

HIERARCHICAL FUZZY PATTERN MATCHING FOR THE  
REGIONAL COMPARISON OF LAND USE MAPS

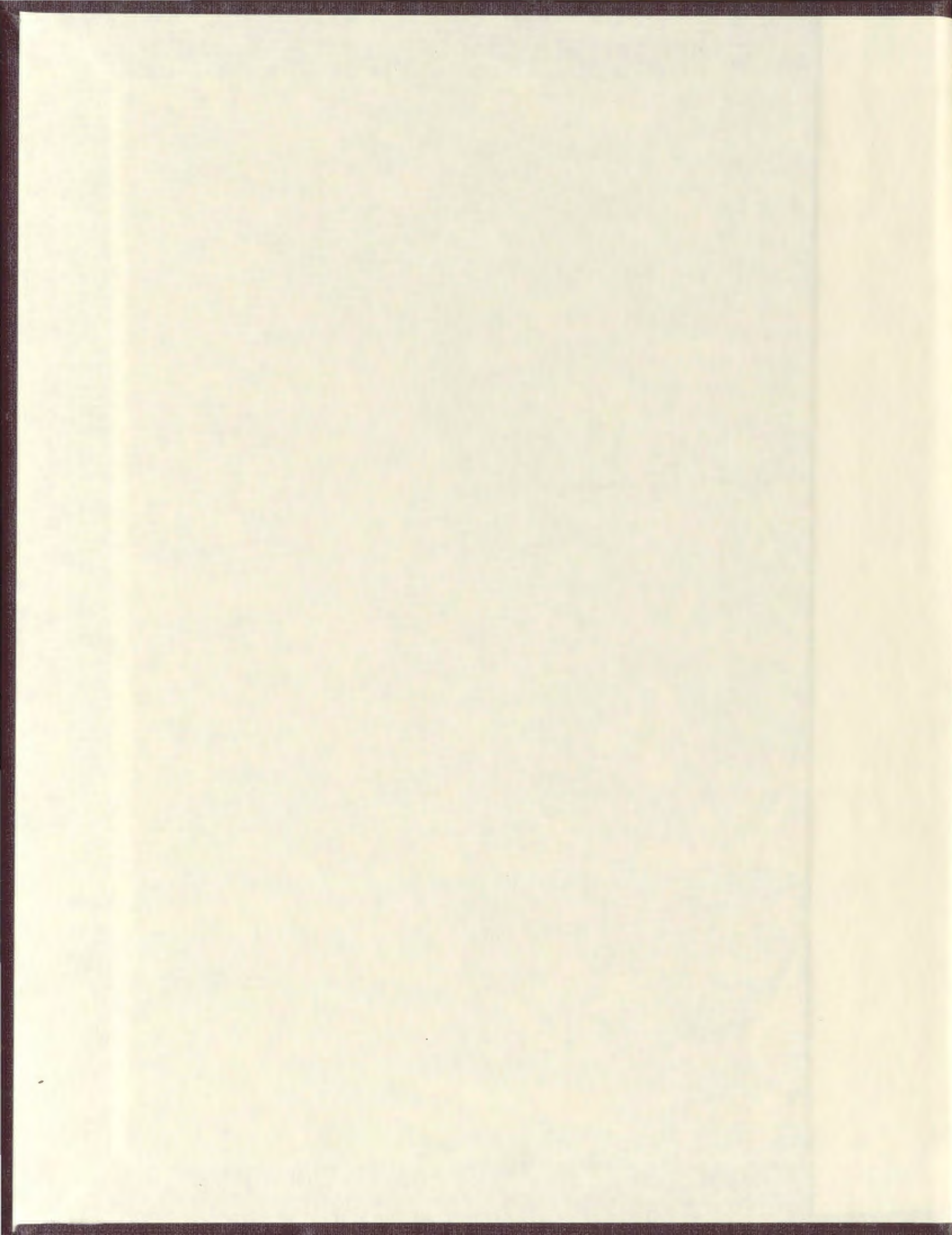
CENTRE FOR NEWFOUNDLAND STUDIES

---

**TOTAL OF 10 PAGES ONLY  
MAY BE XEROXED**

(Without Author's Permission)

CONRAD POWER



## **INFORMATION TO USERS**

**This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.**

**The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.**

**In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.**

**Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.**

**Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.**

**Bell & Howell Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA**

**UMI<sup>®</sup>**  
**800-521-0800**





**National Library  
of Canada**

**Acquisitions and  
Bibliographic Services**

**395 Wellington Street  
Ottawa ON K1A 0N4  
Canada**

**Bibliothèque nationale  
du Canada**

**Acquisitions et  
services bibliographiques**

**395, rue Wellington  
Ottawa ON K1A 0N4  
Canada**

*Your file Votre référence*

*Our file Notre référence*

**The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.**

**The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.**

**L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.**

**L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.**

**0-612-42427-8**





# Memorial

University of Newfoundland

This is to authorize the Dean of Graduate Studies to deposit two copies of my thesis/report entitled

HIERARCHICAL FUZZY PATTERN MATCHING FOR THE  
REGIONAL COMPARISON OF LAND USE MAPS

in the University Library, on the following conditions. I understand that I may choose only ONE of the Options here listed, and may not afterwards apply for any additional restriction. I further understand that the University will not grant any restriction on the publication of thesis/report abstracts.

(After reading the explanatory notes at the foot of this form, delete TWO of (a), (b) and (c), whichever are inapplicable.)

The conditions of deposit are:

~~(a)~~ that two copies are to be made available to users at the discretion of their custodians,

OR

(b) that access to, and quotation from, this thesis/report is to be granted only with my written permission for a period of one year from the date on which the thesis/report, after the approval of the award of a degree, is entrusted to the care of the University, namely, OCTOBER 19 98, after which time the two copies are to be made available to users at the discretion of their custodians,

OR

~~(c)~~ that access to, and quotation from, this thesis/report is to be granted only with my written permission for a period of \_\_\_\_\_ years from the date on which the thesis/report, after approval for the award of a degree, is entrusted to the care of the University; namely, \_\_\_\_\_, 19 \_\_\_\_; after which time two copies are to be made available to users at the discretion of their custodians.

Date August 26, 1998  
C. P. Allen  
Dean of Graduate Studies

Signed Conrad Power  
Witnessed by Alvin Simms

## NOTES

1. Restriction (b) will be granted on application, without reason given.

However, applications for restriction (c) must be accompanied with a detailed explanation, indicating why the restriction is thought to be necessary, and justifying the length of time requested. Restrictions required on the grounds that the thesis is being prepared for publication, or that patents are awaited, will not be permitted to exceed three years.

Restriction (c) can be permitted only by a Committee entrusted by the University with the task of examining such applications, and will be granted only in exceptional circumstances.

2. Thesis writers are reminded that, if they have been engaged in contractual research, they may have already agreed to restrict access to their thesis until the terms of the contract have been fulfilled.

**Hierarchical Fuzzy Pattern Matching for the Regional  
Comparison of Land Use Maps**

**Conrad Power**

**8204968**

**Submitted in partial fulfillment of  
the requirements for the Degree of  
Masters of Science**

**Department of Geography  
Memorial University of Newfoundland**

**© August 24, 1998**

## **Abstract**

The evaluation of the spatial similarities between two raster maps is traditionally based on pixel-by-pixel comparison techniques. These procedures determine the number of cells in agreement for each landuse category and express the overall agreement with a boolean global similarity value. The problem with a pixel-by-pixel comparison is that a small displacement in pixels will be registered as disagreement even though the land use patterns between the maps maybe essentially the same. The issues of unique polygons mapping and hierarchical fuzzy pattern matching, where the maps are compared on both a local and global level, emerge as viable and robust alternatives. The local matchings determine the degree of containment of each unique polygon to its spatial counterparts on the original maps in terms of fuzzy areal intersections. Local agreement values for the unique polygons based on their polygon area containments are calculated from fuzzy logical Max-Min compositional algorithms. A global agreement value is derived by the fuzzy summation of the local matchings. The uses of these basic methods are discussed and further refinements and modeling possibilities are outlined.



## **Table of Contents**

|   |    |
|---|----|
| List of Tables.....   | iv |
| List of Figures.....  | vi |
| 1.0 Introduction.....   | 1  |
| 1.1 Introduction.....   | 1  |
| 1.2 Statement of the Objectives.....                                  | 5  |
| 1.3 Thesis Organization.....  | 7  |
| 2.0 Related Research.....   | 8  |
| 2.1 Introduction.....   | 8  |
| 2.2 Traditional Post Classification Comparison of Thematic Maps.....  | 9  |
| 2.3 The Fundamentals of Fuzzy Set Theory.....                         | 12 |
| 2.3.1 Basic Definitions of Fuzzy Set Properties.....                  | 14 |
| 2.3.2 Operations of Fuzzy Sets.....                                   | 15 |
| 2.4 Spatial Similarity Applications Involving Fuzzy Set Theory.....   | 17 |
| 2.4.1 Fuzzy Set Theory and Land Use Change Detection.....             | 21 |
| 2.5 Conclusion.....   | 25 |
| 3.0 Methodology.....  | 26 |
| 3.1 Introduction.....   | 26 |
| 3.2 Unique Polygons Mapping.....                                      | 28 |
| 3.3 Development of the Fuzzy Inference System for Local Matching..... | 35 |
| 3.3.1 Creation of Membership Functions.....                           | 36 |

|   |     |
|---|-----|
| 3.3.2 Fuzzification.....  | 44  |
| 3.3.3 Rule Based Inference.....   | 45  |
| 3.3.4 Defuzzification.....  | 50  |
| 3.3.5 Global Matching.....  | 51  |
| 3.3.6 Estimation of Uncertainty in the Global Matching.....                 | 53  |
| 3.4 Conclusion.....   | 55  |
| 4.0 Results.....  | 56  |
| 4.1 Introduction.....   | 56  |
| 4.1.1 Description of the Datasets.....                                      | 57  |
| 4.2 Linguistic Evaluations of Map Similarities.....                         | 58  |
| 4.3 Local Matching Results.....   | 60  |
| 4.4 Global Matching Evaluation.....   | 75  |
| 4.4.1 Evaluation of Boolean versus Fuzzy Global Similarity.....             | 79  |
| 4.5 Fuzzy versus Boolean Land Use Comparison Results.....                   | 82  |
| 4.6 Conclusion.....   | 91  |
| 5.0 Discussion and Conclusions.....   | 94  |
| 5.1 Discussion of Results.....  | 94  |
| 5.2 Conclusions.....  | 96  |
| 5.3 Future Research.....  | 97  |
| 6.0 References.....   | 101 |
| Appendix A Unique Polygons Mapping and Fuzzy Inference System Programs..... | 108 |

## List of Tables

|  |    |
|--|----|
| Table 2.2.1 Example Post Classification Agreement Matrix.....  | 10 |
| Table 3.2.1 Example of Updated Unique Polygons Attribute Table for Figure 3.2.1.....                   | 34 |
| Table 3.2.2 Boolean Local Matching Information for Map One .....                                       | 34 |
| Table 3.3.1.1 Input Linguistic Local Matching Interpretations .....                                    | 37 |
| Table 3.3.1.2 Sigmoidal Membership Function Parameters .....   | 39 |
| Table 3.3.1.3 Generalized Bell Membership Function Parameters .....                                    | 41 |
| Table 3.3.1.4 Pixel Group Sigmoidal Function Parameters .....  | 43 |
| Table 3.3.1.5 Linguistic Labels and Parameters for the Output Membership Functions...                  | 44 |
| Table 3.3.3.1 Rules for the Local Matching Fuzzy Inference System .....                                | 47 |
| Table 4.3.1 Local Matching for Land Use Polygons of Cinc1.....   | 65 |
| Table 4.3.2 Evaluation of Matches and Mismatches for Cinc1.....  | 67 |
| Table 4.3.3 Matching Results for Template Polygons for Cinc1 Comprised of 1 or 2<br>Pixels.....        | 67 |
| Table 4.3.4 Local Matching for Land Use Polygons of Cinc2.....   | 68 |
| Table 4.3.5 Evaluation of the Matches and Mismatches of Cinc2.....                                     | 70 |
| Table 4.3.6 Matching Results for Template Polygons of Cinc2 Comprised of 1 or 2<br>Pixels.....         | 71 |
| Table 4.3.7 Local Matching for Land Use Polygons of Forest1.....                                       | 73 |
| Table 4.3.8 Evaluation of the Matches and Mismatches of Forest1.....                                   | 73 |
| Table 4.3.9 Matching results for Template Polygons for Forest1 Comprised of One<br>and Two Pixels..... | 73 |
| Table 4.4.1 Boolean and Fuzzy Global Similarity Statistics.....  | 80 |

|   |           |
|---|-----------|
| <b>Table 4.5.1 Local Matching Possibility Threshold Values.....</b> | <b>83</b> |
|---|-----------|

## List of Figures

|   |    |
|---|----|
| Figure 2.3.1 Membership of X in the Fuzzy Set Adolescent Age.....                     | 13 |
| Figure 3.1.1 Flow Chart of the Methodology for Hierarchical Fuzzy Pattern Matching .. | 29 |
| Figure 3.2.1 Land Use Similarity Polygons.....  | 28 |
| Figure 3.2.2 Structure of a Unique Polygons Map and Attribute Table .....             | 32 |
| Figure 3.3.1 Four Stages of Designing a Mamdani Fuzzy Inference System.....           | 36 |
| Figure 3.3.1.1 Membership Functions for Areal Intersection Linguistic Values.....     | 39 |
| Figure 3.3.1.2 Membership Functions for Pixel Groupings .....                         | 42 |
| Figure 3.3.1.3 Local Matching Output Membership Function .....                        | 44 |
| Figure 3.3.3.1 Rulebase and Inference Structure of the Fuzzy Inference System .....   | 49 |
| Figure 4.2.1 Land Use Maps of Cinc1.....  | 59 |
| Figure 4.2.2 Land Use Maps of Cinc2.....  | 61 |
| Figure 4.2.3 Forest Inventory Maps of Forest1 .....                                   | 62 |
| Figure 4.3.1 Local Template Polygon Matching for Cinc1.....                           | 63 |
| Figure 4.3.2 Lowess Curve of Mean Polygon Size versus Local Matching for Cinc1.....   | 66 |
| Figure 4.3.3 Local Template Polygon Matchings for Cinc2.....                          | 69 |
| Figure 4.3.4 Lowess Curve of Mean Polygon Size versus Local Matching for Cinc2.....   | 70 |
| Figure 4.3.5 Local Template Polygon Matchings for Forest1.....                        | 72 |
| Figure 4.3.6 Lowess Curve of Mean Polygon Size versus Local Matching for Forest1...   | 74 |
| Figure 4.5.1 Fuzzy (A) versus Boolean (B) Agreement for Cinc1.....                    | 84 |
| Figure 4.5.2 Fuzzy (A) versus Boolean (B) Agreement for Cinc2.....                    | 85 |

|   |    |
|---|----|
| Figure 4.5.3 Boolean Change Classes for Cleared and Non-Forested.....     | 88 |
| Figure 4.5.4 Fuzzy Change Classes for Cleared and Non-Forested.....       | 90 |
| Figure 4.5.5 Fuzzy Change Possibilities for Cleared and Non-Forested..... | 92 |

## Acknowledgements

This thesis is the product of the financial and personal support of many people. The primary financial backing was provided by a SSHERC grant with Dr. Roger White and Dr. Alvin Simms. Additional funding was supplied by the School of Graduate Studies at Memorial University of Newfoundland in the form of university fellowships and teaching assistantships from the Department of Geography.

On a personal note, I would especially like to thank my supervisor, Dr. Alvin Simms, for his advice and guidance throughout both my undergraduate and graduate programs. Special thanks to Dr. Roger White for his suggestions concerning the development of the fuzzy inference system and for pushing and prodding me to finish the written section of the thesis. Thanks also to Dr. Elizabeth Simms for the help in the application and presentation of this production. I am grateful to my good friends, Harry Barnes, Andrea Bassan, Karl Keough, and Paul Brett, for their humour and encouragement throughout my time as a graduate student.

Lastly and most importantly, I would like to thank my parents and family who have always been a constant source of support, love, and encouragement.



# **Chapter 1**

## **Introduction**

### **1.1 Introduction**

The study of land use dynamics has been an important theme in geographical research, with the measurement of differences in land use being fundamental to understanding certain socio-economic patterns and processes. The detection of land use differences involves estimating the temporal and spatial extents of the regions of land use disagreement from a direct comparison of two maps, aerial photographs, satellite images, or auxiliary information sources (Singh,1989). Generally, spatial models that perform a comparison of land use maps are designed to either detect land use change between multitemporal maps or to measure the similarities between a predicted map and an actual reference map.

The detection of land use change from a comparison of multitemporal maps is based on identifying and measuring the disagreements in the patterns of land use from one time period to another. The types of change can range from short term phenomena such as snow cover or flooding to long term events such as deforestation or urban fringe development (Cracknell and Hayes,1991). From an interpretation of where the land use maps are spatially different, a researcher can determine the physical, political, and economic factors responsible for the change. Consequently, the comparison of the land use maps is often an essential step needed for modeling spatial land use processes.

A map comparison similarity analysis is conceptually and operationally identical to land use change detection except that the comparison involves atemporal maps. The goal of a similarity analysis is to gauge the performance of a land use prediction or forecasting model, such as a cellular automata, by comparing its predicted results to a map of actual land use. With the degree of similarity between the maps expressed as a quantitative agreement value, an analyst is given a idea of the accuracy of the predicted map.

The simplest way to identify differences between two land use maps is to qualitatively determine the areas of disagreements from a visual comparison of the maps. This approach is commonly used for the interpretation of large scale aerial photographs (Jensen,1981). An analyst would manually delineate the regions that he has subjectively determined to be dissimilar to produce a change or difference layer. Despite its ease of implementation, several authors (Jensen, 1981; Adeniyi, 1980) have stated that the visual comparison of photos is a slow, tiring process that often results in errors of omission. As a result, the preferred methods are those that automatically correlate and compare two maps to detect land use differences.

Since the 1970s, researchers have developed and improved computerized techniques for modeling geographical change. Geographical Information Systems (GIS) and remote sensing research have consequently led to the development of a number of digital change detection algorithms. Although these procedures were primarily developed for land use change detection, they are equally suited for a similarity analysis. The basic premise of using GIS and remote sensing for change detection or similarity analysis is that land use differences will lead to areal disagreements between the maps. A fundamental

assumption of any digital change detection method is that there exists a difference in the value of a pixel on two dates if the land use converts from one type to another (Singh,1989).

Many of the change detection procedures have come from the remote sensing field. These digital methods of discriminating change involve the pairwise comparison of raster land use maps (Lo and Shipman, 1990). In other words, two maps are compared on a pixel-by-pixel basis to identify areas of contrasting reflectance values or thematic attributes. Most of the change detection models discussed in the remote sensing literature identify change as disagreements in spectral reflectance between multitemporal imagery (Jensen,1981; Singh,1989). Two types of procedures for using multispectral reflectance data in change analysis are commonly used. The first type utilizes arithmetic operations between the same spectral bands of bitemporal images (Gong,1993). The most common arithmetic change procedures are image differencing (Singh,1989; Lo and Shipman, 1990; Weismiller *et al.*,1977) and image ratioing (Wickware and Howarth,1981; Renez,1985). Change component images for each spectral band are created by subtracting or dividing the multispectral data of the images. Finally, the component images are classified according to boolean threshold values to indicate whether significant or relevant change has occurred.

The second type of multispectral change analysis utilizes statistical image transformation models, such as change vector analysis (Malila,1980; Johnson,1994; Lamblin,1996) and principal components analysis (Byrne *et al.*,1980; Ingebritsen and Lyon,1985; Gong,1993). Image transformation algorithms analyze the variance in a

multispectral dataset to reduce the number of spectral bands to a relatively smaller number of change components. This enables the dimensionality of the change detection process to be decreased without losing useful information. A land use change map is usually generated through the interpretation and thresholding of the change components.

An important consideration about a multispectral change analysis is that it involves the comparison of unclassified satellite images. Alternative procedures exist that identify change by comparing independently produced classification layers or thematic maps. They are often referred to as post classification comparisons (Wickware and Howarth, 1981; Hodgson *et al.*, 1988; Jensen *et al.*, 1987). Technically, the comparison algorithm performs a pixel-by-pixel crosstabulation of two thematic maps to generate a map and associated statistics that indicate regions of change.

The work discussed in this thesis is concentrated on the use of fuzzy set theory for the development of an improved post classification comparison model. Since the basis of the thesis is the comparison of thematic land use maps, the multitemporal change procedures are not relevant to the research and are not discussed beyond this chapter. However, it should be noted that the methodology presented in subsequent chapters can be adapted to compare multispectral data.

In post classification change detection the accuracy of the results depends on the accuracy of the categorical classifications and geometric registration of the maps. Because the maps are compared on a pixel-by-pixel basis, this method produces poor results if the maps are not extremely accurate, both locationally and thematically. In general, the resultant map produced from the comparison analysis will only be as accurate as the least

accurate input map (Hodgson *et al.*,1988). Another limitation of a post classification comparison is that it can only compare maps that contain boolean categories. Since it can not handle the existence of mixed pixels, the operation of a pairwise comparison often requires complex spatial information to be disregarded or simplified.

The development of a system that can accurately and realistically model land use change dynamics requires it to handle the complex nature of spatial data and to be less sensitive to the map registration problem. First of all, the association procedure should be performed on a regional basis to obtain a more reliable measurement of change between the maps. Therefore, it should involve land use polygons instead of individual pixels. More importantly, the model should be able to handle data uncertainty rather than simply ignore it.

## **1.2 Statement of the Objectives**

The main objective of this research is to develop a fuzzy relational land use change detection model that can produce a qualitative and quantitative description of land use dynamics on a regional scale.

The first step in achieving this objective is to develop a methodology for comparing land use on a polygon to polygon basis. This requires the following tasks:

- (i) determine contiguous groupings of identical land use pixels for each map and calculate the area for each polygon
- (ii) perform an unique polygons analysis to identify which polygons on the maps are locationally correlated

- (iii) determine if the land use is similar for the polygons on map two that intersect with the groupings on map one. Estimate the boolean areal intersection and complement for each polygon on map one.

The next step is the development of the fuzzy change detection inference system that estimates the possibilities of polygon areal containments between the land use maps. Most applications of fuzzy set theory have been based on linguistic interpretation of fuzzy sets (Fedrizzi and Kacprzyk, 1986). For this project, the inference system will emulate a persons' reasoning ability to visually compare maps and intuitively estimate their global and local land use agreements. The creation of the inference system will involve:

- (i) determining fuzzy membership functions to represent the qualitative areal containment interpretations
- (ii) producing fuzzy areal matching rules based on the Max-Min compositional rule of inference and deriving the appropriate fuzzy aggregation and defuzzification algorithms to drive the model.
- (iii) calculating the local matching values for each polygon on the reference map and appraising the applicability of hierarchical fuzzy areal pattern matching for devising a single quantitative agreement value relating the degree of land use matching between the maps.

Finally, the validity of the fuzzy change detection model needs to be measured to determine the significance of the derived global matching values. This requires the calculation of Shannon entropy to estimate the degree of uncertainty in the global containment estimations. In addition, entropy will determine the amount of information lost by transforming the complexity of a qualitative agreement fuzzy set to a quantitative matching value.

### **1.3 Thesis Organization**

This thesis is divided into six chapters. The next chapter presents a review of

- (i) traditional map similarity procedures in geography
- (ii) the fundamentals of fuzzy set theory
- (iii) previous uses of fuzzy set theory in spatial analytical applications, and
- (iv) attempts to improve land use change detection with fuzzy set theory.

Chapter three describes the methodology behind the grouping process and the unique polygons analysis, and gives a detailed description of the development and structure of the fuzzy inference system. In addition, the algorithms for the global matching and entropy analysis will be presented and discussed.

Chapter four explains how the fuzzy areal matching process works by discussing the results for three land use datasets. Maps and tables are included to aid in the description of the land use difference analysis and to show the advantages of hierarchical fuzzy pattern matching over traditional map comparison techniques.

The final chapter contains a discussion of the matching results and an analysis of future research for the expansion of the fuzzy pattern matching process.



## Chapter 2

### Related Research

#### 2.1 Introduction

The existence and importance of uncertainty in many real world problems has been recognized by scientists for some time. Traditionally, the analysis of complex systems was based on precise conventional mathematical operations that often generated unsatisfactory results. A breakthrough occurred in 1965 when Lotfi Zadeh proposed his simple but effective theory of fuzzy sets.

Zadeh (1965) devised his theory of vagueness with the formulation of the Principle of Incompatibility, which states that as the complexity of a problem increases, the ability to analyze it in precise terms decreases. It is generally believed that any model that is able to handle the ambiguity of the real world should emulate the human reasoning method of using qualitative linguistic expressions to describe complex situations and to aggregate information. By representing linguistic expressions as membership functions, fuzzy set theory controls imprecision by grouping elements into classes that do not have sharply defined boundaries (Kandel, 1982).

As an imprecise science, geography is concerned with the modeling of phenomena that are inherently complex and that rarely fit within the confines of standard linear models. Fuzzy set theory is ideally suited for geographical analysis because its inexact structures contain reasoning mechanisms that can preserve the heterogeneous nature of spatial data. More importantly, it allows the geographer to include his qualitative interpretations of a subject into the analysis procedure. Openshaw (1987) and Goodchild

(1988) have pointed out that the descriptive approach of fuzzy set theory may provide a foundation for the development of future spatial analysis procedures that are applicable across all the subdisciplines of geography.

The purpose of this chapter is to describe the traditional post classification map comparison procedure and to review existing fuzzy geographical research. An introduction to the fundamentals of fuzzy sets will be provided before the fuzzy spatial techniques are summarized. The research discussed relates to fuzzy applications in GIS and remote sensing and is divided into fuzzy spatial similarity analysis and fuzzy land use change detection.

## **2.2 Traditional Post Classification Comparison of Thematic Maps**

Singh (1989) states that the pairwise comparison of maps is the preferred and most obvious method for detecting changes or differences between thematic maps. Consequently, it has received considerable evaluation in GIS and remote sensing literature. Jensen *et al.* (1987) assessed wetland changes in South Carolina from change maps and statistics produced by a post classification comparison. Wickware and Howarth (1981) produced supervised classifications of 1973 and 1976 Landsat MSS data and compared the classifications to detect ecological habitat changes in the Peace-Athabasca delta.

The objective of a post classification comparison is to identify areas of categorical disagreement between two maps by determining the pixels with a difference in theme. This entails overlaying or crosstabulating the maps on a pixel-by-pixel basis to produce a map and matrix of site specific differences. The map displays the location and nature of the

differences between the maps while the matrix indicates the total amount of categorical agreements and disagreements for each category. With the matrix representation, it is possible to interpret what the differences are for a certain class from one map to the next. To illustrate, refer to the example post classification matrix in Table 2.2.1. The pixels that are common to both maps are located along the main diagonal of the matrix and represent areas of no change. All nondiagonal values in the matrix represent land use disagreements and can be used to determine the extent of the categorical differences between the maps. For example, seven pixels that were spruce on map one were classified as mixed forest on map two.

Table 2.2.1 Example Post Classification Agreement Matrix

|         |              | Map Two |       |        |      |              |       |
|---------|--------------|---------|-------|--------|------|--------------|-------|
| Map One |              | Water   | Grass | Spruce | Rock | Mixed Forest | Total |
|         | Water        | 11      | 0     | 0      | 0    | 0            | 11    |
|         | Grass        | 0       | 10    | 1      | 0    | 1            | 12    |
|         | Spruce       | 0       | 2     | 6      | 0    | 7            | 15    |
|         | Rock         | 0       | 0     | 0      | 3    | 0            | 3     |
|         | Mixed Forest | 0       | 1     | 4      | 0    | 8            | 13    |
|         | Total        | 11      | 13    | 11     | 3    | 16           | 54    |

From the information in the matrix, summary agreement statistics can be generated to give a measure of areal disagreement. The simplest is the total percentage agreement, which is computed by dividing the sum of the diagonal values by the total number of pixels. Note that Table 2.2.1 has a total agreement value of 84%. Similarly, a more detailed interpretation of map agreement can be obtained from statistics that use all of the entries in the agreement matrix. These measures, such as the Kappa coefficient of

agreement, are used to derive a single quantitative value of global map agreement (Foody,1992; Congalton,1991; Rosenfield and Fitzpatrick-Lins,1986). A thorough review of Kappa and several other global agreement coefficients will be presented in section 4.4 of this thesis.

The need for a better post classification change detection or map similarity procedure relates to the limitations of a pixel-by-pixel comparison. First of all, the procedure is sensitive to image misregistration and the existence of mixed pixels. A pixel-by-pixel comparison of multitemporal maps will detect any misalignment of one or both of the maps as change. In addition, any misclassification of a pixel on either one or both of the maps will be detected as a difference in theme although the disagreement is a result of the inherent errors in the dataset. Secondly, the comparison techniques will often produce results that are significantly different from the actual land use. This is due to their inability to account for the inaccuracies in the maps throughout the comparison operation. Spatial data sets contain errors of various types, and these thematic attribute errors propagate and further reduce the quality of the final product. For example, if the attribute accuracies of two independent maps are 80 % and 75 % respectively, then the accuracy of the comparison map will be only 60 % ( $.80 \times .75 \times 100 \% = 60 \%$ ).

The flexibility offered by a fuzzy representation of spatial data could possibly avoid the problems of the traditional comparison procedure. Therefore, the degrees and types of categorical differences between maps should be determined by a fuzzy post classification comparison. Before the structure of a fuzzy map comparison model can be understood, the fundamentals of fuzzy set theory must be explained.

### 2.3 The Fundamentals of Fuzzy Set Theory

Fuzzy set theory is based on the idea of imprecise membership functions, and was developed as a way to deal with the limitations of traditional data classifiers (Klir,1988). The rigid models consisting of discrete, sharply defined, homogeneous classes ignore the heterogeneous aspects of nature such as measurement error, spatial variability, and complexity (Burrough,1989). Thus, a considerable amount of information is lost when sharp edged entities are combined. Fuzzy set theory provides more appropriate classifiers, because it models cases whose attributes have soft transitional rather than hard boundaries. Therefore, it can deal with the ambiguity and vagueness in mathematical or conceptual models of real phenomena (Burrough,1989).

Zimmerman (1985) explains that fuzzy set theory distinguishes three kinds of inexactness: (1) generality (one concept applies to many situations), (2) ambiguity (a single concept has more than one distinguishable sub-concept), and (3) vagueness (sharp boundaries are not defined). With the possibility of partial memberships, fuzzy sets are generalizations of hard sets to situations where the class boundaries are soft (Kandel,1982).

Mathematically, a fuzzy set A in x is described as a set of pairs

$$A = \{ X, u_a(x) \} \quad x \in X, \quad (2.1)$$

where  $u_a(x)$  is the membership grade of x in A and  $x \in X$  means that x is found in the universe of discourse X. The membership value  $u_a(x)$  ranges from zero to one, where one

represents a full membership in a set and zero nonmembership. In other words, the transition between full membership and no membership is gradual rather than abrupt.

Fuzzy sets are often incorrectly assumed to indicate some form of probability. Despite the fact that they can take on similar values, it is important to realize that membership grades are not probabilities but are a set of possibility values (Zadeh,1987). One significant difference is that the summation of probabilities must equal one, but this is not a requirement for membership grades. Also, an event that is assigned a probability of occurrence is considered to always have full membership in a set (Dubois and Prade,1979).

Although a fuzzy set has inexact boundaries the imprecision can be precisely represented by a membership function. For example, a fuzzy set "adolescent age" may be represented by  $u_a(x)$  as shown in Figure 2.3.1. Thus, the age of 15 fully belongs to A, but the grade of membership decreases as the age differs from 15.

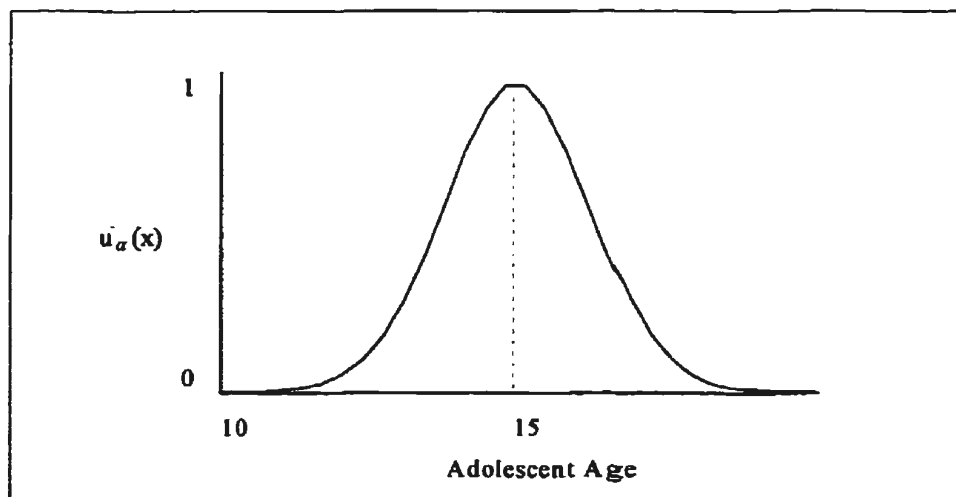


Figure 2.3.1 Membership of X in the fuzzy set Adolescent Age

The selection of the appropriate membership for a fuzzy set is the subjective opinion of the researcher (Zimmerman,1985). However, the structure of the membership function will determine the extent to which the memberships change away from the optimal value (Macmillian.,1978).

### 2.3.1 Basic Definitions of Fuzzy Sets Properties

In this section a number of properties of fuzzy sets are given. The aim is to present the terms and notations necessary to understand how fuzzy sets are used to analyze spatial information. The following definitions are summarized from the work presented by Klir (1988).

- 1) Fuzzy *equality* states that two fuzzy sets  $A, B \in X$  are equal, written  $A = B$ , iff

$$u_a(x) = u_b(x) \text{ for each } x \in X. \quad (2.2)$$

- 2) The *containment* of fuzzy sets is the degree of agreement between their memberships.

A fuzzy set  $A \in X$  is contained in a fuzzy set  $B \in X$ , written  $A \subset B$ , iff

$$u_a(x) \leq u_b(x) \text{ for each } x \in X. \quad (2.3)$$

- 3) The *support* of a fuzzy set  $A$  in  $X$  is the crisp set that contains all the elements that have memberships greater than zero:

$$\text{supp}(A) = \{x \in X : u_a(x) > 0\} \quad (2.4)$$

- 4) The *height* of a fuzzy set is the largest membership grade attained by any element in that set, and a fuzzy set is normalized if at least one element has a membership value of 1.



5) An  $\alpha$ -cut of a fuzzy set A is the set of elements in X that have a membership grade greater than or equal to a specified value of  $\alpha$ . This definition is written as :

$$A_\alpha = \{x \in X: u_a(x) \geq \alpha\}. \quad (2.5)$$

6) The *cardinality* of a fuzzy set A is the summation of the membership grades of all elements in the fuzzy set. Thus

$$\text{card} = |A| = \sum_{x \in X} u_a(x) \quad (2.6)$$

### 2.3.2 Operations on Fuzzy Sets

The concept of a fuzzy set makes it possible to develop the framework of fuzzy set theory from a basic principle that allows the generalization of crisp operations to fuzzy sets known as the Extension Principle (Klir,1988). It provides a means for any function derived for crisp sets to be transformed so that it can also map fuzzy subsets (Zadeh,1987). This makes it possible to extend known algorithms in standard set theory to fuzzy variables (Fedrizzi and Kacprzyk,1986).

As in conventional set theory, the basic operations in fuzzy set theory are also the union, intersection, and complement. The definitions of these fuzzy operators are as follows:

The *union* of A, B  $\in$  X, written  $A \vee B$ , is defined as

$$u_{a \vee b}(x) = \text{Max}(u_a(x), u_b(x)) \quad (2.7)$$

The *intersection* of A, B  $\in$  X, written  $A \wedge B$ , is defined as

$$u_{a \wedge b}(x) = \text{Min} (u_a(x), u_b(x)) \quad (2.8)$$

The *complement* of a fuzzy set  $A \in X$ , written  $A'$ , is defined as

$$u_{a'}(x) = 1 - u_a \text{ for each } x \in X \quad (2.9)$$

Linguistically, the complement corresponds to the negation "not", the union to the logical connective "or", and the intersection to the "and" operator (Fedrizzi and Kacprzyk, 1986).

The use of fuzzy set theory in many fields of science entails the measurement of similarities or dependence between fuzzy sets. For example, several classification and clustering methods attempt to identify and match fuzzy patterns to reference data (Macmillian, 1978). As in conventional mathematics, the basic element of a system of similarity analysis is a fuzzy relation. According to Bandemer (1987), a fuzzy relation is a fuzzy subset of the Cartesian product between two sets A and B. Thus, for the product space  $X_1 \times X_2 = \{x_1, x_2\}$ ,  $x_1 \in X_1$ ,  $x_2 \in X_2$ , a fuzzy relation R in the product space is a set of ordered pairs:

$$R = \{[(x_1, x_2), u_r(x_1, x_2)]\}, x_1 \in X_1, x_2 \in X_2 \quad (2.10)$$

The form of a fuzzy relational equation depends on how the elements of fuzzy sets are structured and what is the best way of measuring their association (Buckles and Petry, 1982). From the Extension Principle, fuzzy arithmetic operators (addition, subtraction, multiplication, and division) combine the fuzzy relations into a single product space (Zimmerman, 1985). The membership functions of each fuzzy set need to be

aggregated to derive an equivalence membership function of the elements between the sets (Zadeh,1987). Pedrycz (1992) has shown that the Max-Min composition rule of fuzzy sets is the preferred method of determining the equivalency relationship function. Formally, the Max-Min possibility measure is

$$u_{A \circ B} = \text{Max} ( u_a (x) \wedge u_b (x) ). \quad (2.11)$$

Through the composition of the fuzzy functions, a link is established between the fuzzy sets that is the maximum association level between them.

## 2.4 Spatial Similarity Applications Involving Fuzzy Set Theory

The notion of a fuzzy set and its implications for spatial modeling was first presented in geographical literature with Gale's (1972) paper about uncertainty analysis in Geographical Analysis. The paper was concerned with inexactness in behavioral geography and had a distinctive message:

"... the fact that inexactness exists must be recognized in terms of both its implications for classification and as an exploratory concept. Any less would undoubtedly leave geography without the theoretical foundation it originally sought to provide." pp. 347

Until recently, few geographers had any interest in fuzzy set theory, which was partially due to a lack of understanding of the basic fundamentals of the theory (Macmillian,1978). With the realization that uncertainty modeling should become a crucial component of spatial analysis procedures, fuzzy set theory has gained increasing support, especially with researchers using geographical information systems and remote sensing.

A number of studies (Cannon *et al.*,1986; Wang,1990; Maselli *et al.*,1996) utilizing fuzzy c-means clustering for remote sensing image classification have shown that fuzzy set theory provides a method of dealing with images containing a complex mixture of spatial and spectral information. A drawback of the traditional clustering algorithms is that a quantity of spectral information is lost when determining the "hard" pixel memberships by allocating each object to exactly one cluster. However, elements often can not be adequately assigned strictly to one cluster because they are spectrally situated between clusters (Cannon *et al.*,1986). In these situations, fuzzy c-means clustering was determined to be a more appropriate classifier since it accounted for the class mixtures and intermediate conditions that occur in remote sensing imagery (Wang,1990). The task of fuzzy clustering is to divide a number of objects into  $c$  ( $2 \leq c < N$ ) soft heterogeneous clusters that contains multiple informational memberships (Zimmerman,1985). The algorithm assigns each observation to all of the fuzzy clusters, with the membership values decreasing with increased distance from the cluster center (McBratney and Moore,1985). Furthermore, cases are allowed to belong totally, partially, or not at all to a set as long as the memberships sum to one. For example, a pixel could be assigned a membership combination of 50% deciduous, 30% coniferous, and 20 % grassland by the fuzzy clustering approach but would be classified as 100% deciduous by the standard clustering procedure. Key *et al.* (1989) used a fuzzy c-means algorithm to classify pixels in AVHRR imagery for the segmentation of sea ice, clouds, and ocean in a polar environment. They found that the fuzzy classifier was sensitive to the range of albedos and physical temperatures encountered in this region due to its ability to classify the well defined

classes (ocean) as well as the categories that fell into intermediate spectral space (sea ice and clouds). Also, McBratney and DeGrujter (1992) utilized this procedure to classify soil conditions in a rural section of the Netherlands. The results showed that the fuzzy grouping technique provided a better overall classification of the study region's main pedological structures than a boolean clustering procedure.

Image interpretation operations based on fuzzy rule based systems have produced more accurate results than standard classification approaches (Blonda *et al.*,1991). In the real world, an expert uses qualitative linguistic terms, such as high, medium or low, as rules to describe the spectral characters of a ground cover (Gopal and Woodcock,1994). These expressions are membership functions in the fuzzy knowledge system and are combined using fuzzy relational and compositional equations to generate the rules for the knowledge bases. The output from the system would be the maximum possibility class value for the combined qualitative interpretation of the scene. Fuzzy sets in this context are used to quantify the linguistic values and to obtain a consensus of the expert evaluations (Blonda *et al.*,1991).

Fuzzy set theory has been used in GIS applications, most notably in the development and manipulation of fuzzy relational databases (Burrough,1989; Wang *et al.*,1990; Kollias and Voliotis,1991; Sui,1992) and the analysis of uncertainty propagation in GIS operations (Veregin,1989). Fuzzy relational databases have been incorporated into GIS modeling because they can handle the imprecision in spatial data representation and manipulation. They also allow for the individualization of the data by a user (Sui,1992). Many fuzzy relational databases are driven by Semantic Import modeling, which uses an  $\alpha$

*priori* membership function to assign a membership grade to individual tuples of the database (Robinson,1988). The membership function identifies the extent to which a particular area belongs to a class based on user submitted criteria (Wang *et al.*,1990). The system is called Semantic Import because the queries submitted to the model are the users' qualitative subjective view of the stored data (Robinson,1988). Consequently, it can process queries that express the uncertainty of the linguistic phrases. An example of a subjective query would be "retrieve all forest stands that are situated on relatively poorly drained soils and fairly flat slopes". The fuzzy relational database converts this question into a fuzzy logical IF-THEN rule comprised of fuzzy relational and compositional operators. This query can rarely be answered by a conventional relational model because it requires queries to be expressed in exact values. Therefore, it is obvious that information is lost when spatial data is retrieved or combined using simple boolean algebra.

GIS functions utilizing fuzzy set theory have been able to manipulate uncertainty as data are transformed during an analysis procedure (Gopal and Woodcock,1994). Spatial data contains a variety of different errors (locational, attributional, etc.) that propagate and contaminate clean datasets and reduce the quality of the final map (Openshaw,1987). For example, a map overlay of two thematic layers results in the creation of a composite map that contains the errors from both input datasets. Veregin (1989) has stated that it is possible to use fuzzy set theory to compute the uncertainty in the output map if the membership values of sites for each category are known for the input maps. Instead of using a traditional boolean overlay to combine themes, a fuzzy relational

equation (fuzzy union or intersection) could aggregate the spatial information while handling the data ambiguity.

#### **2.4.1 Fuzzy Set Theory and Land Use Change Detection**

A review of the literature involving geographical applications based on fuzzy set theory shows that little research has been done on developing an actual fuzzy land use change model. Several authors (Gong,1993; Gopal and Woodcock,1994) have expressed the need for research involving fuzzy sets for map comparison. Also, Edwards and Lowell (1996) suggest that fuzzy set theory should be used to develop a single measure of map accuracy, such as a fuzzy Kappa statistic. However, a number of spatial projects built on fuzzy logic could provide a basis for the creation of an appropriate fuzzy change detection system.

Gong (1993) developed a change detection procedure utilizing satellite imagery that founded its comparison operations on fuzzy set theory. Initially, a principal components analysis was applied to six band-pair difference images (Landsat TM bands one to five and seven) in order to contain the change information into the first few principal components. The difference images were created by subtracting the pixels values for each band of an earlier image from their counterparts on a later one. Through the analysis of the component eigenvalues, it was determined that the first two components contained the change information. The first step in the generation of the change images involved empirically defining the fuzzy membership functions of change by analyzing each principal component difference image and its histogram. From the shape of the histograms,



the parameters of the fuzzy membership function of change were derived. Next, each pixel of the difference component images was assigned a membership value based on where its DN value fell on the membership function, thus creating two change images. Finally, a fuzzy union (MAX) operator was used to combine the change images into a single change image and subsequent change information was extracted from it.

Edwards and Lowell (1996) conducted research on modeling uncertainties in photointerpreted boundaries of forest inventory layers using fuzzy buffers. The problem of two interpreters rarely producing the same results from a common set of photos was handled by a fuzzy boundary width estimator based on texture variability and discrimination. Initially, nine interpreters delineated the forest inventory polygons on three artificially created texture layers. The measured boundary widths for the nine interpreters were determined by generating proximity maps, where each pixel measured the distance from the nearest boundary. The derivation of the fuzzy boundary widths was performed on a local basis to account for local texture variability because the interpretation errors differed in location and were context dependent. An average proximity boundary mask image was produced by averaging the nine proximity widths to acquire a mean measure for each boundary. Subsequently, each of the boundary widths was assigned a plus or negative depending on which side of the average width it was on. The random error between the average and interpreted boundaries was calculated by summing the absolute values of the displacement errors, and this became the desired local fuzzy boundary width. Once the fuzzy widths were estimated, a final fuzzy forest inventory map was created by removing the boundaries of polygons that were affected by local contexts (e.g. slivers and

holes). In relation to land use change detection, it was suggested that a fuzzy intersection operator (MIN) could be used as an overlay technique to detect real change between several multitemporal fuzzy forest species maps.

Another approach using fuzzy sets for comparing multiple spatial data sources consisted of developing a fuzzy method of accuracy assessment of thematic maps (Gopal and Woodcock, 1994). Fuzzy set theory was used to derive fuzzy membership functions, relational equations, and tables to measure the nature, frequency, source, and magnitude of errors in thematic maps. The foundation of the model was a five point linguistic membership scale (1:absolutely wrong, 2:understandable but wrong, 3:acceptable, 4:good answer, and 5:absolutely right) that was utilized to choose the most suitable linguistic value to describe the perceived match between each map class and field sample. Each sample site was assigned one of the linguistic evaluations for each of four possible landcovers, thus providing for the possibility of multiple memberships and classification ambiguity.

The frequency of errors was modeled with fuzzy union and fuzzy cut relational equations. The fuzzy union operator:

$$\text{Max}(u, c) = \begin{cases} 1 & \text{if } u_c(x) \geq u_{c'}(x) \text{ or } 0 \text{ otherwise} \end{cases} \quad (2.12)$$

determined the number of samples sites that were assigned the maximum membership for the correct landcover type. A fuzzy  $\alpha$ -cut grade of three identified the sites whose classification assignment was acceptable. In addition, it identified the sources of errors by analyzing whether a site had multiple memberships (two or more memberships  $\geq$  three).

Summation of the number of matches for all classes represented the global accuracy of the error analysis.

Error magnitude was calculated by a fuzzy Difference functional:

$$D(i) = u_x(x) - \text{Max } u_c(x), \quad (2.13)$$

where  $u_x(x)$  is the membership value of a sample and  $\text{Max } u_c(x)$  is the maximum membership for that category. By computing the average of the difference values for each category, the quality of the classification for a landcover type was measured.

As a final step, the nature of the errors was computed with both a Confusion and an Ambiguity functional:

$$\text{Confusion} = \{ C | c \in C \text{ and } u_c(x) > u_x(x)(X); \quad (2.14)$$

$$\text{Ambiguity} = \{ C | c \in C \text{ and } u_c(x) = u_x(x)(X) \quad (2.15)$$

The confusion functional identified the set of categories whose memberships were greater than that allocated to the map class, while ambiguity identified memberships equal to the category membership. This gave an overview of the between-category misclassification error and the land use types that accounted for most of the error. Comparison of the results to those from a traditional accuracy assessment indicated that the fuzzy analysis system was a more appropriate model because the fuzzy multiple memberships prevented the loss of the land use compositional information that was found to degrade the standard approach.

## 2.5 Conclusion

From an understanding that data uncertainty and complexity are inherent components of spatial information, geographical applications have begun to incorporate fuzzy set theory into their analysis processes. Researchers have realized that the boolean structure of most spatial models fail when ambiguity and vagueness are encountered. Procedures that require the data to be transformed to a normal distribution are in actuality imposing inadequacies upon the representation of geographical objects. Thus, important information is often lost by restricting spatial elements to hard boundaries.

A review of the literature has shown that imprecise results from data retrieval and logical modeling can be addressed with fuzzy set theory. With its capability of multiple and partial memberships and the inclusion of qualitative impressions into an analysis, geographers have little to lose by abandoning nonfuzzy sets in favor of their fuzzy counterparts. Combining fuzzy set theory with spatial modeling increases the latter's ability to deal with uncertainty and also promote further applications of fuzzy sets in spatial decision processes.

## Chapter 3

### Methodology

#### 3.1 Introduction

The basis for the development of a fuzzy inference system to detect land use change is the emulation of human reasoning when comparing multitemporal maps. When asked to relate the level of agreement between maps, a person would use linguistic expressions, such as "very similar" or "significantly different", to describe the perceived similarities between them. Initially, a researcher would notice the overall agreement between the maps but would eventually recognize localized regions of dissimilarities. Consequently, a model that attempts to simulate this visual comparison should be implemented on both a local and global level. For this thesis, the land use map comparison system was modified from the hierarchical fuzzy pattern matching procedure developed by Pedrycz and Roventa (1994). The goal of this thesis is to create a model that can handle the uncertainty of the data and calculate the local and global degrees of containment between two maps.

In this chapter, the methodology utilized in the fuzzy pattern matching for the map comparison analysis is described in three sections. The first explains how the multiple map comparisons are performed on a regional basis with unique polygons mapping. The method used is similar to the procedure presented by Bonham-Carter (1994) and is considered a necessary step to avoid the problems of pixel-to-pixel comparisons. The boolean areal intersections and complements for each polygon on the input land use maps

are calculated from information appended to the attribute table generated during unique polygons mapping. These values are used to derive the local matching values for the polygons in the dataset.

The second section explains the process of developing the fuzzy inference system, from the implementation of the mathematical tool to drive the map comparison model to the selection of an appropriate method of identification of the fuzzy partitioning of the input database (Takagi and Sugeno, 1985). The model was designed using Matlab's Fuzzy Logic Toolbox (1994) software where fuzzy relational equations, compositional rules, and membership functions are selected and assigned to a basic inference shell. Kosko (1992) states that the development of a fuzzy inference system consists of four steps. Step one is the creation of a set of membership functions used to fuzzify the input data. The actual fuzzification of the qualitative interpretation into a membership value of a linguistic variable is the second step. This process depends on the relational structure between the membership functions. Step three is the development of a rule base to interpret the linguistic input in order to generate a qualitative output value. Lastly, this qualitative output value is defuzzified by converting it to a crisp value based on a set of output membership functions.

The final section of the methodology involves the computation of a fuzzy global similarity statistic through the aggregation of the local agreement values. This overall agreement number is an indicator of the general reliability of the derived land use difference map. In addition, an assessment of the uncertainties in the differences maps is performed by computing the Shannon entropy of the global matching. Entropy relates the

degree of uncertainty of the polygon areal pattern matches and determines the reliability of the model's land use disagreement estimations for the study areas. The outcome of a map comparison session is a global agreement value with a specific degree of certainty. For example, the model could predict with 95% certainty that the overall areal agreement between two maps is 90%.

Figure 3.1.1 is a flow chart outlining the steps involved in the local and global matching procedures in the hierarchical fuzzy pattern matching model.

### 3.2 Unique Polygons Mapping

The need for a regional similarity pattern analysis technique for multiple maps stems from the limitations of pixel pairwise comparisons. A traditional pixel-by-pixel overlay procedure fails because it can not preserve the hidden structure of the input maps. The resultant data does not identify the perceptual similarities between the separate land use polygons (Allison,1992). For example, Figure 3.2.1 consists of four polygons with two different land uses.

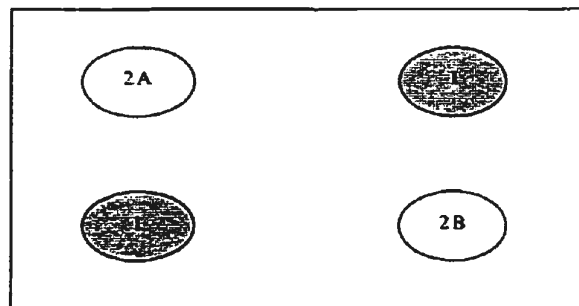


Figure 3.2.1: Land Use Similarity Polygons

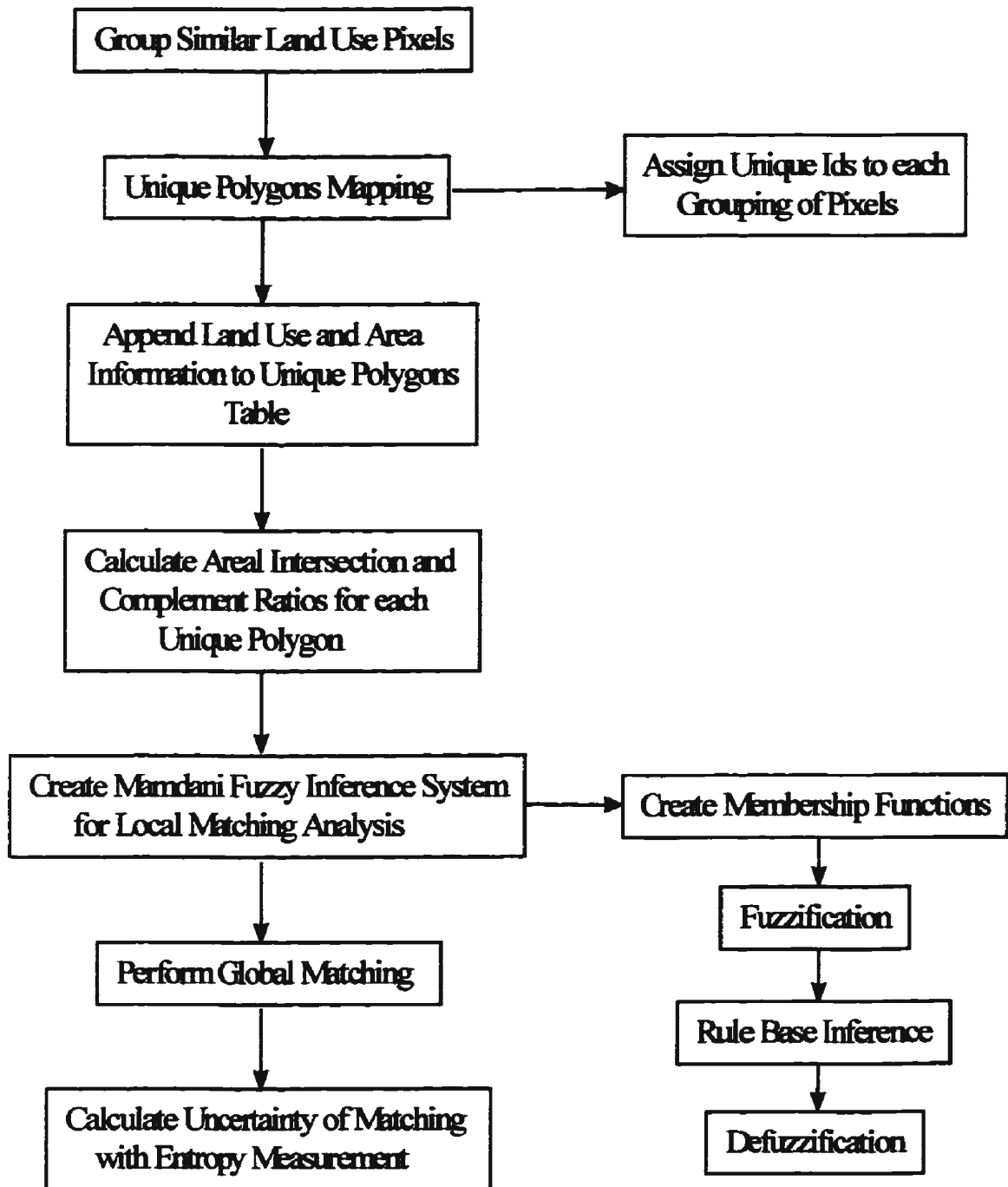


Figure 3.1.1 Flow Chart of the Methodology for Hierarchical Fuzzy Pattern Matching



A conventional measurement of map agreement would combine the pixels of polygons 2A and 2B to determine a single similarity value for land use two. The aggregation of pixels for similar land use polygons prevents the calculation of an agreement value for a single polygon. This prohibits the identification of polygons of high dissimilarity due to actual land use differences or data uncertainty. To characterize the data most of the current map comparison procedures employed in GIS and remote sensing use global measurements, such as the Kappa statistic or Cramer's V (Edwards and Lowell, 1996). Realistically, land use change tends to occur at specific locations because of certain political, economic, and physical factors. Thus, a map comparison model should account for the local variabilities in the data to obtain a more realistic understanding of land use dynamics. A small number of polygons may be responsible for most of the disagreement and their identification and isolation could influence the determination of a global value of similarity between the maps.

The identification of the regional categorical differences between two maps is based on a polygon-by-polygon overlay operation. To avoid the pixel aggregation problem of the standard post classification comparison, it is necessary to delineate the individual polygons of the input maps and to assign them unique identifiers. More importantly, the conversion of the land use maps into unique polygons maps enables the comparison process to be performed on a local level.

Creation of the unique polygons maps was done using the Group module in the GIS software Idrisi (1992). The grouping algorithm determines the contiguous groupings of identically valued pixels in a raster map and assigns them unique integer identifiers. The

derived groups are comprised of pixels that have the same attribute value and contact each other in any of eight possible directions: N, S, E, W, NE, NW, SE, or SW (Eastman,1992).

Unique polygons mapping is performed by overlaying the first grouping map with the second, which creates an overlay image and a relational attribute table. The overlay process generates a series of relational polygons from the intersection of both grouping maps (Bonham-Carter,1994). A unique polygons map is illustrated in Figure 3.2.2, which shows the overlay of map one and map two producing the unique polygons map and attribute table. Each polygon on the map is assigned a unique identifier so that the table has the same number of rows as there are polygons from the overlay process. The attribute table in Figure 3.2.2 contains six rows since the overlay identified six unique class relationships between the maps. For example, the attribute table indicates that unique polygon two was created from the intersection of polygon one on map one and polygon two on map two.

The unavailability of a GIS software to perform unique polygons mapping necessitated the writing of a program, Uniqpoly.bas (see Appendix A), to do the overlay and produce the corresponding unique polygons attribute table. A second program, Uniqmap.bas (Appendix A), was used to produce the unique polygons map from the information in the attribute table by assigning the corresponding unique attribute value to each of the identified intersection polygons.

A unique polygons table is ideally suited to model land use change or map similarities for several reasons. First of all, each unique polygon in the table indirectly

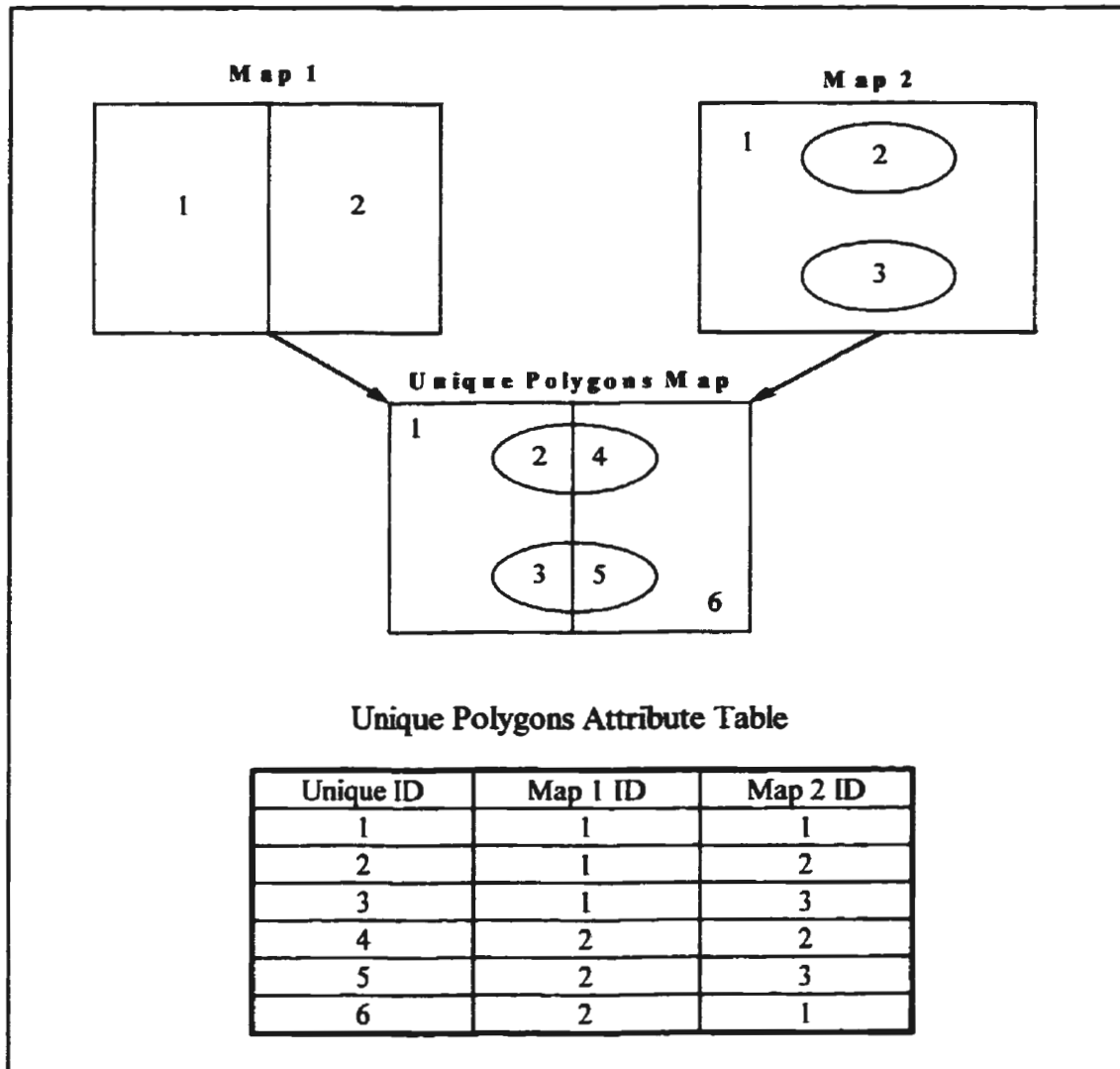


Figure 3.2.2: Structure of a Unique Polygons Map and Attribute Table

represents the degree of containment of the polygons on map one in the polygons on map two. Secondly, auxiliary data that may be needed to monitor categorical differences can be appended to the table as separate columns or fields. It is then a simple matter of reclassifying the unique polygons map with a selected column of the table (Bonham

Carter,1994). Lastly, the table can be imported into other software packages to carry out specialized modeling operations.

The degree of containment for each polygon in the attribute table is utilized to measure the local matching between polygons on the land use maps. Pedrycz and Roventa (1994) point out that the local matching in hierarchical fuzzy pattern matching should involve selected properties of the input fuzzy sets. Therefore, the areal agreement between the polygons is the basis for the local matchings. This requires the areas of the unique polygons and the areas of the intersecting input polygons that make up each unique polygon to be appended to the attribute table.

In the local matching scheme, map one is considered a template or reference map of the land use characteristics of a study area, and map two is a predicted land use layer or an actual land use map at a later date. The attributes and areas of the polygons on map two will be the same as those on the template if no differences in theme are evident. Thus, a unique polygon entry in the attribute table should record identical areas for both land use polygons. The intersection ratio of any two polygons on the maps will be one since map two is perfectly contained within map one. For situations where differences in land use have occurred, the calculated areal intersection ratio will be the local agreements between polygons while the areal complement ratio will represent land use disagreements. It is important to note that the areal intersection and complement ratios are only computed for the unique polygons on map one since it is the template for the matching process.

An initial step in calculating the intersection and complement ratios is to add the land use data for the input polygons to the attribute table. Table 3.2.1 is the updated

attribute table for Figure 3.2.1 and includes the additional area and land use information. The areal intersection ratio is computed by identifying the rows whose land uses are identical for a specific template polygon, summing the unique areas for these rows, and dividing the summed agreement area by the total area for the polygon on the template. For example, refer to the portion of the unique polygons map that represents the agreement of map two with the first template polygon. Rows two and three are regions of land use disagreement or the complement of unique polygon one. The division of the unique area of 3.75 by 5 gives a similarity intersection ratio of 0.75 for the first template, which implies a complement of 0.25. Table 3.2.2 contains the intersections and complements ratios for the polygons on map one.

Table 3.2.1 : Example of Updated Unique Polygons Attribute Table for Figure 3.2.1

| Unique ID | Unique Area | Map 1 ID | Map 1 Area | Map 1 Land Use | Map 2 ID | Map 2 Area | Map 2 Land Use |
|-----------|-------------|----------|------------|----------------|----------|------------|----------------|
| 1         | 3.75        | 1        | 5          | Urban          | 1        | 7.5        | Urban          |
| 2         | .625        | 1        | 5          | Urban          | 2        | 1.25       | Rural          |
| 3         | .625        | 1        | 5          | Urban          | 3        | 1.25       | Rural          |
| 4         | .625        | 2        | 5          | Rural          | 2        | 1.25       | Rural          |
| 5         | .625        | 2        | 5          | Rural          | 3        | 1.25       | Rural          |
| 6         | 3.75        | 2        | 5          | Rural          | 1        | 7.5        | Urban          |

Table 3.2.2: Boolean Local Matching Information for Map One

| Map 1 ID | Intersection | Complement |
|----------|--------------|------------|
| 1        | .75          | .25        |
| 2        | .25          | .75        |

It is at this stage of the procedure that it becomes necessary to account for the uncertainty in the data. The calculated intersections and complements are boolean values

that are computed on the assumption that the unique polygon maps are error free and that real world land use data can be confined to crisp borders. As Edwards and Lowell (1996) have shown, areal spatial information is best represented with boundaries containing fuzzy widths. A more reliable measurement of the local matching between the maps would involve the computation of fuzzy areal intersections and complements. This is accomplished by developing a fuzzy matching system based on a researcher's intuitive understanding of the data and qualitative interpretation of the global and local agreement between the land use maps.

### **3.3 Development of the Fuzzy Inference System for Local Matching**

The purpose of the fuzzy inference system is to describe the regional similarities between land use maps using linguistic terms. This requires a model that can identify a person's intuitive perception of a good match between sections of the maps and convert that expression into a quantitative output value. An analyst will get a feeling for what he believes is the actual local land use agreement of an area from specific criteria (land use type, location, etc) and the visual or quantitative representation of the boolean intersections between polygons. Instead of the local land use analysis depending on a researcher physically interpreting and rating each unique polygon relationship, the fuzzy matching system assigns land use associations to predetermined agreement membership groups.

As with many of the fuzzy inference systems that model imprecise real world processes, the fuzzy inference system for this thesis is based on *Mamdani* inferencing

(Mamdani,1976). Mamdani fuzzy inference systems are rule based decision models that produce mathematical control statements as output membership functions to handle the interactions of the inputs to the system (Jang *et al.*,1997). The design of this system requires the developer to create both input and output membership functions from linguistic interpretations of a subject. Therefore, the designer has to have knowledge of what the output will be for each rule, with the output membership functions being fuzzy sets (Matlab Fuzzy Logic Toolbox Manual,1994). Through fuzzy relational connectives and the compositional rule of inference, Mamdani fuzzy inference systems produce output consequences from the output membership functions (Jang *et al.*,1997).

Figure 3.3.1 is a flowchart of the four basic elements of the Mamdani fuzzy inference system for the matching of the unique polygons. The crisp input values are the calculated areal intersection and complement ratios from the unique polygons mapping.

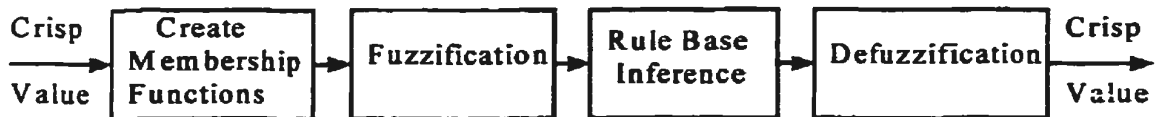


Figure 3.3.1: Four Stages of Designing a Mamdani Fuzzy Inference System

The output local matching values depend on the fuzzy relational and compositional algorithms that comprise and link the sections of the fuzzy inference network.

### 3.3.1 Creation of the Membership Functions

A successful application of fuzzy set theory for map comparisons depends on the appropriateness of the membership functions. Domby (1990) states that both inductive and

deductive approaches have been used to determine membership functions. As a general rule, a suitable membership function is subjectively chosen by the user, who is responsible for the accuracy of the assigned membership grades (Kollias and Voliotis,1991). Several guidelines should be met during the membership function creation process (Burrough,1989). Firstly, the selection of the form and parametric structure of the membership functions must be meaningful in terms of the analysis and should resemble the practical experiences and opinions of the person developing the inference system. Secondly, it is important to choose functions that can define flexible membership grades and that can be easily calculated.

The functionality of the fuzzy inference system requires that membership functions be created for the input information and the predicted output expressions. The initial step in creating the input membership functions is the development of a linguistic scaling of the local matchings for the unique polygons from the boolean areal intersection and complement ratios. Formally, the semantic expressions are needed as answers to the question: " What is the possibility that the land use is similar for a specific localized comparison of unique polygons?". A five point scale is generated ranging from "very low" to "very high". The linguistic values and their descriptions are in Table 3.3.1.1.

Table 3.3.1.1: Input Linguistic Local Matching Interpretations

| Scaling Value | Description  |
|---------------|--|
| Very Low      | definite land use differences; boolean areal intersection is very low          |
| Low           | land use differences very likely; areal intersection is low                    |
| Medium        | possible land use differences; areal intersections and complements are similar |
| High          | land use differences very unlikely; areal agreement is high                    |
| Very High     | land uses are identical; areal agreement close to perfect                      |



From the heterogeneity of the land cover and the believed ambiguity in the areal ratios, the most suitable linguistic label is chosen to describe the nature of the matching between the unique polygons. To transform the crisp intersection and complement numbers into linguistic values, membership functions for each of the qualitative local matching scales are devised. Figure 3.3.1.1 identifies the shape and parameters of the membership functions for the five linguistic scaling expressions for the areal intersection input data. The same membership functions also apply to the areal complements since they are computed from the intersection values. Two distinct types of membership functions are evident: (1) the sigmoidal curve (very low and very high), and (2) the generalized bell curve (low, medium, and high). It should be noted that despite their differences all of the membership functions have the following properties (Dombi,1990): (1) all are continuous, (2) all map an interval  $[a,b]$  to  $[0,1]$ ;  $u_{[a,b]} \rightarrow [0,1]$ , and (3) all are either monotonically increasing, monotonically decreasing, or both.

The sigmoidal curve membership function is one of the most commonly used functions in fuzzy set theory (Eastman,1992). Simpson and Keller (1995) describe a sigmoidal membership function as a asymmetrical curve with regards to its crossover point that assigns values above or below a specific point complete membership or nonmembership. A crossover point of a membership function is a point  $x \in X$  where the membership grade is equal to 0.5 (Jang *et al.*,1997). The formula for a sigmoidal curve is:

$$\text{SIGMF}(X,[A,C] = 1 / (1 + \text{EXP}(-A*(X-C))), \quad (3.1)$$

where  $A$  sets the shape of the curve and  $C$  is the crossover point (Matlab Fuzzy Logic Toolbox Manual, 1994). The parameters for the sigmoidal curves in Figure 3.3.1.1 are found in Table 3.3.1.2. Depending on the sign of  $A$ , a sigmoidal curve is inherently open

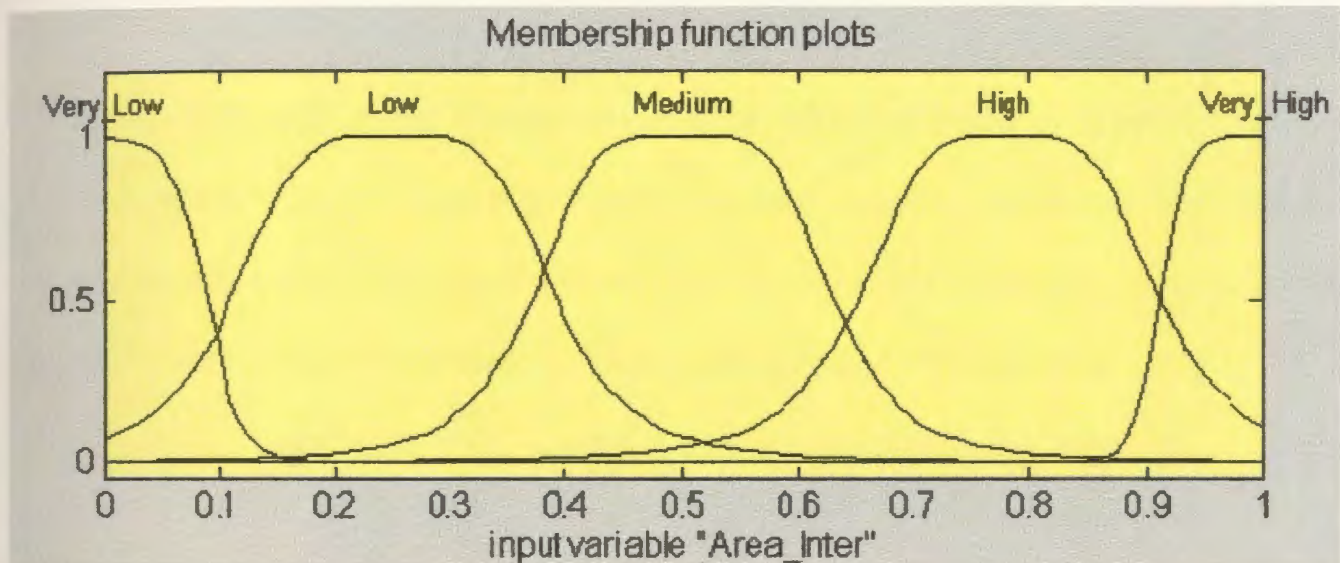


Figure 3.3.1.1 : Membership Functions for Areal Intersection Linguistic Values

left or right so that memberships monotonically rise or fall in one direction towards a plateau of complete membership (Jang *et al.*, 1997).

Table 3.3.1.2: Sigmoidal Membership Function Parameters

| Function Label | $A$    | $C$   |
|----------------|--------|-------|
| Very Low       | -65.41 | 0.091 |
| Very High      | 90.81  | 0.91  |

The asymmetrical open structure of a sigmoidal membership function makes it appropriate for representing concepts such as "very low" or "very high". In terms of localized map comparisons, the sigmoidal curves depict instances where the land use

agreement between maps is known with a high degree of certainty. When an input value falls within the plateau range of either sigmoidal curve, a person intuitively believes that the boolean area measurement represents the actual degree of agreement between the maps.

A generalized bell membership function is a symmetrical closed curve consisting of two transitional membership slopes connected by a total membership plateau. With the presence of two crossover points, membership grades monotonically rise and fall towards the plateau and return again to complete nonmembership. Mathematically, a generalized bell membership curve is specified as a three element vector by the formula:

$$\text{GBELLMF} = (X, [A, B, C] = 1 / ((1 + \text{ABS}(X - C) / A))^{(2 * B)}), \quad (3.2)$$

where A varies the width of the curve at the crossover points, B determines the range of the plateau, and C is the placement of the center of the curve (Matlab Fuzzy Logic Toolbox Manual, 1994). Table 3.3.1.3 contains the parameters for the three input generalized bell functions. It should be mentioned that the shape of the curves and the position of the crossover points can be changed by altering any of the three dispersion values.

The generalized bell functions (low, medium, and high) in Figure 3.3.1.1 represent instances where the user senses that the boolean areal information does not accurately describe the local matching between two maps. A generalized bell function is appropriate for these situations because its two transitional slopes enable it to determine if a boolean area value underestimates or overestimates the actual local agreement. By shifting the

emphasis of gradual membership to the boundaries of the curves, a boolean area value is fuzzified if it falls beyond the lower or upper boundary of the total membership plateau.

Table 3.3.1.3: Generalized Bell Membership Function Parameters

| Function Label | A    | B   | C     |
|----------------|------|-----|-------|
| Low            | 0.14 | 2.2 | 0.252 |
| Medium         | 0.13 | 2.2 | 0.5   |
| High           | 0.13 | 2.2 | 0.78  |

A noticeable characteristic of the membership functions in Figure 3.3.1.1 is that the slopes of the curves overlap. The overlapping of membership functions handles the uncertainty in the linguistic expressions and allows values to have multiple memberships in the function set (Simpson and Keller,1995). For example, it is difficult to define exactly where an analyst's interpretation of low agreement stops and medium begins. The linguistic expressions are intuitive statements that are best represented with fuzzy boundaries. The degree of overlap is subjectively estimated by the user in order to determine the amount of fuzziness desired in the output (Simpson and Keller,1995).

From an analysis of the data and previous research on land use dynamics (White *et al.*,1997), it was thought to be important to include a third set of membership functions to account for the number of pixels comprising the unique polygons. There is a strong possibility that many of the identified polygons from the Grouping procedure will consist of one or two pixels. The problem that arises is whether or not a single pixel disagreement is actually change or a random artifact in the data. The calculation of a global matching value could be adversely affected by assigning the same weight to these small groups and

the larger unique polygons. In addition, the test results provide evidence that the stability and confidence in the derived areal information and local matchings increases as the area for the polygon comparison increases.

Figure 3.3.1.2 displays the pixel group membership functions, both being sigmoidal curves. The input data ranges from one to four since the pixel information is divided into four distinct categories: (1) one pixel, (2) two pixels, (3) three pixels, and (4) > three pixels. Although this discrete input range is different from the scaling for the areal information, the membership grades remain continuous and monotonic.

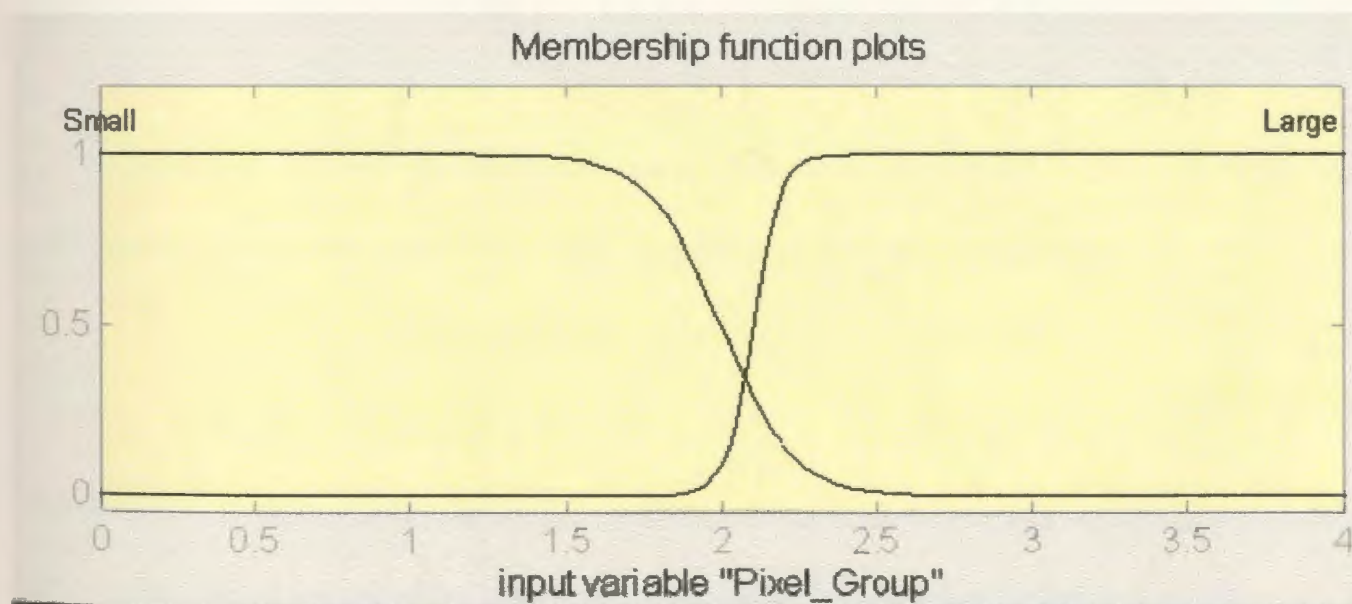


Figure 3.3.1.2 : Membership Functions for Pixel Groupings

The parameters for the sigmoidal curves are found in Table 3.3.1.4. The transition zone between the curves is confined to the 1.75 to 2.25 range since preliminary results indicate that an acceptable degree of certainty that land use disagreement exists requires a minimum grouping of three pixels. Groupings of two pixels could cause

problems because of the possibility that the classification algorithm randomly situated two identically labeled but unrelated cells next to each other.

Table 3.3.1.4: Pixel Group Sigmoidal Function Parameters

| Function Label | A     | C   |
|----------------|-------|-----|
| Small          | -8.85 | 2   |
| Large          | 23    | 2.1 |

The output from the fuzzy inference system is a set of linguistic expressions that describe the local matchings for the unique polygons. As with the input data, the output linguistic statements are based on a five point evaluation scale (Table 3.3.1.5) and require a membership function for each linguistic value. The output membership set (Figure 3.3.1.3) consists of two sigmoidal and three Gaussian membership curves. A Gaussian membership curve is described by a two element vector with the equation:

$$\text{GAUSSMF}(X,[S,C] = \text{EXP}(-(X-C)^2 / (2*S^2)); \quad (3.3)$$

where S is the width of the curve and C is the location of the center of the curve. Table 3.3.1.5 contains the parametric information for the five output membership algorithms

A distinctive feature of the local matching membership set is the amount of overlap between the membership functions produced by the widths of the Gaussian curves. There is no point in the set where the output local matching value can have single membership in a linguistic value. Any derived output value will have multiple membership in the linguistic set, which is necessary to account for any uncertainties in the calculated local matchings.



Table 3.3.1.5: Linguistic Labels and Parameters for the Output Membership Functions

| Linguistic Label | Function Type | A or S | C    |
|------------------|---------------|--------|------|
| Very Poor        | Sigmoidal     | -24.41 | 0.08 |
| Poor             | Gaussian      | 0.1216 | 0.15 |
| Good             | Gaussian      | 0.1623 | 0.49 |
| Very Poor        | Gaussian      | 0.1515 | 0.77 |
| Perfect          | Sigmoidal     | 55.81  | 0.96 |

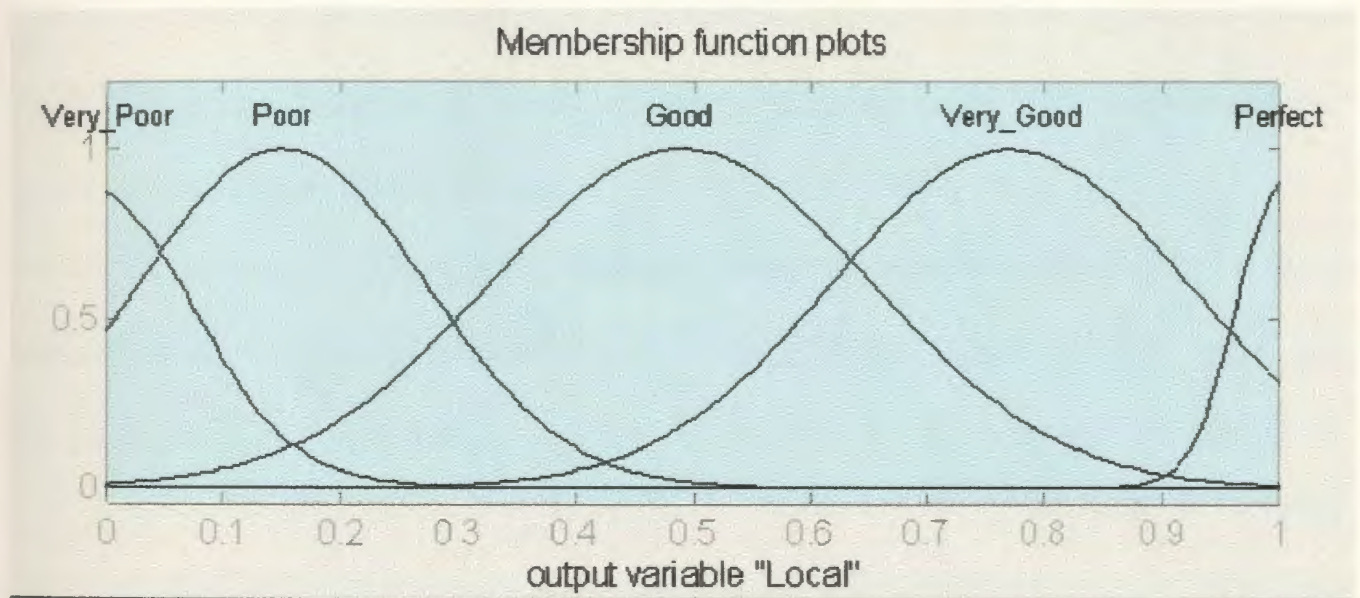


Figure 3.3.1.3 : Local Matching Output Membership Functions

### 3.3.2 Fuzzification

The second step in the development of the fuzzy inference system is the fuzzification of the input data. Fuzzification involves applying the membership functions of the input variables to their actual values to estimate the degree of areal agreement for each unique association. The result from fuzzification is a fuzzy array, a vector of membership strengths of a single crisp input value for each membership function (Simpson and Keller, 1995).

Fuzzification entails locating the crisp input value on the x axis of the membership functions and estimating the corresponding memberships from the y axis. The fuzzy array consists of the memberships for each linguistic expression arranged from left to right. Simpson and Keller (1995) describe a fuzzy vector as:

$$F = [f_1, f_2, f_3, \dots, f_n] \text{ such that } \sum_{i=1}^N f_i \geq 1. \quad (3.4)$$

Since fuzzification produces as many vectors as there are input variables, three fuzzy arrays were generated for the project.

To illustrate, assume that the following input numbers are entered into the fuzzy inference system: (1) areal intersection ratio = 0.4, (2) areal complement ratio = 0.6, and (3) pixel group = 4. Referring to Figures 3.3.1.1 and 3.3.1.3, respectively, the fuzzy vectors for the areal values are  $F_{inter} = [0, 0.44, 0.77, 0, 0]$ ,  $F_{comp} = [0, 0.04, 0.73, 0.23, 0]$ . Also, Figure 3.3.1.2 indicates that the pixel group value results in a fuzzy vector of  $F_{pixel} = [0, 1]$ . Note that the membership strengths do not have to sum to one, which is a requirement for probability values (Klir, 1988).

### 3.3.3 Rule Based Inference

The essential part of a fuzzy inference system is a set of fuzzy rules that are related by means of a fuzzy implication function and a compositional rule of inference (Jang *et al.*, 1997). Fuzzy rules are a collection of linguistic statements that describe how a fuzzy



inference system makes a decision about categorizing an input or controlling an output (Simpson and Keller,1995). The rules are written as If-Then statements in the form:

If (Input 1 is A) and (Input 2 is B) then (Output is C),

where A and B are membership functions for the input values and C is a membership function defined on an output variable. The premise of the rules describes the relationship between the input variables and the firing strength of the rule, while the conclusion determines the characteristics of a membership function for an output variable (Jang *et al.*,1997). The set of rules in the fuzzy inference system is known as a fuzzy rulebase.

With fuzzy rulebase reasoning, the fuzzy rules have to be represented by a fuzzy implication function. Unlike the rules for traditional knowledge based systems, the rules of a fuzzy inference system use linguistic fuzzy sets to quantify the variables in the rules (Pedrycz,1992). The input to a rule consists of values from linguistic membership functions, and the consequent of the rule depends on how the implication process defines the associations between the input membership functions. Thus, the fuzzy implication of a rule depends on its If-Then connective operator, which expresses how a fuzzy rule is delineated by a fuzzy relation (Jang *et al.*,1997). For Mamdani fuzzy inference systems, the premise variables of the rules are connected with a conjunctive T-norm (Dubois and Prade,1979). Jager (1995) describes a T-norm as a fuzzy triangular function that satisfies the condition:

$$I(a,b) = \text{Min}(a,b), \quad (3.5)$$

where Min sets the upper boundary of the function as the intersection of a and b. Formally, a T-norm refers to a logical *And* connective so that fuzzy rules are written as

If A and B then C.

Jager (1995) states that regardless of the implication chosen for the inference system the development of the fuzzy rulebase must ensure that the rules exhibit consistency and completeness. Consistency of a rulebase implies that there are no redundancies in the premises of the rules. The completeness of the rulebase ensures that it does not have blank sections or holes in the knowledge representation. To satisfy these requirements, ten rules (Table 3.3.3.1) were created for the rulebase of the local matching fuzzy inference system. An inference system consisting of two input variables each represented by five membership functions and a third variable with two membership functions would normally require 30 rules for complete coverage of the input knowledge. This was not necessary for this project because the rulebase only needs the rules where the areal intersection and complement ratios are opposites.

Table 3.3.3.1: Rules for the Local Matching Fuzzy Inference System

| Rule # | Rule Structure   |
|--------|--|
| 1      | If(Area_Inter is Very_Low) and (Area_Comp is Very_High) and (Pixel_Group is Small) then (Local is Poor)      |
| 2      | If(Area_Inter is Very_Low) and (Area_Comp is Very_High) and (Pixel_Group is Large) then (Local is Very_Poor) |
| 3      | If(Area_Inter is Low) and (Area_Comp is High) and (Pixel_Group is Small) then (Local is Good)                |
| 4      | If(Area_Inter is Low) and (Area_Comp is High) and (Pixel_Group is Large) then (Local is Poor)                |
| 5      | If(Area_Inter is Medium) and (Area_Comp is Medium) and (Pixel_Group is Small) then (Local is Good)           |
| 6      | If(Area_Inter is Medium) and (Area_Comp is Medium) and (Pixel_Group is Large) then (Local is Good)           |
| 7      | If(Area_Inter is High) and (Area_Comp is Low) and (Pixel_Group is Small) then (Local is Good)                |
| 8      | If(Area_Inter is High) and (Area_Comp is Low) and (Pixel_Group is Large) then (Local is Very_Good)           |
| 9      | If(Area_Inter is Very_High) and (Area_Comp is Very_Low) and (Pixel_Group is Small) then (Local is Perfect)   |
| 10     | If(Area_Inter is Very_High) and (Area_Comp is Very_Low) and (Pixel_Group is Large) then (Local is Perfect)   |

Within Mamdani inferencing, the truth value or strength of the premise of each rule is computed and applied to the conclusion section of each rule (Tanaka *et al.*,1982). This results in a single value being assigned to the output variable of each rule from the intersection of the membership functions for the input variables. The critical step in the implication process is finding the consequence of each rule by combining its strength and output membership function (Jager,1995). The consequence of a rule is computed by clipping an output membership function at the rule strength (Mamdani,1976). Therefore, the output is a membership function with a height equal to the fuzzy support or degree of fulfillment of the premise of a rule. For example, Figure 3.3.3.1 is a graphical representation of the ten rules in the database of the local matching fuzzy inference system. Visual inspection of the output value for each rule shows that the height of the local matching curve is equal to the lowest value of either the areal intersection or areal complement.

An important characteristic of the fuzzy inference engine is that it activates several fuzzy rules during a consultation (see Table 3.3.3.1). Unlike an expert system, the consequence for input data to a fuzzy inference system is rarely obtained from a single rule. Since the purpose of the fuzzy inference system is to map the input variables to an output subset, the consequence of each activated rule needs to be combined into a single output distribution (Jager,1995). Mamdani inference employs the compositional rule of inference for the aggregation of fuzzy rules. More specifically, the inference scheme is applied as (Nguyen and Walker,1997):

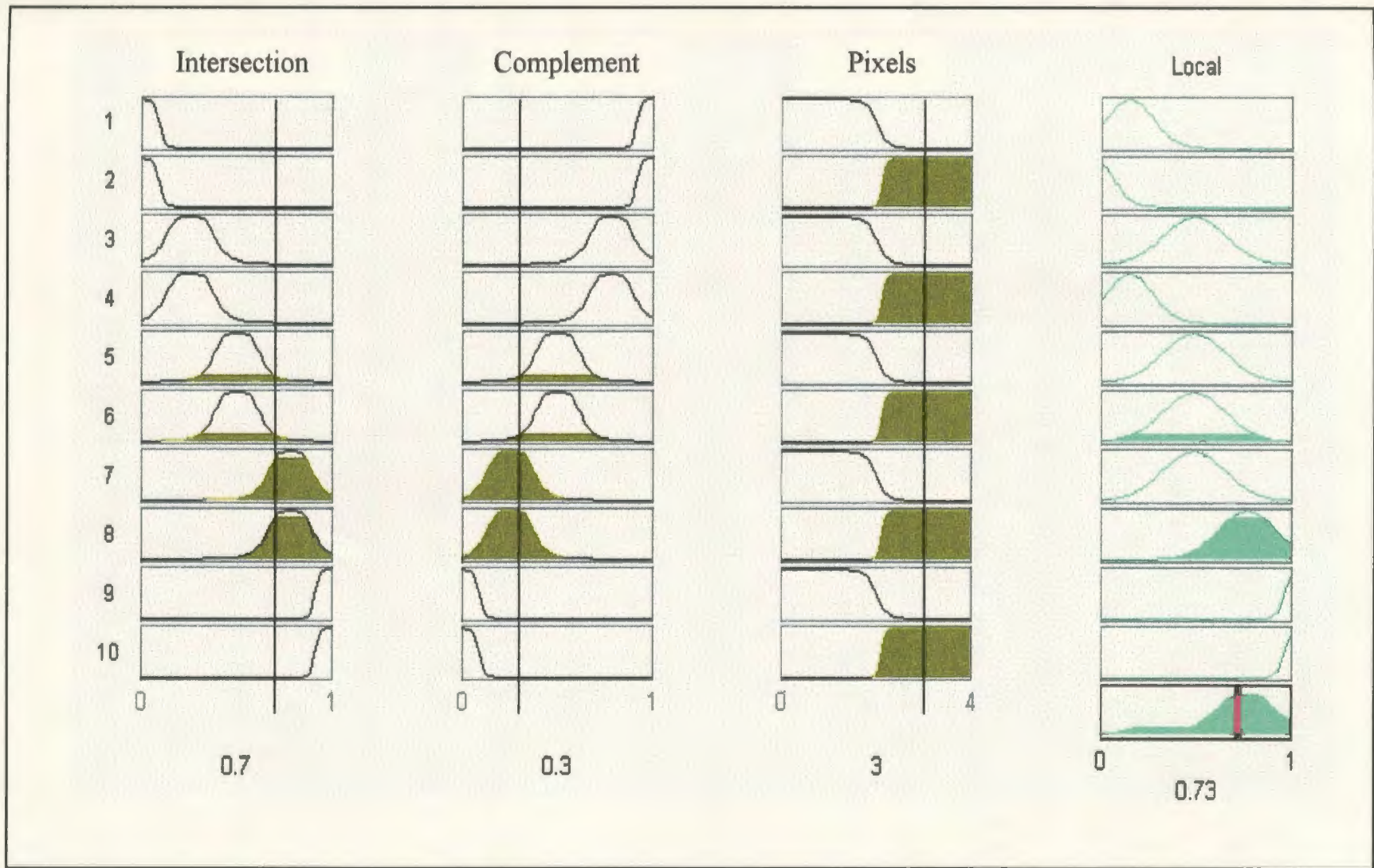


Figure 3.3.3.1: Rulebase and Inference Structure of the Fuzzy Inference System

$$M(x,u) = \text{Max}(A_j(x) \text{ Min } B_j(u)), j = 1,2,...,n, \quad (3.6)$$

where Max and Min are the logical "OR" and "AND" connectives, respectively.

In Max-Min compositional inference, the combined output fuzzy subset is constructed by the union of the fuzzy subsets assigned to a variable by the inference rules. To illustrate, the output distribution in Figure 3.3.3.1 is the union of rules six and eight and represents the maximum relationship between the consequences of the fuzzy rules.

The selection of an appropriate compositional rule of inference is important because it becomes the algorithm that performs the local matchings and produces the output similarity values. With Max-Min composition as the inference rule, the local matching for unique polygon  $X_i$  is expressed as:

$$Lm(X_i) = \text{Max} [\text{Min}(\text{Area\_Inter}(X_i), \text{Area\_Comp}(X_i), \text{Pixel Group}(X_i))] \quad (3.7)$$

### 3.3.4 Defuzzification

To obtain a crisp local matching value, it is necessary to transform the output membership function produced by the inference algorithm into a crisp number. That is, the fuzzy output distribution is defuzzified. Although numerous defuzzification methods have been suggested (Jager,1995; Nguyen and Walker,1997), the centroid of area defuzzification is used to calculate the local matching numbers. Justification of the selection of this defuzzification procedure is supplied by Jager (1995), who states that centroid of area defuzzification is the method of choice when dealing with one dimensional output fuzzy sets. The centroid of area calculates the crisp value of the output variable by

finding the center of gravity value of the aggregated output membership function (Nguyen and Walker,1997). This is computed as follows (Jager,1995):

$$Z(B) = \frac{\int_{\gamma} u_b(y)y \, dy}{\int_{\gamma} u_b(y) \, dy}, \quad (3.8)$$

where  $Z$  is the centroid of area and  $u_b$  is the membership value in the output distribution  $B$ .

For example, return to Figure 3.3.3.1. The centroid of area defuzzification gives a local matching value of 0.73 for this sample unique polygon. The location of the centroid of area of the output distribution is depicted by the red vertical line through the output membership function.

The MATLAB program, Localfis.m, that was used to import the areal data into the model and activate the fuzzy inference system is found in Appendix A.

### 3.3.5 Global Matching

A quantitative measure of the accuracy of thematic classifications or the agreement between maps is typically expressed as a single value. Global statistics, such as the Kappa coefficient of agreement, Cramer's V, and Tau, have been successively utilized to compute the overall agreement between thematic maps and auxiliary classification information (Foody,1992; Ma and Redmond,1995). The importance of a single agreement value for land use change detection and similarity analysis is that it provides the user with a quick interpretation of the land use differences of a study area.

The computation of a fuzzy global similarity number that expresses the overall agreement between two land use maps involves the aggregation of each of the local matchings for the unique polygons. As a first step, the local matching values are appended as a new column to the unique polygons attribute table. The logic behind the aggregation procedure is that a local matching value is a measurement of areal agreement between two land use polygons. By multiplying a local matching number by the area of the unique polygon, an agreement area is calculated. Then, the aggregation of the local matching areas relative to the total area of the unique polygons map produces the global similarity value. This is computed as :

$$g_k = \frac{\sum_{i=1} Lm(x_i) \text{ Area}(x_i)}{\text{Total Area}}, i = 1, 2, \dots, n. \quad (3.9)$$

Several global similarity procedures that have been used in spatial applications have also been used to calculate the similarity between maps for specific land use types. With the land use information in the unique polygons attribute table, the above formula can be modified to compute a land use category global agreement value similar to conditional Kappa.

Hirota and Pedrycz (1991) have stated that an aggregation of fuzzy values will produce an erroneous global matching value because an aggregation formula has no implication support in fuzzy logic. This is not a problem for the above equation because the local matchings are converted from fuzzy to crisp numbers by the defuzzification

algorithm. Since the local matchings numbers and areas are crisp values, the aggregation process of the global matching does not violate any rules of fuzzy set theory.

### **3.3.6 Estimation of Uncertainty in the Global Matching**

The validity of the fuzzy map comparison model depends on the reliability of the derived global matching value. A researcher must have confidence in the matching results before they are used for decision making. Deluca and Termini (1979) describe how a probabilistic measure of certainty is needed for information obtained by global means. In an effort to fit the complexity of a fuzzy set into a global classification, information is lost to transform fuzzy sets into other fuzzy sets. Therefore, the validity of a global similarity value depends on measuring the amount of information lost in the transformation.

Accordingly, Hirota and Pedrycz (1991) relate how Shannon entropy can be used to evaluate the certainty of the matching of two fuzzy sets. Zadeh (1969) describes Shannon entropy as the average amount of information needed to have no uncertainty that a fuzzy event will occur. Referring to pattern matching, entropy would describe the uncertainty that map A matches map B. The higher the value of entropy, the more uncertain the result of matching. By introducing a certainty measurement into the results from the fuzzy inference system, the matchings are expressed as "A matches B to degree C".

The uncertainty of global matching between maps is modified from the shannon entropy measurements proposed by Hirota and Pedrycz (1991). Let the individual local matching numbers be expressed as  $a_k$ , then the probability of map A matching map B is :



$$p_k = (a_k) n_{ak} / \sum_{n=1}^N Lm_n \quad (3.10)$$

where  $n_{ak}$  is the number of template polygons with the similarity value  $a_k$  and  $\sum_{n=1}^N Lm_n$  is the summation of all local matching values for a template layer. Then the Shannon entropy is calculated as:

$$H = - \sum_{k=1}^K (a_k p_k) \log(a_k p_k) + ((1 - a_k) p_k) \log((1 - a_k) p_k). \quad (3.11)$$

Before the uncertainty measurements are computed, it is necessary to normalize the entropy values since they are not restricted to the same [0,1] interval as the global matchings. Normalization of the entropy numbers requires the calculation of the maximum entropy for each matching value with:

$$H_{\max} = \log \sum_{k=1}^K 2^{h(a_k)}. \quad (3.12)$$

From the normalized entropy values, the grade of uncertainty for a global match is computed as

$$\varepsilon = H / H_{\max} \quad (3.13)$$

and the certainty,  $C$ , of matching is equal to

$$C = 1 - \epsilon . \quad (3.14)$$

### 3.4 Conclusion

It was shown how a fuzzy inference system can detect land use differences with hierarchical fuzzy pattern matching. By emulating human reasoning when comparing multiple maps, fuzzy relational modeling utilizing linguistic membership functions can be used to derive local and global matching values. Qualitative interpretations of the areal agreement between two land use maps from unique polygons mapping are the basis for determining the regionalized and overall characteristics of thematic differences between the maps. The advantage of a fuzzy land use map comparison model over the traditional techniques is the ability to handle uncertainties in the dataset by allowing the polygons of the land use maps to have fuzzy boundaries.

## Chapter 4

### Results

#### 4.1 Introduction

The focus of this chapter is the discussion of the land use change and map comparison results from the fuzzy inference system for three experimental datasets. Two of the datasets consist of land use prediction maps from cellular automata, and the third is comprised of multitemporal forest inventory maps. The fuzzy pattern matching results for the datasets are presented in four sections. Since fuzzy local matching emulates human reasoning when comparing multiple land use maps, the first section describes the visual interpretation of the maps and the derivation of a linguistic agreement statement. The qualitative expressions of map similarity should support and approximate the fuzzy global matching values for each dataset.

The purpose of fuzzy pattern matching is to estimate the differences between land use maps in terms of local polygonal matchings. Thus, the second section analyzes the local matches and mismatches to determine the degree and nature of the land use agreement between the maps of the datasets.

From the methodology given in the previous chapter, the calculation of the global matching values and associated entropy measurements for the datasets are discussed in section three. In addition, the utility of the fuzzy global matching values and the ability of the fuzzy matching technique to perform map comparisons are determined from a comparison with traditional two map global similarity statistics.

The last section discusses the advantages of fuzzy land use pattern matching over traditional boolean comparison techniques for both map similarity analysis and land use change detection.

#### **4.1.1 Description of the Datasets**

The data sources for this thesis consist of two sets of atemporal urban land use maps and a set of multitemporal forest inventory maps. The fuzzy inference system compares atemporal maps for a map similarity analysis and multitemporal maps for land use change detection.

Dataset one, which will be referred to as Cinc1, is comprised of simulated and actual land use maps of Cincinnati, Ohio. The actual land use map is derived from the work of Passonneau and Wurman (1966) on land use information for cities in the United States and is the template or reference layer in the matching process. The simulated map was produced by an urban dynamics cellular automata based model developed by White *et al.* (1997). From a set of quasi-deterministic transition rules, the simulated map was generated by ten iterations of the cellular model, with an antecedent land use map as the initial configuration. Both maps are 80 rows by 80 columns rasters at a pixel resolution of 250 meters.

Dataset two, called Cinc2, contains the same maps as Cinc1 except that the simulated map is transposed 90° counterclockwise. The purpose of the transposition was to evaluate the performance of the matching procedure when comparing maps that are visually dissimilar.

The two land use maps in dataset three, called Forest 1, are classified Landsat TM images that were acquired on July 29, 1985 and August 3, 1991, respectively. The images were georegistered with less than 0.5 pixel RMS to the UTM grid on NTS map sheet 12H/04 producing a pixel resolution of 30 meters. A maximum likelihood algorithm was utilized to classify the images into forest inventory types based on field information. However, the forest inventory maps used in this thesis are subscenes containing 334 rows by 222 columns that were extracted from the original imagery and are centered on a region to the Northwest of Pasadena, Newfoundland. For the matching process, the 1985 map is the template, and the 1991 map is the matching layer.

#### **4.2 Linguistic Evaluations of Map Similarities**

Depending on the complexity of the datasets, human reasoning enables a researcher to get an immediate impression or feeling for the degree of similarity between two land use maps. Consequently, human intuition plays an important role in determining land use agreement when maps are visually compared. The basis of the local matching from the fuzzy inference system is to reproduce and utilize this qualitative intuition to obtain an overall agreement statement for the maps. This enables the performance of the model to be determined from how well the linguistic agreement statement approximates and supports the global matching value for a dataset.

Figure 4.2.1 represents the land use maps for Cinc1, where the left map is the template layer. The river, railway, and roads areas on both maps are identical, but several small sections of residential and commercial land use disagreement are noticeable at the

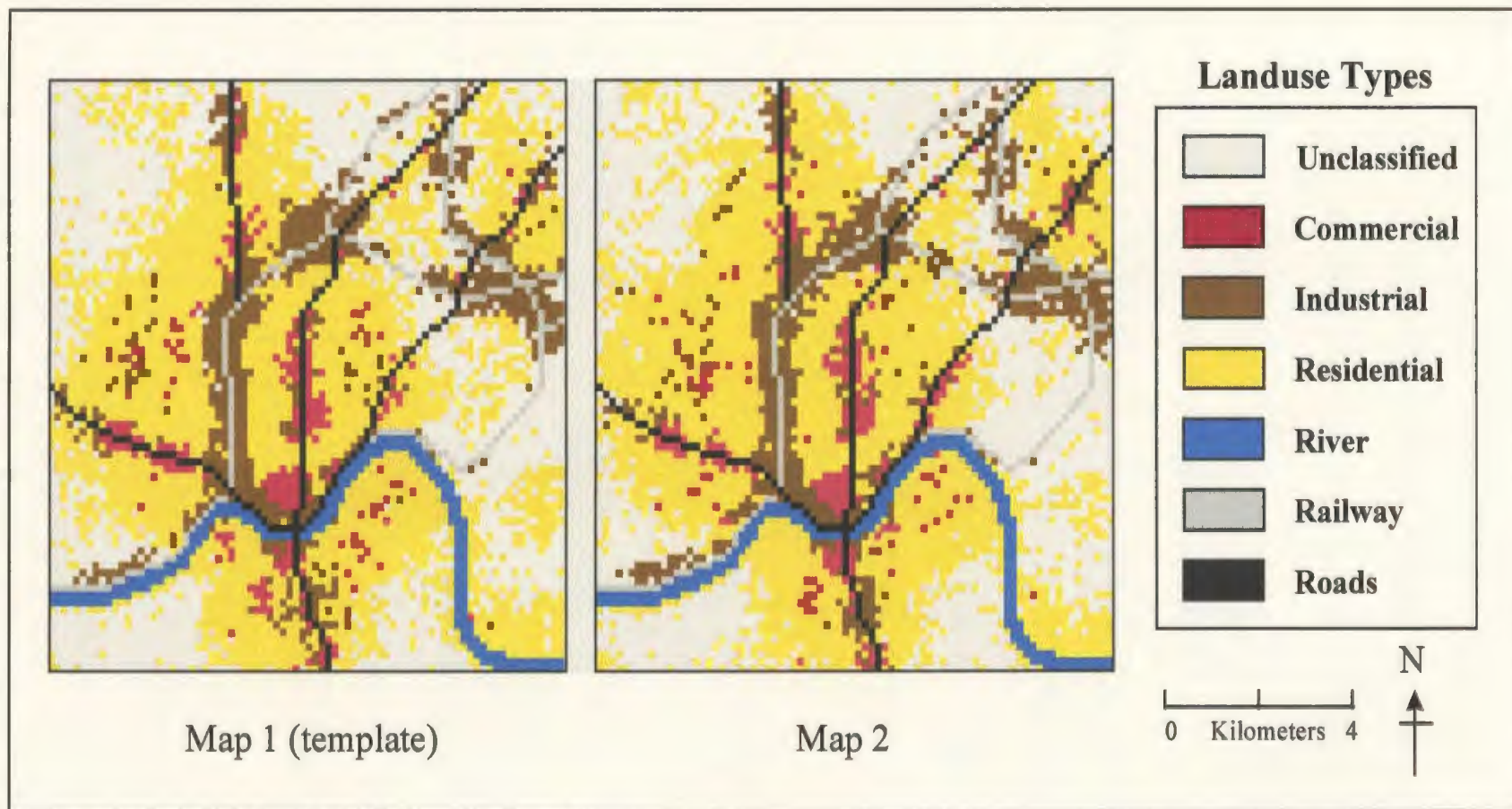


Figure 4.2.1 Land Use Maps of Cinc1

periphery of the city. Therefore, the qualitative interpretation of the agreement between the maps is that "map two is very similar to the template".

Visual inspection of the maps in Cinc2 (Figure 4.2.2) for land use agreement indicates that "map two is considerably different from the template". Most areas of the maps, especially the river and transportation systems, are in total disagreement, and only several small residential and unclassified polygons are similar.

The land use maps in Figure 4.2.3 show that the forest inventory classes have remained predominantly unchanged from 1985 to 1991. Regions of forest change, which include the no data and cleared categories, comprise a small section of the data set. Linguistically, the land use agreement can be expressed as "map two is fairly similar to the template".

### **4.3 Local Matching Results**

The local matching values from the fuzzy inference system are the membership values of the polygons on map two relative to a template map. Therefore, the higher the local matching value the greater the possibility that a polygon on map two is contained within a template polygon. Referring to analysis of land use differences, a low local matching value of 0.3 or less for a polygon represents land use disagreement between the maps.

The local matching values for Cinc1 are illustrated in Figure 4.3.1, which indicates a relatively high degree of agreement between the two maps for most areas. The low

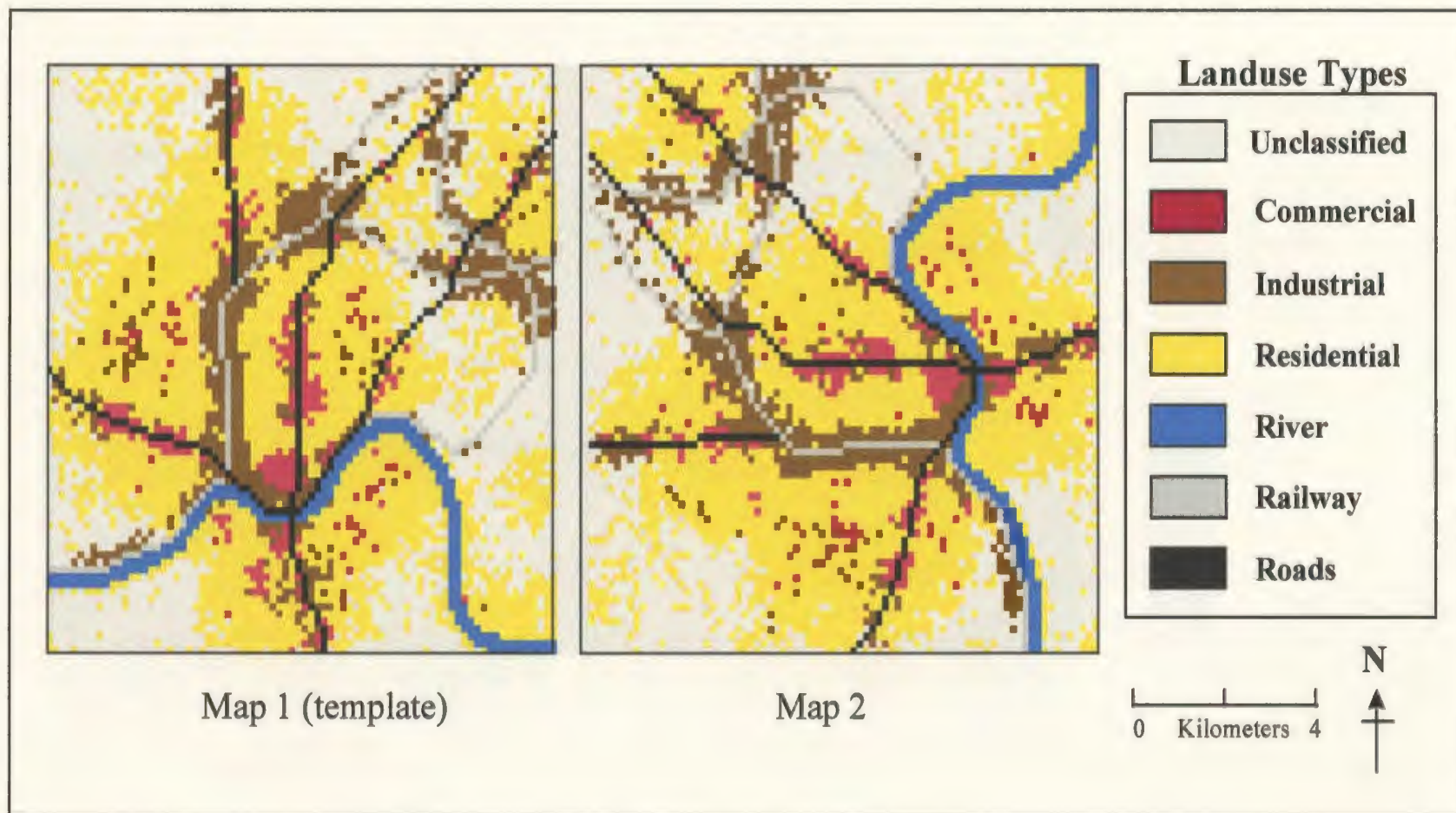


Figure 4.2.2 Land Use Maps of Cinc2



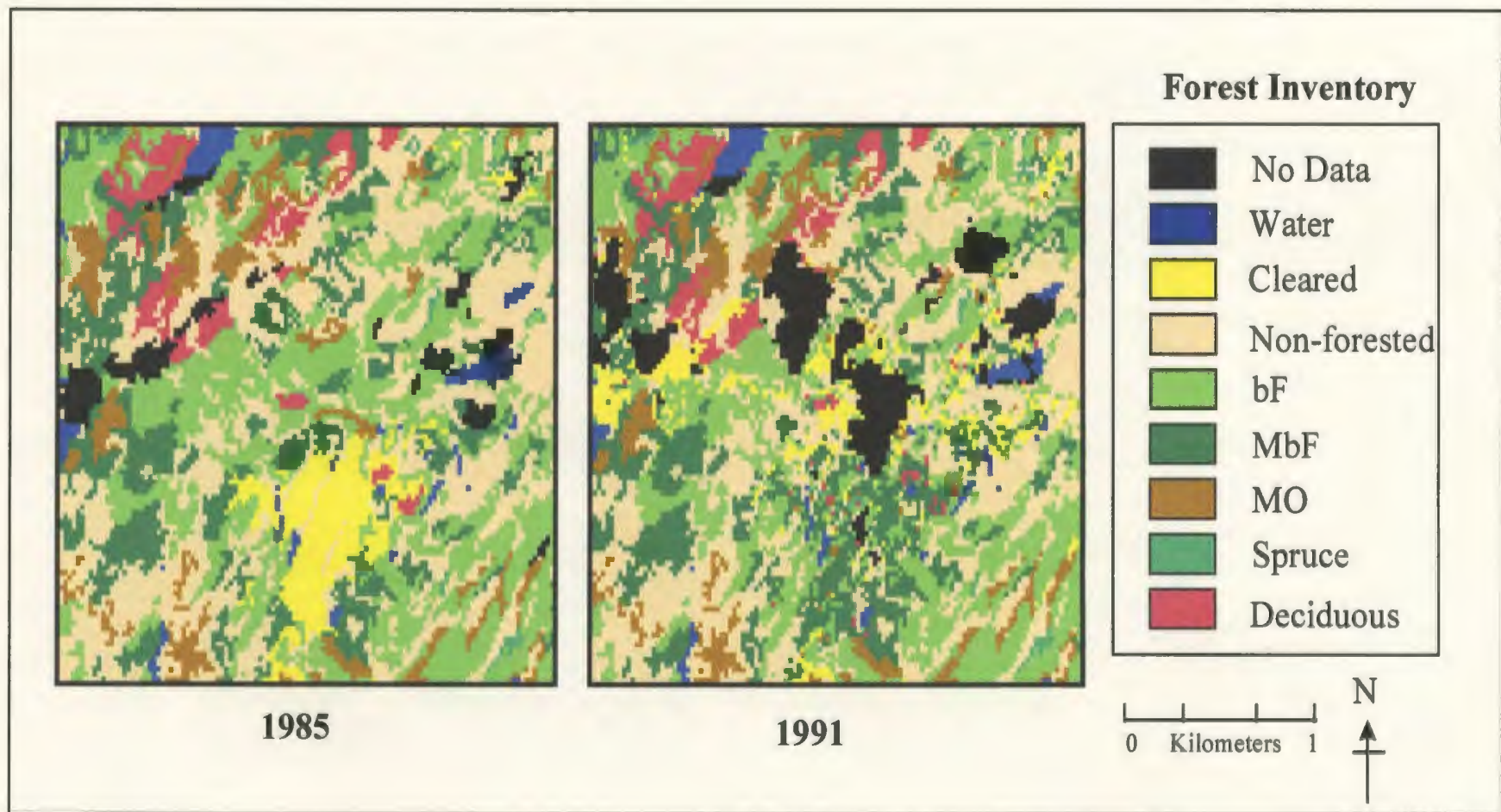


Figure 4.2.3 Forest Inventory Maps of Forest1

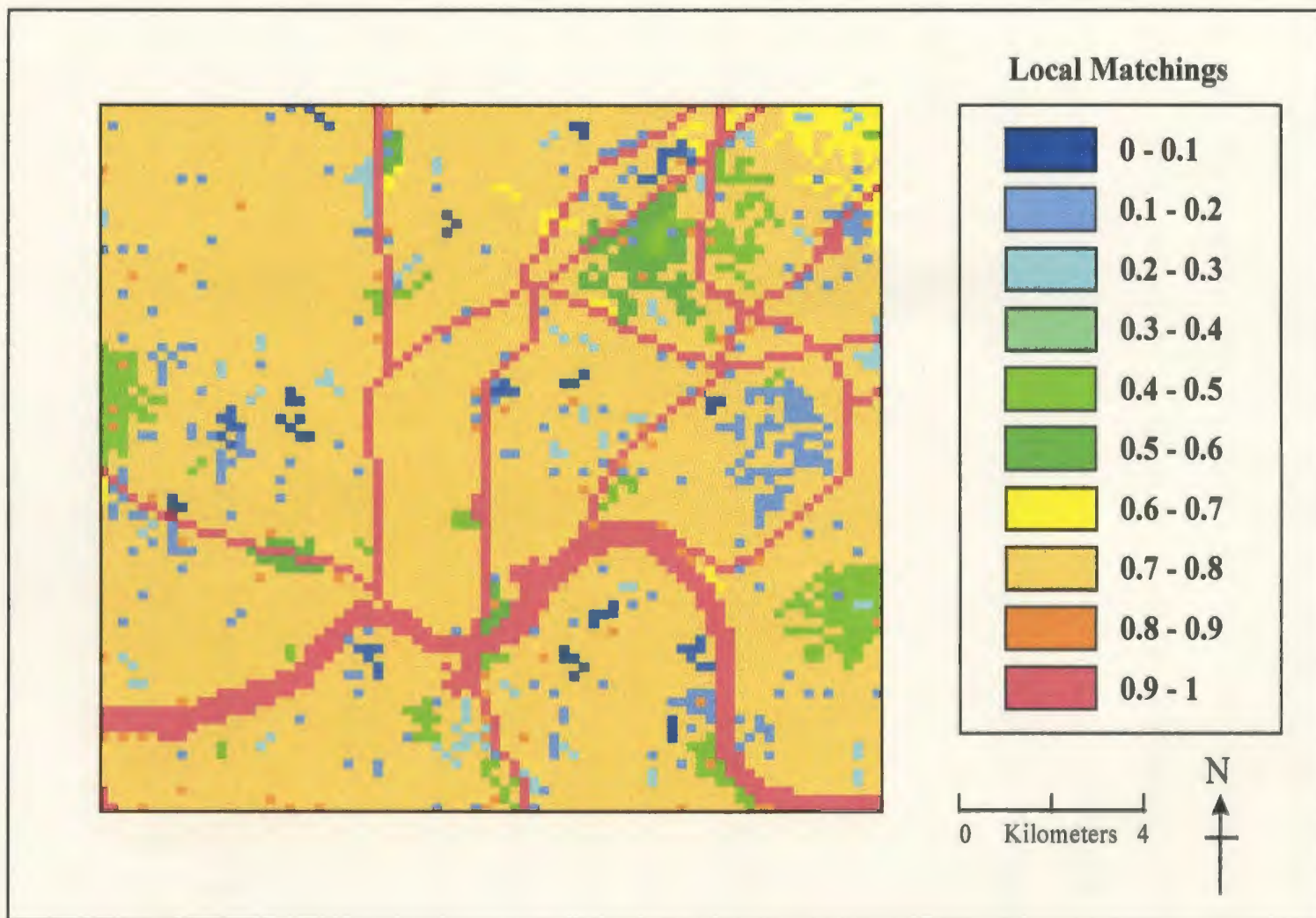


Figure 4.3.1 Local Template Polygon Matchings for Cinc1

matching values generally consist of smaller polygons that are dispersed throughout the study area.

It is expected that a comparison of similar land use maps would identify more land use matches than mismatches. Along the lines of a traditional comparison matrix, the incidences of land use agreement can be measured by a table of frequency of matches and mismatches for each land use category. Similar to the procedure presented by Gopal and Woodcock (1994), a fuzzy cut  $\alpha$  of 0.70 is used to measure the frequency of local matches. Formally,

$$\text{Local Match}(X, \alpha) = \begin{cases} 1 & \text{if } u_c \geq 0.7 \\ 0 & \text{otherwise} \end{cases} \quad (4.1)$$

A land use polygon on a second map is similar to a template polygon if its local matching membership grade is  $\geq 0.70$ .

Table 4.3.1 displays the results for Cinc1 using the fuzzy threshold agreement value. The first column shows the land use type, and the second column displays the total number of polygons for each map category. The matches and mismatches are given as numbers of polygons in columns three and four while the last column shows the percentage of land use agreement for each land use class.

The similarity percentages show that the rivers and transportation systems are in perfect(100%) agreement, but the results for the other categories suggest that substantial land use disagreement is evident. substantial land use disagreement is evident. With 130 matches and 240 mismatches, it is expected that the land use maps should be considerably different. This discrepancy is due to the pixel resolution of the data. The

maps in Cinc1 have a resolution of 250 meters, resulting in most of the polygons in the template layer consisting of one or two pixels. These small template polygons account for most of the mismatches between the maps even though they represent a small portion of the study area.

Table 4.3.1: Local Matchings for Land Use Polygons of Cinc1

| Land Use Type | # of Polygons | Match | Mismatch | Eval. of Similarity |
|---------------|---------------|-------|----------|---------------------|
| Unclassified  | 80            | 19    | 61       | 23.75%              |
| Commercial    | 79            | 32    | 47       | 40.5%               |
| Industrial    | 106           | 35    | 71       | 33.01%              |
| Residential   | 95            | 34    | 61       | 35.8%               |
| River         | 2             | 2     | 0        | 100%                |
| Railway       | 6             | 6     | 0        | 100%                |
| Roads         | 2             | 2     | 0        | 100%                |
| Total         | 370           | 130   | 240      | 61.87%              |

The matching results are further analyzed by fitting a trend curve to the points of a scatterplot (Figure 4.3.2) of mean polygon size and local matching. It is important to note that it was necessary to logarithmically transform the original mean polygon size values for each dataset to normalize their distribution and to enhance comparability with the local matching values. The trend curve was fitted to the points by the robust locally weighted regression smoothing (Lowess) technique of Cleveland (1979). Barringer *et al.*(1994) describe Lowess as a nonparametric iterative method of smoothing the points in a scatterplot where the fitted value is a polynomial fit to the data using a weighted least squares algorithm. The smoothing procedure can be described as

$$y_i = g(x_i) + \varepsilon_i$$

where  $g$  is a smooth function and  $\varepsilon_i$  are random variables with mean 0 and constant scale (Cleveland,1979).

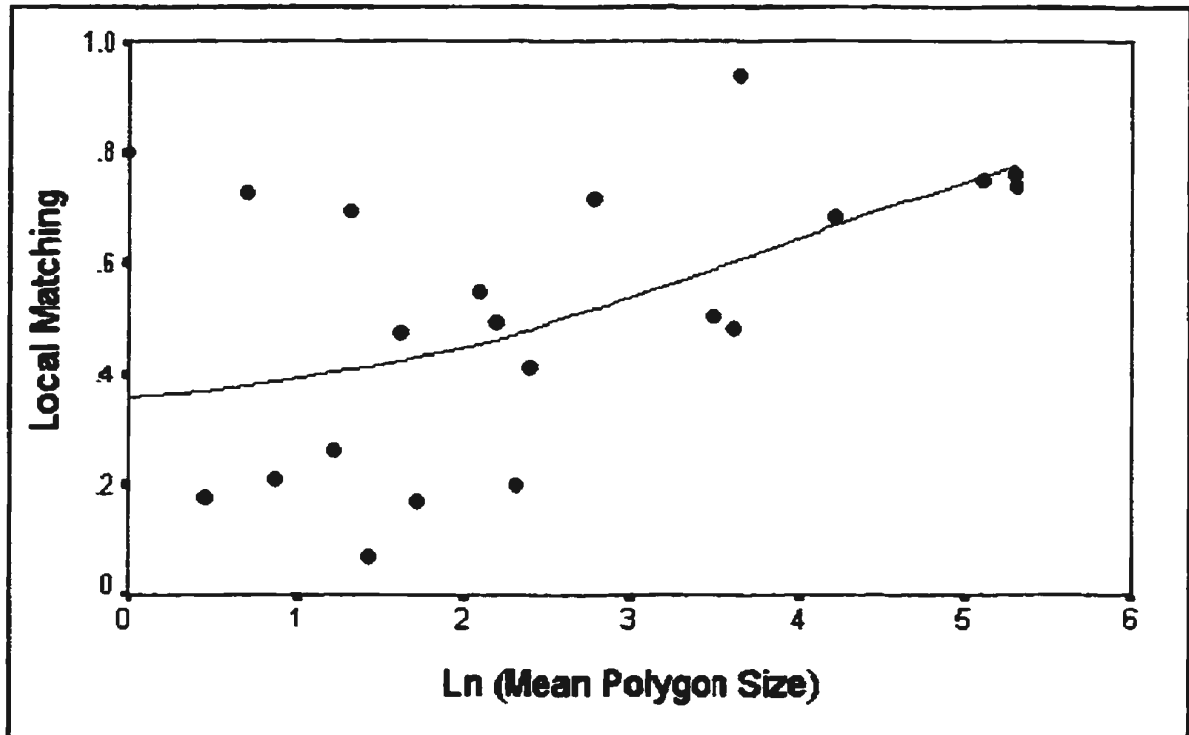


Figure 4.3.2 Lowess Curve of Mean Polygon Size versus Local Matching for Cincinnati

The assumption of smoothness allows points in the neighbourhood of  $(x_i, y_i)$  to be used in forming of the fitted value  $\hat{y}_i$ . The weights are a function of the distance between a point and the center of a smoothing kernel and a neighbour point and the size of the residuals from the previous iterations (Cleveland,1979). Therefore, the weights decrease as the distance between the points and the size of the residuals increases. Smoothing of a scatterplot refers to the formation of new points, which is dependent on the neighbourhood radius of the smoothing kernel (Barringer *et al.*,1994). The statistical software SPSS (1995) was used to produce the Lowess curves for the datasets, which

were fitted to 99 % of the data points with three iterations of weighting adjustments and smoothing.

A Lowess curve is appropriate for analyzing the matching relationship because it is robust to the effects of outliers and can be used to identify clear trends in highly variable datasets. The Lowess curve in Figure 4.3.2 predominantly trends upwards and indicates a positive causal relationship between mean polygon size and local matching. Thus, the matchings ( $\geq 0.7$ ) tend to represent land use agreement for larger polygons, which explains why the total area of agreement for 130 matches is greater than the area of disagreement for 240 mismatches.

The matchings evaluations presented in Table 4.3.2 show that the smaller number of matched polygons accounted for 85.98% of the template area. In addition, 188 of the 240 mismatches (Table 4.3.3) were for one or two pixel polygons that combined occupy only 4.73% of the template map.

Table 4.3.2: Evaluation of Matches and Mismatches for Cinc1

| Definite Matches     | Definite Mismatches  |
|----------------------|----------------------|
| 130                  | 240                  |
| 32.43% of polygons   | 67.57% of polygons   |
| 85.98% of total area | 14.02% of total area |

Table 4.3.3: Matching Results for Template Polygons for Cinc1 comprised of 1 and 2 Pixels

| 221 of 370 polygons are 1 pixel                | 41 of 370 polygons are 2 pixels                |
|--|--|
| 152 are mismatched;<br>3.45% of the total area | 36 are mismatched;<br>1.28 % of the total area |
| 69 are matched                                 | 5 are matched                                  |

The local matchings for most of the template polygons in Cinc2 are low (Figure 4.3.3) because the second land use map was deliberately made to be different from the template. The areas of high matching are limited to several sections of residential and unclassified land use in the center of the template. Table 4.3.4 contains the matches and mismatches information for Cinc2. The residential and unclassified are the only categories that had some degree of matching success. With the transposition of map two, the river and transportation systems are in complete disagreement (0%) and industry is poorly matched (10.4%).

Table 4.3.4 : Local Matchings for Land Use Polygons of Cinc2

| Land Use Type | # of Polygons | Match | Mismatch | Eval. of Similarity |
|---------------|---------------|-------|----------|---------------------|
| Unclassified  | 80            | 38    | 42       | 47.5%               |
| Commercial    | 79            | 3     | 76       | 3.80%               |
| Industrial    | 106           | 11    | 95       | 10.4%               |
| Residential   | 95            | 41    | 54       | 43.16%              |
| River         | 2             | 0     | 2        | 0                   |
| Railway       | 6             | 0     | 6        | 0                   |
| Roads         | 2             | 0     | 2        | 0                   |
| Total         | 370           | 93    | 277      | 15.41%              |

The smoothing trend in the scatterplot of mean polygon size and local matching (Figure 4.3.4) indicates that the effect of pixel resolution on the matching process supports the visual interpretation of differences between the maps. The Lowess curve is negative up to a mean polygon size of three when it becomes neutral for the remaining polygon sizes. This suggests that there is no overall discernible relationship between the variables although the regions that are in agreement represent small template polygons. It is possible that large areas could be both mismatched and matched, and the mismatches could occupy approximately the same portion of the template as the matches. Table 4.3.5 gives a true



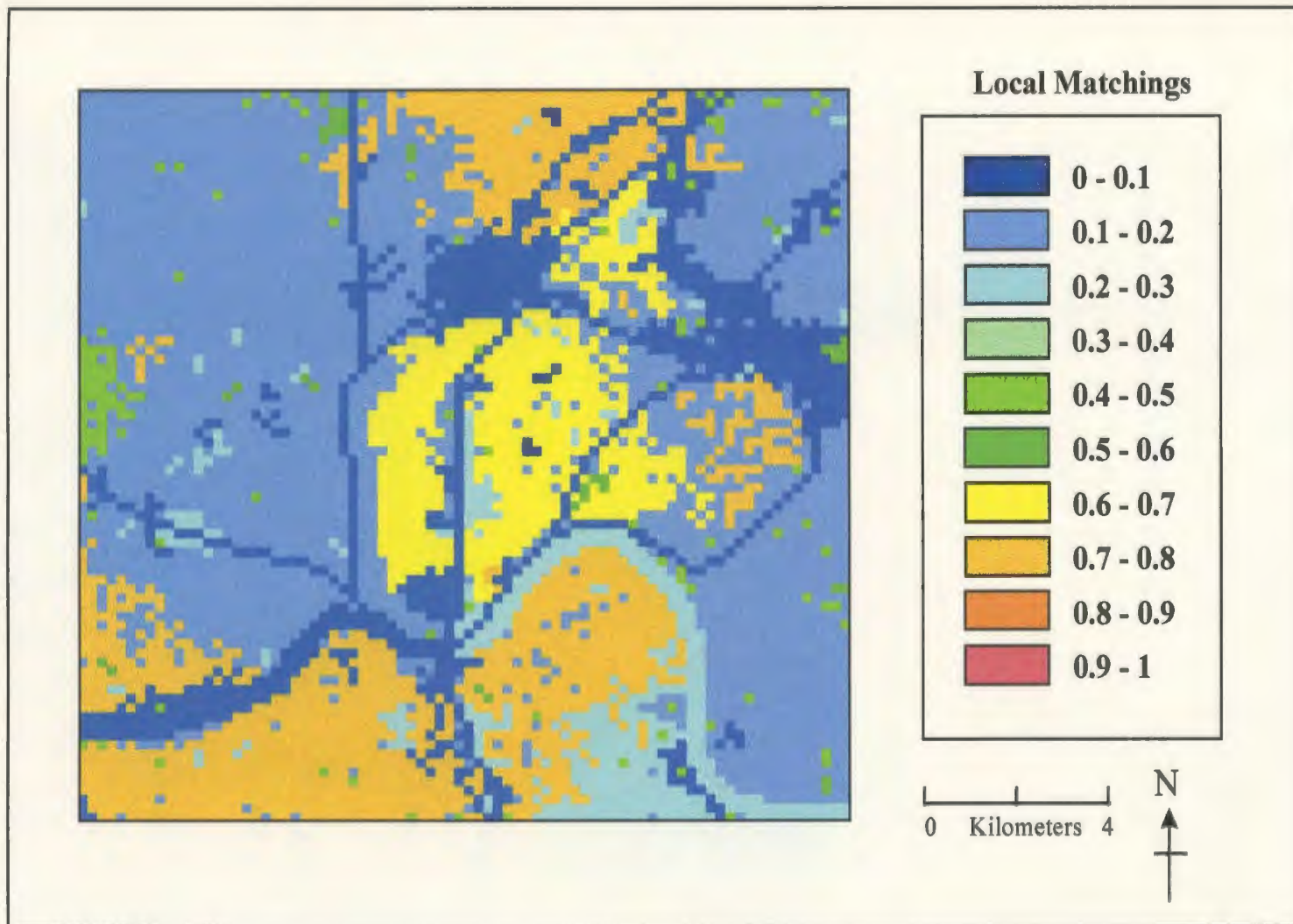


Figure 4.3.3 : Local Template Polygon Matchings for Cinc2



indication of the matching process with 77.55% of the template polygons having a matching value  $< 0.70$ . In addition, 204 of the 277 mismatched polygons consist of one or two pixels (Table 4.3.6) and account for 4.73% of the total area of the template layer. Although fewer in number, the other 73 larger polygons represent a considerable portion (72.82%) of the total mismatched area of the template.

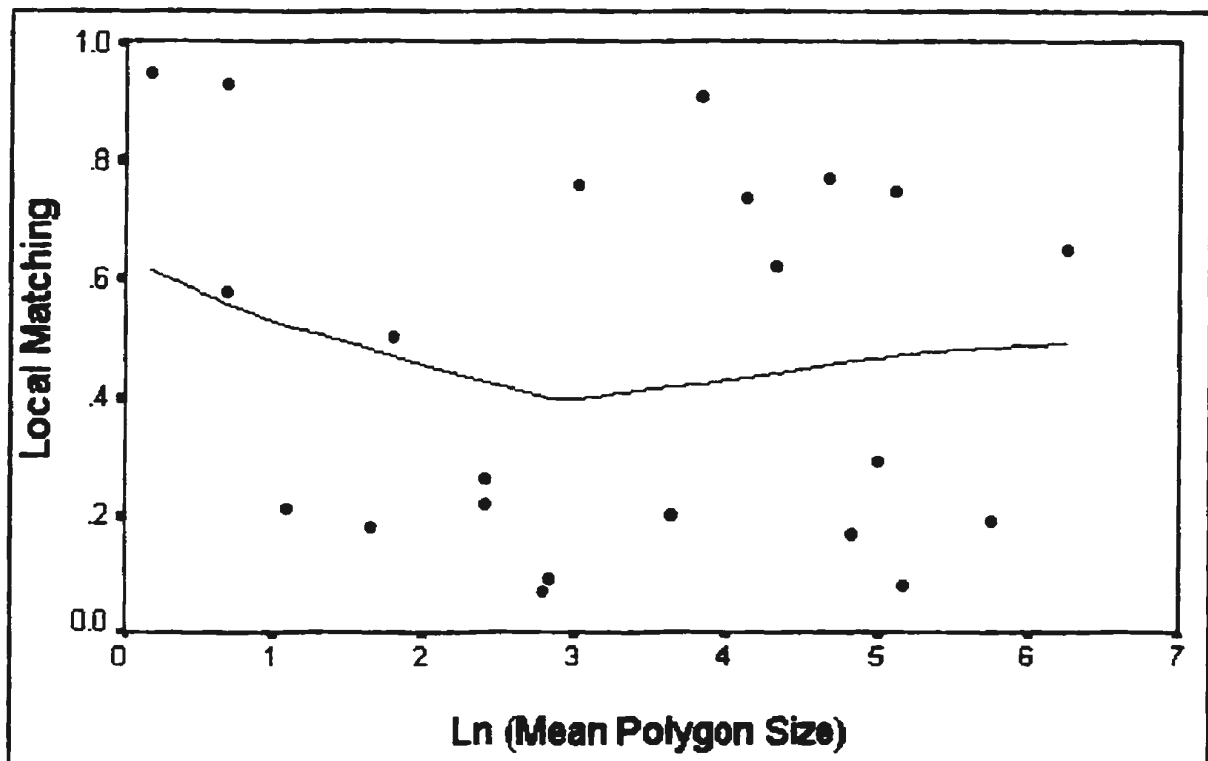


Figure 4.3.4 Lowess Curve of Mean Polygon Size versus Local Matching for Cinc2

Table 4.3.5: Evaluation of the Matches and Mismatches of Cinc2

| Definite Matches       | Definite Mismatches    |
|------------------------|------------------------|
| 93                     | 277                    |
| 25.14% of the polygons | 74.86% of the polygons |
| 22.45% of total area   | 77.55% of total area   |

**Table 4.3.6: Matching Results for Template Polygons for Cinc2 comprised of One and Two Pixels**

|  |   |
|--|---|
| 221 of 370 polygons are 1 pixel                | 41 of 370 polygons are 2 pixels               |
| 168 are mismatched;<br>3.45% of the total area | 36 are mismatched;<br>1.28% of the total area |
| 53 are matched                                 | 5 are matched                                 |

The local matchings for Forest1 indicate that little land use change has occurred from 1985 to 1991 (Figure 4.3.5). A majority of the matchings range from 0.70 to 1, representing a high degree of containment of the 1991 map within the 1985 template layer. The identified areas of change are found at regions where the land use could not be determined by the classification algorithm. These "no data" polygons are situated at different sections of their respective maps and will be mismatched. The potential problem is that these low matching regions may actually be similar but are identified as change because of a fault of the classification procedure. It is beyond the scope of this thesis to develop a method to handle this problem so the local matching values will be accepted as the true template matching results for the dataset.

The matching information in Table 4.3.7 shows that there is a high degree of polygonal land use pattern agreement between the maps. All of the land use categories, except cleared and no data, have similarity percentages greater than 83.0%. Unlike the previous datasets, the matched template polygons outnumber the mismatched polygons. For example, 321 of the 402 template polygons (see Table 4.3.8) matched their counterparts on the 1991 map for a 81.4 % overall areal agreement between the maps. This difference in matching results may be attributed to the 30 meter pixel resolution of

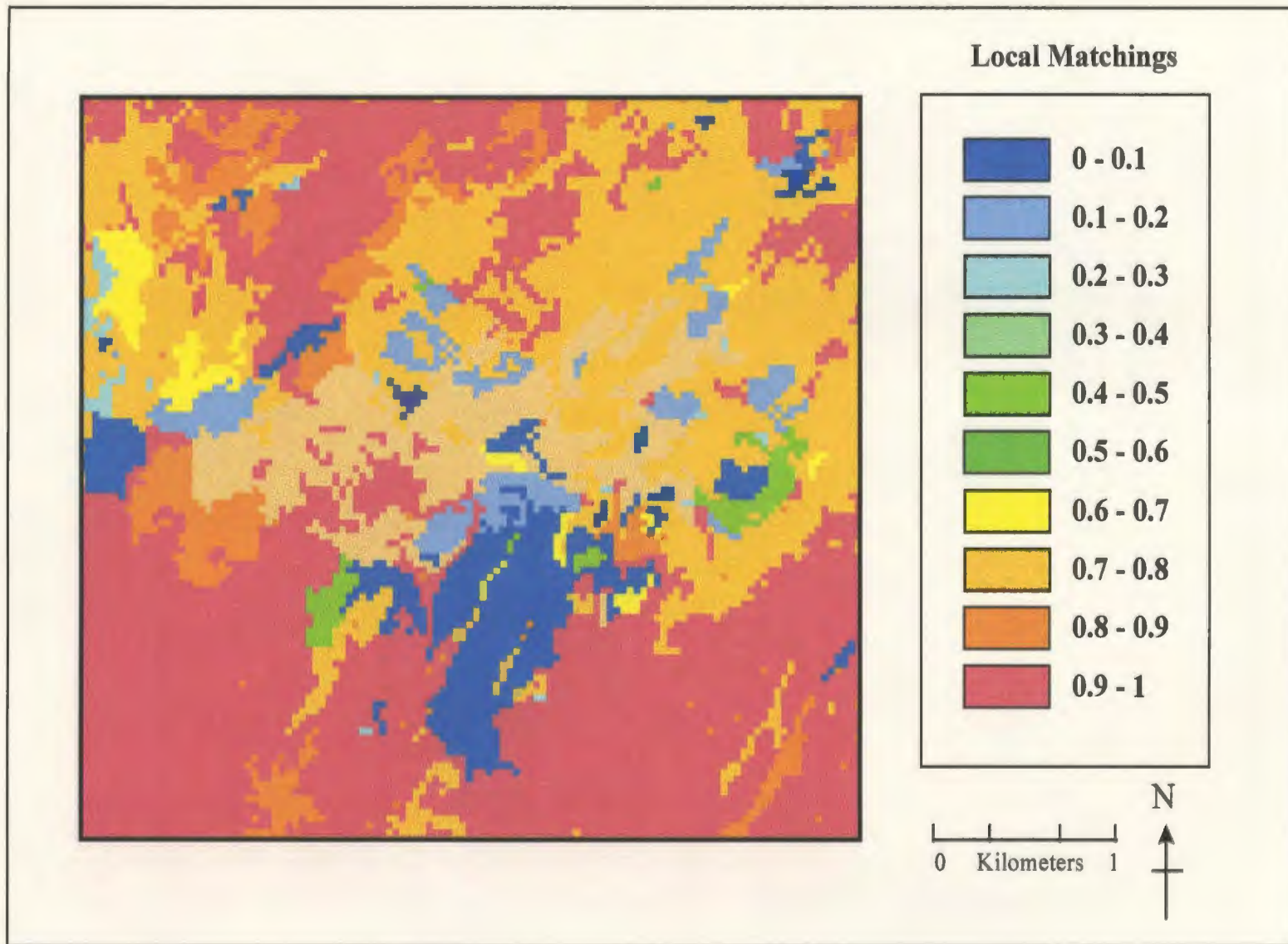


Figure 4.3.5 Local Template Polygon Matchings for Forest1

the land use maps in Forest1. The smaller scale of the template resulted in most of its unique polygons containing more than two pixels. Table 4.3.9 shows that 101 of the 401 polygons consisted of one or two pixels, of which 15 were mismatched.

Table 4.3.7: Local Matchings for Land Use Polygons of Forest1

| Land Use Type | # of Polygons | Match | Mismatch | Eval. of Similarity |
|---------------|---------------|-------|----------|---------------------|
| No Data       | 20            | 2     | 18       | 10.0 %              |
| Water         | 30            | 28    | 2        | 93.3 %              |
| Cleared       | 21            | 4     | 17       | 19.1%               |
| Non-Forested  | 118           | 105   | 13       | 88.9 %              |
| bF            | 60            | 53    | 7        | 88.3 %              |
| MbF           | 75            | 62    | 13       | 82.7%               |
| MO            | 56            | 49    | 7        | 87.5 %              |
| Spruce        | 10            | 10    | 0        | 100.0%              |
| Deciduous     | 12            | 8     | 4        | 66.7%               |
| Total         | 402           | 321   | 81       | 70.7%               |

Table 4.3.8: Evaluation of the Matches and Mismatches of Forest1

| Definite Matches      | Definite Mismatches   |
|-----------------------|-----------------------|
| 321 polygons          | 81 polygons           |
| 79.6% of the polygons | 20.4% of the polygons |
| 81.4% of total area   | 18.6% of total area   |

Table 4.3.9: Matching Results for Template Polygons for Forest1 comprised of One and Two Pixels

| 62 of 402 polygons are 1 pixel                | 39 of 402 polygons are 2 pixels               |
|---|---|
| 6 are mismatched;<br>0.48 % of the total area | 9 are mismatched;<br>0.60 % of the total area |
| 56 are matched                                | 30 are matched                                |

The effect that a decrease in pixel resolution had on the relationship between mean polygon size and local matching is shown on Figure 4.3.6. The data points in the scatterplot have a distinguishable positive orientation, and the trend in the local matching Lowess curve increases monotonically, but at varying rates, as the mean polygon size

increases. The trend shows that the larger polygons on the maps are in agreement and that the mismatched areas are generally comprised of small template polygons. This is due to the pixel group membership functions being sensitive to scale.

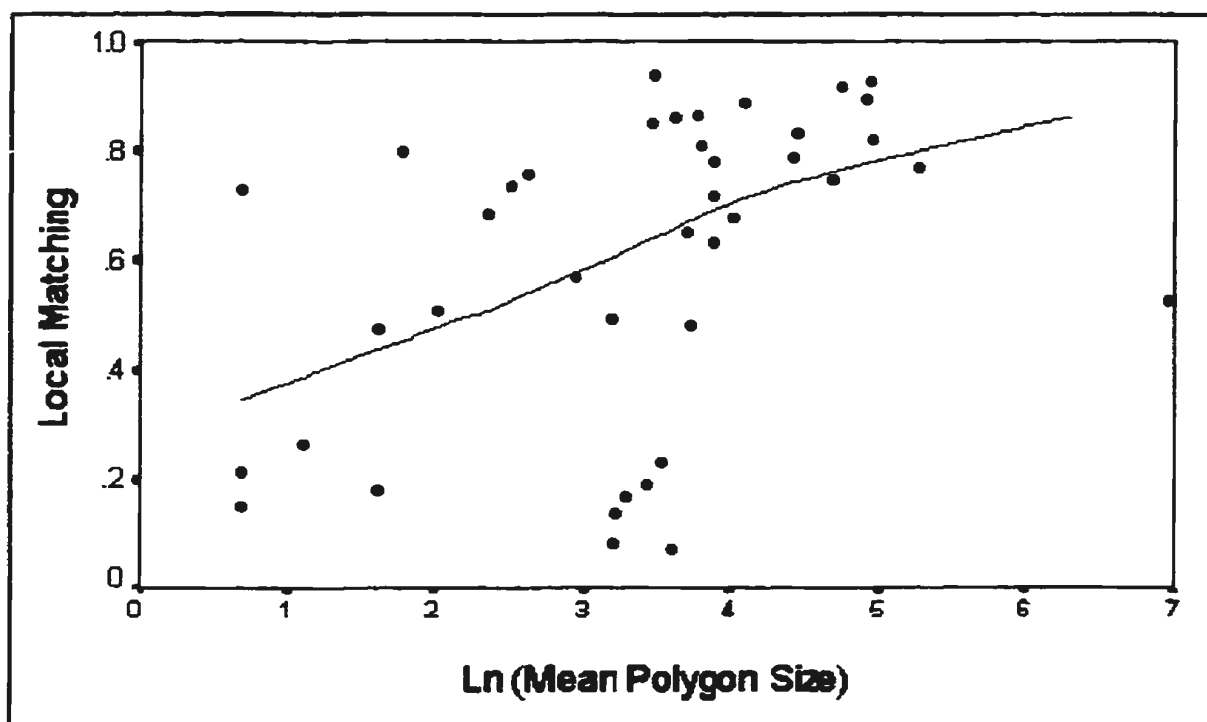


Figure 4.3.6 Lowess Curve of Mean Polygon Size versus Local Matching for Forest1

A local matching value will depend on the perceived possibility that a single pixel disagreement could be random mismatching or actual land use change. At low resolutions, such as 30 meters, the possibility that a single pixel polygon disagreement is a random mismatch increases. The fuzzy inference system handles this uncertainty by computing a slightly higher local matching value than would normally be determined for a mismatch. This would classify the template polygon as moderately instead of poorly

matched. However, a single pixel polygon disagreement for a large pixel resolution dataset is more likely to represent actual land use change than random disagreement. As the pixel resolution increases, the model will compute a local matching value indicative of definite change. For datasets whose land use maps are relatively similar, this could lead to fewer identified mismatches for the lower pixel resolution template.

#### **4.4 Global Matching Evaluation**

As with the traditional similarity statistics, the computation of the global matching value is area based. This is due to the fact that it represents the degree of global areal containment of a map within a template layer. One advantage of the fuzzy pattern matching model over the standard approaches is that Shannon entropy is used to measure the reliability and validity of the global matching value. Consequently, a comparison with a number of common similarity statistics will indicate how the global matching process has performed and whether there is any advantage with applying a fuzzy approach to an overall land use agreement analysis.

Most boolean similarity statistics are derived from coincidence or comparison matrices. Bonham-Carter (1994) states that the coefficient of areal agreement and Cramer's V are commonly used for comparing multiple class maps because of their ease of calculation. For maps that contain the same number of classes, the Kappa and Tau coefficients of agreement have been utilized (Rosenfield and Fitzpatrick-Lins, 1986; Ma and Redmond, 1995).

Taylor (1977) describes the coefficient of areal association (CAA) as a simple measurement of areal agreement derived from the intersection of two maps. The intersection identifies areas of maps that are similar, and the coefficient of areal association measures the observed proportion of agreement between the maps. The diagonal values in the coincidence matrix represent the areal association between the classes, and their aggregation relative to the total area of the maps gives an evaluation of the overall areal agreement. Formally, the coefficient of areal association is computed as:

$$CAA = \sum_{i=1} Area_{ij} / \text{Total Area}, \quad (4.2)$$

where  $Area_{ij}$  is the diagonal value  $ij$  in the coincidence matrix.

Cramer's  $V$  is a similarity measure that is derived from the chi square ( $X^2$ ) statistic, and is used to measure the strength of associations between maps. It should be noted that despite the importance of  $X^2$  in the calculation of Cramer's  $V$ ,  $X^2$  by itself is inappropriate for representing the areal similarity between maps. Chi square is computed from a table of counts which requires an exhaustive mapping rather than a sampling of areas (Goodchild, 1998). It is used to test the statistical significance of associations while Cramer's  $V$  measures the strength of associations in the coincidence matrix. Nevertheless,  $X^2$  can provide an exploratory and descriptive measure of spatial correlations between maps if the analyst recognizes these limitations.

For the chi square statistic, the overall similarity between two maps is based on the observed and expected areal categorical agreements in the coincidence matrix. Bonham Carter (1994) defines chi square as :

$$\chi^2(T) = \sum_{i=1}^N \sum_{j=1}^M \frac{(\text{Observed}(ij) - \text{Expected}(i * j))^2}{\text{Expected}(i * j)}, \quad (4.3)$$

where T is the coincidence matrix.

The observed similarity element  $T_{ij}$  is the agreement value between the  $i^{\text{th}}$  class of map B (rows in T) and the  $j^{\text{th}}$  class of map A (columns of T). The expected areal similarity value  $T_{i*j}$  for the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column is computed as :

$$T_{i*j} = \frac{T_{i.} T_{.j}}{T_{..}}, \quad (4.4)$$

where  $T_{i.}$  is the row total of T,  $T_{.j}$  is the column total of T, and  $T_{..}$  is the grand total summed over T. A similarity measure derived from chi square is Cramer's V:

$$V = \sqrt{\frac{\chi^2}{T_{..} M}} ; M \text{ is the minimum of } (n-1, m-1). \quad (4.5)$$

Cramer's V ranges from 0 (no correlation) to 1 (total correlation).

The Kappa coefficient of agreement is a similarity measurement used to compare two maps that contain the same number of classes and whose class themes are identical. Rosenfield and Fitzpatrick-Lins (1986) describe Kappa as a statistic that measures the agreement between attributes of two maps and corrects for the expected amount of chance agreement.

Kappa has been used extensively in remote sensing for accuracy assessment of classified satellite imagery. The post classification comparison is performed by a crosstabulation of two maps to generate a comparison matrix and associated statistics that indicate regions of change (Singh, 1989). From the comparison matrix, an observed areal



proportion and expected proportion matrix are defined. The observed area proportions  $P_{ii}$  are the diagonal values in the observed proportions matrix determined from  $P_{ii} = T_{ii}/T_{..}$ . The expected area proportions  $Q_{ii}$  are the diagonal values in the expected proportions matrix, and are found with  $Q_{ii} = P_{i.} / P_{.j}$ . The Kappa coefficient is defined as:

$$K = \frac{\sum_{i=1}^N P_{ii} - \sum_{i=1}^N Q_{ii}}{1 - \sum_{i=1}^N Q_{ii}} \quad (4.6)$$

The calculated Kappa value lies typically on a scale between 0 and 1 (total agreement) with 0 representing agreement that is no better than chance (Bonham Carter, 1994).

The Tau coefficient of agreement is similar to Kappa in that it uses the information in a coincidence matrix to measure the improvement of a classification over a random assignment of pixels to groups. Despite the similarities, Tau has several advantages over Kappa. First, it can incorporate either equal or unequal *a priori* probabilities of group membership into the agreement calculation, whereas Kappa uses the *a posteriori* probabilities. Secondly, Tau avoids underestimating the overall similarity accuracy because the probabilities are independent of the error matrix (Ma and Redmond, 1995). The probabilities are computed from an independent source before the error matrix is generated. For thematic maps, the independent source is the known or estimated proportion of the total area of the map occupied by a specific land use. For example, a researcher may know that a study area is comprised of 20% urban, 40% water, 10% grassland, and 30% forest. In this case, the *a priori* probabilities are .20, .40, .10, and .30, respectively.

According to Klecka (1980), Tau is calculated as

$$T = \frac{P_o - P_r}{1 - P_r} \quad (4.7)$$

where  $P_o$  (percentage agreement) =  $\sum_{i=1}^M P_{ii}$  (4.8),  $P_r$  (random agreement) =  $\sum_{i=1}^M (P_{i+})P_i$

(4.9),  $M$  is the number of groups,  $P_i$  is the *a priori* probability and  $P_{i+}$  is the marginal group total in the proportions error matrix.

#### 4.4.1 Evaluation of Boolean versus Fuzzy Global Similarity

The global matching procedure is initially evaluated by determining if the linguistic agreement statements support the global matching values for each dataset. Technically, fuzzy pattern matching translates a visual interpretation of agreement between two maps into a single global agreement value. Therefore, the linguistic agreement statement should qualitatively make sense in describing the global matching value if the matching procedure has performed correctly and accurately. For example, the matching process has failed if a global matching value of 0.1 is computed for maps that are visually very similar. Note that this activity is only a preliminary indicator of the comparison results, and it should not be used to measure the performance of the model. Recall that the linguistic agreement statements for Cinc1, Cinc2, and Forest1 are fairly similar, considerably different, and very similar. From Table 4.4.1, the global matching values are 0.71, 0.35, and 0.78, respectively. Although highly subjective, each of the linguistic agreement statements is an acceptable and sensible description of their associated global matching value. For example,

a label of fairly similar fits a 0.71 global matching value, which validates the performance of the matching process for Cinc1.

**Table 4.4.1 : Boolean and Fuzzy Global Similarity Statistics**

| Dataset | Fuzzy Global Matching | Fuzzy Certainty | Boolean Kappa | Boolean CAA | Boolean Cramer's V | Boolean Tau |
|---------|-----------------------|-----------------|---------------|-------------|--------------------|-------------|
| Cinc1   | 0.71                  | 0.95            | 0.64          | 0.75        | 0.82               | 0.70        |
| Cinc2   | 0.35                  | 0.97            | 0.09          | 0.37        | 0.12               | 0.27        |
| Forest1 | 0.78                  | 0.97            | 0.76          | 0.81        | 0.79               | 0.78        |

A quantitative and more accurate measurement of the performance of the global matching procedure for each dataset can be determined from the certainty measurement for the matching value. For example, Table 4.4.1 indicates that the global matching of the predicted map to the template layer for Cinc1 has a certainty of 0.95 while the certainty of matching for both Cinc2 and Forest1 is 0.97. A certainty value is similar to a probability confidence interval in that it indicates how well the global matchings represent the actual agreement between two maps. With the certainties all  $\geq 0.95$ , the accuracy and reliability of the global matchings are assured.

Furthermore, the reliability of the global matching values can also be estimated from comparisons to a number of standard boolean similarity measures. For the first comparison, notice that the global matching values are less than the Cramer's V values for Cinc1 and Forest1 (See Table 4.4.1) and greater than Cramer's V for Cinc2. These differences may be attributed to the fact that Cramer's V is computed from the chi square statistic. Several authors (Bonham-Carter,1994; Davis,1990) explain how chi square and

its associated statistics are best suited for an exploratory measure of multiple map correlation because they are dependent on area units. When converting to a smaller unit (e.g. km<sup>2</sup> to m<sup>2</sup>), chi square will increase in proportion to the increase in area unit. Therefore, the Cramer's V values are questionable as accurate measures of map similarity.

For the second comparison analysis, the global matchings fall between the calculated Kappa and CAA numbers. For example, the global matching value for Cinc1 is 0.71, which is between the Kappa value of 0.64 and the CAA of 0.75. This is the expected or desired result because of the problems with both Kappa and CAA. Foody (1992) found that Kappa consistently overestimates chance agreement and consequently underestimates map agreement. Congalton *et al.* (1983) state that the CAA is an unreliable measurement of map similarity because it overestimate the agreement between maps by not accounting for chance agreement. Based on this information, a preliminary requirement of an acceptable global similarity procedure is that its output value for a particular map comparison fall between the computed Kappa and the CAA values.

The Tau coefficients and global matchings both satisfy the above requirement. In addition, the Tau values of 0.70, 0.28, and 0.78 are very similar to the global matchings values of 0.71, 0.35, and 0.78 (see Table 4.4.1). Ma and Redmond (1995) describe how the use of Tau over Kappa and CAA is justified for its ability to incorporate probabilities into the calculations which avoids overestimating the random agreement between maps. However, the authors failed to consider that Tau depends on a pixel-by-pixel comparison to obtain the observed agreements for the map categories. Misregistration of one or both of the maps could decrease the computed agreement value. By accounting for locational

and attribute uncertainties in the computation of the local matching, the fuzzy global matching procedure, which derives the overall agreement from the aggregation of the local matchings, is clearly more appropriate and flexible than the boolean methods for the analysis of map similarity. Unlike the boolean approach, chance agreement and misregistration problems are handled by the overlap of the output membership functions for the local matchings. A global matching value may be viewed as a fuzzy coefficient of areal agreement since each local matching is estimated from an intersection of land use polygons.

#### **4.5 Fuzzy versus Boolean Land Use Comparison Results**

The use of fuzzy pattern matching for land use change detection and map similarity analysis will depend on the purpose of the map comparison procedure. Generally, the boolean procedures are sufficient for a map comparison analysis and may be preferred for their ease of understanding and implementation. Fuzzy pattern matching becomes a viable alternate map comparison technique when the complexity of the data and the matching process increases beyond the capability of the structure of the boolean approaches. Therefore, the structure of the map agreement procedure will depend on complexity of the maps and the objectives of the project. This complicates the direct comparison of the agreement results from the boolean and fuzzy models. Beyond the comparison of global agreement values as presented in Section 4.4.1, it is difficult to quantify the advantages of fuzzy pattern matching over the boolean approaches. Consequently, a visual interpretation

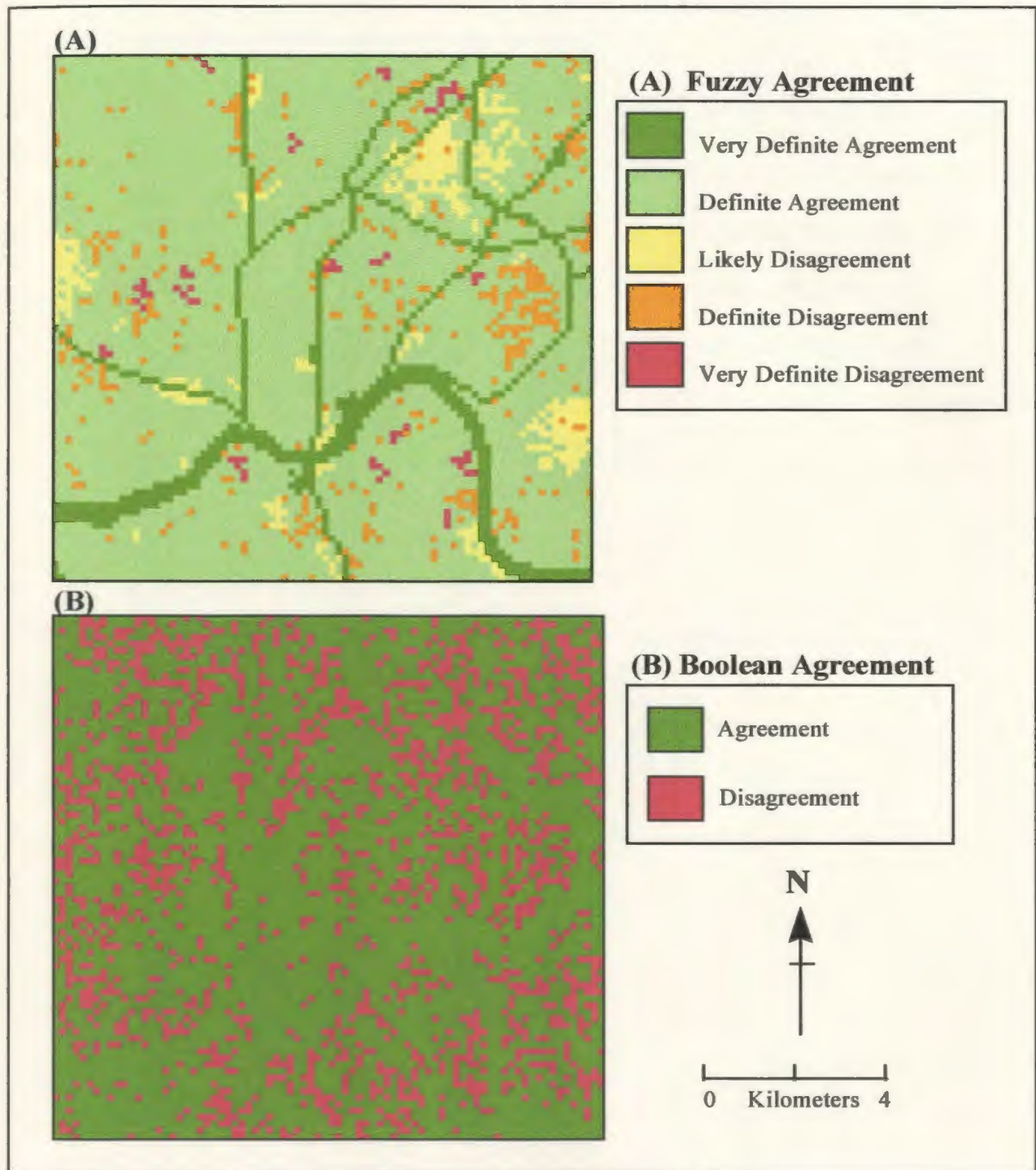
of the differences between the fuzzy and boolean agreement maps is the basis of the discussion of the benefits of fuzzy pattern matching as a map comparison technique.

Figures 4.5.1 and 4.5.2 show the fuzzy and boolean land use similarity maps for Cinc1 and Cinc2, respectively. Map B of both figures is a boolean agreement map produced by reclassifying a crosstabulation map into discrete agreement and disagreement categories. The fuzzy land use agreement possibility layers (Map A) were created by reclassifying the local matching maps according to the band width of the output membership functions (see Table 4.5.1). Recall that the parameters of the output membership functions are subjectively chosen by the developer. Similarly, the labels assigned to the five possibility classes in Table 4.5.1 are subjective linguistic descriptors of what a researcher believes is the degree of local matching between two maps.

Table 4.5.1 Local Matching Possibility Threshold Values

| Possibility Class          | Lower Threshold Value | Upper Threshold Value |
|----------------------------|-----------------------|-----------------------|
| Very Definite Agreement    | 0.91                  | 1.00                  |
| Definite Agreement         | 0.66                  | 0.90                  |
| Likely Agreement           | 0.31                  | 0.65                  |
| Definite Disagreement      | 0.061                 | 0.30                  |
| Very Definite Disagreement | 0.0                   | 0.06                  |

A noticeable difference between the fuzzy and boolean land use agreement maps is the gradual transition of similarity from very definite agreement to very definite disagreement on the fuzzy map. It is possible that an analyst may have difficulty interpreting the meaning of the fuzzy labels whereas the distinction between the boolean agreement and disagreement classes is straightforward. Although the labels may appear similar, the discrimination between the fuzzy classes is based on the certainty of matching.



**Figure 4.5.1 Fuzzy (A) versus Boolean (B) Agreement for Cinc1**



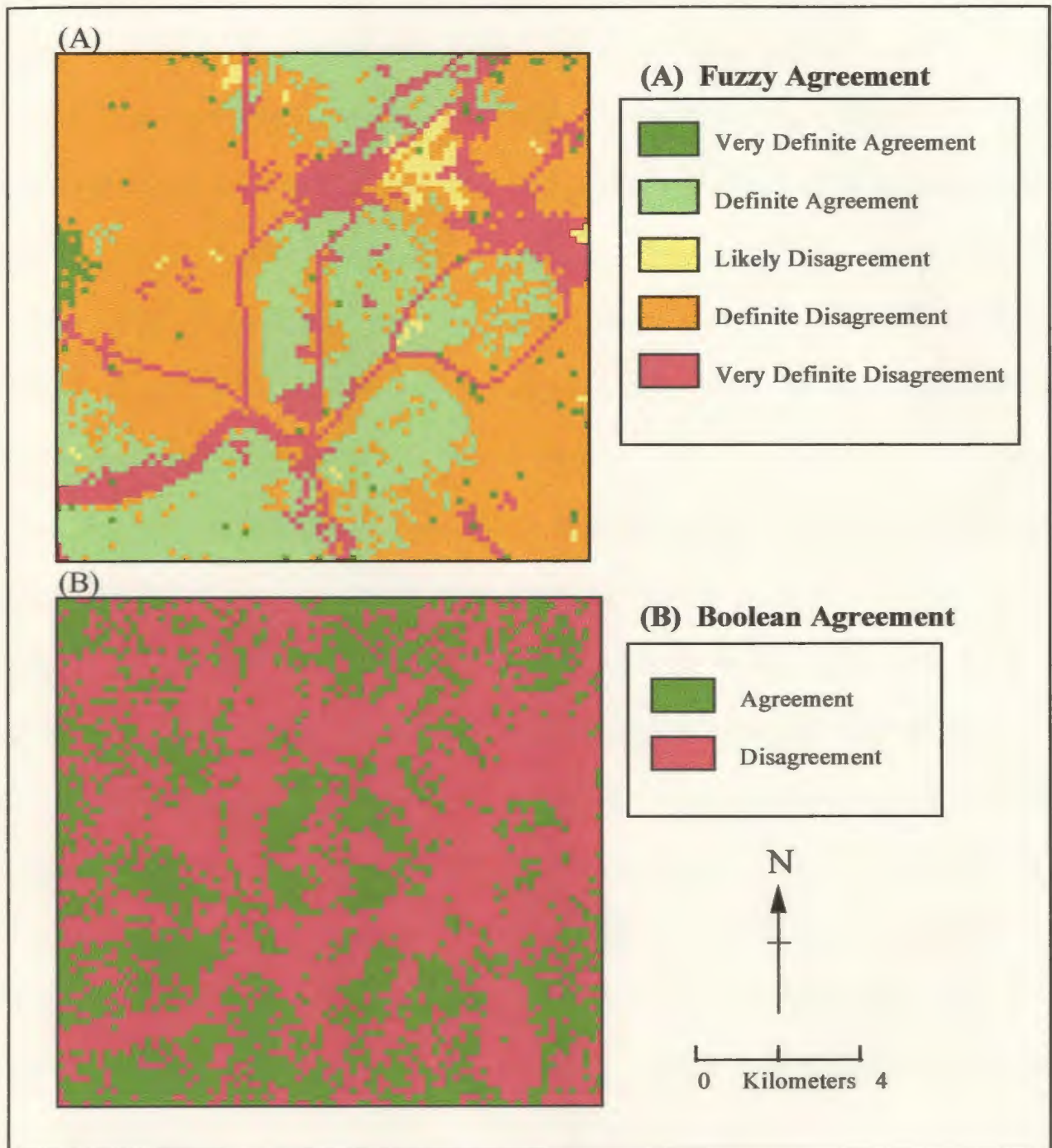


Figure 4.5.2 Fuzzy (A) versus Boolean (B) Agreement for Cinc2



For example, the placement of the linguistic hedge “very” before the descriptor “definite” for the very definite agreement class signifies a reduction in the fuzziness and uncertainty associated with a local matching event. Thus, there is a higher degree of certainty of land use agreement for regions classified as very definite agreement as compared to areas of definite agreement. It may be argued that the choice of labels and threshold values for the fuzzy classes will vary from developer to developer which complicates the interpretability of fuzzy maps. Regardless of the possible differences that could exist, the underlying characteristic of any fuzzy land use agreement map is that the classes are separated and interpreted by the perceived certainty of matching

The primary advantage of a fuzzy agreement map is that it contains more information and gives a more realistic interpretation of the land use characteristics of a dataset. The fuzzy agreement information would permit the user to concentrate on specific characteristics of the results, such as whether a specific land use type accounts for most of the very definite disagreement areas. Since datasets one and two were produced by a cellular automata land use prediction model, an analyst can use the information about the very definite and definite disagreement areas to recalibrate the model to produce better prediction results. This may be difficult or impossible with boolean results because the boolean approaches often lose agreement information when producing dichotomous similarity categories.

A second advantage of the fuzzy agreement map is that it retains the form of the template layer. This gives a better visual impression of where land use differences are situated spatially. For example, the areas of very definite agreement on Figure 4.5.1

represent the river and transportation system of the study area. However, the boolean map consists of a jumbled intermixture of disagreement and agreement areas that make it difficult to relate the result to the original land use maps. It is evident that the discrete classification from the boolean model has simplified the land use similarity results.

The comparison of the forest inventory maps in Forest1 demonstrates how well fuzzy land use pattern matching detected forest succession during the six year study period. It should be mentioned that forest regeneration and succession are complex and complicated processes that are often difficult to model with traditional boolean techniques. This is partially due to the inability of such techniques to represent intermediate growth patterns. Unless a major event, such as an insect infestation or a forest fire, has occurred, the patterns of change in forest inventory over six years will tend to be sporadic and fragmented (Meades and Moores, 1989). It is unlikely that a complete natural transformation from one forest inventory type to another will occur in such a short time span. Even if a fire has taken place, the affected area will experience intermediate primary succession to some degree.

The sensitivity of the fuzzy pattern matching model to complex growth patterns was determined by concentrating the change detection analysis on the cleared and non-forested categories. These forest inventory types were considered to be the ones most likely to produce mixed succession and regeneration results. The boolean classification (Figure 4.5.3) identifies the discrete change and no change classes for each forest inventory type but fails to find areas of mixed change. Thus, the intermediate change

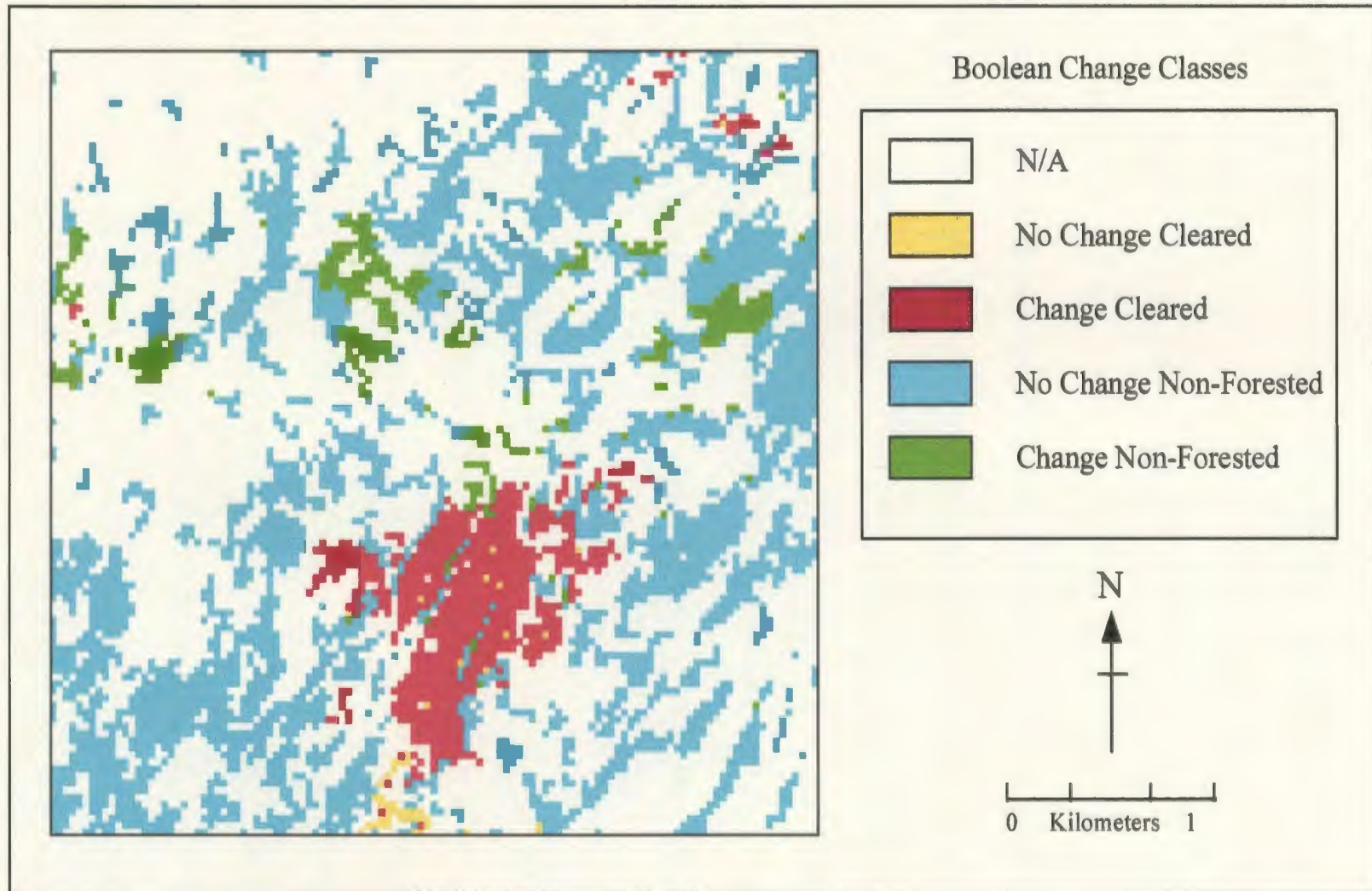


Figure 4.5.3 Boolean Change Classes for Cleared and Non-Forested

information is lost because the boolean approach constrains and simplifies the change detection process.

The transitional range of change on the fuzzy land use possibility map (Figure 4.5.4) shows that the fuzzy model detected intermediate and definitive change patterns. Several areas of intermediate change exist for the non-forested category and are classified as likely change non-forested. These are regions where the matching process has determined that approximately half of a template polygon is contained within the 1991 map. Consider the large polygon of likely change non-forested in the Northwest corner of Figure 4.5.4. The boolean model subdivides this region into areas of definite change and no change. This suggests that entire sections have undergone a complete land use change while other areas have remained unchanged. It is unrealistic that a boolean boundary could separate where forest succession has taken place. In contrast, the likely change classification on the fuzzy map indicates that gradual forest infilling has occurred, but there has not been a complete transformation in forest inventory type. Therefore, the fuzzy map contains more information about the change characteristics of the study area and presents a more appropriate interpretation of dynamics of forest species succession.

A further advantage of using fuzzy change possibilities rather than boolean categories is that there are fewer one pixel agreements and disagreements on the fuzzy map. With the resolution of the input data being 30 meters, the fuzzy inference system is sensitive to the possibility of random disagreements between the maps while the boolean model identifies every pixel-by-pixel disagreement as change.

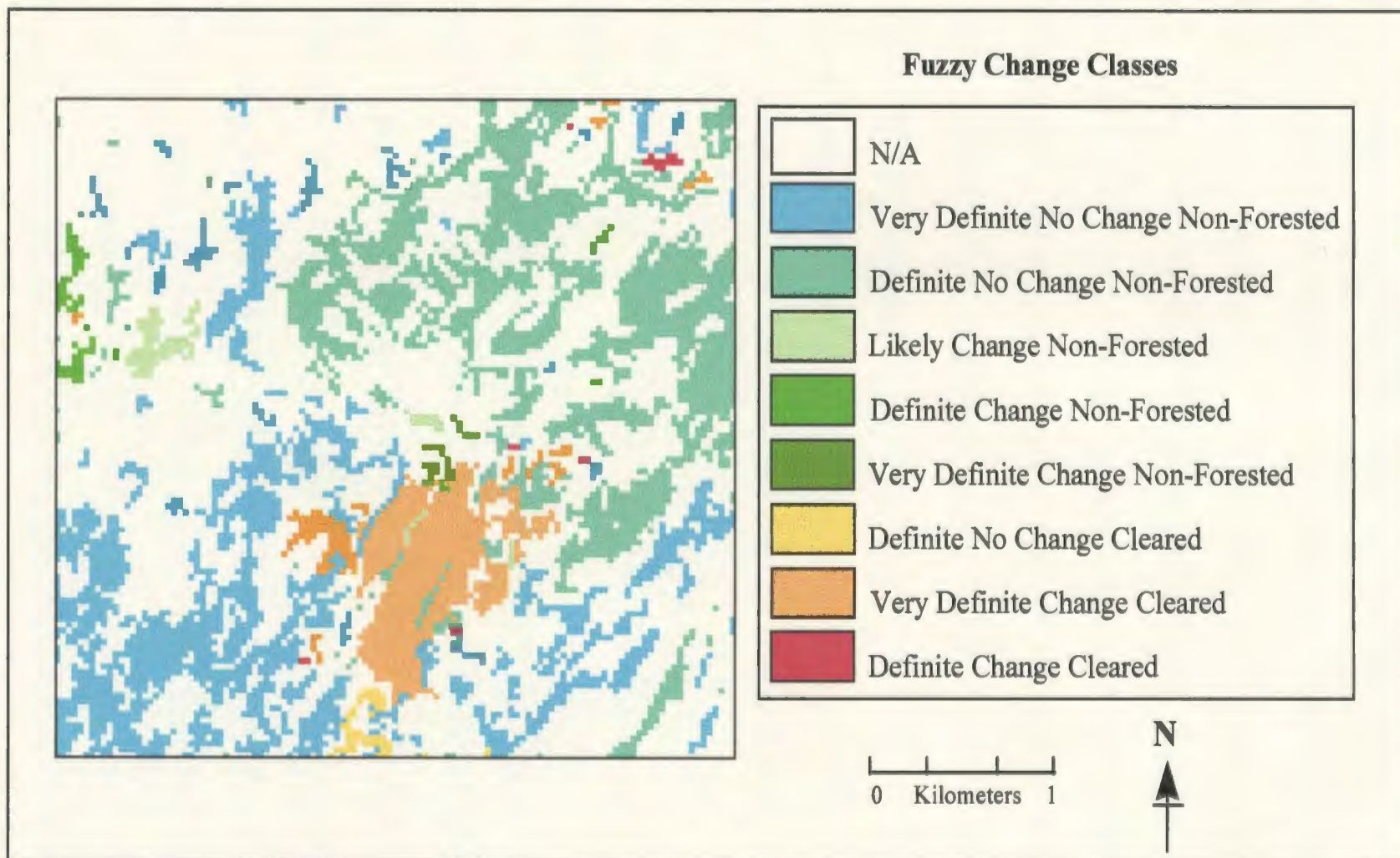


Figure 4.5.4 Fuzzy Change Classes for Cleared and Non-Forested

An important note must be stated about the results from the above section. Recall that the fuzzy agreement maps were reclassified from the membership functions for the purpose of comparing the local matching results to the boolean land use comparison maps. For projects that may be developed for land use change prediction or other forecasting purposes, land use agreement should be displayed as a continuous range of change possibilities. In Figure 4.5.5, for example, the visualization of the change possibilities for cleared and nonforested is based on a gradation in the intensity and hue of the color for the classes. The possibility of change for either class is highest for the darker polygons and decreases as the color lightens. By representing change as a range of possibilities, Figure 4.5.5 still contains the intermediate change information for dataset three but avoids the complications and limitations of reclassifying the matching results into discrete agreement classes.

## 4.6 Conclusion

Fuzzy hierarchical pattern matching was successfully adapted to the analysis of thematic differences between land use maps. The ability of fuzzy membership functions to address the uncertainty concerns of spatial data shows that fuzzy change detection is a viable and flexible alternative to the traditional boolean map comparison procedures.

The results from this chapter provide evidence that a fuzzy map comparison analysis can identify and preserve various types and magnitudes of land use differences between maps. Consequently, land use agreement possibility maps were produced for each dataset. The transitional progression from total agreement to total disagreement enables a



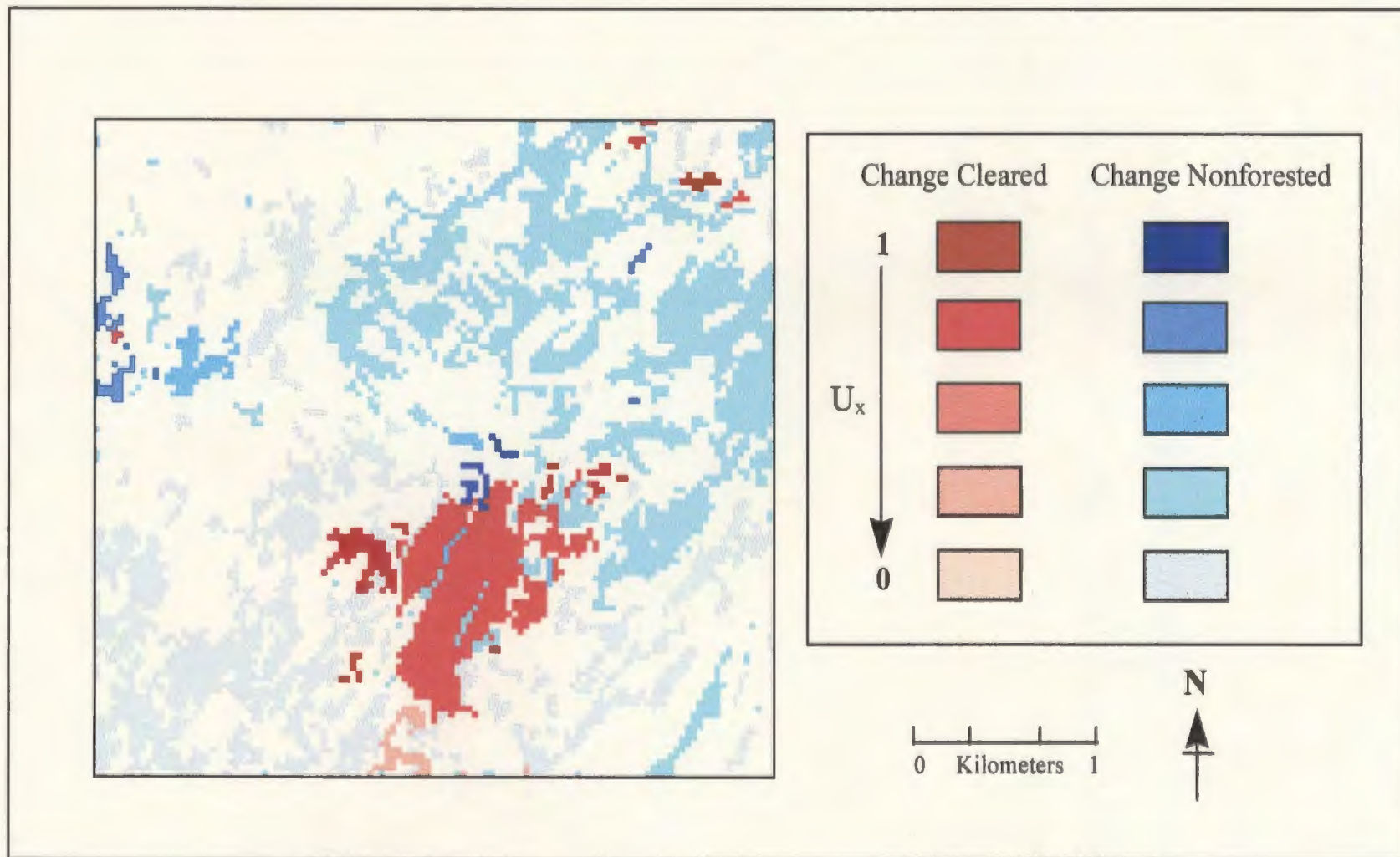


Figure 4.5.5 Fuzzy Change Possibilities for Cleared and Non-Forested

possibility map to present a more realistic interpretation of land use differences to the viewer. Unlike the boolean approaches, the fuzzy model does not lose relevant land use similarity information.

The global matching procedure was found to outperform several traditional similarity statistics. Also, the global values supported the linguistic interpretations of land use change or map similarities for each dataset. Certainties for the global matchings provide a measure of confidence that these values represent the actual overall pattern of land use difference between two land use maps.



## Chapter 5

### Discussion and Conclusion

#### 5.1 Discussion of Results

Historically, the comparison of thematic maps has been the basis for many land use change detection procedures. Traditional pixel-by-pixel comparison map comparison techniques are suspect because of possible map registration and error propagation problems. In addition, these boolean similarity operations can not adequately account for the uncertainty and complexity inherent in spatial information. A fuzzy regional polygon-by-polygon comparison methodology mitigates these difficulties or problems.

The basis of hierarchical fuzzy pattern matching is the emulation of the human reasoning method of expressing the perceived degree of similarity between two maps as a linguistic statement. This qualitative estimation of map agreement is the underlying factor in the development of the fuzzy inference system and the measurement of the performance of the matching process. The parameters of the fuzzy inference system that are selected by the developer are those that produce results that best approximate the linguistic agreement statement. For the datasets used in this thesis, a qualitative interpretation of map agreement from a visual comparison of maps is fairly straightforward because the maps in the datasets are small and contain only a few land use classes. For example, an analyst sees that the transportation systems are identical in the maps of Figure 4.3.1 and expects the local matching results to preserve this observation.

Theoretically, the same fuzzy pattern matching methodology can be applied to larger maps and remote sensing images consisting of many thematic classes. A potential problem that could arise is that a person's ability to linguistically describe the agreement between maps can decrease as the complexity of the maps increase. The visual comparison of two maps comprised of many small irregular polygons can become a complicated task, especially at the local matching level. Without a good mental evaluation of the agreement between the maps, an analyst is constrained in monitoring and fine tuning the fuzzy inference system to produce reliable matching results. Thus, the validation of the matching procedure is based less on expert knowledge and more on the Lowess regression and matches and mismatches analysis. At this point in the research, the effect that the use of large intricate layers will have on the matching process is unknown, but it is a factor that needs to be addressed in future manifestations of the model.

In its present form as a fuzzy areal comparison technique, hierarchical fuzzy pattern matching is a viable alternative to the boolean map comparison procedures. However, an important fact to consider is that the purpose of the local matching process can be altered by expanding the fuzzy model beyond the areal comparison of maps. The fuzzy inference system can be restructured to include membership functions for the matching of complex polygonal properties, such as elongation and fractal dimension. These additional variables could aid in the explanation and description of the differences between maps. For example, an increase in fractal dimension from one year to the next may be the result of an increase in the complexity of the land use pattern of a region due to urban expansion. Therefore, the general benefit of the fuzzy pattern matching model is

that its flexibility could enable the researcher to supplement his knowledge and understanding of the processes of land use change. Therefore, the acceptance of hierarchical fuzzy pattern matching as a practical map comparison procedure should take into account that it is applicable to a wider range of research than what was presented in this thesis.

## 5.2 Conclusions

Results from the analysis reported in this thesis support the following conclusions:

- (i) fuzzy land use change possibility maps provide a better interpretation of the land use agreement characteristics of a dataset than do boolean maps. The transitional change categories portrayed on a fuzzy map contain more change information and better represent the complex and intermediate change conditions. In addition, fuzzy maps give a better visual representation of where change has occurred spatially by retaining the form of the template layers. For example, a researcher can identify the railway, roads, and river as the areas of very definite no change on Map A of Figure 4.5.1, but these land use features are indistinguishable from the boolean change patterns on Map B.
- (ii) the pixel scale membership functions in the fuzzy inference system make the fuzzy land use pattern matching sensitive to scale. Compared to the boolean results, the fuzzy model identified fewer one pixel disagreements as the resolution of the datasets increased from 250 to 30 meters.
- (iii) the fuzzy local polygon-by-polygon land use comparison is less affected by possible map registration problems because the fuzzy inference system indirectly fuzzifies the boundaries of the polygons.
- (iv) the global matching results outperform a number of commonly used overall similarity statistics. By accurately accounting for random agreement, fuzzy global matching can be seen as a reliable method for combining local matching results into a single measure of two map similarities.
- (v) From the overall results of this research, it can be concluded that hierarchical fuzzy pattern matching can be successfully employed to detect both map similarities and land use change in thematic land use layers.

### 5.3 Future Research

The purpose of hierarchical fuzzy land use pattern matching is to determine the degree of similarity between two maps of the same scale. Theoretically, the comparison procedure could be performed on maps with slightly different scales (e.g. a 1:50000 land use reference map and a 30 meter classified Landsat TM image) by fuzzifying the boundaries of the polygons. A consideration for further research could involve estimating the range of differences in scale that maps can have before the matching results are unacceptable. By verifying the flexibility of the model, a researcher would get an indication of what spatial information can be used for a comparison procedure.

A concern with the development of the fuzzy inference system is whether the chosen inference rule and defuzzification algorithm are the best ones to use for the pattern matching procedure. Jang (1997) suggested the use of Max-Min inference and centroid of area defuzzification because of their ease of understanding and implementation. Although both are widely used, they are not easily subjected to mathematical analysis (Klir,1988). Greater mathematical tractability and the optimum inference structure for the model could be achieved from different combinations of inference rules and defuzzification procedures. The best combination would be the one that produces the highest certainty value for the global matching.

Research is required into the implementation of optimization techniques to obtain the best structure for the fuzzy inference system. It is possible that the local matching results are inaccurate because the shape of the membership curves and the amount of overlap between the function are less than optimal. Currently, there is no practical way of

ensuring that the membership functions provide a complete and consistent coverage of the input and output domains. A solution is to replace the fuzzy inference system with an adaptive neuro-fuzzy inference system (ANFIS), which is an adaptive network that is functionally equivalent to a fuzzy inference system. The connective updating capabilities of the ANFIS enable a system to automatically shape its membership functions by training them with input data. Technically, an ANFIS uses backpropagation, a common neural network training algorithm, for bi-directional learning to fine tune the fuzzy inference system by reducing an error measurement between the predicted results and the output of the training set. This procedure progressively reduces the disagreement between the predicted output and the desired output given by the training set. This process continues until the training error reaches a predetermined learning error threshold (Jang *et al.*, 1997). The local matching results from an ANFIS with an optimized input-output structure may be better representations of the patterns of agreement between land use maps.

In its present form, the hierarchical fuzzy pattern matching model is constrained by its concentration on the uncertainty problem of comparing two maps. An important factor that needs to be addressed in future research is the spatial dependency that exists between the land use maps. This would involve calculating measurements of the spatial autocorrelation between the template and comparison polygons and representing these values as membership functions in the fuzzy inference system. The inclusion of spatial autocorrelation into the fuzzy areal map comparison could expand the similarity analysis beyond the direct comparison of polygons to a comparison of the surroundings of the template polygons. This would be similar to a remote sensing analysis of texture or

context on multitemporal images. By enabling the model to be sensitive to spatial dependencies, the map comparison could be performed on highly segmented and fragmented land use patterns that are comprised of a complex intermixture of unique polygons.

The work presented in this study was based on the analysis of land use similarities between two maps. Ideally, the model needs to be expanded to deal with more complex problems, such as the comparison of a sequence of maps for a time series analysis. A fuzzy time series comparison would have to be a "one step ahead" process of matching successive maps to the template layer. This would require a multiple input-one output fuzzy inference system, where several of the inputs are the results from previous matching sessions. Consider the steps required for a change detection analysis involving three maps (a template and two later maps). Initially, the fuzzy inference system performs a comparison between the template and the map at time  $t+1$ . As the matching sequence moves ahead, the local matching results from the first comparison are added as an input variable to the fuzzy inference system before the map at time  $t+2$  is matched to the template. This is essential because the previous matching results are expected to have a moderating effect on the time series analysis since they should not be significantly different from the preceding findings (Bintley, 1987). An important point to consider is that by adding input variables to the fuzzy inference system the dimensionality of the system is increased as additional membership functions and rules are added. The increased complexity of the network can make the system structure difficult to interpret and understand and can increase the computational intensity of the forecasting analysis. It may

be necessary to use an ANFIS for the time series analysis to ensure that membership functions and rulebase are optimized after each new input is added to the system.

Despite the possible changes resulting from future research, the structure of the hierarchical fuzzy pattern matching model discussed in this thesis presents the researcher with a flexible way of handling spatial uncertainty in a map comparison procedure. Even in its present form, hierarchical fuzzy pattern matching produces results that warrant its consideration as a sound and plausible model for both similarity analysis and land use change detection.

## 6.0 References

- Adeniyi, Peter O. (1980) Land Use Change Analysis Using Sequential Aerial Photography and Computer Techniques, Photogrammetric Engineering and Remote Sensing, Vol. 46, No. 11, pp. 1417-1464
- Allison, N.M. (1992) Self Organizing Neural Maps and Their Applications, in Theory and Applications of Neural Networks, edited by J.G. Taylor and C. Mannion, Springer-Verlag Publishing, pp. 101-121
- Bandemer, H. (1987) From Fuzzy Data to Functional Relationships, Mathematical Modeling, Vol. 9, No. 6., pp. 419-426
- Barringer, Thomas H., Robert G. Reiser, and Curtis V. Price (1994) Potential Effects of Development on Flow Characteristics of Two New Jersey Streams, Water Resources Bulletin, Vol. 30, No. 2, pp. 283-295
- Bintley, H. (1987) Times Series Analysis with REVEAL, Fuzzy Sets and Systems, Vol. 23, pp. 97-118
- Blonda, P., G. Pasquarello, S. Losito, A. Mori, F. Posa, and D. Ragno (1991) An Experiment for the Integration of Multitemporal Remotely Sensed Images Based on a Fuzzy Logic Approach, International Journal of Remote Sensing, Vol.12, No.3, pp. 463-476
- Bonham Carter, Graeme (1994) Geographic Information Systems for Geoscientists: modelling with GIS, Pergamon Press, pp. 235-238
- Buckles, B.P., and F.E. Petry (1982) A Fuzzy Representation of Data for Relational Databases, Fuzzy Sets and Systems, Vol. 7, pp. 213-226
- Burrough, A. (1989) Fuzzy Mathematical Methods for Soil Survey and Land Evaluation, Journal of Soil Science, Vol.40, pp. 477-492
- Byrne, G.F., P.F. Crapper, and K.K. Mayo (1980) Monitoring Land-Cover Change by Principal Component Analysis of Multitemporal Landsat Data, Remote Sensing of Environment, Vol. 10, pp. 175-184
- Cannon, R.L., V.D. Jitenda, J.C. Bezdek, and M.M. Trivedi (1986) Segmentation of a Thematic Mapper Image Using the Fuzzy c-Means Clustering Algorithm, IEEE Transactions on Geoscience and Remote Sensing, Vol. GE24, No. 3, pp. 400-408



- Cleveland, William (1979) Robust Locally Weighted Regression and Smoothing Scatterplots, Journal of the American Statistical Association, Vol. 74, No. 368, pp. 829-836
- Congalton, R.G., R. Oderwald, and R.A. Mead (1983) Assessing Landsat Classification Accuracy using Discrete Multivariate Analysis Statistical Techniques, Photogrammetric Engineering and Remote Sensing, Vol. 49, No. 12, pp. 1671-1678
- Congalton, Russell (1991) A Review of Assessing the Accuracy of Classification of Remotely Sensed Data, Remote Sensing of the Environment, Vol. 35, pp. 35-46
- Cracknell, A.P., and W.B. Hayes (1991) Introduction to Remote Sensing, Taylor and Francis Ltd., pp. 224-227
- Davis, John (1990) Statistics and Data Analysis in Geology, John Wiley and Sons, pp. 390-410
- DeLuca, Aldo, and Settino Termini (1979) Entropy and Energy Measures of a Fuzzy Set, in Advances in Fuzzy Set Theory and Applications, Ed. M.M. Gupta, R.K. Ragade, and R.R. Yager, North-Holland Publishing, pp. 321 - 337
- Dombi, J. (1990) Membership Function as an Evaluation, Fuzzy Sets and Systems, Vol. 35, pp. 1 - 21
- Dubois, Didier and Prade, Henri (1979) Advances in Fuzzy Set Theory and Applications, ed. by M.M. Gupta, R.R. Rogade, and R.R. Yager, North Holland Publishing, New York, pp. 27-48
- Eastman, Ronald J. (1992) Idrisi Technical Reference Manual, Clark University Publishing, pp. 77-80
- Edwards, G., and K. Lowell (1996) Modeling Uncertainty in Photointerpreted Boundaries, Photogrammetric Engineering and Remote Sensing, Vol.62, No.4, pp.337-391
- Fedrizzi, Mario, and Janusz Kacprzyk (1986) Essentials of Fuzzy Sets Theory: A Linguistic Perspective, in Fuzzy Sets Applications, Methodological Approaches, and Results, ed. S. Bocklisch, S. Orlovski, M. Peschel, and Y. Nishiwaki, Academic-Verlag, Berlin, pp. 56-63
- Foody, Giles M. (1992) On The Compensation for Chance Agreement in Image Classification Accuracy Assessment, Photogrammetric Engineering and Remote Sensing, Vol. 58, No.10, pp. 1459 - 1460

- Gale, S. (1972) Inexactness, Fuzzy Sets, and the Foundations of Behavioral Geography, Geographical Analysis, Vol. 4, No. 4, pp. 337-349
- Gong, P. (1993) Change Detection Using Principal Components Analysis and Fuzzy Sets Theory, Canadian Journal of Remote Sensing, Vol.19, No.1, pp.22-29
- Goodchild, Michael F. (1988) A Spatial Analytical Perspective on GIS, International Journal of Geographical Information Systems, Vol. 1, pp. 327-334
- Goodchild, Michael F. (1998) Personnel Communication
- Gopal, Sucharita, and Curtis Woodcock (1994) Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets, Photogrammetric Engineering and Remote Sensing, Vol. 60, No.2, pp.181-188
- Hirota, K., and W. Pedrycz (1991) Matching Fuzzy Quantities, I.E.E.E. Transactions on Systems, Man, and Cybernetics, Vol. 21, No. 6, pp. 1580-1586
- Hodgson, Michael E., John R. Jensen, Halkard E. Mackey, and Malcolm C. Coulter (1988) Monitoring Wood Stork Foraging Habitat Using Remote Sensing and Geographic Information Systems, Photogrammetric Engineering and Remote Sensing, Vol. 54, No.11, pp.1601-1607
- Ingebritsen, S.E., and R.J.P. Lyon (1985) Principal Components Analysis of Multitemporal Image Pairs, International Journal of Remote Sensing, Vol. 6, No. 5, pp. 687-696
- Idrisi GIS (1992), Clark University
- Jager, Rene (1995) Fuzzy Logic in Control, Delft University of Technology Publishing, Delft The Netherlands, pp. 44 - 147
- Jang, J.S.R., C.T. Sun, and E. Mizutani (1997) Neuro-Fuzzy and Soft Computing: An Computational Approach to Learning and Machine Intelligence, Prentice Hall, Upper Saddle River New Jersey, pp. 73 - 93
- Jensen, John R. (1981) Urban Change Detection Mapping Using Landsat Digital Data, The American Cartographer, Vol. 8, No. 2, pp. 127-147
- Jensen, John R., Elijah W. Ramsey, Halkard E. Mackey, Eric J. Christensen, and Rebecca R. Sharitz (1987) Inland Wetland Change Detection using Aircraft MSS Data, Photogrammetric Engineering and Remote Sensing, Vol. 53, No.5, pp.521-529

- Johnson, R.D. (1994) Change Vector Analysis for Disaster Assessment: A Case Study of Hurricane Andrew, Geocarta International, Vol. 1, pp. 41-45
- Kandel, Abraham (1982) Fuzzy Techniques in Pattern Recognition, John Wiley and Sons, New York, pp. 22-52
- Key, J.R., J.A. Maslanek, and R. Barry (1989) Cloud Classification from Satellite Data using a Fuzzy Sets Algorithm: A Polar Example, International Journal of Remote Sensing, Vol.10, No.12, pp.1823-1842
- Klecka, W.R. (1980) Discriminant Analysis, SAGE Publication Inc., Beverly Hills, California, pp. 71-72
- Klir, George (1988) Fuzzy Sets, Uncertainty, and Information, Prentice Hall, New Jersey, pp. 2-50
- Kollias, V.J., and A. Voliotis (1991) Fuzzy Reasoning in the Development of Geographical Information Systems FRIS: a Prototype Soil Information System with Fuzzy Retrieval Capabilities, International Journal of Geographical Information Systems, Vol. 5, No. 2, pp. 209-223
- Kosko, Bart (1992) Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence, Prentice Hall, Englewood Cliffs, New Jersey, pp. 20 - 83
- Lamblin, Eric (1996) Land Cover Changes in Africa: Multitemporal Change Vector Analysis at Coarse Scale and Change Processes Categorization with High Spatial Resolution Data, Proceedings of the Second Vegetation Preparatory Programme, Stockholm, Sweden, Nov. 18-20, pp. 1-3
- Lo, C.P., and Robert L. Shipman (1990) A GIS Approach to Land Use Change Dynamics Detection, Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 11, pp. 1483-1491
- Ma, Zhenkui, and R.H. Redmond (1995) Tau Coefficients for Accuracy Assessment of Classification of Remote Sensing Data, Photogrammetric Engineering and Remote Sensing, Vol. 61, No. 4, pp. 435 - 439
- MacMillan, W.D. (1978) An Introduction to the Theory of Fuzzy Sets in the Context of the Construction of Representational Spatial Economic Theory, in Towards the

- Dynamic Analysis of Spatial Systems, Ed. By Martin, R.L., N.J. Thrift, and R.J. Bennett, Pion Limited, London, pp. 36-52
- Malila, W.A. (1980) Change Vector Analysis: An Approach for Detecting Forest Changes with Landsat, 1980 Machine Proceedings of Remotely Sensed Data Symposium, pp. 326-336
- Mamdani, E.H. (1976) Advances in the Linguistic Synthesis of Fuzzy Logic Controllers, International Journal of Man-Machine Studies, Vol.8, pp. 669 - 679
- Maselli, F., A. Rodolph, and C. Conese (1996) Fuzzy Classification of Spatially Degraded Thematic Mapper Data for the Estimation of Sub-Pixel Components, International Journal of Remote Sensing, Vol. 17, No. 3, pp. 537-551
- Matlab Fuzzy Logic Toolbox Software (1994) Mathworks Inc., pp. 5-40
- McBratney, A.B., and A.W. Moore (1985) Application of Fuzzy Sets to Climatic Classification, Agricultural and Forest Meteorology, Vol. 35, pp. 165-185
- McBratney, A.B., and J. Degrujter (1992) A Continuum Approach to Soil Classification by Modified Fuzzy k-Means with Extragrades, Journal of Soil Science, Vol. 43, pp. 159-175
- Meades, W.J., and L. Moores (1989) Forest Site Classification Manual: A Field Guide to the Damman Forest Types of Newfoundland, Forestry Canada, pp. 1.1-3.6
- Nguyen, Hung T., and Elbert A. Walker (1997) A First Course in Fuzzy Logic, CRC Press, pp. 21-60
- Openshaw, Stan (1987) Developing Appropriate Spatial Analysis Methods for GIS, in Geographical Information Systems: Principles and Applications, ed. David J. Maguire, Michael F. Goodchild, and David W. Rhind, Wiley and Sons, Essex, England, pp.389-402
- Pedrycz, Witold (1992) Associations of Fuzzy Sets, I.E.E.E. Transactions on Systems, Man, and Cybernetics, Vol. 22, No. 8, pp. 1438-1488
- Pedrycz, Witold, and Eugene Roventa (1994) A Hierarchical Neural Model of Matching, Information Sciences, Vol. 78, pp. 215-227
- Passonneau, J., and R. Wurman (1966) Urban Atlas: 20 American Cities, Cambridge MIT Press

- Renez, A.N. (1985) Multitemporal Analysis of Landsat Imagery for Monitoring Forest Cutovers in Nova Scotia, Canadian Journal of Remote Sensing, Vol. 11, No. 2, pp. 188-194
- Robinson, V. (1988) Some Implications of Fuzzy Set Theory Applied to Geographic Databases, Computers, Environment, and Urban Systems, Vol. 12, pp. 89-97
- Rosenfield, G.H., and K. Fitzpatrick-Lins (1986) Coefficient of Agreement as a Measure of Thematic Classification Accuracy, Photogrammetric Engineering and Remote Sensing, Vol. 48, No. 1, pp. 131 -137
- Simpson, J.J., and R.H. Keller (1995) An Improved Fuzzy Logic Segmentation of Sea Ice, Clouds, and Ocean in Remotely Sensed Arctic Imagery, Remote Sensing of Environment, Vol. 54, pp. 290 - 315
- Singh, Ashbindu (1989) Digital Change Detection using Remote-Sensing Data, International Journal of Remote Sensing, Vol. 10, No. 6, pp. 989-1003
- SPSS for Windows (1995) Version 6.1.3, SPSS Inc.
- Sui, Daniel (1992) A Fuzzy GIS Modeling Approach for Urban Land Evaluation, Computers, Environment, and Urban Systems, Vol.16, pp.101-115
- Takagi, Tomohiro, and Michio Sugeno (1985) Fuzzy Identification of Systeem and Its Application to Modeling and Control, I.E.E.E. Transactions on Systems, Man, and Cybernetics, Vol. SMC15, No.1, pp. 116-132
- Tanaka, H, S. Uesima, and K. Asai (1982) Linear Regression Analysis with a Fuzzy Model, I.E.E.E Transactions on Systems, Man, and Cybernetics, Vol. 12, pp. 903-907
- Taylor, Peter (1977) Quantitative Methods in Geography, Houghton Mifflin Company, Boston, pp. 177-179
- Veregin, H. (1989) Error Modeling for the Map Overlay Operation, in The Accuracy of Spatial Databases, ed. M. Goodchild and S. Gopal, Taylor Francis, pp. 3-18
- Wang, F. (1990) Improving Remote Sensing Image Analysis Through Fuzzy Information Representation, Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 8, pp. 1163-1169
- Wang, F., G.B. Hall, and P. Subaryano (1990) Fuzzy Information Representation and Processing in Conventional GIS Software: Database Design and Application,

International Journal of Geographical Information Systems, Vol. 4, No. 3, pp. 261-283

Wickware, Gregory M., and Howarth, Philip J. (1981) Procedures for Change Detection using Landsat Digital Data, International Journal of Remote Sensing, Vol. 2, No.3, pp. 277-291

Weismiller, R.A., S.J. Kristof, D.K. Scholz, P.E. Anuta, and S.A. Momin (1977) Change Detection in Coastal Zone Environments, Photogrammetric Engineering and Remote Sensing, Vol. 43, No. 12, pp. 1533-1539

White, Roger, Guy Engelen, and Inge Injee (in press) The Use of Constrained Cellular Automata for High-Resolution Modelling of Urban Land Use Dynamics, Environment and Planning B

Zadeh, Lotfi (1965) Fuzzy Sets, Information and Control, Vol. 8, pp. 338-353

Zadeh, Lotfi A. (1987) Fuzzy Sets as a Basis for a Theory of Possibility, in Fuzzy Sets and Applications: Selected Papers by L.A. Zadeh , ed. by R.R. Yager, S. Ovchinnikov, R.M. Tong, and H.T. Nguyen, John Wiley and Sons, New York, pp. 193-213

Zimmerman, Hans (1985) Fuzzy Set Theory and its Applications, Kluwer Nijhoff Publishing, Boston, pp. 1 - 150

## **Appendix A**

### **Unique Polygons Mapping and Fuzzy Inference System Programs**

**Program 1: Uniqpoly.bas**

- This program identifies the unique intersections between two land use maps

Dim g1, g2, grpmap1, grpmap2, pixels, id as Integer

Open "Landmaps.txt" For Input as #1

- Landmaps.txt is a file that contains the first map in column 1 and the second map in column 2
- Note: Idrisi ascii images are run length encoded files that store the attributes for the pixels in a single column that is (# of rows \* # of columns) in length
- 4 Steps to create Landmaps.txt
- 1) ensure that the Idrisi maps are in integer ascii format; use CONVERT module if necessary
- 2) import the Idrisi maps into a spreadsheet and append the maps into adjoining columns in a new file
- 3) sort the file according to column 1 ascending then column 2 ascending
- 4) save as the file as an ascii text file called Landmaps.txt

Open "uniqinfo.dat" For Output as #2

Input "Enter the number of pixels in the Land Use Maps:", pixels

Input #1, g1, g2

id = 1

Print #2, id; g1; g2

For I = 1 to (pixels - 1)

Input #1, grpmap1, grpmap2

If g1 <> grpmap1 or g2 <> grpmap2 Then

id = id + 1

Print #2, id; grpmap1; grpmap2

g1 = grpmap1

g2 = grpmap2

End If

Next I

Close #1

Close #2

End



**Program 2 : Uniqmap.bas**

- this program creates an unique polygons map

Dim grpid1,grpid2,I,unq1,unq2,unqid,mapid, pixels as integer

Open "map1.img" for input as #1

Open "map2.img" for input as #2

Open "Uniqmap.img" for output as #3

- ensure that both maps are in integer ascii format

Input "Enter the number of pixels in the Land Use Maps:", pixels

For I = 1 to pixels

Input #1, grpid1

Input #2,grpid2

Open "uniqinfo.dat" for input as # 4 ' this is the polygonal information file from Uniqinfo.bas

While Not EOF(4)

Input #4, unqid,unq1,unq2

If unq1 = grpid1 and unq2 = grpid2 then

mapid = unqid

Endif

Wend

Close #4

Print #3, mapid

Next I

Close #1

Close #2

Close #3

End

**Program 3: Localfis.m**

- This program imports the areal information into the fuzzy pattern
- matching model and activates the local matching fuzzy inference
- system

```
load areainfo.dat
```

- areainfo.dat is an ascii text file that contains the areal intersection,
- areal complement, and pixel grouping size for each unique polygon of the
- template map

```
fid = fopen('localmap','w')
```

- localmap is the resultant local matching map

```
fismat = readfis('intcomp');
```

- intcomp is the local matching fuzzy inference file
- the script for intcomp is not listed because it is in byte
- binary format

```
num = length(areainfo);
for k = 1:num
    a = areainfo(k,:);
    output = evalfis(a,fismat)
```

- the evalfis function performs the local matching

```
fprintf(fid,'%6.3f\n',output);
end
status=fclose(fid);
```

