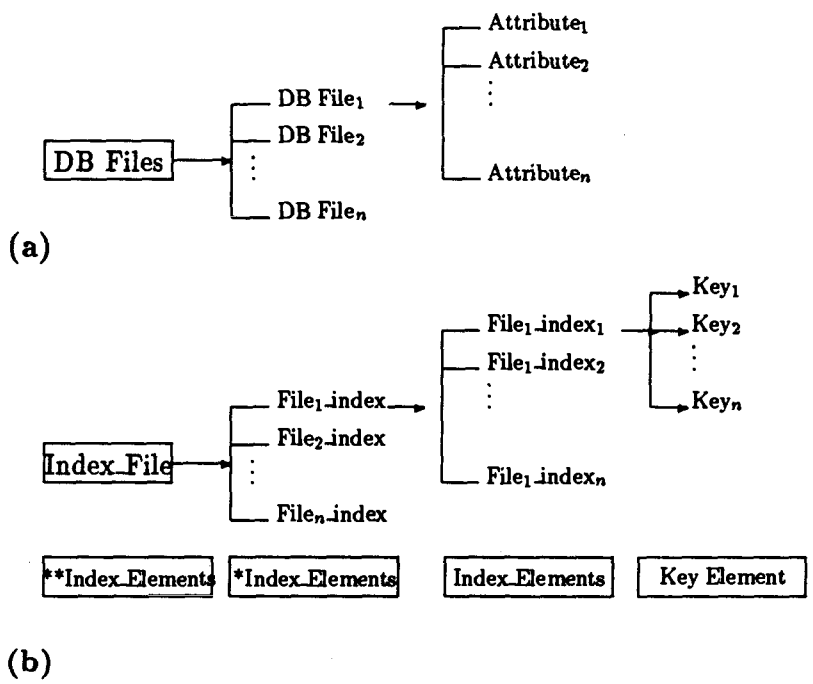


Figure 2.1: (a). The CDATA Data File System Organisation. (b). The CDATA Index File System Structure (After Al Stevens, 1987, page 121).

The main components of the CDATA are (figure 2.2):

1. The CDATA schema and schema compiler.
2. The database manager.
3. The index file manager.
4. The user interface.
5. The database

The CDATA schema facilitates the following database management functions:

- Data definition language (DDL).
- Data manipulation language (DML).

The DDL and DML facilitate a consistent and unified framework for representing real world data and processing transactions on the database (Stevens, 1987). The database manager takes care of the actual transactions on the database. It ensures and enforces validation and consistency during updates on the database. The index file manager facilitates efficient database searches and database updates by keeping track of database files and content. It also provides functions which support insertions, retrieval, and deletions from the database.

The User interface consists of three components:

- The screen manager.
- The query processor.
- The editor.

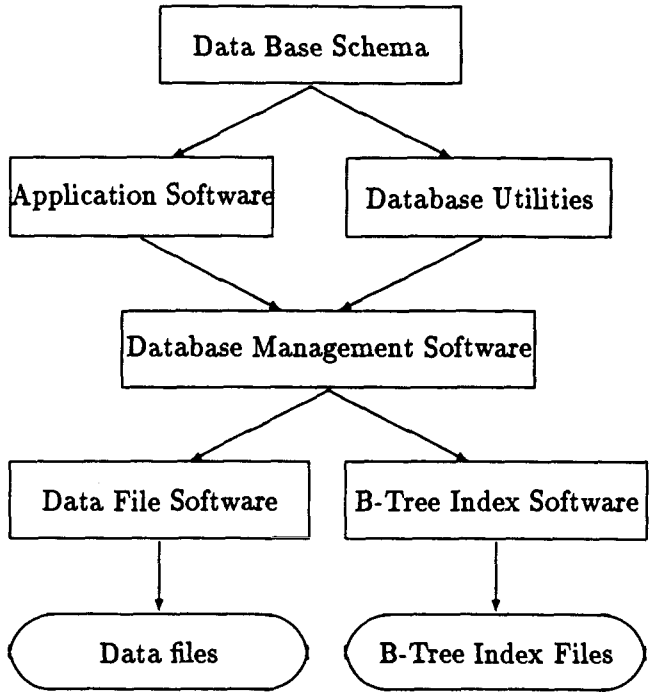


Figure 2.2: The Arctecture of the CDATA (After Al Stevens, 1987,page 122).

The screen manager consists of functions needed for initializing the display screen and displaying database contents during transactions or editing. The query processor facilitates the translation of user queries into the systems internal data definition language for the purpose of accessing the database contents. The Editor supports functions necessary for data input and updates on existing data files.

2.2.1 Modifications on the CDATA.

Two types of modifications were performed on the CDATA to facilitate its installation on the Sun 4 Workstation:

- Portability related modifications.
- Functional modifications.

Portability modifications were performed on the system definition code to facilitate its compilation using the Sun 4 Workstations based UNIX C compilers (SUN4, 1988).

Functional changes were made to the schema as follows:

- Extension of the data model to facilitate generation of data access and editing functions.
- Extension of the data model to recognise system file hierarchy.
- Modification of the schema compiler to support the extended CDATA data model.

The original CDATA schema compiler does not support file hierarchy, all files are treated as flat files unrelated to each other (Stevens, 1987). For the purposes of this research a hierarchical embedding of the relational file structure was desirable as it could be exploited by the knowledge based subsystems. Modification was therefore necessary to enable differentiation of system level files from the data files. The third modification came as a consequence of the first two modifications. It involved rewriting certain portions of the CDATA schema compiler and augmenting it with routines for generating the edit and data access routines as mentioned above.

2.2.2 The CDATA's Data Types.

Four types of data are recognised by the CDATA schema corresponding to the numerical, character, date, and currency domains. These data domains are used by

the schema to perform data validation (Yao and Navathe, 1978; Stevens, 1987). The definition of a CDATA record involves specifying the data element name, the data element domain or data type, the data element field length, and the data element display mask. Special data fields to be used as keys for accessing the database are designated as key elements. The modified schema permits specification of concatenated keys. Valid keys may be formed from up to 3 concatenated elements. Up to 5 keys may be specified per file.

The display mask is used as control for the data input and editing operations. It also serves as a template for database content displays.

2.3 Data Base Prototyping With the CDATA.

Data base prototyping is a process whereby a database schema is designed and then transformed (implemented) into executable code (Stevens, 1987; Dutka, 1989; Whittington, 1988). When appropriately designed the schema ensures and guarantees the integrity of the database and correctness of subsequent database transactions (Loizou and Thanisch, 1984; Stevens, 1987; Whittington, 1988).

Production of the schema requires careful planning for it must satisfy specific requirements for a relational database (Stevens, 1987). In general relational databases must satisfy the relational normal forms (Hotaka, 1978; Kahn, 1978; Kambayashi, 1978; Smith, 1978; Dutka, 1989). In general the design of relational databases is a difficult process requiring complicated dependency analysis of the data and file elements (Feldman, 1984; Loizou and Thanisch, 1984; Dutka, 1989). To facilitate consistency and efficiency this process usually involves automatic dependency analysis (Chen, 1978; Hubbard 1978; Kambayashi, 1978; Hellwig, 1980; Feldman and Fitzgerald, 1985; Loizou and Thanisch, 1984).

Specifically the following subset of the relational database requirements were satisfied in the selection of data elements, file elements and key elements for the SLEMS

schema:

1. Elimination of duplicate key elements.
2. Maintenance of the relations among related data elements in different files.
3. Elimination of hierarchical dependencies between files by introducing connector files where necessary.

Thus the final database schema satisfies the third normal form, and therefore, supports lossless insertion and deletion (Whittington, 1988; Dutka, 1989). Figure 2.3 shows an excerpt of the SLEMS database schema produced by the application of these principles. The steps involved in prototyping the SLEMS Data Base consisted of:

1. Identification of the database files.
2. Identification of database data elements.
3. Identification of key data elements.
4. Compilation of the data element dictionary.
5. Production of the SLEMS schema.
6. Compilation of the schema into the executable database.

After design and production of the schema the database was prototyped in a five phase process:

1. Generation of the source code for data element symbol lists, file and record structures.
2. Production of source code for data element dictionary, and database file lists (C #define statements) and the display masks.
3. Production of source code for index file structures and pointer arrays.

```
#schema SLEM_SCHEMA
#dictionary
#system

;       Define the SLEMS and Component subsystems
        EXPERT,         S,  11, "———"
        LEARN,          S,  11, "———"
        FUZZ,           S,  11, "———"
        DATA_BASE,      S,  11, "———"
#end system
;       Define the DATA_BASE files
        COVER_DATA,     F,  11, "————"
        TOPO_DATA,      F,  11, "————"
        RAIN_DATA,      F,  11, "————"
        RUNOFF_DATA,    F,  11, "————"
        :
;       Define the topographic data elements
        REC_ID,         N,  7,  "——"
        POLY_ID,        N,  7,  "——"
        LABEL,          A,  8,  "——alpha-num."
        AVERAGE_SLOPE,  N,  5,  "—.-%"
        :
#end dictionary
;       Define the file systems
#directory SLEMS_SYS
        EXPERT
        LEARN
        FUZZ
        DATA_BASE
#end file
;
#directory DATA_BASE
        COVER_DATA
        TOPO_DATA
        RAIN_DATA
        :
#end file
        :
;       Define the file access keys
#key TOPO_DATA REC_ID, POLY_ID, LABEL
#key TOPO_DATA REC_ID
#key TOPO_DATA POLY_ID
#key TOPO_DATA LABEL
        :
#end schema SLEMS_SCHEMA
```

Figure 2.3: An Excerpt of the SLEMS Schema Implemented After Al Stevens 1987, pp. 88

4. Production of source code for the data access functions.

5. Generation of source code for the edit functions.

The compilation stage transforms the conceptual database, from schema representation (figure 2.3), into machine readable C source code. Each phase involves the compilation of the database schema using the modified CDATA schema compiler with appropriate options. Options are specified by typing the command <*schema*> <SCHEMA_FILE> $-n$ where *schema* is the executable schema compiler, SCHEMA_FILE is the file containing the database schema in the data definition language (DDL) and $n = \{1, 2, 3, 4, 5\}$ is the desired phase[2].

At the end of these processes five files constituting the SLEMS database's compilable code are produced (see Appendix I). These files are then combined with the database and datafile manager, the screen manager, the index file manager, the display module and editor, and any other SLEMS utility source code files. After compilation and liking of the DBMS source code the executable SLEMS DBMS is produced. To simplify and maintain consistency the whole process of compiling the SLEMS schema and all the system modules has been automated by means of a **makefile** created using the **make** facility of the Sun 4 UNIX workstation programming utilities (SUN4, 1988). The **makefile** contains rules for validating source code files, compiling and liking executable objects.

2.3.1 SLEMS Data Base User Interface.

The user interface of the SLEMS DBMS facilitates the following simple database manipulation functions:

1. Display of the contents of a file.

2. Display of the contents of specified data elements from a specified file.

[2]In the original schema there are only three phases

3. Creation of a new record.
4. Protected/unprotected mode editing of record contents.
5. Deleting file records.
6. Generation of reports from specified files.

The edit functions perform appropriate justification of numeric and text data entries automatically. They also provide checks for duplicate record entries and data domain violation errors. During editing the cursor motion is restricted to only valid field locations. Figure 2.4 shows the display template during an editing session.

```
                    — LEARNING_OBJECTS —

SESSION ID          __________
SUBJECT             ____________________________
VERB                ____________________________
OBJECT              ____________________________
SUBJECT DOMAIN      ____________________________
LEARN STATUS        _ (0,1) only
LOCATION            ____________________________
CERTAINITY          ____ [0-1] decimal%
ATTRIBUTES          ____________________________
```

Figure 2.4: Data Display Template of the SLEMS DBMS

Because the SLEMS DBMS is only regarded as a support facility for the knowledge based subsystems it will not be discussed further in this thesis.

Chapter 3

Principles of Intelligent User Interface Design.

3.1 Introduction.

In chapter one the objectives of the research, data requirements, and general outline of the proposed soil loss estimation and modelling system (SLEMS) were introduced. This chapter first develops and extends the concepts and proposals introduced in the first chapter and then applies them to the development of the LEARN subsystem. In the first two sections of the chapter the necessary knowledge based systems design concepts are discussed in more detail.

Specifically section two deals with principles of knowledge based systems. Section three is devoted to the design and implementation of the LEARN subsystem. Section 4 contains a short demonstration of the LEARN system's application in knowledge acquisition and query processing. The last section briefly introduces the problem of representation of vague or fuzzy data and information which is discussed in more detail in chapter 4.

3.1.1 Basic Principles and Elements of Knowledge Based Systems.

In artificial intelligence (AI) terminology intelligent interfaces are referred to as intelligent front-ends because they are usually patched onto existing database management systems (Minker, 1980). When knowledge based techniques are applied integrally in the development of an information processing system the resulting system is referred to as a Knowledge Based System (KBS). A knowledge based system capable of emulating a human expert within a narrow domain of application is called an expert system (ES)(Tanimoto 1987; Barr et al, 1989).

Artificial intelligence and expert systems technology elements are the basic components of intelligent interfaces. They provide intelligent data and knowledge representation, and facilitate knowledge manipulation in conventional database management systems.

A major difference between conventional programs and knowledge based systems is the separation of procedural knowledge from the declarative knowledge (Martin and Oxman, 1988; Barr et al, 1989; Luger and Stubblefield, 1989). The declarative component of the system consists of the rules or conditional facts residing in the knowledge base. The knowledge base also contains unconditional facts or assertions about domains of interest. More specifically, ES differ significantly from DBMS in two aspects (Barr et al, 1989; Luger and Stubblefield, 1989):

- The ability to explain reasons for actions performed on the knowledge base.
- The knowledge base of an ES is executable while DBMS can only be queried or updated.

Expert systems are characterised by the following theoretical, design and performance characteristics (Barr and Feigenbaum, 1981, 1982, Barr et al, 1989; Cohen et al, 1982; Tanimoto, 1988; Luger and Stubblefield, 1989):

1. They have as their theoretical foundation logic, information theory (IT), and information engineering (EI).
2. Their practical foundation is artificial intelligence (AI) and database management systems (DBMS).
3. They have the following basic components:
   (a) A user friendly user interface.
   (b) A knowledge base consisting of facts and rules on problem solving strategies in a narrow and well defined area of expertise.
   (c) An inference engine which controls selection and execution of problem solving procedures and knowledge base searching.
   (d) An explanation facility to provide users with reasons in support of the systems recommendations or failure to find solutions to user queries.
   (e) An output facility for display and report generation.
4. Generally the knowledge or fact base content consists of:
   (a) Knowledge about knowledge or problem solving strategies (meta knowledge).
   (b) Domain specific knowledge generally prototyped by use of an Expert Shell (an ES with no domain knowledge) through a knowledge engineering process.
5. Basic knowledge representation structures which may be one of:
   (a) Rule based knowledge structures, in which knowledge is captured in the form of production rules or *If ... Then .... rules*(Barr and Feigenbaum, 1981, 1982, Barr et al, 1989; Cohen et al, 1982).
   (b) Frame structures representing a hierarchy of objects and the attributes of objects that can be assigned, inherited from another

frame, or computed through procedures (Barr and Feigenbaum, 1981, 1982; Cohen et al, 1982; Luger and Stubblefield, 1989). Attributes of the object are filled into slots which have the same function as record fields. In this structure knowledge is introduced as structured facts and relations among the knowledge components.

(c) Predicate logic which captures knowledge in the form of predicates or Horn clauses (Minker, 1980; Kowalski, 1984). In Predicate logic-based systems, facts about real world objects and concepts are expressed as predicates (Minker, 1980; Schmidt, 1985). Complex objects are formed by applying logical connectives to atomic predicates. Solutions to queries are obtained by evaluating the truth value of the complex statement by means of truth tables or other strategies.

(d) Semantic networks which capture the inherent hierarchical organisation of real world knowledge. Semantic networks capture knowledge in graphical like structures in which nodes represent objects and inter-node links represent inherent relationships between objects. Relationships typically used as links include *is_a*, *is_part_of*, *etc.* (Fahlman, 1979; Feldman and Fitzgerald, 1985; Schmidt, 1985; Martin and Oxman, 1988). Inheritance relationships are easy to implement with semantic network knowledge representation.

(e) Scripts which are powerful knowledge engineering tools for representation of conceptual knowledge, suited for non-mathematical applications such as text understanding etc.(Barr and Feigenbaum, 1981, 1982; Barr et al, 1989).

(f) Blackboard architectures consisting of complex hierarchical structures whose components may be production rules, frames or semantic networks (Barr et al, 1989; Nii, 1989). The class of black board systems (BBS) differs from the general ES types, in the manner solutions to complex queries are processed. In these systems, solutions to problems are found by cooperating processes (inference engines or experts). Each expert determines when it can make a useful contribution to the solution process. These systems, therefore, use opportunistic rather than deterministic solution strategies (Barr et al, 1989). A blackboard structure represents a hierarchy of classes of the knowledge needed to achieve the desired goal (Barr et al, 1989).

(g) Neural networks which are based on a biological paradigm of knowledge representation and manipulation (Castelaz et al, 1987; Barr et al, 1989; Luger and Stubblefield, 1989)

6. Machine Learning or the ability for the system to achieve self improvement by:

   (a) Rote learning which only involves memorization by the system, of procedures for task performance and control (Schildt, 1987; Minton, 1988; Luger and Stubblefield, 1989).

   (b) Cognitive concept learning which involves learning of new concepts from good and bad examples (Barr and Feigenbaum, 1981, 1982; Schildt, 1987).

   (c) Explanation-based learning (EBL) systems, which are essentially cognitive concept learners able to generate their own examples by analyzing success or failure to solve problems (Minton, 1988).

   (d) Learning by analogy (Minton, 1988).

(e) Learning by discovery based on a few basic axioms and theorems about the domain of knowledge (Davis and Lenat, 1982).

7. Inference schemes and search strategies for the control and execution of the reasoning processes of the expert system, such as:

   (a) Backward chaining also referred to as goal driven strategies (Barr and Feigenbaum, 1982; Minton, 1988; Luger and Stubblefield, 1989).

   (b) Forward chaining or event driven strategies used to specify and control the manner in which the rules in a KBS are executed.

   (c) Unification which involves matching and substitution of variables in objects represented as list structures (Tanimoto, 1987; Walker, 1987).

   (d) Pattern matching procedures used to support most of the inferencing and search procedures in production systems (Barr and Feigenbaum, 1981, 1982).

   (e) Search strategies including, depth first and breadth first methods, and heuristic searches such as shortest path, hill climb, difference reduction, means and ends analysis, hierarchical generate and test methods (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Tanimoto, 1988; Luger and Stubblefield, 1989).

8. Handling of uncertainty or reasoning with uncertain knowledge by various strategies including:

   (a) Certainty factors (CF) (Cohen, 1985; Gale, 1986).

   (b) Possibilistic reasoning systems (Zadeh et al, 1975; Zadeh, 1979; Gale, 1986; Kandel, 1986; Klir and Folger, 1988; Zadeh, 1989).

   (c) Truth maintenance systems (TMS) (Cohen, 1985; Doyle, 1987; De Kleer, 1987).

(d) The theory of endorsements (Cohen, 1985).

(e) Support justification and assumptions (Doyle, 1987; De Kleer, 1987)

Some ES systems are characterised by their ability to learn. Examples of learning systems are PRODIGY (Minton, 1988), and Lennat's FM (Davis and Lenat, 1982) . What characterises learning systems is their ability to perform automated acquisition of domain knowledge. The systems are therefore capable of improving their performance through deduction and inductive inference. Cognitive concept learning systems involve the use of examples and counter-examples of the concept to be learned. The learner uses these examples to make generalizations about the new concept (Michalski et al, 1986; Schildt, 1987). Explanation based learning which includes learning from mistakes and learning from success (Michalski, 1986; Minton, 1988; Barr et al, 1989) involves the analysis or "explanation" of the reasons for success or failure to achieve a goal. Learning from success involves the learner remembering rewarding moves used in the solution and generalizing the solution for future reference. Learning from mistakes is a process in which the learner determines reasons for failure to achieve a desired goal and uses them to correct or refine the solution strategy (Minton, 1988).

The method of learning by analogy involves use of domain knowledge and similarity or dissimilarity among the objects to infer new pieces of knowledge about domain objects in a different context (Cohen et al, 1982; Minton, 1988; Barr et al, 1989).

Learning by discovery was first employed in Lenat's FM (Davis and Lenat, 1982). The major characteristics of this method of learning is that it involves an automatic analysis of the knowledge base contents for the purpose of discovering inherent knowledge structures. Discovered structures are used to generate conjectures on interesting theorems (Davis and Lenat, 1982). Discovery systems use either exhaustive search and permutation of the knowledge base objects in search of interesting conjectures or they perform an informed search using heuristics to weed out uninteresting conjectures (Davis and Lenat, 1982).

Handling of the uncertainty inherent in the facts and rules of a knowledge base system is an important consideration in the design of knowledge based systems. Equally important is the maintenance of the knowledge base integrity through appropriate knowledge updating mechanism. Systems used to facilitate these conditions are generally referred to as truth maintenance systems (TMS)(Doyle, 1987; De Kleer, 1987).

3.2 Knowledge Representation in General.

The most important issue in the design of knowledge systems is knowledge representation (Luger and Stubblefield, 1989). Knowledge representation entails knowledge manipulation in the sense that once the design of the classes of data structures for storing the knowledge has been performed, procedures that allow intelligent manipulation of these data structures can be developed. Knowledge representation is defined (Luger and Stubblefield, 1989) as:

> The identification, of significant objects and relations in the domain of discourse and their mapping into a formal language such that the resulting representation scheme contains sufficient knowledge to support solution of problems in the domain and to facilitate correct inferences from the knowledge efficiently.

Knowledge Representation is also defined in Barr et al (1989) as the combination of data structures (representing facts or rules) and interpretative procedures that, if used in a program will lead to knowledgeable behaviour. Important issues to consider in the planning and design of knowledge representation schemes include, representation language, differentiation between virtual knowledge and explicit knowledge (facts), representation of meta-knowledge (knowledge about knowledge), inheritance relations, default values and exceptions (Cerone, 1980; Minker, 1980; Doyle, 1987).

Procedural attachment to object descriptions in frame and rule based systems is also considered (Luger and Stubblefield, 1989) an important trend in the development of knowledge representation languages. In general knowledge representation addresses

the following issues (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Luger and Stubblefield, 1989):

1. **Knowledge application:** The representation strategy must address the eventual use of the facts and knowledge represented by the knowledge structures in the sense of:
   (a) Retrieval of facts and data required for solving specific queries from the knowledge base.
   (b) Reasoning with uncertainty in the knowledge and facts during the search for solutions.
2. **Retrieval guidelines:** Retrieval must address the issue of the relevance of particular knowledge to specific problems in a situation when the problem solver or learner is presented with, or knows too many things (Barr and Feigenbaum, 1982; Barr et al, 1989; Cohen et al, 1982). Possible solutions to handle such conflicts includes (Luger and Stubblefield, 1989):
   (a) Linking or networking of related concepts and objects at the time of acquisition. This is only possible if it is known in advance that one structure entails another.
   (b) Lumping structures together if they are to be used together in subsequent solutions.
3. **Reasoning about knowledge:** An intelligent system is required to reason or figure out, what it needs to know from what it already knows in order to perform tasks for which explicit instructions are not supplied. The intelligent system should therefore use what it knows to infer new facts and knowledge (Luger and Stubblefield, 1989). Kinds of reasoning performed by machines include (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Michalski, 1986; Schildt, 1987; Minton, 1988; Luger, 1989):

(a) Formal Reasoning involving syntactic manipulation of data and knowledge structures to deduce new structures by applying rules of inference.

(b) Procedural reasoning in which simulation techniques are used to answer questions and solve problems. This method involves the invocation of a specific problem solver to supply the answers needed by the reasoning module to reach a conclusion.

(c) Reasoning by analogy, in which analogically similar cases are studied and used to infer answers to queries in other contexts.

(d) Generalization and abstraction or the natural process in which given sufficiently many instances the general properties of the concept can be abstracted and used to replace the instances.

(e) Meta level reasoning, or reasoning about knowledge and problem solving strategies.

4. **Knowledge Acquisition:** Knowledge acquisition is defined as (Luger and Stubblefield, 1989) the accumulation of facts and ability to relate something new to what is already part of the knowledge base. Knowledge acquisition must address issues of knowledge classification, interaction between new with old structures (validation and truth maintenance) and human interaction factors (natural language interfaces). These issues and the associated processes must take place at the time of the knowledge acquisition to facilitate learning and improvement of the learners knowledgeable behaviour (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Cohen et al, 1982; Luger and Stubblefield, 1989). There are two facets to knowledge acquisition (Barr et al, 1981, 1982):

   (a) Addition of new facts into the knowledge base.

   (b) Learning or self improvement.

5. **Other issues:** Other issues which are important in knowledge representation include (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Cohen et al, 1982):
   (a) Efficacy or the level to which the chosen representation structure fits reality.
   (b) Scope or level of detail of the representation.
   (c) Grain size or how much detail is required by the reasoning mechanism for efficient performance (Barr and Feigenbaum, 1981, 1982; Luger et al, 1989).
   (d) Handling of non-specificity problems in the choice of semantic primitives (Cohen et al, 1982, Luger and Stubblefield, 1989).
   (e) Modularity or ability to add, modify, or delete individual data structures without adverse side effects (Zadeh et al, 1975; Luger et al, 1989).
   (f) Easy comprehension of the representation structures by humans to facilitate their proper design and implementation (Zadeh et al, 1975; Cerone, 1980; Habel et al, 1980; Minker, 1980; Martin, 1984).

Knowledge systems are also characterised by what and how much knowledge is explicitly built into the systems, and by what amount of knowledge is implicit (Luger and Stubblefield, 1989).

Another view of knowledge representation (Luger and Stubblefield, 1989), identifies the following representation schemes:

1. Logical representation schemes based on formal logic such as the first order predicate logic, to facilitate inferencing on the knowledge base. PROLOG is the most widely used implementation of this form of knowledge representation (Kowalski, 1984; Franklin et al, 1986; Walker, 1987). These

systems are also referred to as declarative.

2. Procedural representation schemes in which knowledge is represented by a set of instructions or production rules for solving a problem. Most rule based systems implement this form of knowledge representation using LISP lists (Tanimoto, 1988).

3. Network representation schemes of which the semantic networks scheme is a typical case. These are referred to as declarative representations. Typical examples in this category include conceptual dependencies and conceptual graphs (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Luger and Stubblefield, 1989). Conceptual dependency representation is an extension of the semantic network approach in which standard relations are defined and employed to model actions, objects, modifiers of actions and modifiers of objects or dependency relations as described in Luger and Stubblefield (1989). It is an artificial language for representing natural language structure and meaning in such a way that all sentences that have the same meaning will be represented internally by syntactically and semantically identical graphs (Luger and Stubblefield, 1989).

   Conceptual graphs invented by Sowa (Luger and Stubblefield, 1989) are defined as finite, connected bipartite graphs, and they are employed in a manner such that the nodes of the graph are either concepts or conceptual relations. The conceptual relation nodes represent the relation between objects or concepts. The representation scheme defines rules for forming and manipulating conceptual graphs and the conventions for representing classes, individuals, and relationships. Their main area of application is in natural language representation.

4. Structured representation schemes, regarded as extensions of the network representation schemes which allow nodes in a network to consist of complex structures. Examples of structured representations include frames, scripts, objects and black board architectures (Barr et al, 1989).

Knowledge representation must ensure correctness of inference, implying that the results of inferences must correspond to the result of actions or observations in the real world. This interpretation of knowledge representation appearing in Luger and Stubblefield (1989) and Tanimoto (1987) corresponds to the notion of machine understanding as defined in Rosenberg (1980), Habel et al (1980), and others.

In this research the knowledge representation scheme used is a combination of the semantic network structure and rule based representation schemes. Details of the representation schemes are included in section 3.2.2 of this chapter and in chapter 4.

Principles of cognitive learning by examples is used to facilitate acquisition of knowledge about domain objects by the LEARN subsystem. As explained later (section 3.3.1) the learning algorithm implemented in the SLEMS is the hit and miss strategy given in Schildt (1987). However the SLEMS implementation incorporates a modification to enhance the learning process by introducing the concept of taboo objects.

As implemented in the SLEMS the concept of learning from taboos serves as a powerful device for controlling the inference and learning process. Taboos are forbidden attributes of a concept or object of learning. The analogy upon which the strategy is based is the concept of social taboos which generally play a major role in shaping human ethics and morals. In African societies, for example, taboos play an important role in the learning process of children (personal experience).

Both the LEARN and EXPERT subsystems incorporate a simple explanation mechanism to facilitate simple English language explanation of the results obtained by the inference modules.

3.2.1 The Semantic Network Knowledge Representation Scheme.

The semantic network was invented by Quillian (Barr et al, 1989) to model human thought processes. This structure is based on the assoccianist theory (Luger and Stubblefield, 1989) which defines meaning in terms of a network of associations with other objects in a knowledge base. A semantic network therefore represents knowledge as a graph, with labelled nodes corresponding to facts or concepts and labelled arcs representing relations or associations between the concepts. The scheme developed by Collins and Quillian (Barr et al, 1989; Luger and Stubblefield, 1989) is shown in figure 3.1.

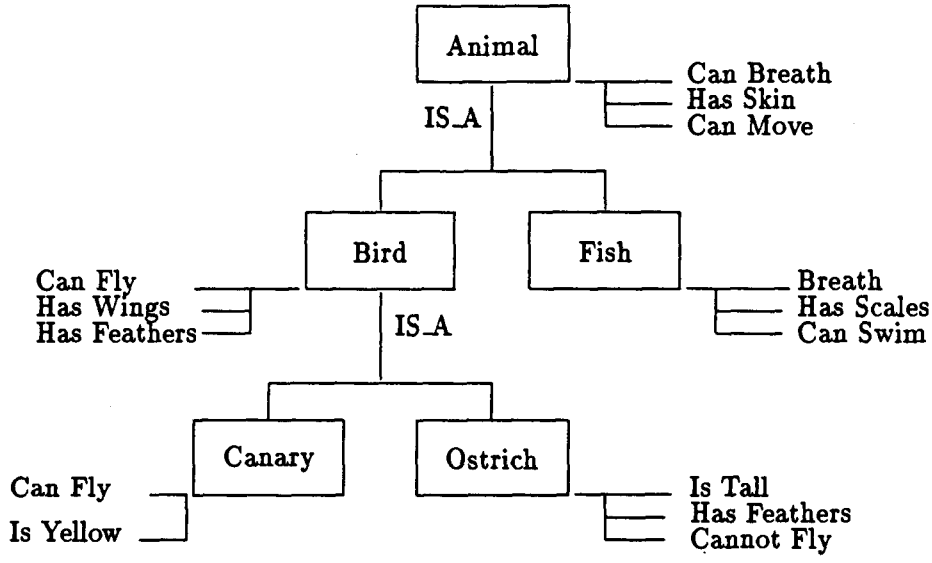


Figure 3.1: Quillian's Semantic Network Model (After Barr et al, 1989, Vol III, page 40).

In a semantic network two types of nodes and relationships (links) may be identified (Shastri, 1989). The first type of nodes consists of classes and categories of individuals. The second type consists of individuals (instances) and their properties.

The first kind of relations represent structural links such as the IS_A or MEMBER link (Fahlman, 1979; Shastri, 1989). The second kind consists of property and individual links such as HAS or HAS_PART. Structural links represent knowledge on the inherent hierarchical subclass - superclass relationship between the domain objects and concepts (Fahlman, 1979; Shastri, 1989). In this hierarchy of concepts, individuals and properties; properties are attached at the highest concept in the conceptual hierarchy to which the property applies (Shastri, 1989). This restriction on the attachment of properties is also referred to in Fisher and Langley (Gale, 1986) as the principle of maximally general discriminant concepts.

Practical systems modelled after the concept of semantic networks are referred to as inheritance systems (Fahlman, 1979; Shastri, 1989) and they have the following characteristics (Shastri, 1989):

- They store information at the highest level of abstraction, thereby reducing the size of stored knowledge.
- Inheritance property helps maintain consistency of the knowledge base when new classes and objects are added by guaranteeing automatic inheritance of all properties of super classes.

More details on the semantic network structure and its formal definition may be obtained in Shastri (1989), Luger and Stubblefield (1989), Fahlman (1979), Cerone (1980) and others.

3.2.2 The SLEMS Knowledge Representation Structures.

The EXPERT subsystem exclusively makes use of the rule based representation scheme. To facilitate retrieval and avoid conflicts the domain knowledge is classified prior to entry. Related subdomains are networked or lumped together. As it will be explained later in the detailed discussion of the EXPERT subsystem in chapter 5, the resulting knowledge structures form a hierarchical network resembling decision

trees (Barr and Feigenbaum, 1981, 1982; Barr et al, 1989; Shapiro, 1987; Luger and Stubblefield, 1989).

The LEARN subsystem exclusively makes use of the semantic network representation scheme. The knowledge base used in this subsystem consists of a network of associative object triplets in which inherent domain knowledge relations are represented by the "is a"[1] link and other relations as explained later in this chapter. Although pre-analysis may facilitate a more efficient network structure the network is not explicitly formed but rather, virtually realised by the learning and inference modules of the LEARN subsystem.

The concept of the associative triplets as introduced in Schildt (1987) is very similar to that of the triple stores described by Martin (1984) and the object attribute value (OAV) triplets discussed in Blais (1987). A collection of these triplets constitute a version space(Tanimoto, 1988)[2] and can effectively be used to define or classify real world objects or concepts and their attributes.

The associative triplets parse the real world knowledge in a manner similar to the parsing of english sentences into subject, verb and object phrases. They are therefore also referred to in this thesis as subject-verb-object (SVO) triplets. Because of their similarity to the triple store structures the SVO triplets inherit the following advantages identified in Martin (1984):

- They support powerful retrieval operations relying only on entity relations and inheritance.
- They facilitate use of generic information (virtual knowledge) for data organisation and hypothesis formation.
- They facilitate use of surrogates, for example:

[1]Within the SLEMS the notation "is a" rather than the usual IS_A, etc. is used because the semantic network used is not composed of symbolic links and nodes, but rather by character strings representing real world relations and facts.

[2]According to Tanimoto (1987), "A version space is a set of rules that is bounded above by the most general rules in the set and that is bounded below by the most specific rules in the set."

<FACT1> <IS> <the rainfall erosivity factor R >

<FACT1> <IS> < 0.5 >

- They allow direct access of the data model by the user for manipulation.
- They give the user freedom in specifying his own conceptual model of data.
- They support flexible user interface.

The associative triplets used in the SLEMS differ from the triple store structure in two aspects:

- They require no mapping of the text input to integers as required in Martin (1984).
- The associative triplet knowledge representation structure forms meaningful strings and it is available for direct manipulation by the user.

In this thesis the associative triplets are referred to as the LEARNER_OBJECT types and they are defined by the Bacus-Naur form (BNF) (Hopcroft et al. 1979) given below.

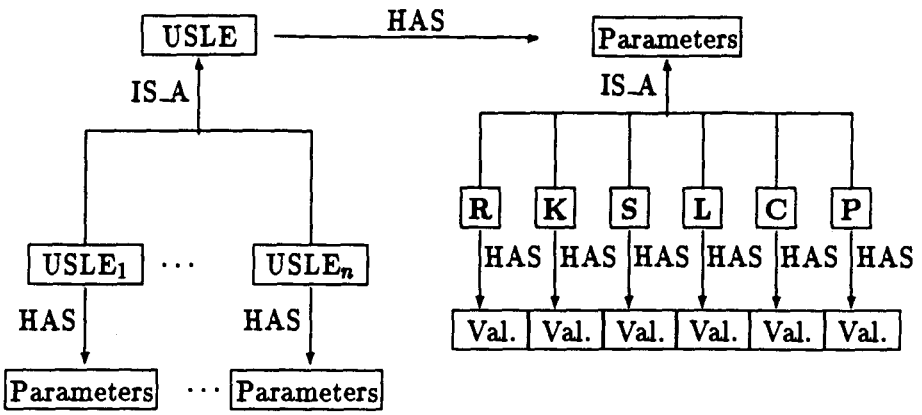
```
LEARNER_OBJECT : <subject_part> <relation> <object_part>
subject_part   : <subject phrase> | <subject of interest>
relation       : <verb phrase> | <any binary relation>
object_part    : <object phrase> | <object of interest> |
                 <operator> <value> <attribute_name>
```

The object attribute (OA) tuples are the basic knowledge representation units in the EXPERT subsystem. The OA-tuples are used to represent both facts and rules for the manipulation of the EXPERT subsystem and they are called EXPERT_OBJECT types. The BNF definition of the EXPERT_OBJECT types is given below.

EXPERT_OBJECT	: <Object> <Attribute_list> | <Conclusion> <Premise_list>
Object	: <object of interest> | <fact>
Attribute_list	: <first attribute... last attribute>
Conclusion	: <rule conclusion> | <preconditioned fact>
Premise_list	: <first premise... last premise>

Internally the EXPERT_OBJECT types are represented by two C structures. The attribute list part is implemented as linked list arrays and the object part is implemented as C array structures as elaborated in chapter 5. Figure 3.2 shows a possible representation of the SLEMS knowledge by a semantic network of the basic structures.

As seen from figure 3.2, the values assigned to the parameters defined at a higher level in this hierarchy would constitute default values inheritable by the lower hierarchy objects, i.e., if some parameters are not defined for USLE_2 these could be inherited from those of the general USLE model, because the USLE_2 model is a subcategory of the universal soil loss model.



KEY:

USLE Universal Soil Loss Equation

Val. Parameter Value

Figure 3.2: A Semantic Network of USLE Concepts.

3.2.3 SLEMS Knowledge Manipulation Language.

SLEMS knowledge manipulation language has very few restriction on the use of the English language words. However some few words specifically the verbs "is a" and "is an", and "has" are assigned special meaning, and they are automatically recognised by the system as class, respectively, value assignment relations.

The relational link "is a" is used in the SLEMS LEARN subsystem as a standard class relationship or parent - sibling relation specifier. The alternate form "is an" is also recognised by the SLEMS inference operators as its equivalent. Typical usage of these special relations to describe a concept is shown in the example below.

USLE_1 "is a" USLE model version
USLE_2 "is a" USLE model version
USLE_3 "is a" USLE model version
⋮
USLE "is a" universal soil loss equation
USLE "has" parameters:S, R, K, L, C, P.

In the above example the verb "has" is used to assign attributes to an object. The relation "has" can also be used to assign values to attributes of an object. As shown in the next example below such use facilitates translation of relational file records into SVO triplets.

polygon A456 "has" rain_intensity:1.7mm./hr.
polygon A456 "has" rain_duration:6hrs.
polygon A456 "has" 30_minute_rain_intensity:2.8mm/hr.
polygon A456 "has" rain_frequency:2yrs.
polygon A456 "has" rain_amount:150mm.

In this fashion a record field and its values are bound. A detailed discussion of the problems and strategies for embedding a relational data structure by the triple store structure is given in Martin (1984).

Assigning attributes and attribute values by means of the special relations "is a" and "has" allows the generalization module, discussed later (section 3.4.3), to condense the knowledge by aggregation to the following form:

polygon A456 "has" rain_intensity:1.7mm/hr. **or**
rain_duration:6hrs. **or**
30_minute_rain_intensity:2.8mm/hr. **or**
rain_frequency:2yrs. **or**
rain_amount:150mm.

The use of symbolic representation of the relations "is a" and "has" permits the designation of other words as class and value relations thereby increasing flexibility in their application. To create a new relation one only needs to re-assign the appropriate symbol to the desired word or string. The special SLEMS relations can be internally represented by the character variables, CLASS_R = "is a", ALT_CLASS_R = "is an", and VALUE_R = "has". For example, to create a new class relation "soil type", both CLASS_R and ALT_CLASS_R can be assigned the string "soil type". This relation may then be used to assign soil types to polygons in the form "Poly456" CLASS_R "podzolic".

3.3 Representation of Domain Knowledge in General.

Knowledge entries in the SLEMS consist of two types of objects as explained above. A SLEMS object is either a LEARNER_OBJECT type or an EXPERT_OBJECT type. A valid SLEMS object is thus represented by the BNF production

SLEMS_OBJECT: <LEARNER_OBJECT> | <EXPERT_OBJECT>

The different object types on the right hand side of this production were specified in terms of BNF productions in section 3.2.2. Because the rule based representation

of soil erosion domain knowledge is discussed in great detail in chapter four this section will concentrate only on the representation of domain knowledge in the LEARN subsystem.

3.3.1 The LEARNER_OBJECT Types.

The LEARN subsystem is primarily a knowledge acquisition tool for the SLEMS system. To facilitate learning and acquisition of domain knowledge it parses the domain knowledge into three categories: the MAY_OBJECTS, the MUST_OBJECTS and the TABOO_OBJECTS object types. Each of these object types consists of associative triplets as explained above. The expanded BNF representation of the LEARNER_OBJECT is given below.

```
LEARNER_OBJECT : <MAY_OBJECTS> |
                 <MUST_OBJECTS> |
                 <TABOO_OBJECTS>
MAY_OBJECTS    : <simple LEARNER_OBJECT> |
                 <generalized LEARNER_OBJECT>
MUST_OBJECTS   : <simple LEARNER_OBJECT> |
                 <generalized LEARNER_OBJECT>
TABOO_OBJECTS  : <simple LEARNER_OBJECT> |
                 <generalized LEARNER_OBJECT>
MUST_OBJECTS   : <subject> <not relation> <object>
```

The MUST_OBJECTS triplet must be the same (with the exception of the negation "not", "no" or "never" of the relation) as a previously learned object to be a valid near-miss example (Schildt, 1987). MAY_OBJECTS and TABOO_OBJECTS have exactly the same BNF as introduced above.

The MAY_OBJECTS can be regarded as the permissible or necessary facts about the concept being being represented. The MUST_OBJECTS constitute the mandatory attributes of the real world concept or object. They represent the attributes which must be present in the object for it to qualify as member of the concept being represented.

The TABOO_OBJECTS are strongly discriminant attributes of the concept or object being introduced. They consist of all associative triplets which invalidate membership of all other real world objects to the object or concept being represented.

The MUST_OBJECTS are automatically derived by the LEARN subsystem's restriction module (see fig. 3.5). The only means for the user to control their input is through the specification of the so called near miss objects (Schildt, 1987). Tanimoto (1987) defines a near miss as a negative example of a concept which is almost true. According to Schildt (1987) and Tanimoto (1987) a new example will have an effect (add to the knowledge) on the concept if it is salient, in the sense that it is either a near miss or an unexpected hit. The unexpected hit is a positive example which is not redundant.

In the SLEMS salient near miss examples are constructed from positive examples by a complementation process. For example, since we know that Landsat MSS *Band 1 has the spectral range:* $0.45\mu m - 0.52\mu m$, then a good near miss example might be *Band 1 has* not *spectral range:* $0.45\mu m - 0.52\mu m$.

Given the above near miss the learner then reasons that since this is an incorrect example of the concept, then the previously given positive example must be mandatory, that is *Band 1 must have spectral range:* $0.45\mu m - 0.52\mu m$, which is therefore a restriction on the original example. Having arrived at this conclusion the learner then moves the original positive example of the concept into the mandatory objects category (MUST_OBJECTS).

Parsing of the input string representing a real world concept or object is a cooperative process between LEARN and the user. LEARN controls the entry

of objects by prompting for the appropriate components of the MAY_OBJECTS, MUST_OBJECTS, and TABOO_OBJECTS required to define the concept. The LEARN subsystem has an absolute faith in the user, and takes whatever the user supplies as subject_part, relation and object_part as is. The assumption is that the human expert will guarantee the input to be logically consistent with the common sense meaning of the concept or object of interest. In short a subject_part, relation, or object_part is whatever the user decides to tell the system. The same applies to the TABOO_OBJECTS.

The loose specification of the SLEMS object components allows a subject_part in one object to appear as an object_part in another object and vice versa, provided they form a meaningful concept description.

During knowledge input the input module automatically collects all terms designated as relations by the knowledge engineer or programmer into special files called the SLEM_RELATIONS. This enables the system to remember them for future references.

The SLEMS EXPERT incorporates modules for facilitating transformation from LEARNER_OBJECT type to EXPERT_OBJECT type. The system is also able to convert conditional facts into rules.

3.4 The SLEMS Knowledge Acquisition and Manipulation Subsystems.

The SLEMS has four subsystems as mentioned in chapter 1. These are the SLEMS DBMS, the LEARN, the EXPERT, and the FUZZ subsystems which are configured as shown in figure 3.3. The LEARN, the EXPERT and the FUZZ subsystems constitute the SLEMS intelligent interface.

Manipulation of SLEMS knowledge structures is done at four different levels. The lowest hierarchy consists of low level routines (Fahlman, 1979), such as simple and

smart[3] string manipulation operators.

The next level in the hierarchy consists of low level operators for performing binary comparisons or matching of elementary objects. These are followed in the next level by operators which perform decision making based on the returned state of the invoked lower hierarchy search operators. This level constitutes the systems inference layer. The highest layer is the user interface through which the user accesses the lower order functions. It therefore hides the details of the database searching and knowledge manipulation operations.

The system to system interface between the LEARN and the CDATA based SLEMS database management system is part of the subsystem interface. Its main task is to facilitate transformation of the SLEMS DBMS data structures into knowledge structures for loading into the LEARN subsystems active memory. Also forming part of the system to system interface is a one way link between the EXPERT and the LEARN subsystems.

[3]Smart string operators include find_token, find_toke, detect_neg, convert_string_to_symb, etc. which perform advanced string matching for the search operators. Simple string operators are the standard C library string operators.

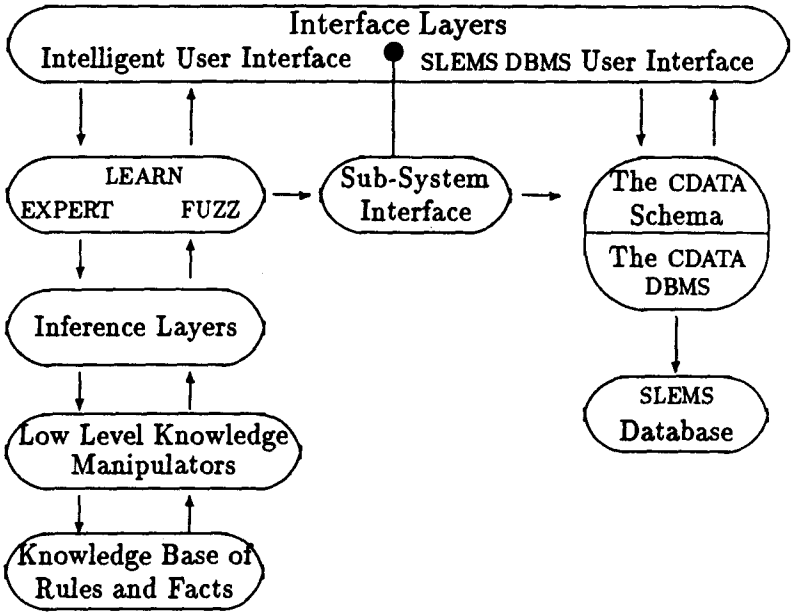


Figure 3.3: The SLEMS Configuration.

3.4.1 The SLEMS LEARN Subsystem.

The LEARN subsystem is the primary knowledge acquisition tool of the SLEMS. It regards each new SVO-triplet as a fact to be learned, performs generalization on its attributes and stores the generalized facts as LEARNER_OBJECT types. Based on the learned concepts the LEARN subsystem is able to answers queries by inductive inference using the stored facts.

The components of the LEARN subsystem are as follows:

1. The knowledge acquisition module consisting of:
   (a) The knowledge entry module.
   (b) The generalisation operator.
   (c) The restriction operator.
   (d) The forbid or taboo operator.
2. The knowledge manipulation modules constituting three layers:
   (a) The inference layer
   (b) The fuzzy knowledge manipulator.
   (c) The low level knowledge manipulation layer
3. The explanation module.
4. The subsystem interfaces.
5. The utility functions.
   (a) The display and report modules.
   (b) The internal system information modules.

The fuzzy knowledge manipulator is essentially a component of the FUZZ subsystem discussed in greater detail in chapter 4, which can be accessed by the LEARN subsystem to handle fuzzy knowledge input and manipulation.

An integral component of the user interface is the explanation module which facilitates user acceptance of the solutions found by the expert system (Schildt, 1987; Shapiro, 1987; Luger and Stubblefield, 1989). In the SLEMS this function is fulfilled by the **slem_how** module in the LEARN subsystem and **reasons** module in the EXPERT subsystem.

The subsystem interface performs data transformations from representation in one subsystem to the other. The system to system interface thus performs an important function designated in Fahlman (1979) and Luger and Stubblefield (1989) as knowledge parsing, for the SLEMS. In the present implementation the subsystem interfaces provide for the EXPERT subsystem to access the knowledge structures generated by the LEARN subsystem and for the LEARN subsystem to access the records stored by the SLEMS DBMS.

Initially a provision for the LEARN subsystem to deposit the knowledge acquired by it in the SLEMS DBMS was provided but later removed because of problems of fitting generalized objects into the (CDATA based) SLEMS DBMS's fixed length records.

The knowledge acquired by the LEARN and EXPERT subsystems is stored and accessed by the SLEMS utility functions which consist of the **module_info** and **learner_info** modules. The I/O and search routines in these modules are based upon the Sun 4 UNIX Workstations database management functions (SUN4, 1988).

3.4.2 The Knowledge Editing Modules.

This section discusses only the functionality of the knowledge editing or transformation modules of the LEARN subsystem. These are the **generalise**, the **restrict** and **forbid** modules. The simple operations embodied in the three modules constitute the knowledge acquisition functions of the LEARN subsystem. Knowledge editing can also be performed during query operations, in which case, knowledge editing consists of addition of new facts derived from the existing knowledge to the active memory.

The query processor has options to suppress or remove invalid objects from the active knowledge base.

In conformity with knowledge based systems design practice (Ebrahimi, 1987; Luger and Stubblefield, 1989) only the knowledge acquisition module can write new knowledge into the permanent or long term memory of the system. In SLEMS the knowledge updates made to the active memory can however be stored at the users discretion at the end of the query session.

3.4.3 The Generalization, Restrict and Forbid Operators.

The algorithm used by this module is the simple hit and miss paradigm(Schildt, 1987; Tanimoto, 1988; Barr et al, 1989) as explained earlier (section 3.3.1). Its main task in the subsystem is the transformation of the simple triplet structures into more complex structures through an aggregation process (Schildt, 1987). The operator analyses multiple instances (SVO-triplets) of an object or concept and creates a more general concept of the object instances by extending either the subject_part or the object_part of the triplet using "or" concatenation (Schildt, 1987).

The conceptual function of this operator is shown in figure 3.4. It takes as input multiple object instances and produces a single generalized object triplet.

In the last example (of figure 3.4) the input objects have no discernible similarities, hence generalization does not produce any results. In this form of generalization the relation part is an invariable parameter.

The restrict operator takes as input two objects, an existing SLEMS object and its near-miss version. It is implemented after the algorithm and modified source code in Schildt (1987) to perform two functional tasks:

- Recognition of near-miss objects.
- The transformation of relevant MAY_OBJECTS to MUST_OBJECTS.

This function is diagrammatically shown in figure 3.5. The output of the restrict operator is placed in the SLEMS MUSTOBJECT or mandatory objects knowledge base if it is not a duplication of some existing mandatory object, and it does not conflict with the established SLEMS taboos.

The purpose of this operator is to prevent the assertion of objects known to be taboo. The operator functions at two levels. During knowledge input it controls the object entries by rejecting objects known to be taboos.

After the initial knowledge input some objects may later be declared taboo. In this case the "forbid" operator, implemented as the "enforce_taboo" routine, scans the knowledge base for inconsistencies with the new state of the knowledge base and removes them. In the event that an existing taboo is removed another operator, "clear_taboo", performs the inverse task of lifting the taboo imposed on previously valid SLEMS objects. This operator has however certain limitations. If the taboo object is embedded in a generalised object one of three outcomes is possible:

- In a situation where the complex object represents a many to many relation over the generalized associative triplets, the automatic decomposition of the object is not possible. The user must manually perform the knowledge update.
- If the complex object is a many to one relation or a one to many relation the operator decomposes the object, extracts and revises the appropriate object component.
- If the object is a simple one to one relation no action is needed since the situation is handled by a simple "kill_taboo" operator.

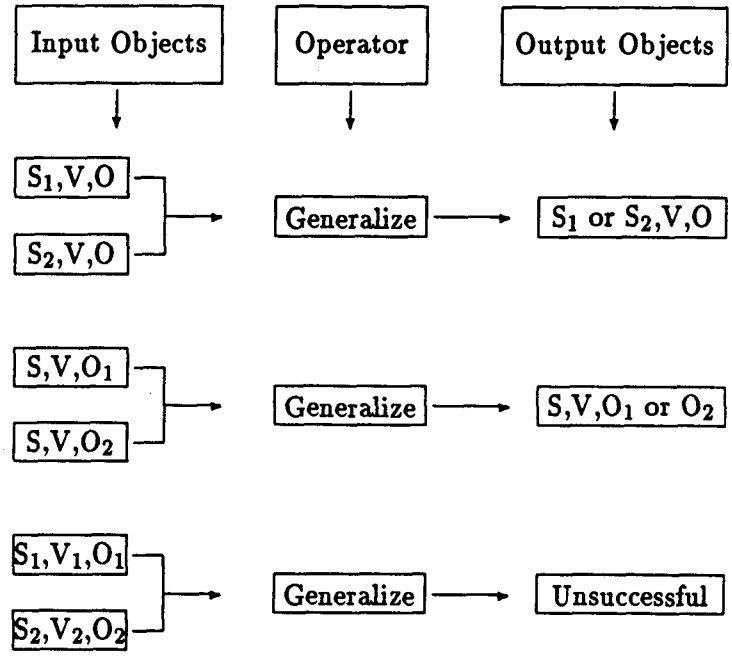


Figure 3.4: The Functional Concept of the Generalization Operator.

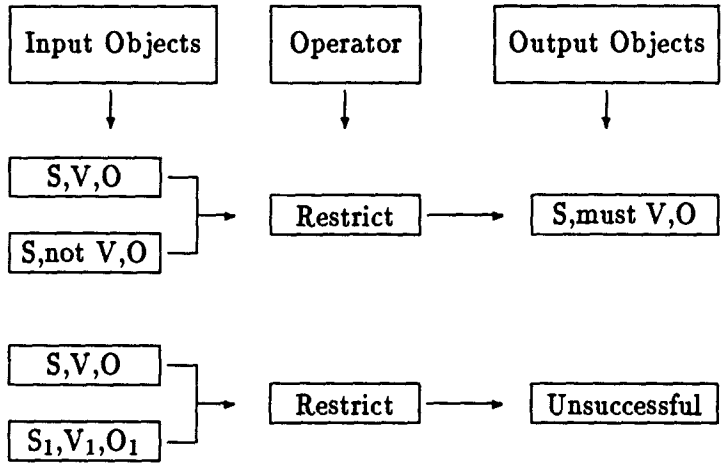


Figure 3.5: The Functional Concept of the Restrict Operator.

3.4.4 The Knowledge Manipulation Operators

The knowledge manipulation operators of the LEARN subsystem can be characterised as default logic based inference operators(Reiter, 1987). However a fundamental difference exists between the objects manipulated in Reiter (1987) and those manipulated by the LEARN subsystem. SLEMS objects are generally not well formed formulae (wff) as in Reiter (1987) but rather associative triplets of meaningful strings of characters.

More specifically the inheritance operators in the LEARN subsystem are based on the principle of modus ponens (Zadeh et al, 1975; Magrez, 1989) and semi-normal defaults as defined in Etherington et al (1987) and Touretzky (1987). Modus pones implication is an inference rule which states that (Zadeh et al, 1975; Delahaye, 1986; Touretzky, 1987; Zadeh, 1989) *if A implies B and A is known to be true then B can be inferred.* Default logic permits situations where even when A is valid B cannot be inferred. Such a situation may arise for example if the assertion of the fact B would conflict with the world knowledge (Etherington et al, 1987).

Normal default inheritance therefore states that (Touretzky, 1987) *If A is valid and B is not known to be in conflict with the known facts, then B can be implied.* Semi-normal default implication provides for additional control by allowing for exceptions to override normal default logic (Touretzky, 1987). This may be illustrated by the implication *If A is true and B is not known to be in conflict with the world knowledge then C may be implied* where A is called the prerequisite or precondition, B the justification or exception and C the consequent or conclusion (Touretzky, 1987).

All SLEMS knowledge manipulation operators implement the normal default by requiring the verification of the results of any implication before they are asserted. The use of taboos to control the inference process in the LEARN subsystem is SLEMS form of implementing the principle of semi-normal defaults espoused in the referenced literature.

The LEARN subsystems knowledge manipulation operators may be categorized

according to the implementation strategy of the inference routines. Using this criteria two categories of operators may be identified:

- Basic Search Operators
- Derived Operators.

Basic search operators perform simple matches on the knowledge base objects and they constitute the low level knowledge manipulation layer (figure 3.6). Derived operators are built from the basic operators and they are used to perform more complex knowledge base manipulations or inference. They constitute the inference layer.

The SLEMS basic operators have a dual relationship as illustrated in figure 3.6.

$$(s,v,o) \xrightarrow{S} (V,O) \qquad (s,v,o) \xrightarrow{VO} (S)$$

$$(s,v,o) \xrightarrow{V} (S,O) \qquad (s,v,o) \xrightarrow{SO} (V)$$

$$(s,v,o) \xrightarrow{O} (S,V) \qquad (s,v,o) \xrightarrow{SV} (O)$$

$$(s,v,o) \xrightarrow{(S,V,O)} (TRUE, FALSE)$$

Figure 3.6: Duality of SLEMS basic operators.

Operators on the left hand side in figure 3.6 transform a one dimensional instance space into a two dimensional solution space while those on the right side transforms a two dimensional instance space into a one dimensional solution space. Knowledge manipulation within the SLEMS LEARN subsystem consists of the following operations:

- Searching for atomic objects or complex objects within the generalized object space.
- Knowledge transformations.

Searching within the generalized object space is achieved by means of search operators. SLEMS search operators are loosely predetermined by the structure of the generalized object space. The search operators take as their input an incomplete or a given associative triplet. Two types of basic search operations are identifiable:

- Finding an unknown object's components.
- Verifying a given objects existence.

Answers to complex queries involve transformations on the knowledge base to produce objects not explicitly stored in the knowledge base. Knowledge transformations implemented in the SLEMS consist of inferencing, explanation and knowledge updating.

The inference operators rely on the "is a" class relation to perform specialized knowledge base matching during which parent-sibling relationships among the knowledge base objects are discovered and used to infer implicit knowledge. The value assignment relationship "has" is used to detect attribute-value relationships which are essential for value comparisons on the knowledge base objects.

Inheritance operators are a specific category of the inference operators. They facilitate the derivation of new knowledge by default reasoning (Etherington et al, 1987; Touretzky, 1987), based on the "is a" hierarchical relationships among the knowledge base objects.

The LEARN subsystem's search operators are mainly based on the breadth first search approach. Given a particular query, the inference modules will attempt to discover all valid solutions to a sub-problem at each stage of the search. If there is more than one solution to a query the particular operator invoked will find all the answers. The problem of conflicting solutions or conflict resolution is not resolved.

Knowledge updates are performed by a special class of operators. Two categories of these operators are identifiable.

- In-build operators.

- User accessible knowledge manipulation operators.

The in-built operators include knowledge input, storage and retrieval modules controlled by the intelligent interface. User accessible functions include, the taboo elimination operator (K), the implication operator (I), the inheritance operator (INH), and the conditional assertion operator (CA) which can be used by the user to modify the active knowledge base content directly. SLEMS explanation module may be regarded as a special type of knowledge base transformation operator. It takes as input a sequence of problem solving steps and produces a simple natural language explanation of the solution strategy. The module **slem_how** implemented as the user accessible command HOW uses knowledge about SLEMS operators and their functions. These are stored in special stacks administered by the **module_info** and **learn_info** modules and they are used to construct english like explanations.

Also belonging to this group of house keeping modules is the knowledge status generator **kne_dump** (ST), and the help functions (H). The status generator shows all objects resident in memory (transient knowledge base) at any particular time with comments on their validity or invalidity. The SLEMS display module (D) displays valid generalized and un-generalized knowledge acquired by the knowledge acquisition module, or down loaded from the fixed knowledge base.

The modules **module_info** and **learn_info** also keep record of the SLEMS numerous routines, and provides quick reference and explanation of their functions and history. This information is kept in special information files and loaded at the commencement of the SLEMS session. The same information is also consulted by the explanation module as explained above. The help function displays a full list of SLEMS knowledge base manipulation operators.

3.4.5 The Basic Search Operators.

The SLEMS LEARN subsystem has eight basic search operators. The first six of these perform binary and unary transformations on object triplets. The sixth is a verification operator. The eighth transformation operator (FUZZC) is a specialized search and verification operator designed mainly to handle objects with imprecise or fuzzy attribute values. This operator is therefore used only for manipulation of objects whose object part are fuzzy attribute value (FAV) pairs.

The first seven knowledge manipulation operators are summarised below with a brief description of their main functions.

- Unary operators:

    1. S: given the subject_part of a SLEMS object produce all valid relation-object tuples.
    2. V: given the relation part produce all related subject-object pairs.
    3. O: given the object_part of a SLEMS object produce all the related subject-relation tuples.

- Binary operators:

    1. SV: given a subject-relation tuple produce all the related object_part(s).
    2. SO: given a subject-object tuple produce all the relation(s) defined (existing) on the tuple.
    3. VO: given a relation-object tuple produce all the related SLEMS subject-_part(s).

- Ternary operator:

    - SVO: given the subject, relation, and object verify the validity of the triplet.

The operator also checks for inconsistencies with the known facts before asserting B. Not withstanding the validity of the precondition, the assertion of B may be blocked if:

- B is an existing knowledge base object (duplicates not allowed).
- B is an existing taboo object (violation of existing taboos).

2. **CA:** The conditional assertion operator facilitates supervised conditional learning of new facts when certain pre-conditions are satisfied. It accepts only valid relations in the precondition. If the relation on the precondition is unknown the operator prompts for its validation by the user. It then checks the facts in the pre-condition and then proceeds to assert or reject the new assertion. It also validates the new assertion against the knowledge base. This operator may therefore be considered an implementation of default logic implication (Touretzky, 1987).

3. **INH:** The simple inheritance operator is based on modus ponens and normal default (Touretzky, 1987). It performs inheritance of the immediate super-class attributes for a designated object. In SLEMS inheritance is not generally automatic but user requested. The **INH** operator uses the in-built knowledge about class relations to facilitate inheritance of super - class objects attributes. The **INH** operator is however designed to automatically inherit the mandatory (MUST_OBJECTS) and taboo attributes (TABOO_OBJECTS) of the super class. Permissible attributes (MAY_OBJECTS) are regarded as weak attributes requiring verification by the user before their assertion.

4. **GINH:** The general implication operator works like the **CA** operator, but it performs more sophisticated examination of the precondition in its attempt to find support for the inheritance of new attributes for the subject_part of the precondition. The inference rule used by this operator is

given by:

$$\begin{array}{c}\text{If } X \text{ is related on } R_i \text{ to } B \\ \text{and } X\ R_j\ C \text{ is not invalid} \\ \text{then } X \text{ is related on } R_j \text{ to } C\end{array}$$

where R_i and R_j are two different relations on the subject X. The **GINH** operator also makes checks similar to those of the **I** and **CA** operators before asserting new facts. It implements the semi-normal default implication (Touretzky, 1987)

Possible usage of the **GINH** operator could be the discovery of new relationships for a given subject. The operator could in this case try to match successive **VO** pairs against the knowledge base accepting those which form meaningful relations with the subject_part of the precondition.

3.4.6 Manipulation of Class and Value Assignment Relations

There are five types of class relation operators apart from the **INH**, **I**, and **GINH** operators discussed above. They take object triplets as input and produces TRUE, FALSE outputs.

1. **SCLA:** Given a particular subject (the subject_part) this operator finds a SLEMS object (object_part) in an "is a" relation with the subject and assigns its class to the subject. If the object is not a predefined class, the operator uses the objects name as the label for the new class assigned to the subject.
2. **ISCLA:** Given a subject (subject_part) find if it corresponds to a SLEMS class as follows:
   - It is a predefined class or so called defined class.

- It is a class by virtue of satisfying other criteria for a domain class (deduced class).

3. **VCLA:** Given a subject and an arbitrary class verify that the subject is a member of the given class.
4. **ISME:** Given a subject find if it is a member of any existing or inferable domain class.
5. **HASME:** Given a subject find, assuming it is a class, if it has any members. This operator tries first to verify that the object is a class either by definition (defined class) or by inference (deduced class), it then proceeds to find objects belonging to its class. If any are found it returns TRUE else it returns FALSE.

Values in the LEARNER_OBJECT's are represented as attribute - value pairs (AV) of the object_part. Operations which may be desirable on these AV-pairs are:

- Extraction of the value part of the pair.
- Verification of an objects attribute value.
- Comparison of the values of two different AV pairs.

These operations constitute the SLEMS basic value manipulation operations. The operators implemented to handle these types of operations are:

1. **EV:** Given a LEARNER_OBJECT instance determine the value associated with a named attribute.
2. **VE:** Given a LEARNER_OBJECT instance verify that its named attribute has a specified value.
3. **FUZZC:** Given two LEARNER_OBJECT instances perform comparisons on their fuzzy valued object-parts. This operator accepts both precise and imprecise or fuzzy value assignments in the generalized object.

A common characteristic of the last three operators is that they all operate on the object_part of the LEARNER_OBJECT's. The FUZZC operator accepts character string inputs and parses it into fuzzy predicate and numeric value parts. The operator recognises and correctly interprets a limited number of fuzzy predicates as given in table 3.1. Detailed discussion of the theoretical basis of the fuzzy comparison operator is included in chapter 4.

FUZZY PREDICATE
more or equal to X
less or equal to X
more or less X
much more than X
much less than X
about X
roughly X
more than X
less than X
slightly more than X
slightly less than X
from X to Y
X to Y
equal to X

Table 3.1: List of Fuzzy Expressions Recognized by the Fuzzy Comparison Operator (FUZZC).

The preprocessor of the FUZZ subsystems parser, consisting of the routines **convert_word_to_symb, replace_word**, invoked by the FUZZC operator permits entry of the fuzzy predicates in table 3.1 exactly the way they are written. The SLEMS implementation for these fuzzy predicates also recognises their many english synonyms. This is achieved by mapping synonyms to the same fuzzy predicates e.g. *greater, above, longer, and wider* are mapped to *more* while *lower, below, shorter, narrower, smaller* are mapped to *less*. The mapping of similar predicates into one constitutes a knowledge compression operation and it reduces the amount of knowledge directly

built into the system.

3.5 Knowledge Extraction and Query Processing With the LEARN Subsystem.

A typical LEARNER_OBJECT instance is constructed by generalization of the knowledge acquired by the LEARN subsystem. Knowledge is introduced in the form of simple objects as seen in the extract of the learning session shown below.

> *Knowledge domain:*

SPECTRAL_SENSING

> *Enter your example:*

> *Subject:*

BAND 1

> *Relation:*

has

> *Object:*

wavelength 0.45 to 0.52

> *Your example was:*

BAND 1 has wavelength 0.45 to 0.52.

> *Enter next example*

> *Subject:*

BAND 1

> *Relation:*

has

> *Object:*

blue nominal spectral location

> *your example was:*

BAND 1 has blue nominal spectral location.

In this example the prompt *Relation* stands for the *verb_part* of the SVO-triplets discussed in section 3.2.2. As stated earlier the user interface controls the input such that the object is parsed according to the syntax <subject_part> <relation> <object_part> at the time of entry. After applying the generalization process an extended MAY_OBJECTS instance is obtained as shown in the extract presented below.

> *May have extended:*

BAND 1 has wavelength 0.45 to 0.52 or
nominal spectral location blue

Notice that the aggregated object has the form:

<BAND 1> < has > < wavelength 0.45 to 0.52 or
nominal spectral location blue. >

The generalized object_part is in this case a disjunction of two atomic object_part(s). User entries may consist of simple objects or complex objects formed with the "or" concatenation.

Simple queries are executed by the basic search operators. A typical simple query and solution is shown below

> *Option:*

O

> *Enter Object Name:*

spectral location

> *Query Solution:*

BAND 1 has nominal spectral location blue.
BAND 2 has nominal spectral location green.

BAND 3 has nominal spectral location red.

⋮

In this example the search operator (**find_may_on_object**) has retrieved all the MAY_OBJECTS instances in which the sub-object "spectral location" appears in their *object_part* Complex queries must first be pre-processed by the SLEMS query parsers. The objective of query processing is two fold: first the query processor must break down the complex query into its constituent components; secondly the query processor must retrieve knowledge base objects according to the search criteria inherent in the query, break them down into their constituent components, and then pass the constituent components to the inference layer. In the implementation of the SLEMS query processor, three possible states may be returned for the component and overall query evaluation; TRUE, PARTIAL_MATCH, or FALSE. Other status, such as invalid query syntax etc., are dealt with separately. The inference layer uses the returned states to draw conclusions about the state of the search and produce desirable system response. If the state returned is TRUE., the system reports success and presents its solution set to the user. If the state is FALSE the system reports. If the state returned is PARTIAL_MATCH, the system presents the solutions found and cautions that these were found on partial match. This leaves the onus to accept or reject the solution set on the user. This state was incorporated to allow the query processor to accept and process incomplete queries as explained in section 3.5.1. The subsystem also explains reasons for the failure or success on request as shown in the typical "HOW" query type below.

> *Option:*

INH

> *Enter subject to inherit:*

tricycle

> *Enter class to inherit:*

animal

> *The subject may inherit the class attributes.*

HOW

> *How I arrived at:*

[tricycle is a beast] is a [taboo object]
is a [valid] answer to your query?]

> *well, to begin with,*

[tricycle is a beast] was found invalid with
help of task module
[find may_on_subject_attributes] using
the [permissible category of SLEMS knowledge base],
which [established that the object is neither member
of SLEMS simple nor extended objects category] as
partial solution to your query.

> *Moreover the assertion [tricycle is a beast] is a*

[taboo object] was found [valid] with [100.0]
certainty.

The bracketed contents in the above example are those retrieved from the SLEMS knowledge base by appropriate search operators. The rest are from a static template used to construct the explanation. In absence of expert supplied certainty factors the system adopts the default certainty factor of 100%. Finally an example in which the query consists of a fuzzy valued object is presented below.

> *Option:*

FUZZC

> *Enter subject:*

BAND1

> *Enter relation:*

has

> *Enter object:*

slightly less than 0.45 wavelength or

more than 0.52 wavelength

> *Sorry, that fuzzy valued attribute is in the*

taboo range for that object.

For such fuzzy queries the fuzzy comparison operator (FUZZC) which is the main module of the FUZZ subsystem invokes parsers which pre-process the fuzzy object and pass the fuzzy object components to a fuzzy object recognition module (**geoperator**). Once the fuzzy object is recognised its value part is retrieved and then passed to another module (**is_within_range**) for the actual comparison. Details of the theory and implementation of the fuzzy comparison operator are discussed in chapter 4.

3.5.1 Incomplete Queries and Partial Match Situations.

A partial match corresponds to the situation in which part of the query object being used as the search key is successfully matched Samet (1989). Alternatively it is defined in De Michiel (1989)[4] as the situation in which a real attribute value in a domain cannot be mapped into a single definite value, giving rise to a partial value instead. De Michiel (1989) also describes the results of search operations over partial match values as producing "maybe results" which is probably a misnomer for the term fuzzy results.

Within the context of the SLEMS query processor two aspects of the problem of incompleteness of queries are considered. The first deals with handling of partially specified query objects and has similar connotations to Samet's view of the partial match situation. In such cases no object in the knowledge base corresponds exactly

[4]De Michiel, 1989,..."A maybe tuple, is a tuple that cannot be excluded from the results of a query but that is not known with certainty to belong to it.... In a partial result, the partial value is characterised by a set of values of which exactly one must be correct."

to the query object because either the query contains less information or more information than the corresponding database objects. The SLEMS handles this type of incomplete query by matching on partial objects in a manner similar to those employed in information retrieval systems (Yager, 1989), which allow matching on partial words or keys.

The other aspect of query incompleteness and partial match situations arises when fuzzy queries or crisp queries and fuzzy knowledge base objects are involved. This situation corresponds more to De Michiel's (1989) view of partial match situations in query processing, and is handled by the SLEMS special fuzzy comparison operator FUZZC as explained in the above subsection.

The design of the SLEMS LEARN subsystem therefore provides a partial solution to some aspects of the partial match and incomplete query problems. The fuzzy comparison operator which facilitates the representation of fuzzy objects and searching operations on a data base of fuzzy objects is first introduced in the next section and the discussed in greater detail in chapter 4.

3. The FUZZ Subsystem in Brief.

In Samet (1989) a database query is defined as a request for all records that satisfy a predicate or have specific values or range of values for specified keys. Within the context of Samet's (1989) definition of database query, a direct link between the process of executing a query and that of determining set membership exists. Thus the criteria for satisfying a predicate or range of values for a specified key may be interpreted as a fuzzy set membership determination problem where the term set membership is used in the context of the definition of fuzzy sets and membership functions (Zadeh et al, 1975; Klahr, 1980; Kandel, 1986; Klir and Folger, 1988; Zadeh, 1989).

More specifically, the SLEMS knowledge base which consists of objects with fuzzy

valued attributes, can be partitioned for query processing purposes, into subsets with non-precisely specified bounds. Membership of object instances to these subsets define fuzzy partitions of the search space. In the SLEMS such fuzzy partitions are generated and then used to facilitate knowledge base searching and inferencing by the FUZZ subsystem.

The FUZZ subsystem consists of two separate, but cooperating modules, the **range_compare** and **range_taboo_compare** modules. The former performs searches on the permissible (MAY_OBJECTS) and mandatory (MUST_OBJECTS) active memory objects and attempts to match the fuzzy query object. The latter performs searches of the TABOO_OBJECTS and attempts to prove that the query object is an invalid domain object. If the object is not a fuzzy object the module attempts an ordinary match by invoking the **find_toke** and **find_token** smart string matching operators. If the query object is proved to be a TABOO_OBJECTS instance the system rejects the query immediately (see chapter 4 figure 4.8). If a valid fuzzy object is found, the operator invokes a special parser (**fuzzy_compare** and **is_within_range**) which parses the query object and performs the comparison. Special functions are invoked by the parser to evaluate the meaning of each fuzzy value involved in the comparison as explained later in chapter 4.

Chapter 4

A Fuzzy Partitions Approach to Fuzzy Query Processing.

4.0.1 Introduction.

Conventional databases consist of precisely specified facts and/or numerical values. Database query languages used to facilitate the storage and retrieval in databases, such as the SQL, impose a strict format for data entry and query, and does not permit ambiguity(Cuff, 1984; Feldman, 1984, Jiang and Lavington, 1985; Kowalski, 1984; Laender, 1984; Loizou and Thanisch, 1984; Martin, 1984; Schmidt, 1985; Whittington, 1988; De Michiel, 1989; Samet, 1990a, 1990b).

Real world data or knowledge, however, consists of precise and vague, ambiguous kinds. Methods exist for modelling concise real world knowledge and transforming it into the database management systems internal format (Whittington, 1988; Dutka, 1989; Samet, 1990a, 1990b). However, vague and imprecise real world knowledge does not lend itself to manipulation by the conventional methods of conceptual and physical data modelling (Zadeh et al, 1975; Zadeh, 1979, 1989; Whittington, 1988; De Michiel, 1989).

Until the introduction of the fuzzy sets theory by Zadeh in 1965 (Zadeh et al,

1975; Dubois and Prade, 1980) it can be said that there was no adequate method for the representation of vague information in databases. Since the introduction of the fuzzy set theory, however, much effort has been put into the development of theories and practical methods for representing fuzzy, non-precise, vague facts and facilitating intelligent queries to databases consisting of such objects (Zadeh et al, 1975; Zadeh, 1979, 1989; Dubois and Prade, 1979, 1980; Magrez and Smets, 1989; Dubois and Prade, 1990; Di Nola, 1990). Fuzzy set theory also provides a versatile tool for studying and analyzing complex systems (Zadeh et al, 1975; Baldwin, 1979; Dubois and Prade, 1979, 1980; Kandel, 1979; Adamo, 1980; Giles, 1980; Buckley, 1985; Chen, 1985; Scwartz, 1985; Baldwin, 1986; Piasecki, 1986; Zadeh, 1989; Dubois and Prade, 1990).

Once fuzzy knowledge has been meaningfully stored in a database, the database can be intelligently queried to provide solutions to real world problems through fuzzy inference (Zadeh et al, 1975; Zadeh, 1979, 1989; Magrez and Smets, 1989). Two principal fuzzy inference techniques exist. Zadeh's generalized modus ponens (GMP), which is the most widely used fuzzy inference method, is a forward inference scheme which works from premises to the conclusion. Modus tollens, on the other hand, is a reverse inference method which assumes the conclusion and infers the preconditions or premises (Zadeh et al, 1975; Cohen, 1985).

Knowledge representation entails representation of the uncertainty inherent in real world knowledge. Because uncertainty or vagueness cannot be assumed to have a normal distribution, the theory of probability is not appropriate for its study and manipulation (Zadeh et al, 1975; Cohen, 1985; Piasecki, 1986, Baldwin, 1986; Zadeh, 1979). Other theories have therefore evolved to handle uncertainty in real world knowledge (Baldwin, 1979, 1986; Cohen, 1985; Piasecki, 1986; Magrez and Smets, 1989; Zadeh, 1989). Most uncertainty handling methods are based on the Dempster-Shafer theory of evidence (Klir and Folger, 1988; Cohen, 1985; Caudil, 1990). The Dempster-Shafer theory provides a unified framework for thinking about uncertainty

and it complements the fuzzy set theory.

In this research a new fuzzy knowledge representation method is introduced. The method, referred to as the *fuzzy geometric partitions method*, is intended for the representation of vague numerical data such as "about 2", "more or less 7" etc. Such fuzzy expressions are said to constitute elastic constraints on the set of admissible real numbers (Dubois and Prade, 1980; Kandel, 1986). An arbitrary real number x which satisfies the elastic constraint is a "generic value" of the vague expression.

Theories on the representation of linguistically expressed knowledge and analysis of human perception of vagueness or fuzziness have been advanced by a number of researchers including Dahlgren (1988) and Nowakoska(1979). From Dahlgren's (1988) theory of naive semantics, generic information may be defined as the commonsense knowledge associated with nouns, locative relationships, typical materials forming an object, sizes of objects etc. In this study the term *generic value* is used to characterise crisp values lying within the vague intervals or fuzzy partitions associated with vague or fuzzy numbers such as "about 2", and "more or less 7".

As it will be shown later, the proposed fuzzy partitions approach differs substantially from other methods. The basic principle upon which the new method is founded is the simple, intuitive idea that, in commonsense reasoning, a vague statement such as "about 2" invokes a mental band of uncertain but reasonably narrow width around the crisp number 2, representing a vague set of reals.

It is assumed that, the human mind realises the vague interval by a process in which values picked out from the domain of real numbers are subconsciously compared to the crisp value 2 and rejected if they differ "too much" from it. The vague number, "about 2", can therefore, be represented by a generic binary relation $about(x, 2)$ where x is an arbitrary value from the real numbers which may or may not be equal to the generic number "about 2" depending on its "distance" from the crisp value 2. In this respect the binary relation constitutes a fuzzy partition of the real numbers space.

The partitions induced by fuzzy restrictions, and the derivation of generic objects

which are valid members of such fuzzy objects, constitute the basis of the fuzzy geometrical partitions method, and its use for the representation and comparison of fuzzy objects in connection with knowledge base inferencing and database search.

4.0.2 Motivation for the New Approach

Fuzzy set theory provides a precise mathematical framework for thinking about and representing vague knowledge and concepts (Zadeh et al, 1975; Dubois and Prade, 1980). For example, a vague predicate such as "about 2", can be expressed as a fuzzy number (Zadeh et al, 1975; Mizumoto et al, 1979; Mizumoto and Tanaka, 1979; Dubois and Prade, 1980).

Within the soil erosion management community vague expressions such as "annual soil loss is about 300 tons per acre" or "slope gradient of 2.5% to 7.1%" are frequently used to assign values to parameters used to model and compute soil loss in farmlands, rangelands and forest areas (Wischmeier and Smith, 1957; Smith and Wischmeier, 1957; Hóly, 1980; Goldman et al, 1986; Morgan, 1986).

The purpose of the suggested method is to store such vague data in its original format, in order to maintain its original meaning, and to perform searches on the resulting fuzzy database, to provide answers to queries which may also be vague. To achieve this, one must be able to perform direct comparison of fuzzy valued expressions.

To clarify, the nature and scope of the problem, consider a situation where an expert in the soil erosion field specifies that all areas with "annual soil loss much greater than 50 tons per acre" have a "serious erosion hazard potential". Assume that a data base of annual soil loss, referenced by spatial location, is available. To identify all areas which fall into the category "serious erosion hazard potential", would require the fuzzy restriction, "much greater than 50", to be evaluated so that a search routine can use it as the basis for comparison of all the soil loss values in the database.

In short the problem, which must be addressed, is the comparison of fuzzy numbers, and it includes determination of fuzzy number equality and inequality. A suitable program to do this must understand the meaning of such fuzzy valued expressions, and determine whether one fuzzy object is included in, equal to, or contains another fuzzy object.

Solutions to problems of this type involve modelling the fuzzy objects by standard membership functions (Dubois and Prade, 1980, Kandel, 1986), whose characteristic parameters are subjectively determined, on the basis of experience and desired results (Zadeh et al, 1975; Baldwin, 1979, 1986; Zadeh, 1979). Membership functions can however, also be constructed from other theories of uncertainty such as the probability theory (Kandel, 1979, 1986; Civanlar, 1986; Wang, 1990a, 1990b).

Once the membership functions have been constructed fuzzy sets and fuzzy numbers can be compared or tested for equality, using existing theories of fuzzy sets inclusion and composition of fuzzy relations (Zadeh et al, 1975; Kandel, 1979, 1986; Zadeh, 1979, 1989; Dubois and Prade, 1980; Klir and Folger, 1988). Relevant mathematical criteria and formulae for comparing fuzzy numbers, based on their standard characteristic functions, can be found in, for example, Dubois and Prade (1980, 1990) and Chen (1985a, 1985b). In this research, an attempt has been made to model fuzzy objects by fuzzy geometrical partitions.

The practicality of the suggested method as a tool for the direct manipulation of fuzzy query and knowledge base objects has been demonstrated on a test database. The concept of the fuzzy partitions induced by fuzzy restrictions on the elements of the universe X has been developed and implemented as the FUZZ subsystem of the SLEMS to handle vague or fuzzy knowledge input and fuzzy query processing.

4.0.3 Partitions Induced by Binary Relations in General.

The concise definition of the term fuzzy partition can be found in Dubois and Prade (1980). For the purposes of this research, fuzzy partitions associated with fuzzy

restrictions such as "much greater than 50", "slightly more than 18", "about 100" etc., will be defined and used to characterise and represent the meaning of these common vague numerical expressions. It will be shown later that, when appropriately specified, the fuzzy geometrical partitions provide the means for performing (direct) approximate comparison of the fuzzy objects for database search purposes.

Binary relations play an important role in the design of algorithms for database search and information retrieval (Minker, 1980; De Michiel, 1989). In Dowsing et al (1986) the equality operator is characterised as a binary operator which induces a diagonal subset or partition in the domain of discourse X. De Michiel (1989) extended the equality operator to facilitate comparison of definite and partial values.

Conventional manipulation of fuzzy knowledge is generally based on fuzzy sets theory. Such methods include the generalized modus ponens based on max-min composition of fuzzy relations (Zadeh et al, 1975; Baldwin, 1979; Kandel, 1979; Dubois and Prade 1980; Kandel, 1986; Klir and Folger, 1988; Zadeh, 1989; Di Nola and Sessa, 1990) and fuzzy modus ponens (Magrez and Smets, 1989).

This work builds upon the idea of the diagonal subset generated by the equality operator (Dowsing et al, 1986) and extends it to a general fuzzy comparison operator (GFCO). The GFCO, which is in general, a fuzzy binary relation or fuzzy restriction on some crisp argument, is defined and characterised in terms of the radial subsets or partitions induced by it in the universe of discourse X (figure 4.1).

It will be shown that, proper assignment of certain fuzzy constants which control the size of the partitions induced by each fuzzy restriction on the elements of the universe X, allows direct comparison of fuzzy objects and hence facilitates retrieval of vague database objects. Alternatively, fuzzy membership functions can be directly constructed from such partitions for the manipulation of the fuzzy objects by existing fuzzy set methods (Zadeh et al, 1975; Tsukamoto, 1979; Zadeh, 1979, 1989; Dubois and Prade, 1980; Kandel, 1986; Klir and Folger, 1988). A brief introduction to elementary fuzzy set theory concepts is given in the next section to provide a background

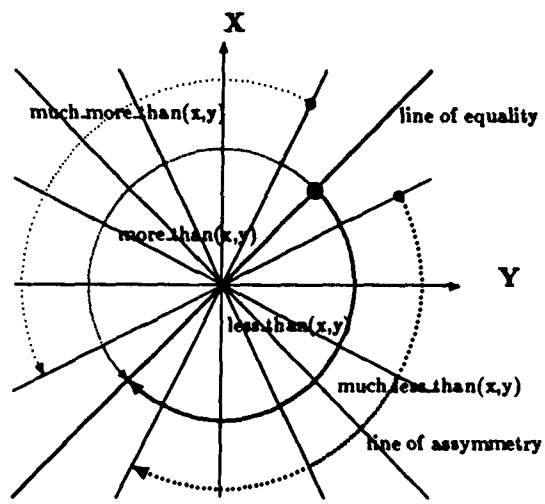


Figure 4.1: General Scheme of the Partitions Induced by Fuzzy Binary Relations

to the introduction of the fuzzy geometrical partitions method.

4.0.4 Comparison of Fuzzy Sets and Fuzzy Numbers in General.

A Fuzzy restriction, denoted by $R(x)$, is a relation R, that acts as an elastic constraint on the values which the variable $x \in X$ may take (Dubois and Prade, 1980; Klir and Folger, 1988). For example, the expression "much greater than x", is a fuzzy restriction on the values which x may take. When x is a numerical argument, the fuzzy restriction is a fuzzy number.

Generally the comparison of fuzzy sets and fuzzy numbers is achieved by the analysis of fuzzy sets inclusion (Dubois and Prade 1979, 1989) and fuzzy relations composition (Mizumoto et al, 1979; Dubois and Prade, 1980, 1989). Various forms of fuzzy inclusion theorems exist. They include Zadeh's inclusion theorem which states that

$$A \subseteq B \text{ iff } \forall\, x \in X,\ \mu_A(x) \leq \mu_B(x)$$

and fuzzy set weak inclusion defined by

$$A \leftarrow_\alpha B \text{ iff } x \in (\bar{A} \cup B)_\alpha \,\forall\, x \in X.$$

The Zadeh inclusion theorem is a direct extension of the classical set inclusion theorem. Weak inclusion states in effect that A is weakly included in B as soon as all elements of X α-belong to $\bar{A}$ or to B where an element is said to α-belong to a fuzzy set A *iff* $x \in A_\alpha$ (Dubois and Prade, 1979, 1980). Details on other inclusion theorems and comparison of fuzzy sets and fuzzy numbers may be found in Dubois and Prade (1980).

The membership value $\mu_A(x)$ expresses the degree of compatibility of the fuzzy set A with the element x of X which is non-fuzzy. The compatibility of one fuzzy set

to another is defined (Dubois and Prade, 1980) by

$$\mu_\tau(u) = \sup_{x:u=\mu_A(x)} \mu_B(x) \; \forall u \in [0,1],$$

or

$$\tau = \mu_A(B) = \sum_X \mu_B(x)/\mu_A(x).$$

It expresses the membership value in A of the element that mostly belongs to B and is thus a fuzzy inclusion index (Dubois and Prade, 1980).

Assuming that the characteristic functions $\mu_M(x)$ and $\mu_N(y)$ of two fuzzy restrictions are known, their comparison can be achieved by means of equations 4.1 and 4.2 derived in Dubois and Prade (1980).

$$\mu_{M \geq N}(x,y) = \sup_{(x,y):x \geq y} \min(\mu_M(x), \mu_N(y)) \tag{4.1}$$

$$\begin{cases} N \geq_\theta M, & \text{iff } n - m > \beta + \gamma, \\ M \geq_\theta N, & \text{iff } m - n > \alpha + \delta. \end{cases} \tag{4.2}$$

where $M = (m, \alpha, \beta)_{LR}$, $N = (n, \gamma, \delta)_{RL}$, are RL and LR type (Dubois and Prade, 1980) membership functions; m is the peak-point, and α and β are respectively the left and right spreads of M; n is the peak-point and γ and δ are respectively the left and right spreads of N; and θ is some appropriate threshold (ie. an α-cut). The fuzzy numbers M and N are said to be approximately equal when $\min(\mu_{(M \geq N)}, \mu_{(N \geq M)}) \geq \theta$.

4.0.5 Fuzzy Logic and Fuzzy Inference.

Fuzzy logic is a special kind of many valued logic (Baldwin, 1979; Kandel 1986). Many valued logic is a type of logic which admits of the possibility of vague or multiple truth values. Because in fuzzy logic truth value is a vague variable, classical inference rules and implication are not applicable (Zadeh et al, 1975; Goguen, 1979).

Fuzzy logic implication provides the means for inferring new information or making conclusions from some existing fuzzy premises (Zadeh et al, 1975; Magrez and Smets,

1989; Zadeh, 1989). Two type of implication are generally used, *modus ponens* and *modus tollens* (Klir and Folger, 1988; Magrez and Smets, 1989). Modus ponens, defined in Eq. 4.4, forms the basis of Zadeh's generalized modus ponens (GMP) implication (Zadeh et al, 1975; Dubois and Prade, 1980; Magrez and Smets, 1989; Zadeh, 1979, 1989).

$$\begin{aligned} \text{Implication rule} &: \text{"}x \text{ is } A \rightarrow y \text{ is } B\text{"} \\ \text{given fact} &: \text{"}x \text{ is } A^*\text{"} \\ \text{conclusion} &: \text{"}y \text{ is } B^*\text{"} \end{aligned} \tag{4.3}$$

where A^* and B^* are fuzzy predicates.

An approach to fuzzy inference recently proposed by Magrez and Smets (1989) differs from Zadeh's method substantially. The new method called the fuzzy modus ponens (FMP) is based on four principles called the four favourable characteristics of a fuzzy inference operator (Magrez and Smets, 1989). The first property, called, the fundamental property of inference, states that the inference operator should preserve information even when the predicates are fuzzy. The second property, referred to as total indetermination, states that, a complete indetermination on the consequent domain must be the result of the inference. The third property is the subset property, which requires that, the consequent of an implication may not be more precise than the fact from which it was derived. The fourth property, called the shape of indetermination, implies an independence of the shape of indetermination of the consequent from that of the fact used to derive it.

Magrez and Smets (1989), claims that the characteristics briefly outlined above, produce an inference operator which captures the commonsense meaning of the original fuzzy predicates.

Whether one uses the GMP or the FMP approach to fuzzy inference the first stage in the process of fuzzy knowledge representation is the transformation of the linguistic fuzzy predicates involved into fuzzy sets or fuzzy numbers. This process

involves a subjective assignment of membership characteristic functions or fuzzy truth values as explained in various literature sources including, Zadeh (1975, 1979, 1989), Kaufmann (1975), Dubois and Prade (1979, 1980, 1989), Goguen (1979), Baldwin (1979), Mizumoto, Fukami, and Tanaka (1979) and Tanaka et al (1979).

4.1 Theoretical Basis of the Fuzzy Partitions Method.

The theoretical basis for the method proposed in this thesis are the axioms of first order predicate calculus and fuzzy set theory. In Kaufmann (1975), the concept of the fuzzy subset M, induced by the binary relation $y \gg x$ in the cartesian $R_+ \times R_+$ space, is demonstrated for $y = kx$, where $k \geq 1$, and shown to constitute a radial partition of the $R_+ \times R_+$ where R_+ are the non-negative real numbers.

Using axioms of first-order predicate calculus, outlined in Dowsing et al (1986), Delahaye (1986) and others, extensions can be made to the predicate (PRED) and function symbol (FN) subsets of the predicate language $\mathcal{L}$ as follows:

- Let the set FUZ_PRED be defined as

  FUZ_PRED ={EQUAL_TO, GREATER_THAN, LESS_THAN,
  SLIGHTLY_LESS_THAN, SLIGHTLY_MORE_THAN,
  MUCH_LESS_THAN, MUCH_GREATER_THAN, ABOUT,
  MORE_OR_LESS, ROUGHLY}

  Then the extended predicate subset PRED of the language $\mathcal{L}$ is

  $$PRED = \{PRED_o + FUZ\_PRED\}$$

  where $PRED_o$ is the original set of predicates in Dowsing et al (1986).

- Further more if it is explicitly required that the set of function symbols FN of $\mathcal{L}$ include the new functions $FNNEW = \{partition\}$ where *partition* is a special

function which assigns to a partition label from the PARTT set, the partition induced by the predicate in X. Then the extended set of function symbols FN is

$$FN = \{FN_o + FNNEW\}$$

where FN_o is the set of original function symbols used in Dowsing et al (1986).

Also, let R denote any predicate or relation in the PRED set, a new subset PARTT, is added to the disjoint subsets of the predicate language $\mathcal{L}$ as defined in Dowsing et al (1986), such that PARTT_R is the label of the radial partition induced by the predicate in X, and it is given by

$$PARTT = \{PARTT_R \mid R \in PRED\}.$$

In the first order predicate logic, an interpretation, is the process of specifying the means by which predicates are to be evaluated and the values which their variables may take. An interpretation, I of a predicate language $\mathcal{L}$, is defined in terms of four rules specified in Dowsing et al (1986) and Delahaye (1986). Essentially, when these four rules are satisfied, the result is a specification or description of the variables and the universe in which they are valid.

To assign values to the variables requires another process called valuation, which has a component in both X and the truth valuation domain. Procedures for undertaking valuations are described in Dowsing et al (1986) and Delahaye (1986).

4.1.1 Definition of the Equality Operator

Dowsing et al (1986) describes and defines the equality operator, =, as follows:

Definition 4.1.1 (The Equality Operator (=)) *For an interpretation I, with universe X, the set $=_I$ on which $=$ is to be true must be the diagonal subset $\{(x, x) \mid x \in X\}$ of $X \times X$.*

Figure 4.2 shows the diagonal set induced by the equality operator, obviously $x = y$ in the diagonal set.

Dowsing et al (1986) also gives the following interpretation to the operator $>$, which correspond to the predicate GREATER_THAN.

Definition 4.1.2 (The $>$ Operator) *If* greater, *is interpreted as the relation $>$ on real numbers then* greater *will be associated with the pairs (x, y) of real with $x > y$ in the usual sense.*

In other words this is the set on which the predicate *greater* is intended to be true. Further more *greater*(1^{st}term, 2^{nd}term) is evaluated true *iff* 1^{st}term $>$ 2^{nd}term is found to be true in the universe of discourse (Dowsing et al, 1986). These definitions are used in the next section as the basis for defining the general fuzzy comparison operator (GFCO).

4.1.2 Definition of the General Fuzzy Comparison Operator.

Within the target application area of soil loss estimation and modelling, crisp and fuzzy values involved are usually positive quantities. All definitions and valuation sets necessary for the comparison of fuzzy valued objects in this study are, therefore, restricted to the first quadrant of the XY plane (see Fig. 4.1). Although it is possible to extend the definitions to include all quadrants, this will not be pursued at this stage.

Referring to definition 4.1.1 and 4.1.2, it is easy to see that the following definitions for the predicates GREATER and LESS are natural extensions of the Dowsing et al (1986) definition of the equality operator.

Definition 4.1.3 (GREATER_THAN) *For an interpretation I, with the universe X, the set $>_I$ on which $>$ is to be true must be the upper diagonal subset $\{(x, y) \mid x \in X,\ y \in X,\ x > y\}$ of X.*

Definition 4.1.4 (LESS_THAN) *For an interpretation I, with the universe* **X**, *the set $<_I$ on which $<$ is to be true must be the lower diagonal subset $\{(x,y) \mid x \in X,\ y \in X,\ x < y\}$ of $X \times X$.*

Let any predicate in set $PRED$ be denoted by the symbol R, then R is a general fuzzy comparison operator. By a natural extension of the definition of the equality operator (def. 4.1.1), considered for this purpose a special case of the general fuzzy comparison operator, the general comparison operator is now defined.

Definition 4.1.5 (General Comparison Operator) *For an interpretation I, with the universe X, the set R_I on which R is to be true must be a radial or sectoral subset $\{(x,y) \mid x \in X,\ y \in X,\ xRy\}$ of $X \times X$.*

Based on definition 4.1.5) members of subset R_I are now concretely specified.

Definition 4.1.6 (MUCH_GREATER_THAN) *Let R, in definition 4.1.5, be the predicate MUCH_GREATER_THAN of the PRED set, then:*

$$MUCH\_GREATER\_THAN_I = \{(x,y) \mid x \in X,\ y \in X,\ x \gg y\}$$

where $\gg$ has the usual meaning much greater than.

Definition 4.1.7 (MUCH_LESS_THAN) *Let R, in definition 4.1.5, be the predicate MUCH_LESS_THAN of the PRED set, then:*

$$MUCH\_LESS\_THAN_I = \{(x,y) \mid x \in X,\ y \in X,\ x \ll y\}$$

where $\ll$ means much less than.

When referring to linguistic fuzzy predicates such as "more or less", etc. in thesis, the convention *more_or_less* etc. will be used. Note *greater_than*(x) induces a partition which is the superset of the one induced by *much_greater_than*(x) (figure 4.3). The full set of definitions for the comparison operators in the set $PRED$ is given in

table 4.2. Each radial subset defined in table 4.2 corresponds to a radial partition of the universe X, such that, to each fuzzy predicate $R \in PRED$, there corresponds a unique partition, $\text{PARTT}_R \in \text{PARTT}$. Some of these partitions are shown in figure 4.3 and figure 4.4.

The partitions in figure 4.4 represent terms whose commonsense meaning induces a symmetrical "stretch" about the crisp argument x of the object, such as, *about*(x), *more_or_less*(x), *roughly_equal_to*(x), and *approximately*(x). Note also that unlike the partitions in Kaufmann (1975), these partitions are generated in the universe X and not in the valuation space.

Each partition is identified by a unique label (Eq. 4.4) from the PARTT set, and its boundaries are specified in terms of the angular parameters shown in figure 4.5.

$$PARTT_R \mid PARTT_R \in PARTT;\ PARTT_R \rightarrow R \in PRED\}. \tag{4.4}$$

The notation $PARTT_R$ will be used every where to denote the label of the partition induced by the relation R. Note that the partition induced by a relation R, is defined on the domain $[0, \pi/2]$, while the partition label is defined in the PARTT subset. R will be used to represent both a general predicate in the PRED set and a general fuzzy comparison operator.

Using this notation the diagonal set, $=_I$, induced by the equality operator (Dowsing et al, 1986), corresponds to the partition label $PARTT_=$ of the PARTT set. Similarly $PARTT_>$, $PARTT_<$, $PARTT_{\gg}$, and $PARTT_{\ll}$ are the labels for the partitions induced by GREATER_THAN, LESS_THAN, MUCH_GREATER_THAN, and MUCH_LESS_THAN respectively.

In the next section an adhoc application of the definitions introduced above is used to describe and specify the operators and valuation sets necessary for the proposed method.

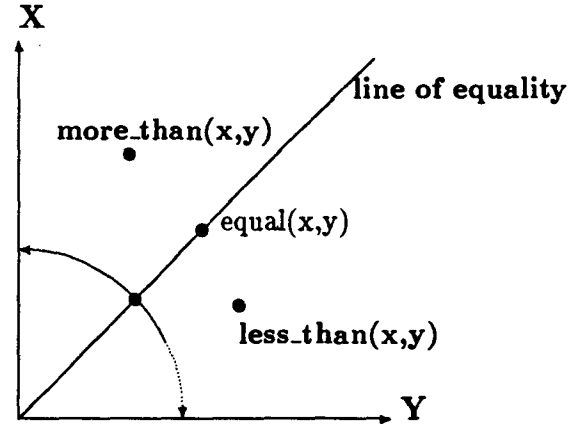
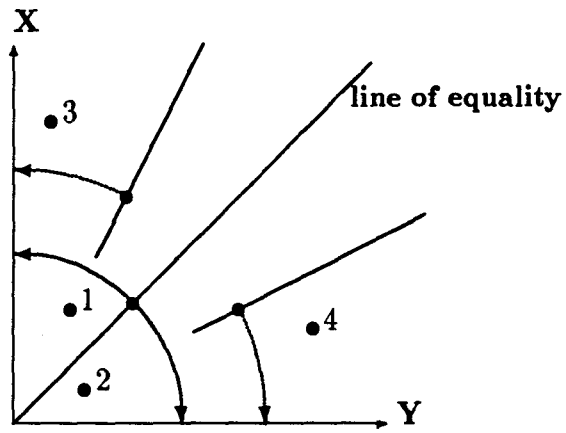
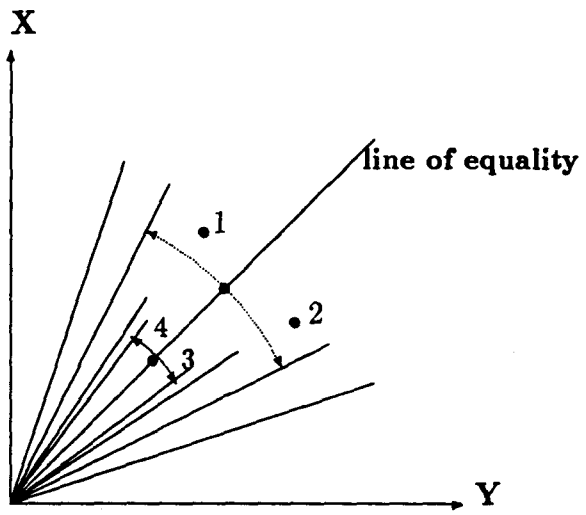


Figure 4.2: The Diagonal Partition and Basic Fuzzy Partitions Induced by the Equality Operator.



1: more_than(x,y)
2: less_than(x,y)
3: much_more_than(x,y)
4: much_less_than(x,y)

Figure 4.3: Some Common Fuzzy Restrictions Corresponding to a Refinement of the Basic Fuzzy Partitions.



1,2: about(x,y)
4: slightly_more_than(x,y)
3: slightly_less_than(x,y)

Figure 4.4: Refinement of the Fuzzy Partitions of the Real Numbers Universe of Discourse.

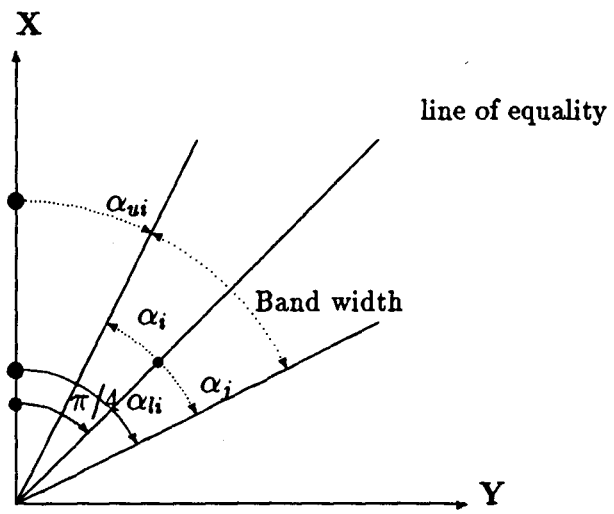


Figure 4.5: The Geometrical Parameters of the Partitions Induced by an Arbitrary Fuzzy Binary Relation

4.2 Specification and Assignment of the Fuzzy Partitions.

4.2.1 Specification.

For the purposes of specifying and assigning the partitions induced by a fuzzy predicate R, some associated geometrical quantities are first defined in figure 4.5. The partition induced by the fuzzy relation R is then specified by equations 4.5 and 4.6,

$$\alpha_{ui} = \pi/4 - \alpha_i. \tag{4.5}$$

$$\alpha_{li} = \pi/4 + \alpha_j. \tag{4.6}$$

where α_i, α_{ui}, α_{li}, and α_j are the angular parameters defining the radial fuzzy partitions induced on X by the arbitrary fuzzy relation R(Fig. 4.5).

Further more, from Eq. 4.5 and Eq. 4.6, the following geometrical relationships are obvious:

$$\alpha_i = \pi/4 - \alpha_{ui}. \tag{4.7}$$

$$\alpha_j = \alpha_{li} - \pi/4. \tag{4.8}$$

where $\pi = 3.14159$. Using Eq. 4.5 to Eq. 4.8 the half width of the symmetric partition about the line of equality, induced by the fuzzy relation R on X can be computed. The radial partition corresponding to a partition label $PARTT_R$ in set PARTT is given by

$$partition(PARTT_R) = \{\alpha_{uR}, \alpha_{lR}\}. \tag{4.9}$$

where, the subscripts u and l, stand for upper bound and lower bound of the partition respectively (figure 4.5). By this definition, the apex of the sector representing the partition, is at the origin of the universe, where origin is the point $(x = 0, y = 0)$ in X.

Considering only the partitions in the first quadrant of the XY plane (figure 4.5), let the set of partitions labelled by the $PARTT$ set be evaluated in the domain $[0, \pi/2]$, such that $\alpha_u \in [0, \pi/2]$ and $\alpha_l \in [0, \pi/2]$. Then the partition corresponding to the diagonal set $=_I$ is

$$partition(PARTT_{=}) = \{\pi/4, \pi/4\}.$$

Note the use of the convention $\pi/4, \pi/4$ rather than simply $\pi/4$ to reflect the fact that the 45°-line (line of equality) is for this purpose, the limiting case as the width of a symmetrical partition about the line of equality tends to zero (Eq. 4.10).

$$partition(PARTT_{=}) = \lim_{\substack{\alpha_u \to \pi/4, \\ \alpha_l \to \pi/4}} partition(PARTT_R). \tag{4.10}$$

Using similar arguments the subsets induced by GREATER_THAN, ($>_I$), and LESS_THAN, ($<_I$), in X are given by Eq. 4.11 and Eq. 4.12 respectively.

$$\begin{aligned} partition(PARTT_{>}) &= \lim_{\substack{\alpha_u \to 0, \\ \alpha_l \to \pi/4-}} partition(PARTT_R) \\ &= \{0, \pi/4-\}. \end{aligned} \tag{4.11}$$

$$\begin{aligned} partition(PARTT_{<}) &= \lim_{\substack{\alpha_u \to \pi/4+, \\ \alpha_l \to \pi/2}} partition(PARTT_R) \\ &= \{\pi/4+, \pi/2\}. \end{aligned} \tag{4.12}$$

where $\pi/4-$ is infinitesimally smaller than $\pi/4$ and $\pi/4+$ is infinitesimally greater than $\pi/4$.

The partitions corresponding to the labels $PARTT_{=}$, $PARTT_{>}$ and $PARTT_{<}$ defined above are referred to in this study as basic partitions. Using the same process refinements of the basic partitions can be achieved (figures 4.3 and 4.4). The width of the partition induced by the fuzzy predicates can be controlled by subjectively changing the constant parameters in table 4.1. In the next section some guidelines for the assignment of these constants, which serve as control for the size of the partitions generated by each fuzzy relation in X, are given.

4.2.2 Considerations for Assigning the Fuzzy Partitions.

Based on the Magrez and Smets (1989) favourable properties of a fuzzy inference operator, the best definition of a particular fuzzy object will be given by the partition which captures best the commonsense meaning of the fuzzy restriction representing the object. How to achieve or determine the best partition using this principle is a problem which is not fully addressed in this research but this does not detract from the potential use of the concept to provide intuitive guidelines.

When assigning the partitions it is useful to consider consistency with conventional knowledge representation by fuzzy membership functions as specified in Zadeh (1989). As noted in Klir and Folger (1988) and Kandel (1986), the usefulness of fuzzy sets for modelling a concept, class or a linguistic label depends on the appropriateness of its membership function. Tentatively two ways for managing this problem are proposed. In the first method specification of the partitions is done arbitrarily, semi-guided by the desire to obtain partitions which agree with the common sense meaning of the vague expressions. This is the approach implemented in SLEMS as a first trial. Also one may construct fuzzy membership functions from the partitions to facilitate valuation of the fuzzy partition assignments. This approach has also been used in the study and is discussed more in section 4.5.1.

The second approach requires the use of techniques involving direct calibration of the user to establish his/her impressions of what the meaning of the fuzzy predicates are. After specifying the initial partitions one can undertake numerous statistical tests as suggested in Klir and Folger (1988) and Nowakoska (1979), to determine the most acceptable meaning for each fuzzy object. This alternative has not been developed in this study and is discussed briefly in the conclusion part.

4.3 Practical Assignment of the Fuzzy Partitions.

From figure 4.4 and figure 4.5 it is clear that the larger the difference between the conceptual values of the fuzzy objects involved in a fuzzy comparison the larger the parameters α_i and α_j . Subjective values can therefore be assigned to these angular parameters, for each fuzzy predicate such that this qualitative relation is preserved.

Using these general guidelines the characteristic values for each partition were assigned in terms of fuzzy constants WIDE CLOSE, VERYWIDE, VERYCLOSE, etc. as shown in table 4.1 and table 4.2. These constants can be thought of as generic band widths associated with the characteristic functions of the fuzzy sets "wide", "close", "very wide", "very close" etc. Their values can therefore be changed to accommodate a different interpretation of the fuzzy sets they define.

Definition 4.3.1 (Generic Value) *Let R be any predicate in the PRED set. Then the unary fuzzy predicate $R(x)$ is said to induce a generic value $\mathbf{x} \in X$ from $x \in X$, such that the expression $\mathbf{x} = R(x)$ or equal$(\mathbf{x}, R(x))$ evaluated over X is true.*

Note that, a predicate such as *greater_than*(x), has a binary connotation represented by the "mental" binary relation *greater_than*$(\mathbf{x}, x)$ where $\mathbf{x}$ is a generic value induced by the relation *greater_than* on its crisp argument x. In table 4.2 the angles corre-

	VERYWIDE	WIDE	CLOSE_1	CLOSE	VERYCLOSE
WIDTH	$\pi/3$	$\pi/6$	$\pi/12$	$\pi/24$	$\pi/48$

Table 4.1: Specification for the Constant Fuzzy Terms.

sponding to boundaries of the fuzzy partitions i.e. α_u and α_l are measured clockwise from the X-axis (figure 4.5). The notation WIDE, VERYWIDE etc is used to denote partition labels while the emphasized notation $WIDE$, $VERYWIDE$, etc. represents (absolute) angular measures from the line of equality.

fuzzy predicate	lower bound α_l	upper bound α_u
more_than (x)	$\pi/4-$	0
less_than (x)	$\pi/4+$	$\pi/2$
much_more_than (x)	$\pi/4 - VERYWIDE$	0
much_less_than (x)	$\pi/2$	$\pi/4 + VERYWIDE$
slightly_more_than(x)	$\pi/4 - VERYCLOSE$	$\pi/4+$
slightly_less_than(x)	$\pi/4 + VERYCLOSE$	$\pi/4-$
more_or_less (x)	$\pi/4 + CLOSE_1$	$\pi/4 - CLOSE_1$
about (x)	$\pi/4 + CLOSE$	$\pi/4 - CLOSE$
roughly (x)	$\pi/4 + WIDE$	$\pi/4 - WIDE$

Table 4.2: Fuzzy Partitions Induced by Various Fuzzy Predicates as Functions of Generic Band Widths.

4.3.1 Partitions Induced by Specific Fuzzy Predicates

Using the parameters in tables 4.1 and 4.2, partitions induced by specific fuzzy predicates can be derived. For example, the partition induced by the fuzzy predicate MUCH_GREATER_THAN(x) or $\gg x$), is given by Eq. 4.13

$$\begin{aligned} partition(PARTT_{\gg}) &= \lim_{\substack{\alpha_u \to 0, \\ \alpha_l \to \pi/4 - VERYWIDE}} partition(PARTT_R) \\ &= \{0, \pi/4 - VERYWIDE\}. \end{aligned} \tag{4.13}$$

and for MUCH_LESS_THAN(x) or $\ll x$ by Eq. 4.14.

$$\begin{aligned} partition(PARTT_{\ll}) &= \lim_{\substack{\alpha_u \to \pi/4 + VERYWIDE, \\ \alpha_l \to \pi/2}} partition(PARTT_R) \\ &= \{\pi/4 + VERYWIDE, \pi/2\}. \end{aligned} \tag{4.14}$$

Similarly partitions induced by the other predicates in the PRED set can be obtained (table 4.2). Synonyms of the natural language predicates corresponding to the predicates in the PRED set such as *more_than*, *longer_than*, *wider_than* etc. and *lower_than*, *shorter_than*, *thinner_than*, *narrower_than* etc, are mapped into the same partitions as those for *greater_than* and *less_than* respectively.

4.3.2 Refinement of the Fuzzy Partitions Specifications.

During practical implementation and testing of the method, it was observed that, simply using the angular values in table 4.2 gave partitions which produced unsatisfactory results. A decision was, therefore, made to introduce further refinements into their specification by logarithmically scaling the base parameters. For example, as seen in the equations in table 4.3, the definition of *more_or_less* uses three different variations in which $CLOSE$ is defined as a piecewise continuous trigonometric logarithmic function. These variations were introduced to accommodate the change in perception of the differences between numerical quantities as the magnitudes involved change from very small to very large.

The decision to apply logarithmic scaling is purely intuitive, and is based on similar practice in image and photographic processing in which logarithmic functions are used to model the human physiological response to light stimulus (Land et al, 1989). In addition the partitions actually employed to compute the meaning of vague objects in the SLEMS, are modifications of the constant parameters in table 4.2, as shown in table 4.3.

Fuzzy Predicate	α_i or αj	Symbolic value
slightly_more_than(x)	$VERYCLOSE/(2+\log(x))$	$VERYCLOSE'$
veryclose_to(x)	$VERYCLOSE/(4+\log(x))$	$VERYCLOSE''$
much_more_than(x)	$VERYWIDE/2+$ $VERYCLOSE/(4+\log(x))$	$VERYWIDE'$
much_less_than(x)	$VERYWIDE/2+$ $VERYCLOSE/(4+\log(x))$	$VERYWIDE'$
roughly(x) for $x<5$	$CLOSE/(1+\log(x))$	$CLOSE''$
for $x<10$	$CLOSE/(2+\log(x))$	
for $x>10$	$CLOSE/(3+\log(x))$	
about(x)	$CLOSE/(2(1+\log(x)))$	$CLOSE'$
more_or_less(x) for $x<5$	$5\times CLOSE_1/(12+\log(x))$	$CLOSE_1'$
for $x<10$	$CLOSE_1(3+\log(x))$	
for $x>10$	$CLOSE_1(4+\log(x))$	

Table 4.3: Computation of Generic Band Widths From the Basic Parameters

4.3.3 Scope and Limitations of Application of the Method.

The main concern in this chapter is the comparison of fuzzy valued expressions. No attempt will therefore be made to use this partition scheme for the representation of fuzzy expressions such as *very_beautiful*(x) and *very_wide*(x) where x has no numeric evaluation. Evaluation of such fuzzy terms can be made using rules of translation of fuzzy quantifiers (Kaufmann, 1975; Dubois and Prade, 1979; Kickert, 1979; Shmucker, 1984; Kandel, 1986; Klir and Folger, 1988; Zadeh, 1989). Evaluation of fuzzy relations such as, "much_longer(1^{st}slope element, 2^{nd}slope element)", cause no problem since in common sense reasoning what is actually being compared are the slope gradients of the two slope elements.

4.4 Fuzzy Partitions Based Comparison of Fuzzy Objects.

For the purposes of evaluating fuzzy restrictions on the elements of the universe X and comparing fuzzy numeric valued objects, direct use of the partitions defined in the domain $[0, \pi/2]$ can be applied. In this respect the comparison of fuzzy valued objects can be reduced to the determination of the partition corresponding to the fuzzy predicate. This can be done by means of the concept of the generic value of a fuzzy restriction defined (def. 4.3.1) in section 4.3.

By means of the generic value concept, direct comparison of fuzzy restrictions on the elements of X is possible. For example, if the query object is *much_greater_than*(x_1), and the database contains the fuzzy object *greater_than*(y_1), the query will be satisfied provided that, a generic value **y**, exists in the subset $\{y > y_1\}$, induced by the predicate *greater_than*(y_1) in X, such that it falls in the partition $\{x \gg x_1\}$ induced by the predicate *much_greater_than*(x_1) . In other words if **x** is the generic value induced by *much greater than*(x_1), then $\mathbf{y} \geq \mathbf{x}$ is a valid

solution to the query.

It is easy to demonstrate the validity of the above argument by substituting, say, 5 for x_1 and 50 for y_1, and adopting the convention that, "one thing is much greater than another one if it is at least more than ten times as much". Then obviously "greater than 50" is in this case much greater than 5 and hence the query is satisfied. Note however that "equal to 50" would not satisfy the criteria for selection in this case.

Using the generic value it is thus possible to make approximate comparison of fuzzy numbers. The fuzzy partitions based solution, therefore, avoids some of the problems elaborated in Magrez and Smets (1989), and Mizumoto et al (1979), concerning the use of the generalized modus ponens. Specifically, this method satisfies the third favourable characteristic of a fuzzy inference rule (Magrez and Smets, 1989) which requires that, the hedge "very", should not be transmitted to the conclusion. Also, since in this method the comparison of fuzzy objects does not involve use of the characteristic membership function, the method satisfies the fourth favourable characteristic of fuzzy inference(Magrez and Smets, 1989) which states that the shape of indetermination (uncertainty) is not relevant to the conclusion. In other words if one is interested in objects which are "much greater than 5" in the above sense, then whether an object is equal to 51 or 100 is immaterial since they both equally satisfy the selection criteria.

4.4.1 Formulae for the Computation of the Fuzzy Partitions.

The functions used by the SLEMS fuzzy comparison module to compute generic values induced by the fuzzy predicates of the PRED set are summarised in table 4.4. These functions serve as the basis for the approximate comparisons of fuzzy valued predicates by the SLEMS general fuzzy comparison operator implemented as the

FUZZ sub-system in SLEMS.

To retrieve a desired fuzzy object the fuzzy comparison operator first constructs the partitions induced by the corresponding fuzzy predicate. Noting for example, that the predicate *equal to* x, induces a generic value $\mathbf{x}$ as defined above, any object $\mathbf{y}$ in X, such that it is equal to the generic value $\mathbf{x}$ is a solution to the query. The tuple $(\mathbf{x}, \mathbf{y})$, is the desired solution and it satisfies the condition;

$$\arctan(\mathbf{y}/\mathbf{x}) = \pi/4.$$

This is so because the two crisp values $\mathbf{x}$ and $\mathbf{y}$ can only be equal if the point $(\mathbf{x}, \mathbf{y})$ is member of the diagonal set, according to the axioms of equality (Dowsing et al, 1986). Also, because the predicate *equal_to* does not actually fuzzify the crisp value x, this computation is fairly straight forward.

In support of this simple approach to the comparison of fuzzy valued objects reference is made again to the fourth favourable property of fuzzy inference rules (Magrez and Smets, 1989) according to which the shape of indetermination is irrelevant to the shape of the conclusion.

4.4.2 Comparison of Mixed Fuzzy Predicates.

Figures 4.6 and 4.7 shows general geometrical relationship between crisp values and induced generic values based on two specific fuzzy restrictions. Comparison of mixed fuzzy objects using generic values can be easily demonstrated. Consider the predicate *more_than*(x_1) or *less_than*(x_1) representing a query, and the database objects *greater-_than*(y_1) and *less_than*(y_1). Taking an arbitrary generic value $\mathbf{x}$ associated with the fuzzy restriction *greater_than*(x_1), the condition which must be satisfied for $\mathbf{x}$ to be an acceptable generic value of the fuzzy query object is (Eq. 4.15)

$$\arctan(x_1/\mathbf{x}) = \alpha_l < \pi/4. \tag{4.15}$$

where, α_l (refer to fig. 4.5), is the angle of the lower bound of the partition induced by *greater_than*(x_1).

Similarly, for the fuzzy query object *less_than*(x_1) the condition (Eq. 4.16)

$$\arctan(x_1/\mathbf{x}) = \alpha_u > \pi/4. \tag{4.16}$$

must be satisfied, where α_u (refer to Fig. 4.5), is the angle of the upper bound of the partition induced by *less_than*(x_1); $\mathbf{x}$ is taken along the generic values axis and x_1 along the crisp values axis respectivelly (see figure 4.6).

The same procedure is applied to the crisp arguments of the database objects *greater_than*(y_1) and *less_than*(y_1) to derive the corresponding generic objects $\mathbf{y}$. The comparison is then performed with the generic values, $\mathbf{x}$ and $\mathbf{y}$, induced from the crisp arguments of the fuzzy query and database objects respectively.

Since the lower and upper bounds of the fuzzy predicates *greater_than*(x_1) and *less_than*(y_1) are respectively x_1 and y_1, the comparison of the generic values can in this case be approximated with a comparison on the crisp arguments of the two predicates. More precisely, the condition to be satisfied, for the comparison of the query object *greater_than*(x_1) and the fuzzy database object *less_than*(y_1) to return success, is given by (Eqs. 4.17 and 4.18)

$$\mathbf{y} > \mathbf{x} \tag{4.17}$$

implying,

$$\arctan(\mathbf{x}/\mathbf{y}) < \pi/4. \tag{4.18}$$

where, $\mathbf{y}$, the generic object corresponding to the "unknown" fuzzy database object is, in this case, taken along the generic objects axis while $\mathbf{x}$ the generic value corresponding to the users "hazzy" query value is taken along the crisp objects axis (see figure 4.6 for a description of the axes).

By substituting 30 for x_1 and 50 for y_1 it is clear that the object "less than 50" satisfies the query "greater than 30" even if the user was actually thinking of, for example, the value 40.

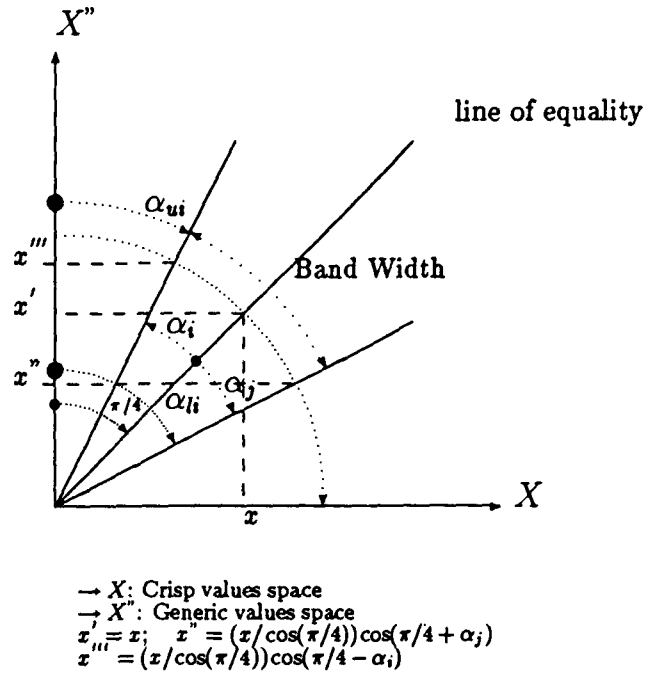
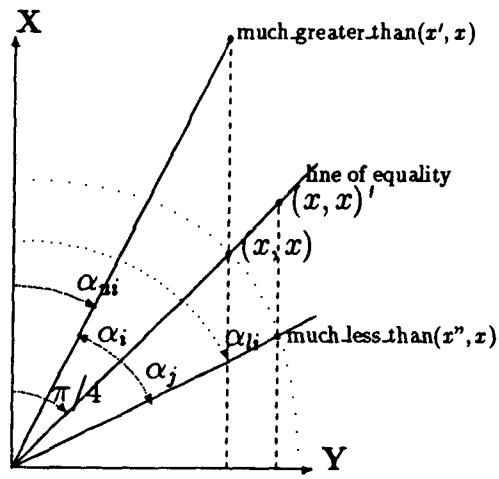


Figure 4.6: The Geometric Relation Between Crisp Values and Generic Values Induced by a Symmetric Fuzzy Restriction on the Elements of the Real Numbers Universe of Discourse



$$x' = (x/\tan(\pi/4 - \alpha_i)$$
$$x" = (x/\tan(\pi/4 + \alpha_j)$$

Figure 4.7: The Geometric Relation Between Crisp Values and Generic Values Induced by the Fuzzy Restrictions much_greater_than(x) and much_less_than(x) on the Elements of the Real Numbers Universe of Discourse

Fuzzy Predicate	Equation for the Generic value
slightly_less_than(x)	$x = x - D(x'')$; $D(x'') = abs(x'' - x)$; $x'' = x/\cos(\pi/4)\cos(\pi/4 - VERYCLOSE')$
slightly_more_than(x)	$x = x + D(x'')$; $D(x'') = abs(x'' - x)$; $x'' = x/\cos(\pi/4)\cos(\pi/4 + VERYCLOSE')$
very_close(x)	$x = x \pm D(x'')$; $D(x'') = abs(x'' - x)$; $x'' = x/\cos(\pi/4)\cos(\pi/4 - VERYCLOSE'')$
much_less_than(x)	$x = x - D(x'')$; $D(x'') = abs(x'' - x)$; $x'' = x/\tan(\pi/4 + VERYWIDE')$
much_more_than(x)	$x = x + D(x'')$; $D(x'') = abs(x'' - x)$; $x'' = x/\tan(\pi/4 - VERYWIDE')$
roughly(x)	$x = x \pm D(x'')$; $D(x'') = abs(x'' - x)$; $x'' = x/\cos(\pi/4)\cos(\pi/4 - CLOSE'')$
more_or_less(x)	$x = x \pm D(x'')$; $D(x'') = abs(x'' - x)$; $x'' = x/\cos(\pi/4)\cos(\pi/4 - CLOSE_1')$
about(x)	$x = x \pm D(x'')$; $D(x'') = abs(x'' - x)$; $x'' = x/\cos(\pi/4)\cos(\pi/4 - CLOSE')$
from_to(x,range)	x; $range_{low} < x < range_{hi}$
on_average(x)	x =about(x)
above(x)	$x = x + D(x'')$; $D(x'') = (R > 0)$; R = real number
below(x)	$x = x - D(x'')$; $D(x'') = (R < 0)$; R = real number;

Table 4.4: Computation of Specific Generic Values Using the Generic Band Widths

For the query object *much_greater_than*(x_1) and database object *less_than*(y_1) the condition in equation 4.19 must be satisfied.

$$\arctan(\mathbf{y} > \mathbf{x}). \tag{4.19}$$

This in turn implies that

$$\arctan(\mathbf{x}/\mathbf{y}) \leq \pi/4 - VERYWIDE' \tag{4.20}$$

where $\mathbf{x}$ = *much_greater_than*(x_1) (see Fig. 4.6), and $\mathbf{y}$ = *less_than*(y_1) is the generic value of the fuzzy database object which satisfies the query.

Figures 4.6 and 4.7 above, illustrate the important geometric relationships in the derivation of generic values for *much_greater_than*(x) and *much_less_than*(x) respectively.

Referring to figures 4.5 and tables 4.2, and 4.3, it is clear that, the generic query object $\mathbf{x}$ is in fact an interval about the user's "hazy value" of x_1. For example, by denoting the cutout points of the fuzzy restriction *more_or_less*(x_1) by $\mathbf{x}_{l1}$ and $\mathbf{x}_{u1}$, the generic value $\mathbf{y}$ of the fuzzy database object satisfying the query "more or less x_1" must lie between the cutout points.

Referring to figure 4.5 and table 4.2, possible generic values for the fuzzy query object *more_or_less*(x_1), are specified by equations 4.21 and 4.22

$$\mathbf{x}_{u1} \geq \mathbf{y} \geq \mathbf{x}_{l1} \tag{4.21}$$

implying,

$$\begin{cases} \arctan(\mathbf{x}_{u1}/\mathbf{y}) \geq \pi/4 - CLOSE_1' & \text{for upper bound} \\ \arctan(\mathbf{x}_{l1}/\mathbf{y}) \leq \pi/4 + CLOSE_1' & \text{for lower bound.} \end{cases} \tag{4.22}$$

where, $\mathbf{y}$, the generic value of the unknown fuzzy object satisfying the fuzzy query, is taken along the generic objects axis (Fig. 4.6). This is the condition which must now be used to select database objects satisfying the query "more or less x_1". The generic values $\mathbf{x}_{l1}$ and $\mathbf{x}_{u1}$ are computed from the equations given in table 4.4.

Thus, by means of the partitions induced in X, the original fuzzy query is transformed into an interval comparison problem in which the interval bounds correspond to the cutout points, peak points or any other desirable characteristic points of the membership function of the fuzzy object (Kandel, 1986, Klir and Folger, 1988). However, manipulation of fuzzy numeric data by interval arithmetic can only be tolerated for low precision requirements (Kandel, 1986; Klir and Folger, 1988). For higher precision fuzzy membership functions must be constructed from the fuzzy partitions to facilitate rigorous solutions. The general procedure illustrated in these examples is applicable for all predicates in the PRED set (table 4.2).

Other expressions not shown in table 4.4 include variations of *equal_to*(x), *above*(x) (same as *greater_than*(x)) and *below*(x) (same as *less_than*(x)) which fuzzify the strict interpretation (table 4.4) of these predicates. For example, the implemented function for *equal_to*(x), permits values falling within a very narrow band width around the crisp argument x. Thus "equal to x" is specified as

$$\begin{aligned} \mathbf{x} &= x \pm D(x"); D(x") = abs(x" - x); \\ & x" = x/\cos(\pi/4)\cos(VERYCLOSE/(32 + \log(x))). \end{aligned}$$

Similarly for *more_than*(x) and *less_than*(x) the acceptable generic values are respectively specified by

$$\begin{aligned} \mathbf{x} &\geq x + D(x"); D(x") = abs(x" - x); \\ & x" = x/\cos(\pi/4)\cos(VERYCLOSE/(32 + \log(x))) \end{aligned}$$

and

$$\begin{aligned} \mathbf{x} &\leq x - D(x"); D(x") = abs(x" - x); \\ & x" = x/\cos(\pi/4)\cos(VERYCLOSE/(32 + \log(x))). \end{aligned}$$

4.5 Truth Valuation by Use of Pseudo-Truth Tables.

It is common practice, in first order predicate logic, to compute the truth values of logical functions in terms of truth tables (Dowsing et al, 1986; Delahaye, 1986; Turner, 1985). Truth tables can also be applied in fuzzy logic (Mizumoto et al, 1979, Mizumoto and Tanaka, 1979).

To facilitate truth valuation for fuzzy query processing, the conditions which must be satisfied for various combinations of predicates have been developed in the form of pseudo-truth tables. The relations necessary for the computation of generic values of the fuzzy predicates involved were summarised in table 4.4. Some of the implemented cases are presented in summary form in table 4.5.

The fuzzy comparison operator is a C language implementation of the generic value computation formulae in table 4.4 and the truth table (table 4.5).

In table 4.5 the bold **x**, and **y** notation represent generic objects, x and y denote the crisp argument of the fuzzy predicate, y_{lo} and $\mathbf{y}_{lo}$ represent lower bounds, while y_{hi} and $\mathbf{y}_{hi}$ are the higher bounds of crisp and generic objects respectively. Figure 4.8 shows the design of the fuzzy comparison operator as implemented in SLEMS FUZZ subsystem.

4.5.1 Direct Derivation of Membership Functions From the Fuzzy Partitions.

This section presents a method for constructing fuzzy membership functions from the partitions induced by fuzzy predicates. The main objective for the development of the method is to provide a tool for analyzing the theoretical and practical validity of the proposed fuzzy partition method for knowledge representation and query processing in the SLEMS. Once membership functions are available they can also be used to

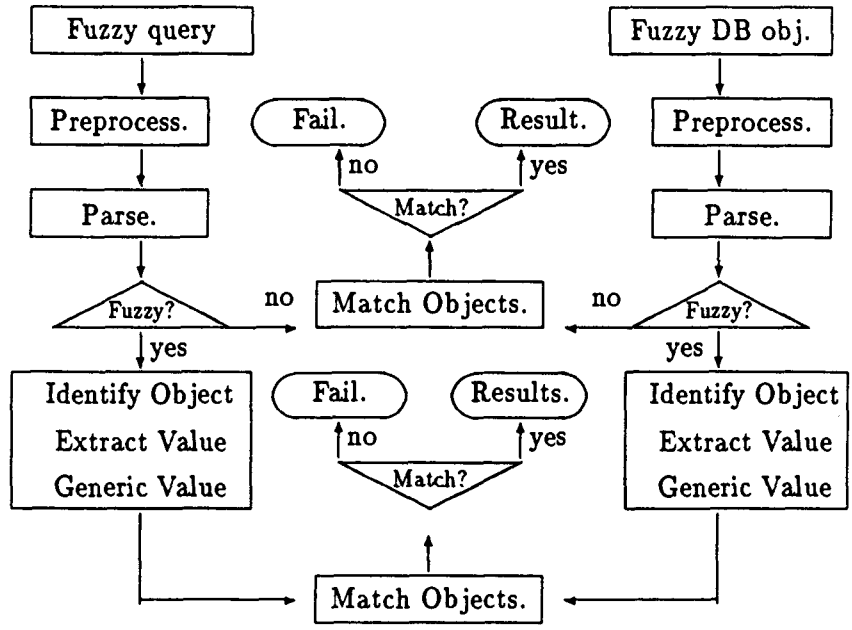


Figure 4.8: The SLEMS FUZZ Subsystem Design.

Database Object	Generic Object	T Condition	F Condition
equal_to(y)	$\mathbf{y} = y$	$\mathbf{y} \leq \mathbf{x}$	$\mathbf{y} > \mathbf{x}$
less_than(y)	$\mathbf{y} < y$	$\mathbf{y} \leq \mathbf{x}$	$\mathbf{y} > \mathbf{x}$
more_than(y)	$\mathbf{y} > y$	$\mathbf{y} < \mathbf{x}$	$\mathbf{y} > \mathbf{x}$
slightly_less_than(y)	$\mathbf{y} < y$	$\mathbf{y} \leq \mathbf{x}$	$\mathbf{y} > \mathbf{x}$
slightly_more_than(y)	$\mathbf{y} > y$	$\mathbf{y} \leq \mathbf{x}$	$\mathbf{y} > \mathbf{x}$
less_or_equal(y)	$\mathbf{y} \leq y$	$\mathbf{y} \leq \mathbf{x}$	$\mathbf{y} > \mathbf{x}$
more_or_equal(y)	$\mathbf{y} \geq y$	$\mathbf{y} \leq \mathbf{x}$	$\mathbf{y} > \mathbf{x}$
more_or_less(y)	$\mathbf{y}_{lo} \leq \mathbf{y} \leq \mathbf{y}_{hi}$	$\mathbf{y}_{hi} \leq \mathbf{x}$	$\mathbf{y}_{hi} > \mathbf{x}$
about(y)	$\mathbf{y}_{lo} \leq \mathbf{y} \leq \mathbf{y}_{hi}$	$\mathbf{y}_{hi} \leq \mathbf{x}$	$\mathbf{y}_{hi} > \mathbf{x}$
on_average(y)	$\mathbf{y}_{lo} \leq \mathbf{y} \leq \mathbf{y}_{hi}$	$\mathbf{y}_{hi} \leq \mathbf{x}$	$\mathbf{y}_{hi} > \mathbf{x}$
roughly(y)	$\mathbf{y}_{lo} \leq \mathbf{y} \leq \mathbf{y}_{hi}$	$\mathbf{y}_{hi} \leq \mathbf{x}$	$\mathbf{y}_{hi} > \mathbf{x}$
from_to(y_{lo}, y_{hi})	$y_{lo} \leq \mathbf{y} \leq y_{hi}$	$\mathbf{y}_{hi} \leq \mathbf{x}$	$\mathbf{y}_{hi} > \mathbf{x}$

Table 4.5: Pseudo_truth Table for the Query Object *much_less_than*(x) for Various Database Objects; $\mathbf{x}$ is the Generic Query Object.

manipulate fuzzy objects using conventional fuzzy set theory principles.

The desirable characteristics of fuzzy membership functions as presented in Kandel (1986), Klir and Folger (1988), and Zadeh et al (1975) are:

- The membership function must map the set of objects in the universe of discourse into the interval $[0, 1]$.
- It must also satisfy the fuzzy set theoretic properties of fuzzy membership functions with respect to fuzzy union, intersection, complementation etc.

Let $R(x)$ represent a general fuzzy predicate where $x \in X$ is some crisp numeric value in the database. Let, also, a generic value induced by the fuzzy predicate R from x be $\mathbf{x} \in X$, such that $\mathbf{x}$, lies on the boundary of the partition induced by the fuzzy predicate R. Further more let D denote the amount by which the fuzzy predicate R "stretches" the crisp value x. Then D is the *width* of the fuzzy partition generated

by R and it is given by the equation

$$D = |\mathbf{x} - x|$$

for symmetric partitions of the search space X.

Let $R(y)$ represent some other database object such that $y \in X$ is a crisp numerical value which may or may not be equal to x. Let the width of the partition induced by R on y be denoted by d. Then, by the definition of set membership, the generic object $\mathbf{y} \in X$, induced by the fuzzy restriction $R(y)$ is in the partition induced by the fuzzy restriction $R(x)$ if the condition $d = |\mathbf{y} - y| < D$ is satisfied. This is a sufficient condition for membership in $R(x)$, because it guarantees that all points or members of $R(y)$ fall within the "stretch" of $R(x)$.

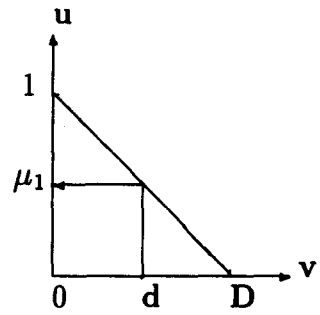
A possible function for mapping the set induced by $R(x)$ onto the interval $[0, 1]$ can now be constructed on the basis of figure 4.9. Let v represent an axis in the direction of the stretch induced by $R(x)$ and u be the number line oriented perpendicular to v.

1. Using an appropriate scale set the width D of the partition induced by $R(x)$ on the v-axis and denote it with $\mathbf{D}$.
2. Set out on the u axis a unit line segment such that the origin is at the intersection of u and v
3. Link the end of the unit line with the point located at distance D along v from the common origin.
4. Plot the width d induced by the predicate R on y along v in the same manner and denote it by $\mathbf{d}$.
5. Mirror project the end point of segment d located on axis v onto the unit line perpendicularly and denote the projection point by μ_1.
6. The distance of the projection point μ_1 from the origin of the unit line is proportional to the strength with which d and by analogy y is contained

in the partition induced by R on X. It may therefore be taken as a first approximation to the membership function.

7. Modify μ_1 by applying dilation, intensification, concentration, or other fuzzy set theoretic operations (Schmucker, 1984; Kandel, 1986; Klir and Folger, 1988) to arrive at an intuitionally appropriate shape of the membership function.

Figure 4.9 is a schematic depiction of the fuzzy membership function construction process.



u = membership valuation space
v = measurement space (crisp values)
d = input value (measurement)
D = partition bandwidth in measurement space
μ_1 = approximate membership value

Figure 4.9: Construction of Approximate Membership Values From Fuzzy Partitions.

An approximate formula for the computation of the fuzzy membership values based on this procedure is the similarity relation

$$(1-\mu_1)/d = 1/D$$

and μ_1 is thus given by

$$\mu_1 = 1 - d/D \tag{4.23}$$

To enforce the fuzzy membership requirement that μ_1 must be a value in the range $[0,1]$, the additional requirement that the fraction d/D be always less or equal to 1 is introduced. This requirement is not arbitrary but is based on the fact that, if the partition induced by the fuzzy restriction $R(y)$ has a width greater than that induced by $R(x)$ then y is not contained it. In such case it is required that the membership function get a value of 0.

Using this additional criteria, the fuzzy membership for an arbitrary fuzzy restriction may be computed by equation 4.24

$$\mu_1 = \begin{cases} 0, & \text{if } d/D > 1. \\ 1 - d/D, & \text{otherwise.} \end{cases} \tag{4.24}$$

For practical use appropriate functions, for the generation of the fuzzy partitions, are substituted for d and D in these equations. Further more the ratio d/D is replaced by an appropriately chosen $F(d/D)$ where F is a *dilation, concentration, intensification,* etc. (Schmucker, 1984; Kandel, 1986) fuzzy set theoretic operator.

To illustrate the procedure the fuzzy membership function for *much_less_than*(y) is derived below. Referring to table 4.4 and figure 4.7 the stretch induced by the fuzzy predicate *much_less_than* on y is computed as

$$D = abs(y'' - y)$$

where

$$y'' = y/\tan(\pi/4 + VERYWIDE').$$

From table 4.3

$$VERYWIDE' = VERYWIDE/2 + VERYCLOSE/(4 + \log(y))$$

can be derived and after substitution of $WIDE$ and $VERYCLOSE$ from table 4.2 this gives

$$VERYWIDE' = \pi/6 + (\pi/48)/(4 + \log(y)).$$

Thus

$$y" = y/\tan(\pi/4 + \pi/6 + (\pi/48)/(4 + \log(y)))$$

and after simplification the equation

$$y" = y/\tan(\frac{\pi}{48}(20 + 1/(4 + \log(y)))$$

is obtained.

Now let y_1 be an arbitrary database object. It is required to find the degree to which y_1 is compatible with the predicate *much_less_than*(y). Using equation 4.23 the approximate membership function is

$$\mu_1 = 1 - \frac{abs(y_1 - y)}{abs(y" - y)}.$$

Assuming a square relationship for the compatibility of y_1 and *much_less_than*(y) the membership function can be expressed as

$$\mu = 1 - \left(\frac{abs(y_1 - y)}{abs(y" - y)}\right)^2.$$

Taking y as a unit value and substituting into the above equation produces

$$\mu = 1 - \left(\frac{abs(y_1 - 1)}{abs(y" - 1)}\right)^2.$$

By a similar process the membership functions for the rest of the predicates in set PRED can be obtained. The graphical plots of the membership functions for some of the predicates in PRED, are shown in figures 4.11 to 4.14 of section 4.6 below.

4.6 Testing and Analysis of the Theory on a Test Database.

The functions derived in sections 4.4.2 and 4.5 were used to partition an artificial data set, generated by the LOTUS 123 spread sheet package, consisting of crisp values ranging from 1 to 300.

Considering a query object, such as, *more_or_less*$(y, 50)$ the set **y**, of all database values in the range [1,300] satisfying the query criteria were generated. The partitions induced in the test database by the fuzzy predicates *more_or_less*$(y, 50)$, *much_greater_than*$(y, 50)$, *much_less_than*$(y, 50)$, *slightly_less_than*$(y, 50)$ and *slightly_more_than*$(y, 50)$ where, y, represents an arbitrary database value which satisfies the fuzzy query, were plotted against the database crisp values. The results of this experiment are graphically presented in figure 4.10.

The XY plane in figure 4.10 is the plane containing all the tuples (y_i, y_j) for $i = 1, \ldots, n$, $j = 1, \ldots, n$ where n is the total number of objects in the database. It can be seen from these results, that the partitions induced on the simulated data base closely resemble those in figures 4.5, 4.3, and 4.4 as required from the theoretical considerations.

Having demonstrated that the proposed procedure produces reasonable partitions, fuzzy membership functions based on the partitions, were then constructed. The results of this test are summarised in the next section.

4.6.1 Evaluation of Directly Computed Membership Functions.

The fuzzy membership functions generated from the test database and query objects can be put into two categories

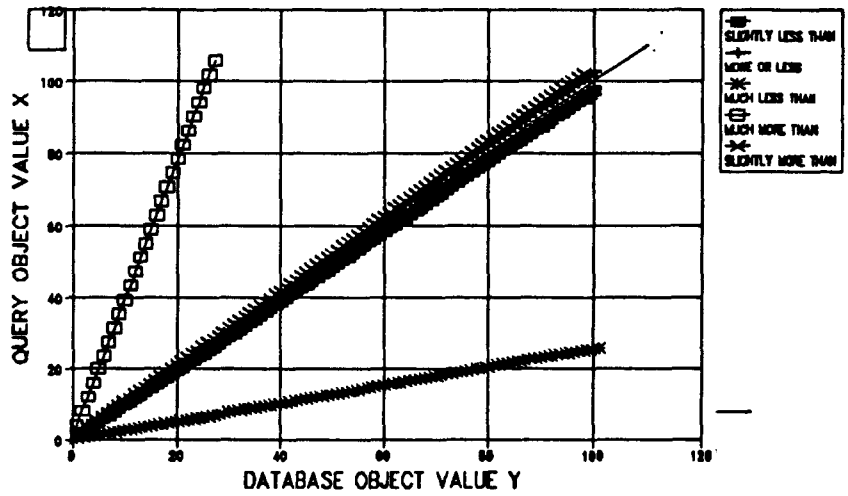


Figure 4.10: Boundaries of Partitions Induced on the Test Database by Some Fuzzy Predicates.

- Triangular membership functions corresponding to the predicates *more_or_less*(x), *slightly_more_than*(x), *slightly_less_than*(x) and *about*(x).

- Higher degree functions, corresponding to the S-function and the Z-function (Kandel, 1986), for the predicates *much_greater_than*(x) and *much_less_than*(x) respectivelly.

The membership function for *slightly_more_than*(x) shown in figure 4.11 reflects what was expected i.e a triangular function with values distributed close to the crisp argument of the fuzzy predicate. The same case applies to the predicate *more_or_less*(x) shown in figure 4.12. However, in existing literature (Zadeh et al, 1975; Klir and Folger, 1988; Kandel, 1986, Schmucker,1984) the membership function for the linguistic predicate "more or less A" is interpreted as $\sqrt{\mu_A(x)}$ where $\mu_A(x)$ is the membership of x in the fuzzy set A. This is not a problem, however, because in Baldwin (1979), Mizumoto et al (1979), Schmucker (1984) and Kandel (1986) the above fuzzy expressions are used as modifiers to fuzzy sets while in the present study they are used to fuzzify crisp objects in X. The result for *more_or_less*(x) in figure 4.12 corresponds well with the interpretation of the predicate, that is, it induces triangular membership function with a narrow symmetric stretch about the crisp argument x of the fuzzy predicate.

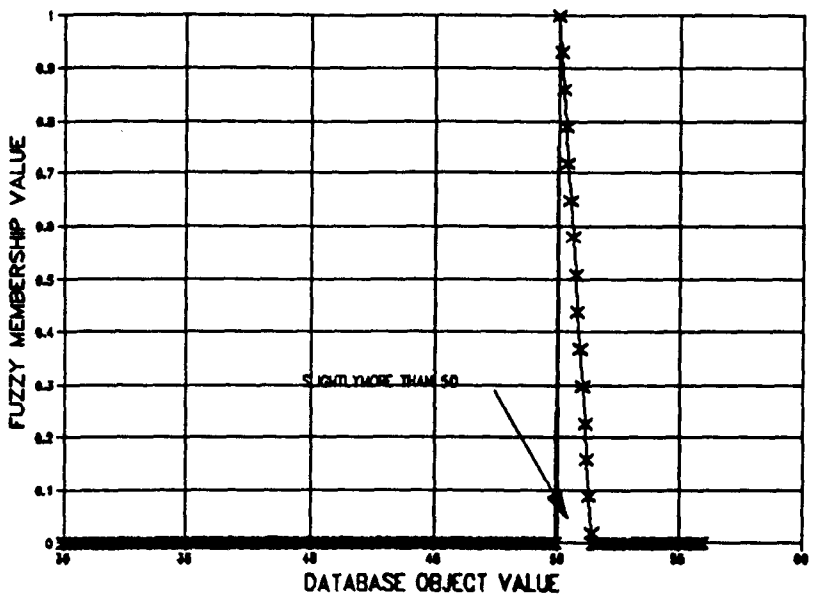


Figure 4.11: Membership Function for *slightly_more_than*$(y, 50)$

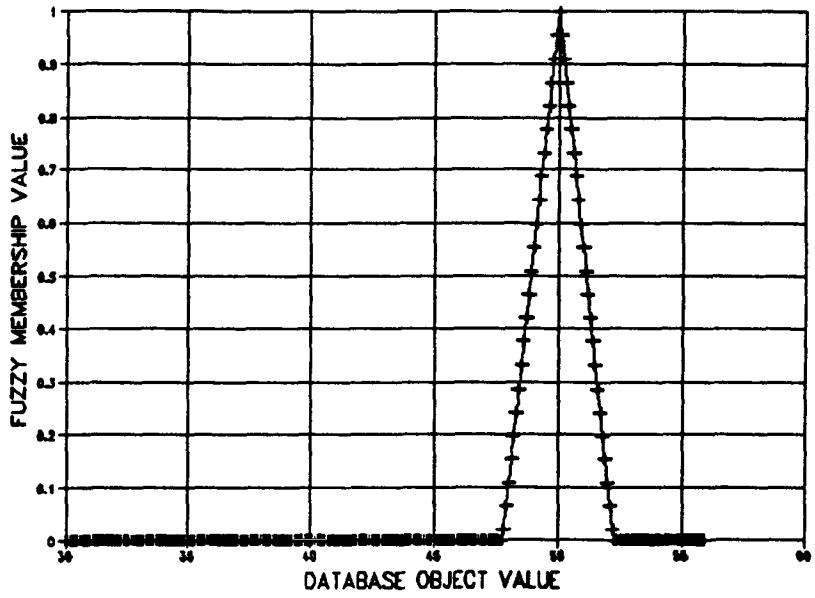


Figure 4.12: Membership Function for *more_or_less*$(y, 50)$

4.6.2 Further Examination of the Constructed Membership Functions.

As a further test of the validity of the fuzzy membership functions directly computed from the fuzzy partitions, the characteristic functions of a fuzzy restrictions such as "more or less 50", were used to model the same fuzzy relation applied to different crisp arguments such as "more or less 100". The test results also indicate that fuzzy restrictions such as "more or less 100" can be adequately modelled by the membership characteristic function derived using a different crisp argument. It is therefore safe to say that the fuzzy geometrical partitions-based membership functions exhibit an independence from the crisp argument of the fuzzy predicate, and that their shape depends only on the fuzzy relation involved. By the same argument it can be said, that, the resulting membership function only reflects and therefore models, the shape of indetermination in the associated fuzzy relation. Figure 4.15 illustrates the possible use of such membership functions to facilitate database selection of fuzzy objects.

The results presented in figures 4.10 to, 4.14 therefore strengthen the initial argument that membership functions for the fuzzy relations discussed in this study can, in general, be directly constructed from their induced partitions in the universe of discourse. This conclusion is also supported by the fact that there appears to be a similarity between the shape of these membership functions and those obtained by conventional fuzzy sets methods. For example, the shape of the membership function for "about 2" in Mizumoto et al (1979), and that for "about 50" in figure 4.16 are obviously similar.

Overall these tentative results are encouraging as they seem to concur with my expectations and interpretation of the commonsense interpretation of the fuzzy predicates examined. At the same time results on retrieval of fuzzy objects based on the direct use of fuzzy partitions, presented earlier in chapter 3 of the thesis, have been found to be reasonable.

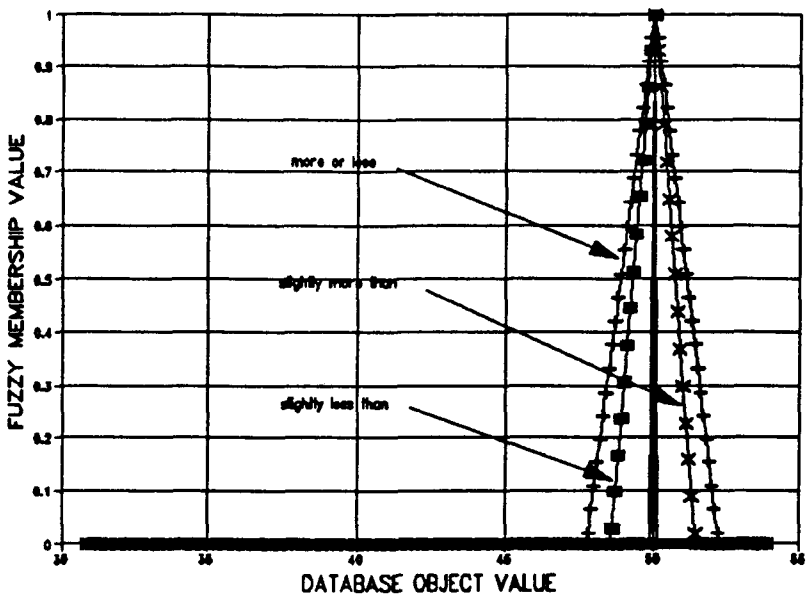


Figure 4.13: Membership Functions for Some Fuzzy Restrictions.

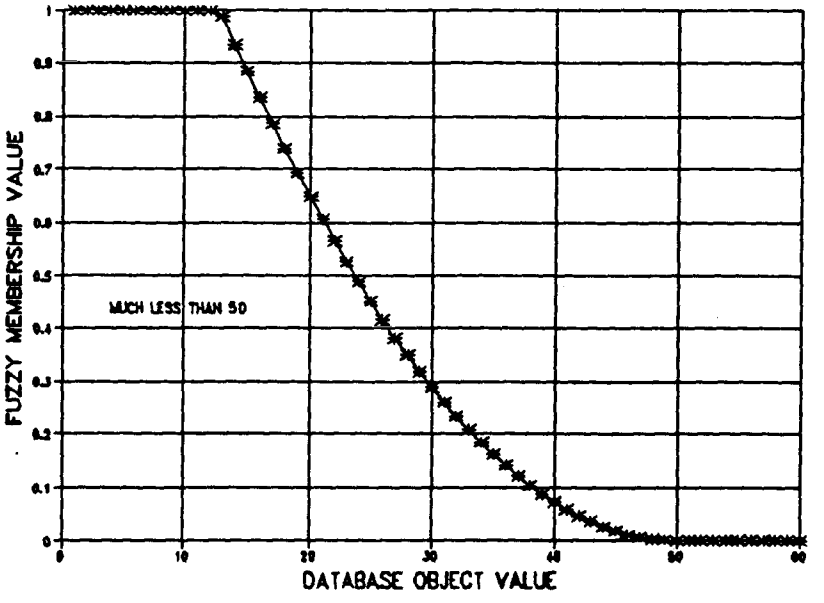


Figure 4.14: Membership Function for the Fuzzy Restriction *much_less_than*(x)

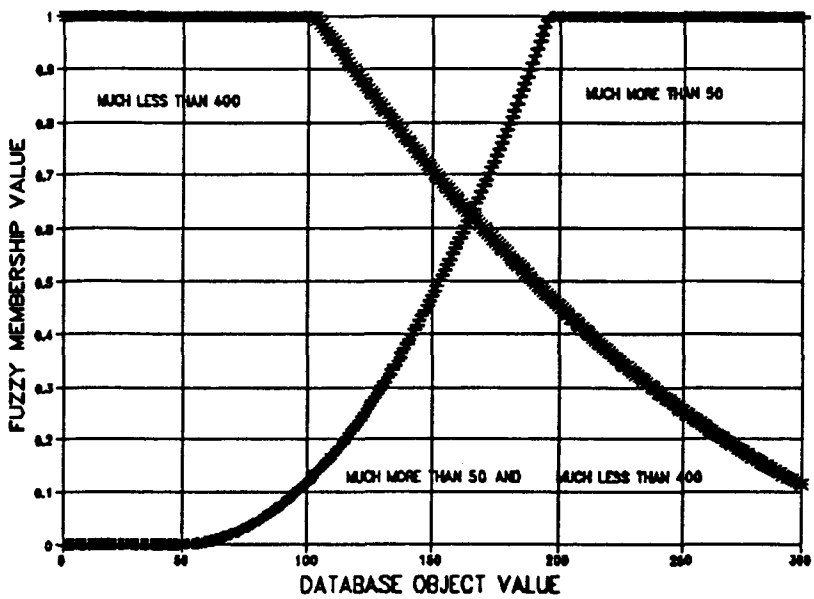


Figure 4.15: Possible Use of the Membership Functions to Facilitate Retrieval of Fuzzy Objects.

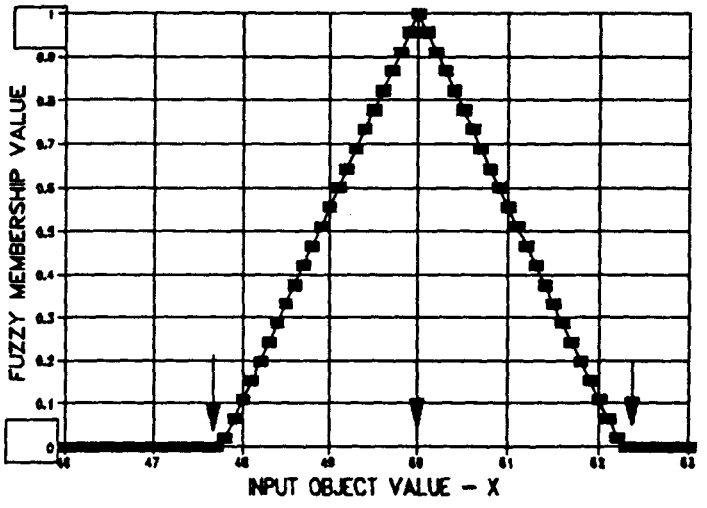


Figure 4.16: Membership Function for about(50)

Chapter 5

An Expert System for Soil Erosion Estimation and Modelling Using Remote Sensing and GIS Information.

5.1 Introduction.

Soil erosion estimation and modelling is a complex problem requiring information and expertise spanning over a wide field of knowledge. Information sources and information processing methods, for soil erosion monitoring, modelling and estimation were discussed in chapter one. They can generally be placed in five categories as:

- Conventional field and laboratory methods, conventional mapping information such as topographic and thematic maps (Smith and Wischmeier, 1957; Wischmeier and Smith, 1957; Bali and Karale, 1977; Bobrovitskaya et al, 1977; Lal et al, 1977; Hóly, 1980; Lal, 1981; Goldman et al, 1986; Morgan, 1986; Foster and Wischmeier, 1974; Foster et al, 1981; Foster and McCool, 1987; McCool et

al, 1987; McIsaac et al, 1987; Mills, 1986).

- Aerial photographs using conventional photo-interpretation methods (Stephens et al, 1985, Bondelid et al, 1980; Sneddon and Lutze, 1989).

- Satellite imagery such as the NOAA AVHRR, Landsat MSS, Landsat TM and SPOT images, using spectral image analysis and classification techniques (Morris-Jones and Kiefer, 1978; Schmugge et al, 1979; Scoffield and Oliver, 1979; Bondelid et al, 1980).

- Geographic information systems using data and information from all the previous sources (Morris-Jones and Kiefer, 1978; Logan et al, 1982; Spanner, 1983; Best and Westin, 1984; ESRI, 1984; Walsh, 1985; Hart et al, 1985; Ventura et al, 1988; Lynn, 1989).

- Knowledge Based Systems or Expert Systems (Morrison et al, 1989).

At present very few studies have been conducted in the application or use of knowledge based methods in the solution of soil loss problems. However, many knowledge based solutions to the problem of data acquisition from aerial photography, remotely sensed imagery and spatial databases exist today (Goodenough, 1987, 1988; Ripple et al, 1987; Robinson et al, 1987; Argialas et al, 1988; Swann et al, 1988; Corr et al, 1989; Schowengerdt and Wang, 1989; Argialas and Harlow, 1990; Egenhofer and Frank, 1990; Hadipriono et al, 1990; Kretsch and Mikhail, 1990; Mehldau, 1990; M-Jensen et al, 1990; Schenk and Zilberstein, 1990; Srinivasan and Richards, 1990).

PLANTING, the expert system developed by Morrison et al (1989) for the USDA-SCS, is a clear indication on the future direction of soil loss estimation and modelling in the USA. As mentioned in chapter two, the system uses agricultural and soil erosion domain knowledge to assist farmers in the USA to select suitable farming equipment for conservation planting.

The solution proposed in this research differs from the one used in PLANTING by the fact that it is not based on an existing expert shell. As it will become evident later the EXPERT subsystem is an expert shell implemented from scratch in the C language. Currently it runs as an independent subsystem of the SLEMS on a Sun 4 UNIX workstation, but there are plans for porting it onto an IBM compatible PC running MS-DOS.

It is hoped that the simple general approach to the soil loss estimation and modelling problem proposed and implemented in this research will contribute to more understanding of the potential use of the the expert systems technology in solving soil erosion problems. The main emphasis in the presentation is therefore the management and utilization of knowledge about data requirements, procedures for acquiring conventional mapping, remote sensing and GIS data, and application of the information in soil loss estimation and modelling problems.

5.2 The SLEMS EXPERT Subsystem in General.

The EXPERT is a subsystem of the soil loss estimation and modelling system (SLEMS). The EXPERT subsystem is a simple domain independent rule based intelligent program which can be used to acquire and query domain knowledge in the form of human readable rules.

The knowledge structure used to store the SLEMS rule base is a hierarchical network of the < Object > < Attribute-list > tuples (OA-tuples) introduced in chapter two. In this format the rule conclusion becomes the object part while the rule premises constitute the attribute-list of the object-attribute(OA) tuple.

Essentially the EXPERT subsystem is a simple expert shell built around a backward chaining search control strategy (Schildt, 1987). The inference mechanism therefore recursivelly attempts to satisfy a hypothesis by matching the goal or its sub-goals against the database objects (Barr and Feigenbaum, 1981, 1982; Cohen, 1985; Schildt,

1987; Barr et al, 1989; Luger and Stubblefield, 1989). As pointed out in Cohen (1985) and others, backward chaining is a good strategy for rule based systems and it facilitates an easy way to explain how the inference mechanism achieved a given result, or why it requires a certain piece of information to resolve a dead-end. Essentially the explanation mechanism keeps a history of the sub-goals and returned condition (acceptance or rejection) of hypotheses and reads back the history on request.

Because expert systems are meant to manipulate common sense knowledge which is often vague, the design of expert systems must incorporate the means for representing and manipulating uncertainty (Schmucker, 1984; Cohen, 1985; Kandel, 1986; Klir and Folger, 1988). Management or control of uncertainty in the EXPERT subsystem is provided by a simple method of propagating and pooling certainty factors similar to the solution used in MYCIN (Barr and Feigenbaum, 1981, 1982; Cohen, 1985; Luger and Stubblefield, 1989). Although this is not the best way to handle uncertainty in expert systems it has been used in the EXPERT because of its simplicity and to avoid overly complex system design. Detailed discussion of the EXPERTS four main modules and uncertainty management follows in the next two sections.

5.2.1 Components of the SLEMS EXPERT.

In this section specific algorithms used to implement the SLEMS EXPERT are discussed in more detail. The four main modules constituting the EXPERT subsystem are the rule entry module (**enter**), the query module (**query_rules**), the explanation module (**reasons**), and the rule base loading module (**load_rules**) built around source code from Schildt (1987).

The query module is the inference mechanism of the EXPERT subsystem. The inference method used to execute queries in the rule base is a simple backward chaining or goal driven strategy (Schildt, 1987). The essence of this strategy is that the inference module first guesses a solution (the goal) and then attempts to show that this solution satisfies all the conditions of the query. If any of the conditions fail the

module discards the current solution and makes another guess.

To prevent cycling back to the old goals the module keeps track of all invalid objects. The objects examined by the query module are rules, so when an object satisfying all the conditions in the query is found the rule conclusion is presented as the solution to the query.

Detailed discussion of the principles and applications of goal driven inferencing and other inference strategies can be found in numerous expert systems and artificial intelligence sources such as Luger (1989), Walker (1987), Barr and Feigenbaum (1981), Barr et al (1989), Davis and Lenat (1982), Rolston(1988), Cohen (1985) and others. The backward chaining algorithm implemented for the EXPERT subsystems inference module is based on Schildt (1987), and is presented below in pseudo-code form.

```
Begin
For all rules in the rule base
    Try a rule and see if it matches the query
    For each rule premise retrieve the premise
        If the rule premise is on the invalid list
            Then the rule fails.
        End if
        If the rule premise is not on the valid list
            Then the rule fails.
        End if
        If the rule premise has not been examined and
        The premise cannot be verified from existing facts.
            Then ask user for verification
        End if
        If the user asks why
            Then explain reasons for requesting verification.
        End if
```

```
        If the premise is not verified
            Then remember this case as invalid.
        End if
        If the rule premise is verified as valid
            Then update rule current certainty.
            Remember this case as valid
        End if
    End for each premise.
    If all premises are valid and
        Current rule has sufficient certainty
        Then accept the rule conclusion as the query solution.
        Search for any other solutions if required by user
    Else the rule conclusion is not a solution to the query.
    End if
End for all rules.
If no valid rule was found
    Then the query has no solution.
End
```

The rule entry module performs rule editing functions for the EXPERT subsystem. It facilitates controlled rule base editing by means of prompts according to the syntax:

SLEMS rule:

```
<rule domain>
<rule explanation>
<rule conclusion>
[<rule premise><premise certainty>]
<rule certainty>.
```

where the square bracket indicates that a rule may have several premises. At the end

of rule entry session the module prompts for a file name under which the rules should be stored. According to the semantics of the SLEMS, rule base file names correspond to specific subject areas of the knowledge domain.

The required domain knowledge is loaded into the EXPERT's active memory by the rule base loader at the beginning of the query session according to the following algorithm:

```
Begin
Prompt for the domain of interest
Prompt for subject area (rule file name)
For each rule in the rule base
    If rule domain is the same as domain of interest
        Then load the rule into memory
        Else ignore
    End if
End for.
End
```

A fundamental characteristic of an intelligent program is the ability to explain its reasoning (Minton, 1988; Rolston, 1988; Luger and Stubblefield, 1989). In the SLEMS EXPERT this function is performed by the explanation module using the following algorithm:

```
Begin
If the user asks why
  Backtrack to the rule recently processed
  Display the rule
  For all valid and invalid premises.
      Display all verified premises for the rule
      Display all the premises so far not verified for the rule
```

End for

Display all the rules so far rejected and

The reasons for their rejection.

End if.

End

These main modules rely on several other routines which perform the actual rule base searching and other tasks required by the main modules. Appendix II contains the relevant source code for the four main modules.

5.2.2 Handling of Uncertainty in the EXPERT Subsystem

To facilitate the management of uncertainty in the rule conclusion and its premises the EXPERT subsystem employs a procedure similar to MYCIN's propagation of certainty factors (Barr and Feigenbaum, 1981; Cohen, 1985; Barr et al, 1989). This requires that, parallel to the evaluation of the validity of the rule premises, an uncertainty combination function called in SLEMS "comp_prob", calculate current certainty values for the rule conclusion. The computed value is used together with the status (FALSE or TRUE) of the premises to control the inference module's search path. If the computed certainty factor is below some pre-established threshold the rule is rejected irrespective of the status of its premises.

The EXPERT subsystem has an option allowing the user to select one of three approaches of treating uncertainty in the premises. These options are classical probability, the weakest link method and the strongest link methods discussed below. In the implemented version a threshold of acceptance for a conclusion is fixed at the commencement of the consultation session or fixed at 0.5 by default. This method of handling uncertainty is referred to in Cohen (1985) as parallel certainty inference method. The weakest link method is a minimum law of propagation of uncertainty

(Dubois and Prade, 1980; Cohen, 1985; Schildt, 1987) which assumes that in a conjunctive body of evidence the piece of evidence with the least degree of belief determines the degree belief of the total body of evidence.

The strongest link method (Schildt, 1987) assumes that in a disjunctive body of evidence, the piece of evidence with the highest certainty value is the determining evidence. It therefore takes the maximum of the certainty factor values of the premises as the certainty factor of the conclusion of the rule.

The concepts of weakest link, and strongest link methods of propagation have a close association with the min, and max fuzzy set operators used to compute the intersection and union of fuzzy sets as explained in Kandel (1986), and Klir and Folger (1988). To demonstrate how certainty factors (CF) are employed in the SLEMS EXPERT consider for example the rule (Eq. 5.1):

$$\begin{array}{ll} \text{IF} & A\,[CF_A] \text{ AND } B\,[CF_B] \text{ AND } C\,[CF_C] \text{ AND } D\,[CF_D] \\ \text{THEN} & E\,[CF_E] \end{array} \tag{5.1}$$

where CF_A, CF_B, CF_C and CF_D are the degrees of belief (certainty factors) in A, B, C, and D respectively and CF_E is the degree of belief in the conclusion E. Then by the method of parallel certainty inference (Cohen, 1985) the degree of belief in the conclusion must be modified by the uncertainty expressed in the premises. Propagation of the uncertainty of the premises to the conclusion is achieved by taking the minimum certainty factor of the conjunctive premises (Cohen, 1985). This approach is the weakest link method (Schildt, 1987; Klir and Folger, 1988). The pooling of the propagated and given certainty factors of the rule is done by a multiplication combination rule (Cohen, 1985). Thus denoting the propagated uncertainty with 'CF_E' and the updated uncertainty factor by "CF_E" the procedure applied in the EXPERT is given by

$$'CF_E' = \min\{CF_A, CF_B, CF_C, CF_D\} \tag{5.2}$$

$$"CF_E" = CF_E \times 'CF_E'. \tag{5.3}$$

Example 1 is a simile of the process of propagating and pooling uncertainty in the conclusion of a rule for selecting combined USLE crop and conservation management factor (CP factor). The premises and conclusion of the rule are based on an actual look-up table method for the selection of USLE parameters in the Dane County Soil Erosion Control Plan by Ventura et al (1988).

The CF values for the premises have been arbitrarily assigned to model a hypothetical degree of belief in the contribution of each premise to the conclusion. In practice a criteria which could be used in determining appropriate values might be for example, the certainty or ease with which the information contained in the premises could be derived by remote sensing information processing procedures. The CF of the rule conclusion could be thought of as a measure of the strength of the rule's overall validity and would normally be subjectively assigned by a soil erosion conservation specialist based on experience.

Example 1.

Assume that the variables in the rule specified by Eq. 5.1 are substituted for as follows:

$$\begin{aligned}
A &= \text{slope class } 1\% - 2\%, \text{with } CF_A = 0.9 \\
B &= \text{crop rotation is C-S-O, with } CF_B = 0.7 \\
C &= \text{crop stage is no till, with } CF_C = 0.8 \\
D &= \text{conservation is contour strip cropping, with } CF_D = 0.7 \\
E &= \text{CP is } 0.027, \text{with } CF_E = 0.8
\end{aligned}$$

where C = corn, S = soy beans, O = oats or other small grain. Then from Eq. 5.2, the propagated certainty factor is given by

$$\begin{aligned}
'CF_E' &= \min\{CF_A, CF_B, CF_C, CF_D\} \qquad (5.4) \\
&= \min\{0.9, 0.7, 0.8, 0.7\} \\
&= 0.7.
\end{aligned}$$

and the modified degree of belief in the conclusion (Eq. 5.3) is therefore

$$\begin{aligned} "CF_E" &= CF_E \times 'CF_E' \\ &= 0.8 \times 0.7 \\ &= 0.56 \end{aligned} \tag{5.5}$$

This means that a CP factor of 0.027 can be applied with a certainty factor of 0.56 when all the premises $A - D$ have been verified by the inference module. It is important to notice that the certainty factors are not probabilities since clearly the sum of the CFs is not necessarily 1 as required in probability theory (Zadeh et al, 1975; Cohen, 1985; Klir and Folger, 1988; Zadeh, 1989).

This scheme of uncertainty handling as employed in the EXPERT is a simplified form of the approach used in MYCIN (Barr and Feigenbaum, 1981; Barr et al, 1989; Cohen, 1985). It has no mechanism for conflict resolution, and in case of multiple solutions it presents all the solutions to the user.

5.3 General Structure of the SLEMS Rule Base.

In discussing the architecture of the SLEMS rule base the following conventions and definitions will be used:

A **domain** is a category of knowledge spread over several rule base files but semantically pooled together through high level links.

A **sub-domain** is a sub-category of knowledge required to solve a group of related problems. A domain may contain several sub-domains.

A **subject area** is a body of knowledge required to solve a specific problem of a sub-domain or domain. A domain or sub-domain may consist of several subject areas.

Procedural knowledge is knowledge on task performance or problem solving routines.

A **link file** is a special body of knowledge representing relations among the real world domain or sub-domain knowledge.

The concept of domain, subdomain, and subject area as defined above constitutes a "chunking" of related bodies of knowledge according to the principle of knowledge representation of clumping together associated knowledge pieces, espoused in chapter 2 section 3.2. It also exploits the modularity inherent in domain knowledge as elaborated in Stricklen et al (1987).

5.3.1 Structure of the Active Memory.

The active memory stacks into which the rules are loaded from the rule base are organised in the form of C structures at two levels. The first level consists of C structures stacked in array form and they hold the rule domain, rule explanation, rule conclusion part and a pointer to the memory location of the rule premises. The second level structures are linked-list C-structures which hold the rule premises and the certainty values associated with each premise. Figure 5.1 represents the C language implementation of the two level knowledge structures.

5.3.2 The Rule Base Structure.

The SLEMS rule base consists of a hierarchical file structure, in which the highest level link files contain relations or dependencies among different knowledge domains, high level files contain domain and sub-domain specific knowledge and the low level files consists of subject area knowledge. Where a subject contains many details it may be up graded to a sub-domain.

At the highest level of the hierarchy is a special link file called the SLEMS system file which contains knowledge on the system's main components and organisation.

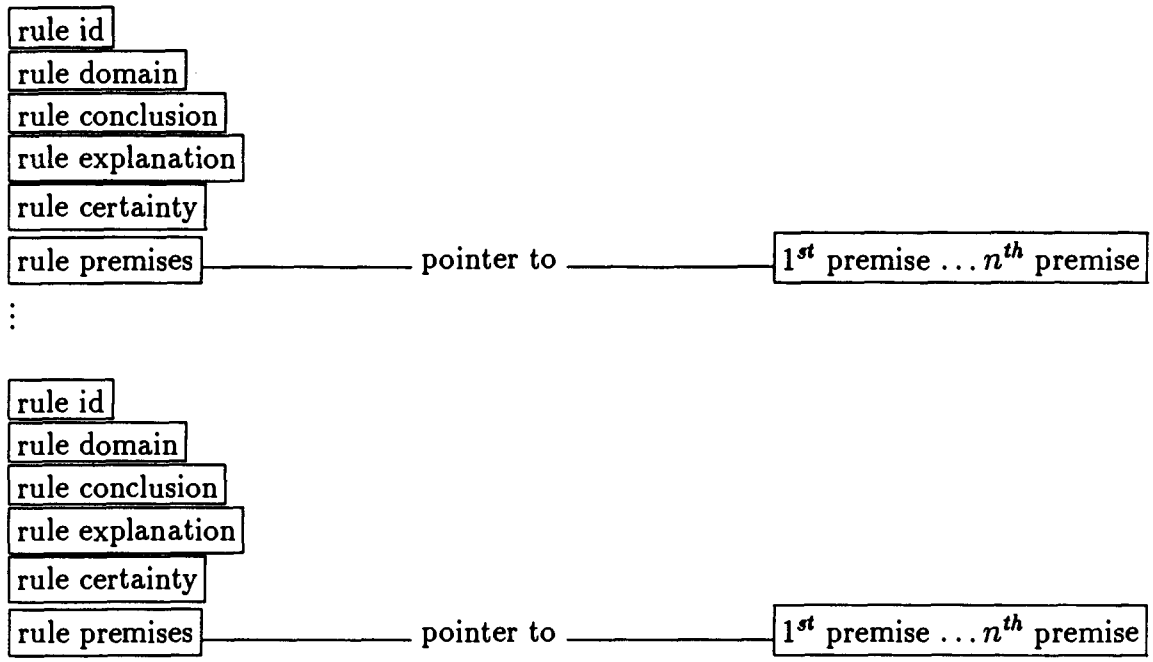


Figure 5.1: Structure of the Active Memory

The purpose of this file is to provide the user with ready information about the SLEMS rule base and give guidance on how to select appropriate domains for the interactive discourse with the EXPERT sub-system.

Each relevant real world knowledge category is represented by a domain file such as EROSION, SOIL, SPECTRAL_SENSING, etc. (figure 5.2). Also at this level there are files which contain knowledge on important domain procedures or task performance such as SELECT_MODEL (figure 5.2). This file contains knowledge on different soil loss estimation models, USLE parameter selection rules, and external knowledge input requirements from other domains or sub-domains.

This kind of knowledge organisation resembles structures employed in knowledge-based simulation applications such as the ABLE's current rule-base in Round (Barr et al, 1989) and Stricklen et al (1987). At the bottom of the hierarchy are subject specific rule files such as USDA_CN containing knowledge on the determination of runoff curve numbers (CN) by the USDA-SCS AMC-II method (Bondelid et al, 1980;

Stephens et al, 1985), and C_FACTOR containing knowledge on the selection of crop management factors (C factors) for the universal soil loss equation (USLE) (Stephens et al, 1985).

Some files may be entirely composed of meta rules which capture the dependencies between knowledge at a higher and lower hierarchies. Similarly some rules may be used to provide a link across real world knowledge domains. For example, the COVER file contains not only knowledge on ground cover types (AGRI_CLASS), classification methods (DI_INTERPRETATION), cover parameters (DENSITY), etc., but also rules which link this domain file with SPECTRAL_SENSING to facilitate answers on ground cover related spectral sensing queries.

Artificial or contrived links may be introduced to constrain the possible search paths. For example, the rule base file SELECT_MODEL is, essentially, a link file whose rules specify how to go about selecting various models and parameters using knowledge stored in other domain or subject specific files. A user who is experienced may decide to directly load subject specific rules, such as, rules for selecting runoff curve number (CN) values by the USDA-SCS AMC-II method (Bondelid et al, 1980) residing in the USDA_CN rule base file (figure 5.2). The USDA-SCS AMC-II curve number estimation method and others, such as, the USGS LUDA CN and the LANDSAT CN methods are explained later in section 5.6.2.

The USDA_CN file contains rules for assigning default CN values based on the USGS Level I and Level II landuse classification scheme (Bondelid et al, 1980, 1981). Knowledge on the USGS landuse classification scheme is also included in the USGS_CLASS and LANDUSE_CATEGORY (not shown) rule base files under the LAND_USE domain file. Similarly the AGRI_CLASS file contains knowledge on the Ontario Agricultural Resources Classification scheme (see section 5.6.3 fig. 5.8). Obviously other classification schemes may be similarly incorporated.

Since soil erosion depends on a wide number of factors, adequate modelling of the knowledge on soil erosion requires rule files spanning several related domains.

The rule base file SPECTRAL_SENSING contains knowledge on procedures and resources for spectral analysis and classification of ground cover, soil types, soil characteristics etc., which are essential inputs to the soil loss estimation and modelling process. Examples of subject specific rule base files under this knowledge category are, SPECTRAL_BANDS, containing knowledge on the designation of spectral channels, and BAND_APPLICATIONS, containing default rules for selecting spectral channels for various applications. COVER contains knowledge on vegetation and crop cover classification systems which are used in modelling default USLE model parameters. Procedures for systematic aerial photo interpretation such as the dichotomous crop interpretation key can also be easily programmed, e.g. SLEMS rule base file DI_INTERPRETATION (fig 5.2) under the COVER domain. The hierarchical organisation of the knowledge in this file is shown later in section 5.6.3 figure 5.7.

The SOIL file contains knowledge on soil classification systems including the Canadian taxonomic soil classification system (ACECS, 1987) contained in the CLASS_SYSTEM and CLASS_SCHEME (not shown on fig 5.2) files. Knowledge on various soil properties used to assign soil classes, such as, soil horizons and soil texture are also included in the rule base files HORIZONS (not shown) and TEXTURE respectively. The TEXTURE file, for example, contains two sets of default rules for designating soil texture according to soil particle size and application domain i.e. agricultural usage or engineering usage. The organisation of the SOIL domain file is shown in section 5.6.3 figure 5.9. These files constitute an important class of knowledge required by conservation experts for the determination of soil erodibility parameters and erosion susceptibility.

Other files under the EROSION domain not appearing in figure 5.2 include USLE_MODEL, P_DFACTOR, EROSION_HAZARD, DRAINAGE, and ALPHA_RULE, CHERZY_RULE containing samples of various kinds of knowledge needed to classify and assess erosion hazard potential. In addition other files such as RAIN, WIND, etc. shown in figure 5.2, may be used to input and store knowledge

on other factors influencing soil erosion.

Each knowledge base file is in ASCII format meaning that, the rule components are stored as character strings. There is, therefore, no requirement for preprocessing, to translate the real world knowledge represented by tuples of character strings into integer tuples as, it is done in Martin (1984) in conjunction with triple stores. The general structure of the rule base is shown in figure 5.2.

Searching of the knowledge base is performed sequentially, until the inference module dictates branching to a new starting point. If the branching rule is a meta rule, searching will proceed in a separate rule base file. This is illustrated by the links (figure 5.2) between the high level domain file EROSION, and the subject specific rule files for CN estimation (USDA_CN and CN_LANDSAT). Rules for the estimation of default USLE parameters represented by S_FACTOR, R_FACTOR, K_FACTOR, L_FACTOR, C_FACTOR and P_FACTOR are similarly linked to the EROSION file. These files contain default rules for selecting USLE parameters organised, through use of appropriate meta-level premises, by source, method, and location.

The files shown in figure 5.2 by no means exhaust the required knowledge. They are only representative of the variety of knowledge required for soil loss modelling and estimation. During practical use only those files which are essential at any particular level in the knowledge acquisition process need be established. As the need for additional domain and subject area specific knowledge arises, new files can be dynamically added at any level in the hierarchy of rule base files.

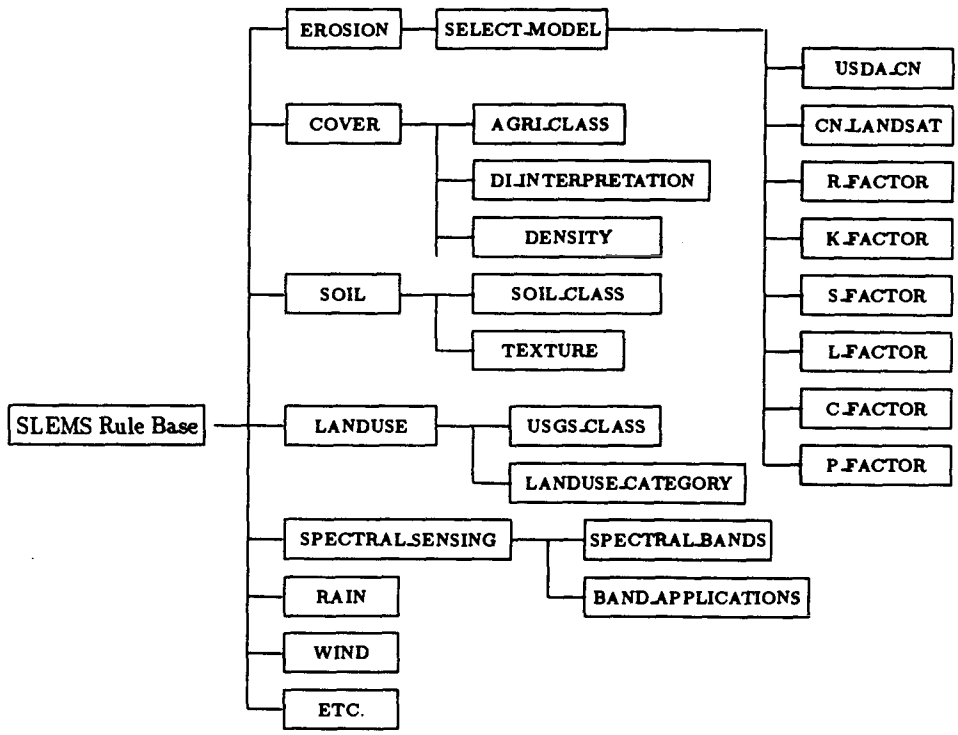


Figure 5.2: The Structure of the SLEMS Rule Base.

Generally files which serve as links across rule base files represent semantic links among the real world domain objects. The structure formed by the domain links, and subject links constitutes a semantic network. Thus both from the design point of view and implementation point of view, the SLEMS rule base may be characterised as a rule based semantic network knowledge representation structure.

The hierarchical structure of the rule base facilitates the intelligent representation and manipulation of domain knowledge as explained in the next section.

5.4 Representation of Domain Knowledge in the EXPERT.

The C language implementation of the algorithms and knowledge structures discussed in sections 5.2.1 and 5.3.1 facilitate the means for restructuring the domain knowledge into a form suitable for simple inferencing by the query module. This is achieved through a combination of the hierarchical organisation provided by the active memory structures and the appropriate organisation of the domain knowledge for input. The hierarchical structuring of the domain knowledge is achieved through the appropriate selection of knowledge categories, subject area, and subject specific attributes or premises at the time of rule entry or knowledge engineering stage in accordance with the inherent structures of the domain knowledge (Barr and Feigenbaum, 1981, 1982; Stricklen et al 1987)

Two types of rules are used to achieve the efficient representation of knowledge by the SLEMS EXPERT. Rules of the first kind represent knowledge about the SLEMS knowledge base organisation and components. Figure 5.3 shows the organisation of this type of knowledge. For example, after loading the rules in the system file (SLEMS), the EXPERT subsystem can use appropriate cues such as "soil loss", "ground cover" etc. to establish the area of interest of the user. Once this is established the EXPERT subsystem then advices the user on which domain and domain

rules he should use to facilitate appropriate answers to his queries. This process could easily be automated so that the system loads the appropriate rules instead of advising the user to load them.

Rules of the second type represent knowledge on the domain of expertise and application. These rules may either be global, spanning over several subjects, or localised within a specific subject area of the domain. Figure 5.4 depicts the organisation of this kind of knowledge at various levels of the soil loss estimation and modelling process. It represents an abbreviated form of the knowledge programmed into the SELECT_MODEL rule base file appearing also in figure 5.2.

After loading the SELECT_MODEL file, the EXPERT subsystem uses its rules to establish what the user is specifically interested in, i.e. selection of some USLE model or estimation of some specific USLE parameter. If, for example, the user says he wants to select a specific USLE model, the next meta-level prompt might be whether he wishes to make use of the irregular terrain variant of the USLE (Foster and Wischmeier, 1974) or the modified USLE model (MUSLE) (Bondelid et al, 1980, 1981; Mills, 1986). If he chooses MUSLE the system then asks whether CN values are available since they are essential for this model. Depending on the users response the system would then proceed along the assumption that CN values are known or unknown. If the CN values are unknown the system would advice the user to estimate them using either the USDA-SCS AMC-II, the USGS LUDA or the LANDSAT CN methods.

An example of the organisation of the knowledge in a subject-specific rule base file is shown in figure 5.5 representing knowledge on the selection of C-factors for the USLE model (Wischmeier and Smith, 1957; Bondelid et al, 1980, 1981; Hóly, 1980; Stephens et al, 1985).

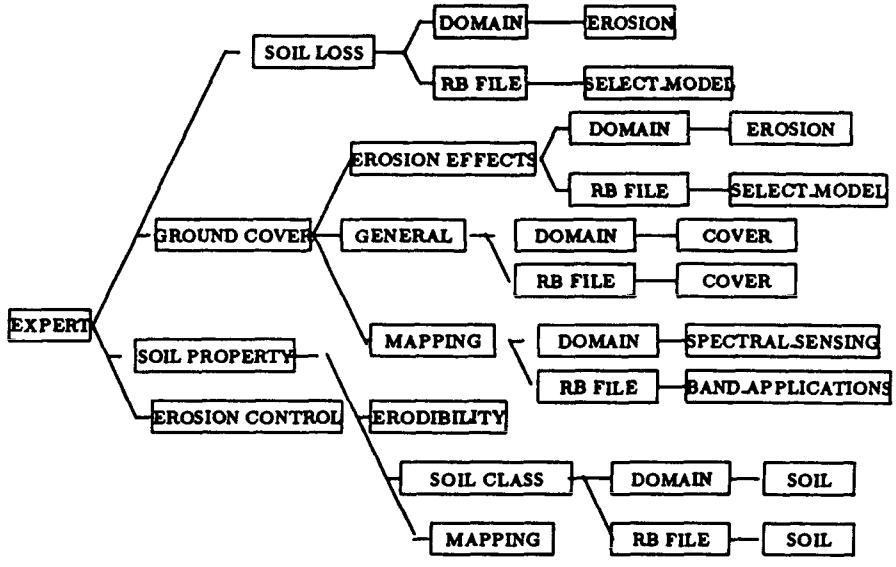


Figure 5.3: Meta-Knowledge About the SLEMS Knowledge Base Structure and Applications.

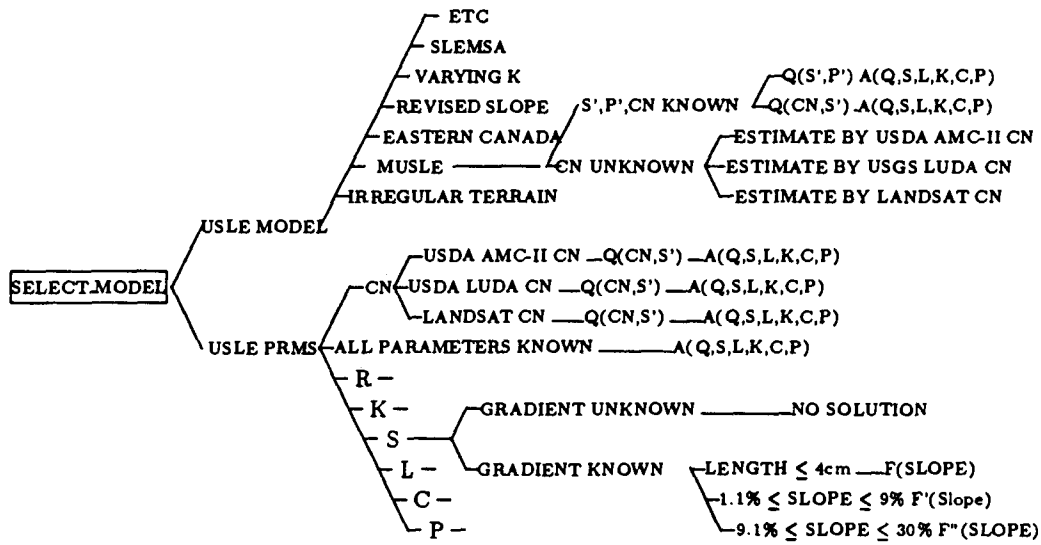
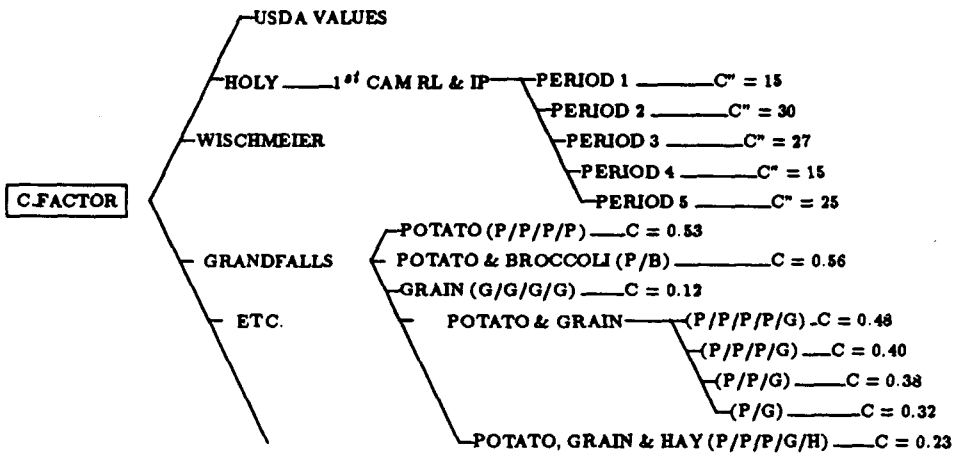


Figure 5.4: Knowledge About the Soil Loss Modelling Process.



(P/G...): Crop rotation; e.g. Grain after potato
1st CAM RL & IP: first year corn after meadow, residue left and incorporated by ploughing
PERIOD 1, PERIOD 2, etc.: Crop stage.

Figure 5.5: A Segment of Knowledge on Selection of USLE C FACTORS.

To improve searching, meta-level premises such as method inventor (HOLY, WISCHMEIER, etc.), or organisation (USDA defaults), and location (GRANDFALLS values) are used to restrict the domain and subject level categories of knowledge relevant to a query solution. The SLEMS scheme of rule base knowledge entry, facilitates coherent and systematic organisation of the knowledge base and reduces the size of each rule file. The scheme also guarantees efficient searching of the rule base by restricting the number of rules which need to be loaded at the beginning of the query session.

Each SLEMS rule base file dynamically governs the behaviour of the inference module and acts as a search control at query time. The efficiency and versatility with which each rule base file is compiled determines the level of intelligence demonstrated by the inference module and subsequently by the EXPERT subsystem. Issues related to the optimization of search control and the rule base are discussed only briefly in the next section. Relevant information and details on general principles and methods for optimization of search control in rule based inference systems can be obtained from various Artificial Intelligence and Expert Systems literature sources including Minton (1988), Ginsburg et al (1987), Doyle (1987), and Cohen (1985). An excellent presentation and formalization of the principles of default reasoning with semantic networks is given in Shastri (1989).

5.5 Computation Complexity Analysis Issues and Rule Base Optimization.

Considerable speeding up of the searching process is achieved by introducing premises reflecting higher level domain knowledge in the specification of lower hierarchy rules. Such premises facilitate discrimination of knowledge into knowledge clusters related through the higher level premises and achieves the desirable effect advocated in the knowledge representation principle of maximum discriminant concepts elaborated in

Fisher and Langley (Gale, 1986) and Luger and Stubblefield (1989).

Premises about higher level knowledge strategically placed at the beginning of a rule's premise list serve as meta rules for search control. In this way it is possible to reduce the number of wasteful prompts for verification of rule premises. The gain in search efficiency is demonstrated by considering that (figure 5.6) if location, method inventor, and owner organisation are not incorporated as meta level premises, the inference module would have to prompt the user for verification of all the premises in each of the first two categories even if the user was only interested in using the USDA conservation management factors (P factors). By first asking the user if he is interested in the Grandfalls data set, or Wischmeiers method, the module skips all rules on Grandfalls and Wischmeiers method if the answers is no. Assuming n rules in each category, and m categories of rules, then the worst case scenario is $O(n.(m-1))$ wasteful prompts.

The worst case time complexity for processing a query in a knowledge base of n rules where each rule has a maximum of m premises is of the order $O(n.(m-1))$ assuming that rules are either highly correlated or the user does not respond to prompts correctly. This case may arise, for example, if each time the user responds with a yes for each premise but the last of each rule, and assuming the solution is found in the last rule.

It is clear, that, on the basis of the computational complexity analysis, the search time can be improved if:

1. The number of the rules in the knowledge base is kept small.
2. The number of premises per rule is kept to a minimum.
3. Premises which discriminate knowledge at hierarchically higher levels are introduced and placed at the front of individual rule premises to act as meta rules for search control.

4. The design and implementation of the knowledge base is such that, correlation among the premises of the different rules is eliminated or kept to a minimum.

The current implementation of the SLEMS is not efficient in terms of storage space requirements. This is obvious, because similar premises appearing in several rules are stored multiple times. This means that if a total of n rules share m premises each requiring k bytes of storage, then the number of bytes wasted is $(n-1).m.k$. Since k is normally constant the storage inefficiency can be considered of the order $O((n-1).m) \approx O(n.m)$.

A lesser but significant storage redundancy also occurs when several rules share the same conclusion. In this case the redundancy is $O(n-1)$ where n is the total number of rules sharing the same conclusion.

The problem of redundant storage in the rule base may somewhat be reduced by implementing the recommendation in item 4 above. Alternatively new storage structures must be implemented to facilitate the elimination or reduction of the number of redundant premises or conclusions stored. A possible strategy to achieve this is outlined in section 6.

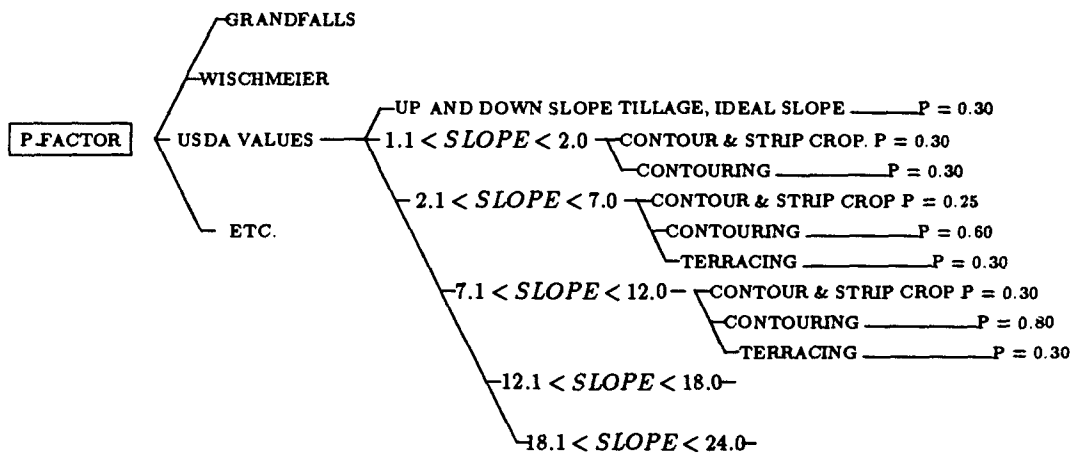


Figure 5.6: A Segment of Knowledge on Selection of USLE P FACTORS Using Location, Inventor or Organisation as Meta Level Premises for Partitioning the Knowledge Base and Shortening the Search Path.

5.6 Knowledge Compilation With the EXPERT Subsystem.

Knowledge compilation using the EXPERT is a two stage process. Assuming a specific soil erosion problem, the first stage is to analyze the resource and procedural knowledge requirements for its general solution. Next the logical relationship among the various knowledge components are examined to discover the inherent dependencies. At this stage dependency graphs akin to those in figures 5.1, 5.2, 5.3, and 5.5 are constructed and used to determine appropriate knowledge structures for implementation.

After preliminary analysis the knowledge structures are translated into human readable rules according to the syntax of the SLEMS rule base. Once all the domain specific rules and necessary link files have been established, the SLEMS system file is updated to reflect the new knowledge base entries.

Actual compilation consists of an interactive session with the EXPERT in which the EXPERT prompts for the components of each rule until the knowledge for the specific domain is all entered. At the end of the editing session the domain rules are stored under the specified subject. A typical rule entry is shown in example 2 below.

Example 2:

Rule Entry Session With the EXPERT's Editor

enter next object <blank> to quit

Domain EROSION: ID 71

please enter the object name:

> CN = 91

enter rule explanation:

> Rule for selecting default runoff curve number (CN)

enter object attributes:

> CN by USDA-SCS AMC-II method

enter uncertainty for this attribute: range [0-1]

> 1

enter next attribute <blank> to skip:

> Landuse is streets and roads

enter uncertainty for this attribute: range [0-1]

> 1

enter next attribute <blank> to skip:

> Cover is gravel

enter uncertainty for this attribute: range [0-1]

> 1

enter next attribute <blank> to skip:

> Hydrologic soil group is D

enter uncertainty for this attribute: range [0-1]

> 1

enter next attribute <blank> to skip:

>

enter next object <blank> *to quit:*

>

please remember to save the rules into the rule base

In example 2 above the EXPERT subsystem first establishes the domain of interest of the user. It then displays the domain of interest and current rule identification number (automatically assigned ID number) and prompts the user to enter the rule conclusion, which the system interprets as the object part (object name) of the object-attribute (OA) tuples of the EXPERT_OBJECT types discussed earlier in chapter 2 section 3.2.2. The system then prompts for the explanation of the rule, which is essentially, a short description of the rule purpose. Next the system prompts for the rule premises and their certainty factors. It interprets rule premises as the attribute

part of the of the OA-tuple.

In general the certainty factors (CF) are subjectively assigned by the domain expert to reflect the importance or contribution of the premise to the rule conclusion. In the example they have been arbitrarily assigned the value of 1 to reflect absolute certainty in the premise's contribution.

The value, CN = 91, specified as the rule conclusion would normally be selected by the conservation specialist depending on hydrologic, vegetative, and hydrologic soil group factors predominant in the land unit under consideration. In example 2 this value has been adopted from USDA-SCS AMC-II CN values documented in Bondelid et al (1980, 1981). If the rule entry session is broken for some reason, the partial knowledge can be saved and the entry resumed at a later date.

Although the format of the entry used in example 2 mentions nothing about rules, each attribute entered is a precondition for the value CN = 91 to be used. In the EXPERT subsystem the same OA-tuple structure is used to hold both facts and rules as explained earlier. In particular the attribute or premise "CN by USDA-SCS AMC-II method" represents a meta rule and it restricts the CN values having this precondition from being used for application in curve number estimation methods other than the USDA-SCS AMC-II method.

Using the approach suggested above, representative knowledge in various domains related to the soil erosion problem was compiled and programmed into the SLEMS rule base. The bulk of this knowledge is contained in the rule base files discussed in the preceding sections (figure 5.2). It includes the knowledge directly relevant to soil loss estimation and modelling consisting of rule base files EROSION, SELECT_MODEL, CN_LANDSAT, USDA_CN, S_FACTOR, K_FACTOR, R_FACTOR, L_FACTOR, C_FACTOR, and P_FACTOR; knowledge needed to model the USLE parameters consisting of rule base files COVER, SOIL, LAND_USE etc., and peripheral knowledge about procedures and requirements for acquiring data about soil, cover, landuse, etc.

contained in rule base files SPECTRAL_SENSING, DI_INTERPRETATION, SPECTRAL_BANDS, BAND_APPLICATIONS, etc. As pointed out earlier this knowledge is not exhaustive. It was compiled quickly, with no input from domain experts, to demonstrate the practicality of the concepts embodied in the design of the SLEMS EXPERT subsystem.

5.6.1 Practical Use of the EXPERT in Soil Loss Study.

In this section several examples are presented to demonstrate the practical application of the SLEMS EXPERT for soil loss estimation and modelling related problems. These problems generally fall into four categories (Lal, 1977; Hóly, 1980; Goldman et al, 1986; Morgan, 1986; Ventura et al, 1988).

- selection of an appropriate soil loss estimation model and soil loss computation for known conditions and geographic location.
- selection of the soil loss model parameters under known geomorphology, soil, ground cover, and erosion control practices.
- selection of a conservation practice for a known soil loss tolerance under various conditions.
- Classification of soil erosion potential hazard and impact.

Assuming that the domain knowledge necessary to solve the three problems above has been programmed into the SLEMS rule base, the session begins with the EXPERT prompting for the domain and subject area (file name) of interest. Before commencing the actual knowledge base search, the system further prompts the user to select a desirable method for handling uncertainty. Three different approaches may be used to propagate uncertainty; conventional probabilistic approach if the uncertainty is believed to be normative, and the MYCIN-like (Barr and Feigenbaum, 1981; Cohen, 1985; Barr et al, 1989) approach of weakest link and strongest link methods (Schildt,

1987), depending on whether the aggregation of knowledge and uncertainty is considered conjunctive or disjunctive respectively. After finishing with the preliminaries the system is ready to perform interactive consultation with the user. Example 3 demonstrates the use of the EXPERT in the selection of C factors for the USLE soil loss estimation model. In all the examples, some of the dialogue has been streamlined or omitted to highlight the main line of thought. Each emphasized output is a system prompt requiring simple yes/no user response.

Example 3:

Selection of the USLE C Factor.

Is this attribute valid for the query object [Y/N/W]
Default values for the Central Indiana (US)
are adequate
> yes
Landuse is cultivation
> yes
Crop type is grass and legume meadow (M), Oats and
Spring seed small grain (O)
> no
Crop type is grass and legume meadow (M), Oats and spring
Seed small grain (O), wheat or fall seeded grain (W)
> no
Crop type is grass and legume meadow (M), wheat or
Fall seeded grain (W), corn (C)
> yes
Crop rotation is M-M-M-C-W
> no
Crop rotation is M-M-C-W

> no

Crop rotation is M-C-W

> yes

Hay yield is 2 to 3 tons, corn yield 60 to 74 bushels

> yes

Residue left on surface over winter, ploughed under or disked and left on surface in spring

> yes

By domain EROSION rule no. 34

Crop cover factor C = 0.090

The current facts fit the rule with

a certainty factor of 100%

EXPERT: continue ?

In example 3 above the system starts by asking the user whether he considers default values for Central Indiana (US) to be appropriate for his use. If the user wishes to know why the system requires this information he could respond with why (W), in which case the system would respond by displaying the current rule it was examining. Otherwise the system follows with a series of prompts corresponding to specific premises of the current rule, in the SLEMS rule base, which is being examined. By keeping track of the users responses the system determines which rule satisfies the users query (the one in which all responses were "yes"). It then displays the rule conclusion as the answer to the query. In this example the system uses knowledge on crop classification units, e.g. "crop type is grass and legume meadow"; crop rotation sequence, e.g. "M-C-W" representing rotations of meadow (M) followed by corn (C), and fall seeded grain (W); yield per acre and tillage system, e.g. "residue left on surface over winter..."; to determine an appropriate crop cover factor (C = 0.090). The system also informs the user about the domain whose rules were used (EROSION),

and the identification number of the rule (rule no. 34) used to assign the C-factor value. In this case the rule consists of pre-conditions for assigning C factor values based on Wischmeier and Smith (1957) default values. The certainty factor of the conclusion (i.e 100%) was in this case arrived at by propagating the certainty factors (expressed in %), of the premises of the successful rule which, in this case, appear to have all been set to the value of 1 at knowledge compilation time. The rule used in this example is shown below.

Domain EROSION: ID 34

IF default values for Central Indiana (US) are adequate
with a certainty of 100%
AND landuse is cultivated
with a certainty of 100%
AND crop type is grass & legume meadow (M), wheat or fall
seeded grain (W) and corn (C) with a certainty of 100%
AND crop rotation is M-C-W with a certainty of 100%
AND hay yield is 2 to 3 tons, corn yield 60 to 74 bushels
AND residue left on surface over winter, plowed under or
disked and left on surface in spring
with a certainty of 100%
THEN crop cover factor C = 0.090.

The certainty values of 100% for the premises of the rule have been subjectively assigned. In this example no certainty value has been attached to the rule conclusion. This means that the computed certainty of the conclusion will only depend on the propagated certainty factors of the premises.

5.6.2 Modelling Runoff and Soil Loss With Curve Numbers From Landsat Data.

An important hydrologic parameter in the modelling of watershed runoff is the runoff curve number CN (Bondelid et al, 1980; Mills, 1986). This parameter is dependent on the basin retention parameter S and the precipitation parameter P (Mills, 1986). S depends on a number of basin characteristics including, soil type, landuse, and antecedent moisture. The curve number is used in the computation of the volume of runoff volume Q (Eq. 5.6).

$$\begin{aligned} Q &= \begin{cases} \frac{(P-I)^2}{(P-I)+S}, & \text{for } P \geq I; \\ 0, & \text{for } P < I. \end{cases} \\ I &= 0.2 \times S \\ S &= 2.54 \times (\frac{1000}{CN} - 10) \end{aligned} \tag{5.6}$$

where Q is annual runoff volume in cm of depth over the watershed, P is the annual rainfall depth in cm, I is the initial abstraction in cm and S is the watershed retention (storage) capacity in cm (Bondelid et al, 1980; Mills, 1986).

The volume of runoff estimated in this way is used in the estimation of soil yield by the modified USLE or MUSLE (Mills, 1986) as given by Eq. 5.7.

$$Y = 11.8 \times (Q \times q)^{0.56} \times K \times SL \times C \times P \tag{5.7}$$

where Y is the soil yield from individual storm in metric tones, Q is the runoff volume in m^3 and p is the peak runoff volume in m^3 and K, S, L, C, and P are the standard parameters of the USLE (Wischmeier, 1984; Mills, 1986).

Several methods exist for the estimation of CN values differing in the data set, landuse classification, and method used to model the basin retention parameter S (Bondelid et al, 1980; Mills, 1986). Currently limited knowledge on the USDA-SCS AMC-II method, USGS LUDA method, and the LANDSAT CN method (Bondelid et al, 1980, 1981) has been programmed into the SLEMS knowledge base. The USDA-SCS AMC-II method is based on medium to large scale topographic maps and the

USGS level I and level II landuse classification. Default CN values for this method are assigned by soil erosion and conservation specialists based on cover type, hydrologic soil groups, and antecedent moisture conditions (AMC) designated class AMC-II (Bondelid et al 1980, 1981). The USGS LUDA curve numbers are based on USGS land use development agency (LUDA) maps and the USGS level I and level II land use classification. The LANDSAT CN method differs from the first two methods both in the resolution and detail of land use classification and the map base used for their compilation. The land use classification used in this method is based on maps compiled from Landsat MSS image classification and is, therefore, more coarse. However, Bondelid et al (1980, 1981) have demonstrated that CN values estimated by the three methods are comparable.

The USGS LUDA method rules have been merged into the LANDSAT rule base file because they constituted only a small file. Example 4 shows how the USDA_SCS AMC-II method (Bondelid et al, 1980) is used by the EXPERT to select CN values for an urban residential area runoff prediction.

Example 4:

Selection of runoff CN values for urban area runoff volume prediction using USDA-SCS land use classification and AMC-II antecedent soil moisture conditions.

Landuse is industrial district

> no

Land use is residential

> yes

Average plot size is equal or less than 1/8 acre

> no

Average plot size is equal or less than 1/4 acre

> no

Average plot size is about 1/3 acre

> yes

Cover is about 30% impervious

> yes

Hydrological soil group is A

> no

Hydrological soil group is B

> no

Hydrological soil group is C

> no

Hydrological soil group is D

> yes

Management is good condition, lawns in good pasture conditions, roof water directed off lawns

> yes

By domain EROSION rule no. 54

Runoff curve number CN = 86

The current facts fit the rule with

a certainty factor of 100%

EXPERT: continue ?

The system prompts in example 4 above consists of specific landuse categories at USGS level II classification, land parcel size, predominant cover type, and hydrologic soil group. Like the first example, these prompts correspond to premises of rules in the SLEMS rule base, in this case the USDA_CN file. A yes response for all the prompts is interpreted by the system to mean that the rule currently being examined by it satisfies the users query. If a premise is rejected, then depending on the certainty factor of the premise, the rule is discarded and the system proceeds to examine another

rule.

Once the CN value is determined the user can then consult the SLEMS EXPERT on the appropriate model for runoff and soil loss computation from runoff curve numbers (CN) as demonstrated in example 5 below.

Example 5:

Model for computing runoff volume (Q) from curve numbers (CN).

Soil loss estimation by modified USLE (MUSLE)

> yes

Estimation of volume of runoff Q from curve number CN

> yes

Runoff curve number CN is known

> yes

Retention parameter S is unknown and precipitation parameter P is known

> yes

By domain EROSION rule no. 5

Runoff volume Q = (P - 0.2*S)2 / (P + 0.8*S),**

where S = 254/CN - 10

The current facts fit the rule with

a certainty factor of 100%

EXPERT: continue ?

The solution in this case consists of an advice on the computational formulae suitable for computing volume of runoff based on the precipitation, P, basin retention parameter, S, and the curve number, CN. Alternativelly the system could use this formulae to compute the required quantities (not yet implemented).

Examples 1, 2 and 3 given above illustrate well the variety and nature of knowledge required by soil erosion experts for the appropriate selection of parameters for empirical hydrologic and soil loss modelling. The same knowledge is used by the EXPERT subsystem to facilitate interactive selection of CN values using the USDA-SCS landuse classification scheme and data for the AMC-II antecedent soil moisture conditions.

Clearly a sizable chunk of the information required by the EXPERT is the type which is easily available from remote sensing and GIS sources. These examples, also serve to show the relevance of these two technologies in the estimation and modelling of soil loss.

5.6.3 Knowledge on Remote Sensing Resources and Applications.

The final example in this section illustrates the possible use of the SLEMS EXPERT to compile and use knowledge on remote sensing methods and procedures. The EXPERT subsystem can in this respect be used as a guide to other domain experts not conversant with remote sensing methods on the selection of suitable sensors, spectral bands, spectral analysis procedures and classification methods. Example 6 demonstrates an application of the system for the selection of spectral bands appropriate for sensing vegetation and soil moisture discrimination.

Example 6:

Selection of spectral sensing bands for vegetation cover
and soil moisture sensing

Satellite image is Landsat TM

> yes

Designated use is sensing in the chlorophyll

absorption region

> no

Application is determination of vegetation type and vigour

> yes

Application is determination of vegetation biomass content

> yes

Application is delineation of water bodies

> yes

Application is soil moisture discrimination

> yes

By domain SPECTRAL_SENSING rule no. 5

Use Landsat TM BAND 4

The current facts fit the rule with

a certainty factor of 100%

EXPERT: continue

The rule whose premises are used by the system to elicit user response is shown below and is self explanatory. The prompts for which the user responded in the affirmative correspond to the premises of rule number 5, in the BAND_APPLICATIONS file of the SPECTRAL_SENSING domain, given below.

Domain SPECTRAL_SENSING: ID 5

IF Satellite image is Landsat TM with a certainty of 100%

AND Application is determination of vegetation type and vigour with a certainty of 100%

AND Application is determination of vegetation biomass

content with a certainty of 100%

AND Application is delineation of water bodies

with a certainty of 100%

AND Application is soil moisture discrimination

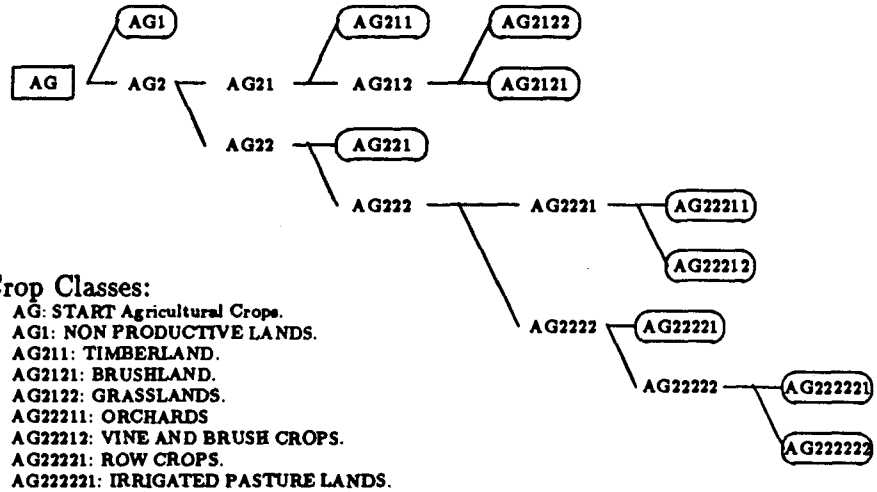
with a certainty of 100%

THEN Use Landsat TM BAND 4.

The premises in the sample rule above would normally be determined on the basis of experience or experimental evidence of the suitability of the selected spectral channel for mapping the various types of information classes. The certainty factors may similarly be assigned subjectively based on experience or on the basis of classification error analysis in controlled test applications. A designated application is the use for which specific spectral channels were initially designed.

Figure 5.7 shows the classification decision tree used to implement knowledge on di-chotomous photo-interpretation of crops. In this method an interpreter proceeds by picking one of two possible descriptions of the scene appearing on the photograph at each node in the decision tree. As the distance from the root node (AG) increases the scene descriptions become more specific until there is no more branching. The lowest level nodes or leaf nodes such as AG1, AG2121, AG2122, etc, correspond to the actual names of the identified ground cover categories. Inner nodes in the decision tree correspond to the preconditions for branching to either of the pair of scene descriptions. In the actual implementation of the dichotomous key the symbols AG1, AG2, etc are replaced by the actual strings shown on the key associated with figure 5.7.

Similarly figures 5.8 and 5.9 show how the system can be used to input and manipulate agricultural and soil classification knowledge. This kind of application can be exploited by less experienced personnel to make quick expert level decisions based on the compiled experience and knowledge of domain experts.



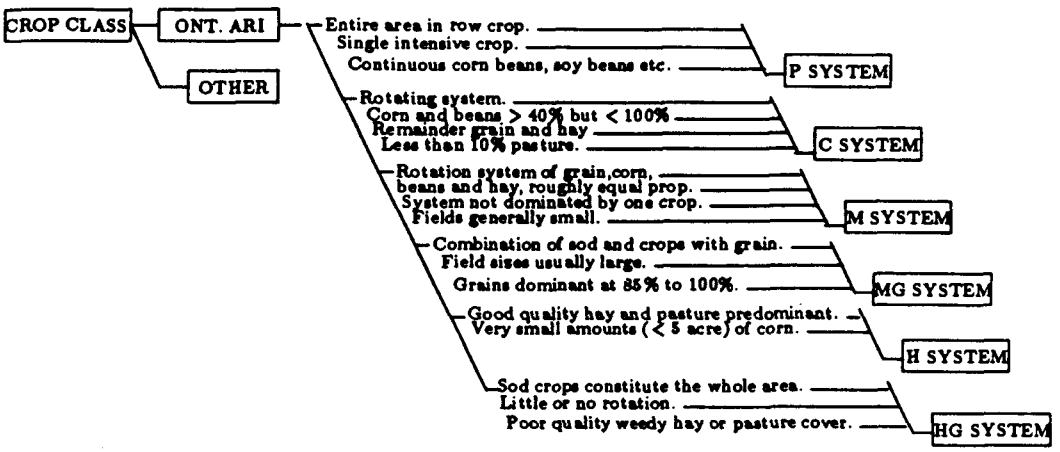
Crop Classes:

AG: START Agricultural Crops.
AG1: NON PRODUCTIVE LANDS.
AG211: TIMBERLAND.
AG2121: BRUSHLAND.
AG2122: GRASSLANDS.
AG22211: ORCHARDS
AG22212: VINE AND BRUSH CROPS.
AG22221: ROW CROPS.
AG222221: IRRIGATED PASTURE LANDS.
AG222222: CONTINUOUS COVER CROPS, SMALL GRAINS, HAY ETC..

Clues for Interpretation.

AG1: Vegetation & soil absent or obscured by rock etc.
AG2: vegetation clearly discernible on photos.
AG21: cultivation pattern absent; field boundaries irregularly shaped
AG22: cultivation pattern absent; field boundaries regularly shaped.
AG211: trees present, covering most of the ground surface
AG212: trees absent or widely scattered, ground covered by low lying vegetation.
AG2121: crowns of individual plants discernible; texture coarse and mottled
⋮

Figure 5.7: A Segment of Knowledge on Aerial Photo Interpretation of Crops Using a Dichotomous Crop Interpretation Key.



C: Corn ONT. ARI : Ontario Agricultural resource Inventory
M: Mixed H: Hay
MG: Grain HG: Pasture P: Continuous row crop

Figure 5.8: Segment of Knowledge on Agricultural Crop Classification Systems.

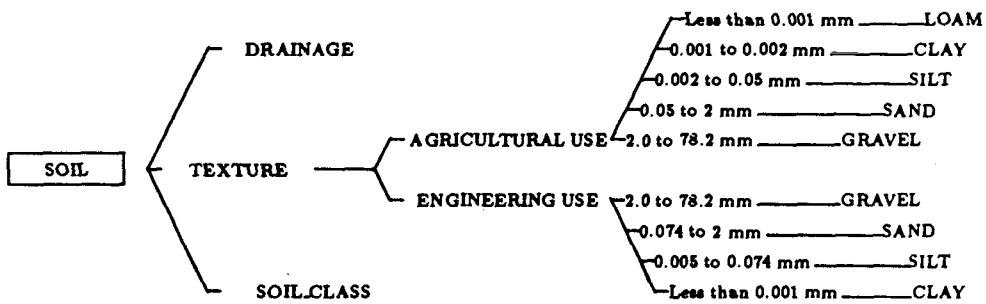


Figure 5.9: A Segment of Knowledge on Soil Classification by Texture.

The EXPERT subsystem can therefore function as a tool for facilitating inter-domain knowledge transfer. Also, as it can be seen in the examples given in the previous sections, the EXPERT subsystem satisfies most of the specifications and objectives put forth in chapter one regarding knowledge representation and system functionality.

Chapter 6

Future Enhancements and Conclusions.

6.1 Proposals for Enhancing the SLEMS Subsystems.

The SLEMS current implementation faces three serious problems which may reduce its usefulness in real life applications. The first problem concerns the issues raised in conjunction with the fuzzy comparison operator in chapter four. Specifically the fuzzy comparison operator has not been tested in real life application.

The second problem concerns inefficiency of the current inference module and the simple user interface. It also encompasses the problem briefly discussed in section 5.5 regarding storage space inefficiency. These problems are discussed further in the next two sections where some possible solutions are suggested. The third problem is the lack of spatial data handling capability. Solution to this problem can be found in the form appropriate interfacing to some existing spatial information analysis system, such as the PCI EASI/PACE geographic and image analysis package or the CARIS geographic information system, residing on the Sun 4 Workstation.

6.2 Future Improvements to the Fuzzy Comparison Operator.

To resolve the problem of subjectivity in the assignment of the values of the fuzzy partitions, calibration of the user perception of the fuzzy restrictions was earlier proposed as a possible solution. Such development would require the system to initially present a few queries to the user to establish enough data for calibration of the partitions. Possible strategy might consist of presentation of simple assertions containing the fuzzy predicates most likely to be encountered by the user. The user would have the option to agree with the suggested propositions or modify them to suit his own understanding.

Using the calibration data the system would then modify internal parameters which control and set the partitions of the search space to conform with the users subjective perception of the fuzzy predicates. However some difficult problems have to be solved before an efficient user calibration procedure is obtained. First a sufficiently robust testing procedure has to be designed to ensure that the calibration process sufficiently well models the users perception. Secondly the process assumes that the system is able to predict in advance the type of fuzzy predicates most likely to be used by the user. This is a complex problem requiring modelling of user expectations (Kass and Finin, 1988; Finin, 1988). The alternative is to test the user on all the possible fuzzy restrictions. This is not a viable solution as it would consume too much of the users time resource.

6.2.1 Proposals for Improvements to the EXPERT Inference Module.

The current implementation of the SLEMS EXPERT subsystem relies too much on user responses to guide the searching process. To make the searching process more

automatic requires an interface module which will search existing databases and translate the existing facts into appropriate responses for the inference module. This would have the desirable effect of reducing the number of times the system has to prompt the user. However implementation of such an interface is not a trivial matter and requires the introduction of complex natural language query systems (NLQS) based solutions (Obermeier, 1990).

Problems with the NLQS approach include the inherent complexity of natural language understanding systems, and the multiplicity of database query languages as discussed recently by Obermeier in the August issue of the Byte (Obermeier, 1990). Other problems and solutions for natural language interfaces have been discussed in great detail in Hahn et al (1980), Dahlgren (1988), Minker (1980), Cerone (1980) and others in Bolc (1980a, 1980b, 1980c, 1980d). The second approach assumes a knowledge based solution. Assuming that all the required knowledge is compiled by the SLEMS knowledge acquisition modules, improvements on the system's performance may be achieved in two ways:

- The user first inputs a query list consisting of observed real world object attributes and the module verifies the attributes against existing facts (the LEARNER_OBJECT types) until they are exhausted before it resorts to user prompts.

- Knowledge about the real world is compiled in the form of LEARNER_OBJECT beforehand using the SLEMS LEARN subsystem, the EXPERT subsystem is then used to automatically analyze and classify the real world knowledge.

The first aspect of this solution requires only cosmetic user interface improvements and has in fact been partially dealt with in the current implementation of the EXPERT subsystem. When in use it enhances the consultation session by suppressing the number of user prompts.

The second item of the problem above requires more substantial improvements to the user interface and a sophisticated system to system interface for interfacing the EXPERT and other run-time programs. Such programs could deposit their results into files which would be accessed by the LEARN module and organized into knowledge triplets (LEARNER_OBJECT) ready for analysis and classification by the EXPERT using knowledge on domain classification rules.

6.2.2 Proposed Solution to the Redundant Storage Problem.

It was briefly stated in section 5.5 that the current SLEMS implementation suffers from the problem storage inefficiency due to redundant storage of rule components. To solve this problem future enhancements of the SLEMS should eliminate the need to store multiple copies of the same premises or rule conclusions. Because of the simple inference mechanism of the EXPERT subsystem general methods for handling problems of this nature such as unification cannot be used. Two strategies are therefore proposed to solve the above problem. Both of the proposed solutions require the introduction of two new storage file structures. One designated PREMISES, to store all the unique premises of all the rules, and another called CONCLUSIONS, to store all the unique rule conclusions of all the rules in each knowledge category file. The necessary sorting and identification of unique rule elements could be performed by two simple algorithms.

The algorithm for extracting the rule premises and storing them in the PREMISES rule base file is as follows,

Begin

For all rule base files

    Pick a rule base file

    For all rules

```
        Get the rule premise list
        For all premises in the list
            Examine a premise
            If the premise has been seen
                Then skip
            Else
                Save the premise in PREMISES
                Update premise counter
            End if
        End for all premises
    End for all rules
End for all rule base files.
End
```

Similarly the conclusions in each rule can be mapped into the CONCLUSION rule base file as follows,

```
Begin
For all rule base files
    Pick a rule base file
    For all rules
        Examine the rule conclusion
        If the conclusion has been seen
            Then skip
        Else
            Save the conclusion in CONCLUSIONS
        End if
    End for all rules
End for all rule base files.
```

End

At the same time new active memory structures (CONCLUSIONS and PREMISES) must be introduced to hold the character string arrays representing the unique premises and conclusions of the original rules. Using these structures the original rules can then be reorganised by replacing the rule components with pointers to their respective locations in PREMISES and CONCLUSIONS structures. This requires changes to the original active memory rule structures (figure 5.2), so that instead of holding character strings they now holds pointers to the CONCLUSIONS and PREMISES.

The processing necessary to set the pointers to their respective locations in CONCLUSIONS and PREMISES can be done by the following algorithm:

Begin

```
For all rule base files
    Pick a file
    For all rules in the file
        Match the conclusion against CONCLUSIONS
        If it matches
            Then replace conclusion with pointer to CONCLUSIONS
        Else signal unexpected failure (each conclusion must have
                a counterpart in CONCLUSIONS!)
        End if

        For all premises in the rule
            Match the premise against PREMISES
            If it matches
                Then replace premise with pointer to PREMISES
            Else signal unexpected problem (see comment above)
```

End for all premises in the rule

End for all rules in the file

End for all rule base files.

End

After checking and verification of the re-structured rule base the original rules can then be deleted. These modifications can be introduced without changes to the inference module because the new rule structure essentially retains the original form shown in figure 5.1 of section 5.3.1. Only the rule base loader and storage modules need to be modified to incorporate the PREMISES and CONCLUSIONS structures.

6.3 Concluding Remarks.

Testing of the SLEMS EXPERT, LEARN and FUZZ subsystem has only been demonstrated on a few examples based on knowledge generally available in soil erosion literature (Wischmeier and Smith, 1957; Bondelid et al, 1980; Hóly, 1980; Goldman et al, 1986; Morgan, 1986; Ventura et al, 1988). More elaborate testing involving domain experts is therefore required before any reliable conclusions can be made regarding its performance. Therefore statements regarding the performance of the system must be viewed in this context.

Based on the tentative results obtained in this research it is now possible to conclude that the knowledge based approach, proposed in this thesis, to the solution of soil loss related problem is viable. During the course of presentation of the theory and research results it has been abundantly demonstrated that the simple knowledge based tools developed in this research facilitate soil erosion domain knowledge compilation and manipulation. Also as mentioned in chapter 3 and 5 these tools essentially constitute a simple expert shell and can therefore be used to compile knowledge in any other domain.

In this respect the EXPERT subsystem facilitates knowledge capture and manipulation through simple natural language rules. It was adequately demonstrated, that by means of rules consisting of *if...then* EXPERT_OBJECT types, both factual and procedural knowledge about soil loss modelling and estimation processes can be entered into the system and queried.

The LEARN subsystem, on the other hand, constitutes the SLEMS primary knowledge acquisition tool. As explained in chapter 3, this subsystem contains a considerable number of class and inheritance operators which handle queries to the knowledge residing in the LEARNER_OBJECT types such as the MAY_OBJECTS the MUST_OBJECTS and the TABOO_OBJECTS. It has also been shown that, by exploiting the "is a" class, and "has" value assignment relationships explicit in the semantic network of SVO-triplets, the LEARN subsystem can give answers to queries requiring knowledge not directly entered into the system.

The fuzzy geometric partitions-based representation of fuzzy objects and manipulation of databases containing vague or fuzzy objects, has been demonstrated as a viable alternative for representing and processing fuzzy knowledge commonly encountered in the soil erosion business. The FUZZ subsystem and the fuzzy comparison operator (FUZZC) discussed in chapters 3 and 4 facilitate the necessary processing of fuzzy knowledge within the SLEMS system. Where higher accuracies are required membership functions, for modelling fuzzy restrictions on the domain of discourse, can be directly derived from the fuzzy partitions and used (not yet implemented) in the manipulation of the database objects by standard fuzzy sets methods. On the basis of the design criteria put forth in chapter one of the thesis the SLEMS performance may be summarised as follows.

- **Design concept:**

  - The SLEMS system in its current implementation retains modularity and simplicity as perceived at the design stage. The modular structure of the system permits the introduction of new sub-systems and modules as the need arises without substantial rewriting of the original code. The implementation of the whole system in standard C guarantees its portability.

  - An advantage of storing and retrieving knowledge base facts in natural language makes the whole process of knowledge engineering, compilation and query processing easy to grasp by the inexperienced user.

  - The main objective set out at the design stage was the acquisition of a system capable of compiling soil erosion domain knowledge using simple natural language rules and facts. Using a minimum of two types of structures (the OA tuples and SVO triplets) this goal has been moderately achieved. The simple knowledge structure used enables knowledge compilation with very few knowledge engineering processes. Also, as demonstrated in the examples given in chapters 3 and 5, the system is essentially an expert shell capable of acquiring domain knowledge from scratch. This extends its usefulness to other fields of application.

- **System performance:**

  - Based on the examples given in the previous chapters, the practical usefulness of the system has been demonstrated. It has been shown that it is capable of acquiring knowledge on a specific domain and disseminating that knowledge in a manner helpful to the inexperienced user. The system makes it possible to compile knowledge on task performance procedures from experienced experts and act as a consultant or advisor to the less experienced.

- A disadvantage of the use of natural language input with no translation into symbolic internal representation language is that it requires large amounts of storage resources. However this is balanced by the absence of preprocessing for knowledge input and query output.

- **Solution of the problem**

  Part of the objectives of the design of the SLEMS was to demonstrate an effective method for utilizing remote sensing and GIS knowledge in the soil loss estimation and modelling process. These objectives have been achieved in more than one way:

  - From a systems design approach the SLEMS essentially uses knowledge based methods to address the solution of spatial attributive information management for soil erosion studies. The current implementation does not incorporate linkage to graphical and image bases but these can be built in by appropriate interfaces and rule base files. The same applies for procedural attachment to facilitate computation of model parameters and actual soil loss.
  - Clearly methods and procedures for acquiring basic data necessary for soil erosion studies using remote sensing (example 6 in chapter 5) and geographic information systems can be effectively compiled into the system using appropriate rules as demonstrated by the given examples.
  - Vague information and knowledge frequently used by practitioners in the soil erosion field can be modelled by the FUZZ subsystem which makes possible the compilation and manipulation of vague facts or fuzzy predicates.

Thus in conclusion it is appropriate to state that all the objectives set out in the thesis proposal have been achieved.

References

[1] (ACECS) Agriculture Canada Expert Committee on Soil Survey, (1987). *The Canadian System of Soil Classification*, 2nd ed. Agric. Can. Publ. 1646 164 pp.

[2] Adamo, J. M. (1980). "Fuzzy decision trees," *Fuzzy Sets and Systems*, 4 (1980) 207-219, North-Holland.

[3] Aidoo, J. G. (1987). "Application of GIS to the assessment of land degredation and associated hazards in the Upper East Region of Ghana," *Proceedings of the Second Annual International Conference, Exhibits and Workshops on Geographic Information Systems*, GIS 87-SAN FRANSISCO, ASPR and ACSM, University of California, Berkeley, pp. 487-497.

[4] Anderson, D.G. (1979) "Satellite versus conventional methods in hydrology," In: M. Deutch; D.R. Wiesnet; A. Rango (Eds.) *Satellite Hydrology, Proceedings of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing*, American Water Resources Association, Sioux Falls South Dakota, June 10 - 15, pp. 33-36.

[5] Ardhi Institute (ARI) (1980-1987). Advanced diploma reports (Unpublished), Dare es Salaam.

[6] Argialas, D.P.; J.G. Lyon; O.W. Mintzer (1988). "Quantitative description and classification of drainage patterns," *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 4, April, pp. 805-809.

[7] Argialas, D.P.; C.A. Harlow (1990). "Computational image interpretation models: An overview and perspective," *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 6, January, pp. 871-886.

[8] Askari, F.; N.L.Faust (1981). "Automated classification of runoff coefficients for watersheds from landsat multispectral data," *Technical Papers of the Americal Society of Photogrammetry*, ASP-ACSM Convention, Washington D.C., Feb 22 - 27, Falls Church, pp 426-439.

[9] Asrar, G. (1989). *The Theory and Application of Optical Remote Sensing*, John-Willey & Sons, N.Y., Chichester.

[10] Bali, Y. P.; R.L. Karale (1977). "A sediment yield index as a criterion for choosing priority basins," In: *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July, pp. 180-188.

[11] Baldwin, J. F. (1979). "Fuzzy logic and its application to fuzzy reasoning," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp. 93-115.

[12] Baldwin, J. F. (1986). "Automated fuzzy and probabilistic inference" *Fuzzy sets and systems* Vol. 18, 1986, pp 219-235.

[13] Barret, E.C. (1979). "Satellite rainfall estimation by cloud indexing methods for desert locust survey and control," In: M. Deutch; D.R. Wiesnet; A. Rango (Eds.) *Satellite Hydrology, Proceedings of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing*, American WAter Resources Association, Sioux Falls South Dakota, June 10 - 15, 1979, pp. 92-100.

[14] Barr, A.; E.A. Feigenbaum (1981). *The Handbook of Artificial Intelligence*, Vol. I, Addison-Wesley Publishing Company, Inc., Reading, Massachussets, Menlo Park.

[15] Barr, A.; E.A. Feigenbaum (1982). *The Handbook of Artificial Intelligence*, Vol. II, Addison-Wesley Publishing Company, Inc., Reading, Massachussets, Menlo Park.

[16] Barr, A.; P.R. Cohen; E.A. Feigenbaum (1989). *The Handbook of Artificial Intelligence*, Vol. IV, Addison-Wesley Publishing Company, Inc., Reading, Massachussets.

[17] Blais, J. A. R. (1987). *Lecture Notes in Digital Mapping and Land Information*, University of Calgary.

[18] Best, R. G.; F.C. Westin (1984). "GIS for soils and range management," *PECORA 9 Proceedings, Spatial Information Technologies for Remote Sensing Today and Tomorrow*, October 2-4, Sioux Falls, SD. pp. 70 - 74.

[19] Bobrovitskaya, N. N.; I.N. Bogolubova; S.N. Tumanovskaya (1977). "Methods for computation of runoff and sediment yield from slopes used

for agricultural needs," In: *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July, pp. 62-66.

[20] Bolc, L. (Ed.) (1980). *Natural Language Based Computer Systems*, Carl Hanser Verlag, München Wien.

[21] Bolc, L. (1980). *Representation and Processing of Natural Language*, Carl Hanser Verlag, München Wien.

[22] Bolc L. (1980). *Natural Language Answering Systems*, Carl Hanser Verlag, München Wien.

[23] Bondelid, T.R.; T.J. Jackson; R.H. McCuen (1980). "Comparison of conventional and remote sensing estimates of runoff curve numbers in Southeastern Pennsylvania," *Technical Papers of the* 46_{th} *Annual Meeting of the American Society of Photogrammetry*, Washington,D.C, March, pp. 81-95.

[24] Bondelid, T.R.; T.J. Jackson; R.H. McCuen (1981). "Efficient curve number estimation using digital data bases," *Technical Papers of the Americal Society of Photogrammetry*, ASP-ACSM Convention, Washington D.C., February 22-27, Falls Church, pp. 440-449.

[25] Buckley, J.J. (1985). "Fuzzy hierarchical analysis," *Fuzzy Sets and Systems*, 17 (1985) 233-247.

[26] Burrough, P.A. (1986). *Principles of Geographic Information Systems for Land Resources Assessment*, Clarendon Press, Oxford.

[27] Castelaz, P.; J. Angus; J. Mahoney (1987). "Application of Neural Networks to Expert Systems and Command and Control Systems," *Proceedings of the Western Conference on Expert Systems*, WESTEX-87, June 2-4, Anaheim California, pp. 118-125.

[28] Caudil, M. (1990). "Using neural nets: Fuzzy decisions," *AI Expert*, April, pp. 59-64.

[29] Cermak, R.J.; A. Feldman; R.P. Webb (1979). Hydrological landuse classification using Landsat," In: M. Deutch; D.R. Wiesnet; A. Rango (Eds.) *Satellite Hydrology, Proceedings of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing*, American Water Resources Association, Sioux Falls South Dakota, June 10 - 15, pp. 262-269.

[30] Cerone, N. (1980). "The representation and use of knowledge in an associative network for the automatic comprehension of natural language," In: L. Bolc (Ed.) *Representation and Processing of Natural Language*, Carl Hanser Verlag, München Wien, pp. 121-205.

[31] Chen, P.P. (1978). "Application of the entity relationship model," In: S.B. Yao; S.B. Navathe; J.L. Weldon; T.L. Kunii(Eds.) *Data Base Design Techniques I: Requirements and Logical Structures*, Proceedings, New York, May, Springer-Verlag, Berlin, Heidelberg, N.Y. 1982, pp. 87-113.

[32] Chen, S. (1985a). "Ranking fuzzy numbers with maximizing set and minimizing set," *Fuzzy sets and systems*, Vol. 17, No.2, pp. 113-129.

[33] Chen, W. L. (1985b). "Fuzzy match and floating threshold strategy for expert system in traditional Chinese medicine," *Fuzzy sets and systems*, Vol. 17, No. 1, pp. 143-151.

[34] Civanlar, M.R.; H.J. Trusell (1986). "Constructing membership functions using statistical data",*In Fuzzy Sets and Systems*, 18 (1986) 1-13.

[35] Cohen P.R.; E.A. Feigenbaum (1982). *The Handbook of Artificial Intelligence*, Vol. III, Addison-Wesley Publishing Company, Inc., Reading, Massachussests, Menlo Park.

[36] Cohen, P. R. (1985). *Heuristic Reasoning about Uncertainity: An Artificial Intelligence Approach*, Morgan Kaufmann Publishers, Inc., Los Altos, California.

[37] Corr, D. G.; A.M. Tailor; A. Cross; D.C. Hogg; D.H. Lawrence; D.C. Mason; M. Petrou (1989). "Progress in automatic analysis of multi-temporal remotely sensed data," *International Journal of Remote Sensing*, Vol. 10, No. 7, 1989, pp.1175-1195.

[38] Cuff, R. N. (1984). "HERCULES: Database query using natural language fragments," In: J. Longstaff (Ed.) *Proceedings of the Third British National Conference on Databases (BNCOD3)*, Leeds, 11-13 July, Cambridge University Press, London, NY., pp. 133-149.

[39] Dahlgren, K. (1988). *Naive Semantics for Natural Language Understanding*, Kluwer Academic Publishers, Boston, Dordrecht, London.

[40] Davis, R.; D. Lenat (1982). *Knowledge Based Systems in Artificial Intelligence*, McGraw-Hill International Book Co., NY.

[41] Dean, J.D.; W.M. Snyder (1977). "Temporally and areally distributed rainfall," *Journal of the Irrigation and Drainage Division*, IR2, June, pp. 221-229.

[42] Delahaye, J. P. (1987). *Formal Methods in Artificial Intelligence*, Translated by, J. Howett, North Oxford Academic Publishers Ltd., {Outils Logiques Pour l'Intelligence Artificielle}, Editions Eyrolles Paris.

[43] Dowsing, R.; V.J. Rayward-Smith; C.D. Walter (1986). *A First Course in Formal Logic and its Applications in Computer Science*, Blackwell Scientific Publications, Oxford, London, Edinburg.

[44] Doyle, J. (1987). "A truth maintenance system," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Ginsberg, M. L. (ed.), Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 259-279.

[45] Dubois D.; H. Prade. (1980). *Fuzzy sets and Systems: Theory and applications*, Academic Press, N.Y., London.

[46] Dubois, D.; H. Prade (1989). "Processing fuzzy temporal knowledge,"*IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 4, July/Aug, pp. 729-743.

[47] Dubois, D.; H. Prade (1990). "A discussion of uncertainty handling in support logic programming," *International Journal of Intelligent Systems*, Vol. 5, pp. 15-42, 1990.

[48] Dutka, A.F.; H.H. Hanson (1989) *Fundamentals of Data Normalization.*

[49] Ebrahimi, M. (1987). "A Structured approach to expert system design, "*Proceedings of the Western Conference on Expert Systems*, WESTEX-87, June 2-4, Anaheim, California, pp. 18-23.

[50] Egenhofer, M. J.; A.U. Frank (1990). "LOBSTER: combining AI and data base techniques for GIS," *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 6, pp. 919-926.

[51] ESRI, (1984). "A map of desertification hazards: Explanatory note," *Environmental Systems Research Institute (ESRI)*, United Nations Environmental Programme, The F.A.O. U.N.

[52] Etherington, D. W.; R. Reiter (1987). "On inheritance hierarchies with exceptions," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 101-105.

[53] Fahlman S. E. *NETL: A System for Representing and Using Real-world Knowledge*, The MIT Press, Cambridge, Mass., London.

[54] Fansworth, R.K.; R.P. Canterford (1980)."Satellite rain estimation for hydrologic forecasting," *Technical Papers of the* 46_{th} *Meeting of the American Society of Photogrammetry*, Washington D.C., pp. 97-105.

[55] Feldman, D. (1984). "A diagrammer for the automatic production of entity type models," In: J. Longstaff (Ed.) *Proceedings of the Third British National Conference on Databases (BNCOD3)*, Leeds, 11-13 July, Cambridge University Press, London, NY., pp. 57-70.

[56] Feldman, P.; G. Fitzgerald (1985). "Action modelling: A symmetry of data and behaviour modelling," In: A.F. Gundy (Ed.) *Proceedings of the Fourth British National Conference on Databases*, (BNCOD4). University of Keele, 10 - 12 July, Cambridge University Press, London, NY., pp. 81-104.

[57] Finin, T. (1988). "Default reasoning and stereotypes in user Modelling," *International Journal of Expert Systems, Research, and Applications*, Vol 1, No. 2, pp. 131-158.

[58] Fisher, D.; P. Langley (1986). "Conceptual clustering and its relations to numerical taxonomy," In: W.A. Gale (Ed.) *Artificial intelligence and statistics*, Addison-Wesley Publishing Company, Reading, Mass., pp. 77-116.

[59] Fleming, G. (1981). "Data requirements for sediment, erosion, and transport simulation," In: *Erosion and Sediment Transport Measurement Symposium* Proceedings of the Florence Symposium, 22-26 June, IAHS Publication No.133, pp. 321-327.

[60] Foster G.R.; W.H. Wischmeier (1974). "Evaluating irregular slopes for soil loss prediction" *The Transactions of the ASAE*, pp.305-309

[61] Foster, G.R.; L.S. Lane; J.D. Nowlin; J.M. Laflen; A. Young (1981). "Estimating erosion and sedment yield on field sized areas," *Transactions of the ASAE*, Vol. 24, pp.1253-1262.

[62] Franklin, W. R.; P.Y.F. Wu; S. Samaddar; M. Nichols (1986). "Prolog and geometry projects," *IEEE CG&A*, Vol. 6, Nov., pp. 46-55

[63] Gale, W. A. (1986). *Artificial intelligence and statistics*, Addison-Wesley Publishing Company, Reading, Mass., 1986.

[64] Giles, R. (1980). "A computer program for fuzzy reasoning," *Fuzzy Sets and Systems*, 4 (1980) 221-234.

[65] Goguen, J. (1979). "Fuzzy sets and the social nature of truth," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp. 49-67.

[66] Goldman, S.J.; K. Jackson; T.A. Bursztynsky (1986). *Erosion and Sediment Control Handbook*, Mc Graw Hill Book Company, N.Y.

[67] Goodenough, D.G. (1987). "An Expert system for remote sensing,"In: *IEEE Transactions on Geoscience and Remote Sensing*, Vol GE-25 No.3, May, pp. 349-359

[68] Goodenough, D. G. "Thematic Mapper and SPOT integration with a geographic information system," *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 2, Feb. 1988, pp. 167-176.

[69] Goosen, D. (1964). "The use of aerial photography for the development of soil resources in South America," *Photogrammetria* Commision VII, Lisbon, pp. 104-

[70] Green, C. J. R.; M.M. Keyes (1987). "Verification and validation of expert systems," *Proceedings of the Western Conference on Expert Systems*, WESTEX-87, June 2-4, Anaheim, California, pp. 38-43.

[71] Grundy, A.F. (ed.) (1985). *Proceedings of the Fourth British National Conference on Databases (BNCOD4)*, University of Keele, 10 - 12 July, Cambridge University Press, London, N.Y.

[72] Habel U.; C-R. Rollinger; A. Schmidt; H-J. Schneider (1980). "A logic-oriented approach to automatic text understanding," In: L. Bolc (Ed.) *Natural Language Based Computer Systems*, Carl Hanser Verlag, München Wien, pp. 57-117.

[73] Hadipriono, F. C.; J.G. Lyon; L.W.H. Thomas (1990). "The development of a knowledge based expert system for analysis of drainage patterns," *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 6, (1990), pp. 905-909.

[74] von Hahn, M.; W. Hoeppner; A. Jameson; W. Wahlster (1980). "The anatomy of the natural language dialogue system HAM-RPM," In: L. Bolc (Ed.) *Natural Language Based Computer Systems*, Carl Hanser Verlag, München Wien, pp. 119-253.

[75] Hart, J. A.; D.B. Wherry; S. Bain (1985). "An operational GIS for Flathead national forest," *Proceedings of Autocarto 7*. ASP and ACSM, Washington D.C., March 11-14, pp. 244-253.

[76] Heilman, J.L; D.G. Moore (1979) "Soil moisture applications of heat capacity mapping mission," In: M. Deutch; D.R. Wiesnet; A. Rango (Eds.) *Satellite Hydrology, Proceedings of the Fifth Annual William T. Pecora*

Memorial Symposium on Remote Sensing, American WAter Resources Association, Sioux Falls South Dakota, June 10 - 15, pp. 371-376.

[77] Hellwig, P. (1980). "PLAIN - A program system for dependency analysis and for simulating natural language inference," In: L. Bolc (Ed.) *Representation and Processing of Natural Language*, Carl Hanser Verlag, München Wien, pp. 271-376.

[78] Hóly, M. (1980) *Erosion Environment*, Environmental Sciences and Applications, A. Biswas; M.R. Biswas (Eds.), Vol. 9, Translated by, J. Ondráčková, Pergamon Press Ltd., Oxford, NY.

[79] Hotaka, R. (1978) "The design of an integrated data dictionary directory system," In: S.B. Yao; S.B. Navathe; J.L. Weldon; T.L. Kunii (Eds.) *Data Base Design Techniques I: Requirements and Logical Structures*, Proceedings, New York, May, Springer-Verlag, Berlin, Heidelberg, N.Y. 1982, pp. 56-71.

[80] Hubbard, G.U. (1978) "Technique for automated logical database design," In: S.B. Yao; S.B. Navathe; J.L. Weldon; T.L. Kunii (Eds.) *Data Base Design Techniques I: Requirements and Logical Structures*, Proceedings, New York, May, Springer-Verlag, Berlin, Heidelberg, N.Y. 1982, pp. 219-227.

[81] Jiang, Y.J.; S.H. Lavington (1985). "The qualified binary relationship model of information," In: A.F. Gundy (Ed.) *Proceedings of the Fourth British National Conference on Databases (BNCOD4)*, University of Keele, 10 - 12 July, Cambridge University Press, London, NY., pp. 61-79

[82] Kahn, B. K. (1978) "LDDM–A structured logical database design methodology," In: S.B. Yao; S.B. Navathe; J.L. Weldon; T.L. Kunii (Eds.) *Data Base Design Techniques I: Requirements and Logical Structures*, Proceedings, New York, May, Springer-Verlag, Berlin, Heidelberg, N.Y. 1982, pp.31-55.

[83] Kambayashi, Y; K. Tanaka; S. Yajima (1978). "Problems of relational database design," In: S.B. Yao; S.B. Navathe; J.L. Weldon; T.L. Kunii (Eds.) *Data Base Design Techniques I: Requirements and Logical Structures*, Proceedings, New York, May, Springer-Verlag, Berlin, Heidelberg, N.Y. 1982, pp. 172-214.

[84] Kandel A. (1979). "On fuzzy statitstics," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 181-199.

[85] Kandel, A. (1986). *Fuzzy Mathematical Techniques with Applications*, Adison-Wesley Publishing Company, Reading, Massachussets.

[86] Kass, R.;T.Finin (1988). "The need for user models in generating expert system explanations,"International Journal of Expert Systems, Research, and Applications, Vol 1, No. 4, pp 345-375.

[87] Kaufmann, A. (1975). "Fuzzy graphs and fuzzy relations," In: A. Kaufman *Introduction to the Theory of Fuzzy Subsets: Fundamental Theoretical Elements*, Vol. I., Academic Press, N.Y., San Fransisco, London, pp. 41-89.

[88] Kickert, W. J. M. (1979). "An example of linguistic modelling: the case of Mulder's theory of power," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 519-540.

[89] De Kleer, J. (1987). "An assumption based TMS," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 280-297.

[90] Klahr, P.; L. Travis; C. Kellog (1980). "A deductive system for natural-language question answering," In: L. Bolc (Ed.) *Natural Language Answering Systems*, Carl Hanser Verlag, München Wien, pp. 73-136.

[91] Klir, G.J.; T.A. Folger (1988). *Fuzzy Sets, Uncertainty and Information*, Prentice Hall, Englewood Cliffs, N.J.

[92] Kolář, V. (1977). "Recconnaissance measurements and sedment estimates in basins with insufficient data," In: *Erosion and Solid Matter Transportation in Inland Waters Symposium*, Proceedings of the Paris Symposium, IAHS-AISH Publication No. 122, July, Paris, pp. 3-10.

[93] Kowalski, R. (1984). "Logic as a database language," In: J. Longstaff (Ed.) *Proceedings of the Third British National Conference on Databases (BNCOD3)*, Leeds, 11-13 July, Cambridge University Press, London, NY., pp. 103-132.

[94] Kretsch, J. L.; E.M. Mikhail (1990). "Decision considerations arising from error prapagation through belief calculations," *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 6, (1990) pp. 927 - 931.

[95] Lal, R. (1981). "Analyses of different processes governing soil erosion by water in the tropics," In: *Erosion and Sediment Transport Measurement Symposium*, Proceedings of the Florence Symposium, 22-26 June, IAHS Publication No.133, pp. 361-364.

[96] Lal, V. B.; J. Banerji, J. Narayan (1977). "Sediment yield in relation to drainage basin characteristics in some Indian river valley projects," In: *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July, pp. 93-97.

[97] Lillesand, T.M.; R.W. Kiefer (1987). *Remote Sensing and Image Interpretation*, Second Ed.; John Willey & Sons; NY, Chichester, Brisbane.

[98] Logan, T.J.; D.R. Urban; J.R. Adams; S.M. Yaksich (1982). "Erosion control potential with conservation tillage in the Lake Erie basin: Estimates using the universal soil loss equation and the Land Resource Information System (LRIS)," *Journal of Soil and Water Conservation*, Vol. 37, No. 1, (1982), pp. 50-55.

[99] Loizou, G.; P. Thanisch (1984). "Do embedded dependencies allways lead to wild goose chase?," In: J. Longstaff (Ed.) *Proceedings of the Third British National Conference on Databases (BNCOD3)*, Leeds, 11-13 July, Cambridge University Press, London, N.Y., pp. 41-56.

[100] Longstaff, J. (ed.) (1984). *Proceedings of the Third British National Conference on Databases (BNCOD3)*, Leeds, 11-13 July, Cambridge University Press, London, N.Y.,

[101] Lowry, M. (1989). "Knowledge-based software engineering,"In: A. Barr; Cohen, P.R.; E.A. Feigenbaum (Eds.) *The Handbook of Artificial Intelligence*, Vol. IV. Addison-Wesley Publishing Company, Inc., Reading, pp.149-192.

[102] Luger, G. F.; W.A. Stubblefield (1989). *Artificial Intelligence and the Design of Expert Systems*, The Benjamin/Cummings Publishing Company, Inc., Redwood City, NY., 1989.

[103] De Luka, A.; S. Termini (1979). "Entropy and energy measures of a fuzzy set," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 321-338.

[104] L'vovich, M.I.; N.L. Bratseva; G. Ya Karasik; G.P. Medvedeva (1990). "A map of contemporary erosion of the surface of the earths surface," *Mapping Sciences and Remote Sensing*, Vol 27, No. 1, pp. 51-67.

[105] Lynn, E.J. (1989). "MAPHYD - A digital map-based hydrologic modelling system," *Photogrammetric Engineering and Remote Sensing*, Vol. 55, No. 6, June, pp.911-917.

[106] Mabuchi, S. (1988). "An approach to the comparison of fuzzy subsets with an α-cut dependent index," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 18, No. 2, pp. 264-272.

[107] Magrez, P.; P. Smelt (1989). "Fuzzy modus pones: A new model suitable for applications in knowledge-based systems," *International Journal of Intelligent Systems*. Vol. 4, pp. 181-200.

[108] Martin, N.J. (1984). "The construction of interfaces to triple based databases," In: J. Longstaff (Ed.) *Proceedings of the Third British National Conference on Databases (BNCOD3)*, Leeds, 11-13 July, Cambridge University Press, London, N.Y., pp. 151-172.

[109] Martin, J.; S. Oxman (1988). *Building Expert Systems: A Tutorial*, Prentice Hall, Englewood Cliffs, N.J.

[110] McCool D.K,; L.C. Brown; G.R. Foster; C.K. Mutchler; L.D. Meyer (1987). "Revised slope steepness factor for the universal soil loss equation" *Transacttions of the ASAE*, Vol. 30, No.5, October, pp. 1387-1395.

[111] McIsaac, G.F.; J.K. Mitchell; M.C. Hirschi (1987). "Slope steepness effects on soil loss from disturbed lands," *Transactions of the ASAE*, Vol. 30, No. 4, July - August, pp. 1005-1013.

[112] McKeown, D. M.; W.A. Harvey; L.E. Wixson (1989). "Automating knowledge acquiostion for aerial image interpretation," *Computer Vision, Graphics, and Image Processing*, Vol. 46, 1989, pp. 37-81.

[113] Mehldau, G. (1990). "A C - extension for rule-based image classification systems," *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No.6, pp. 887-892.

[114] Meyer, L.D. (1984). "Evolution of the universal soil loss equation," *Journal of Soil and Water Conservation*, March-April, pp. 99-104.

[115] De Michiel, L. G. (1989). "Resolving database incompatibility: An approach to performing relational operations over mismatched domains," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 1. No. 4. December, pp. 485-493.

[116] Michalski, R.; J.G. Carbonell; T.M. Mitchell (1986). "Understanding the nature of learning: issues and research directions," In: R.S. Michalski; J.G. Carbonell; T.M. Mitchell (Eds.) *Machine Learning: An Artificial Intelligence Approach*, Vol. II, Morgan Kaufmann Publishers, Inc., pp. 3-26.

[117] Mills W. C. (1986). "Estimating soil loss probabilities for Southern Piedmont cropping-tillage systems," *Transactions of the ASAE*, Vol.29 (4):July-August, pp. 948-954.

[118] Minker, J.; P. Powell (1980). "Answer extraction, natural language, and voice output for deductive relational data bases," In: L. Bolc (Ed.) *Natural Language Based Computer Systems*, Carl Hanser Verlag, München Wien, pp. 9-56.

[119] Minton, S. (1988). *Learning Search Control Knowledge: An Explanation-Based Approach*, Kluwer Academic Publishers, Boston, London.

[120] Mizumoto, M.; S. Fukami; K. Tanaka (1979). "Some methods of fuzzy reasoning," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 116-136.

[121] Mizumoto M., K. Tanaka (1979). "Some properties of fuzzy numbers," In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 153-164.

[122] Moller-Jensen, L. (1990). "Knowledge-based classification of urban area using texture and context information in Landsat-TM imagery," *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 6, (1990), pp. 899-904.

[123] Morgan, R.P.C. (1986). *Soil Erosion and its Control*, Van Nostrand Reinhold Company, N.Y.

[124] Morris-Jones, D.R.; R.W. Kiefer (1978) "Application of remote sensing to estimating soil erosion Potential" In: F. Shahrokhi (ed.), *Remote Sensing of Earth Resources*, Technical Papers Selected From the Seventh Annual Remote Sensing of Earth Resources Conference, Tullahoma, Tennessee, March 27-29, Vol. VII, pp. 159-186.

[125] Morrison, J. E.; S.H. Parker; A.C. Jones; R.R. Allen; D.E. Wilkins; G.M. Powell; R. Grisco; D.C. Erbach; L.P. Herndon; D.L. Murray; G.E. Formanek; D.L. Pfost; M.M. Herron; D.J. Baumert (1989). "Expert system for selecting conservation planting machines: 'PLANTING'," *Transactions of the ASAE*, Vol. 32, No. 2, March-April, pp. 397-401.

[126] Di Nola, A.; S. Sessa (1990). "Modus ponens for fuzzy data realized via equations with equality operators," *International Journal of Intelligent Systems*, Vol. 5, No. 1, (1990), pp. 1-14.

[127] Nowakowska, M. (1979). "Fuzzy concepts: Their structure and problems of measurement," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 361-387.

[128] Nii, H.P. (1989). "Blackboard systems,"In: A. Barr; Cohen, P.R.; E.A. Feigenbaum (Eds.) *The Handbook of Artificial Intelligence*, Vol IV. Addison-Wesley Publishing Company, Inc., Reading, pp. 1-82.

[129] Overmars, M.H. (1983). *The Design of Dynamic Data Structures*, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, NY.

[130] Parry, J. T.; J.A, Beswick (1973). "Application of two morphometric terrain-classification systems using aerial photo interpretation methods," *Photogrammetria*, Vol. 29, pp. 153-186.

[131] Pfaltz, J. L. (1977) *Computer Data Structures*. McGraw-Hill Book Company Inc.

[132] Piasecki, K. (1986). "On the Bayes formula for fuzzy prbability measures," *Fuzzy Sets and Systems*, 18 (1986) 183-185.

[133] Ralescu D. (1979). "A survey of the representation fuzzy concepts and its applications," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp. 77-91.

[134] Rapp, (1972). "Soil erosion and sedmentation in Tanzania and Lesotho," In: R.P.C. Morgan (Ed.) *Soil Erosion and its Control*, Van Nostrand Reinhold Company; NY., pp. 76-85.

[135] Reiter, R. (1987). "A Logic for default reasoning," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 68-93.

[136] Ringrose, S.; W. Matheson (1987). "Spectral assessment of indicators for range degradation in the Botswana Hardveld Environment," *Remote Sensing of the Environment*, Vol. 23, pp. 379-396.

[137] Ripple, W. J.; V.S. Ulshoefer (1987). "Expert Systems and Spatial Data Models for Efficient Geographic Data Handling," *Photogrammetric Engineering and Remote Sensing*, Vol 53, No. 10, pp. 1431-1433.

[138] Robinson, A. R. (1977). "Relationship between soil erosion and sediment delivery," In: *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July, pp. 159-167.

[139] Robinson, V. B.; A.V. Frank (1987). "Expert systems for Geographic Information Systems", *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 10, pp. 1435-1441.

[140] Robinson, G.B. (1988). "Remote sensing: The unheralded component of geographic information systems," *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No.2, Feb., pp. 195-199.

[141] Rolston, D. W. (1988). *Principles of Artificial Intelligence and Expert Systems Development*, McGraw-Hill Book Co., N.Y., St. Louis.

[142] Rosenberg, R. S. (1980). "Approaching discourse computationally: A review," In: L. Bolc (Ed.) *Representation and Processing of Natural Language*, Carl Hanser Verlag, München Wien, pp. 11-83.

[143] Round, A. (1989). "Knowledge-Based Simulation, "In: A. Barr; Cohen, P.R.; E.A. Feigenbaum (Eds.) *The Handbook of Artificial Intelligence*, Vol IV. Addison-Wesley Publishing Company, Inc. Reading, pp. 493-500.

[144] Samet, H. (1990a). *The Design and Analysis of Spatial Data Structures*, Addison-Wesley Publishing Co., Inc., Reading, Mass.

[145] Samet H. (1990b). *Applications of Spatial Data Structures : Computer Graphics, Image Processing, and GIS*, Addison-Wesley Publishing Co., Reading, Mass.

[146] Schenk, T.; O. Zilberstein (1990). "Experiments with a Rule-Based System for Interpreting Linear Map Features," *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 6, pp. 911 - 917.

[147] Schildt, H. (1987). *Artificial Intelligence Using C: The C Programmers Guide to AI Techniques*, Osborne McGraw-Hill, Berkeley, California, pp. 239-265.

[148] Schmidt, J.W.; V. Linnemann (1985). "Higher level relational objects," In: A.F. Gundy (Ed.) *Proceedings of the Fourth British National Conference on Databases (BNCOD4)*, University of Keele, 10-12 July, Cambridge University Press, London, N.Y., pp. 1-23.

[149] Schmugge, T. J.; T. J. Jackson; H. L. McKim (1979). "Survey of in-situ and remote sensing methods for soil moisture determination," In: M. Deutch; D.R. Wiesnet; A. Rango (Eds.) *Satellite Hydrology, Proceedings of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing*, American WAter Resources Association, Sioux Falls South Dakota, June 10-15 1979, pp. 333-352.

[150] Schmucker, K. J. (1984). *Fuzzy sets natural language computations, and risk analysis*, The Computer Science Press Inc.,1984.

[151] Schneider, S.R.; D.F. McGinnis; J.A. Pritchard (1979). "Delineation of drainage and physiographic features in North and South Dacota using NOAA-5 infrared data," In: M. Deutch; D.R. Wiesnet; A. Rango (Eds.) *Satellite Hydrology, Proceedings of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing*, American WAter Resources Association, Sioux Falls South Dakota, June 10-15, pp. 324-330.

[152] Schowengerdt, R. A.; H. Wang (1989). "A general purpose expert system for image processing," *Photogrammetric Engineering and Remote Sensing*, Vol. 55, No. 9, pp. 1277-1284.

[153] Schwartz, D. G. (1985). "The case for an interval-based representation of linguistic truth," *Fuzzy sets and systems*, Vol. 17, No. 2, pp. 153-165.

[154] Scoffield, R.A.; V.J. Oliver (1979). "A satellite derived technique for estimating rainfall from thunderstorms and hurricanes," In: M. Deutch; D.R. Wiesnet; A. Rango (Eds.) *Satellite Hydrology, Proceedings of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing*, American WAter Resources Association, Sioux Falls South Dakota, June 10-15, pp. 70-76.

[155] Shapiro, A. D. (1987). *Structured Induction in Expert Systems*, Addison-Wesley Publishing Co., Workingham, Reading.

[156] Shastri, L. (1989). "Default reasoning in semantic networks: A formalization of recognition and inheritance," *Artificial Intelligence*, Vol. 39, pp. 283-355.

[157] Simonett O.; F. Turyatunga; R. Witt, (1987). "Environmental database development for assessment of deforestation, soil erosion hazards and crop suitability: A joint Uganda-UNEP/GEMS/GRID Study," *Proceedings of the Second Annual International Conference, Exhibits and Workshops on Geographic Information Systems*, GIS 87-SAN FRANSISCO, ASPR and ACSM, University of California, Berkeley, pp.544-553.

[158] Smith, D.D; W.H. Wischmeier (1957). "Factors affecting sheet and rill erosion" *Transactions of the American Geophysical Union*, Vol. 38, No. 6, December, pp. 889-896.

[159] Smith, J. M.; D.C.P. Smith (1978). "Principles of Database conceptual design," In: S.B. Yao; S.B. Navathe; J.L. Weldon; T.L. Kunii (Eds.)

Database Design Techniques I: Requirements and Logical Structures, Proceedings, New York, May, Springer-Verlag, Berlin, Heidelberg, 1982, pp. 114-146.

[160] Sneddon, J.; T.A. Lutze (1989). "Close-range photogrammetric measurement of erosion in coarse-grained soils," *Photogrammetric Engineering and Remote Sensing*, Vol. 55, No. 5, May, pp. 597-600.

[161] Spanner, M. A. (1983). "The use of digital elevation model topographic data for soil erosion modeling within a geographic information system," *Technical Papers of the* 49_{th} *Annual Meeting of the American Society of Photogrammetry*, March 13 - 18, WAshington, D.C., pp. 314-321.

[162] Hon. Sparrow, H.O. (1984). *Soil at Risk: Canada's Eroding Future* Extract from the Minutes of Proceedings of the Senate of Tuesday, February 7, Standing Committee on Agriculture, Fisheries and Forestry, Committees and Private Legislation Branch., The Senate of Canada, Ottawa, Ont.

[163] Speight, J. G. (1977). "Landform pattern description from aerial photographs," *Photogrammetria*, 32, 1977, pp. 161-182.

[164] Srinivasan, A.; J.A. Richards (1990). "Knowledge-based techniques for multi-source classification," *International Journal of Remote Sensing*, Vol. 11, No. 3, 1990, pp. 505-525.

[165] Stephens P. R.; J.K. McMillan; J.L. Daigle; J. Cihlar (1985). "Estimating universal soil loss equation factor values with aerial photography", *Journal of soil and water conservation*, May June, pp. 293-296.

[166] Stevens, Al (1987). *C Data Base Development*, Management Information, Source, Inc., MIS Press.

[167] Stonebraker, M. (1989). "Future trends in database systems," *IEEE transactions on knowledge and data engineering*, Vol. 1. No. 1., March, pp. 33-40.

[168] Stricklen, J; J.W. Smith; B. Chandrasekar; J. R. Josephson (1987). "Modularity of domain knowledge," *International Journal of Expert Systems*, Vol. 1, No. 1, pp. 1-15.

[169] (SUN4) Sun Microsystems Inc. (1988). *C Programmers Guide*, Sun Microsystems Inc., 1987.

[170] Swann, R.; D. Hawking; A. W.-Roper, W. Johnston; McD. Detwiller (1988). "The potential for automated mapping from geocoded digital image data", *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 2, Feb., pp. 187-193.

[171] H. Tanaka, T. Okuda, K, Asai (1979a). "Fuzzy information and decision in statistical Model", In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Set Theory and Application*, North Holland Publishing Company, Amsterdam, pp. 303-320.

[172] Tanaka, H., T. Okuda; K. Asai (1979b). "Fuzzy information and decision in statistical model," In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, 1979, pp. 303-320.

[173] Tanimoto, S.T. (1987). *The Elements of Artificial Intelligence: An introduction using LISP*, Computer Science Press, Inc., Seatle Washington.

[174] Temple, P.H. (1972). "Measurement of runoff and soil erosion at an erosion plot scale with particular reference to Tanzania," In: Morgan, R.P.C. (ed) *Soil Erosion and its Control*, Van Nostrand Reinhold Company, N.Y. pp. 184-210.

[175] Touretzky, D.S. (1987). "An implicit ordering of defaults in inheritance systems," In: M.L. Ginsberg (Ed.) *Readings in Monotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 106-109.

[176] Turner, R. (1985). *Logics for artificial intelligence*, Ellis Horwood Limited, Sussex.

[177] Ullaby, F.T.; G.A. Bradley; M.C. Dobson (1979). "Potential applications of satellite radar to monitor soil moisture," In: M. Deutch; D.R. Wiesnet; A. Rango (Eds.) *Satellite Hydrology, Proceedings of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing*, American WAter Resources Association, Sioux Falls South Dakota, June 10-15, pp. 363-370.

[178] Ventura, S. (1988). *Dane County Soil Erosion Control Plan*, Dane County Land Conservation Committee, February, 1988, Dane County, Wisconsin.

[179] Walker, A. (ed.) (1987). *A Logical Approach to Expert Systems and Natural Language Processing: Knowledge Systems and Prolog*, Addison-Wesley Publishing Company, Inc., Reading, Mass.

[180] Walsh, S. J. (1985). "Geographic information systems for natural resources management," *Journal of soil and water conservation*, Vol. 40, No. 2, (1985). pp. 202-205.

[181] Wang, F. (1990a). "Fuzzy supervised classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 28, No. 2, (1990), pp. 194-201.

[182] Wang, F. (1990b) "Improving remote sensing image analysis through fuzzy information representation," *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 8, August, pp. 1163-1169.

[183] Webster, R.; I.F.T. Wong (1969). "A numerical procedure for testing soil boundaries interpreted from aerial photographs," *Photogrammetria*, Vol. 24, pp. 59-72.

[184] Webster R.; P.H.T. Beckett (1970). "Terrain classification and evaluation using aerial photography: A review of recent work at Oxford," *Photogrammetria*, (1970) pp. 51-74.

[185] Wenstøp, F. "Exploring linguistic consequences of assertions in social sciences," In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, 1979, pp 501-518.

[186] Whittington, R.P. (1988). *Database Systems Engineering*, Clarendon Press, Oxford.

[187] Williams, J.R. (1977). "Sediment delivery ratios determined with sediment and runoff models," In: *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July, pp. 168-179.

[188] Wischmeier W.H.; D.D. Smith (1957). "Soil-loss estimation as a tool in soil and water management planning", *International Association of Scientific Hydrology*, Commision of Land Erosion, Extract of Publication No. 59., pp. 148-158

[189] Wischmeier, W.H. (1984). "The USLE: Some reflections," *Journal of Soil and Water Conservation*, March-April, pp. 105-107.

[190] Yager, R.R. (1986). "A characterisation of the extension principle", *Fuzzy sets and systems* 18 (1986) pp. 205-217.

[191] Yamakawa, T. (1988). "Intrinsic fuzzy electronic circuits for sixth generation computer," In: M.M. Gupta; T. Yamakawa (Eds.) *Fuzzy Computing Theory, Hardware and Applications*, North-Holland, Amsterdam, pp. 157-171.

[192] Yao, S. B.; S.B. Navathe; J.L. Weldon; T.L. Kunii (Eds.) (1982). *Data Base Design Techniques I: Requirements and Logical Structures*, Proceedings, New York, May 1978 Springer-Verlag, Berlin, Heidelberg, N.Y.

[193] Yao, S.B.; S.B. Navathe (1978). "An integrated approach to database design," In: S.B. Yao; S.B. Navathe; J.L. Weldon; T.L. Kunii (Eds.) *Data Base Design Techniques I: Requirements and Logical Structures*, Proceedings, New York, May, Springer-Verlag, Berlin, Heidelberg, N.Y. 1982, pp. 1-30.

[194] Zadeh, L. A.; K-S. Fu; K. Tanaka; M. Shimura (1975), *Fuzzy sets and their applications to cognitive and decision processes*, Proceedings of the US-Japan Seminar on Fuzzy Sets and Their Applications, The University of California Berkeley, July 1-4 1974, Academic Press Inc., N.Y., San Francisco, London.

[195] Zadeh, L. (1979). "Fuzzy sets and information granularity,"In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 3-18.

[196] Zadeh, L. A., (1989). "Knowledge representation in fuzzy logic," *IEEE Transactions on knowledge and data engineering*, Vol. 1, No. 1, March, pp. 89-98.

[197] Zenner, R.B.R.C.; R.M.M. De Caluwe; E.E. Kerre (1985). "A new approach to information retrieval systems using fuzzy expressions," *Fuzzy sets and systems*, Vol. 17, No. 1, pp. 9-22.

Appendix I

Extracts of the Listing of the Compiled SLEMS DBMS schema

I/1

```
/* ------------------------ SLEM_SCHEMA ----------------------- */

/**************** SLEM SYSTEM directory structures ***************/

struct slem_system {
char expert [12];
char converser [12];
};

struct expert {
   char rule_base [12];
   char facts_base [12];
   char slem_objects [12];
   char learning_objects [12];
   char must_objects [12];
   char taboo_objects [12];
   char graph_base [12];
   char locations [12];
   char attribute [12];
   char q_base [12];
   char q_history [12];
   char query_solution [12];
   char dialog_base [12];
   char kne_eng [12];
   char reason [12];
```

```
    char expert_prog [12];
};

struct converser {
    char commands [12];
    char modifiers [12];
    char operators [12];
    char quantifiers [12];
    char word_dictionary [12];
    char slem_relations [12];
    char qualifiers [12];
    char parser [12];
    char nd_parse [12];
    char trans_tense [12];
    char trans_verb [12];
    char trans_focus [12];
};
```

I/2

```
/* -------------------- SLEM_SCHEMA ----------------------- */

/********** SLEM SYSTEM directory file  ASCII names ************/


char *dedirectory [] = {
   "RULE_BASE",
   "FACTS_BASE",
   "SLEM_OBJECTS",
   "ATTRIBUTE",
   "LEARNING_OBJECTS",
   "MUST_OBJECTS",
   "TABOO_OBJECTS",
   "GRAPH_BASE",
   "LOCATIONS",
   "Q_BASE",
   "Q_HISTORY",
   "QUERY_SOLUTION",
   "DIALOG_BASE",
   0
};

/************** SLEM SYSTEM directory file  types **************/

char filetype [] = "FFFFFFFFFFFFFXXXFFFFFFFXXXXXFXAA";
```

```
/*********** SLEM SYSTEM directory file display masks **********/

char *direlmask [] = {
   "-----------",
   "-----------",
   "-----------",
   "-----------",
   "-----------",
   "-----------",
};


/*************** SLEM SYSTEM file name strings ****************/

char *db_dirs [] = {
   "EXPERT",
   "CONVERSER",
   0
};
```

I/3

```
/************* SLEM SYSTEM data element ASCII names **********/

char *denames [] = {
   "RELATION_ID",
   "RULE_ID",
   "RULE_LOC",
   "RULE_SUBJECT",
   "RULE_DOMAIN",
   "RULE_PREMISE",
   "RULE_NAME",
   "RULE_TYPE",
   "RULE_PREDICATE",
   "RULE_EXPLAIN",
   "RULE_ACTION",
   "RULE_CERTAINITY",
   0
};

/*************** SLEM SYSTEM data element types ***************/

char eltype [] =
"ANAAAAAAAAANNANAAAANNNNANAAAANNNNANNNNAAAANAANNNNNNNADAAAAAAAANA
AADNNAAAAANANAAAAANANAAADNAANNNNNZZNANANANNNNNNNNNNNNNAANNNAAANAA
AAANNAANAANAAAANNAANNAAAANANNNNNADAAAADAAADAAAAANAAAAAAAAANNNNNN
AAAAAANNNANNNNNNN";
```

```
/*************** SLEM SYSTEM data element display masks *******/

char *elmask [] = {
   "____________",
   "___________",
   "___________",
   "_______________________________",
   "_______________________________",
   "____ [0-1] decimal%",
   "_ [0,1]",
   "_ [0,1] only",
};
```

I/4

```
/*************** SLEM SYSTEM file name strings ***************/

char *dbfiles [] = {
   "RULE_BASE",
   "FACTS_BASE",
   "SLEM_OBJECTS",
   "LEARNING_OBJECTS",
   "MUST_OBJECTS",
   "TABOO_OBJECTS",
   "GRAPH_BASE",
   "LOCATIONS",
   "ATTRIBUTE",
   "Q_BASE",
   "Q_HISTORY",
   "QUERY_SOLUTION",
   0
};

/* ----------------------- SLEM_SCHEMA -------------------- */

/************* SLEM_SYSTEM directory name lengths ***********/

int direllen [] = {
   11,11,11,11,11,11,11,11,11,11,11
};
```

```
/*************** SLEM_SYSTEM dir list pointer array *********/

int *dir_ele [] = {
   d_expert,
   d_converser,
   0
};

/************** SLEM_SYSTEM file content arrays ************/

int d_expert [] = {
   LEARNING_OBJECTS,
   MUST_OBJECTS,
   TABOO_OBJECTS,
   0
};

 I/5

/*************** SLEM_SYSTEM dir index arrays *************/

int xx1_expert [] = {
   LEARNING_OBJECTS,
   0
};
int xx2_expert [] = {
   MUST_OBJECTS,
```

```
    0
};
int xx3_expert [] = {
    TABOO_OBJECTS,
    0
};
int *xx_expert [] = {
    xx1_expert,
    xx2_expert,
    xx3_expert,
    0
};
```

```
/********* The pointer to index pointer arrays *************/

int **dirindex_ele [] = {
   xx_expert,
   xx_converser,
   0
};

/************ SLEM_SYSTEM data element lengths **************/

int ellen [] = {
1,11,30,30,30,30,30,30,30,30,30,4,11,30,11,30,30,
30,30,4,1,1,1,30,11,30,30,30,30,1,11,11,11,...
};
```

I/6

```
/************** SLEM_SYSTEM file content arrays ************/

int f_learning_objects [] = {
   SESSION_ID,
   SUBJECT,
   VERB,
   OBJECT,
   SUBJECT_DOMAIN,
   LEARN_STATUS,
   LOCATION,
   CERTAINITY,
```

```
    ATTRIBUTES,
    0
};
int f_must_objects [] = {
    MUST_ID,
    SUBJECT,
    VERB,
    OBJECT,
    SUBJECT_DOMAIN,
    LEARN_STATUS,
    LOCATION,
    CERTAINITY,
    ATTRIBUTES,
    0
};
int f_taboo_objects [] = {
    TABOO_ID,
    SUBJECT,
    VERB,
    OBJECT,
    SUBJECT_DOMAIN,
    LEARN_STATUS,
    LOCATION,
    CERTAINITY,
    ATTRIBUTES,
    0
};
```

I/7

```
/**************** SLEM_SYSTEM file index arrays *************/

int x1_learning_objects [] = {
   SESSION_ID,
   0
};
int x2_learning_objects [] = {
   SUBJECT,
   0
};
int x3_learning_objects [] = {
   VERB,
   0
};
int x4_learning_objects [] = {
   OBJECT,
   0
};
int x5_learning_objects [] = {
   SUBJECT_DOMAIN,
   0
};
int *x_learning_objects [] = {
   x1_learning_objects,
   x2_learning_objects,
   x3_learning_objects,
```

```
   x4_learning_objects,
   x5_learning_objects,
   0
};
int x1_must_objects [] = {
   MUST_ID,
   0
};
int x2_must_objects [] = {
   SUBJECT,
   0
};
int x3_must_objects [] = {
   VERB,
   0
};
int x4_must_objects [] = {
   OBJECT,
   0
};
int x5_must_objects [] = {
   SUBJECT_DOMAIN,
   0
};
```

I/8

```
/*************** SLEM_SYSTEM file index arrays *************/
```

```
int *x_must_objects [] = {
   x1_must_objects,
   x2_must_objects,
   x3_must_objects,
   x4_must_objects,
   x5_must_objects,
   0
};

int x1_taboo_objects [] = {
   TABOO_ID,
   0
};

int x2_taboo_objects [] = {
   SUBJECT,
   0
};

int x3_taboo_objects [] = {
   VERB,
   0
};

int x4_taboo_objects [] = {
   OBJECT,
   0
```

```
};

int x5_taboo_objects [] = {
   SUBJECT_DOMAIN,
   0
};

int *x_taboo_objects [] = {
   x1_taboo_objects,
   x2_taboo_objects,
   x3_taboo_objects,
   x4_taboo_objects,
   x5_taboo_objects,
   0
};

/************** The pointer to index pointer arrays ************/

int **index_ele [] = {
   x_learning_objects,
   x_must_objects,
   x_taboo_objects,
   0
};
```

Appendix II

Listing of the Source Code for the Main Modules of the EXPERT Subsystem

II/O

```
/*****************************************************************/
/* enter.c                                                       */
/* Source: Schildt, 1987.                                        */
/* Modifications: E.G.Mtalo, UNB, April and May 1989:            */
/*                                                               */
/*****************************************************************/
/* facilty for  entering  rules and facts into the  SLEMS Rule   */
/* Base. Facts consists of  assertions  about domain objects     */
/* Rules have the general form:                                  */
/*   CONCLUDE:  SUBJECT X                                        */
/*   IF :  ASSERTION Y1, AND                                     */
/*     ASSERTION Y2, AND                                         */
/*      ldots                                                    */
/*   The assertion part consists of the                          */
/*  <VERB> <OBJECT_PART>                                   */
/* sentential form.                                              */
/* The corresponding form of SLEMS RULES is:                     */
/* SLEM RULE:   <PREDICATE> <CONDITION>                   */
/* PREDICATE:   <SUBJECT_PART>                              */
/* CONDITION:   <VERB> <OBJECT_PART>                      */
/*****************************************************************/
/*****************************************************************/
```

II/1

```
#include"expt.h" /* header files for structs, global variables  */
                 /*   and standard library <include> files       */

int get_next(); /* routine for computing next stack position    */
extern subj_dom[81];      /* char global string for domain name. */
int try_rules();          /* the inference module                */
void clear_Rbase(),kb_talk(),err_msg(); /* utility routines     */

/-------------------------enter.c-------------------------------*/

enter(subj_dom)
char *subj_dom;
{
   char spik = 's';
   char ob[80];       /* for the object name                    */
   char attrib[80];   /* hold attributes                        */
   int ch;            /* hold key board character input         */
   char temp_dom[81];  /* hold domain name                      */
   char buf[8];       /* hold input                             */
   int t,i;           /* integer counters                       */
   struct attributes *p,*oldp; /* pointers to attribute struct  */
   static char EXPERTS[] = "enter: ";/* static prompt string    */

   getchar();       /* flush input channel                      */
                 /* display input instructions                  */
   fputs("|     SLEMS KNOWLEDGE ENTRY MODULE |\n",
```

```
stdout);
fputs("| The SLEMS knowledge consists  of <FACT>
<ATTRIBUTE>  |\n", stdout);
fputs("| objects. The ATTRIBUTE part is a list of atributes of
|\n", stdout);
fputs("| the object or concept being introduced.|\n",stdout);
fputs("| The module prompts for each required component  until
|\n",stdout);
fputs("| the user enters a blank. It then prompts for the next
|\n",stdout);
fputs("| piece of knowledge, or exits if the
previous prompt|\n",stdout);
fputs("| was also responded to by a blank.|\n",stdout);
fputs("hit <RTN> to continue...\n",stdout);
getchar();
fputs("| please set options: N[ew], <RTN> O[ld]:|/
n",stdout);
fputs("| "N" clears the stacks, "O" adds to stack  |/
n",stdout);
gets(buf);
if (buf[0] == 'N' || buf[0] == 'n')  {
   fputs("|   ARE YOU SURE??  |\n", stdout);
   gets(buf);
   if (buf[0] == 'y' || buf[0] == 'Y')
      l_pos = -1; /* overwrite current stack                    */
   }
```

```
fputs("|   enter name of domain of interest:  |\n",stdout);
gets(temp_dom);
if (!*temp_dom) {
   fputs("| assume all domains? <RTN> or choice: |/
   n",stdout);
   gets(temp_dom);
   if (*temp_dom) strcpy (subj_dom, temp_dom);
               /* set global domain to current                 */
   else strcpy(subj_dom, "ALL");/* global to all domains      */
   subj_dom[strlen(subj_dom)] = '\0';/* null char!             */
   }
   do {
      t = get_next();/* next position on rule stack            */
      fprintf(stdout, "Domain  %s: ID:  %d\n",subj_dom,t);
                     /* display index of next available space
                                  for current rule              */
      if (t == -1) {   /* no more space in the rule base       */
         err_msg(EXPERTS,NO_MEM);
          /* no memmory: post an error message                  */
         return;
         }
   fputs("| please enter the object name |\n",stdout);
                     /* object = rule                          */
   gets(K_base[t].name);/* get rule conclusion                 */
   if (!K_base[t].name[0]) {/* not entered? verify exit        */
      fputs("| abort session? <RTN> or
      name:|\n",stdout);
      gets(K_base[t].name);
```

```
                    /* prompt for rule conclusion                  */
        }
if (!*K_base[t].name) {       /* invalid object entry              */
   l_pos--;          /* reset to previous stack position           */
   fputs("| please choose [S] to save anyrules:|\n",stdout);
   return;
}
strcpy(K_base[t].object_domain, subj_dom); /* copy domain name
                        into active memory                          */
p = (struct attributes *) malloc (sizeof(at));/* memory for
                                  attribute list                    */
if (p == '\0') {              /* memory request failed             */
   err_msg(EXPERTS,NO_MEM);
                    /* no memmory: post a message                  */
   return;
}
K_base[t].alist = p;    /* set the active memory premise
                        listpointer to the allocated pointer        */
                    /*initialize  with blanks                       */
for (i=0; i < sizeof(p->attrib);i++) p->attrib[i] = ' ';
for (i=0; i < sizeof(p->objct); i++) p->objct[i] = ' ';
          /* prompt for attributes: quit if not given              */
do {                      /* prompt for attributes                 */
```

```
fputs("| enter object attributes: |\n",stdout);
gets(p->attrib);
if (!*p->attrib) break; /* no premise entered? end       */
puts(p->attrib);
fputs("| uncertainity for this attribute: range [0-1]: |/
n",stderr);
gets(buf);
if (!*buf) {   /*user selected default value? verify       */
   fputs("| assume default? <RTN> or
   choice:|\n",stdout);
   gets(buf);
   }
if (*buf)      /* user sets the uncertainty handling       */
   sscanf(buf, "%f", &p->prob);
else p->prob = 0.5;
while (p->prob < 0  || p->prob > 1) {   /*
                              get correct certainty factors!  */
   fputs("\n",stdout);
   fputs("| invalid probability: choose [0-1] range
   only: |/n",stdout);
   gets(buf);       /* get the certainty factor for premise*/
   if (*buf)
      sscanf(buf, "%f",&p->prob);
   else p->prob = 0.5;
   }                /* end while incorrect certainty input */
if (*ob)           /* A valid SLEMS rule or object entered */
   sprintf(p->objct,"%s", K_base[t].name);
oldp = p;          /* attribute pointer to the new object  */
```

```
      p->next = (struct attributes *) malloc (sizeof(at));
      if (p->next == '\0') {      /* malloc has no more memory */
         err_msg(EXPERTS,NO_MEM);
                    /* insufficient memory: post an error message*/
         return;
         }
         p = p->next;
         p->next = '\0';                    /* append null char */
         for (i = 0; i< sizeof(p->attrib); i++)
            p->attrib[i] = ' ';
                 /* fill the rest of the record space with blanks */
         fputs("| enter next attribute? <y> <blank> to skip
         |\n",stdout);
       } while (TRUE);
      oldp->next = '\0';
      K_base[t].id = t;
      fputs("\n",stdout);

      fputs("| enter next object? <y> <blank>
      to quit|/n",stdout);
   } while (TRUE);
   return;
}
```

II/2

```
/*--------------------------load_rules.c----------------------*/

/*****************************************************************/
/* load_rules.c                                                  */
/* Source E.G. Mtalo, UNB, 1989; After load.c by H. Schildt      */
/*,1987                                                          */
/* facilty for down-loading rules from the SLEMS RULE_BASE       */
/*****************************************************************/

load_rules(subj_dom)
char *subj_dom;
{
   char spik = 's';
   int t,x;
   struct attributes *p, *oldp;/* pointer to attribute struct   */
   FILE *filep;   /* file descriptor for opening rule file      */
   int dbase;   /* integer descrip[tor                           */
   char tmp[81];   /* hold string                                */
   int a_pos = 0;   /* position attributes stack                 */
   static char   EXPERTS[] = "RULES: ";/* default prompt         */
   fputs("\n Enter File Name to load\n>",stderr);/* prompt
                        for subject area or filename             */
   gets(filenam);
   if (!*filenam) {                    /* not given             */
      fputs ("load default (domain_rules)? choice,
```

```
   <RTN>confirm/n>",stderr);
   gets(filenam);          /* get subject area or filename    */
   if (!*filenam) strcpy(filenam,subj_dom);  /*  default
                        is current domain filename            */
}
if ((filep = fopen (filenam,"r")) == 0) {
   err_msg(EXPERTS,CANT_LOAD);
                /* report file opening unsuccessful          */
   return;
}
kb_talk(EXPERTS,KB_SAY_LOA,spik);
                /* report file successfully opened           */
                 *free any old lists in the active memmory   */
clear_Rbase();            /* clear the active memory stacks  */
```

```
for (t = 0; t < DBSZMAX; ++t) {
   fscanf(filep,"%s\n",RULES[t].object_domain);
   if (RULES[t].object_domain[0] == '\0') break;
   fscanf(filep, "%s\n", RULES[t].name);
   if (!RULES[t].name[0]) break;
   RULES[t].alist = (struct attributes *)malloc(sizeof(at));
   p = RULES[t].alist;
   if (!p) {
      err_msg(EXPERTS,NO_MEM);
      return;
      }
   for (;;) {
      fscanf(p->attrib,sizeof (p->attrib),filep);
      if (*p->attrib)
         fprintf (stderr, "\n IF : %s \n AND
         \n",p->attrib);
      else break;
      if (!p->attrib[0]) {
         oldp->next = '\0';
         break;
         }
      fscanf (filep,"%f\n", &p->prob);
      p->next = (struct attributes *)malloc(sizeof(at));
      if (!p->next) {
         break;
      }
   oldp = p;
   p = p->next;
```

```
        }
    }
    fclose (filep);
    rul_pos = t - 1;
}
```

```
/---------------------------query_rules.c---------------------*/

/*****************************************************************/
/* query_rules.c                                                 */
/* facility for querying the SLEMS rule base                     */
/* Source: E.G. Mtalo, UNB, 1989; After query.c by H.Schildt,    */
/* 1987                                                          */
/* Modified to allow for selection of uncertainity propagation   */
/* formulae                                                      */
/* and examination of only a restricted chunk of Knowledge Base */
/* rules.                                                        */
/*****************************************************************/


query_rules(subjdom)
char *subjdom;
{
   int t;
   static int all_dom = 0;/* if all domains are to be searched */
   static int found = 0;  /* if a specified domain is found    */
   float *p1;
   char buf[8];
   char spik = 's';
   int  ch, in;
   char *ch1;
   char temp_dom[81];                          /* domain name      */
```

```
struct attributes *p;            /* pointer to active objects  */
static charEXPERTS[] = "EXPERT: ";

p1 = &p_cut;                 /* pointer to current threshhold   */
ch1 = &ptyp;            /* current uncertainty management method*/
if (!*subj_dom) {
   puts("   Enter Domain of Interest or <RTN> [ALL]:");
   gets(temp_dom);                /* get domain name           */
 }
if (!*temp_dom && !*subjdom) {
   puts("assume default? <RTN> or choice:");
   gets(temp_dom);
   if (*temp_dom) strcpy (subjdom, temp_dom);
   else all_dom = TRUE;         /* default to all if not given */
}
fprintf(stderr, "Using [%s] RULES\n",subjdom);
kb_talk(EXPERTS,UNCERT_TYP,spik); /* prompt uncertainty type  */
```

```
gets(buf);                    /* get uncertainty handling method*/
if (!*buf) {                 /* not given? use default          */
   puts("assume deafult? <RTN> or choice:");
   gets(buf);
}
if (*buf) *ch1 = isupper(buf[0]) ? tolower(buf[0]) : buf[0];
else *ch1 = 'c';
                    /* default uncertainty handling is classical*/
kb_talk(EXPERTS,UNCERT_THR,spik);  /* prompt threshold level  */
printf("\nClassical probability and 0.5 level will
be used if none picked\n>");
gets(buf);                   /* get threshold setting           */
if (*buf) sscanf(buf, "%f",p1);
else *p1 = 0.5;                  /* user accepts 50% default   */
printf ("\n successful setting of uncertainty cotrol
mechanism\n");
for (t = 0; t < rul_pos; t++) {
   if (!all_dom){/* for specific domain verify rule domain    */
      if (strcmp(RULES[t].object_domain, subjdom)) {
                                 /* no match! skip it          */
      continue;
      }
else {
    p = RULES[t].alist;                      /* specific rules  */
   }
}
else p = RULES[t].alist;          /* if no domain given use all*/
if (ptyp){                   /* uncertainty option              */
```

```
   switch (ptyp) {          /* set appropriate method          */
      case 's':
         prob = 0.0;      /* set up for the most compelling
                            argument or stongest link method
                             see Klir,1988,                     */
      break;
      default:
         prob = 1.0;             /* same initial setting for
                                 weakest link method            */
      break;
      }
                 /* set up classic odds else the weakest link */
   }
else {
   prob = 1.0;         /* same initial setting as for classical*/
   }
```

```
                    /* try_rules is the main inference engine!   */
if (try_rules(p,RULES[t])) {   /* try a rule in the rule base */
                    /* the rule satisfies the query              */
   fprintf(stdout, "\n| By domain [%s] rule no [%d] \n the
   conclusion is [%s] ",RULES[t].object_domain,     /* domain */
   RULES[t].id, RULES[t].name);
                    /* rule conclusion = object name             */
   fputs("\n| the current facts fit the rule with
   a \n",stdout);
    fprintf(stdout, "| certainity factor of%5.2
   f%%",100*prob);    /* express uncert. in percent             */
                           /* prob is global, and propagated  */
   fputs("\n| EXPERT: continue ?\n",stdout);
                       /* more solutions on the way!            */
   gets(buf);

   ch = isupper(buf[0]) ? tolower(buf[0]) : buf[0];

   if (!ch) {        /* if solution is good user must say so!*/
      fputs("\n",stdout);
      fputs("| Please I need an answer!|\n", stdout);
      gets(buf);
      ch = isupper(buf[0]) ? tolower(buf[0]) : buf[0];
      }
      if (ch == 'n'){     /* not interested in more solutions*/
         return;
      }
   }
```

```
    }
    kb_talk(EXPERTS,KB_SAY_END,spik);    /* query is finished    */
}
```

```
/*--------------------------reasons.c-------------------------*/

/*****************************************************************/
/* reasons.c                                                     */
/* Source: H Schildt, 1987 "Artificial Intelligence in C"        */
/* Modified by: E.G. Mtalo, UNB, 1989,                           */
/* Facility for explaining reasons for specific user prompts     */
/* Modifications done to enhance the simple explanation          */
/* algorithm in the original source                              */
/*****************************************************************/

reasons (obj)
struct object *obj;
{
   char spik = 's';
   struct attributes *t;
   int i;
   char ans[4];
   static char   EXPERTS[] = "EXPERT: ";
   fprintf(stdout,"|I was trying [%s] ", obj->name);
   if (valid)            /* belongs to valid premises            */
      fputs("\n| these are its attribute(s) as validated: ",
      stdout);
   t = valid;              /* remember this was found to be valid*/
   while(t) {    /* repeat as long as valid attributes are found */
```

```
   fprintf (stdout, "\n| [%s] \n| is a valid attribute of
   [%s] \n|",t->attrib, obj->name);
   t = t->next;                    /* check the next attribute  */
   }
if (invalid)                        /* report attribute invalid  */
   fprintf(stdout,"\n| and these were determined invalid/
   n|");
t = invalid;                    /* remember attribute was invalid */
while (t) {
   fprintf (stdout, "\n| [%s] ",t->attrib);
                          /* display all the matches            */
   t  = t->next;
   }
if (r_base[0].name[0]) {
   fputs("\n|\n| these trials were rejected as
   indicated:\n|", stdout);
   for (i = 0; i < r_pos; i++) {
      fprintf(stdout, "\n| [%s] %s ",r_base[i].name,
      r_base[i].name[0] ? " rejected because " : "");
      if (r_base[i].condition == 'n') {      /* invalid
                                              attributes         */
         fprintf (stdout, "\n| [%s] ",r_base[i].attrib);
         fprintf(stdout,"\n| attribute was validated by
          neither %s,\n| %s nor %s
          \n|",by_who[BY_OBSERV],
          by_who[BY_LEARNER],by_who[BY_USER]);
      }
      else if (r_base[i].condition == 'y')
```

```
            fprintf(stdout, "| [%s] \n| %s \n|
            [%f%%]according to %s", r_base[i].attrib,
            "attribute is  required but has low
                    certainity",r_base[i].prob,
             by_who[r_base[i].by]);
         else {                    /* required attributes        */
            fprintf(stdout, "| [%s]   ",r_base[i].attrib);
            fputs("\n| has low possibility ", stdout);
         }
      }
   }
   fputs(" hit <RTN> to continue...", stdout);
   getchar();
   return;
}
```

Bibliography

[1] S. B. Ahuja; W-Y Soh; A. Schwartz (1989). "A connectionist processing metaphor for the diagnostic reasoning," *International Journal of Intelligent Systems*, Vol. 4, pp. 155-180.

[2] Akinyokun, O.C.; P.M. Stocker; M.R.S. Borges (1985). "Bidirectional mapping between a user oriented conceptual schema and a target logical schema: the ACS," In: A.F. Gundy (Ed.)*Proceedings of the Fourth British National Conference on Databases (BNCOD4)*, University of Keele, 10 - 12 July, Cambridge University Press, London, N.Y., pp. 105-130.

[3] Allen, J. (1987). *Natural language understanding*, Benjamin/Cummings Publishing Company, Inc., Menlo Park, California.

[4] Ammar, R.A.; T.L. Booth (1988). "Software optimization using user models," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 18, No. 4, July/Aug., pp. 552-560.

[5] Anderson, J.A. (1990). "Data representation in neural networks," AI EXPERT, June, pp. 30-37.

[6] Anderson, J.R. (1986). "Knowledge compilation: The general learning mechanism," In: R.S. Michalski; J.G. Carbonell; T.M. Mitchell (Eds.) *Machine Learning: An artificial intelligence approach*, Vol. II, Morgan Kaufmann Publishers, Inc., pp. 289-310.

[7] Ateshian, K.H. (1974). "Estimation of rainfall erosion index," *Journal of the Irrigation and Drainage Division*, IR3, Sept., pp. 293-307.

[8] Astroth, J. H. (1990) "A retrospective analysis of GIS performance: The Umatilla Basin Revisited," *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 3, pp. 359-363.

[9] Aydin, K. (1987). "Dual polarization radar estimation of rainfall parameters compared with ground based disdrometer measurements," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-25, No 6, pp****

[10] Bachman, C. W. (1989). "A personal chronicle: Creating better information systems, with some guiding principles," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 1, No. 1, pp. 17-32.

[11] Backer, E.; J.J. Gerbrands; J.H.C.Reiber; A.E.M. Reijs; W. Krijgsman; H.J. Van Dem Henrick (1987). "Modelling uncertainity in ESATS by classification inference: An expert system for the analysis of thallium-201 scintigrams," *Pattern recognition letters*, Vol. 8, No. 2, pp. 113-121.

[12] Barber, T.J.; G. Marshall; J.T. Boardman, (1988). "Tutorial - philosophy and archtecture for a rule-based frame system: RBFS," *Engineering Application of AI*, Vol. 1, June, pp. 67-86.

[13] Barr, A.; P.R. Cohen; E.A. Feigenbaum (1982). "Generalization Methods", and "Learning Programs." In: Cohen, P.R.; E.A. Feigenbaum (Eds.) *Handbook of Artificial Intelligence*, Vol. III, Addison-Wesley Publishing Company, Inc., NY., pp. 360-491.

[14] Barr, A.; P.R. Cohen; E.A. Feigenbaum (1989). "qualitative calculus" In: A. Barr; Cohen, P.R.; E.A. Feigenbaum (Eds.) *The Handbook of artificial intelligence*, Vol. 4, Addison-Wesley Publishing Company, Inc., pp. 339-350.

[15] Basu, A.; A. Dutta (1989). "Reasoning with imprecise knowledge to enhance intelligent decision support," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 4, pp. 756-770.

[16] Bentley, J.L.; J.H. Friedman (1979) "Data structures for range searching," *Computing Surveys*, Vol. II, No.4, December, pp. 397-409.

[17] Bergano, F.; A. Giordana (1989). "Concept recognition:An approximate reasoning frame work," *International Journal of Intelligent Systems*, Vol. 4, pp. 23-44.

[18] Berry-Rogghe, G. L.; M. Colvenbach; H-D. Lutz (1980). "Interacting with PLIDIS a deductive question answering system for german," In: L. Bolc (Ed.) *Natural Language Answering Systems*, Carl Hanser Verlag, München Wien, pp. 137-216.

[19] Berwick, R.C. (1986). "Learning from positive only examples: The subset principle and three case studies,"In: R.S. Michalski; J.G. Carbonell; T.M. Mitchell (Eds.) *Machine Learning: An Artificial Intelligence Approach*, Vol. II, Morgan Kaufmann Publishers, Inc., pp 625-645.

[20] Bezdek, J. C.; S.K. Chuah; D. Leep (1986). "Generalized k-nearest neighbour rules," *Fuzzy sets and systems*, Vol. 18, No.3, pp. 237-256.

[21] Bhattacharya, P. "An algebraic environment to process fuzzy images," *Pattern recognition letters*, Vol. 8, No. 2, pp. 29-33.

[22] Blais, J.A.R.; J.C. Muller (1987). "Infological modelling concepts for land related information systems," In: J.A.R. Blais (Ed.) *Lecture Notes in Digital Mapping and Land Information*, University of Calgary, pp. 124-145.

[23] Bronneberg, W. J. H. J.; H.C. Bunt; S.P.J. Landsbergen, R.J.H. Scha; W.J. Schoenmakers; E.P.C. Van Utteren (1980). "The question answering system PHLIQA1," In: L. Bolc (Ed.) *Natural Language Answering Systems*, Carl Hanser Verlag, München Wien, pp. 217-305.

[24] Brown, F. M.; C.B. Schwind (1980). "An integrated theory of natural language understanding," In: L. Bolc (Ed.) *Representation and Processing of Natural Language*, Carl Hanser Verlag, München Wien, pp. 85-120.

[25] Buchanan, B. G.; R. Smith (1989). "Fundamentals of Expert Systems,"In: A. Barr; Cohen, P.R.; E.A. Feigenbaum (Eds.) *The Handbook of Artificial Intelligence*, Vol IV. Addison-Wesley Publishing Company, Inc., Reading, Massachussets, pp. 149-192.

[26] Buckley, J.J.; W. Siler (1988). "Fuzzy Numbers for Expert Systems," In: M.M. Gupta; T, Yamakawa (Eds.) *Fuzzy Logic in Knowledge-based Systems, Decision and Control*, Elsevier Science Publishers, North-Holland, Amsterdam, pp. 153-172.

[27] Bundy, A.; V. Kumar (1987). "Evidential reasoning using stochastic simulation of causal models," *Artificial Intelligence*, Vol. 32, pp. 245-257.

[28] Büttner, G.; Hajo's T.; Kora'ndi M. (1989). "Improvements to the effectiveness of supervised training procedures," *International Journal of Remote Sensing*, Vol. 10, No 6, pp. 1005-1013.

[29] Campbell, W.J.; R.F. Cromp (1990) "Evolution of an intelligent information fusion system," *Photogrammetric Engineering and Remote Sensing* Vol. 56, No.6, June, pp. 867-870.

[30] Campbell, I. A. (1981). "Spatial and temporal variability in errosion," In: *Erosion and Sediment Transport Measurement Symposium*, Proceedings of the Florence Symposium, 22-26 June, IAHS Publication No.133, pp. 447-456.

[31] Capece, J.C.; K.L. Campbell; L.B. Baldwin; K.D. Konyha (1987). "Estimating runoff volumes from flat, high-water table watersheds," *Transactions of the ASAE*, Vol. 30, No. 5, Sept - Oct., pp. 1397-1405.

[32] Carrol, R. (1983). "Automated gully delineation using digital elevation data," *Technical Papers of the* 49_{th} *Annual Meeting of the American Society of Photogrammetry*, March 13-18, Washington D.C., pp. 144-151.

[33] Chattopadhyay, S.; S.M. Chattopadhyay (1989). "Application of remote sensing in terrain analysis, A case study of the Wayanad Plateau, Kerala, India," *International Journal of Remote Sensing*, Vol. 10, No. 6, pp.1075-1084.

[34] Chen, H.S. (1985). *Space Remote Sensing:An Introduction*, Academic Press, Inc., London, Ltd.

[35] Chen, F.; Hsu, W-H.; Chen, C-A. (1989). "Fuzzy approach to solve the recognition problem of handwritten Chinese characters," *Pattern Recognition*, Vol. 22, No.2, pp. 133-141.

[36] Chen, Y. Y.; T.C. Tsao (1989). "A description of dynamic behaviour of fuzzy systems," *IEEE transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 4, pp. 745-755.

[37] Chisci, G. (1981). "Upland erosion: evaluation and measurement," In: *Erosion and Sediment Transport Measurement Symposium*, Proceedings of the Florence Symposium, 22-26 June, IAHS Publication No.133, pp. 331-349.

[38] Cleynenbreugel, J. Van; F. Fierens; P. Suetens; Oosterlinck (1990). "Delineating Road Structures on Satellite Imagery by a GIS Guided Technique," *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 6, pp. 893-898.

[39] Clark, K. L. (1987). "Negation as failure," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 311-325.

[40] Coleman, T.L. (1987). "Soil moisture and iron content effect on spectral characteristics of selected Vertisols and Alfisols in Alabama," *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 12, pp. 1659-1663.

[41] Contzen, J. P. (1989). "The benefits of remote sensing activities of the joint research centre (JRC) in Europe and Africa," *Photogrammetria (PRS)*, Vol. 43, pp. 311-322.

[42] D'Ambrosio, B. (1989). "Extending the mathematics in qualitative Process Theory," *International Journal of Intelligent Systems*, Vol. 4, pp. 55-80.

[43] D'Appolonio, V.; Fuggetta, A.; Lazzarini, P.;, Negri, M.; Pelagatti, G. (1985). "The integration of the network and relational approaches in a DBMS," In: A.F. Gundy (Ed.) *Proceedings of the Fourth British National Conference on Databases*, (BNCOD4), University of Keele, 10 - 12 July, Cambridge University Press, London, N.Y., pp. 179-197.

[44] D'Costa, V.P.; C.J. Johansen; M.F. Baumgardner (1986). "Using Landsat TM data as an aid for detailed soil survey in Kenya," In: PECORA XI *Satellite Remote Sensing: Current Programs and a Look to the Future*, PECORA XI Symposium Proceedings, The ASP, Falls Church, Virginia, pp. 325.

[45] Das, D.C. (1977). "Soil erosion conservation practices and erosion control in India - A case study," In: R.P.C. Morgan (Ed.) *Soil Erosion and Control*, Van Nostrand Reinhold Company, NY., pp.118-150.

[46] De Gloria, S. D.; A.S. Benson (1987). "Interpretability of advanced SPOT film products for forest and agricultural survey," *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 1, Jan, pp. 37-44.

[47] Dellepiane, S.; C. Regazzoni; S.B. Serpico; G. Vernazza (1988). "Extension of IBIS for 3D organ recognition in NMR multislices," *Pattern recognition letters*, Vol. 8, No. 2, pp. 65-72.

[48] Denzo, S.D. (1987) "Gaussian maximum likelihood and contextual classification algorithms for multicrop classification," *IEEE Transactions on Geoscience and Remote Sensing*, Vol GE-25, No. 6.

[49] Di Manzo, M.; E. Trucco; F. Giuchiglia; F. Ricci (1989). "FUR: Understanding functional reasoning," *International Journal of Intelligent Systems*, Vol. 4, pp. 431-457.

[50] Dougherty, E. R.; C.R. Giardina (1988). *Mathematical methods for Artificial Intelligence and Autonomous Systems*, Prentice Hall, Englewood Cliffs, NJ., pp. 59-94.

[51] Drummond, J. (1987). "A framework for handling error in geographic data manipulation," ITC Journal 1987-I, pp. 73-81.

[52] Dubois D.; H. Prade (1979). "Outline of fuzzy set theory: An introduction," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 27-48.

[53] Dubois, D., H. Prade (1979). "Decision-making under fuzziness," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 279-302.

[54] Duffin, P. H. (1988). *Knowledge-based Systems Application in Administrative Government*, Ellis Horwood Limited.

[55] Dune, T. (1977). "Studying patterns of soil erosion in Kenya," In: R.P.C. Morgan (Ed.) *Soil Erosion and its Control*, Van Nostrand Reinhold Company; N.Y., pp. 62-73.

[56] Edwards, D.L. (1983). "Terrain analysis database generation through computer-assisted photo interpretation," *Technical Papers of the* 49_{th} *Annual Meeting of the American Society of Photogrammetry*, March 13-18, Washington D.C., pp. 152-158.

[57] Efraim, T.; P.R. Watkins (eds.) (1988). *Applied Expert Systems*, Elsevier Science Publishers, North-Holland, Amsterdam, NY., Oxford.

[58] Esogbue, A. O. (1986). "Optimal clustering of fuzzy data via fuzzy dynamic programming," *Fuzzy sets and systems*, Vol. 18, No. 3, pp. 283-298.

[59] Even, S. (1979) *Graph Algorithms*, Computer Science Press, Inc.

[60] Farhan, Y.I. (1983). "Soil erosion mapping and evaluation: Photo-geomorphological approach," *Technical Papers of the* 49_{th} *Annual Meeting of the American Society of Photogrammetry*, March 13-18, Washington D.C., pp. 131-143.

[61] Fordon, W.A.; J.C. Bezdek (1979). "The application of fuzzy set theory to medical diagnosis," In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Set Theory Applications*, North-Holland Publishing Company, Amsterdam, N.Y., Oxford, pp. 445-461.

[62] Fowler, M.G.; H.h.K. Burke; K.R. Hardy; N.K. Tripp (1979) "The estimation of rain rate over land from space-borne passive microwave sensors," In: M. Deutch; D.R. Wiesnet; A. Rango (Eds.) *Satellite Hydrology, Proceedings of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing*, American WAter Resources Association, Sioux Falls South Dakota, June 10 - 15, pp. 101-108.

[63] Franklin, J.; A.H. Strahler (1988). "Invertible Canopy reflectance modeling of vegetation structure in semi-arid woodland," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 26, No. 6, Nov., pp. 809-825.

[64] Fukami, S., M. Mizumoto; K. Tanaka (1980). "Some considerations on fuzzy conditional inference," *Fuzzy sets and systems*, Vol. 4, pp. 243-273.

[65] Garey, M.R.; D.S. Johnson (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, N.Y.

[66] Giakoumakis, E.; G. Papaconstantinou; E. Skordalakis (1987). "Rule-based systems and pattern recognition,"*Pattern Recognition Letters*, 5, pp. 267-272.

[67] Gil, M. A.; N. Corral; M.R. Casals (1989). "The likelihood ratio test for goodness of fit with fuzzy experimental observations," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, No. 4, pp. 771-779.

[68] Ginsberg, M. L. (Ed.)(1987). *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California.

[69] Ginsberg, M. L. (1987). "Multi-valued logics," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Ginsberg, M. L. (ed.), Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 251-258.

[70] Ginsberg, M. (1987). "Reasoning about action I:A possible worlds approach," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 434-463.

[71] Godo, L.; R.L. Mantaras; C. Siera; A. Verdaguer (1989). "MILORD: The archtecture and the management of Linguisticall expressed uncertainty," *International journal of Intelligent Systems*, Vol. 4, pp. 471-501.

[72] Escalada-Imaz, G.; M. Ghallab (1988). "A practically efficient and almost linear unification algorithm," *Artificial Intelligence*, Vol. 36, pp. 249-263.

[73] Gooley, M. M.; B.W. Wah (1989). "Efficient reordering of prolog programs," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 1, No. 4, pp. 475-481.

[74] Gugan, D.J. (1988). "Accuracy and Completeness of topographic mapping from SPOT imagery," *Photogrammetric Record*, 12 (72):787-796, Oct.

[75] Gupta, M.M.; T. Yamakawa (eds.) (1988). *Fuzzy computing theory, hardware, and applications*, Elsevier Science Publishers, North-Holland, Amsterdam, N.Y., Oxford.

[76] Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) (1979). *Advances in Fuzzy Set Theory Applications*, North-Holland Publishing Company, Amsterdam, N.Y., Oxford.

[77] Gustafson, D. E.; W. C. Kessel (1979). "Fuzzy clustering with a fuzzy covariance matrix," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 605-620.

[78] Guy, H. P. (1977). "Sediment information for an environmental impact statement regarding a surface coal mine, western United States," In: *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July, pp. 98-108.

[79] Hahn, V. (1989). "Making Understanders out of Parsers: Semantically Driven Parsing as a Key Concept for Realistic Text Understanding Applications," *International Journal of Intelligent Systems*, Vol. 4, No. 3, Fall, pp. 345-393.

[80] Hayes-Roth, F. (1989). "Towards benchmarks for knowledge systems and their implications for data engineering," *IEEE Transactions on knowledge and data engineering*, Vol. 1, No. 1, pp. 101-110

[81] Haggerstone, B. (1983). "Drainage pattern delineation: A function of image scale," *Technical Papers of the* 49_{th} *Annual Meeting of the American Society of Photogrammetry*, March 13 - 18, Washington D.C., pp. 529-533.

[82] Hartley, D.M. (1987a). "A simplified process model for water sedment yield from a single storm: Part I - Modified formulation," *Transactions of the ASAE*, Vol. 30 (3), May - June, pp. 710-717.

[83] Hartley, D.M. (1987b). "A simplified process model for water sedment yield from a single storm: Part II - Performance," *Transactions of the ASAE*, Vol. 30 (3), May- June, pp. 718-723.

[84] Hathaway, R.; J.W. Davenport; J.C. Bezdek (1989). "Relational duals of the c-means clustering algorithms," *Pattern recognition*, Vol. 22, No. 2, pp. 205-212.

[85] Hersch, H. M.; A. Caramazza; H.H. Brownell (1979). "Effects of context on fuzzy membership functions," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 389-408.

[86] Hisdal, E. (1979). "Possibilistically dependent variables and general theory of fuzzy sets," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 215-234.

[87] Holeman, J. N. (1981). "The national erosion inventory of the Soil Conservation Service, US Department of Agriculture, 1977-1979," In: *Erosion and Sediment Transport Measurement Symposium*, Proceedings of the Florence Symposium, 22-26 June, IAHS Publication No.133, pp. 316-319.

[88] Hopcroft, J. E.; J.D. Ullman (1979). *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley Publishing Co., Inc., Reading, Mass., Melo Park.

[89] IAHS-AISH (1981), *Erosion and Sediment Transport Measurement Symposium*, Proceedings of the Florence Symposium, 22-26 June, IAHS-AISH Publication No.133.

[90] IAHS-AISH, (1977). *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July.

[91] Jackson, T.J.; J.C. Ritchie; J. White; L. le Schack (1988). "Airborne laser profile data for measuring ephemeral gully erosion," *Photogrammetric Engineering and Remote Sensing*, Vol.54, No. 8, August, pp. 1181-1185

[92] Jackson, T.J.; P.E. O'Neill (1983). "Aircraft scatterometer observations of the soil moisture on rangeland watersheds," *Technical Papers of the* 49_{th} *Annual Meeting of the American Society of Photogrammetry*, March 13 - 18, WAshington D.C., pp. 198-211.

[93] Jain, R. (1979). "Application of fuzzy sets for the analysis of complex scenes," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 577-587.

[94] Jenson, S.K.; J.O. Domingue (1987). "Extracting topographic structure from digital elevation data for geographic information system analysis," *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 9, Sept., pp. 1593-1600.

[95] Johnson, G.R. (1984). "Integrating data and meta data to enhance the user interface," In: J. Longstaff (Ed.) *Proceedings of the Third British National Conference on Databases (BNCOD3)*, Leeds, 11-13 July, Cambridge University Press, London, NY., pp. 30-39.

[96] Jolly, J.P. (1979) "Utilizing GOES imagery to determine convective storm characteristics in data defficient regions," In: M. Deutch; D.R. Wiesnet; A. Rango (Eds.) *Satellite Hydrology, Proceedings of the Fifth Annual William T. Pecora Memorial Symposium on Remote Sensing*, American WAter Resources Association, Sioux Falls South Dakota, June 10-15, pp. 77-83.

[97] Joshua, W. D. (1977). "Soil erosive power of rainfall in the different climatic zones of Sri Lanka," In: *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July, pp. 51-61.

[98] Kacprzyk, J. (1986a). "Towards 'human-consistent' multistage decision making and control models using fuzzy sets and fuzzy logic," *Fuzzy sets and systems*, Vol. 18, No. 3, pp. 299-314.

[99] Kacprzyk, J. (1986b). "Group decision making with a fuzzy linguistic majority," *Fuzzy sets and systems*, Vol. 18, No. 2, pp. 105-118.

[100] Kandel, A. (1979). "On the theory of fuzzy switching mechanisms (FSM's)," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 247-262.

[101] Kandel, A.; W.J. Byatt (1980). "Fuzzy processes," *Fuzzy Sets and Systems*, 4, pp. 117-152.

[102] Kang, Y.; T. Bahill (1990). "A tool for detecting expert-system errors," *AI EXPERT*, Feb., pp. 46-51

[103] Kaufmann, H.; M.M. Gupta (1988). *Fuzzy Mathematical Models in Engineering and Management Science*, Elsevier Science Publishers, North-Holland, Amsterdam, N.Y.

[104] Keller, J. M.; C.L. Carpentier (1990). "Image segmentation in the presence of uncertainity," *International Journal of Intelligent Systems*, Vol. 5, pp.193-208.

[105] Kent, J.T.; K.V. Mardia (1988). "Spatial classification using fuzzy membership models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol., 10, No. 5, Sept., pp. 659-671.

[106] Keskey, P.E.; D.J. Sykes (1987). "Expert systems in aircraft maintanance training," *Proceedings of the Western Conference on Expert Systems*, WESTEX-87, June 2-4, Anaheim California, pp. 50-56.

[107] Kerr, Y.H. (1989). "NOAA AVHRR and its uses for rainfall and evapotranspiration monitoring," *International Journal of Remote Sensing*, Vol. 10, No.4 and 5, pp. 847-854.

[108] Key, J.; J.A. Maslanik; A.J. Schweiger (1989). "Classification of merged AVHRR and SMMR arctic data with neural networks," *Photogrammetric Engineering and Remote Sensing*, Vol. 55, No. 9, pp. 1331-1338.

[109] Kim, T., J.C. Bezdek; R.J. Hathaway (1989). "Optimality tests for fixed points of the fuzzy c-means algorithm," *Pattern Recognition*, Vol. 21, No. 6, pp. 651-663.

[110] Klinger, A.; M.L. Rhodes (1979). "Organisation and access of image data by areas," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-1, No.1, pp. 50-59.

[111] Kóczy, L.T. (1988). "On the description of relative position of fuzzy patterns," *Pattern recognition letters*, Vol. 8, No. 2, pp. 21-28.

[112] Kodratoff, Y.; JG. Ganascia (1986). "Improving the generalization step in learning," In: R.S. Michalski; J.G. Carbonell; T.M. Mitchell (Eds.) *Machine Learning: An Artificial Intelligence Approach*, Vol. II, Morgan Kaufmann Publishers, Inc., pp. 215-244.

[113] Kondrat'yev, K. Ya.; S.M. Somova; P.P. Fedechenko (1986). "The study of changes in soil quality by remote sensing methods," *Mapping Sciences and Remote Sensing*, Vol. 23, No. 3, pp. 207-216.

[114] Kuz'min, V. B.; S.V. Ovchinnokov; (1980). "Group decisions in arbitrary spaces of fuzzybinary relations," *Fuzzy sets and systems*, Vol. 4, pp. 53-62.

[115] Laender, A. H. F. (1984). "An approach to interactive definition of database views," In: J. Longstaff (Ed.) *Proceedings of the Third British National Conference on Databases (BNCOD3)*, Leeds, 11-13 July, Cambridge University Press, London, N.Y., pp. 173-183.

[116] Larner, D.L. (1990). "Factories, Objects, and Blackboards," *AI EXPERT*, April, pp. 38-45.

[117] Lee, S.; Shan, M-K. (1990). "Access methods of image database," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 4, No. 1, pp.27-44.

[118] Lee, C.N.; P. Liu; M. Chiu (1987). "Dynamic causal model construction for model based reasoning," *Proceedings of the Western Conference on Expert Systems*, WESTEX-87, June 2-4, Anaheim, California, pp. 90-95.

[119] Lehnert , W. G. (1980). "Question answering in natural language processing," In: L. Bolc (Ed.) *Natural Language Answering Systems*, Carl Hanser Verlag, München Wien, pp. 9-71.

[120] Ley, R. G. (1988). "Some aspects of height extraction from SPOT imagery," Photogrammetric Record, Vol. 12, No. 72, pp. 823-832.

[121] Lisytsina, K. N. (1977). "The effect of exogeneous and endogeneous factors on water erosion," In: *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July, pp. 67-74.

[122] Lodwick, G. D.; A.W.L. Colijn (1987). "Expert systems and land-related information systems," In: J.A.R. Blais (Ed.) *Lecture Notes in Digital Mapping and Land Information*, University of Calgary, pp. 168-198.

[123] Looney, C. G. (1988). "Fuzzy petri nets for rule-based decisionmaking," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 18, No. 1, pp. 178-183.

[124] Luhandjula, M.K. (1986). "On Possibilistic linear programming," *Fuzzy Sets and Systems*, 18, pp. 15-30.

[125] Madramotoo, C. A. (1988). "Rainfall and runoff erosion indices for Eastern Canada," *Transactions of the ASAE*, Vol. 31., No. 1, Jan.-Feb., pp.107-110.

[126] Manola, F. (1990a). "Object - Oriented knowledge bases: Part I," *AI EXPERT*, March 1990, pp. 26-36.

[127] Manola, F. (1990b). "Object - Oriented knowledge bases: Part II," *AI EXPERT*, April, pp. 46-57.

[128] Maurer, H. (1974). "Quantification of textures- textural parameters and their significance for classifying agricultural crop types from colour aerial photographs," *Photogrammetria*, Vol. 30, pp. 21-40.

[129] Matsuyama, T. (1987). "A knowledge based aerial image understanding system and expert system for image processing," IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-25, pp.

[130] McCoy, M. S.; R.R. Levary (1988). "Augmenting knowledge acquisition processes of expert systems with human performance modelling techniques," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 18, No.3, May/June, pp. 467-470.

[131] McDermott, D.; J. Doyle (1987). "Non-Monotonic Logic I," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 111-126.

[132] McHenry, J. R.; J.C. Ritchie (1977). "Estimating field erosion losses from fallout cesium-137 measurements," In: *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July, pp. 26-33.

[133] McCauley, J. D. (1981). *Everything that Linguists have allways wanted to know about Logic*, The University of Chicago Press.

[134] McKim, H.L; S.G. Ungar; W.J. O'Donoghue; M.S. Miller; C.J. Merry (1983) "Integration of Landsat data into the Saginaw river basin geographic information system," *Technical Papers of the* 49_{th} *Annual Meeting of the American Society of Photogrammetry*, March 13 - 18, Washington D.C., pp. 212-224.

[135] Meyer, L.D.; W.H. Wischmeier (1969). "Mathematical simulation of the process of soil erosion by water," In: R.P.C. Morgan (Ed.) *Soil Erosion and its Control*, Van Nostrand Reinhold Company; N.Y., 1986, pp. 227-232.

[136] Minker, J. (1987). "On indefinite databases and the closed world assumption," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 326-336.

[137] Mirtskhoulava, TS. E. (1981). "Land erosion, research equipment, forecasting methods and prospects for their improvement," In: *Erosion and Sediment Transport Measurement Symposium*, Proceedings of the Florence Symposium, 22-26 June, IAHS Publication No.133, pp. 521-526.

[138] Morgan, R.P.C.; D.D.V. Morgan; B.J. Finney, (1984). "A predictive model for the soil erosion risk," In: R.P.C. Morgan (Ed.) *Soil Erosion and its Control*, Van Nostrand Reinhold Company; N.Y., 1986, pp. 251-259.

[139] Morgan, R.P.C. (1981). "Field measurement of splash erosion," In: *Erosion and Sediment Transport Measurement Symposium*, Proceedings of the Florence Symposium, 22-26 June, IAHS Publication No.133, pp. 373-382.

[140] Mou, J. (1981). "The establishment of experimental plots for studying runoff and soil loss in the rolling loess regions of China," In: *Erosion and Sediment Transport Measurement Symposium*, Proceedings of the Florence Symposium, 22-26 June, IAHS Publication No.133, pp. 467-477.

[141] Nath, A. K.; S.W. Lin; T.T. Lee (1985). "On some properties of a linguistic classifier," *Fuzzy sets and systems*, Vol. 17, No.3, pp. 297-311.

[142] Negoita C. V. (1979). "On fuzzy systems," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 69-75.

[143] Newman, I. A.; J. Sethi (1985). "Some observations on user interface design and user performance," In: A.F. Gundy (Ed.) *Proceedings of the Fourth British National Conference on Databases (BNCOD4)*, University of Keele, 10 - 12 July, Cambridge University Press, London, N.Y., pp. 199-213.

[144] Ngu, A. H. H. (1989). "Conceptual Transaction Modelling," *IEEE Transactions on knowledge and data engineering*, Vol. 1 No. 4, pp. 508-519.

[145] Nguyen, H. T. (1979). "Towards a calculus of the mathematical notion of possibility," In: Gupta, M. M.; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 235-246.

[146] Nicoln, B. (1987). "A knowledge based system for the analysis of aerial images," *IEEE Transactions of Geoscience and Remote Sensing*, Vol. GE-25, No.3, pp.

[147] Niemann, B.J.; J.G. Sullivan; S.J. Ventura; N.R. Christman (1987). "Results of the Dane County Land Records Project," *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 10, pp. 1371-1378.

[148] Noble, H. (1984). "Occurence dependencies, linkages and the structure of the relational data base," In: J. Longstaff (Ed.) *Proceedings of the Third British National Conference on Databases (BNCOD3)*, Leeds, 11-13 July, Cambridge University Press, London, N.Y., pp. 15-27.

[149] O'Leary, D. E.; N.A. Kandelin (1988). "Validating the weights in a rule-based expert system,"International Journal of Expert Systems, Research and Applications Vol. 1, No 3. pp. 253-279.

[150] Onstad, C. A.; A.J. Bowie (1977). "Basin sediment yield modelling using hydrrological variables," In: *Erosion and Solid Matter Transport in Inland Waters Symposium*, Proceedings of the Paris Symposium, July, pp. 191-202.

[151] Östman, A. (1987). "Photogrammetrically sampled digital elevation models," *Photogrammetric Record*, Vol 12, pp. 333-340.

[152] Page-Jones, M. (1988). *The Practical Guide to Structured Systems Design*, Sec. Ed., Youdon Press, Englewood Cliffs, N.J.

[153] Pal, S. K. (1989). "Fuzzy skeletonization of an image," *Pattern recognition letters*, Vol. 10, No. 1, pp. 17 - 23.

[154] Palmer, I.R. (1978). "Practicalities in applying a formal methodology to data analysis," In: S.B. Yao; S.B. Navathe; J.L. Weldon; T.L. Kunii (Eds.) *Data Base Design Techniques I: Requirements and Logical Structures*, Proceedings, New York, May, Springer-Verlag, Berlin, Heidelberg, N.Y. 1982., pp. 147-171.sub-

[155] Parsaye, K.; M. Chignell; S. Khoshafian; H. Wong (1990). "Intelligent Databases," *AI EXPERT*, March, pp. 38-47.

[156] Pearl, J. (1987). "Evidential Reasoning Using Stochastic Simulation of Causal Models," *Artificial Intelligence*, Vol. 32, pp. 245-257

[157] Pech, R. P. (1987). "Reflectance modelling of semi-arid woodlands," *Remote Sensing of Environment*, pp. 367-377.

[158] Philipson, W. R. (1988). "Operational interpretation of AVHRR vegetation indices for world crop information," *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 1, pp. 55-59.

[159] Pilon, P. G.; P.J. Howarth; R.A. Bullock (1988). "An enhanced classification approach to change detection in semi-arid environments," *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 12, pp. 1709-1716.

[160] Piasecki, K. (1986). "Probability of fuzzy events defined as denumerable additivity measure," *Fuzzy Sets and Systems*, 17, pp. 271-284.

[161] Pittarelli, M. (1990). "Probabilistic Databases for Decision Analysis," *International Journal of Intelligent Systems*, Vol. 5, pp.209-236.

[162] Poiker, T. K.; M. Feuchwanger (1987). "Infological and Datalogical Spatial Models," In: J.A.R. Blais (Ed.) *Lecture Notes in Digital Mapping and Land Information*, University of Calgary, pp. 146-167.

[163] Poirier, S. (1989). "SAR applications in agriculture," *International Journal of Remote Sensing*, Vol. 10, No. 6, pp. 1085-

[164] Price, S. (1989). "Modelling the Temporal element in land info system," *International Journal of Geographic Information Systems* Vol. 3, No. 3, pp.233-243

[165] Prisley, S. P.; J.L. Smith (1987). "Using classification error matrices to improve the accuracy of weighted land-cover models," *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 9, pp. 1259-1263.

[166] Ramstein, G.; M. Raffy (1989). "Analysis of the structure of radiometric remotely-sensed images," *International Journal of Remote Sensing*, Vol. 10, No. 6, pp. 1049-1073.

[167] Rau, L.F.; P.S. Jacobs (1989). "NL ∩ IR: NAtural Language for Information Retrieval," *International Journal of Intelligent Systems*, Vol. 4, No. 3, Fall, pp. 319-343.

[168] Reiter, R. (1987a). "A theory of diagnosis from first principles," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 352-371.

[169] Reiter, R. (1987b). "On closed world data bases," In: M.L. Ginsberg (Ed.) *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann Publishers, Inc., Los Altos, California, pp. 300-309.

[170] Riggins, R.E.; J.T. Bundy (1981). "R-Factors for soil loss impact prediction," *Transactions of the ASCE*, August, pp. 851-857.

[171] Sader, S. A. (1987). "Forest Biomass, Canopy Structure, and Species Composition Relationships With Multipolarization L-Band Synthetic Aperture Radar Data," *PE & RS*, Vol. 53, No. 2, February, pp. 193-202.

[172] Sagerer, G. (1987). "Automatic interpretation of medical image sequences," *Pattern recognition letters*, Vol. 8, No.2, pp. 87-102.

[173] Sammut, C.; R.B. Banerji (1986). "Learning concepts by asking questions," In: R.S. Michalski; J.G. Carbonell; T.M. Mitchell (Eds.) *Machine Learning: An Artificial Intelligence Approach*, Vol. II, Morgan Kaufmann Publishers, Inc., pp. 167-191.

[174] Sanchez, E. (1979). "Composition of fuzzy relations," In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 421-433.

[175] Schneider, M.; A. Kandel (1988). "Application of Fuzzy Expected Interval to Fuzzy Expert Systems," *International Journal of Expert Systems, Research and Applications* ,Vol. 1, No. 2, pp. 169-186.

[176] Shapiro, L. G. (1980). "Design of spatial information system," *Map Data Processing*, Academic Press, pp. 101-117.

[177] Shekar, B.; M. Narasimha; G. Krishna (1987). "A knowledge based clustering scheme," *Pattern Recognition Letters*, Vol. 5, pp. 253-259.

[178] Silver, B. (1986). "Precondition Analysis: Learning Control Information," In: R.S. Michalski; J.G. Carbonell; T.M. Mitchell (Eds.) *Machine Learning: An Artificial Intelligence Approach*, pp. 647-670.

[179] Slater, P. N. (1980). *Remote Sensing: Optics and Optical Systems*, Addison-Wesley Publishing Company, London, Amsterdam.

[180] Spillman, R. (1990). "Managing uncertainty with belief functions," AI EXPERT, May, pp. 44-49.

[181] Thomson, A.G. (1989) "Population density of sheep related to Landsat Thematic Mapper radiance," *International Journal of Remote Sensing*, Vol.10, No.12, pp.1907-1917.

[182] Tong, R. M. (1979). "The construction and evaluation of fuzzy models," In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 559-576.

[183] Tsukamoto, Y. (1979). "An approach to fuzzy reasoning method," In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp. 137-149.

[184] Usery, E. L.; P. Altheide (1988). "Knowledge-based GIS techniques applied to geological engineering," *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 11, Nov., pp.1623-1628.

[185] Vopenka, P.; P. Hajek (1972). *The theory of semi sets*, North-Holland Publishing Company, Amsterdam, London.

[186] Wahome, E.K. (1989). *Soil Erosion Measurement Under Natural Rainfall for Evaluating the USLE in Manitoba*, University of Manitoba Msc. Thesis, March 7.

[187] Waugh, T.C. (1987). "The GEOVIEW design: A relational database approach to geographic data handling," *Int. Journal of Geographical Information Systems*, Vol. I, No. 2, pp. 101-118

[188] Wechler, W. (1979). "Fuzzy sets and languages," In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 263-278.

[189] Weiss, J.J; M.L. Donnel (1979). "A general purpose policy capturing device using fuzzy production rules," In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 589-604

[190] Willaeys, D.; M. Malvache (1979). "Contribution of fuzzy sets theory to man-machine system," In: M.M. Gupta; R.K. Ragade; R.R. Yager (Eds.) *Advances in Fuzzy Sets Theory and Applications*, North-Holland Publishing Company, Amsterdam, pp 481-499.

[191] Williams, V. L.; W.R. Philipson; W.D. Philpot, (1987). "Identifying vegetable crops with Landsat Mapper Data," *Photogrammetric Engineering and Remote Sensing* Vol. 53, No.2, Feb, pp. 187-191.

[192] Wygralak, M. (1986). "Fuzzy cardinals based on the generalized equality of fuzzy subsets," *Fuzzy sets and systems*, Vol. 18, pp. 143-158.

[193] Yager, R. R. (1980). "On the lack of inverses in fuzzy arithmetic," *Fuzzy sets and systems*, Vol. 4, pp 73-82.

[194] Yager, R. R., (Ed.) (1989). *International Journal of Intelligent Systems*, Vol. 4, No. 3, John Wiley and Sons, NY., Fall.

VITA

Candidate's full name	: Elifuraha Gerard B. Mtalo
Place and date of birth	: Uru East, Kilimanjaro, Tanzania,
	September 12, 1949.
Permanent address	: Ardhi Institute,
	P.O. Box 35176,
	Dar Es Salaam, Tanzania.
Schools attended	: Primary School
	Materuni, Kilimanjaro,
	1958 - 1962.
	: Middle School
	Uru, Kilimanjaro,
	1963 - 1965.
	Secondary School
	Kibaha, Dar Es Salaam,
	1966 - 1969.
	: High School
	Kibaha, Dar Es Salaam,
	1970 - 1972.
Institutes Attended	: Ardhi Institute
	Dar Es Salaam,
	1972 - 1973.
	: International Preparatory Institute
	Budapest, Hungary,
	1973 - 1974.
Universities attended	: Technical University of Budapest
	Budapest, Hungary,
	MSc.(Civil Eng.)
	1974 - 1979.