THE UNIVERSITY OF CALGARY

UNCERTAINTY MANAGEMENT USING FUZZY SETS
IN A
COMMERCIAL VECTOR GIS

by

SYLVIA LAM

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

AUGUST, 1992

ISBN 0-315-79185-3

Name SYLVIA LAM

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

SYSTEM ~~SEHN~~ SCIENCE

SUBJECT TERM

0	7	9	0

SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
- Architecture 0729
- Art History 0377
- Cinema 0900
- Dance 0378
- Fine Arts 0357
- Information Science 0723
- Journalism 0391
- Library Science 0399
- Mass Communications 0708
- Music 0413
- Speech Communication 0459
- Theater 0465

EDUCATION
- General 0515
- Administration 0514
- Adult and Continuing 0516
- Agricultural 0517
- Art 0273
- Bilingual and Multicultural 0282
- Business 0688
- Community College 0275
- Curriculum and Instruction 0727
- Early Childhood 0518
- Elementary 0524
- Finance 0277
- Guidance and Counseling 0519
- Health 0680
- Higher 0745
- History of 0520
- Home Economics 0278
- Industrial 0521
- Language and Literature 0279
- Mathematics 0280
- Music 0522
- Philosophy of 0998
- Physical 0523
- Psychology 0525
- Reading 0535
- Religious 0527
- Sciences 0714
- Secondary 0533
- Social Sciences 0534
- Sociology of 0340
- Special 0529
- Teacher Training 0530
- Technology 0710
- Tests and Measurements 0288
- Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
- Language
  - General 0679
  - Ancient 0289
  - Linguistics 0290
  - Modern 0291
- Literature
  - General 0401
  - Classical 0294
  - Comparative 0295
  - Medieval 0297
  - Modern 0298
  - African 0316
  - American 0591
  - Asian 0305
  - Canadian (English) 0352
  - Canadian (French) 0355
  - English 0593
  - Germanic 0311
  - Latin American 0312
  - Middle Eastern 0315
  - Romance 0313
  - Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
- Philosophy 0422
- Religion
  - General 0318
  - Biblical Studies 0321
  - Clergy 0319
  - History of 0320
  - Philosophy of 0322
- Theology 0469

SOCIAL SCIENCES
- American Studies 0323
- Anthropology
  - Archaeology 0324
  - Cultural 0326
  - Physical 0327
- Business Administration
  - General 0310
  - Accounting 0272
  - Banking 0770
  - Management 0454
  - Marketing 0338
- Canadian Studies 0385
- Economics
  - General 0501
  - Agricultural 0503
  - Commerce-Business 0505
  - Finance 0508
  - History 0509
  - Labor 0510
  - Theory 0511
- Folklore 0358
- Geography 0366
- Gerontology 0351
- History
  - General 0578
  - Ancient 0579
  - Medieval 0581
  - Modern 0582
  - Black 0328
  - African 0331
  - Asia, Australia and Oceania 0332
  - Canadian 0334
  - European 0335
  - Latin American 0336
  - Middle Eastern 0333
  - United States 0337
- History of Science 0585
- Law 0398
- Political Science
  - General 0615
  - International Law and Relations 0616
  - Public Administration 0617
- Recreation 0814
- Social Work 0452
- Sociology
  - General 0626
  - Criminology and Penology 0627
  - Demography 0938
  - Ethnic and Racial Studies 0631
  - Individual and Family Studies 0628
  - Industrial and Labor Relations 0629
  - Public and Social Welfare 0630
  - Social Structure and Development 0700
  - Theory and Methods 0344
- Transportation 0709
- Urban and Regional Planning 0999
- Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
- Agriculture
  - General 0473
  - Agronomy 0285
  - Animal Culture and Nutrition 0475
  - Animal Pathology 0476
  - Food Science and Technology 0359
  - Forestry and Wildlife 0478
  - Plant Culture 0479
  - Plant Pathology 0480
  - Plant Physiology 0817
  - Range Management 0777
  - Wood Technology 0746
- Biology
  - General 0306
  - Anatomy 0287
  - Biostatistics 0308
  - Botany 0309
  - Cell 0379
  - Ecology 0329
  - Entomology 0353
  - Genetics 0369
  - Limnology 0793
  - Microbiology 0410
  - Molecular 0307
  - Neuroscience 0317
  - Oceanography 0416
  - Physiology 0433
  - Radiation 0821
  - Veterinary Science 0778
  - Zoology 0472
- Biophysics
  - General 0786
  - Medical 0760

EARTH SCIENCES
- Biogeochemistry 0425
- Geochemistry 0996
- Geodesy 0370
- Geology 0372
- Geophysics 0373
- Hydrology 0388
- Mineralogy 0411
- Paleobotany 0345
- Paleoecology 0426
- Paleontology 0418
- Paleozoology 0985
- Palynology 0427
- Physical Geography 0368
- Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
- Environmental Sciences 0768
- Health Sciences
  - General 0566
  - Audiology 0300
  - Chemotherapy 0992
  - Dentistry 0567
  - Education 0350
  - Hospital Management 0769
  - Human Development 0758
  - Immunology 0982
  - Medicine and Surgery 0564
  - Mental Health 0347
  - Nursing 0569
  - Nutrition 0570
  - Obstetrics and Gynecology 0380
  - Occupational Health and Therapy 0354
  - Ophthalmology 0381
  - Pathology 0571
  - Pharmacology 0419
  - Pharmacy 0572
  - Physical Therapy 0382
  - Public Health 0573
  - Radiology 0574
  - Recreation 0575
  - Speech Pathology 0460
  - Toxicology 0383
- Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences
- Chemistry
  - General 0485
  - Agricultural 0749
  - Analytical 0486
  - Biochemistry 0487
  - Inorganic 0488
  - Nuclear 0738
  - Organic 0490
  - Pharmaceutical 0491
  - Physical 0494
  - Polymer 0495
  - Radiation 0754
- Mathematics 0405
- Physics
  - General 0605
  - Acoustics 0986
  - Astronomy and Astrophysics 0606
  - Atmospheric Science 0608
  - Atomic 0748
  - Electronics and Electricity 0607
  - Elementary Particles and High Energy 0798
  - Fluid and Plasma 0759
  - Molecular 0609
  - Nuclear 0610
  - Optics 0752
  - Radiation 0756
  - Solid State 0611
- Statistics 0463

Applied Sciences
- Applied Mechanics 0346
- Computer Science 0984
- Engineering
  - General 0537
  - Aerospace 0538
  - Agricultural 0539
  - Automotive 0540
  - Biomedical 0541
  - Chemical 0542
  - Civil 0543
  - Electronics and Electrical 0544
  - Heat and Thermodynamics 0348
  - Hydraulic 0545
  - Industrial 0546
  - Marine 0547
  - Materials Science 0794
  - Mechanical 0548
  - Metallurgy 0743
  - Mining 0551
  - Nuclear 0552
  - Packaging 0549
  - Petroleum 0765
  - Sanitary and Municipal 0554
  - System Science 0790
- Geotechnology 0428
- Operations Research 0796
- Plastics Technology 0795
- Textile Technology 0994

PSYCHOLOGY
- General 0621
- Behavioral 0384
- Clinical 0622
- Developmental 0620
- Experimental 0623
- Industrial 0624
- Personality 0625
- Physiological 0989
- Psychobiology 0349
- Psychometrics 0632
- Social 0451

Nom ______________________________

Dissertation Abstracts International est organisé en catégories de sujets. Veuillez s.v.p. choisir le sujet qui décrit le mieux votre thèse et inscrivez le code numérique approprié dans l'espace réservé ci-dessous.

______________________________ ☐☐☐☐ U·M·I

SUJET — CODE DE SUJET

Catégories par sujets

HUMANITÉS ET SCIENCES SOCIALES

COMMUNICATIONS ET LES ARTS

- Architecture 0729
- Beaux-arts 0357
- Bibliothéconomie 0399
- Cinéma 0900
- Communication verbale 0459
- Communications 0708
- Danse 0378
- Histoire de l'art 0377
- Journalisme 0391
- Musique 0413
- Sciences de l'information 0723
- Théâtre 0465

ÉDUCATION

- Généralités 515
- Administration 0514
- Art 0273
- Collèges communautaires 0275
- Commerce 0688
- Économie domestique 0278
- Éducation permanente 0516
- Éducation préscolaire 0518
- Éducation sanitaire 0680
- Enseignement agricole 0517
- Enseignement bilingue et multiculturel 0282
- Enseignement industriel 0521
- Enseignement primaire. 0524
- Enseignement professionnel 0747
- Enseignement religieux 0527
- Enseignement secondaire 0533
- Enseignement spécial 0529
- Enseignement supérieur 0745
- Évaluation 0288
- Finances 0277
- Formation des enseignants 0530
- Histoire de l'éducation 0520
- Langues et littérature 0279
- Lecture 0535
- Mathématiques 0280
- Musique 0522
- Orientation et consultation 0519
- Philosophie de l'éducation 0998
- Physique 0523
- Programmes d'études et enseignement 0727
- Psychologie 0525
- Sciences 0714
- Sciences sociales 0534
- Sociologie de l'éducation 0340
- Technologie 0710

LANGUE, LITTÉRATURE ET LINGUISTIQUE

- Langues
  - Généralités 0679
  - Anciennes 0289
  - Linguistique 0290
  - Modernes 0291
- Littérature
  - Généralités 0401
  - Anciennes 0294
  - Comparée 0295
  - Mediévale 0297
  - Moderne 0298
  - Africaine 0316
  - Américaine 0591
  - Anglaise 0593
  - Asiatique 0305
  - Canadienne (Anglaise) 0352
  - Canadienne (Française) 0355
  - Germanique 0311
  - Latino-américaine 0312
  - Moyen-orientale 0315
  - Romane 0313
  - Slave et est-européenne 0314

PHILOSOPHIE, RELIGION ET THEOLOGIE

- Philosophie 0422
- Religion
  - Généralités 0318
  - Clergé 0319
  - Études bibliques 0321
  - Histoire des religions 0320
  - Philosophie de la religion 0322
- Théologie 0469

SCIENCES SOCIALES

- Anthropologie
  - Archéologie 0324
  - Culturelle 0326
  - Physique 0327
- Droit 0398
- Économie
  - Généralités 0501
  - Commerce-Affaires 0505
  - Économie agricole 0503
  - Économie du travail 0510
  - Finances 0508
  - Histoire 0509
  - Théorie 0511
- Études américaines 0323
- Études canadiennes 0385
- Études féministes 0453
- Folklore 0358
- Géographie 0366
- Gérontologie 0351
- Gestion des affaires
  - Généralités 0310
  - Administration 0454
  - Banques 0770
  - Comptabilité 0272
  - Marketing 0338
- Histoire
  - Histoire générale 0578
  - Ancienne 0579
  - Médiévale 0581
  - Moderne 0582
  - Histoire des noirs 0328
  - Africaine 0331
  - Canadienne 0334
  - États-Unis 0337
  - Européenne 0335
  - Moyen-orientale 0333
  - Latino-américaine 0336
  - Asie, Australie et Océanie 0332
- Histoire des sciences 0585
- Loisirs 0814
- Planification urbaine et régionale 0999
- Science politique
  - Généralités 0615
  - Administration publique 0617
  - Droit et relations internationales 0616
- Sociologie
  - Généralités 0626
  - Aide et bien-àtre social 0630
  - Criminologie et établissements pénitentiaires 0627
  - Démographie 0938
  - Études de l' individu et de la famille 0628
  - Études des relations interethniques et des relations raciales 0631
  - Structure et développement social 0700
  - Théorie et méthodes 0344
  - Travail et relations industrielles 0629
- Transports 0709
- Travail social 0452

SCIENCES ET INGÉNIERIE

SCIENCES BIOLOGIQUES

- Agriculture
  - Généralités 0473
  - Agronomie. 0285
  - Alimentation et technologie alimentaire 0359
  - Culture 0479
  - Élevage et alimentation 0475
  - Exploitation des péturages 0777
  - Pathologie animale 0476
  - Pathologie végétale 0480
  - Physiologie végétale 0817
  - Sylviculture et faune 0478
  - Technologie du bois 0746
- Biologie
  - Généralités 0306
  - Anatomie 0287
  - Biologie (Statistiques) 0308
  - Biologie moléculaire 0307
  - Botanique 0309
  - Cellule 0379
  - Écologie 0329
  - Entomologie 0353
  - Génétique 0369
  - Limnologie 0793
  - Microbiologie 0410
  - Neurologie 0317
  - Océanographie 0416
  - Physiologie 0433
  - Radiation 0821
  - Science vétérinaire 0778
  - Zoologie 0472
- Biophysique
  - Généralités 0786
  - Medicale 0760

SCIENCES DE LA TERRE

- Biogéochimie 0425
- Géochimie 0996
- Géodésie 0370
- Géographie physique 0368
- Géologie 0372
- Géophysique 0373
- Hydrologie 0388
- Minéralogie 0411
- Océanographie physique 0415
- Paléobotanique 0345
- Paléoécologie 0426
- Paléontologie 0418
- Paléozoologie 0985
- Palynologie 0427

SCIENCES DE LA SANTÉ ET DE L'ENVIRONNEMENT

- Économie domestique 0386
- Sciences de l'environnement 0768
- Sciences de la santé
  - Généralités 0566
  - Administration des hipitaux 0769
  - Alimentation et nutrition 0570
  - Audiologie 0300
  - Chimiothérapie 0992
  - Dentisterie 0567
  - Développement humain 0758
  - Enseignement 0350
  - Immunologie 0982
  - Loisirs 0575
  - Médecine du travail et thérapie 0354
  - Médecine et chirurgie 0564
  - Obstétrique et gynécologie 0380
  - Ophtalmologie 0381
  - Orthophonie 0460
  - Pathologie 0571
  - Pharmacie 0572
  - Pharmacologie 0419
  - Physiothérapie 0382
  - Radiologie 0574
  - Santé mentale 0347
  - Santé publique 0573
  - Soins infirmiers 0569
  - Toxicologie 0383

SCIENCES PHYSIQUES

Sciences Pures

- Chimie
  - Genéralités 0485
  - Biochimie 487
  - Chimie agricole 0749
  - Chimie analytique 0486
  - Chimie minérale 0488
  - Chimie nucléaire 0738
  - Chimie organique 0490
  - Chimie pharmaceutique 0491
  - Physique 0494
  - PolymÇres 0495
  - Radiation 0754
- Mathématiques 0405
- Physique
  - Généralités 0605
  - Acoustique 0986
  - Astronomie et astrophysique 0606
  - Electronique et électricité 0607
  - Fluides et plasma 0759
  - Météorologie 0608
  - Optique 0752
  - Particules (Physique nucléaire) 0798
  - Physique atomique 0748
  - Physique de l'état solide 0611
  - Physique moléculaire 0609
  - Physique nucléaire 0610
  - Radiation 0756
- Statistiques 0463

Sciences Appliqués Et Technologie

- Informatique 0984
- Ingénierie
  - Généralités 0537
  - Agricole 0539
  - Automobile 0540
  - Biomédicale 0541
  - Chaleur et ther modynamique 0348
  - Conditionnement (Emballage) 0549
  - Génie aérospatial 0538
  - Génie chimique 0542
  - Génie civil 0543
  - Génie électronique et électrique 0544
  - Génie industriel 0546
  - Génie mécanique 0548
  - Génie nucléaire 0552
  - Ingénierie des systàmes 0790
  - Mécanique navale 0547
  - Métallurgie 0743
  - Science des matériaux 0794
  - Technique du pétrole 0765
  - Technique minière 0551
  - Techniques sanitaires et municipales 0554
  - Technologie hydraulique 0545
- Mécanique appliquée 0346
- Géotechnologie 0428
- Matières plastiques (Technologie) 0795
- Recherche opérationnelle 0796
- Textiles et tissus (Technologie) 0794

PSYCHOLOGIE

- Généralités 0621
- Personnalité 0625
- Psychobiologie 0349
- Psychologie clinique 0622
- Psychologie du comportement 0384
- Psychologie du développement 0620
- Psychologie expérimentale 0623
- Psychologie industrielle 0624
- Psychologie physiologique 0989
- Psychologie sociale 0451
- Psychométrie 0632

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Uncertainty Management using Fuzzy Sets in a Commercial Vector GIS" submitted by Sylvia Lam in partial fulfilment of the requirements for the degree of Master of Science.

Supervisor, Dr. I. K. Crain
Department of Geomatics Engineering

Dr. J.A.R. Blais
Department of Geomatics Engineering

Dr. P. Gong
Department of Geomatics Engineering

Dr. N.M. Waters
Department of Geography

August 26, 1992
(Date)

ABSTRACT

This thesis examines the management of imprecision and vagueness in geographical databases using fuzzy sets. Several fuzzy data modelling concepts were implemented in a prototype forest ecological geographic information system (GIS). The concepts of linguistic variables and fuzzy numbers were applied to represent the imprecision and vagueness in quantitative and qualitative attributes. Three fuzzy comparison operators were implemented to facilitate information retrieval with fuzzy criteria. Furthermore, a set of classification procedures has been developed to handle ecological classification with imprecise ecosystem definitions. Polygon overlay and consolidation spatial operations were also implemented to handle fuzzy attributes. Graphical display of uncertainty information is also illustrated.

This research has demonstrated that fuzzy sets provide better management of imprecise and ambiguous information than conventional techniques. In many natural resource databases, because much data are inherently fuzzy, the techniques presented in this thesis can be applied to enhance the performance of decision support GIS.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Ian K. Crain, for introducing me to the interesting topics of uncertainty management and fuzzy logic. I also want to thank him and members of my supervisory committee Dr. J.A.R. Blais, Dr. P. Gong and Dr. N.M. Waters for their support and encouragement throughout my M.Sc. program. My thanks are also extended to my fellow students and friends, particularly Mr. W.K. Lam and Mr. P.A. Robbins, for their friendship and assistance on many occasions. Expert advice on forest ecology and related issues from Dr. Keith Jones and Dr. Ian Corns and financial support from the University of Calgary, the Natural Sciences and Engineering Research Council of Canada and Hughes Aircraft of Canada are also acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

As Geographic Information Systems (GIS) assume increasingly important roles in planning and decision making, the quality of data becomes a prime concern. Quality of data directly affects the reliability of analysis results. However, much data in GIS, especially natural resources data, are inexact and sometimes error-loaded. Burrough [1986] listed many sources of errors in a GIS. This list is by no means exhaustive, but it reveals the extent of the data quality problem in GIS. The study of uncertainty in GIS deals with the identification, reduction, modelling and representation of the inexactness, inaccuracy and incompleteness in geographical databases.

1.1 Some Limitations of Conventional Information Processing Techniques for GIS Applications

For many years scientists and researchers have been studying different natural and economic phenomena using the precise tools of conventional mathematics [Zadeh, 1976]. Consequently, the development of computer data processing and database techniques have also followed the stringent rules in conventional mathematics. One limitation of these techniques is that they are designed for modelling exact data. Attributes are assumed to be well represented by a single value and the domains of these attributes are assumed to be well-defined. These techniques work well in scientific and engineering applications in which data are exact quantities. Yet, some geographical data are best described in qualitative or linguistic terms such as *well-drained,* describing soil drainage. Linguistic terms are imprecise symbols representing a range of similar values. Therefore, conventional techniques are inadequate for representing the vagueness in linguistic values.

In some situations, qualitative terms are preferred even when quantitative information is available. For instance, it is often better to describe soil texture as sandy, loamy or silty than to provide the exact percentage contents of sand, silt and clay of a soil sample. Using conventional techniques one has to express the classification criteria in exact terms, for example, the soil texture triangle shows that if the percentage of sand is more than 70% sand and less than 15% clay then the sample is labelled loamy sand. This rigid condition does not reflect the gradual transition of the soil texture. While the difference between a sample with 69% sand and one with 70% sand is minimal, they are described by different terms. As the difference between the terms is a matter of degree rather than of kind, the boundary between the two classes should be gradual rather than clear-cut. Conventional techniques do not allow easy representation of this gradual transition from one class to another. Representing qualitative values without losing significant information presents a challenge to conventional techniques.

In addition, conventional techniques are designed to analyze orderly or *mechanistic* systems. Analyzing complex and ill-defined systems using conventional quantitative techniques is often ineffective. For instance, the description of a Lower Boreal Cordilleran Ecoregion (LBC) [Corns & Annas 1986] reads as follows:

> *"The LBC Ecoregion.... occurs at elevations of 800 to 1150 m, **mainly** on ground moraine of Continental origin. The LBC/Upper Boreal Cordilleran (UBC) boundary is lower in the **moister** and/or **cooler** parts of the zone, occurring at 900 to 1050 m on the north slope.... On the south slopes .... the LBC/UBC boundary is **much higher** (up to 1150 m)...."*

Translating this linguistic description into a set of precise *if-then* rules requires one to quantify the terms *mainly, moister, cooler* and *much higher*. Furthermore, two-valued or Boolean logical operations, which are fundamental to conventional techniques, are too rigid to accommodate sites with less-than-perfect match. This could result in many sites being unclassified. As expressed in the *principle of incompatibility* [Zadeh, 1973], the more complex a system, the lesser our ability to make precise and yet significant statements about its behavior. In fact, the ability to summarize information constitutes one difference between human and artificial intelligence (AI). To allow GIS to become more powerful, intelligent decision support tools for solving less structured problems, a human-like approach to information modelling and processing should be adapted.

Among many theories developed from AI research, one appropriate theory which has been applied to the development of inexact data modelling and processing is the *fuzzy set theory* introduced by L.A. Zadeh [1965]. Fuzzy set theory is a mathematical theory developed to represent formally and consistently inexact or *fuzzy* information such as the ambiguity found in linguistic values. Borrowing from the concept of fuzzy set theory, Buckles and Petry [1982] proposed a fuzzy relational database (FRDB) model to incorporate fuzzy data within the specification of a relational database, and to develop mechanisms for the manipulation of the fuzzy data. The model was designed to handle both non-fuzzy and fuzzy data, with non-fuzzy data being treated as a special case of the fuzzy representation. Since then, additional work has been carried out by Buckles and Petry [1983, 1984], Semankova-Leech and Kandel [1985], Shenoi and Melton [1989] Lui and Li [1990] to improve upon the original model.

1.2 Uncertainty in Geographic Information

One research focus in GIS that has been receiving more attention is the management of uncertainty [Robinson & Strahler, 1984; Robinson & Frank, 1985; Bédard 1987; Stoms 1987; Walsh et al 1987; Leung, 1988; Miller et al, 1989; Wang et al, 1990; Sui, 1990]. The main concerns in this research area are collecting, tracking, encoding, modelling and reporting uncertainty in a GIS environment [Guptill, 1989]. The term *uncertainty*, has been used to refer to vagueness, ambiguity, generality, incompleteness, imprecision and inaccuracy in data values and computer models. In AI research, uncertainties in data and models have been important research topics to improve the performance of intelligent computer systems. From the AI perspective, Stoms [1987] summarized data uncertainties by 1) uncertainty due to variability or error, 2) imprecision due to vagueness and 3) incompleteness due to inadequate sampling frequency or missing variables. In the context of expert systems, uncertainty can be interpreted as uncertainty of how to combine multiple conditions to assert an overall strength of the antecedent part of an *if-then-rule*, uncertainty of the strength of the rules itself in asserting its consequence, and uncertainty in resolving conflicts between different rules or in assigning overall confidence when several rules of varying strength assert the same conclusion. Various theories have been developed to handle both uncertainty in data and in reasoning. Some well studied theories are probability theory [Hartigan, 1983], information theory [Shannon, 1948], the mathematical theory of evidence [Shafer, 1976], and the fuzzy set theory [Zadeh, 1965]. Stoms [1987] concluded that probability theory is designed to handle uncertainty due to randomness. The theory of evidence is best suited to handle uncertainty caused by incompleteness, and fuzzy set theory was developed to handle uncertainty due to imprecision. Nevertheless, these theories can be combined to handle more intricate problems [Bouchon-Meunier et al, 1991].

From the GIS perspective, Bédard [1987] provided another framework for the study of uncertainty. He classified uncertainty into conceptual uncertainty, descriptive uncertainty, locational uncertainty and meta-uncertainty. Conceptual uncertainty refers to the fuzziness in identification of the observed reality, such as determining the vegetation type of an area according to a certain classification scheme. Descriptive uncertainty refers to the uncertainty in the attribute values of an observed reality. This includes the data uncertainty mentioned above. Locational uncertainty refers to uncertainty concerning spatial attributes. This aspect is unique and important to GIS. Meta-uncertainty refers to the uncertainty in the knowledge of the previous three types. All these uncertainties directly affect the performance of GIS. The issue of uncertainty becomes more prominent when GIS is used for planning and decision making. Therefore, uncertainty management should be a high priority in the design of decision support GIS. Recently, some researchers have applied fuzzy set theory in the management of geographical information [Leung 1988; Robinson 1988; Burrough 1989; Sui 1990; Wang et al 1990; Kollias & Voliotis 1991]. Their research demonstrates that fuzzy data modelling techniques reduce information loss and provide more flexible representation of geographical phenomena and relationships. A review of this research can be found in Section 2.6.

1.3 Research Objectives and Thesis Organization

The objectives of this research project are to apply fuzzy data management techniques to handle vagueness and imprecision in attribute values and to the management of uncertainty in information processing in a commercial vector GIS. To achieve these objectives a prototype system has been implemented in conjunction with a commercial vector GIS, the ARC/INFO software package, to evaluate the fuzzy techniques. (ARC/INFO is a registered trade mark of Environmental Systems Research Institute, Redlands California, U.S.A). A forest ecology database was used for illustration purposes.

This research project is part of the Naia Project, a forestry expert GIS project, currently being developed in Calgary, Alberta by Hughes Aircraft of Canada, Alberta Research Council, and the University of Calgary, Department of Geomatics Engineering. The Naia Project aims at developing a decision support system with expert system capability to assist foresters and related professionals in the management of the forest ecology. Several forestry management areas in Northwestern Alberta have been chosen as test areas for the Naia Project. Forest, soil and topographical information was provided by Weldwood Forestry and Canadian Forestry products (Canfor). In this thesis, only part of the Canfor data set was used for the illustration of concepts and techniques.

The organization of this thesis is as follows: Chapter 2 discusses the relevant concepts in the theory of fuzzy sets and reviews the current state of research in the application of fuzzy set theory in GIS. Chapters 3 to 6 describe the specifics of this research. Chapter 3 describes the details of the design, methodology and test data of this research. Chapter 4 explains the representation of attribute uncertainty in a relational database. Chapter 5 presents several fuzzy techniques for modelling and classifying ill-defined objects. Chapter 6 describes the application of fuzzy set theory to spatial operations. Conclusions and recommendations are presented in Chapter 7.

CHAPTER 2

Fuzzy Set Theory and Its Application in GIS

Inappropriate data representation is one of the many sources of uncertainty in GIS. Poor representation can cause misinterpretation and loss of information, but conventional database systems have little capability to represent imprecision or ambiguity in attribute values. This chapter introduces the concept of fuzzy sets and reviews some applications of fuzzy sets in data modelling and uncertainty management in GIS.

2.1 Fuzzy Sets

Information representation and reasoning are two fundamental research areas in artificial intelligence. Due to the underlying vagueness of knowledge and inexactness of human reasoning, two-value and multi-value logic are too precise and too limited to model ill-defined systems such as those for economic forecasting and landuse classification. In a search for better tools for information representation, Zadeh developed the *fuzzy set theory* [Zadeh 1965].

The theory of fuzzy sets is a mathematical theory developed from conventional set theory. It provides a formal and consistent way to represent and process inexact information and vague concepts. In a conventional or non-fuzzy set, such as the set of soil types, a sample plot must either be a member or not a member of a soil type. However, in a fuzzy set a sample can be a partial member of a soil type. The degree of belonging of each element to a set is indicated by a membership grade, which is usually a real number ranging from 0 to 1. The higher the membership grade, the more an element belongs to the set. A fuzzy set is made up of a set of ordered pairs. Each pair consists of an element from the universe of discourse and a membership grade. Consider a fuzzy set for the linguistic term *hilly*, which can be expressed as a function of the percent gradient of the slopes,g :

$$\text{hilly}(g): \begin{cases} 0 & \text{if } g = 0\% \\ \dfrac{g}{35\%} & \text{if } 0\% < g < 35\% \\ 1 & \text{if } g \geq 35\% \end{cases}$$

While slopes with gradients greater than or equal to 35% are best described as hilly (indicated by a membership grade of 1.0), slopes with gradients between 15% and35 % can be described as hilly to some degree.

Fuzzy set theory incorporates conventional set theory as a special case. Thus the membership grade of an element in a conventional set is either 0 or 1. The mathematical definition of a fuzzy set, A, is as follows:

$$\{[x_i/\mu_A(x_i)]\} \; \forall \; x_i \in U \qquad (2.1)$$

where "$\forall$ x" denotes "for all x" and "/" is a separator to separate a set element (left) and its membership grade (right). U is the universe of discourse, x_i is an element of U, and μ_A is a membership function of x_i, which maps x_i into $\mu_A(x_i)$ in an ordered membership set, M. If M ranges from 0 to 1, the set is called a normalized fuzzy set. The universe of discourse, U, is an ordinary set of discrete elements and fuzzy sets are subsets of this discrete universe. Figure 2.1 is a graphical representation of a normalized fuzzy set.

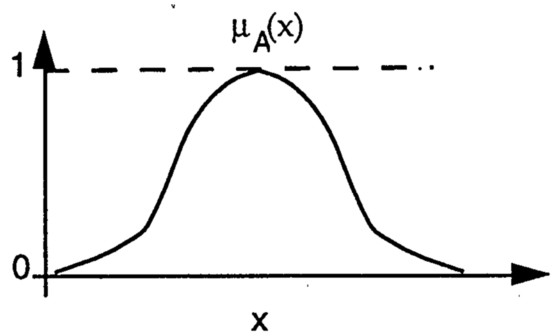

Figure 2.1 Graphical representation of a normal fuzzy set

The membership function of the fuzzy set A, μ_A, defines the degree to which an element belongs to the set. Membership functions are often derived empirically and are context dependent. Some researchers believe that the derivation of membership functions is crucial in fuzzy information processing and that the lack of simple and generally acceptable methods to build membership functions cause it to compare less favorably with other techniques [Kruse et al, 1991]. Turksen [1986] noted that membership functions can be determined either normatively or empirically [Turksen, 1986]. The normative approach is commonly used for deriving membership functions for linguistic values because imprecisions inherent to these values are subjective and, thus, should be defined by the system users. However, this lack of objectivity sometimes raises the concern over the scientific merit of fuzzy set theory. The empirical approach follows the objective experimental procedures of the scientific methods found in measurement theory [Krantz et al, 1971] but little work has been done on the empirical derivation of membership functions. Membership functions used in many fuzzy systems are based mainly on expert knowledge and/or statistics. Despite the lack of scientific foundation, many fuzzy systems have demonstrated satisfactory performance when compared with two-valued logic systems [Chatterji, 1985; Sui, 1990].

2.2 Fuzzy Set Operators

This section provides the original definitions of some fuzzy operations given by Zadeh [1965]. The Min and Max operators are the basic tools used for the aggregation of fuzzy sets. Other operators used in this project are introduced at the appropriate sections.

Let A and B be two fuzzy subsets of the universe, U:

Inclusion:

A is included in B if and only if $\forall x \in U$

$$\mu_A(x) \leq \mu_B(x) \tag{2.2}$$

Equality:

A and B are equal iff $\forall x \in U$

$$\mu_A(x) = \mu_B(x). \tag{2.3}$$

Complement:

Let B be the complement of A, $\forall x \in U$ and $M = [0,1]$,

$$\mu_B(x) = 1-\mu_A(x). \tag{2.4}$$

Intersection:

$$\mu_{A \cap B}(x) = \mathrm{Min}(\mu_A(x), \mu_B(x)). \tag{2.5}$$

Union:

$$\mu_{A \cup B}(x) = \mathrm{Max}(\mu_A(x), \mu_B(x)). \tag{2.6}$$

Depending on the problem situations, other definitions of the intersection and union operators have been presented [Dubois & Prade, 1985; Leung, 1988]. However, Bellman & Giertz [1973] proved that the Min and Max operators are not only natural but also the only operators that possess all of the following properties:

Commutative:

$$\begin{aligned} X \cup Y &= Y \cup X \\ X \cap Y &= Y \cap X \end{aligned} \tag{2.7}$$

Associative:

$$\begin{aligned} X \cup (Y \cup Z) &= (X \cup Y) \cup Z \\ X \cap (Y \cap Z) &= (X \cap Y) \cap Z \end{aligned} \tag{2.8}$$

De Morgan's Laws:

$$\begin{aligned} X \cup (Y \cap Z) &= (X \cup Y) \cap (X \cup Z) \\ X \cap (Y \cup Z) &= (X \cap Y) \cup (X \cap Z) \end{aligned} \tag{2.9}$$

Non-decreasing:

$\mu_{X \cup Y}(x)$ OR $\mu_{X \cap Y}(x)$ cannot decrease if $\mu_X(x)$ or $\mu_Y(x)$ increases. (2.10)

Strictly increasing:

$$\text{if} \quad \mu_X(x_1) = \mu_Y(x_1) > \mu_X(x_2) = \mu_Y(x_2) \text{ then}$$

$$\mu_{X \cup Y}(x_1) > \mu_{X \cup Y}(x_2)$$

$$\mu_{X \cap Y}(x_1) > \mu_{X \cap Y}(x_2) \qquad (2.11)$$

Consistency:

$$\mu_X(x) = 0 \text{ and } \mu_Y(x) = 0 \text{ then } \mu_{X \cup Y}(x) = 0$$

$$\mu_X(x) = 1 \text{ and } \mu_Y(x) = 1 \text{ then } \mu_{X \cap Y}(x) = 1 \qquad (2.12)$$

2.3 Similarity Relations and Similarity Matrices

Let x and y be arbitrary elements of a scalar domain U, S(x,y) is a similarity relation which exhibits the following properties [Zemankova-Leech and Kandel, 1985]:

Symmetry:

$$S(x,y) = S(y,x) \qquad (2.13)$$

Reflexivity:

$$S(x,x) = 1 \qquad (2.14)$$

Max-Min Transitivity:

$$S(x,z) \geq \text{Max}_y[\text{Min}(S(x,y), S(y,z))] \qquad (2.15)$$

The property of symmetry ensures the degree of similarity is the same regardless of the order of the pair. However, in spatial analysis, there are cases where symmetry is inappropriate. These include migration pattern, traffic flow volume, and commodity flow [Leung, 1988]. Reflexivity meets the logical assumption that an element is totally similar to itself. The Max-Min transitivity states that the similarity between two elements x and z, S(x,z), in the universe of discourse should be at least as high as the lower of the S(x,y) and S(y,z).

A similarity matrix can be used to express the degree of similarity between pairs of elements of a similarity relation. Table 2.1 shows a similarity matrix of soil drainage. To illustrate the concept of Max-Min transitivity using Table 2.1, S(Rapid,Imperfect) should be at least as high as the lower of S(Rapid,Well) and S(Well,Imperfect). That is, S(Rapid,Well) must be greater than or equal to 0.2, and one cannot assign a value smaller than 0.2 to (Rapid, Imperfect) even though Rapid and Imperfect share no similarities. Shenoi & Melton [1989] commented that Max-Min transitivity is a very restrictive constraint, and sometimes may be counter-intuitive for certain domains. Zemankova-Leech and Kandel [1985] also noted that no generally acceptable transitivity rules can be easily established; therefore, the Max-Min transitivity property is not always enforced in fuzzy database systems.

	Rapid	**Well**	**Imperfect**	**Poor**
Rapid	1.0	0.3	0.2	0.2
Well	0.3	1.0	0.2	0.2
Imperfect	0.2	0.2	1.0	0.3
Poor	0.2	0.2	0.3	1.0

Table 2.1 A similarity matrix for soil drainage.

2.4 The Extension Principles

The extension principles in fuzzy set theory define the membership grades of the fuzzy elements when mapping from one universe to another [Zadeh,1975]. They are defined as follows:

Extension Principle I:

Let A be a discrete fuzzy set in a universe, U_1, f is a mapping function which maps elements in A into another universe U_2, then the fuzzy set B = f(A) in U_2 is defined by:

$$B = f(A) = \cup_x f(x)/\mu_A(x). \tag{2.16}$$

Example:

Let $f : x \to 4*x$

$A = \{3/0.8, 4/1.0, 5/0.9\}$

$B = \{12/0.8, 16/1.0, 20/0.9\}$

Extension Principle II:

Let U be a Cartesian product of universes $U = U_1 \times U_2 \times \ldots \times U_n$, and $A_1, A_2, \ldots, A_n$ be n discrete fuzzy sets in $U_1, U_2, \ldots U_n$, respectively. f is a mapping function from U to another universe V and $y = f(x_1, x_2, \ldots x_n)$, then a fuzzy set B in υ is defined as:

$$B = \cup_y \, y/\mu_B(y)$$

$$\begin{aligned} \mu_B(y) &= \{\text{MaxMin}_y\{\mu_{A_1}(x_1), \ldots, \mu_{A_n}(x_n)\} && \text{for } f^{-1}(y) \neq 0, \\ \mu_B(y) &= 0 && \text{otherwise} \end{aligned} \tag{2.17}$$

and $f^{-1}(y)$ is the set of points in U which are mapped into V by f.

Example:

$$A = \{4/0.8, 5/1.0, 6/0.9\}$$
$$C = \{5/0.7, 6/1.0, 7/0.8\}$$

$$\begin{aligned} A*C &= (4*5)/\mathrm{Min}(0.8,0.7),\ (4*6)/\mathrm{Min}(0.8,1.0),\ (4*7)/\mathrm{Min}(0.8,0.8), \\ &\quad (5*5)/\mathrm{Min}(1.0,0.7),\ (5*6)/\mathrm{Min}(1.0,1.0),\ (5*7)/\mathrm{Min}(1.0,0.8), \\ &\quad (6*5)/\mathrm{Min}(0.9,0.7),\ (6*6)/\mathrm{Min}(0.9,1.0),\ (6*7)/\mathrm{Min}(0.9,0.8) \\ &= (20/0.7, 24/0.8, 25/0.7, 28/0.8, 30/1.0, 35/0.8, 36/0.9, 42/0.8) \end{aligned}$$

2.5 The Possibility Theory

2.5.1 Possibility versus Probability

Probability and possibility are two related terms used to describe uncertainty. As Leung [1988] stated, probability theory is a tool to study randomness and possibility theory is the tool to study imprecision. While the uncertainty in probability models is caused by randomness, the uncertainty in possibility models is due to the incompleteness and imprecision in information, which prevent the drawing of indisputable conclusions. Zadeh [1978] noted that while probability and information theory measure the quantity of information, possibility and fuzzy set theory study the semantics of information. The difference between the two can be elucidated by the following example which shows the probability and possibility distribution of a fair die:

	1	**2**	**3**	**4**	**5**	**6**
P(x)	0.167	0.167	0.167	0.167	0.167	0.167
π(x)	0.167	0.167	0.167	0.167	0.167	0.167

However, if the die is loaded, the possibility of having any number is still the same because still there are six faces on the die, but the probability distribution shows that the die is loaded to give a higher occurrence of 1.

	1	2	3	4	5	6
P(x)	0.5	0.1	0.1	0.1	0.1	0.1
π(x)	0.167	0.167	0.167	0.167	0.167	0.167

When using fuzzy sets to represent possibility functions, π(x) is always expressed as a normalized fuzzy set. Therefore we have the following representation:

	1	2	3	4	5	6
P(x)	0.5	0.1	0.1	0.1	0.1	0.1
π(x)	1.0	1.0	1.0	1.0	1.0	1.0

In the example π(x) is undoubtedly 1 for all six faces. For events which the outcomes are not as clear-cut as the faces on a die, the possibility can be expressed as any real number from 0 to 1. For instance, the possibility of a person being called *young* depends on the definition of the concept young. Because the transition from *young* to *not young* is gradual, π(x) can be expressed as continuous function of the age of a person. Since much of the information humans use for decision making is possibilistic in nature, Zadeh [1978] believes that possibility theory should be the framework for information analysis.

2.5.2 Randomness versus Imprecision

Another pair of concepts worth mentioning is randomness and imprecision. In a mathematical sense, randomness is characterized by chaotic, stochastic and unpredictable behavior. Randomness is the uncertainty arising from the unpredictable element in a deterministic model. Fuzziness is the uncertainty arising from the lack of precise information. Leung [1988] encapsulated the two concepts precisely as follows:

> ".... randomness is an uncertainty resulting directly from the breakdown of deterministic cause-effect relationships, and imprecision is an uncertainty resulting directly from the breakdown of the law of the excluded-middle."

As with probability and randomness being the foundation of statistics, possibility and fuzziness are the foundation of the fuzzy information processing. Information modelling in computer systems has become more sophisticated and both possibility and probability theories share important roles in modelling uncertainty in this imperfect reality.

Possibility theory can be expressed with the concepts and tools in fuzzy sets. The membership function μ_A acts as a *fuzzy restriction* to restrain the values that may be assigned to x. Thus the proposition "x is A" is postulated to be equal to $\mu_A(x)$, indicating the possibility or the truth value of "x is A". Therefore in this context, μ_A can be interpreted as π_A [Zadeh, 1978].

2.6 State of Current Research on Application of Fuzzy Set Theory to Uncertainty Management in GIS

Research on the application of fuzzy set theory to manage uncertainty in geographic information is limited due to the short history of GIS. Nevertheless, some interesting work has been done in cartographic modelling, database management systems and approximate reasoning in expert GIS. The following is an overview of the current research on these topics.

2.6.1 Modelling Fuzziness in Data

Due to the complex nature of many natural phenomena, most geographic data, especially those in natural resource management, are inexact and context dependent. In addition, geographical data are often categorized into classes with qualitative values. Furthermore, because geographical data are usually a scaled representation of the real phenomena, they are inherently inexact. Converting fuzzy data to fit into the conventional exact data processing framework is inappropriate.

To better understand the nature of fuzzy data, Robinson [1988] (following Sack et al [1983]) summarized fuzzy data models into four cases: nonfuzzy schema/nonfuzzy data, nonfuzzy schema/fuzzy data, fuzzy schema/nonfuzzy data and fuzzy schema/fuzzy data. The first case is the conventional database model in which domains are discrete and data values are exact. Boolean logic is sufficient to handle this type of data. As we will see in the sequel, this case can also be treated properly within a fuzzy data model since fuzzy set theory contains conventional set theory. Nonfuzzy schema/fuzzy data has discrete domains but data values cannot be captured exactly. Fuzzy schema/nonfuzzy data refers to data models with inexact domains but where exact data values can be obtained. Fuzzy schema/fuzzy data is the most generalized model in which both domains and data values cannot be expressed exactly. Little work has been done in modelling fuzziness at the data level.

Two models proposed for fuzzy schema/nonfuzzy domain are reviewed in Robinson's paper [1988]. The *import semantic* model [Baldwin & Zhou, 1984] can be used to store the degree of fuzziness in data values. This model simply attaches a column next to each attribute to store the corresponding membership grade (Table 2.2).

The *similarity relation model* [Buckles and Petry 1982] makes use of the similarity matrices to represent the fuzziness between linguistic terms. The elements in a similarity matrix can be interpreted as the degree of overlap between the meanings of any two terms. Although Robinson used this model to represent nonfuzzy schema/fuzzy data, it seems to be more appropriate to represent fuzziness in schema because the similarity relation does not provide information on fuzziness of the individual values as those provided by the import semantic model, but it indicates the similarity of the definitions of the attribute terms.

Most of the current research emphasizes the modelling of fuzziness in cartographic models rather than the fuzziness of the data themselves. One reason could be the lack of awareness in the importance of data modelling. Another reason could be that users are accustomed to absorbing uncertainty in data. The classification of fuzzy data serves as a framework for further investigation into modelling fuzzy data in a GIS context.

Sample ID	Drainage	μ
1	well	0.7
2	imperfect	0.9
3	well	0.8
4	rapid	0.6
5	rapid	1.0
:	:	:

Table 2.2 *Drainage* represented as the Import semantic model.

2.6.2 Fuzzy Cartographic Modelling

Fuzzy data models have been applied to soil analysis [Burrough 1989], suitability analysis [Wang et al 1990], economic regionalization [Leung 1988], and modelling gradual change of urban land values [Sui 1990].

Sui [1990] studied urban land value for the city of Jining in the P.R. China. Conventional techniques classify lands by a multi-section linear function, in which a constant threshold is set for each section. For instance, 0 to 500 meters from a shopping center would be classified as first class land, 501 to 1000 meters, second class, and so on. This results in a step function, rather than a gradual change, in land value. To model the gradual change, Sui first represented the study area in raster format. He then created a set of fuzzy matrices, one for each attribute. The elements of each fuzzy matrix represented the membership grade of each pixel in each class of that attribute. Using the fuzzy operations, a final matrix with quantitative values showing the degree of each pixel belonging to a certain category. Compared with the evaluation done with conventional methods, both methods gave similar results with respect to the general pattern, but the fuzzy technique provided details about the graduated evaluation.

Another approach to a similar problem is presented in Leung [1988]. Leung recognized that most phenomena vary over space in a more or less continuous manner. Therefore, he defined a set of regionalization procedures using fuzzy set theory which allow for smooth transition from one region to another. He demonstrated the versatility of this procedure in the classification of climatic regions. Expressing attributes like *warm* and *humid* as fuzzy sets, Leung determined the fuzzy boundaries of the regions using fuzzy operations. The objective of regionalization is to determining the *edge*, the *boundary* and the *core* of a region. An edge is the outermost boundary of a region beyond which the area is not likely to be classified as the region. The core depicts the area of a region whose characteristics are most compatible with the definition of the region. The area between the core and the edge is the boundary. Because the location of a boundary can be fuzzy, a boundary is represented by a gradient rather than a line. It can be interpreted as a zone within which all points are more or less compatible to the characteristics of a region. Thus if we consider a region as a fuzzy set, the points in the core have a membership grade of 1.0, and all points outside the edge have a membership grade of 0.0. The points in the boundary zone have a membership value the range of 0.0 to 1.0. In the event that a precise boundary must be established, an α-boundary can be established by restricting the membership grade to a specific value, α.

Wang et al's suitability study [1990] used fuzzy techniques to classify a study area in Indonesia by its suitability for several farming activities. The authors noted that the physiographic characteristics of an area do not completely match the classification requirements, and thus fuzzy techniques were used. The classification technique used is a pattern recognition method which comprehensively takes into consideration all its characteristics and all the suitability classes. By representing areas as vectors in a feature space and the growing conditions for each crop as *prototype vectors*, the authors calculated the mathematical distances (Euclidean distance) between the prototype vectors and the area vectors. The greater the distance, the less suitable is the area for the crop being considered. Wang concluded that this technique reduced the loss of information and provided indications of the appropriateness of the classification.

In Wang et al's research, no data uncertainty was considered. In fact, Boolean logic was used in the comparison of the area vectors with the prototype vectors. For instance, annual average temperature for class S_1 is 25 to 29 degree Celsius and for class S_2 is 30 to 32 degree Celsius. If a area measured 29 degree, it would match S_1. However, a one degree difference would fail to match with S_1. This could be a potential problem particularly when average temperature is used. Using fuzzy logic in the matching process could provide more insight to the classification. Chapter 5 of this thesis discusses the use of fuzzy logic operators to perform feature matching.

Burrough [1989] demonstrated the use of fuzzy data models to study soil condition in several study areas in Venezuela and Kenya. He expressed the uncertainty in the eleven soil attributes as fuzzy sets. The membership functions represent the soil specialists' knowledge and experience in soil characteristics in the study areas. Based on the membership functions, Burrough calculated a membership value for each attribute, which was interpreted as the possibility of a particular attribute value being found in a certain layer of a soil profile. To demonstrate the robustness of fuzzy operations, Burrough formulated retrieval criteria which involved all eleven attributes in the data set. He showed that using Boolean logic and exact criteria to retrieve information resulted in a futile operation. The problem can be visualized by overlaying eleven raster binary maps and finding their intersection. Since attributes tend to cancel each other out, the final map is more or less blank. However, with fuzzy operations, retrieval is done by degree of match rather than by exact match on values. The final map showed degree of match of each pixel with respect to the retrieval criteria. Users can then decide which areas meet the criteria better.

Burrough's work demonstrated that when queries cannot be stated exactly or queries are too complex, fuzzy operations provide more informative results than when using Boolean operations. In this example, only numerical attributes were used. Yet, it is not uncommon to find qualitative attributes in geographical databases. Handling uncertainty in qualitative attributes presents a greater challenge than handling quantitative attributes. Chapter 4 presents a detailed discussion on modelling qualitative uncertainty.

2.6.3 Fuzzy Database Management System

A more complete prototype database management system was developed by Kollias and Voliotis [1991]. FRSIS, a prototype soil information system with fuzzy retrieval capabilities, provides formal definition and manipulation commands to manage incomplete and imprecise data and queries. In addition to common relational database functions, FRSIS provides some new data type definitions and fuzzy operation commands. Three data types can be defined in FRSIS. They are 1) normal, i.e. those provided by INGRES (a commercial database management system), 2) possibility distribution of normal type domain value, and 3) membership in a set. Additional manipulation commands include creation of fuzzy relations, definition of fuzzy sets, normalization of fuzzy sets, fuzzy retrieval, deletion, replacement, fuzzy qualifiers and fuzzy comparison operators. The grammar of these commands is described in their paper. The following is an example of retrieval using fuzzy conditions:

> 'Retrieve the code number and the number of horizons of the soil profiles that are ***quite shallow*** and have been developed on ***flat*** alluvial terraces'.

FRSIS performs this operation according to the definitions of ***quite*** *shallow* and *flat* in FRSIS. Again, the membership functions for *quite shallow, flat* and the like were defined by users.

FRSIS was built on top of the INGRES DBMS on a UNIX platform. It extends the INGRES *QUEL* DBMS language by incorporating a language pre-processor to QUEL, hence, created a new DBMS language called *FQUEL*. The pre-processor translates all fuzzy operations in FQUEL into standard QUEL commands which can then be processed in INGRES.

With the capabilities to manage incomplete and imprecise data and allowing users to express their subjective view on the data, FRSIS provides a more human-like approach to information retrieval. Yet, to meet the needs of GIS users, spatial analysis and mapping capabilities must be added. One approach is to interface FRSIS with other software that provides these functions.

2.6.4 Approximate Reasoning Using Fuzzy Logic

Fuzzy set theory was developed with a goal to improve the performance of artificial intelligence in reasoning. Leung's [1990] research focus is on the application of fuzzy set theory to build an expert GIS with a high level of intelligence, and he proposed a framework for a rule-based expert GIS based on fuzzy logic. Combining fuzzy set theory with the rule bases, the fuzzy inference engine can perform inferences such as the following:

Major premise: If the temperature is high, then the pressure is low.

Minor premise: The the temperature in area X is *quite* high.

Approximate conclusion: The pressure in area X should be quite low.

In rule-based systems that use Boolean logic, the production rule in the major premise is activated only when the database entry equals *high*. To be able to handle all possible conditions, the set of production rules must exhaust all possible values for the attribute *temperature*. This could be very inefficient when there is a large set of attribute values. Furthermore, conventional systems cannot handle database entries with fuzzy values. In rule-based systems that use fuzzy logic, the production rule is triggered when the temperature of X is roughly equals to *high*. This depends on the definition of *high* and its relation to other terms describing temperature. For instance, *very high* and *quite high* can be considered *high*. The production rule in the major premise is sufficient to handle all these data base entries. Hence, approximate reasoning using fuzzy logic is a more flexible approach to rule-based systems.

Nevertheless, one should note that the performance of fuzzy inference depends very much on the knowledge stored in the knowledge base and the validity and appropriateness of the membership functions. The advantage of using fuzzy logic is that users are not limiting ourselves to finding the perfect solution but to obtaining a set of acceptable solutions.

CHAPTER 3

Research Project Design and Test Data

The objective of this research is to investigate the effectiveness of fuzzy set theory in the management of uncertainty in a commercial GIS environment. This prototype system uses fuzzy set theory-based techniques to handle the fuzziness in attributes, objects and operations. Fuzziness in attributes refers to the imprecision and vagueness in quantitative values and qualitative terms. Fuzziness in objects relates to the indefinite descriptions of the entities being studied. Fuzzy operations are operations that involve operands with fuzzy values. Examples are *approximately equal*, *much less than* and *much more than*. A forest ecological database has been chosen to illustrate these functions in this prototype system.

3.1 Goals of the Prototype System

This prototype system was designed as an add-on module to a commercial GIS which is capable of the following functions:

1. Representing imprecise values and qualitative terms using fuzzy sets and similarity matrices.
2. Modelling fuzzy entities and fuzzy attributes.
3. Processing queries with fuzzy criteria.

4. Performing fuzzy ecological classification.

5. Extending spatial operations to handle fuzzy attributes.

The choice of a forest ecological database for this prototype system is appropriate because the database contains both quantitative and qualitative attributes which can be modelled by fuzzy sets and similarity matrices. In addition, because the classification scheme was derived from a limited number of ground sample plots, it can only serve as a rough indicator to the classification of ecosystem association. Fuzzy techniques surpass conventional techniques in handling databases with incomplete information.

Two terms in forest ecology require formal definition. In the *Field Guide to Forest Ecosystems of West-Central Alberta* [Corns & Annas, 1986], *ecoregions* are geographical areas that have a distinctive, mature ecosystem plus specified edaphic variation as a result of a given regional macroclimate. An *Ecosystem association* is an abstract taxonomic unit within an ecoregion. It includes all land areas with the potential of supporting plant communities with similar successional development belonging to the same plant association. For each ecosystem association, a set of forest management guidelines is provided. This research project applies fuzzy set theory to the classification of forest land into ecosystem associations so that foresters can evaluate their management and planning activities according to the suggestions in the field guide. Forest ecosystem classification serves as an important guideline for the planning, harvesting and regeneration, release and tending stages of forestry management. The standardized classification also facilitates communication among various specialists in related fields.

3.2 Functions of the Prototype System

This prototype system consists of four modules: the fuzzy attribute definitions module, the fuzzy query module, the ecosystem classification module and the fuzzy spatial operation module. The following subsections describe the components of a fuzzy relational database and the functions of the modules.

3.2.1 The Fuzzy Relational Database Model

This prototype system follows the concept of a Fuzzy Relational Database Model (FRDB) presented in Buckle & Petry [1982] and Zemankova-Leech & Kandel [1985]. The database consists of three components: the value database (VDB), the explanatory databases (EDB) and a set of interpretation rules. The VDB is the "normal" database which stores the original values of the attributes. The EDB stores the fuzzy attributes and their meanings in the fuzzy attribute dictionary. For instance, the attribute "slope" can be redefined as a fuzzy attribute with qualitative terms such as *flat, gentle and steep*. The meaning of these terms, i.e. the membership functions, are stored in the fuzzy attribute dictionary. The EDB contains a fuzzy attribute dictionary, a modifier dictionary and similarity matrices. The fuzzy attribute dictionary stores the name and the membership functions of the fuzzy attributes. The meanings of fuzzy attributes can be changed slightly by modifiers such as *very* and *approximately*. The names and the functions of these modifiers are stored in the modifier dictionary. The interpretation rules are algorithmic procedures used to compute the membership grades of the fuzzy attributes based on the values in the VDB. Figure 3.1 shows the components of the database.

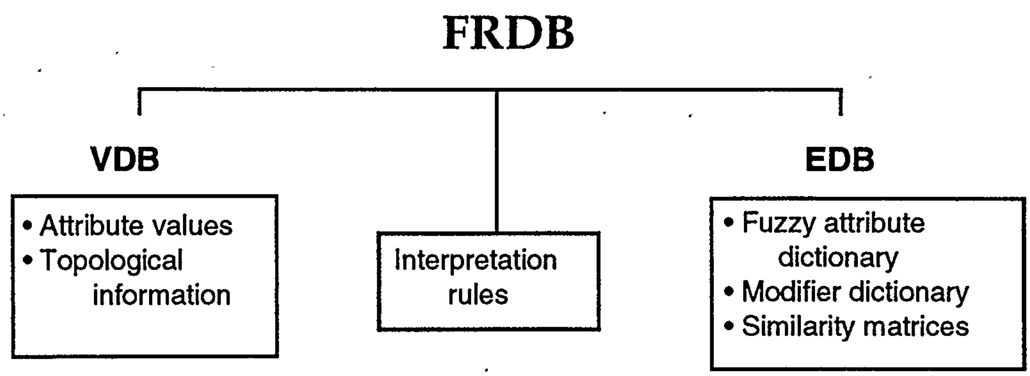

Figure 3.1 Organization of a fuzzy database.

3.2.2 Fuzzy Attribute Definition

The fuzzy attribute definition module was developed from the concept of *linguistic variables* [Zadeh, 1973] described in Chapter 4. This module allows users to define fuzzy attributes as characteristic functions (membership functions). Because the meaning of these qualitative terms are often context dependent and sometimes subjective, the system allows users to define and change the definitions of the terms without altering the original values in the database. For illustration purposes, only triangular or trapezoidal-shaped functions were used in this prototype system. These functions can be represented by at most four parameters and require less computation. To represent these functions in the fuzzy attribute dictionary, only the function parameters are stored. An example of a fuzzy attribute is *slopes*. In the Canfor data set *slopes* are described by three terms, *flat, moderate* and *steep*. *Slopes* with gradients less than 30% are labelled *flat*, 31% to 45% are labelled *moderate*, and over 45% are labelled steep. This is similar to the categorisation precdeure using conventional techniques. However, conventional techniques impose a clear-cut boundary between classes which usually result in loss of information after categorisation. Figure 3.2 illustrate the difference between conventional representation and fuzzy representation. Fuzzy techniques allow the expression the gradual transition from one term to another by using different shapes of characteristic functions.

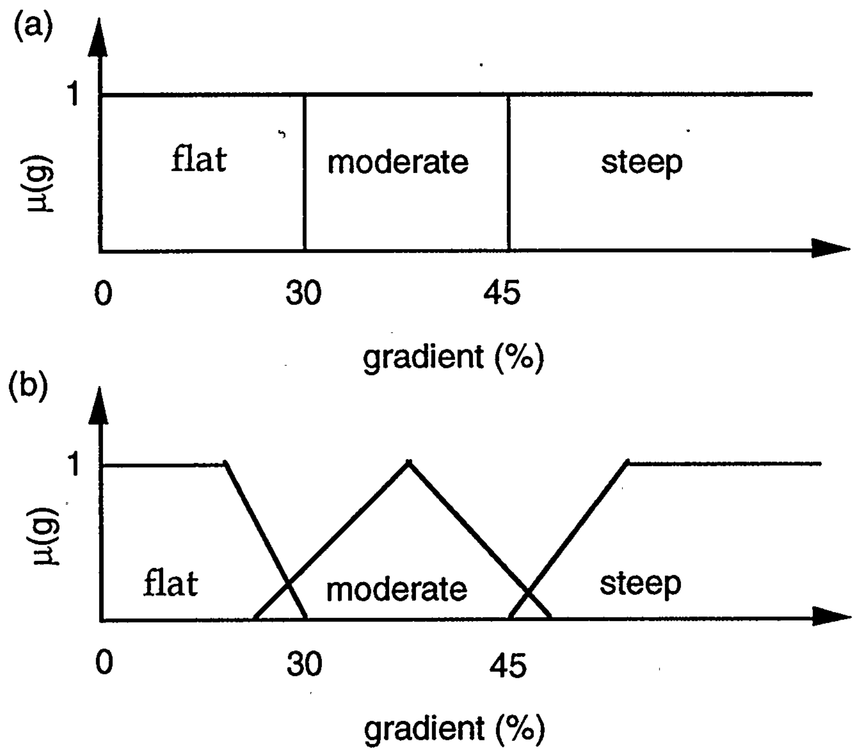

Figure 3.2 (a) Conventional representation of *slope*. (b) Fuzzy set representation of *slope*.

In addition to defining the meaning of the linguistic terms, users can also associate modifiers with these terms. A modifier alters the meaning of a term by changing the values of the parameters thus generating new membership functions. When defining a modifier, the user will be asked to input the name of the modifier and its effect on the parameters of the functions.

3.2.2 Fuzzy Query

In conventional databases, queries are formulated with conventional logic operators like *equal, less than*, and *greater than*. However, there are times when the user wants to perform less precise queries so as to accommodate less-than-perfect matches or when the user does not know exactly what to retrieve. This system allows him/her to retrieve information using fuzzy criteria. For instance, in a conventional system, if the user wants to retrieve all areas that are *approximately* 800 to 900 meters in elevation, he/she has to provide exactly the elevation range to be retrieved. However, *approximately* is only a measure of degree, thus this system not only retrieves the records that have elevation between 800 to 900 meters, but also the records that fall close to the range specified. The system assigns 1.0 to the degree of match for areas that are within the range and a smaller value to areas that are outside the specified range. Users can also specify the degree of match at retrieval to limit the number of cases being retrieved.

The fuzzy operators implemented in this project are *approximately equal* (~=), *much less than* (<<), and *much greater than* (>>). Each of these operations takes two operands: a precise value and a fuzzy value. The precise operand comes from the entry in the VDB, and the fuzzy operand comes from the retrieval criteria. The records retrieved are stored in a separate database file for further processing.

3.2.4 Ecosystem Classification

In forest management, forest lands are classified into ecosystem associations so as to facilitate management and planning. To facilitate classification, a set of attributes are identified in the field guide [Corns & Annas, 1986]. However, due to the complexity of the ecosystem, the values of the characterizing attributes given in the field guide are only approximate values. Therefore, these associations were treated as fuzzy objects and a fuzzy classification procedure has been developed to perform ecosystem classification. Due to its imprecise definition, it is likely that a forest area will not match the definitions of any associations perfectly. Thus, the fuzzy classification techniques evaluate the degree of match of a forest to each of the ecosystem associations. The result of this fuzzy classification provides several possible interpretations with their respective membership grades. Different operators were used to determine the classification, and the comparison of the results are described in Section 5.4 and 5.5, respectively.

To perform classification, users have first to provide the definitions of the ecosystem associations and these definitions are kept in a database file. Users then invoke the fuzzy classifier. Upon termination of the classification process, the results and the degree of match of each attribute will be saved in a separate database file for further analysis.

3.2.5 Fuzzy Spatial Operations

Conventional spatial operations such as the overlay and the polygon merging functions use Boolean logic to process the overlay or merging criteria. These operations can be extended to handle fuzzy attributes. In this prototype system, the fuzzy overlay and polygon merging functions were implemented as extensions to the operations provided in a commercial GIS. These fuzzy overlay operation propagates the membership grades of the attributes from two maps to the resultant map. The fuzzy merging function consolidates adjacent polygons that have the same attribute value with a membership grade greater than or equal to a user-specified level. For instance, if the adjacent polygons both have sandy texture but the membership grades are 0.6 and 0.8, the two polygons will be merged only if the user-specified level is 0.5 or smaller. Detailed description of these two fuzzy spatial operations is presented in Chapter 6.

3.3 Software and Hardware Requirements

This prototype system has been implemented on a PC-386 micro-computer. PC ARC/INFO version 3.4D was used in conjunction with dBASE IV database management system version 1.1 (dBASE IV is a registered trademark of Ashton-Tate Corporation, Torrance, California U.S.A), with PC ARC/INFO serves as the host system. The system is driven by a multi-level bar menu, which is written in ARC's Simple Macro Language (SML). Due to the limited programming ability of SML, all four modules are written in the dBASE programming language. Because this version of ARC/INFO stores both topological and attribute information in dBASE files, access to data files and topological information is straightforward. Although using the dBASE programming language to implement these fuzzy techniques is less efficient when compared with other languages like the C programming language, it was chosen because of the easy access to data files. Furthermore, it is also desirable to limit the number of software packages for implementation so as to maintain generality in the computing environment and to avoid interface problems.

3.4 Study Area and Sources of Data

The study area for this research project is located approximately 120 km south east of Grande Prairie, Alberta, Canada (Figure 3.3). A pilot project for ecological classification in this area was carried out by the joint effort of the Canfor, Forestry Canada and Alberta Research Council. The data were collected in the summer of 1990 using both aerial photographs and data from ground surveys. Statistical analyses and interpretation were performed by Forestry Canada. All attribute data provided are stored in dBASE format. Digital maps of the ecological boundaries and point samples were provided in Intergraph IGDS format and they were subsequently translated into PC ARC/INFO format.

The attribute data file of ground survey samples contains 151 attributes which include soil characteristics, forest cover and understory vegetation. Figure 3.4 shows a predictive site mapping form used for data collection. To reduce the complexity of the classification process, forest ecologists have identified several important attributes to be used in this prototype system. Fifty-nine samples points were assigned an ecosystem association. These samples were used in the fuzzy classification procedure to establish membership functions and to evaluate the classification results generated by different fuzzy operators.

3.5 Evaluation of the Prototype System

The nature of this research is to demonstrate the usefulness of fuzzy sets in the management of fuzziness in attribute, objects and operations. In this prototype system, little consideration was given to the storage and processing efficiency of the system. The main concern is to evaluate the effectiveness of fuzzy techniques over conventional techniques. To show that fuzzy techniques are effective tools for managing uncertainty, the difference between conventional techniques and fuzzy techniques are compared in various sections of this thesis. The objective is to show that in many situations, fuzzy techniques provide more flexibility than conventional techniques in a natural resource database such as the forest ecological database used in this thesis.

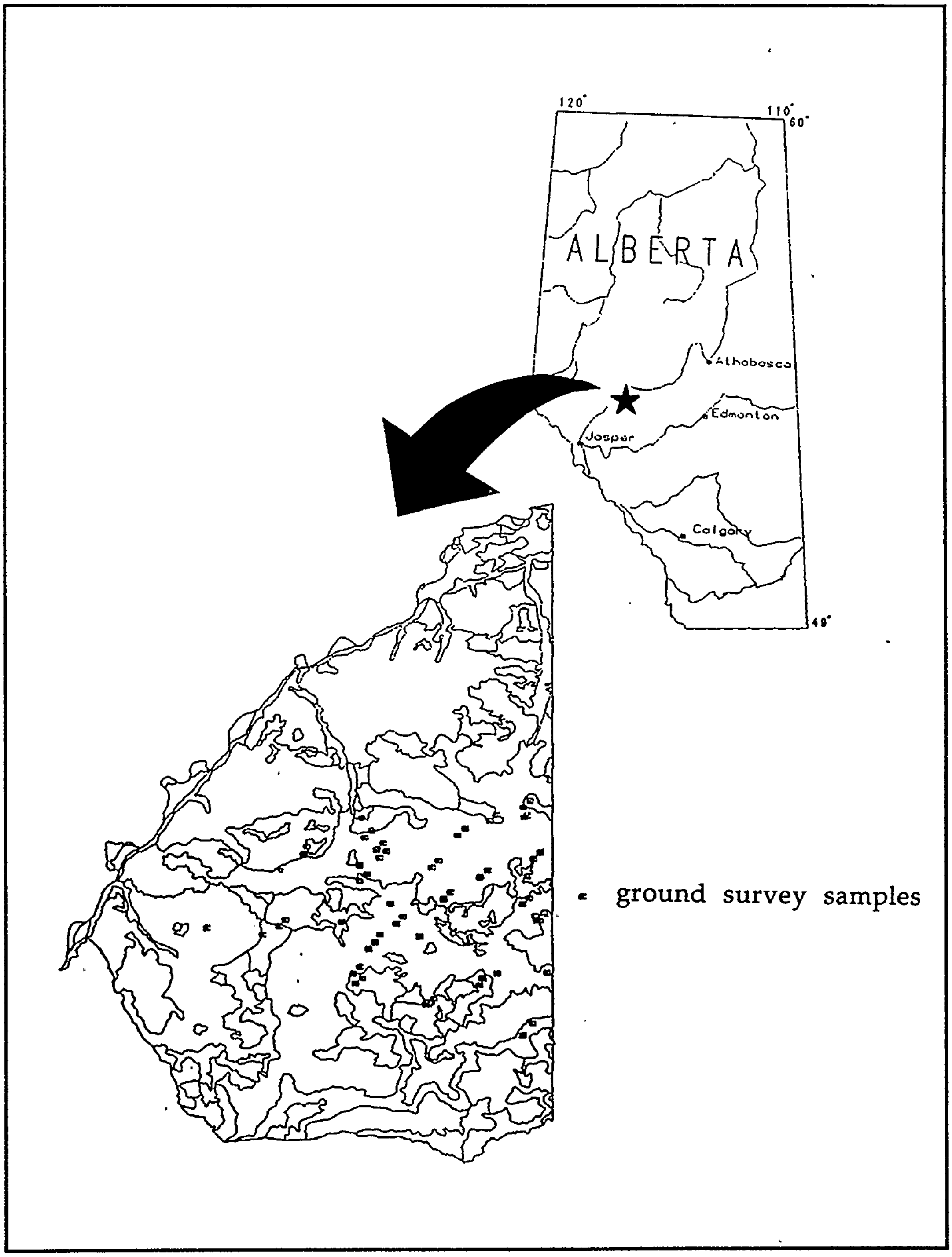

Figure 3.3 Location of the study area.

Predictive Site Mapping Form

Site Info. (Office)
Plot no.: ____
Date: ____
Surveyor: ____

TWP: ____
Range: ____
Meridian: ____
Section: ____
LSD: ____
Mngt Unit: ____
Compart No.: ____
Air Photo No.: ____
Elevation: ____m ____ft
Ecoregion:
- <800m-BMW
- 800-1150m - LBC
- 1150-1300 - UBC
- >1300m - SA

Soil type: ____
Forest cover type: ____

Site Info. (Field)
Ecosystem Assoc.: ____
Soil Type: ____

Slope %: ____
Slope Position: C U M Low T D Lev.
Aspect (degrees): ____

Forest cover type: ____

Soils
Parent Material:

Drainage: VR R W I P VP

Layer 1 text: ____
Layer 1 thick: ____cm

Layer 2 text: ____

Humus Form: ____

Depth org. layer: ____cm

Depth of soil: ____cm

Depth to mottles: ____cm

Depth to gley.: ____cm

Depth to calc.: ____cm

Depth of root: ____cm

Seepage present?: Y N

C.F. Content %: ____

No. Trees:

Tree #	Dom.	Sp.	Ht.	DBH	Age
1					
2					
3					
Mean	—	—			

Vegetation Structure	Sw	Sb	Pl	Pj	Fb	Lt	Aw	Pb	Bw	Other
Tree layer (% for each):										
Adv. growth 2-5m:										
Adv. growth <2m:										

Understory cover class:	Shrub	Herb	Grass	Moss/lichen

Shrub layer cover: (% cover class) a (<1) b (1-5) c (6-20) d (21-50) e (>50)

White Rhododendron	(Rhodalb/137)	Honeysuckle	(Loninv/134)
False Azalea	(Menzgla/135)	Black Currant	(Ribelac/138)
Low-bush Cranberry	(Vibuedu/151)	Red Currant	(Ribetri/139)
Green Alder	(Alnucri/128)	Prickly Rose	(Rosaaci/140)
Red Osier Dogwood	(Cornsto/130)	Wild Raspberry	(Rubuida/141)
Devil's Club	(Oplohor/136)	Thimble Berry	(Rubupar/142)
Labrador Tea	(Ledugro/133)	Buffalo-berry	(Shepcan/146)
Tall Bilberry	(Vaccmem/149)	Bog Cranberry	(Vaccvit/156)
Blueberry	(Vaccmyr/150)		
Kinnikinnick	(Arctuva/152)		

Herb layer cover:

Common Horsetail	(Equiarv/158)	Heart-leaf Arnica	(Arnicor/166)
Meadow Horsetail	(Equipra/159)	Lindley's Aster	(Asteci/167)
Woodland Horsetail	(Equisyl/160)	Showy Aster	(Astecon/168)
Cloudberry	(Rubucha/186)	Bunchberry	(Corncan/170)
Wild Sarsaparilla	(Aralnud/165)	Fireweed	(Epilang/172)
Five Leaved Bramble	(Rubuped/187)	Wild Strawberry	(Fragvir/173)
Running Raspberry	(Rubupub/188)	Twin-flower	(Linnbor/153)
Enchanters Nightshade	(Circalp/ -)	Lily-of-the-valley	(Malacan/178)
Oak Fern	(Gymndry/161)	Tall Mertensia	(Mertpan/179)
Hairy Wild Rye	(Elyminn/197)	Palmate Coltsfoot	(Petapal/183)
Marsh Reed Grass	(Calacan/196)	Bishop's-cap	(Mitenud/180)
Common Pink Wintrgrn.	(Pyroasa/184)		

Moss/lichen layer cover:

Feathermoss	(Pleusch/205)	Knight's Plume	(Ptilcri/207)
Stairstep Moss	(Hylospl/204)	Haircap Moss	(Polycom/206)
Peat Moss	(Sphaspp/208)	Green Dog Lichen	(Peltapt/201)
Reindeer Lichen	(Cladspp/200)		

Notes:

Tiepoints:
Landform:

Sketch:

Figure 3.4 Predictive site mapping form used for field data collection [Corns and Annas, 1986].

CHAPTER 4

Storage and Retrieval of Fuzzy Data

Miller et al [1989] noted that uncertainty reduction and absorption are two common treatments for data uncertainty in geographical databases. Reducing uncertainty in data is the ideal treatment. However, it is often impractical or physically impossible to obtain precise data. Uncertainty absorption assumes data in the databases are in their best possible quality, and users are left to absorb any unexplained errors. These two treatments are by no means the best treatments. A more reasonable approach is to present to the users the available information on data uncertainty so as to allow users to make decisions based on the reliability of the data. This chapter presents the concepts of linguistic variables and fuzzy numbers, which are appropriate to model imprecision in qualitative and quantitative data, respectively [Boy & Kuss, 1986]. Tools for manipulating fuzzy data in a relational database are also discussed.

4.1 The Concept of Linguistic Variables

Linguistic variables are well-defined data structures developed from the fuzzy set theory. They differ from numerical variables in that they contain linguistic terms rather than numerical values. Also linguistic variables are associated with syntactical and semantic rules. A linguistic variable is characterized by a quintuple (X,T(X),U,G,M). *X* is the name of the variable;

T(X) is called the *term-set* of X, that is, the set of linguistic values of X. Each linguistic value is a fuzzy set ranging over a universe of discourse U. G is a set of syntactic rules for generating the term-set, and M is the semantic rule for each term in the term-set. For example, to represent the depth of organic soil as a linguistic variable, the linguistic variable *Organic thickness* as is defined as follows:

X = *Organic thickness*

T(X) = {thin, thick, very thin, very thick, not thin and not thick, ...}

U: Base variable: depth
Range : 0 to 200 cm
G: Atomic term: {thin, thick}
Modifier : {very, not}
Syntax :
g1 = Modifier(atomic term) or atomic terms
g2 = Modifier(T) or T
M: Semantics :

μ_{thin}(depth):

$$\begin{cases} 1 & \text{if depth} < P2 \\ \dfrac{P3\text{-depth}}{P3\text{-}P2} & \text{if } P2 \le \text{depth} < P3 \\ 0 & \text{if depth} \ge P3 \end{cases}$$

P1 = 0, P2 = 20, P3 = 30

μ_{thick}(depth):

$$\begin{cases} 0 & \text{if depth} < P1 \\ \dfrac{P1\text{+depth}}{P2\text{-}P1} & \text{if } P1 \le \text{depth} < P2 \\ 1 & \text{if depth} \ge P2 \end{cases}$$

P1 = 25, P2 = 70, P3 = 200

Very(t) : P1-5, P2-5, P3-5 if t= thin
P1+10; P2+10, P3+10 if t = thick

Not(t) : 1 - μ(t)

The linguistic variable *Organic thickness* is defined by a base variable called *depth*, which is a numeric variable with value ranging from 0 to 200 cm. The term-set T(X) contains the valid terms for *Organic thickness*. T(X), in this case, is an infinite set as explained below. *Atomic terms* in a term-set are the terms that function as single units. Here *thin* and *thick* are two units of *Organic thickness*. The meanings of these units are defined by the membership functions μ_{thin}(depth) and μ_{thick}(depth). The syntactic rule g1 allows a modifier to precede an atomic term to form *a composite term*. Rule g2 further allows a modifier to precede a composite term. By recursively applying rule g2, an infinite term-set can be generated. The meaning of the composite terms are computed by applying the effect of the modifiers to the membership functions of the atomic terms. For example, the modifier *very* shifts the membership function of *thin* 5 cm to the left, and the membership function of *thick* 10 cm to the right. P1, P2, and P3 are parameters of the membership functions as shown in Figure 4.1.

A linguistic variable is called a structured linguistic variable if T(X) and M can be characterized algorithmically. This implies that a structured linguistic variable relies on algorithmic procedures to generate the term-set and to compute the meaning for each term. *Organic thickness* is a structured linguistic variable. Some qualitative attributes, such as *vegetation groups* or *moisture regime*, cannot be easily expressed in terms of a discrete measurable base variable. Therefore, it is difficult to define the semantics as mathematical functions. Subjective assessments are usually used to assign the membership grade for each database entry [Zadeh 1975]. Variables of this type are referred to as unstructured linguistic variables.

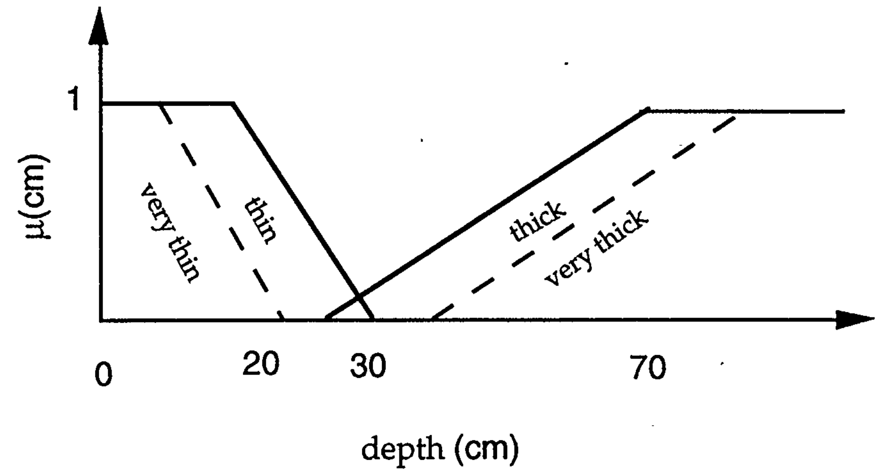

Figure 4.1 Graphical representation of *Organic thickness*.

4.2. Database Representation of Semantics

4.2.1 Structured Linguistic Variables

To represent a structured linguistic variable in a relational database, the semantics must be represented in the form of relational tables. One approach is described in Zemankova-Leech & Kandel [1985]. Here the semantics are stored as a look-up table (Table 4.1a). From the table, a site with 22 cm of organic material is assigned to *thin* with a membership grade of 0.8. Look-up tables are suitable for linguistic variables with small term-sets and from discrete universes. For structured semantics which can be characterized by a few parameters, the parameters can be stored in a relational table (Table 4.1b) and the membership grades computed during processing.

(a) relation *thin*:

CM	μ
20	1.0
22	0.8
24	0.6
26	0.4
28	0.2
30	0.0

(b)

Term	p1	p2	p3
thin	0	10	30
thick	25	70	200

Table 4.1 (a) A semantic look-up table. (b)A semantic function table.

4.2.2 Unstructured Linguistic Variables

For unstructured linguistic variables, the import semantic model [Baldwin & Zhou, 1984] can be used to store the membership grades. This model simply attaches a column next to each linguistic variable for storing the corresponding membership grade (Table 4.2). Because this type of linguistic variable lacks well defined semantics, subjective opinion and/or numerical methods are used to determine the membership grade.

The import semantic model only provides information on the compatibility of the numerical values with the linguistic terms. To represent the fuzziness between the linguistic terms, we use the similarity relation model [Buckles and Petry 1982]. Table 4.3 shows a similarity matrix of the linguistic variable *Drainage*. The elements in the matrix indicate the degree of overlap between two terms. For instance, the overlap between *well-drained* (W) and *imperfect* (I) is 0.1, indicating a slight overlap in the meanings of the two terms. In a query which requires consideration of all *well-drained* sites one will have to include some of the *imperfect* plots. Similarity matrices, in another sense, define the search domain for retrieval of unstructured variables.

Sample	Drainage	μ
1	well	0.7
2	imperfect	0.9
3	well	0.8
4	rapid	0.6
5	rapid	1.0
:	:	:

Table 4.2 *Drainage* represented as the import semantic structure.

Drainage:

	VR	R	W	MW	I	P	VP
VR	1.00	0.30	0.10	0.00	0.00	0.00	0.00
R	0.30	1.00	0.30	0.10	0.00	0.00	0.00
W	0.10	0.30	1.00	0.30	0.10	0.00	0.00
MW	0.00	0.10	0.30	1.00	0.30	0.10	0.00
I	0.00	0.00	0.10	0.30	1.00	0.30	0.10
P	0.00	0.00	0.00	0.10	0.30	1.00	0.30
VP	0.00	0.00	0.00	0.00	0.10	0.30	1.00

Table 4.3 A similarity matrix for *Drainage*.

4.3 Fuzzy Numbers

Based on the fuzzy set theory, a fuzzy number can be described as a normal fuzzy subset in the real number space. In addition to being a quantitative measurement, it includes a qualitative valuation described by the function $\mu(x)$. A fuzzy number is represented as a fuzzy set characterized by a membership function. Figure 4.2 shows a graphical representation of a triangle fuzzy number. Each number has an expected value and two values to define the upper (a_2^{α}) and lower (a_1^{α}) bounds. The expected value, E(a), has a membership grade of 1, while the bounding values has a membership grade of alpha (α). This is similar to the concept of a confidence interval in statistics. However, a confidence interval always has a preset significance level (e.g. α=0.05). A fuzzy number goes beyond being just a confidence interval because it is defined at all alpha levels from 0 to 1.

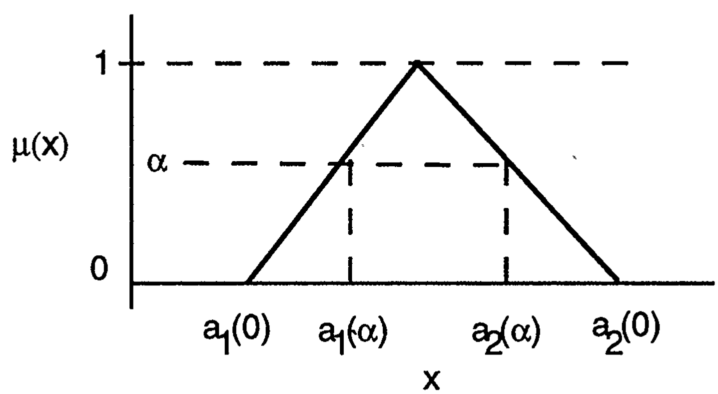

Figure 4.2. Graphical representation of a fuzzy number.

Fuzzy numbers can be classified by the shape of the membership function. If a fuzzy number is characterized by a triangular membership function, it is called a triangular fuzzy number (TFN), and it requires three parameters (Figure 4.3) to define the membership function. Another common shape is the trapezium (TzFN), which is defined by four parameters.

An important property of fuzzy numbers is their closure under linear combinations, that is, only fuzzy numbers of the same type can be operated on, and the resultant value will also be of the same type of fuzzy number. This simplifies the computation and makes it possible to carry a small number of parameters [Kaufmann and Gupta, 1985].

In some cases, numerical attributes are better represented as fuzzy numbers when the values represent averages of the measurement within a specific area rather than precise measurements. Examples are the density of crown closure of a forest and the depth of organic soil of a particular soil type.

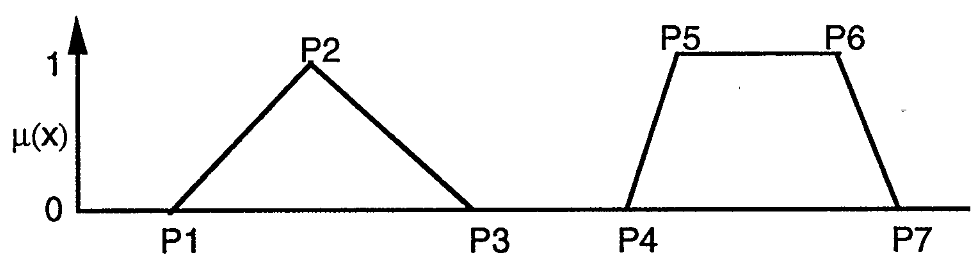

Figure 4.3. Triangular and trapezoidal fuzzy numbers.

4.3.1 Fuzzy Arithmetic

The definition of fuzzy arithmetic came soon after the introduction of fuzzy set theory and fuzzy numbers. Fuzzy arithmetic is considered an extension to the classical arithmetic because it augments the arithmetic operators to handle the extra information provided by fuzzy numbers. Compared with classical arithmetic, fuzzy arithmetic is a more expressive tool because real world phenomena are rarely as exact as most mathematical models assume. With fuzzy arithmetic, the degree of fuzziness of the operands can be represented and propagated onto the resultant values in a formal manner. In addition, as shown in Figure 4.2, the presumption level, α, can be varied in each operation, providing analysts with an extra parameter to control the degree of fuzziness they are willing to absorb. The basic fuzzy arithmetic operators are defined as follows:

Given two triangular fuzzy numbers $A(a_1^\alpha, a_2^\alpha)$, $B(b_1^\alpha, b_2^\alpha)$ in the real number space, evaluated at all alpha level:

Addition:

$$A^\alpha + B^\alpha = C\{a_1^\alpha + b_1^\alpha, a_2^\alpha + b_2^\alpha\} \tag{4.1}$$

Subtraction:

$$A^\alpha - B^\alpha = C\{a_1^\alpha - b_2^\alpha, a_2^\alpha - b_1^\alpha\} \tag{4.2}$$

Multiplication:

$$A^\alpha * B^\alpha = C\{a_1^\alpha * b_1^\alpha, a_2^\alpha * b_2^\alpha\} \tag{4.3}$$

Division:

$$A^\alpha / B^\alpha = C\{a_1^\alpha / b_2^\alpha, a_2^\alpha / b_1^\alpha\} \tag{4.4}$$

Multiplication by a constant C:

$$A^\alpha * D = C\{a_1^\alpha * D, a_2^\alpha * D\} \tag{4.5}$$

Fuzzy arithmetic operators are not implemented in this prototype system because few numerical operations are required in this research project. However, a good example of the application of fuzzy arithmetic can be found in Bardossy et al [1989]. Bardossy presented a fuzzy kriging technique for predicting soil liner permeability from imprecise data points. This technique allows analysts to incorporate fuzzy data points into the kriging process when not enough good data points are available to perform normal kriging. The authors concluded that fuzzy arithmetic can be used to incorporate imprecise data in a geo-statistical analysis, and the example of permeability field prediction for a soil liner is a typical problem where imprecise data is available and should be utilized.

4.4 Fuzzy Logical Operations

The conventional logical operators such as *greater than, less than* and *equal* take precise values as operands. In fuzzy comparison, three situations can occur: 1) both operands are precise, 2) one operand is precise and one is fuzzy, and 3) both operands are fuzzy. The first case can be handled by conventional operators. Because the latter two cases involve fuzzy sets, the conventional logical operators must be modified. These extended logical operators return the degree of truth (T) of the comparison. Equation 4.6, 4,7 and 4.8 are the extended logical operators developed to handle case 2. In the equations, fv and pv stand for fuzzy value and precise value respectively.

Approximately equal:

$$fv \sim= pv: T = \mu_{fv}(pv) \qquad (4.6)$$

Fuzzy less than:

$$pv << fv: T = \frac{\int_{pv}^{\infty} \mu_{fv}(x)\,dx}{\int_{\upsilon} \mu_{fv}(x)\,dx} \tag{4.7}$$

Fuzzy greater than:

$$pv >> fv: T = \frac{\int_{-\infty}^{pv} \mu_{fv}(x)\,dx}{\int_{\upsilon} \mu_{fv}(x)\,dx} \tag{4.8}$$

Equation 4.6 simply computes the membership grade of pv in the fuzzy set fv. Equation 4.7 computes the proportion of the area under the membership function of fv up to but not including the value of pv. If pv equals the expected value of fv, E(fv), then the truth value equals 0.5. The further away pv from E(fv) the higher the truth value. The reverse applies to the *less than* operator. Figure 4.4 graphically illustrates these operations. These operators have been implemented in this prototype system.

To extend the logical operators to handle two fuzzy operands. The concept of *degree of coincidence* [Lui & Li, 1990] is introduced. The degree of coincidence of A with respect to B is:

$$\omega(A,B) = \frac{\int_{\upsilon} \min(\mu_A(x), \mu_B(x))\,dx}{\int_{\upsilon} \mu_B(x)\,dx} \tag{4.9}$$

which is the proportion of A in B. The definitions of these three operators are defined as follows:

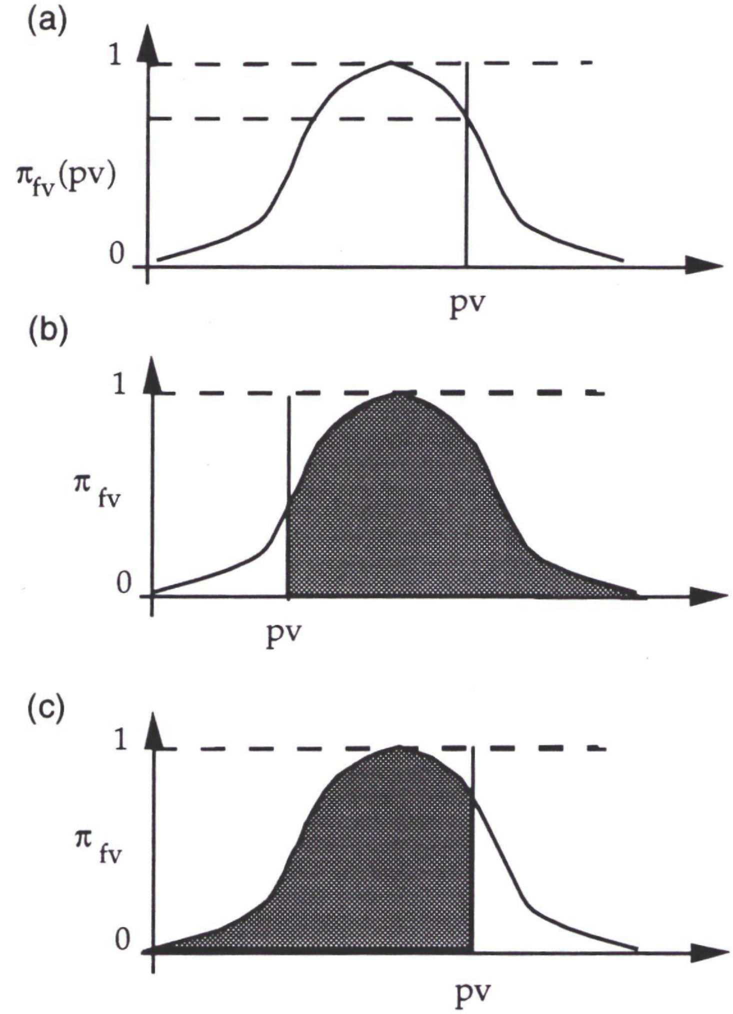

Figure 4.4 Graphical illustration of the fuzzy comparison operators. (a) Approximately equal, (b) fuzzy less than, (c) fuzzy greater than.

Approximately equal:

$$f_1 \sim= f_2 : T = \omega(f_1, f_2) \tag{4.10}$$

Much less than:

$$f_1 << f_2 : T = \mu_{MLT}(E(f_1), E(f_2)) \tag{4.11}$$

Much greater than:

$$f_1 >> f_2 : T = \mu_{MGT}(E(f_1), E(f_2)) \tag{4.12}$$

E(x) is the expected value of a fuzzy set and MLT and MGT are fuzzy sets. Depending on the applications, these functions can be designed to fit particular scenarios. Liu and Li [1990] suggested the following function for modelling the *much greater than* operation:

$$\begin{aligned} \mu_{MGT}(x,y) &= 0 && \text{if } x < y \\ &= [1+c(y-x)^{-2}]^{-1} && \text{if } x > y \end{aligned} \tag{4.13}$$

Since most comparisons in this research project involve one fuzzy operand and one precise operand, only equations 4.6, 4.7, 4.8 were implemented in the current prototype system.

4.5 Fuzzy Retrievals

An amazing aspect of human reasoning is its ability to process a considerable amount of vague information and yet make precise and important decisions. Zadeh [1984] asserted that fuzzy logic is a more realistic tool to provide computers with this human-like reasoning capability than two-valued or multi-valued logic. The reason being that most human reasoning processes are imprecise and that fuzzy logic is oriented towards the processing of fuzzy information. The use of linguistic variables facilitates both fuzzy retrieval as well as approximate reasoning, which are considered a more human-like approach to information processing [Zadeh, 1984]. Without leaving the scope of this thesis, only the fuzzy retrieval techniques in a relational database are presented.

4.5.1 Lexical Matching and Semantic Matching

Conventional retrieval procedures perform what is called *lexical matching*. Lexical matching matches the retrieval criteria with the value in the database. For instance, in a conventional database, qualitative terms are stored as lexical symbols. To retrieve all areas with *thick* to *very thick* organic soil, the system matches the word *thick* and *very thick* with the database entries and retrieve records that match the word exactly. On the other hand, fuzzy retrieval performs *semantic matching*. In semantic matching, semantic functions of the terms in the retrieval criteria are compared with the semantic functions or exact values of the attributes in the database. This is similar to the fuzzy operations presented in Section 4.4. The users can also limit the set of records being retrieved by setting the acceptable degree of match for a particular query.

4.6 Comparison of Conventional and Fuzzy Retrieval

To demonstrate the differences between fuzzy retrieval and conventional retrieval, several examples are presented in this section.

1) In conventional systems, to retrieve all areas that are facing South, the user has to specify the range of degrees that he/she would consider as

South. To be exact, South denotes 180 degrees from North. However, one can also define South in the range from 90 to 270 degrees from North. In a fuzzy system, the user can define south as a triangular membership function ranging from 90 to 270 with the expected value at 180 degree (Figure 4.5). Table 4.4 shows the results of these retrievals. It shows that using the most exact query, only four records were retrieved. Both the second and third methods (column 2 and 3) retrieved the same set of records, yet fuzzy retrieval provides more information than the conventional query.

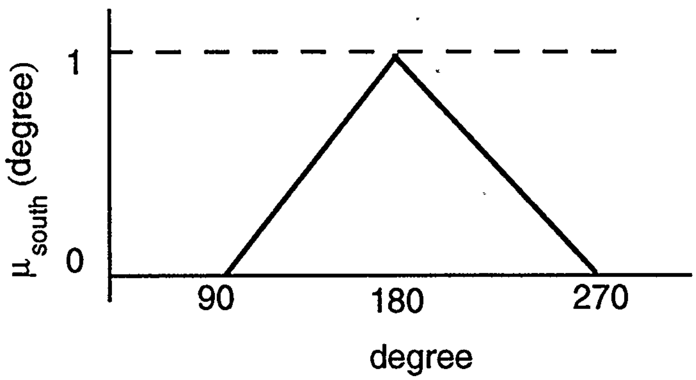

Figure 4.5 The characteristic function of *South.*

2) To illustrate the robustness of fuzzy comparison operators, a query was submitted to retrieve all polygons that are much lower than 1150 meters in elevation. Such an operation cannot be easily achieved in conventional systems unless a threshold, say 1000 meters, is provided. In fuzzy retrieval, a fuzzy number 1150 was declared and the fuzzy less than operator (<<) was used in the query. Table 4.5 is a partial listing of the query result. It compares the records retrieved by conventional method and fuzzy method. Using fuzzy retrieval, the farther away from 1150 meters, the greater the membership

grade. Not only are more records being retrieved, but the fuzzy technique provides the membership grades of each record so that the user can decide which records are really required.

ID	ASPECT	180 degrees	90 to 270 degrees	μ_{South}
28	180	x	x	1.00
48	260		x	0.11
50	225		x	0.50
84	225		x	0.50
95	240		x	0.33
96	225		x	0.50
111	95		x	0.06
171	210		x	0.60
177	180	x	x	1.00
178	180	x	x	1.00
185	180	x	x	1.00
186	160		x	0.78

Table 4.4. Records retrieved by query 1. 'x' indicates record retrieved.

3) To demonstrate the propagation of uncertainty in retrieval with a compound condition, consider the following query:

Retrieve records with drainage ~= Imperfect (cf = 1.0) and *Organic thickness* >> 10 cm (cf = 0.6)

In the query, the confidence factor (cf) for each attribute can be specified. The system first retrieved all records with *imperfect* drainage, then those records with *organic thickness* greater than 10 cm and with membership a grade greater than 0.6 were retrieved. The system propagated the uncertainty with the Min operator.$\mu_T(x)$ in Table 4.6 represents the propagated uncertainty associated with each record.

Fuzzy retrievals are particularly useful in GIS because of the imprecise and the continuous nature of geographical information. When compared with the conventional technique, fuzzy retrieval provides more information to the GIS users. As noted before, fuzzy information processing does not attempt to provide the user with the ideal solution but rather to supply him/her with possible solutions. Therefore, a well designed fuzzy decision support GIS will provide more detailed information to assist the analysts and policy makers to make important decisions.

ID	Elev.	Elev. < 1000	$\mu_{<<1150}$
48	1012		0.92
73	899	x	1.00
81	1021		0.90
84	1006		0.93
95	1010		0.92
103	1021		0.90
106	968	x	0.97
111	930		0.99
112	768	x	1.00
129	1021		0.90
130	1006		0.93
133	1000		0.93
134	1000		0.93
136	991	x	0.94
138	957	x	0.98
150	920	x	1.00

Table 4.5 Partial listing of records much less than 1150 meters.

ID	DR	μ_{DR}	OT	μ_{OT}	μ_T
28	I	1.0			0.00
50	I	1.0	15	0.75	0.75
81	I	1.0	20	0.89	0.89
84	I	1.0	15	0.75	0.75
89	I	1.0			0.00
150	I	1.0			0.00
155	I	1.0	20	0.89	0.89
156	I	1.0	40	1.00	1.00
171	I	1.0			0.00
177	I	1.0	17	0.81	0.81
27			12	0.64	0.00
72			20	0.89	0.00

Table 4.6 partial listing of a compound fuzzy query.
DR=drainage, I=imperfect. μ_{DR}=membership grade for DR. OT=Organic thickness in cm.μ_{OT}=membership grade for OT. μ_T= propagated membership grade.

CHAPTER 5

Modelling and Classifying Fuzzy Objects

Many GIS applications deal with classifications of natural or socio-economic phenomena or objects. These phenomena are sometimes abstract concepts like ecological regions or wild animal habitats. In this prototype system these phenomena and concepts are referred to as objects. These objects must be represented realistically in the computer system for further processing. The more realistic the object models, the less uncertainty will be introduced in the modelling process and the more effective is the GIS. As GIS are used for repositories of information as well as decision support tools, realistic object modelling is of prime importance. This chapter introduces the concept of linguistic modelling and compares different fuzzy set aggregation operators for classifying fuzzy objects.

5.1 Abstract Concepts versus Computer Models

Many commercial GIS use conventional relational database tools for data management. Because of the complexity of some phenomena and the abstraction of many geographic concepts, realistic representation in these GIS is not easily achievable. Conventional relational database techniques are based on Boolean logic, and the *law of the excluded middle* makes modelling non-mutually exclusive concepts impossible. However, many concepts in geographical analyses cannot be easily represented by a few attributes, especially when the domains are continuous. Ecoregions are examples because the transition from one ecoregion to another is usually gradual rather than abrupt. Thus, no mutually-exclusive boundaries can be easily drawn between two regions.

5.2 The Fuzzy Set Approach to Object Modelling

The fuzzy-algorithmic approach to object modelling was presented by Zadeh [1973 1975, 1976]. Zadeh called this the *linguistic approach* because objects are described by sentences which consist of linguistic variables joined together by fuzzy connectives. This approach can be considered an extension to the concept of linguistic variables discussed in Chapter Four. To illustrate, *soil type A* can be modelled as a fuzzy object if it is described by fuzzy attributes as follows:

"*fresh* soils with *sandy* or *loamy sand* parent material"

which can be translated into a formal representation:

Soil type A :: Moisture_regime(Fresh) and (texture(sandy or loamy_sand)).

Where *moisture_regime* and *texture* are linguistic variables previously defined in the system, *and dry, fresh, sandy* and *loamy sand* are terms of these variables. *And* and *or* are fuzzy connectives which connect the attributes to form the definition of Soil type A.

Zadeh's linguistic approach [1976] defines an object by a set of *atomic* questions. An atomic question, Q, is a question containing only one constituent. An atomic question has an object-set, X, the body of the question, B, which is the label of a fuzzy set, and the answer-set, A. Q can be denoted by $Q(X,B,A)$. The following is an example of an atomic question:

Q: "*Is sample #2 dominated by white spruce?*"

X: All samples in the data set.

B: dominated by white spruce.

A: A truth value in the range of 0 to 1.

The body of this question is a fuzzy set denoting the percentage cover of white spruce in the sample plot. The answer-set A can be a set of truth values in the range of 0 to 1, or linguistic terms like *true, undecided* and *false.*

An object is defined by a composition of atomic questions. Putting the questions in an *analytic representation* [Zadeh, 1976], an object can be described as:

$$\text{Object}_x = B_1 \bullet_1 B_2 \bullet_2 \cdots \bullet_{n-1} B_n \tag{5.1}$$

B_i is the bodies of Q_i, $\bullet_i$ are connecting operators, and n is the number of atomic questions. To evaluate whether an object, y, belongs to Object_x, the following function is used:

$$\text{Object}_x = t(Q_1(y)) \bullet_1 t(Q_2(y)) \bullet_2 \cdots \bullet_{n-1} t(Q_n(y)) \tag{5.2}$$

$t(Q_i(y))$ is the truth value for atomic question i when applied to y, and $\bullet_i$ is the connecting operator to combine the truth values.

5.3 Representation of Fuzzy Ecosystem Associations

The Field Guide to Forest Ecosystems of West-Central Alberta [Corns & Annas, 1986] has described the following four ecosystems in that area: the Boreal Mixedwood ecoregion (BMW), the Lower Boreal Cordillearn ecoregion (LBC), the Upper Boreal Cordilleran ecoregion (UBC) and the Subalpine ecoregion (SA). These ecoregions are further subdivided into ecosystem associations (EA). Ground survey in the vicinity of the study area was carried out in the summer of 1990. Among the 247 sample plots, 59 were located inside our study area (Figure 3.4). Based on these 59 samples fourteen EA were identified, covering the LBC and UBC ecosystems. These fourteen EA were modelled with the linguistic approach.

Forest ecologists identified these nine attributes as indicating features for the EA: *elevation, aspect, thickness of organic material, soil drainage, soil texture, percentage covers of white spruce, black spruce, lodgepole pine and poplar.* These tree species are considered indicator species for the EA. All attributes except *soil drainage* and *texture* are numerical attributes. To illustrate the modelling concepts without involving complicated procedures, all attributes were assumed independent. Figure 5.2 shows the attribute values for an EA called the Lower Boreal Cordilleran association 1 (LBC 1). The values in the table were obtained from the field guide [Corns & Annas, 1986] and from statistics of the sample data. These values were represented as fuzzy sets. Numerical attributes in Figure 5.2 were represented by triangular or trapezodial-shaped characteristic functions and qualitative attributes were stored as a look-up tables. Examples of these representation are given in Table 5.1 and 5.2.

To illustrate, according to the field guide, the LBC 1 association has an elevation range of 850 to 1150 meters. Since the range for the LBC ecoregion is between 500 to 1150 meters, the possibility of finding LBC 1 outside the specified range is acknowledged by extending the possibility function to cover this range. Thus a trapezoidal-shaped function was used. For other numerical attributes, triangular-shaped characteristic functions were used, with the given values as the expected values (peak of the triangles). The fuzziness in these numerical variables were derived from the frequency distribution of the attributes. For discrete qualitative attributes, i.e. soil texture and drainage, discrete possibility scores are used to represent the possibility distribution. In a relational database, a table was created for each attribute and stored the possibility distribution functions by the EA. Table 5.1 and 5.2 show examples of the representation of *elevation* and *drainage* as relational tables.

5.4 Classification of Ecosystem Associations

The fuzzy classification procedure involves three steps: feature evaluation, pattern matching and categorisation [Oden & Lopes, 1982]. Feature evaluation determines how well an attribute value fits the values given in the object definition. Pattern matching combines the information obtained from feature evaluation to assess the overall fitness of the instance to the object. Categorisation uses the information in pattern matching to determine the category or categories of the instance. Figure 5.1 shows a flow chart of this process. Two fuzzy pattern matching approaches are evaluated in the sequel. They are the fuzzy propositional approach and the distance approach. These two approaches differ in the interpretation of the feature information as well as the operators used to aggregate fuzzy sets. The fuzzy propositional approach is presented in Oden & Lopes [1982] and the distance approach is based on a modified version of that described in Wang et al [1990]. The comparison of classification results generated by the two methods are given in Section 5.5.

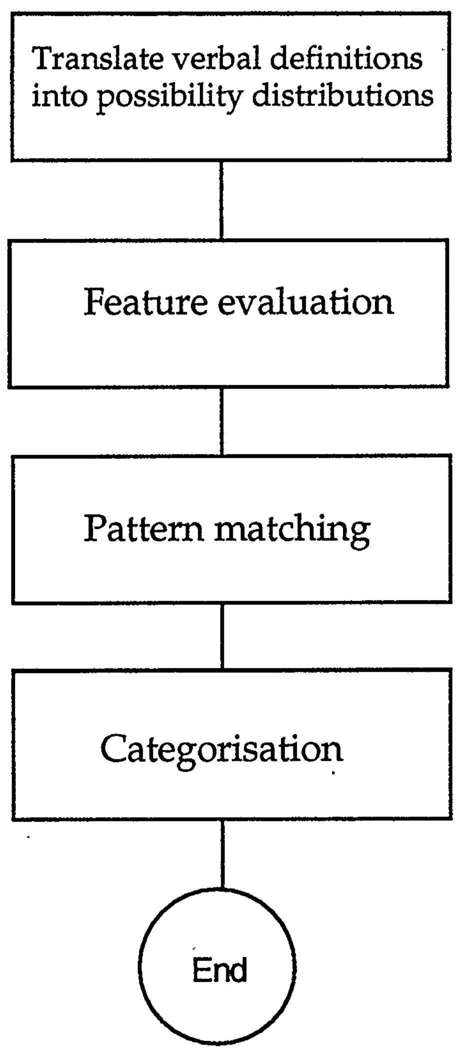

Figure 5.1 The fuzzy classification procedure.

Species	
%white Spruce	~2%
%black Spruce	~7%
%lodgepole Pine	~18%
%poplar	~0%
Elevation	850 to 1150 meters
Aspect	Mostly North
Drainage	Moderate well to imperfect
Texture	Clay loam
Organic Thickness	~17 cm

Figure 5.2. Linguistic descriptions of LBC 1. "~"denotes a fuzzy quantity.

EA	P1	P2	P3	P4
LBC1	500.00	850.00	1150.00	1150.00
LBC2	500.00	650.00	910.00	1150.00
LBC3	500.00	740.00	1150.00	1150.00
LBC4a	500.00	670.00	1070.00	1150.00
LBC4b	500.00	800.00	1140.00	1150.00
LBC4c	500.00	710.00	1010.00	1150.00
LBC5b	500.00	780.00	1060.00	1150.00
LBC5c	500.00	520.00	1050.00	1150.00
LBC7	500.00	800.00	1200.00	1200.00
LBC9	500.00	670.00	950.00	1150.00
LBC10	500.00	600.00	1180.00	1180.00
UBC2	900.00	1000.00	1290.00	1500.00
UBC3	900.00	980.00	1460.00	1500.00
UBC4	900.00	930.00	1370.00	1500.00

Table 5.1 Possibility distribution functions for elevation range for all EA.
P1, P2, P3, P4 are parameters defining the possibility distribution function.

TERM	SDY	SIY	FLMY	CLMY	CLY	UNM	ORG	UNDO
LBC1	0.24	0.52	0.76	1.00	0.00	0.24	0.00	0.24
LBC2	0.07	0.21	0.33	1.00	0.00	0.07	0.00	0.07
LBC3	0.00	0.33	0.33	1.00	0.08	0.22	0.00	0.22
LBC4a	0.00	0.39	0.39	0.58	1.00	0.07	0.00	0.19
LBC4b	0.34	0.56	1.00	0.77	0.11	0.34	0.00	0.34
LBC4c	0.00	0.00	0.50	0.33	1.00	0.17	0.00	0.17
LBC5b	0.26	0.50	1.00	0.74	0.26	0.26	0.00	0.26
LBC5c	0.15	1.00	0.35	1.00	1.00	0.15	0.00	0.15
LBC7	0.00	0.21	0.61	1.00	0.82	0.21	0.00	0.21
LBC9	0.00	0.00	0.50	0.17	1.00	0.17	0.17	0.17
LBC10	0.07	0.21	0.33	1.00	0.00	0.07	0.00	0.07
UBC2	0.12	0.76	0.50	1.00	0.24	0.12	0.12	0.12
UBC3	0.05	0.20	1.00	0.25	0.20	0.05	0.00	0.05
UBC4	0.13	0.13	0.64	0.64	1.00	0.13	0.00	0.13

Table 5.2 Possibility distribution of texture for all EA.
SDY=sandy, SIY=silty,FLMY=Fairly loamy, CLMY = clay loamy, CLY=clay,
UNM = undetermined, ORG=organic, UNDO = not observed.

5.4.1 Feature Evaluation

Section 5.2 presented an approach to modelling complex concepts by a set of atomic questions. For each attribute listed in Figure 5.2, an atomic question is formed:

Is white spruce dominant in the location of sample x?
Is black spruce absent in sample X?
Is lodgepole pine dominant in sample x?
Is poplar absent in sample x?
Does sample x fall within 850 to 1150 meters in elevation?
Does sample x faces North?
Does sample x have well-drained to imperfectly drained soil?
Does sample x have 3 to 4 inches of organic material?

This set of questions evaluates whether an instance matches the description of LBC 1. Since there are fourteen EA, fourteen sets of questions were derived. The feature evaluation process determines the truth values or possibility, $\pi(x)$, of these atomic questions. Using Boolean logic the answers to these questions are either yes or no. With fuzzy sets, the answers are truth values in the range of 0 to 1, with 0 representing a *definite NO* and 1 representing a *definite YES*. Table 5.3 shows the result of sample # 73 after the feature evaluation process. The values shown in the table are truth values of individual attributes being evaluated for the fourteen EA.

EA	SW	SB	PL	PO	TXT	OT	ELEV	DR	ASP
LBC1	1.00	0.91	1.00	1.00	0.48	0.00	1.00	1.00	1.00
LBC2	0.00	0.00	0.00	0.38	0.79	1.00	1.00	0.50	0.30
LBC3	1.00	0.40	0.50	0.83	0.67	0.60	1.00	0.50	1.00
LBC4a	0.00	0.00	0.00	1.00	0.61	0.95	1.00	0.50	0.00
LBC4b	1.00	0.00	0.50	0.00	0.44	0.00	1.00	1.00	0.00
LBC4c	1.00	0.00	0.00	0.00	1.00	1.00	1.00	0.50	0.50
LBC5b	1.00	0.00	0.20	0.00	0.50	0.50	1.00	1.00	0.00
LBC5c	1.00	0.00	0.00	0.00	0.00	0.50	1.00	1.00	0.00
LBC7	1.00	0.00	0.00	0.00	0.79	1.00	0.50	0.50	0.00
LBC9	0.00	0.00	0.15	0.00	1.00	0.95	1.00	0.00	0.00
LBC10	0.00	0.00	0.00	0.00	0.79	0.95	1.00	0.50	0.00
UBC2	0.00	0.00	0.00	0.00	0.24	0.92	0.00	1.00	0.00
UBC3	1.00	1.00	0.75	0.00	0.80	0.60	0.00	0.50	1.00
UBC4	1.00	0.00	0.00	0.00	0.87	0.00	0.00	0.50	0.00

Table 5.3 Result of pattern matching for sample #73.

5.4.2 Pattern Matching

5.4.2.1 Pattern Matching using Fuzzy Proposition

The fuzzy propositional approach evaluates the degree of match of the proposition and an instance. Lopes & Oden [1982] noted that fuzzy proposition combines both the semantic network representation and the logical structures which allows the expression of continuity in many natural concepts. Using the fuzzy propositional approach, each object is described by a formal proposition such as the follows:

LBC1:: White spruce(~2%) • Black spruce (~7%) • Lodgepole pine(~18%) • Poplar (~6%) • Elevation (850 to 1150) • Aspect (Mostly N) • Soil drainage(Mod. well to Imperfect) • Texture (Clay loam) • Organic Thickness (17 cm)

• is a connective for combining all truth values to derive at the possibility for the conclusion. Contrary to conventional set theory, no well defined operators exist for fuzzy sets. Criteria for selecting the appropriate operator depends on the problem situation. In much of the research on uncertainty propagation, the triangular norms (t-norms) and conorms (t-conorms) are often used [Dubois & Prade, 1985; Buckley [1990]; Kruse et al, 1991].

A t-norm, T, is a mapping function T:[0,1]x[0,1]->[0,1] which satisfies the following axioms:

1. $T(a,b) = T(b,a)$; (5.3)

2. $T(0,0) = 0$, $T(a,1) = a$; (5.4)

3. $T(a,b) \leq T(c,d)$ if $a \leq c$ and $b \leq d$ and (5.5)

4. $T(a,T(b,c)) = T(T(a,b),c)$. (5.6)

A t-conorm, C, is a mapping function C:[0,1]x[0,1]->[0,1] which satisfies the following axioms:

1. $C(a,b) = C(b,a)$; (5.7)

2. $C(1,1) = 1$, $C(a,0) = a$; (5.8)

3. $C(a,b) \leq C(c,d)$ if $a \leq c$ and $b \leq d$ and (5.9)

4. $C(a,C(b,c)) = C(C(a,b),c)$. (5.10)

Property 1 ensures that the order of evaluation does not affect the final result. Property 2 reflects the intersection (t-norm) or union (t-conorm) operations. Property 3 ensures monotonicity of the function and property 4 allows the extension of the function to more than two arguments. Kruse et al [1991] pointed out that, for a rule based system, the t-norms can be applied to combine the certainty of the premise of a production rule with the certainty of a rule. The t-conorms, on the other hand, can be used to combine the certainties of different rules to obtain the certainty of the conclusion. Therefore, the t-conorms can be used in this classification process to obtain the certainty of pattern matching.

Buckley [1990] reviewed three sets of t-norms and t-conorms. They are the Max and Min operators, the probabilistic AND (PAND) and OR (POR), the Lukasiewicz AND (LAND) and OR (LOR) operators. The definition of these three operators are as follows:

Min and Max operators:

$$\mathrm{Min}(a,b) = a \text{ if } b > a;\ b \text{ if } b < a$$

$$\mathrm{Max}(a,b) = a \text{ if } a > b;\ b \text{ if } a < b \qquad (5.11)$$

LAND and LOR operators:

$$\mathrm{LAND}(a,b) = \mathrm{Max}(a+b-1,0)$$

$$\mathrm{LOR}(a,b) = \mathrm{Min}(a+b,1) \qquad (5.12)$$

PAND and POR operators

$$\mathrm{PAND}(a,b) = a*b$$

$$\mathrm{POR}(a,b) = a+b-ab \qquad (5.13)$$

These operators satisfy the inequalities

$$\mathrm{LAND}(a,b) \leq \mathrm{PAND}(a,b) \leq \mathrm{Min}(a,b)$$

$$\mathrm{Max}(a,b) \leq \mathrm{POR}(a,b) \leq \mathrm{LOR}(a,b) \tag{5.14}$$

These three t-conorms were used to aggregate the possibilities and the results are presented in section 5.5.

5.4.2.2 Pattern Matching using Semantic Distances

The semantic distance approach evaluates the difference in the semantics of two terms. It is similar to the semantic matching process presented in section 4.5.1. Wang et al [1990] applied this method to perform suitability analysis. Wang et al created prototype vectors describing the best conditions for growing several crops. The prototype vectors contain the representative values for characteristic attributes. Each area was also represented by a feature vector and the Euclidean distance between two vectors was used to evaluate the suitability of each crop being grown in the area. Nevertheless, in Wang's research, no fuzzy attributes were involved. Applying this method to evaluate fuzzy ecosystem association requires some modifications.

Fourteen prototype vectors containing fuzzy attributes were created. The semantic distance, D, between a sample plot's feature vector and an EA prototype vector was computed by aggregating the semantic distance of of individual attribute, d. Several methods for computing distances between two vectors were presented in Lui and Li [1990]. They are:

the Hamming distance:

$$D(A,B) = \sum_{i=1}^{n} |(\mu_A(x_i)-\mu_B(x_i))| \tag{5.15}$$

the Euclidean distance:

$$D(A,B) = \sum_{i=1}^{n} (\mu_A(x_i)-\mu_B(x_i))^2 \tag{5.16}$$

the Chebyshev's distance:

$$D(A,B) = \underset{1<i<n}{\mathrm{Max}} \; |(\mu_A(x_i),\mu_B(x_i))| \tag{5.17}$$

Where n is the number of features used for pattern matching. A restriction must be added to handle the situation when $A \cap B = 0$. Therefore,

$$\text{If } A \cap B = 0, \text{ then } D(A,B) = n. \tag{5.18}$$

Examining equations 5.11, 5.12, 5.15 and 5.17, one will note that the Hamming distance and the Chebyshev's distance correspond to the LOR and the Max operations, respectively. The two sets of equations only differ in the range of the resultant values. In this research, equations 5.15, 5.16 and 5.17 cannot be directly applied because this problem requires the computation of distance between a precise value (the sample plot's attribute values) and a fuzzy value (the EA definition). The modified versions of equations 5.15, 5.16, 5.17 become:

the Hamming distance:

$$D(EA,x) = \sum_{i=1}^{n} |(1 - \mu_{EA}(x_i))| \tag{5.19}$$

the Euclidean distance:

$$D(EA,x) = \sum_{i=1}^{n} ((1-\mu_{EA}(x_i))^2 \qquad (5.20)$$

the Chebyshev's distance:

$$D(EA,x) = \underset{1<i<n}{Max} \; |(1 - \mu_{EA}(x_i))| \qquad (5.21)$$

These equations were used to compute the semantic distances between EA, the ecosystem association, and x, the sample plots being considered.

5.4.3 Categorisation

From the pattern matching results, one can identify the EA assignment for each sample plot by examining the truth values or the semantic distances. Nevertheless, the absolute scores do not indicate the relationship among all the possible EA assignments for each sample plot. To illustrate, Table 5.4 shows the pattern matching results for sample # 125 and # 133. Based on the highest absolute scores, both plots were assigned to LBC 7, which matched the experts' assignment. From the table, one can also see that the difference between the first two assignments for #125 is greater than that for # 133, but without standardized scores, it is difficult to compare the appropriateness of the assignment. Therefore, a relative score, R, is introduced. R is an assessment of the object's degree of belonging to each candidate category relative to that for all the other possible categories. Adopting the definition from Oden & Lopes [1982] and Wang et al [1990], $R_j(x)$ is defined as:

$$R_j(x) = \frac{S(EA_j(x))}{\sum_{i=1}^{c} S(EA_i(x))} \qquad (5.22)$$

where c is the number of classes that x can be classified to belong to. The numerator is the similarity score for sample plot x and EA_j, and the denominator is the summation of the similarity of EA that sample plot x could be classified into. $S(EA_j(x))$ is the similarity measure between sample plot x and EA_i, and is computed by equation 5.23.

$$S(EA_i(x)) = 1 - D(EA_i, x) \qquad (5.23)$$

$S(EA_i(x))$ can be replaced by $t(EA_i(x))$ if the propositional approach is used. The rationale behind this relative measure, as stated by Oden & Lopes [1982], is that not being in one category should count in favor of it being in another category.

ID	EA	R_D	R_P	Cheb	Ham	Eud	LOR	POR
125	LBC7	0.12	0.10	1.00	1.61	0.84	0.81	0.57
	LBC3	0.07	0.06	1.00	2.82	1.54	0.68	0.51
	LBC9	0.07	0.06	1.00	3.05	1.58	0.65	0.50
133	LBC7	0.07	0.06	1.00	3.71	1.81	0.58	0.46
	LBC5b	0.06	0.05	1.00	4.5	1.94	0.05	0.41
	LBC3	0.06	0.06	1.00	4.55	1.86	0.49	0.40

Table 5.4 Partial listing of classification result for #125 and #133.
R_D= relative measure of Ham. dist., R_P = relative measure for LOR.

5.5 Results and Discussion

Both the fuzzy propositional and distance approaches were used to classify the 59 sample plots (Figure 3.4) into fourteen EA. For comparison purposes, the operators described in Sections 5.4.2.1 and 5.4.2.2 were used to aggregate the truth values or the semantic distances, respectively. While conventional techniques aim at finding the most appropriate EA for each sample plot, the fuzzy techniques provide a set of possible classifications. The EA with the highest R was considered the most appropriate EA for the sample plots. Depending on the situation, one can also consider the next few most appropriate EA for particular application.

Table 5.5 presents classification results generated by different operators. These results were compared with the original EA provided by the forest ecologists. The first column represents the highest R which matches the experts' assignment. The second and third columns represent the second and third highest R which match the experts' assignment. Note that in this classification both the Max operator and the Chebyshev's distance failed to classify. The Max (and Min) operator is called a non-interactive operator [Leung, 1988] because a high truth value or semantic distance for one attribute does not compensate for a low value for another attribute. In this classification procedure, a perfect match of any one attribute resulted in a perfect match for the EA being evaluated. Intuitively this is not correct. In this case, because among the nine attributes there are always some attributes with truth value equal to 1, the Max operator will only register 1.0 as the "aggregated value". Thus, the "aggregated" truth values are almost always the same. Despite Bellman's [1973] argument on the appropriateness of the Max and Min operators as the only natural, and in some situations the only possible operators to model conjunction and disjunction of fuzzy sets, the non-interactive nature of the Max (and Min) operator does not perform well in these types of problems. In fact, based on empirical studies, Dubois and Prade [1985] and other authors [Thöel et al, 1979] commented that the Min operator does not always reflect users' attitude when aggregating truth values. In this exercise, it shows that the Max operator also is ineffective in the assignment of EA.

The classification result showed that the four operators basically generate the same results despite the difference in their numerical values, i.e. the order of relative similarity is the same although the distances or the truth values do not demonstrate similar difference in magnitude.

Table 5.7 shows a confusion matrix which compares the experts' EA assignment with the fuzzy classification. The former is presented on the column and the latter is presented on the row. The numbers in the diagonal of the matrix indicate the number of cases that matched the experts' assignment. Since fuzzy techniques were used, the top three EA assigned by the system were used for comparison. If any of the first three assignment matched the experts' assignment, it is considered a match. Summing up the numbers on the diagonal gives the percentage of correct assignment. Forty-nine of the 59 cases or 81% of the cases were matched correctly.

Numbers off the diagonal are mismatched cases. Unfortunately, experts were unavailable to examine the mismatched cases. Several reasons could have contributed to the failure of these cases. First, some EA such as the LBC3 and UBC 3 are basically identical associations except for their elevations.

If the sample sites are located in the fuzzy boundary of the two regions, it could be difficult to decide the actual EA. Arbitrary assignment could have been given. Second, in this research, experts suggested only nine attributes for prototype system testing. Compared to the 151 attributes used in the original assignment, it is not surprising to have some mismatched cases. Third, the experts noted that subjectivity involved in the original classification could have contributed to the mismatches. Last, the choice of membership functions and fuzzy connectives could have significant implication on the classification results. However, the application of fuzzy sets to ecological classification is relatively new and little experience on the derivation of membership functions and the selection of fuzzy aggregation operators is available. The extra information provided by the fuzzy classification procedures provided insight into the classification. For instance, Table 5.6 shows the classification for sample #47 and #131, which are both classified as LBC3 by ecologists and the fuzzy classifiers. For sample #47, both the hamming distance and the Euclidean distance show that LBC3 is about half the distance to the definition vector than UBC3. Compare to sample #131, where the difference in distance between LBC3 and LBC5b is only 14%. This suggests that sample #47 is a more distinct LBC3 than sample #131. In addition, the Hamming distance for #47 is 1.47 and 2.38 for 131. This again shows that sample #47 matches the definition of LBC3 better than 131. The truth values LOR and POR endorse the conclusion. Since the truth values are expressed on a standardized scale, they provide a better indication on the degree of match than using the distances.

Method	1st	2nd	3rd	4+
LOR	29	16	8	6
POR	29	16	8	6
Max	-	-	-	-
Hamming	29	16	8	6
Euclidean	28	17	8	6
Cheb	-	-	-	-

Table 5.5 Classification results using Fuzzy Proposition and Distances approach.

LBC3 #47

EA	Cheb	Hamm	Euclid	LOR	POR
LBC3	1.00	1.47	0.84	0.83	0.58
UBC3	1.00	2.95	1.61	0.67	0.50
LBC4c	1.00	3.41	1.66	0.61	0.48

LBC3 #131

EA	Cheb	Hamm	Euclid	LOR	POR
LBC3	1.00	2.38	1.21	0.73	0.53
LBC5b	1.00	2.76	1.48	0.69	0.51
LBC4b	1.00	3.34	1.59	0.62	0.48

Table 5.6 Truth values and distances of two LBC3 plots.
EA=ecosystem association, Cheb=Chebyshev's distance, Hamm=Hamming distance, Euclid=Euclidean distance, LOR=Lukasiewicz OR, POR=Probability OR.

Experts' Assignment

Fuzzy Assignment	L1	L2	L3	L4a	L4b	L4c	L5b	L5c	L7	L9	L10	U2	U3	U4	
L1	4														4
L2		1													1
L3			12												12
L4a															0
L4b					2		1		1						4
L4c						5									5
L5b							2								2
L5c								3							3
L7			1	1			1		10						13
L9										2					2
L10											1				1
U2												1			1
U3			3		1								6		10
U4														1	1
TTL	4	1	16	1	3	5	4	3	11	2	1	1	6	1	59

Table 5.7 Confusion matrix of classification results.
L=LBC, U = UBC, TTL= Total

CHAPTER 6

Fuzzy Spatial Operations

One advantage of GIS over paper maps is their convenience in manipulating, analysis and displaying information. With a fuzzy database, some spatial operations can be extended to work with fuzzy data. Among many spatial operations in a GIS, the *overlay* and *polygon consolidation* functions are the two most used operations. The *overlay* function allows users to combine information from different maps of the same scale onto one map. The *polygon consolidation* or the *dissolve* function simplifies maps by merging adjacent polygons with the same attribute values together. In the ARC/INFO GIS software, there are three overlay functions: *Identity*, *Intersect*, and *Union*. *Identity* overlays points, lines or polygons on polygons and keeps all input map or coverage features. (A coverage is an ARC/INFO term for a digital version of a single map together with attribute and topological information.) *Intersect* overlays points, lines and polygons but keeps only features that fall within areas that are common to both coverages. *Union* overlays polygons and keeps all areas in both coverages. Despite their differences in the operations in the spatial domain, the attributes are managed in the same manner, i.e. the resultant attribute table contains both topological information and attributes values from the two input coverages. The *dissolve* function in ARC/INFO eliminates the boundary between two polygons that have the same attribute values. The resultant attribute table contains new topological information as well as the values of the attribute used for merging. In this prototype system, these functions are modified to handle uncertainty propagation.

6.1 The Fuzzy Overlay Functions

The fuzzy overlay function allows users to overlay two coverages and to specify the operator to propagate the uncertainty of the attributes. Because an ARC/INFO overlay function only generates geometric intersections and copies the attribute values from the two input coverages onto the resultant coverage, the propagation of attribute uncertainty has to be done separately in the database module. In this system, fuzzy attributes in the attribute table are accompanied by their membership grades, which are treated as an indicator to the certainty of the attribute (Table 6.1). The higher the membership grade, the higher the certainty. The fuzzy overlay function first performs an ARC/INFO overlay and then switches into the database module to calculate the propagated uncertainty. In this prototype system only the *AND* and *OR* operators are implemented. Figure 6.1 shows the input screen for the fuzzy overlay function. The user has to provide the input, overlay and output coverage names as required by the normal ARC/INFO overlay functions. In addition, the user has to specify which attributes will be propagated to the output coverage and how the uncertainty should be propagated. The *AND* and *OR* operators are implemented using the Max and Min operations. To overlay the forest coverage with the soil coverage, the overlay operation propagates the uncertainties of SP50 (the dominant species in the forest cover), and TYPE (the soil type) using the AND operator. Table 6.2 shows a partial listing of the polygon attribute table after the overlay operation. The column OPOSS indicates the propagated uncertainty. Some columns in the polygon attribute table have been deleted for clarity of presentation.

(a)

AREA	PERIMETER	FCCLIP_	FCCLIP_ID	SP50	PSP50
590728.50	5683.40	2	1	PL	0.80
26350.27	777.08	3	2	PL	0.60
498084.10	386.42	4	3	SW	0.80
389.95	8.77	5	4	PL	0.80
179157.50	2338.26	6	5	SW	0.50
452904.20	4179.47	7	6	SW	0.80
1132191.00	9806.21	8	7	SB	0.70
428275.50	4495.43	9	8	SW	0.40
136903.30	2702.70	10	9	PL	0.80
19884.88	676.66	11	10	PL	0.70

(b)

AREA	PERIMETER	SCCLIP_	SCCLIP_ID	TYPE	PTYPE
2624024.00	12424.69	2	1	EDS2	0.90
550968.00	8197.86	3	2	BKM1	0.80
32209.48	732.90	4	3	BKM1	0.60
27935.42	780.66	5	4	BMK2	0.80
156540.70	1594.69	6	5	BKM2	0.90

Table 6.1 Partial listing of polygon attribute files for (a) forest cover (b) soil type. SP50 = species with more than 50% coverage. PSP50 = certainty of SP50. TYPE = soil groups, PTYPE = certainty of TYPE.

```
Fuzzy Overlay

Input Coverage        :FCCLIP
Overlay Coverage      :SCCLIP
Output Coverage       :OVERLAY1
Field 1               :TYPE
Operator              :AND
Field 2               :SP50

Identity   Intersect   Union   CANCEL
```

Figure 6.1 Input screen for the fuzzy overlay function.

AREA	PERIM.	SP50	PSP50	TYPE	PTYPE	OPOSS
35186.65	940.97	PL	0.80	EDS2	0.90	0.80
424417.40	4061.37	PL	0.80	BKM1	0.80	0.80
32209.48	732.90	PL	0.80	BKM1	0.60	0.60
23677.91	713.75	PL	0.60	BKM1	0.80	0.60
28053.02	734.91	PL	0.80	BKM1	0.80	0.80
267535.60	2433.19	SW	0.80	BKM1	0.80	0.80
230549.00	3652.87	SW	0.80	EDS2	0.90	0.80
389.95	88.77	PL	0.80	EDS2	0.90	0.80
162544.80	2208.63	SW	0.50	EDS2	0.90	0.50

Table 6.2 Partial listing of a PAT after the fuzzy overlay operation.

6.2 The Fuzzy Dissolve Function

The polygon consolidation function is often used to simplify the coverage after analysis. A simplified coverage not only enhances visual interpretation but also improves storage and processing efficiencies. The *dissolve* function in ARC/INFO merges neighbouring polygons with equal values for one or more selected attributes. This operation is extended to handle consolidation with fuzzy attributes. Figure 6.3 shows a small area of the forest coverage with the species and the certainty scores. When performing the fuzzy dissolve function, the user has to specify the attributes for merging and the minimum level of certainty to be considered in dissolving boundaries.

Figure 6.2 shows the input screen of the fuzzy dissolve function. The user is asked to supply the input and output coverage names. Only adjacent polygons with the same attribute value and the same or higher alpha level will be merged. If alpha is set to 0.0, the fuzzy dissolve operation is equivalent to the standard ARC/INFO dissolve function. Figure 6.4 shows the forest coverage after the fuzzy dissolve operation. Note that the certainty values for merged polygons are set to the specified alpha level to indicate the minimum confidence level for the polygons.

```
Fuzzy Dissolve

Input Coverage       :FCCLIP
Output Coverage      :DISOLVE5
Merge Field          :SP50
Alpha level          :0.7

Process   CANCEL
```

Figure 6.2 Input screen for the fuzzy dissolve function.

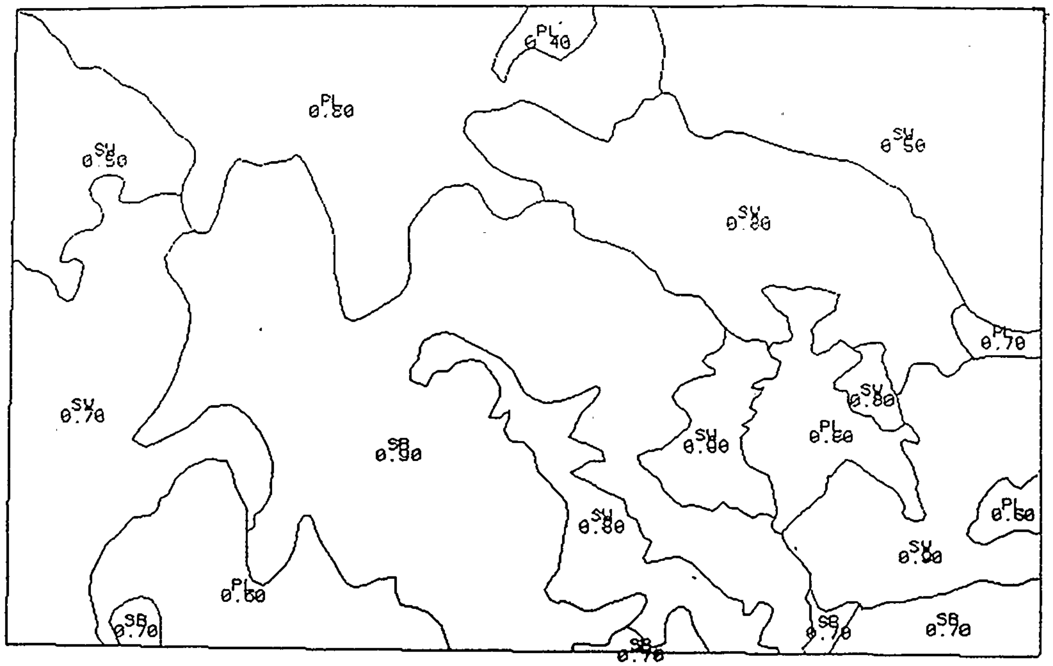

Figure 6.3 Clipped area of the forest coverage.

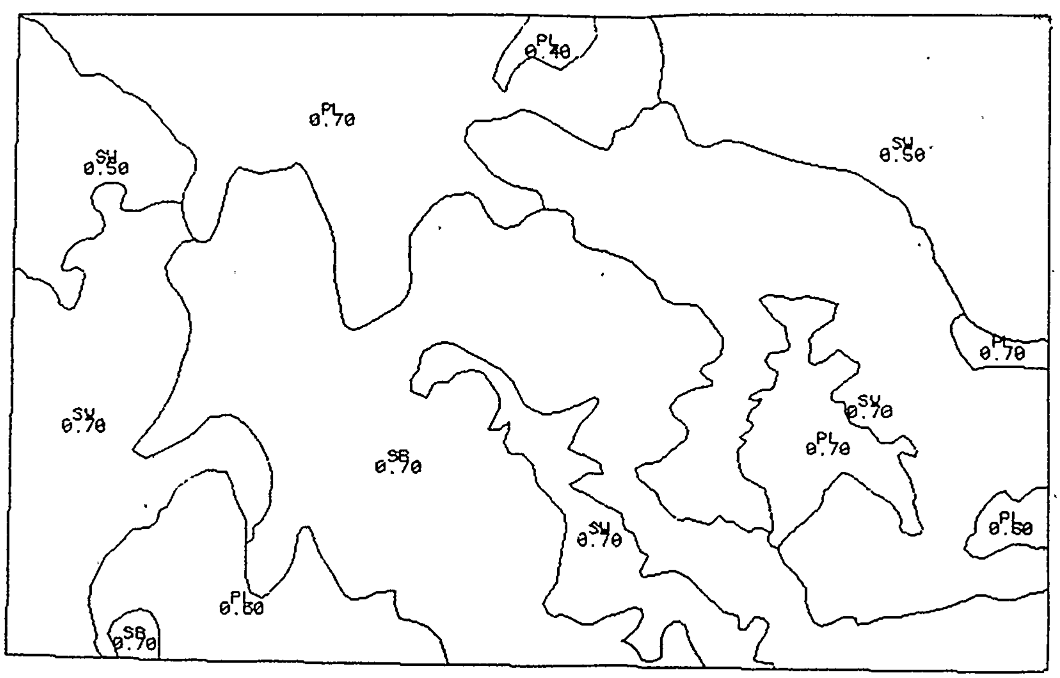

Figure 6.4 Forest coverage after the fuzzy dissolve operation.

6.3 The Display of Uncertainty Information

Graphical display of geographical information is an essential element in GIS. In this system, because uncertainty information is treated as ordinary attributes in the polygon attribute file, they can be manipulated as any other attribute. To achieve effective presentation, three options are provided to display attribute uncertainty. Option one uses the *identify* command in ARC/INFO's ARCPLOT to allow the user to obtain the certainty value of an individual polygon with the cursor. Figure 6.5 shows the screen after the execution of the *identify* option on the a portion of the forest coverage. The attribute value and the certainty value are displayed in the dialogue box on top of the screen. Option two allows the user to view the distribution of uncertainty by a selected attribute value. Figure 6.6 displays the distribution of certainty for white spruce being the dominant species in the forest coverage. Option three allows users to identify all polygons of a particular tree species with a minimum certainty value. The result is shown in Figure 6.7.

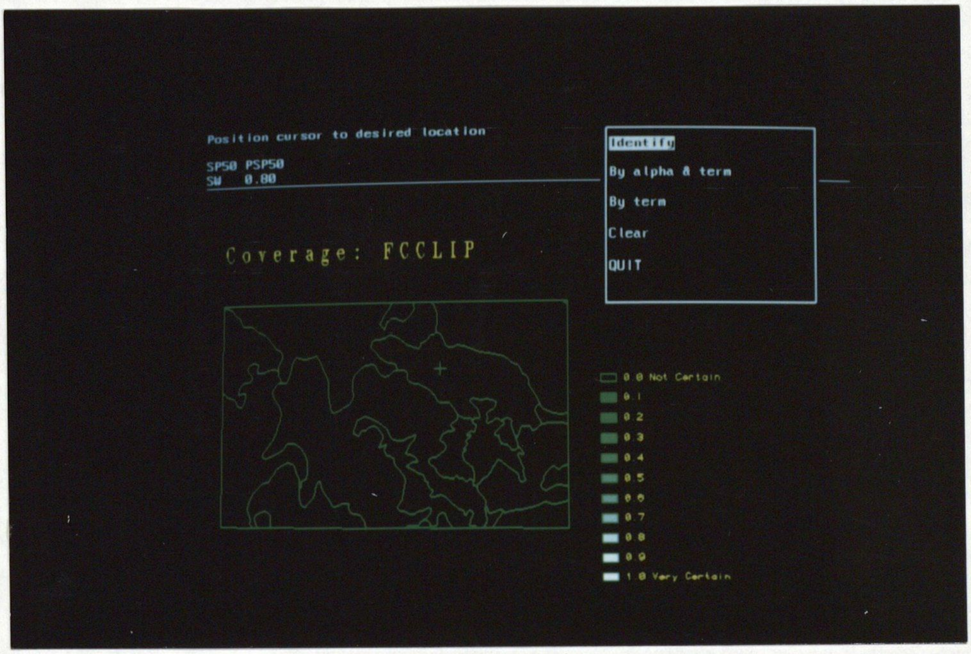

Figure 6.5 Computer screen showing the execution of the *identify* operation.

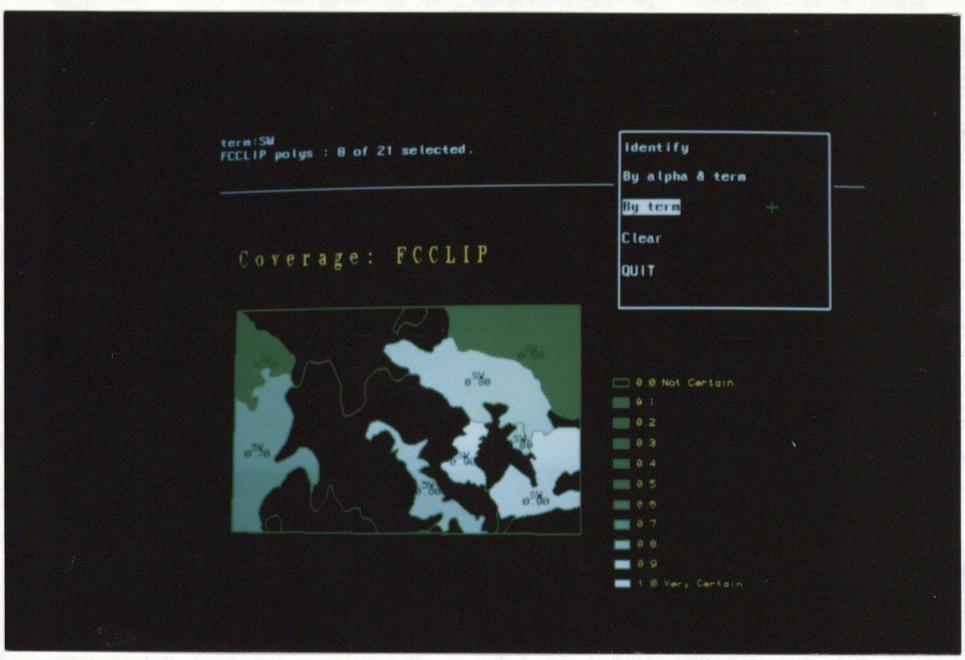

Figure 6.6 Computer screen showing the uncertainty of white spruce.

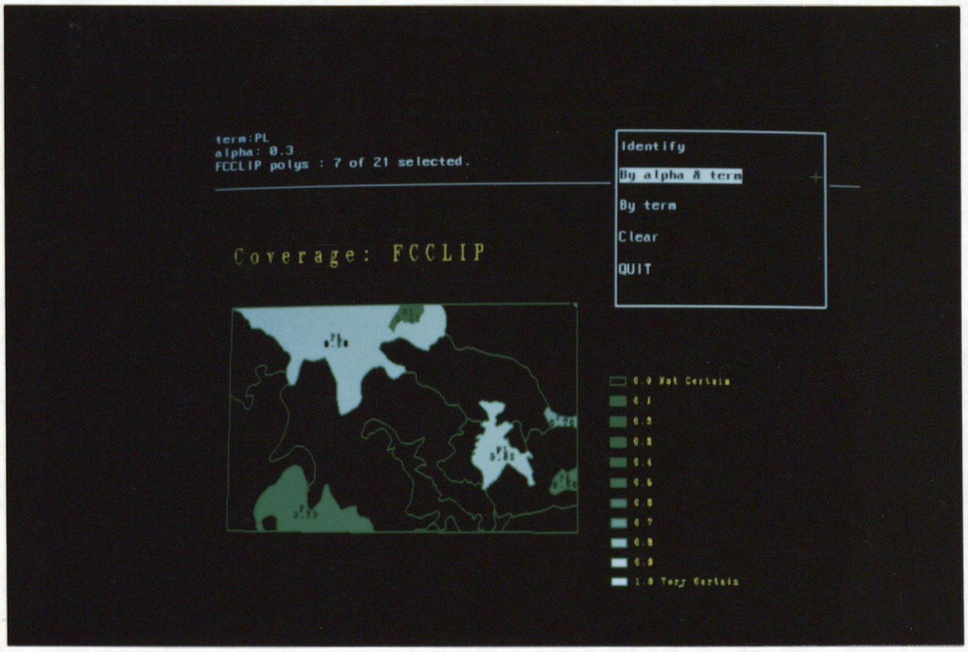

Figure 6.7 Computer screen showing the polygons dominated by lodgepole pine with alpha = 0.3.

6.4 Implementation Considerations

PC ARC/INFO and dBase IV provide sufficient programming functions to implement the basic fuzzy spatial operations. Although not very efficient, these fuzzy spatial operations are not difficult to implement. PC ARC/INFO's overlay functions and dissolve function operate on the spatial domain only. Users are left to handle the attribute separately from the database. This allows the prototype system to store and manipulate uncertainty scores as ordinary attributes. Because the interface between dBase IV and PC ARC/INFO has been built in the ARC shell, access to database is available. This has facilitated the prototype system implementation

Although PC ARC/INFO has a relatively limited macro language, SML has provided useful tools to develop the menu user interface. The menu creation functions enable fast development of friendly and effective user interface. These functions are indispensable tools for system prototyping.

The ARC/PLOT module in PC ARC/INFO provides sufficient color and symbols for the display of uncertainty information. It also allows users to modify color palette and shade symbols with simple commands. The creation of a continuous scale of shades of green (Figure 6.6) involved a few simple commands, which can be stored in a macro program for execution and future reference.

The major difficulties encountered in the implementation of these fuzzy spatial operations in PC ARC/INFO are the limited programming capabilities of SML and the inefficient access to the database from SML. These cause the system to switch back and forth between ARC and dBase, displaying the dBase copyright message every time the program accesses the database. In addition, dBase runs extremely slowly inside the PC ARC/INFO shell. This slowed down the classification process significantly. Therefore, the classification function is programmed to run both inside and outside of the shell. For large processes (more than 50 polygons) the classification process should be run outside the shell. In addition, because dBase cannot be accessed through the ARCPLOT module, database search cannot be efficiently performed unless one leaves ARCPLOT. The display options could be improved if database access were more efficient. To illustrate, in the database the classification results were stored in a table as shown in Table 6.3. If the user wants to view all the polygons that could be assigned to c1, regardless of the alpha level, the system should search all three columns and display all four polygons. Nevertheless, due to the limited access to the database from ARCPLOT, the current display options can only select polygons by the most possible class, i.e. A1, and only polygon 1 would be displayed. It should be noted that all these limitations exist in PC ARC/INFO and dBase only. ARC/INFO running on the workstation platforms can be linked to different database management systems which have a smoother interface between the two software packages. Furthermore, the workstation version of ARC/INFO has a more flexible macro language (Arc Macro Language or AML) which provides more efficient access to the database.

polygon	A1	A1 poss.	A2	A2 poss.	A3	A3 poss.
1	c1	0.8	c5	0.4	c3	0.1
2	c4	0.6	c2	0.2	c1	0.1
3	c2	0.6	c1	0.3	c2	0.1
4	c2	0.7	c3	0.5	c1	0.2

Table 6.3 Example of a fuzzy attribute with multiple assignment.
A1=first assignment, A2=second assignment, A3= third assignment.

CHAPTER 7

Conclusions and Recommendations

7.1 Conclusions

This thesis introduces the theory of fuzzy sets and applies several fuzzy data management techniques to handle uncertainty due to vagueness and imprecision. Although vagueness and imprecision in quantitative and qualitative data represent only one of many sources of uncertainty identified in GIS, they are by no means negligible. This thesis has concentrated on the management of this aspect of uncertainty in GIS and has shown that fuzzy techniques can be applied to handle fuzziness in data, conceptual objects and operations in a forest ecological databases.

1) Research into the application of fuzzy set theory in GIS has demonstrated that fuzzy techniques allow the representation of continuity in cartographic modelling [Sui, 1990], reduce information loss [Wang et al., 1990] and provide superior performance over conventional techniques in data retrieval and processing [Robinson, 1984; Burrough, 1989]. This research project has also showed that fuzzy sets permit flexible and realistic representations of the subjectivity and vagueness in qualitative terms, imprecision in quantitative values as well as ill-defined concepts. Thus, fuzzy set-based techniques should be considered desirable tools in decision supports GIS.

2) At present, no commercial GIS software provides fuzzy set-based tools to its users. The prototype system developed in this research project has provided some insights into the incorporation of such techniques in a commercial GIS software. In general, any GIS package that interfaces with a database management system which has programming capability will be able to implement some fuzzy set-based techniques. The dBASE IV programming language and the SML macro language provide sufficient tools to achieve the goals listed in Section 3.1. Although dBASE IV and SML are not the best available database management system to implement these fuzzy techniques, they do have some functions which are useful for implementing these procedures. Examples are the Max and Min functions, and the flexible indexed search function for quick access to look-up tables. On the other hand, the limited programming capability and database access of SML complicated the implementation. This has caused PC ARC/INFO to be a less favourable choice for the practical implementation of a fuzzy logic-based system.

3) The main concerns in the implementation of fuzzy set-based techniques are storage and processing efficiency. Simulating fuzzy logic in a Boolean logic-based environment can be effective but not efficient. To utilize fully the power of fuzzy logic, hardware and software should be redesigned [Zadeh, 1984]. In this project, only a small data set was used for demonstration, and the processing time is still acceptable. The results show however that for a realistic GIS application with large data sets (e.g. several thousand polygons), the computation load may be very high and may result in unacceptably slow performance. In addition, as most GIS are already suffering from information overload, the extra information available in the

fuzzy environment will add to storage and processing requirements of GIS. This would indicate that implementors must consider the most efficient possible internal integration of the fuzzy logic processes, and the use of fast powerful processors. This would probably preclude effective implementation on current "PC" technology.

4) The biggest concern in implementing a fuzzy system is the formulation of the membership functions. As many researchers have commented, the need to develop some general guidelines for the derivation of membership functions should be addressed in order to increase the credibility of fuzzy systems. Turksen [1986] noted two approaches to determine membership functions. The normative approach is suited for deriving membership functions that are inherently subjective. An example is linguistic terms of human languages. The empirical approach follows the objective experimental procedures of the scientific methods in measurement theory used in mathematical psychology, but little work has been published on this subject. Like other studies, the membership functions in this research were derived by the forest ecologists based on published information [Corns & Annas, 1986] and statistics on field data. Though not all the membership functions were derived objectively, many of them are supported by the published data and descriptive statistics. Therefore, subjectivity has only minor influence on this prototype system.

7.2 Recommendations for Future Research

1) As noted by many fuzzy set researchers, one urgent need in the future research is to develop some general guide-lines for the formulation of the membership functions. Although fuzzy sets can be used to express subjective interpretations, objective formulation should be used whenever available. It would be interesting and useful to derive some frequently used membership functions for certain applications. For instance, semantics for soil drainage and texture should be relatively standardized. In fact many of these commonly used scales are published. An example is the soil texture triangle. Deriving a generally accepted fuzzy representation such as a fuzzified soil texture triangle will be useful in the development of fuzzy systems.

2) Another research recommendation is to apply these concepts to manage spatial uncertainty. Spatial uncertainty is unique to GIS and relatively little research has been carried out to model its uncertainty. A more comprehensive research topic is to study and model the combined effect of spatial and attribute uncertainty. In many instances, spatial and attribute uncertainty are dependent of each other. For instance, the boundary between two distinctly different soil types can be determined more accurately than that of two similar soil types. The combination of spatial and attribute uncertainty will provide the user with sufficient information to assess the reliability of geographical databases.

3) The third research recommendation is to extend this prototype system with fuzzy reasoning ability. The concepts of linguistic variables and fuzzy concept modelling were developed to facilitate human-like reasoning in a computer environment. With these concepts already implemented, it is

reasonable to fully utilize these structures to increase the intelligence of the system. Zadeh [1976] believes that fuzzy set theory and the associated concepts in the right direction to develop a more human-like system.

4) The implementation of fuzzy data management techniques as an add-on module in a commercial vector GIS such as PC ARC/INFO may not be the best approach for building a fuzzy logic-based GIS. This is because a fuzzy system requires much more computational power and extra storage. At present, most commercial GIS are already suffering from slow processing and information overload. Using the macro languages provided by the GIS software or database management systems requires an extra level of interpretation and drastically slow down the process of fuzzy operations. An alternative is to implement the fuzzy operations as built-in functions. This will require redesigning of the GIS to allow optimal incorporation fuzzy operations. Effort should be directed to investigate the incorporation of fuzzy operations as built-in functions in GIS.

REFERENCES

Baldwin JF & Zhou SQ, 1984. A Fuzzy Relational Inference Language. *Fuzzy Sets and Systems,* 14, 155-174.

Bardossy P, Dogardi JA, Kelly L, 1989. Geostatistics Utilizing Imprecise (Fuzzy) Information. *Fuzzy Sets and Systems* , Vol. 31, 331-328.

Bellman R & Giertz M, 1973. On the Analytic Formalism of the Theory of Fuzzy Sets. *Information Science 5,* 149-156.

Bédard Y, 1987. Uncertainties in Land Information Systems Databases. *Proceedings Auto-Carto 8,* 175-184.

Bouchon-Meunier B, Yager RR & Zadeh LA, 1991. Hybrid Approaches to Uncertainty. *Uncertainty in Knowledge Bases, Chapter 7. Third International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU'90 Paris France,* 360-413.

Boy GA & Kuss PM, 1986. A Fuzzy Method for the Modelling of Human-Computer Interaction in Information Retrieval Tasks. *Application of Fuzzy Set Theory in Human Factors.* Karwowski & Mital eds. Elsevier Science Publishers, Amsterdam, 189-199.

Buckles BP & Petry FE, 1982. A Fuzzy Representation of Data for Relational Databases. *Fuzzy Sets and Systems,* 7, 213-226.

Buckles BP & Petry FE, 1983. Information-theoretic Characterization of Fuzzy Relational Database. *IEEE Trans. Systems Man and Cybernet. SMC-*13(1), 74-77.

Buckles BP & Petry FE, 1984. Extending the Fuzzy Database with Fuzzy Numbers. *Information Science* 34(2), 145-155.

Buckley JJ, 1990. Managing Uncertainty in a Fuzzy Expert System. *Machine Learning and Uncertainty Reasoning*.. Gaines & Boose eds. Academic Press, 191-210.

Burrough PA, 1986 *Principles of Geographic Information Systems for Land Resources Assessment*. Clarendon Press, Oxford.

Burrough PA, 1989. Fuzzy Mathematical Methods for Soil Survey and Land Evaluation. *Journal of Soil Science*, 40, 477-492.

Chatterji BN, 1985. Fuzzy Set Theoretic Approach for Scene Analysis. *Approximate Reasoning in Expert Systems*. Gupta, Kandel, Bandler & Kiszka eds. Elsevier Science Publishers B.V. North-Holland. 673-654.

Corns IGW & Annas RM, 1986. *Field Guide to Forest Ecosystems of West-Central Alberta*. Canadian Forestry Services North Forestry Center Edmonton Canada.

Dubois D & Prade H, 1985. A review of Fuzzy set Aggregation Connectives. *Information Sciences*, 36, 85-121.

ESRI, 1989. *ARC/INFO Users Guide*, Vol. 1. Environmental Systems Research Institute, Inc., Redlands, CA. USA.

Guptill SC, 1989. Inclusion of Accuracy Data in a Feature Based, Object-oriented Data Model. *The Accuracy of Spatial Databases*. M Goodchild & S Gopal eds. Taylor and Francis, 91-98.

Hartigan JA, 1983. *Bayes Theory*. Springer-Verlag. New York Inc.

Kaufmann A & Gupta MM, 1985. *Introduction to Fuzzy Arithmetic: Theory and Applications*. Van Nostrand Reinhold Company, New York.

Kollias VJ & Voliotis A, 1991. FRSIS: a Soil Information System with Fuzzy Retrieval Capability. *Int. J. Geographic Information Systems* 5(2), 209-224.

Krantz DH, Luce RD, Suppes P & Tversky A. *Foundations of Measurement.* Vol. 1, Academic Press. New York.

Kruse R, Schwecke E & Heinsohn J, 1991. *Uncertainty and Vagueness in Knowledge Based Systems.* Springer-Verlag Berlin Heidelberg. 262-277.

Leung Y 1988. *Spatial Analysis and Planning under Imprecision,* Elsevier Science Publishers B.V. North-Holland.

Leung Y 1990. Fuzzy Logic and Knowledge-based GIS. *Proceedings, IGRASS* 1989 Vol. 1, 47-50.

Lui D & Li Y, 1990. A *Fuzzy Prolog Relation Database.* Academic Press.

Miller RG, Karimi HA & Feuchtwanger M, 1989. Uncertainty and Its Management in Geographic Information Systems. *Proceedings of National GIS Conference,* Ottawa, Canada. 252-260.

Norwich AM & Turksen IB, 1982. The Construction of Membership Functions. *Fuzzy Set and Possibility Theory: Recent Developments.* Yager ed, Pergamon Press Inc. 110-145.

Oden GC & Lopes LL, 1982. On the Internal Structure of Fuzzy Subjective Categories. *Fuzzy Set and Possibility Theory: Recent Developments.* Yager ed, Pergamon Press Inc. 75-89.

Robinson VB & Frank AU, 1985. About Different Kinds of Uncertainty in Collections of Spatial Data. *Proceedings of Auto-Carto 7,* 440-449.

Robinson VB & Strahler AH, 1984. Issues in Designing Geographic Information Systems Under Conditions of Inexactness. *Proceedings, 10th Intern Symp Machine Processing of Remotely Sensed Data,* 198-204.

Robinson VB, 1988. Some Implications of Fuzzy Set Theory Applied to Geographic Databases. *Comput. Environ. and Urban Systems*, Vol. 12, 89-97.

Sack IH, Giardina DR & Sinha D, 1983. Investigations into Fuzzy Relational Database. *Paper presented at Meeting of North America Fuzzy Information Processing Society*, Schenectady, New York.

Shafer G, 1976. *A Mathematical Theory of Evidence.* Princeton University Press.

Shannon CE, 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379-423, 623-656.

Shenoi S & Melton A, 1989. Proximity Relations in the Fuzzy Relational Database Model. *Fuzzy Sets and Systems* 31(3), 285-296.

Stoms D, 1987. Reasoning with Uncertainty in Intelligent Geographical Information Systems. *Second International Conference Exhibits & Workshops on Geographic Information Systems.* San Francisco, 693-700.

Sui D, 1990. A Fuzzy Mathematical Modelling Technique and Its Implementation in Geographic Information System: Toward a Fuzzy Cartographic Modelling Technique. *GIS/LIS Proceedings*, Anaheim California. Vol.1, 234-244.

Thöel U, Zimmermann HJ and Zysno P, 1979. On the Suitability of Minimum and Product Operators for the Intersection of Fuzzy Sets, *Fuzzy Sets and Systems* 2, 167-180.

Turksen IB, 1986. Measurement of Membership Functions. *Applications of Fuzzy Set Theory in Human Factors.* W Karwowski & A Mital eds. Elsevier Science Publishers, Amsterdam.

Walsh SJ, Lightfoot DR & Bulter DR, 1987. Recognition and Assessment of Error in Geographic Information Systems. *Photogrammetric Engineering and Remote Sensing*. Vol. 53, No. 10, 1423-1430.

Wang F, Hall GB & Subaryono, 1990. Fuzzy Information Representation and Processing in Conventional GIS software: Database Design and Application. *Int. J. Geographical Information Systems* Vol. 4 #3, 261-283.

Zadeh LA, 1965. Fuzzy Sets. *Information and Control* 8, 338-353.

Zadeh LA, 1971. Similarity Relations and Fuzzy Orderings. *Information Sciences,* 3, 177-200.

Zadeh LA, 1973. Outline of a New Approach to the Analysis of Complex Systems and Decision Process. *IEEE Trans. Systems, Man, and Cybernetics,* SMC-3(1973), 28-44.

Zadeh LA, 1975. The Concept of a Linguistic Variable and its Application to Approximate Reasoning. Part 1 and 2. *Information Sciences,* 8, 199-249, 201-357.

Zadeh LA, 1976. A Fuzzy-Algorithmic Approach to the Definition of Complex or Imprecise Concepts. *Int. J. Man-Machine Studies,* 8, 249-291.

Zadeh LA, 1978. Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems,* 1, 3-28.

Zadeh LA, 1984. Coping with the Imprecision of the Real World: An Interview with Lotfi A. Zadeh. *Comm. ACM,* 27, 304-311.

Zemankova-Leech M and Kandel A, 1985. Uncertainty Propagation to Expert Systems. *Approximate Reasoning in Expert Systems.* Gupta, Kandel, Bandler & Kiszka eds. Elsevier Science Publishers B.V. North-Holland, 529-548.