CHANGE DETECTION IN REMOTE SENSING USING SUPERVISED FUZZY CLASSIFICATION

Peter Deer
peter.deer@dsto.defence.gov.au

Department of Geography and Department of Computer Science
The University of Adelaide

SUBMITTED FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

1999

Abstract

Change detection is an important analytic function in pattern recognition and image processing. A variety of change detection approaches and techniques have been reported. Because it also identifies the nature of the change, the most appealing approach is to independently classify each image, and then compare the class labels. Unfortunately, this approach has shown relatively poor accuracy. This is, in part, because traditional classification approaches are themselves relatively inaccurate, and individual classification errors compound in the change detection process.

Traditional classification approaches seek to allocate each observation (e.g. pixel spectral values in an image) to one of a set of mutually exclusive classes, usually based on some maximum likelihood/least risk criterion. Unfortunately, classes commonly do not have distinct class 'boundaries'. Even human experts in the field will often disagree on class assignment e.g. when distinguishing between pasture and thinly wooded areas. Moreover, a single pixel in a remotely sensed image may contain a number of classes (the mixed pixel problem).

Fuzzy set theory was introduced to address issues such as class or set vagueness. Using fuzzy set theory, we can determine and reason with the grade of membership of a particular pixel in a number of classes. This provides an approach to partial class assignment in regions where there is a gradual transition from one class to another.

There is also support in the literature to its use in addressing the mixed pixel problem.

This thesis examines a fuzzy approach to Post Classification Comparison for change detection in digital remotely sensed imagery. Although this approach has wider application, the thesis focusses on its use for environmental monitoring.

A supervised fuzzy classifier was implemented by an adaptation of the unsupervised fuzzy c-means clustering algorithm. Class means and covariance matrices are determined from training data. Fuzzy class memberships are calculated using the normalised weighted reciprocals of the Mahalanobis Distances of each pixel from each training class mean. Two approaches to comparing fuzzy class labels are examined: an arithmetic approach, and the use of fuzzy logic operators. The fuzzy logic approach is shown to be superior. It is then compared with the Boolean logic approach of traditional Post Classification Comparison, with highly favourable results.

The Mahalanobis Distance fuzzy classifier requires the *a priori* selection of a fuzzy weighting parameter. Previous work has provided only limited guidance on this matter (usually in the context of the unsupervised fuzzy c-means classifier). Suitable values for the parameter are investigated, both empirically and analytically, under the requirement that the fuzzy memberships reflect the proportions of class representation in each pixel. It is concluded that it is only possible to satisfy this requirement precisely, for all mixture proportions, in a special case. For the general case, the sensitivity of the fuzzy classifier to the selection of this parameter is investigated. This work also suggests an interpretation of the physical significance of the suggested range of suitable values.

It has been suggested that fuzzy classifications should be based on class posterior probabilities. It is shown that this approach gives poor results, producing a very 'hard' classification. Variants which produce more 'fuzzy' values are examined, but

these show a close correspondence to the Mahalanobis Distance approach, and offer no apparent advantage.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I acknowledge the efforts, support and contributions of my academic supervisors: Prof. Peter Eklund, Dr. Stephen Kirkby and Dr. Maureen Longmore. Maureen was my sole supervisor when I commenced this research, part-time, over 6 years ago. She took on the task knowing that, with our disparate backgrounds, interests and attitudes, it would not be easy. I just wanted to understand and develop algorithms: she quite rightly insisted that I develop and maintain a focus on applications. As the work progressed, Peter Eklund expressed an interest, and became joint supervisor. The breadth and depth of his interests amazed and guided me. When the end was in sight (at least, I thought it was at the time!), Maureen's interests took her elsewhere. I was fortunate in getting Steve Kirkby as a replacement joint supervisor. His enthusiasm and work ethic were an inspiration. Without his cry of "Focus, man!" ringing continually in my ears, I would still be working on "just one more thing", and this thesis would probably never have been completed.

One hears numerous horror stories about relationships between research students and their supervisors. Mine were highly pleasurable (to me at least) and without rancour: I count my supervisors amongst my friends.

I undertook this research part-time, while in full-time employment with the

Defence Science and Technology Organisation (DSTO). This would not have been possible without the support of a number of people in DSTO management, most notably: Dr. Ian Chessell, Dr. Stephen Hood, Mr. Eric Youle, Mr. Peter Calder and Dr. Lakshmi Narasimhan. I thank them.

Some of the work described herein was undertaken in collaboration with fellow students Mr. Brenton Norman and Ms. Katrina Opperman. Some also benefitted from discussions with work colleagues Mr. Greg Chase, Dr. Mark Nelson and Dr. Andre Yakovleff. Comments by Dr. Norm Campbell, CSIRO, led to a considerable portion of the work described in Chapter 8. Although it did not find its way into this thesis, I was also fortunate to collaborate on related work with Dr. Salim Bouzerdoum, Mr. Richard Cole, and Dr. Garry Newsam.

The environment and facilities of the Geographic Information Systems Cooperative Adelaide (GISCA), and the support and assistance of its Director Professor Graeme Hugo, Manager Mr. Errol Bamford, and staff members Messrs. Simon Pollitt and Simon Jacobs, were greatly appreciated. The final thesis writing was undertaken at Griffith University: I thank Professor Eklund for making various facilities available to me, and for both insisting that I change my thesis to LaTeX, and for his considerable assistance in the conversion process.

It would be an understatement to suggest that undertaking a PhD part-time, while satisfying the requirements of full-time employment, requires sacrifices from one's family. To my sons, Randall and Michael, I apologise for any inconvenience and embarrassment I may have caused as a fellow university student. To my wife Trudy, thank you for your support and encouragement over the years.

Declaration

I hereby declare that none of the material contained in this thesis has been accepted for the award of any other degree or diploma in any institution and that, to the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except where due reference has been made in the text of the thesis.

I consent to this thesis being made available for photocopying and loan, if applicable, and if it is accepted for the award of the degree.

Peter Deer

Paradise Point, Queensland

February 1999

Chapter 1

INTRODUCTION

1.1 General

It is now relatively commonplace for satellites, travelling at kilometres per second, at altitudes of hundreds of kilometres, to capture and transmit images of the Earth. These images represent areas hundreds or thousands of kilometres square. It is also possible to capture imagery from airborne platforms, potentially using hundreds of spectral bands. Image capture traditionally uses a variety of naturally occurring radiation sources (such as the Sun, or 'black body' temperature emissions from an object). We are, however, no longer constrained to naturally occurring radiation sources: we can illuminate the object with radar, and image at night and through cloud, light rain, haze and sometimes smoke. We can gain further advantage by controlling the radiation source, and synthesising antenna lengths to achieve much higher spatial resolutions than the physical laws would seem to permit.

A noteworthy characteristic of many of these images is that they are captured in digital form. It is therefore possible to process them with computers. This considerably changes both the nature and quantity of the processing that can be undertaken. Indeed, it can be argued that it is a very limiting viewpoint to consider

these data as an 'image'. Rather, it is a digital data file of numbers representing the strength of upwelling radiation in whatever frequency bands we desire, arranged in a well ordered spatial structure, containing much information about the nature of objects, that can be viewed as an image by a human analyst if so desired. It is noted that the budget of a nation can be viewed as a pie-chart, and the performance of the stock market can be viewed as a graph. These are both valid and useful views of the data, but they are limited views. There is much more underlying information than can be represented in these forms. This is also the case with digital imagery. The human eye-brain has many remarkable qualities and attributes (most particularly, in this context, its ability to discern patterns and shapes in spatial data), but it has a number of limitations (such as an inability to input more than three spectral bands, to process numerically, or to make adequate use of a large range of brightness values).

A major use of remotely sensed imagery is for environmental monitoring. Studies with this aim will generally either seek to understand the image content, or to determine those images and areas in which change has occurred over some time interval. The latter method goes under the rubric of digital change detection, and there exists a solid body of literature on its use for environmental applications in remote sensing. The former method is less successful and well established, with the reported research commonly bearing titles related to 'image understanding'. This thesis is concerned with the use of digital change detection techniques.

1.2 Change Detection

Change detection is the process of identifying differences in the state of an object or phenomenon by observing it at different times. The basic premise in using remotely sensed data for change detection is that changes in the object of interest will result in

changes in radiance values or local texture that are separable from changes caused by other factors, such as differences in atmospheric conditions, illumination and viewing angle, soil moisture, etc. It is also necessary that changes of interest be separable from expected or uninteresting events, such as seasonal, weather, tidal or diurnal effects.

Satellite remote sensing offers a potentially powerful method of monitoring changes in imagery at higher temporal resolution and lower costs than those associated with traditional methods (Martin, 1989). The attributes of remotely sensed satellite data include a synoptic view, a high frequency of revisitation, relative cheapness, and its digital nature (rendering machine assistance or analysis possible). Images can be taken of areas that are difficult or impossible to access by other means. The capabilities of different sensing technologies can be utilised, to some extent, to overcome problems of day/night observation, and obscuration by smoke, rain, cloud, and haze.

Clearly, the aspects of change that are of interest are:

1. has it occurred (in some specific time interval)? (detection);
2. where? (location and extent);
3. what change occurred? (identification); and
4. what are the causes and implications of this? (analysis)

The term change detection is variously and loosely applied in the literature. It invariably involves the first of these aspects, normally the second, and sometimes the third. Although the fourth aspect is normally left to the human analyst, it is noted that Dreschler-Fischer *et al.* (1993) attempts to capture this process in the knowledge rules of an expert system.

A number of researchers have used remotely sensed satellite data for change detection, and a variety of approaches and techniques have been developed. Because

it addresses all of the relevant aspects of change, the most appealing approach is to independently classify each image, and then compare the class labels. Unfortunately, this approach, commonly known as Post Classification Comparison, has shown relatively poor accuracy. This is, in part, because traditional classification approaches are themselves relatively inaccurate, and individual classification errors compound in the change detection process.

Traditional classification approaches seek to allocate each observation (e.g. pixel in an image) to one of a set of mutually exclusive classes, usually based on some maximum likelihood/least risk criterion. Unfortunately, classes commonly do not have distinct 'boundaries'. Even human experts in the field will often disagree on class assignment e.g. when distinguishing between 'stressed' and 'not-stressed' vegetation, or between pasture and thinly wooded areas. Moreover, a single pixel in a remotely sensed image may contain a number of classes (the so-called mixed pixel, or 'mixel' problem).

Fuzzy set theory was introduced to address issues such as class or set 'vagueness'. Using fuzzy set theory, we can determine and reason with the grade of membership of a particular pixel in a number of classes. This provides an approach to partial class assignment in regions where there is a gradual transition from one class to another. There is also support in the literature to its use in addressing the mixed pixel problem.

1.3 Motivation/Objective of this Thesis

The objective of this thesis is to examine the use of fuzzy classification fir detecting environmental changes in remotely sensed imagery.

1.4 Structure and Contributions of this Thesis

The literature on change detection is reviewed in Chapter 2. The major change detection techniques are described and compared. Chapter 2 also includes a discussion of a number of important associated issues: registration of imagery, radiometric correction, threshold value selection, and accuracy assessment.

The study area, data and image processing environment used are described in Chapter 3.

The common 'hard' classification approaches, both supervised and unsupervised, are examined and compared in Chapter 4. It is shown that, for this application at least, the Minimum Mahalanobis Distance Classifier produces a highly comparable result to the complete supervised Maximum Likelihood Classifier (under the multivariate Gaussian distribution assumption, and incorporating terms for global class prior probabilities and the determinant of the class covariance matrices). Chapter 4 also describes an approach to post-processing for 'noise' and error reduction: morphological image filtering.

Chapter 5 introduces fuzzy set theory and its notation, reviews the use of fuzzy sets in image processing and remote sensing, and discusses approaches to fuzzy classification. The approach adopted for this thesis (a supervised variant of the fuzzy c-means clustering algorithm) is described. The chapter includes a preliminary study, undertaken to investigate the feasibility of the proposed approach, and to gain insight into choices for one of its key parameters. Armed with this insight, the major images of interest are classified using the fuzzy approach. The accuracy assessment of fuzzy classifications is also discussed.

Chapter 6 addresses the key issue of this thesis: change detection using Post

Classification Comparison. First, it shows the results of using Post Classification Comparison with 'hard' classifications, observing that this approach can be formulated as a problem in logical querying. It then makes an original contribution in the area of change detection with fuzzy classifications. It discusses two approaches: differencing of fuzzy membership values between two dates, and change detection through fuzzy logical querying. The latter approach requires the development of theory drawn from the formal fuzzy logic literature, and is shown to be highly functional, flexible, simple to implement, computationally efficient, and superior to the traditional approach of 'hard' Post Classification Comparison.

The chosen fuzzy classification algorithm requires the selection of a key parameter: the 'fuzzy exponent' in the fuzzy c-means clustering algorithm. There have been a number of studies of this subject, with somewhat equivocal or inconclusive results. Chapter 7 makes a contribution to this literature by examining the selection of this parameter, both empirically and analytically, under the requirement that the fuzzy memberships reflect the proportion of class representation in each pixel. It is concluded that it is only possible to satisfy this requirement precisely, for all mixture proportions, in a special case. For the general case, the sensitivity of the fuzzy classifier to the selection of this parameter is investigated.

There has been a position advocated in the literature that fuzzy memberships, if used, should be based on the posterior probabilities determined by a Maximum Likelihood Classifier. This position is examined in Chapter 8. Through analysis of both the resultant imagery, and various measures of 'fuzziness', it is shown that this is not a satisfactory approach. Some variants of the posterior probability fuzzy classifier in which the degree of fuzziness is increased by either exponentiation or by direct scaling of the class covariance matrices are introduced and discussed. Their

relationship to each other, and to the Mahalanobis Distance Fuzzy Classifier, is examined both empirically and analytically.

Chapter 9 concludes the thesis, and contains a description of future work.

Chapter 2

DIGITAL CHANGE DETECTION TECHNIQUES

Summary: This chapter reviews the literature on change detection, and describes and compares the major change detection techniques. Some observations are made about the essential similarity of a number of techniques that appear, prima facie, to be quite different. The notion of pixel, feature and object level analysis is introduced. It is argued that an object level, or classification-based, approach to change detection is highly desirable. A number of issues highly relevant to change detection are also discussed: registration of imagery, radiometric correction, threshold value selection, and accuracy assessment.

2.1 Approaches to Digital Change Detection

Singh (1989) categorises digital change detection techniques into two basic approaches: the comparative analysis of independently produced thematic labellings or classifications of imagery from different dates; and the simultaneous analysis of multi-temporal datasets. Later in this chapter, we will consider a view, or categorisation, of the techniques in terms of whether they operate over: spectral

values (whether 'raw' or radiometrically corrected); features processed from these spectral values (e.g. vegetation indices); or object regions or classifications of the data, however derived.

Irrespective of how we choose to categorise approaches, there are a number of methods and techniques. The more common of these are reviewed in the following section.

2.2 Review of Techniques

2.2.1 Post Classification Comparison

Post Classification Comparison involves the classification of each image independently, followed by a comparison of the corresponding pixel (thematic) labels to identify areas where change has occurred. The process of classification itself can be undertaken in either a supervised or an unsupervised manner (these are discussed in Chapter 4). Post Classification Comparison is the most obvious method of detecting change. It not only reveals that change has occurred, and where; it also reveals the precise nature of the change (assuming, of course, that the classification process itself was sufficiently accurate and class-specific).

Comparison of the separately classified images can be carried out visually, or by a computer. Computers are considerably better at quantitative analysis, but humans possess a superior ability to discern patterns and shapes, and are generally less susceptible to error effects caused by misregistration of the images. If the comparison of images is to be carried out by computer, it may be assisted by the use of a geographic information system (GIS).

The major disadvantages of Post Classification Comparison are:

1. Classification techniques are relatively expensive (in both time and cost). There is normally a substantial requirement for 'ground truthing' to reduce uncertainty and error. Any subsequent classification and comparison is effectively constrained to the initial set of class labels, or supersets of them.
2. A classification process results in a class label being assigned to a pixel (provided some threshold condition of 'closeness' or 'typicality' is satisfied). The labels assigned are normally discrete, and information about the certainty of the assignment is normally not kept. A pixel may fall into one category on one date, and fall into a different category on a second date. A Post Classification Comparison will indicate a class change, but there may be, according to the probabilities associated with the individual classifications, only a small, statistically significant indication of change.
3. Errors in classification have a compounding effect (Pilon *et al.*, 1988; Quarmby and Cushnie, 1989). Consider images on two dates of the same area, with no change having occurred. Assume perfect image to image registration. Let the accuracy for each independent classification be 90% (a generous allowance). A pixel-wise comparison of the two classifications could reveal, in the worst case, that 19% of the scene has undergone change (1 - 0.9*0.9 = 0.19). We know that no change has occurred, yet this method could indicate that up to 19% of the scene has changed: this is an alarmingly high false detection rate! Clearly, this simple analysis assumes no correlation between the two sets of erroneous classifications, and is offered for illustrative purposes only. Notwithstanding this, Stow *et al.* (1980) found that a change map produced from two independent classifications exhibited similar accuracy to the product of multiplying the individual accuracies.

Riordan (1980) used unsupervised classifications of 1973 and 1978 Landsat MSS data to detect non-urban to urban change, and reported an accuracy of 67%. Joyce *et al.* (1980) concluded that Post Classification Comparison "appears suitable for detecting land cover change with Landsat MSS data in sites where large areas of forestland are being converted to cropland". Gordon (1980), however, used this method to monitor land use change in Ohio and, after a rigorous quantitative assessment, observed " ... we must conclude that substantial errors are associated with the use of Landsat data for land cover and change analysis". Fisher and Pathirana (1993) found that (conventional) Post Classification Comparison indicated that over 40% of the land area studied had undergone change, whereas it was known that little change had actually occurred.

Jensen *et al.* (1987) used Post Classification Comparison with unsupervised classification of aircraft MSS data for wetland change detection, but achieved limited success. Jensen *et al.* (1995) reports on a later study also using Post Classification Comparison for wetland change detection, and Hoffer and Lee (1989) used it with synthetic aperture radar data for forest change detection.

2.2.2 Direct Multi-Date Classification

This method (sometimes also referred to as spectral/temporal classification) uses a single analysis of a combined dataset of two or more dates to identify areas of change. To illustrate, in a study involving two dates using four band Landsat MSS imagery, a single dataset of eight bands is produced. It may then be classified in either a supervised or an unsupervised mode. In the supervised approach, training sets pertaining to change and no-change areas are used to derive statistics to define sub-spaces of the feature (normally spectral) space. In the unsupervised approach,

spectral classes are determined by cluster analysis, and subsequent inspection should reveal where changes have occurred. In either case, change classes are expected to display significantly different statistics from no-change classes. The principal disadvantage of this approach is that adequately sized training areas for each change class need to be accurately identified. A second disadvantage is that the data needs to be re-processed whenever a new image is added.

Weismiller *et al.* (1977) used a clustering technique combined with layered spectral/temporal classification to detect change in a Texas coastal area. Hoffer and Lee (1989) used Direct Multi-Date Classification for forest change detection, and achieved an accuracy of 90.6% for a full (12-band) image, and 90.4% with a reduced, 6-band, dataset. The data reduction was undertaken with a Principal Components technique (which is discussed in 2.2.7).

2.2.3 Image Differencing

In this technique, images of two different dates are spatially registered. The corresponding pixel values are subtracted to produce a new image that represents the change between the two. Pixels exhibiting a significant radiance change can be expected to lie in the tails of the distributions of the difference image, whereas the remaining pixels should be grouped about the mean. It is noted, however, that "the bias and variance ... are unknown and might be substantial" (Van Deusen, 1994). The technique may be applied to a single band (in which case it is called Uni-variate Image Differencing) or to multiple bands. Some form of radiometric standardisation is normally applied to reduce the effects of illumination angle, intensity (including path effects) and viewing angle.

Weismiller *et al.* (1977) found that change detection using Image Differencing

worked well in the Texas coastal zone environment (although many small areas of change were not identified accurately). Miller *et al.* (1978) applied Landsat Image Dfferencing successfully to the mapping of changes in tropical forest cover in northern Thailand. Williams and Stauffer (1978) employed this technique to monitor gypsy moth defoliation in the forests of Pennsylvania; Brera and Shahrokhi (1978) used it for the analysis of desertification in the Sahara; and Vogelmann (1988) investigated the detection of change in temperate forests. Singh (1989) mentions his earlier use (1986) of this technique for monitoring changes attributable to shifting cultivation in a tropical forest environment. Manavalan *et al.* (1995) report on its use for the analysis of irrigated crops.

2.2.4 Image Regression

Image Regression can be used to account for differences in the mean and variance between pixel values for different dates, thus reducing the effects of different atmospheric conditions and sun angle. Pixel values from one time are assumed to be a (normally) linear function of pixel values at some other time. The function that relates the pixel values can be determined e.g. using a least squares regression. Pixels that have changed between the two dates will exhibit values that differ significantly from that predicted by the determined regression function. Thresholding is applied to detect areas of change. Alternatively, the actual and predicted values are differenced (or ratioed), and, again, a threshold is applied. Reported examples of the use of this method are not common, but include Singh (1986), Hanaizumi *et al.* (1991) (using both regression and segmentation), and Jha and Unni (1994).

2.2.5 Image Ratioing

Image Ratioing calculates the ratio of the values of corresponding pixels of registered images of different dates, on a band by band basis. If no change has occurred, it is expected that the ratio of corresponding pixels, in corresponding bands, will be near unity. If change has occurred in a particular pixel, the ratio is expected to be either considerably more than or considerably less than unity (depending on the 'direction' of the change). Some data standardisation or radiometric correction between dates may be necessary. Todd (1977) used Image Ratioing to determine urban change in Atlanta, Georgia, and determined that 91.4% of land cover change was correctly identified. Weydahl (1991) states that ratioing is less sensitive than differencing to multiplicative noise in SAR imagery.

The main criticism of Image Ratioing centres on the statistical distributions of ratioed images (Robinson 1979). Because functions of the standard deviation are normally used to threshold change, non-symmetric distributions will result in unequal areas on either side of the mean. Therefore, the error rates on either side of the mean will not be equal. This is generally undesirable. Further work is also required on more robust thresholding proposals.

2.2.6 Vegetation Index Differencing

Digital spectral radiance values can be analysed independently on a band by band basis, or in combinations of two or more bands. One of the most commonly used band combination techniques involves the calculation of vegetation indices. These have been developed to enhance spectral features on the basis of the strong absorption of red and strong reflectance of near-IR by vegetation. They are most commonly either linear combinations of bands, or band ratios, or a combination of both. It

has been found that the ratio of near-IR to red reflectance is significantly correlated with green leaf biomass (Tucker, 1979). There are a number of vegetation indices in use (see e.g. Guoling, 1989; or Elvidge and Chen, 1995), but the most common is the normalised difference vegetation index (NDVI), given by (near-IR minus red reflectance)/(near-IR plus red reflectance).

It is considered that normalising or ratioing spectral bands negates the effect of any extraneous multiplicative factors in sensor data that act equally in all bands (Lillesand and Kiefer, 1987). It should be noted, however, that it may enhance noise that is not correlated in different bands (Singh 1989). Qi *et al.* (1993) advocate a correction to account for first-order soil background effects when using vegetation indices. In a recent paper on the use of vegetation indices for change detection, Lyon *et al.* (1998) found that NDVI produced superior results to those of other vegetation indices.

For change detection purposes, the difference between vegetation indices (or indeed, between any other indices) of two or more dates should give a reasonable indication of change in the vegetation canopy (or some other condition of interest). Coiner (1980) used vegetation index differencing to study desertification, but provided no accuracy assessment or comparison with other techniques. Nelson (1983) assessed this method quantitatively in the study of gypsy moth defoliation in Pennsylvania.

2.2.7 Principal Components Analysis

The principal components transformation (sometimes termed the *Hotelling* or discrete *Karhunen-Loeve* Transformation) is a linear transformation that defines a new, orthogonal coordinate system such that the data can be represented without correlation. Both the correlation and covariance matrices in the new co-ordinate

system are therefore diagonal.

The transformation that achieves this can be found from the eigenvectors of either the covariance or correlation matrices of the original data. The axes of the new, orthogonal co-ordinate system are defined by the eigenvectors. Each individual pixel is transformed by vector multiplication of its original vector and the eigenvectors, resulting in coordinates in the new space i.e. a new pixel vector. Each eigenvector can be thought of as defining a new band, and the coordinates of an individual pixel can be thought of as its brightness in that band. The amount of the total scene variance that is represented by each new band is given by the eigenvalue of the corresponding eigenvector.

Principal Components Analysis (PCA) can also be applied to image datasets comprising bands from two or more dates i.e. to multi-temporal image data (e.g. Lodwick, 1979; Byrne *et al.*, 1980; Richards, 1984; Ingebritsen and Lyon, 1985; Fung and LeDrew, 1987). There is a high correlation between image data for regions that have not changed significantly and a relatively low correlation between regions that have changed substantially. Provided the major portion of the variance in a multi-temporal image dataset is associated with constant cover types, regions of localised change will be enhanced in the higher numbered Principal Components (PC), while brightness changes between the two dates attributable to atmospheric conditions or sun angles can be expected to be represented in the lower PC (Richards, 1986).

It was stated earlier that PCA can be undertaken using either the covariance matrix or the correlation matrix, but it should be noted that the derived principal components will be different. A detailed treatment of this matter is given in Singh and Harrison (1985). Fung and LeDrew (1987) state that the use of correlation PC is "especially useful in multi-temporal analysis because standardisation can minimise

the differences due to atmospheric conditions or sun angles". It is, however, noted that Singh (1989) reports greater accuracy in change detection when using the unstandardised (i.e. covariance) PC.

Fung and LeDrew (1987) used PCA on a primarily urban and agricultural area with MSS data from two dates, and found that the correlation PC were essentially indicating greenness, brightness, change in greenness between the two dates, and change in brightness (in order). Ingebritsen and Lyon (1985) generally found the PC to be brightness, greenness, change in brightness and change in greenness, although they noted that the order after brightness was determined by the amount and nature of the change between dates. Deer and Longmore (1994) found that the order of PC of like-season imagery in southern Australia was brightness, greenness, change in brightness and change in greenness, and noted that there are similarities between PCA and Vegetation Index Differencing for applications involving changes to vegetation.

2.2.8 Change Vector Analysis

Multi-spectral remotely sensed image data can be represented by constructing a vector space with as many axes or dimensions as there are spectral components (bands) associated with each pixel. A particular pixel in an image can be plotted as a point in this vector space with co-ordinates that correspond to its brightness values in the appropriate spectral components. The data values associated with each pixel thus define a vector in the multi-dimensional space. If a pixel undergoes a change from time t_1 to time t_2, a vector describing the change can be defined by the subtraction of the vector at t_1 from the vector at t_2. This is called the spectral change vector. It may be calculated from either the original or transformed (e.g. PCA) data, and using either individual pixels or clusters formed by a spectral clustering or spatial segmentation

algorithm. If the magnitude of the computed spectral change vector exceeds some specified threshold criterion, it may be concluded that change has occurred. The direction of the vector contains information about the type of change.

Change Vector Analysis (CVA) was applied to forest change detection in northern Idaho (Malila, 1980) and in South Carolina (Colwell and Weber, 1981), and to general change in the Ann Arbor, Michigan area (Virag and Colwell, 1987). An extension of the concept to the vector differencing of time series of observations has been used to monitor land use in West Africa (Lambin and Strahler, 1994).

2.2.9 Statistical Tests

These involve examining the statistics of two or more image files, of the same area, taken at different times. Eghbali (1970) used the *Kalmogorov-Smirnov* test to determine whether two samples (i.e. two dates of imagery of the same location) had been drawn from the same population. With this method, the maximum difference between the cumulative distributions of the two datasets is calculated. If it is above some threshold, change is concluded to have occurred. Coiner (1980) used the correlation coefficient between the datasets of two dates as an indicator of change. Townshend *et al.* (1992) used the semivariance (the sum of the squares of the differences in pixel values) of multi-date NDVI images.

These techniques do not yield much useful information on the nature of the change, or its specific location within the image: they merely indicate that a statistically significant change has occurred somewhere in the image or region under examination. They offer some advantage, however, in that they are likely to be less affected by image misregistration.

2.2.10 Miscellaneous Techniques

There are a number of reported studies announcing or utilising techniques that are not sufficiently widespread to warrant a sub-section, but which never the less deserve mention.

There seems to be some potential for new change detection techniques to emerge from areas such as computer vision, image understanding, knowledge-based systems or fuzzy set theory. For example, Choo *et al.* (1989) sought to identify change in an image sequence using shape analysis. Researchers in the computer vision area use local texture features for segmentation, and it would seem that multi-temporal images so segmented would be amenable to change detection (see e.g. Lazaroff and Brennan, 1992). Although not focussed on change detection, Matsuyama (1987) provides an excellent review on the growing use of artificial intelligence (AI) techniques for analysis in remote sensing. Dai and Khorram (1997b) describes a change detection system using artificial neural networks. Adams *et al.* (1995) reports on cover change through the analysis of 'end-members' (i.e. proportions in mixed pixels, or 'mixels'). Chellapa (1997) discusses model-based change detection.

This thesis has a particular interest in fuzzy set theoretic approaches to change detection. There have been a number of studies reporting fuzzy classification e.g. Kent and Mardia (1988), Key *et al.* (1989), Wang (1990a), Blonda *et al.* (1991), Foody (1992), Fisher and Pathirana (1990), Nishida *et al.* (1993), and Al-Sultan and Selim (1993). Fisher and Pathirana (1993) reported on a study that used fuzzy classification in change detection. This work is discussed in greater detail in Chapter 6. Palubinskas *et al.* (1995) suggested that recent work on fuzzy classification algorithms would seem to offer some promise of improvement in the accuracy of

Post Classification Comparison.

2.3 Comparison of Techniques

2.3.1 "What's in a Name"

In order to discuss and compare techniques, it is useful to generalise and categorise individual studies and approaches. Authors choose a word or phrase to describe their approach, but, fundamentally, there are often great similarities between techniques that appear, *prima facie*, to be quite different. Consider, for example, Image Differencing. It will be called Image Differencing if the differencing is between temporal bands in the original spectral space. It will be called Vegetation Image Differencing if the differencing is performed in a transformed spectral space chosen to highlight vegetation features i.e. features that exploit the relatively high infra-red compared to red reflectance of actively photosynthesising vegetation. A transformation of the spectral space may be chosen to highlight some other specific feature, using, e.g., a canonical variate or some other (directed) statistical orthogonalisation technique (e.g. *Gramm-Schmidt* (Collins and Woodcock, 1994), or directed PCA). If the resultant transformed images are then differenced, the technique may be given a title based either on the orthogonalisation technique used e.g. the Multi-variate Alteration Detection, or MAD, transformation (Nielsen, 1996), or on the specific feature highlighted.

If a transformation of the spectral space is chosen in some undirected manner (e.g. PCA, or an unsupervised clustering technique), and the resultant images then differenced, the technique may be given another title. If there is some preprocessing of the spectral data, before differencing, to reduce unwanted radiometric effects, the change detection technique may gain its title from the

normalisation/standardisation/correction approach e.g. regression analysis. If the differencing is undertaken in the vector space (rather than individual bands or features being differenced), the approach may be called Change Vector Analysis, either in the original or in a transformed spectral space.

There is increasing awareness in remote sensing classification studies of the value of incorporating local spatial features (such as texture measures), contextual or ancillary knowledge (e.g. soil type, elevation, slope, aspect, or rainfall), and time series observations. Change detection research is beginning to reflect this increased awareness. We may therefore anticipate titles of techniques including synonyms for some or all of the following: spectral/spatial/temporal/contextual feature differencing. Fundamentally, all of the above approaches are differencing of one form or another.

This illustrates the case where different names describe essentially the same process. There are also cases where the same name describes different processes. The title Principal Components Analysis is used to describe studies in which Principal Components Analysis is used as a data reduction tool for each image separately, with the resultant images then being differenced. The term Principal Components Analysis is also used to describe studies in which one or more of the Principal Components determined from a single multi-temporal dataset shows changes directly.

In the following comparison of techniques, the names used by the authors have been generally adhered to, excepting those cases where such use would potentially cause confusion.

2.3.2 Qualitative Comparisons

The majority of the literature reviewed dealt with the use of a single change detection technique, in a particular application. Few comparative studies are available, and the majority of these do not support their conclusions by quantitative analysis. Colwell and Weber (1981) used Post Classification Comparison, Change Vector Analysis and visual estimates for forest change detection, but no ground reference accuracy assessment was done. Howarth and Wickware (1981) qualitatively compared the Image Ratioing and Post Classification Comparison methods for environmental change detection. Toll *et al.* (1980) used Uni-variate Image Differencing, Post Classification Comparison and Principal Component differencing for urban change detection, but again, only qualitative analysis was undertaken. Weismiller *et al.* (1977) used Image Differencing, Post Classification Comparison and Multi-Date Classification for change detection in a coastal zone environment, but there was no coincident ground truth available for accuracy assessment. Yasuoka *et al.* (1993) reports briefly on the use of Principal Components and "spectral signature similarity" (a correlation technique) for vegetation mapping and change detection using NOAA AVHRR LAC imagery, but did not assess the accuracy of either method. Several remote sensing change detection techniques for use in updating maps were briefly examined in Sloggett *et al.* (1994). No quantitative assessment of any of the techniques was presented, but comments were offered on the advantages and disadvantages of each. El-Raey *et al.* (1995) found that Image Differencing and Image Ratioing did not distinguish between areas of erosion and accretion in the Nile Delta, but that PCA did.

2.3.3 Quantitative Comparisons

There are, however, a number of comparative studies that do offer quantitative assessment of change detection techniques. Nelson (1983) found that Vegetation Index Differencing was superior to both Image Differencing and Ratioing for detecting gypsy moth defoliation. Banner and Lynham (1981), however, found Vegetation Index Differencing less accurate than the Multi-Date Classification approach for forest mapping. Singh (1989) summarises some previous work (Singh, 1986) on the objective evaluation of automated methods for forest change detection. Uni-variate Image Differencing, Image Ratioing, normalised Vegetation Index Differencing, Image Regression, Principal Components Analysis, Post Classification Comparison and Multi-Date Classification were compared. A number of local spatial processing techniques such as image smoothing, background subtraction, edge enhancement and texture defined by standard deviation were also investigated. A thresholding technique was applied and a number of standard deviation threshold levels were tested in the upper and lower tails of the distribution in order to find a threshold value which produced the highest change classification accuracy. The conclusions were as follows:

1. The Image Regression method using Landsat MSS band 2 produced the highest change detection accuracy.
2. Image Ratioing and Image Differencing produced the next highest change detection accuracies.
3. The various local spatial processing techniques did not improve the change detection accuracy.
4. The Post Classification Comparison approach produced the lowest change

classification accuracy.

Fung and LeDrew (1988) combined an investigation of thresholding with the change detection techniques of Image Differencing, Image Ratioing and Principal Components Analysis. They concluded that Principal Components Analysis gave the best overall results, but noted that some specific change categories were identified with greater accuracy by other techniques. Jiaju (1988) also found Principal Components Analysis superior to Image Differencing and Image Ratioing. On the other hand, Stow *et al.* (1990) found that Image Ratioing produced higher change detection accuracies than did Principal Components Analysis, and Muchoney and Haack (1994) found that Image Differencing was better than PCA, but both were better than Post Classification Comparison or Multi-Date Classification. Finally, Martin (1989) concluded Post Classification Comparison gave better results than either Multi-Date Classification or Principal Components Analysis for change detection in the rural-urban fringe.

The results of the quantitative comparisons for forestry applications are summarised in Table 2.1.

Table 2.1: Ranked Order of Change Detection Techniques for Forestry Applications

Ranking			
1	**2**	**3**	**Reference, Remarks**
VI diff	diff	ratio	Nelson (1983)
MD class	VI diff		Banner and Lynham (1981)
regression	ratio, diff	PCC	Singh (1989)
PCA	ratio, diff		Fung and LeDrew (1988)
PCA	ratio, diff		Jiaju (1988)
ratio	PCA		Stow *et al.* (1990)
diff	PCA	PCC, MD class	Muchoney and Haack (1994)
VI diff	PCA	MD class	Michener and Houhoulis (1997)
diff	PCC	MD class	Flemons (1996)

(Table 2.1 Legend: VI diff — Vegetation Index Differencing; diff — Image Differencing; ratio — Image Ratioing; MD class — Multi-Date Classification; regression — Image Regression; PCC — Post Classification Comparison; PCA — Principal Components Analysis)

2.3.4 Summary of Comparison

This review has shown that there is considerable conflict between the results of the comparative studies, even those that support their conclusions with quantitative accuracy assessment. It is not clear from the literature that a universally 'optimal' change detection technique exists: the choice appears dependent upon the application (and possibly on the specific dataset). This is not a new view. It is supported by Virag and Colwell (1987), who state "the procedure that is most appropriate to use in a given situation depends upon the specific application (type of environment, targets of interest), and the amount of detail required", and Singh (1989) "the fundamental conclusion is that even in the same environment various techniques may yield different results". There does not appear to have been much real progress since these statements. Further, it seems that claims in the early literature (specifically relating to classification) that accuracies would improve with greater spatial and spectral resolution are no longer being made with such confidence.

2.3.5 An Alternate Categorisation of Techniques

In this sub-section, we will develop a categorisation of change detection techniques that differs from Singh's (given at the beginning of this chapter as either the comparative analysis of independently produced thematic labelling or classifications of imagery from different dates; or the simultaneous analysis of multi-temporal datasets). We choose to consider change detection techniques as either pixel, feature, or object level.

In this context, pixel level refers to numerical values attached to each band of each pixel in an image. These could be the simple band recorded digital numbers (DN), or a radiometrically corrected form of these DN, or an absolute reflectance value derived from these recorded numbers.

Feature level refers to some transformation of pixel level values. The transformation could be either radiometric, or spatial, or both. We could combine information from the original image, either radiometrically (across bands) or spatially (across local neighbourhoods), or both, in order to enhance some feature of the image. Pratt (1991) defines an image feature as any distinguishing primitive characteristic or attribute of an image. Following this definition strictly, a single band may be considered an image feature, as, of course, may multi-spectral transformations such as vegetation indices. For our purposes, an image feature will be considered to be at a higher level of processing or abstraction than a simple image band, and will have required some transformation of the original data values. This is consistent with the position adopted in Bezdek and Pal (1992). The enhanced feature may have some meaningful real-world interpretation (such as vegetation indices in the radiometric domain, or edges and lines in the spatial domain), or it may not (e.g. principal components in the radiometric domain, or texture in the spatial domain). For any given application, the literature abounds with candidate features. The numbers of vegetation indices and texture measures are illustrative. Treatments of vegetation indices are contained in Guoling (1989) and Elvidge and Chen (1995). Texture measures are described in Haralick and Shanmugan (1974), Haralick (1979), and Borne (1994). Comparisons of image features often result in conflicting findings, and the 'best' feature for a given application is highly dependent on the particular dataset(s).

Finally, object level refers to symbolic or thematic labellings, or classifications, of the data. It implies an understanding of what the imaged object is. We could thus choose to categorise change detection techniques by whether they are pixel, feature or object level; i.e. by whether they operate on pixel, feature or object values between multitemporal images.

In image processing, simple band fusion (same sensor or sensor type) is commonly undertaken, using a wide variety of (predominantly statistical) techniques. There are also a number of studies that combine information of different types (e.g. electro-optical and synthetic aperture radar (SAR) imagery, or imagery with ground elevation data or magnetic readings), but 'fusion' is generally at the raw pixel level. This is referred to as image fusion, or low-level fusion, in Zhou (1994). The objective is that the fused image will reveal "new information concerning features that cannot be perceived in individual sensor images" (Zhou, 1994). Examples of this approach include Duane (1988) and Nezry *et al.* (1993).

There are also studies that attempt image fusion at a higher level of abstraction. Depending upon the nature of the approach, and the background of the researcher, this might be termed object or feature level rather than pixel level, or decision level rather than pixel level (Schistad Solberg *et al.*, 1994), or symbolic rather than sensor level (Izraelivitz and Cochand, 1990), or semantic or syntactical rather than physical level (Clement and Thonnet, 1993).

For our purposes, it is desirable to clearly distinguish between feature and object level. Object level implies that some meaningful label has been attached to the pixel. This is not a number computed from the pixel's radiometric or local spatial neighbourhood values, but a label associated with the pixel that has real world meaning. Moreover, this real world meaning is expressed in terms of object

identification, rather than merely observing that "vegetation content is high". The pixel label reveals, to some level of both accuracy and precision (meant, in this context, to be the antonym of generalisation), what the real world object being imaged (in the particular image pixel) was thought to be, at that time.

By this categorisation of change detection techniques, Image Differencing and Ratioing are pixel level techniques. Principal Components Analysis, and Vegetation Index Differencing, are feature level. Post Classification Comparison and Multi-Date Classification are object level. The value of this categorisation is discussed in the following sub-section.

2.3.6 Utility of Techniques

The comparison of change detection techniques presented earlier essentially focussed on the relative accuracy of each technique. The utility of each technique is a different matter (although there is undoubtedly some 'coupling' between these two issues). A change detection technique might aim simply to say "there has been a change in the spectral values of this pixel over time", without making any attempt to identify the nature of that change (this might be categorised as a pixel-level technique). As an example, Image Differencing does essentially this. It therefore fails to address one of the key aspects of change: identification. We might reasonably expect a technique that was attempting to be less specific (i.e. less precise) to be more accurate. In amplification, consider that we have a standing prediction that it will rain tomorrow. This prediction will be correct more often than a more specific prediction that it will rain between 1:00pm and 2:00pm tomorrow, or that the amount of rain that will fall tomorrow will be between 20mm and 30mm. The prediction that is correct more often (the less specific prediction) will be adjudged as the more accurate. Analogously, the

less specific classifier and the less specific change detection algorithm will be more accurate than their more specific counterparts.

Once a relatively low utility technique such as Image Differencing has detected change, further processing (or field survey) effort must be expended in order to identify that change. It is noted that these subsequent processes may well contribute error, but this error is not normally represented in a comparison of the change detection techniques.

At a slightly higher level of sophistication, a change detection technique could operate at the feature level, either on transformed spectral values, or on features computed over a local spatial region. It may be possible to associate some 'real world' meaning to the computed feature (e.g. a vegetation index), or it may not be. If it is not, we will not be able to draw meaningful inference from the feature level change detection technique, and we are left in essentially the same position as if we had employed a pixel-level technique. If it is possible to associate some 'real world' meaning to the computed feature, we will be able to draw some inference about the nature of the change e.g. "actively photosynthesising vegetative cover has declined". This level of change identification may prove adequate for some environmental monitoring applications, but it is not adequate for all. Again, we may need to undertake a second level of processing, or field surveys, to identify the change, and, again, the accuracy assessment of these techniques may be optimistic.

Many environmental monitoring applications need to identify change more specifically than is offered by the pixel and feature level techniques. We may ask: "where has native forest been cleared?" or, even more specifically, "where has native forest been converted to pasture?" For this, if we wish to detect changes with a single processing pass, and without extensive field surveys, we need an object level,

or classification based, change detection technique. There are two: Post Classification Comparison, and Multi-Date Classification. The former is more flexible, and is easier and simpler to implement. It is reasonable to expect that object level techniques will exhibit generally lower accuracies than many of the other techniques, as they are aiming at higher specificity. In compensation, their utility is comparatively high: the result, as presented by the change detection technique, is generally useful 'as-is'. Further, a complete 'end-to-end' analysis of change that required precise identification may reveal the object level techniques to be more accurate and/or less costly to implement than a multi-level approach using a low utility technique for change detection, followed by a post-processing identification phase.

Fisher and Pathirana (1993) argue convincingly for the use of classification based approaches to change detection. They cite some earlier literature (such as Jensen, 1986), and conclude "It has been suggested in the literature that when multitemporal imagery is examined, the most useful approach is Post Classification Comparison, since this gives direct extraction of thematic information". Fisher and Pathirana (1993) demonstrate that the use of fuzzy classification in a classification based approach to change detection produces markedly superior results to the use of traditional classification. The work in this thesis rests on these same two fundamental underpinnings. First, that a classification based approach to change detection is best (because it reveals all that we wish to know about the change), and second, that fuzzy classification produces markedly superior results to those using traditional classification. There are, however, substantial differences between Fisher and Pathirana (1993) and this current work: these are discussed in Chapter 6.

2.4 Other Issues

Before digital change detection can be undertaken, several potential problem areas must be addressed. They are image registration and rectification, selection of threshold values, radiometric correction/calibration, and accuracy assessment issues. These are discussed below.

2.4.1 Image Registration and Rectification

Importance to Change Detection

The literature is in common agreement that digital change detection requires accurate spatial registration of the multi-temporal images (e.g. Jensen, 1986; Singh, 1989; White, 1991). If high registration accuracy is not achieved, widespread changes will seem to occur over the entire scene. These are error artifacts caused by misregistration of the images (Milne, 1987). There have not, however, been many attempts to quantify 'high registration accuracy'. Jensen *et al.* (1987) stated that an accuracy of 2.26 pixels was "just adequate" (this was using aircraft remote sensing data, and considerable effort was required to get even this level of accuracy). El-Raey *et al.* (1995) suggested that an accuracy of registration within 2 pixels "represents reasonable accuracy". Milne (1988) suggested one pixel or less. Tanaka *et al.* (1992) provided some analysis and comment on required registration accuracy, and suggested that " ... 0.5 pixels for practical applications would be appropriate accuracy of registration for land change detection". Jensen (1986) advocated one half or preferably one-quarter of a pixel. Other authors have also suggested that sub-pixel accuracy is required (see e.g. Manavalan *et al.*, 1995).

Furthermore, the method normally used to estimate registration error is somewhat optimistic: true registration accuracy is almost assuredly less than that quoted in

most studies. We return to this point later in the thesis.

Little work has been done on the error effects if high registration accuracy is not achieved, and there have been few detailed, quantitative studies on the effect of registration accuracy upon change detection. One such study was Townshend *et al.* (1992). They found that misregistrations of the order of 0.2 of a pixel resulted in errors in change detection accuracies of up to 10%. It should be noted, however, that they used only one change detection method (and an uncommon one at that): a semivariagram method. These findings have been subsequently supported in Dai and Khorram (1997a), who (again using the semivariagram method) also found that a registration accuracy of less than 0.2 of a pixel was required to achieve a change detection error of less than 10%.

This matter is of considerable importance to change detection. There have not been sufficient studies to clearly indicate what level of registration accuracy we really require for a desired level of change detection accuracy, nor the relative sensitivity of various techniques to varying degrees of registration accuracy. Future work in this area is clearly warranted, but is beyond our current scope. In this thesis, the common approaches to image registration will be described, and registrations undertaken as accurately as reasonably possible.

It should be noted that the earlier conclusions regarding uncertainty of the best technique for a given application and/or dataset did not take into account the factor of registration accuracy or the relative sensitivity of various techniques to it: the situation regarding comparisons is likely to be even less clear if these issues were also considered.

On Image Registration and Rectification

Image registration is the process of geometrically aligning two or more sets of image data so that resolution cells for each imaged area can be superimposed (Swain and Davis, 1978), whereas image rectification involves correcting distortions in the imagery so that its geometry accurately represents the geometric features of the Earth's surface (Harrison and Jupp, 1989). For our purposes, rectification is essentially registration of an image to an agreed map (thus changing the image geometry to that of a map projection).

Registration and rectification aim to remove the geometric distortions that exist between image and image, and image and map, respectively. Registration will provide corresponding pixel addresses (e.g. 42,103; line or row 42, column or sometimes called pixel 103), relative to a common origin (normally the top left corner of one of the images), whereas rectification will provide absolute pixel location, in an agreed projection and coordinate system (such as a Universal Transverse Mercator (UTM) Australian Map Grid (AMG) grid reference, or a latitude and longitude).

Survey and review papers on image registration techniques include Brown (1992) and Novak (1992).

The two main cases of interest in this research are a scanned image of a map (or information from a vector GIS converted to a raster image), and an image of the real world. These cases differ significantly. A scanned map should be essentially geometrically correct, in terms of a particular projection. Imagery of the real world contains significant geometric errors or distortions.

A map needs only, at most, to be re-projected, and there are numerous texts discussing the re-projection of maps (e.g. Richardus and Adler, 1972). A scan of

a map also, of course, needs spatial referencing as well as (possibly) re-projection. There may also be a slight rotation required as a result of placing the map askew on the scanner. But, all in all, the 'geo-referencing' of a scanned map is relatively easy. If no re-projection is needed, all that is required is origin translation, scaling, and possibly rotation (to allow for some angular error in placing the map on the scanner) and reflection (e.g. to convert an image coordinate system, numbered from top left, to a map coordinate system, numbered from bottom left). If the map is in a different projection, it is arguably best to re-project before undertaking registration.

If, however, we are dealing with imagery (e.g. photographs, remotely sensed imagery), these simple corrections will not normally be adequate. There are numerous sources of geometric distortion in remotely sensed imagery. These sources include (Richards, 1986):

1. rotation of the Earth,
2. sensor scan rate,
3. sensor field of view,
4. curvature of the Earth,
5. sensor non-idealities (non-linearities?),
6. variation in the platform altitude, attitude and velocity, and
7. panoramic effects related to imaging geometry.

Relief displacement can also have an effect, particularly in areas of significant relief variation, at the edges of imagery, or if the imaging system is employing off-nadir viewing (sometimes called 'oblique' imagery, although it should be noted that there is an element of 'obliqueness' in all imagery). It is noted that relief displacement effects may be particularly significant in the case of synthetic aperture radar (SAR) imagery.

Therefore some or all of the geometric distortions inherent in imagery must be removed or reduced before comparison of multi-temporal imagery is possible by other than visual means. The two principal methods of removing geometric distortions are:

1. model the nature and magnitude of the distortions, and then use these models to correct the imagery; and/or
2. establish empirically derived mathematical relationships between pixel addresses in an image, and either pixel addresses in another image (registration), or map coordinates (rectification).

It is also sometimes appropriate to correct for topographic effects (by use of digital elevation models). This is common in the case of radar imaging because of the viewing angles used.

The two methods for removing geometric distortions can be, and commonly are, used in conjunction. The first method, using platform parameters (such as position and attitude) and knowledge of the viewing geometry and sensor characteristics, is often used to get the general geometry desired, but it is not able to achieve the required accuracy for many applications. It is therefore often followed by the second method to achieve a more precise registration. Corrections possible by the first method are commonly applied by image vendors prior to delivery of the imagery.

The most commonly used procedure for the second method is as follows:

1. Define two Cartesian coordinate systems such that the location of a point in an image is given by the coordinate pair (u, v), and the location of the same point on a corresponding map is given by the coordinate pair (x, y).
2. Assume that the two Cartesian coordinate systems can be related by the mapping functions f and g such that $u = f(x, y), v = g(x, y)$.

3. Since the explicit form of these functions are not known, we normally assume them to be simple polynomials of some chosen order.
4. Estimate the polynomial coefficients by substituting into the chosen polynomials the corresponding coordinate pairs of a number of measured ground control points (GCP), using some error minimisation technique if there are more points than are strictly required.
5. Estimate the accuracy that would be achieved if the mapping was undertaken using this transformation (note that this is related to GCP transformation error, not general pixel transformation error) (this step serves to assist in identifying any mis-identified or mis-located pixel/map coordinate pairs).
6. Resample the image to the map coordinate system in some manner (e.g. nearest neighbour, bilinear interpolation, cubic convolution).
7. (Optionally), estimate the accuracy that has been achieved in using this transformation (i.e. general pixel transformation error). This is normally achieved from an analysis of the 'residuals' of the GCP used to determine the transformaton. It is, however, noted that a more accurate estimation procedure would involve the use of an independent test set of GCP.

A number of issues need to be addressed. These relate to both the acceptable accuracy and confidence limits on the spatial positioning of pixels in the resultant registered images, and the radiometric values to be assigned to each of those pixels. More specifically, we need to address:

1. The number and spatial distribution of GCP.
2. The accuracy of measurement of pixel and map coordinates of these GCP.
3. The order of the mapping polynomial, and whether or not to include particular polynomial terms.

4. The resampling method.
5. The resampled pixel size.
6. Global or local transformations.
7. Differences between aircraft and space platforms.

Several of these issues are examined in Mather (1995).

The number and spatial distribution of GCP. Schlien (1979) noted that "The geometric accuracy ... depends directly on the accuracy, number and distribution of ground control points acquired". In order to solve for the transformation polynomial coefficients, we require a certain minimum number of GCP. Or, to phrase this another way, the minimum number of GCP required is determined by the chosen order of the mapping polynomial. In the case of the commonly used third order transformation (or mapping) polynomial, at least ten GCP are required (i.e. at least one GCP for each term in the polynomial). GCP in excess of the minimum required number permit the estimation of the coefficients based upon some error minimisation technique (such as minimising the sum of the squared errors, or minimising the maximum error), and allows the calculation of the magnitude of the errors that will occur, at least in so far as the GCP are concerned, if that mapping polynomial is used to transform the image. Ford and Zanelli (1985) conclude that, in order for the mean square error to be less than 0.25 pixel, the number of ground control points needs to be four times the number of coefficients in the least-squares polynomial function (this conclusion is, of course, dependent on the spatial distribution of the GCP).

As stated ealier, a number of GCP may be held separate to permit a calculation of transformation errors independently of those used to determine the mapping polynomial. A practice similar to this is commonly used by cartographers, but it

is, perhaps surprisingly, not normally employed in image registration/rectification. Gong *et al.* (1992) note that "The so-called 'standard error' or 'average residual error' provided by existing geometric correction software are only estimates from many individual pixels (ground control points) selected from both images. This type of index inevitably underestimates the actual registration noise in difference images". It therefore seems prudent to estimate the error with an independent set of selected data points.

The manual identification and measurement of GCP is a time consuming business. Accordingly, there is considerable research interest in automating this process (see e.g. Allison *et al.*, 1991; Holm, 1991; or Brivio *et al.*, 1992).

The spatial distribution of GCP is also important, and this is again influenced by the chosen order of the mapping polynomial. A detailed treatment of this subject is beyond the scope of this paper (see e.g. Orti, 1981), but the following general observations and 'rules' are offered:

1. Because the 'influence' of a GCP in 'tying down' the mapping polynomial decreases with distance from the GCP, GCP should be selected throughout the regions for which 'accurate' mapping transformation is desired. Arguments as to the desired statistical distributions (e.g. regular or random) are largely academic, as practical considerations of locating GCP in images normally dictate where, within a region of an image, the GCP are selected.
2. The higher the order of the mapping polynomial, the less 'well behaved' the polynomial is outside the minimum polygon enclosing the set of GCP. Thus some GCP should be selected on the periphery of the image area of interest, or even outside it. The higher the order of the mapping polynomial, the more important this 'rule' becomes.

The literature sometimes uses expressions like 'a good distribution', or 'well distributed' without explaining what is meant by these expressions. It is not even always clear whether the criteria for such judgements are based upon the mathematical or common English usages of these terms. A random distribution would resolve questions of bias in GCP selection, but such a distribution of GCP will result in varying 'accuracy' of the mapping transformation across the image (it is conceded that all distributions of GCP will result in varying 'accuracy' of the mapping transformation across the image; some distributions will, however, be more predictable, and others will ensure accuracies within certain predefined limits). GCP will, of course, never be selected at random locations in practice, because the process requires the precise identification and location of points on both map and image (or image and image). The nature of most areas and materials is such that not all points are uniquely identifiable. Thus the process cannot be random. It can similarly not be regular (although regular would confer certain advantages with respect to 'uniformity' of the transformation accuracy across the image).

Thus it would seem that a 'good' distribution is stratified in nature (for reasons of 'uniformity' and predictability of the transformation accuracy across the image), with GCP selection in each stratified sub-area as 'un-biased' as possible.

The accuracy of GCP measurement. The accuracy of maps is commonly stated upon them. Therefore, although it can in principle be stated that it is desirable to make highly accurate measurements of the location of GCP in both map and image coordinate systems, this must clearly be influenced by the accuracy of the available reference maps. For 1:50,000 topographic maps, the stated accuracy (in spatial location) rarely exceeds +/- 25 metres. Further, cartographers are known

to use some licence when it come to placing (in particular) human-made features on maps. This is particularly noteworthy as human-made features are commonly selected as GCP due to the relative ease of locating them on both images and maps.

Therefore to quote map coordinates implying an accuracy exceeding +/- 25 metres could be misleading. When dealing with imagery with a pixel size of the same order (i.e. 25 metres), we might reasonably expect various parts of the rectified imagery to exhibit 'errors' of the general order of one pixel. It is noted that this error exceeds the recommendation for change detection work. Selection of image coordinates should obviously be to the achievable accuracy. In a number of image processing systems, these coordinate values must be integral, which provides an implied spatial accuracy of not better than the order of +/- one half a pixel size. Again, we can expect difficulty in achieving the recommended accuracy for change detection.

The order of the mapping polynomial. The higher the order of the mapping polynomial, the better the 'fit' that can be achieved in mapping the observed GCP image coordinates to the chosen coordinate system. Unfortunately, the higher the order of the mapping polynomial, the more likely it is that the mapping polynomial will deviate from its desired value as its distance from the GCP increases. This effect can be particularly pronounced outside the minimum polygon enclosing the set of GCP, as observed earlier.

If we know that our starting image is a scanned map, already in the desired projection, we will only need transformation operations such as translation, rotation, scaling and reflection. These can be effected by an affine transformation of the form:

$$X' = A.X + B.Y + D$$

$$Y' = K.X + L.Y + N$$

The values D and N account for origin translation in X and Y respectively, the coefficients A and L account for re-scaling in X and Y respectively and/or reflections and rotations, and the coefficients B and K account for skewing in X and Y respectively and/or reflections and rotations.

To see that this is so, observe that

$X' = X + D$

$Y' = Y + N$

account for origin translation.

$X' = A.X$

$Y' = L.Y$

account for re-scaling.

$X' = A.X + B.Y$

$Y' = K.X + L.Y$

account for a rotation of θ if $A = L = cos\theta, B = -K = sin\theta$.

$X' = A.X + B.Y$

$Y' = KX + L.Y$

account for a reflection in the line $y = x\ tan\alpha$, if $A = -L = cos2\alpha$, $B = K = sin2\alpha$.

We can also account for skewing by introducing an additional term in each equation, giving us:

$X' = A.X + B.Y + C.X.Y + D$

$Y' = K.X + L.Y + M.X.Y + N$

where

$X' = X + B.Y$

$Y' = K.X + Y$

account for skewing (in one direction only), and

$X' = C.X.Y$

$Y' = M.X.Y$

account for bi-directional skewing.

If, however, we are dealing with a 'real world' image, we may (and normally do) require higher order terms. These will permit warping the image along some curved shapes or lines, and can be used to account for distortions caused by variations in platform parameters and non-linearities during the scanning process (in the case of a scanning imager), or for obliquity in the case of photographs. The higher order terms can be used to correct skews and distortions that are not linear across the image. A quadratic form will permit correction of one change in each parameter, a cubic form permits two changes in each parameter, etc. It has not normally been found necessary (or practical) to account for more than two changes. We may therefore choose a transformation of the form:

$$X' = A.X + B.Y + C.X.Y + D + E.X^2 + F.Y^2 + G.X^2.Y + H.X.Y^2 + I.X^3 + J.Y^3$$

$$Y' = K.X + L.Y + M.X.Y + N + O.X^2 + P.Y^2 + Q.X^2.Y + R.X.Y^2 + S.X^3 + T.Y^3$$

Equations of this form can also be used to change maps from one projection to another if, for some reason, it is not possible to use a precise mathematical transform.

Whether or not to include particular polynomial terms. While the order of the mapping polynomial may have been chosen, it is not necessary (or even, in some cases, desirable) to include all of the terms of that polynomial: e.g. we may choose to leave out the line squared times pixel term (i.e. the $X^2.Y$ term). Guidance on this matter can be obtained by calculating the "t-values" of the terms. The "t-value" is the ratio of the absolute value of the coefficient to its standard error.

If the "t-value" is low for a particular term, this term may safely be assessed as being of low significance, and should be omitted from any subsequent calculations. Alternatively, an iterative approach could be adopted whereby the coefficients are calculated without a particular term, and the errors calculated and compared with the errors associated with the mapping polynomial determined by including the term. This approach clearly significantly increases computational effort.

The resampling method. Resampling methods are normally either interpolative, or nearest neighbour. With nearest neighbour resampling, the pixel that has its centre nearest to the desired map point is effectively transferred to that point, with brightness values unchanged. Richards (1986) states that this is the preferred technique if some form of image processing, particularly classification, is to be performed on the resultant image. The reason given is that, for classification, the brightness values should be the actual radiance values detected by the original sensing system. This argument does not seem sufficient. It is understood that the radiance values detected by the sensing system have contributions not only from the pixel in the IFOV of the sensor, but also from neighbouring pixels (and indeed, also from the atmosphere etc.) in a weighted additive manner that differs for each sensing system (sometimes referred to as "the point spread function", or PSF). This effect is mathematically similar to interpolation. Thus it might be argued that interpolation should be the preferred method to estimate the brightness values that would have been detected by the sensor, had it actually been 'looking' at the desired precise area. This should be irrespective of the intended application. Notwithstanding this, the literature generally recommends nearest neighbour resampling for classification purposes. Since an investigation of this matter is beyond the scope of our current

work, the general recommendation of the literature is accepted.

Interpolation effectively involves fitting a two dimensional surface through some number of surrounding pixel centres, thus estimating the brightness values for a pixel at the desired location. The two dimensional surface is normally of first order (i.e. bilinear), or third order (i.e. bicubic).

Interpolation is meaningful if the digital numbers recorded in the imagery are some measure of brightness value. This is the case in 'raw' remotely sensed imagery, but is often not the case in classified (i.e. thematic) imagery, scanned maps etc. How does one interpolate between class 1 and class 3 in classified imagery, or between orange and blue on a scanned map? In the case of scanned maps, the problem is actually worse than it first appears. As in the case of all 'raster' image files, a decision must be taken on how many bits will be used to store the 'value' of each pixel. The final display device must take this 'value', and decide what to plot on the screen. Using 24 bits of data for each pixel maps nicely to display devices, in that 8 bits can be used each for RGB, or HSV (depending upon the colour model being used), but a particular display device may only be 8 bit and not able to use the 24 bit precision of the data. We have thereby wasted data storage capacity. If only 8 bits are stored, it is common to use some form of colour look up table (CLUT) to make more effective use of these 8 bits. Interpolative methods can therefore not be used on colour images with associated CLUT.

Interpolation can make sense if the image is thought of as being comprised of pixels with values representing brightness in some ordered manner (e.g. monochrome images, or colour images with 24 bit values, with the first 8 bits representing ordered red brightness values, the next 8 bits representing ordered green brightness values, and the final 8 bits representing ordered blue brightness values), but interpolation is

inappropriate otherwise. It is therefore suggested that interpolative methods should only be used for greyscale (or other monochrome) images and 24 bit colour images, with the nearest neighbour method being used for 8 bit colour images (and in any case where the preservation of original pixel values is desired).

The resampled pixel size. Selection of the resampled pixel size is influenced by the resampling method chosen. It is most significant in the case of nearest neighbour resampling. Problems of both the duplication of pixels and the loss of complete pixels can occur, even if the resampled pixel size is chosen wisely. The safest course is to resample to the original pixel size, or smaller.

It is sometimes desired to 'thin' the data somewhat. Interpolative resampling methods might then use simple or weighted averaging (with the weights determined by Euclidean distance from the new pixel centre), and the nearest neighbour method might be either used as is, or replaced with a modal filter value from some chosen spatial neighbourhood.

The choice of resampled pixel size may also be influenced by the sophistication of the intended eventual display system. Zooming in is easy (particularly to areas decreasing as the reciprocal of 2^n, where n is a non-negative integer); zooming out is less so. If one pixel on the screen is required to represent some number of pixels in the stored image, some decisions and/or processing will be required. This problem may be largely avoided by choosing a data pixel 'size' in the stored image such that the image, when viewed at the maximum likely 'zoom out', is displayed with one data value corresponding to one screen pixel. Thus decisions of this nature clearly affect data pixel 'size', and the required sophistication of the intended eventual display system. They therefore affect decisions about the resampled pixel size.

Global or local transformations. A transformation may be termed global if a single transformation is used to map all pixels in the original image to the new image. A transformation may be termed local if it only acts upon a limited part of the original image. The other regions of the original image must therefore be transformed by other (local) transformations. Historically, most registration transformations have been global (the techniques described above generally fall into this category). Local transformations, however, offer potential for improved accuracy in each region, at the expense of additional computational effort. A review of approaches to local transformations can be found in Brown (1992).

Differences between aircraft and space platforms. Satellite platforms are relatively stable. A simple least-squares regression approach is usually reasonably successful in removing image distortions, and image-to-image registrations to within one pixel are usually possible (Weismiller *et al.*, 1977; Jensen *et al.*, 1987). With remotely sensed data from aircraft, image distortion can be a much greater problem. Winds and turbulence can cause significant aircraft movements or changes in velocity while the sensor is scanning. In general, significantly poorer registration results are obtained with aircraft data, even if three or more times as many ground control points are used (Jensen *et al.*, 1987).

Research Directions. A major area of research has been previously mentioned: automatic GCP selection and registration (see e.g. Lee *et al.*, 1993; Zheng and Chellapa, 1993, 1994; Hanaizumi and Kanemoto, 1996; or Bartl *et al.*, 1996). There is also research into a number of other approaches to image registration (see Fonseca and Manjunath, 1996, for a review). Some of these proposed techniques require

'approximate' registration first, with high (normally sub-pixel) accuracies achieved in this preliminary registration. A second processing step, using a different approach, is then used to refine the registration hierarchically.

After a review of the literature, it is concluded that, while these research directions offer promise for increased ease, speed and accuracy for image registration, they are not yet sufficiently mature for operational use. The standard error minimisation global polynomial mapping transformation approach with manual GCP selection has therefore been used for the work reported in this thesis.

2.4.2 Selection of Threshold Value

Most digital change detection techniques require the selection of a threshold value in order to determine when change is considered to have occurred. The two main methods currently in use for selecting the threshold value are:

1. inter-actively, whereby the analyst adjusts the value until satisfied with the result; or
2. using some statistical measure, such as standard deviation, or the Mahalanobis Distance from the mean of difference images (under the assumption that such distances will follow a chi-squared distribution, with the number of degrees of freedom equal to one less than the number of spectral bands).

Essentially, selection of a threshold value to detect change implies a Boolean decision: it is assessed that change either has, or has not, occurred. The approach developed later in this thesis aims at determining the extent to which change is believed to have occurred in any given pixel. It seeks to avoid this type of Boolean decision. The subject of selection of threshold values is therefore not discussed further. Further details may, however, be found in Fung and LeDrew (1988).

2.4.3 Radiometric Correction/Calibration

Some digital change detection techniques (e.g. image differencing, change vector analysis) give improved results if the radiometric data are corrected and/or calibrated in some manner. It was stated earlier that a fundamental premise of using remotely sensed data to detect change was that changes in the object of interest will result in changes in radiance values that are large compared to radiance changes caused by other factors, such as differences in atmospheric conditions, illumination angle and soil moisture. If the change detection technique is sensitive to these other factors, they need to be corrected for in some manner.

If meteorological information for the time of data acquisition is available, atmospheric corrections can be made to the raw data. Topographic information can also be used (e.g. Pons and Sole-Sugranes, 1994). It has, however, been more common in the past to attempt to standardise one dataset to another rather than to try and apply the complex modeling that is required to correct for variations in atmospheric transmission and path radiance (Milne, 1987).

Changes in the solar illumination angle of a surface brought about both by the time of day of the image acquisition and by the apparent summer-winter migration of the sun have an obvious effect on the digital numbers recorded by the sensor system. As many satellite systems are in a sun-synchronous orbit, and pass over the same latitudinal band at the same local ground time on each overpass, there are no variations in the data caused by daily changes in solar elevation for each latitude. However, seasonal changes in sun angle effect both the intensity of energy received at a surface and the component of shadow included in the reflectance values recorded. Graetz and Gentle (1982) showed that for rangeland environments in Western New

South Wales, the dynamic range of Landsat digital count values was 2.6 times greater in summer than in winter and the shadow component associated with a metre high salt-bush plant increased from 8% in summer to 35% in winter.

Surface reflectivities vary with the stage of phenological development. Phenologic change associated with events such as crop growth or ephemeral vegetation flush may obscure long term changes to surface types. While there are software routines to correct for sun angle changes, the only truly effective way of dealing with the effects of seasonal phenologic change is by selecting data collected on or near anniversary dates (Milne, 1987).

The comparison of pixel values derived from different sensors can lead to poor results unless some form of calibration is used to standardise the responses of multi-spectral scanners involved. Tables have been published in Robinove (1982) and Ahern (1985) to convert the digital numbers for individual bands on each of the Landsat multi-spectral scanners to radiance or reflectance values.

Another method of calibration between different dates and/or sensors is by reference to regions in the image in which no significant change is believed to have occurred. A regression approach (Virag and Colwell, 1987; Vogelmann, 1988) or image histogram matching (Richards, 1986) may then be applied. A useful paper on reflectance calibration for change detection is Olsson (1995).

2.4.4 Accuracy Assessment

Accuracy Assessment in Classification

The traditional method of determining the accuracy of a thematic map is by comparison with corresponding reference data. It is normal to create a (normally square) matrix, termed an error or confusion matrix, with the assigned class labels

on one side (usually the rows of the matrix), and the 'true' classes on the other (usually the columns of the matrix). The correct classifications are on the principal diagonal of the matrix: all other entries are errors. The row entries (excepting the principal diagonal entry) reveal errors where the classifier has labeled pixels as a particular class, whereas they actually 'belong' to a different class. This is the "commission", or "user's" error; the total of which is found by dividing the correct classifications, the principal diagonal entry, by the total number of pixels assigned to that class, the row total. There are analogous errors to be found in columns, where the 'true' class was misidentified (excepting, of course, the principal diagonal entry). These are referred to as "omission", or "producer's" errors. Again, the total is found by dividing the correct classifications, the principal diagonal entry, by the total number of pixels actually in that class, the column total.

A single figure is sometimes quoted for accuracy. This is the number of correct classifications, the sum of the diagonal elements, divided by the total number of classifications (revealed by either summing the sums of rows, or summing the sums of columns).

It has been argued (Rosenfield 1981; Congalton *et al.*, 1983; Rosenfield and Fitzpatrick-Lins, 1986; and Fitzgerald and Lees, 1994) that the kappa statistic should be used, thereby eliminating the element of chance agreement. See Hudson and Ramm (1987) for calculating this statistic. Other useful papers on assessing traditional classifier accuracy include Zhuang *et al.* (1995) and Richards (1996).

The principal limitation of these traditional methods of accuracy assessment is that they assume that each area on the ground can be assigned a single label. This raises the problem of vagueness of class definitions (e.g. what size and density of trees is needed in an area for that area to be classified as forest). This is addressed

in Gopal and Woodcock (1994), and Foody (1996). Further problems include:

1. Accuracy assessment can only be undertaken for the exact classes or superclasses represented in the reference data. Obviously, if reference data is collected for a particular study, efforts can be made to ensure that the reference data and image classification classes coincide precisely. This is not the case when historical reference data is used (as is often the case in change detection studies).
2. The mixed pixel problem arises. Training and testing sites are normally selected to avoid physical boundary regions. Practically, there are a significant number of pixels in an image which correspond to physical boundary regions, and therefore comprise mixed pixels. Accuracy assessments will therefore be questionable, as will any areal estimates derived from the classification (see Cross *et al.*, 1991; or Foody and Cox, 1994).
3. Limited information is provided about the 'seriousness' of the errors, or their spatial distribution. All errors are normally considered to be equal in terms of class mis-assignment, and the location of errors is not normally reported or considered.

Accuracy Assessment in Change Detection

The accuracy assessment of change detection techniques has normally been subjective and qualitative. This is attributable, in large part, to the absence of the reliable ground truth data necessary to undertake more objective evaluation. Ground truth data is required for two or more dates, corresponding to the dates of the chosen or available imagery. Often, reasonably reliable ground truth data can be obtained for one date (usually, the most recent), but it is rarely available for any earlier period. An approach that is sometimes adopted is to undertake visual inspection and

classification of high (spatial) resolution aircraft imagery, but this approach is not without flaws (and, of course, it requires that such imagery be available). Congalton (1991) notes that assumption that photointerpretation is completely correct "is rarely valid and can lead to a rather poor and unfair assessment of the digital classification". Similarly, digitisation of old maps will carry forward the errors of the old maps, and invariably introduce some more.

Where quantitative accuracy assessment of change detection techniques has been undertaken, a Boolean approach has generally been adopted. That is, does the change detection technique agree with the (hard) reference data? There is obviously scope within this paradigm to report both omission and commission errors, although this is rarely done. There has been little research reported on the use of fuzzy change detection techniques, and the matter of accuracy assessment in fuzzy change detection requires further work. If a classification based approach to fuzzy change detection is proposed, Foody (1995) represents an excellent starting point.

There are a number of factors influencing the accuracy of change detection techniques. A highly significant one is registration error, the effects of which, as previously stated, are not well known. The image registrations in this research were imperfect, and there was an absence of reliable ground truth data for both dates. Therefore only qualitative assessment of the approach described in this thesis was possible.

Chapter 3

STUDY AREA, MATERIALS AND ENVIRONMENT

Summary: In this chapter, the study area, data and image processing environment are described and shown.

3.1 Study Area

The study area was located towards the western end of Kangaroo Island, South Australia, and broadly bounded by latitudes 35 degrees 40 minutes S and 36 degrees 03 minutes S, and by longitudes 136 degrees 45 minutes E and 137 degrees 05 minutes E. It included a large part of the Flinders Chase National Park.

The region of interest is primarily native forest, with the dominant canopy species being eucalypts (in particular: *Eucalyptus remota* and *E. baxterii* on the lateritic plateau; *E. diversifolia*, *E. rugosa* and *E. lansdowneana* around coastal areas and areas of calcareous sands; and *E. cladocalyx*, *E. fasciculosa* and *E. viminalis ssp cygnetensis* in creek and drainage lines). There are, however, numerous cleared areas. These are variously natural or improved pasture, cereal crops, or used for the production of hay. The pastures are used as grazing paddocks for various livestock.

3.2 Materials

The following materials were used for various parts of the research associated with this thesis:

1. A Landsat MSS image (Path 98 Row 85, acquisition date 8-3-89).
2. Two Landsat 5 TM (partial) images, 1024 by 1024 pixels, bands 2, 3, 4 and 5, nominally centred 35.50323 degrees S latitude, 137.1208 degrees E longitude (Path 98 Row 85, acquisition dates 8-3-89 and 29-1-93).
3. Maps (Topographic Survey, Australia 1:100,000, Series R641, Sheets 6226 Borda and 6326 Vivonne).
4. Ground truth data (digital), nominally 1993.

The reasons for using only TM Bands 2 to 5 (and not Bands 1, 6, and 7) were as follows. Horler and Ahern (1986) found that, for general forestry applications, TM Bands 3, 4, and 5 were the best three band combination, and that they performed almost as well as the first three Principal Components. Horler and Ahern (1986) also noted that for analysis of softwoods (a small area of which is present in the Kangaroo Island study area), TM Bands 1 and 2 become more important than TM Band 3. It was therefore considered that using only TM Bands 2 to 5 for forestry classification should result in little or no appreciable information loss. These findings are convenient for change detection study purposes, because TM Bands 2, 3, 4, and 5 correspond spectrally to MSS Bands 1, 2, 3, and 4, and there is archival MSS imagery dating back to the early 1970s. In the desire to make any findings of this study readily extensible to the archival MSS imagery, it was decided to use only TM Bands 2, 3, 4, and 5. No attempt is made in this thesis to further investigate the findings of Horler and Ahern (1986) in the context of the chosen approaches to classification of

Australian mixed forest and pasture landscapes.

3.2.1 Images

The MSS89 and TM89 images displayed regions of cloud and cloud shadow, but the TM93 image was entirely cloud free. Remotely sensed imagery often contains spatial striping patterns caused by the miscalibration of multiple sensors. The period of the striping is related to the number of sensors in the scanning system. There was clear evidence of such striping in the TM89 image. Destriping was attempted using a Fourier transform/periodic noise removal approach. A number of spatial frequencies (from three to ten) were tried, but none favourably affected the data without introducing other undesired effects. The TM89 image was therefore used as supplied.

3.2.2 Ground Truth Data

The 'ground truth' data were obtained from the South Australian Department of Housing and Urban Development, in Arc Interchange format (a proprietary data exchange format from ESRI). Thes data were released in 1993, and are therefore described as nominally 1993. The dataset was, however, compiled over the preceding 18 months by interpretation of photographic imagery supported by limited field survey. The effect of this is discussed later. The data were imported as vector layers into ARC/INFO, then converted to a thematic raster image, with a pixel size of approximately 12 metres. There were 16 classes in the original data, but this reduced to 12 when only a (geographic) subset of the full ground truth data was used.

The 12 remaining classes were:

forest (1) - AS2482 code 60010, condition 1 (normal)

forest (40) - AS2482 code 60010, condition 40 (uncertain/unknown)

forest (61) - AS2482 code 60010, condition 61 (modified)

natural vegetation (other) - AS2482 code 60001

sea - AS2482 code 40011

swamp - AS2482 code 44190

lake - AS2482 code 44010

pine - AS2482 code 65010

pasture/cropland - AS2482 code 65060

3 road classes - AS2482 codes 20130 (vehicular track), 20300 (sealed road), and 20330 (unsealed road)

The final subsetted ground truth raster image is shown as Figure 3.1.

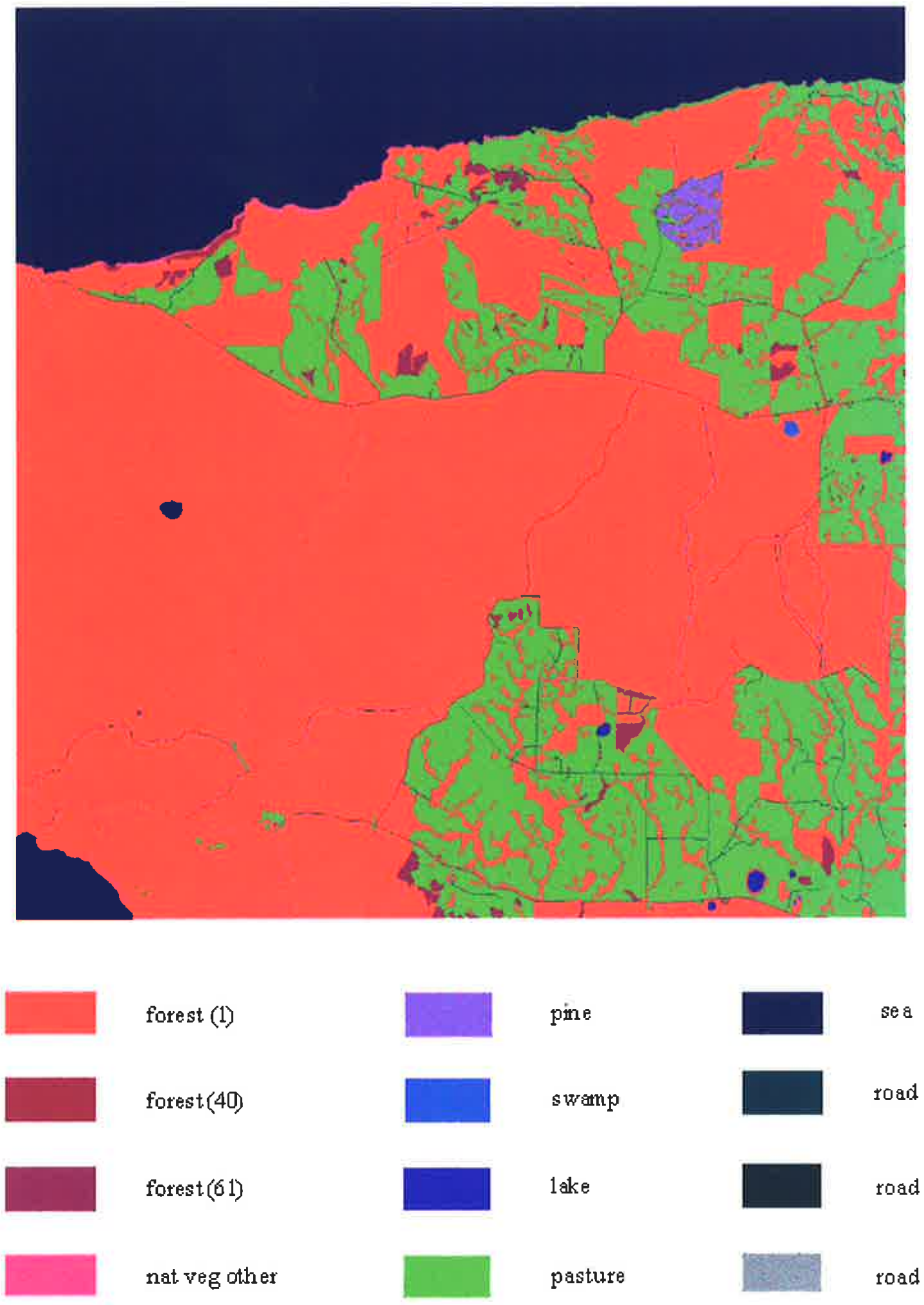

Figure 3.1: Ground Truth Raster Image, 1993

As is often the case, there were errors associated with these ground truth data. They appear to have arisen largely through data collection occurring over a period of time. The dataset, when produced, was dated 1993, but it was already out of date at that time. Some areas had apparently been cleared of native forest, but were not shown as such in the ground truth data. These errors show in the accuracy assessment of classifications discussed in Chapter 4. There was no reliable ground truth dataset available for any earlier date.

3.3 Image Processing and Computing Environment

The majority of the image processing work described in this thesis was undertaken on Silicon Graphics machines, running ERDAS IMAGINE (V8.2/8.3) under IRIX (V5.3/6.2). Some of the later work was on WindowsNT (V4.0) machines, also running ERDAS IMAGINE. Most of the software was developed using the graphical modeling and script language environments provided in ERDAS IMAGINE. The productivity increase that these environments provided, over the traditional software development environments for image processing, was considerable.

Some associated research reported earlier (Deer *et al.*, 1996a,b) was undertaken using software written in C, developed in the Khoros environment, and run on a SUN SPARCstation IPC operating under SunOS version 4.1.4. Khoros is a software development environment which is distributed through the Internet via anonymous ftp as Free Access Software (see http://www.khoros.unm.edu). Source code is available, so developers can write their own applications and 'link' them into Khoros. The Khoros environment provides a number of useful facilities (including image management and display), and a visual programming environment, Cantata.

This work also required the use of the MicroBRIAN Image Processing System, with peripheral equipment, for loading image tapes.

Considerable use was made of Microsoft Excel for calculations and the production of tables. MATLAB, Excel, and the MacIntosh Graph Calculator were used for the production of graphs.

3.4 Image Registration/Rectification

The importance to change detection studies of accurately registering image to image, or image to map, has been previously discussed. The normal approach was described in Section 2.4.1: manual selection of corresponding points in images and maps (GCP); calculation of a global polynomial transformation, using a minimum squared 'error' strategy; then warping the image in accordance with this transformation. This approach has been adopted in this thesis. Both the TM89 and TM93 images were rectified/registered to the ground truth image described earlier using the nearest neighbour resampling method. The images were resampled to correspond to the pixel size of the ground truth image. This caused a data expansion to nearly 6.5 million image pixels (from the original number of slightly over 1 million). No other form of pre-processing of the images was undertaken (i.e. no smoothing, radiometric enhancement or equalisation, atmospheric corrections, or conversion to absolute reflectances). The results of the registration process are given and discussed below.

It was also stated earlier that estimation of the accuracy of the transformation achieved is normally taken from the 'residuals' of the GCP used to determine the transformation. The errors are often quoted as the mean square or root mean square (RMS) of these residuals. That is, the points used to determine the polynomial are then used to estimate the accuracy achieved. Recalling that the transformation

(i.e. polynomial) coefficients were chosen to minimise this figure, it might reasonably be expected that the 'true' error would be higher than any estimate so derived. Accordingly, we have used a number of independent test points to estimate error.

3.4.1 TM89 Image

50 GCP points were manually located in both the original TM and ground truth images. There were sufficient well defined artificial features (e.g. road junctions, boundary fences) to be satisfied that these GCP were identified and located with reasonable accuracy. Transformations of various orders (up to and including third order) were calculated. Some iteration (i.e. re-examination of GCP, and adjustment if necessary) occurred as residuals were calculated. A third order transformation, and a file to map transformation type (map projection UTM, Australian National spheroid, UTM zone 53, South of the Equator), were finally selected. The transformation coefficients, and RMS residuals, are shown in Table 3.1.

Table 3.1: Transformation Coefficients for TM89 Image

TM89 Transformation Coefficients		
	X'	Y'
constant	-319.175642464	-922.299848186
$\mathbf{X}$	0.575676178	0.095812959
$\mathbf{Y}$	-0.081351155	0.567495900
$\mathbf{X^2}$	0.000022888	0.000020850
$\mathbf{X \cdot Y}$	-0.000016612	-0.000019265
$\mathbf{Y^2}$	-0.000011791	0.000017173
$\mathbf{X^3}$	-0.000000009	-0.000000006
$\mathbf{X^2 \cdot Y}$	0.000000014	0.000000007
$\mathbf{X \cdot Y^2}$	-0.000000009	-0.000000001
$\mathbf{Y^3}$	0.000000005	-0.000000002
RMS Residuals		
$\mathbf{X}, \mathbf{Y}$ RMS	0.471792000	0.510891000
Total RMS	0.695412000	

Information on the errors of the 10 independent test points after transformation is shown in Table 3.2.

Table 3.2: Errors of the Independent Test Points - TM89 Image

GCP Number	X error	Y error
GCP #59	0.0800	-1.6650
GCP #60	3.2490	-0.0090
GCP #61	-4.0600	-0.9490
GCP #62	-0.5110	-1.4820
GCP #63	1.5300	-0.9010
GCP #64	1.7820	-0.2070
GCP #65	2.9960	0.6950
GCP #66	-0.4890	1.2100
GCP #67	2.7320	1.0130
GCP #68	0.7310	1.2300
X, Y RMS	2.2369	1.0588
Total RMS		2.4748

Ford and Zanelli (1985) recommend that, in order to achieve a registration mean square error (MSE) of 0.25, the number of GCP must be at least four times the number of polynomial coefficients. For a third order polynomial (if all terms are included), there are 10 coefficients. The number of GCP selected (50) exceeded the recommendation. It is noted that none of the 50 GCP used in determining the mapping polynomial had a residual of one pixel (or more), in either the x or y directions, and the RMS residual was 0.70 (MSE of 0.5). However, every one of the independent test points has an error, in either x or y, of at least one pixel. The RMS error was 2.47 (MSE of over 6). There is clearly a different indication of registration error provided by 'GCP residuals' and 'independent test set' methods.

The rectified TM89 image is shown at Figure 3.2.

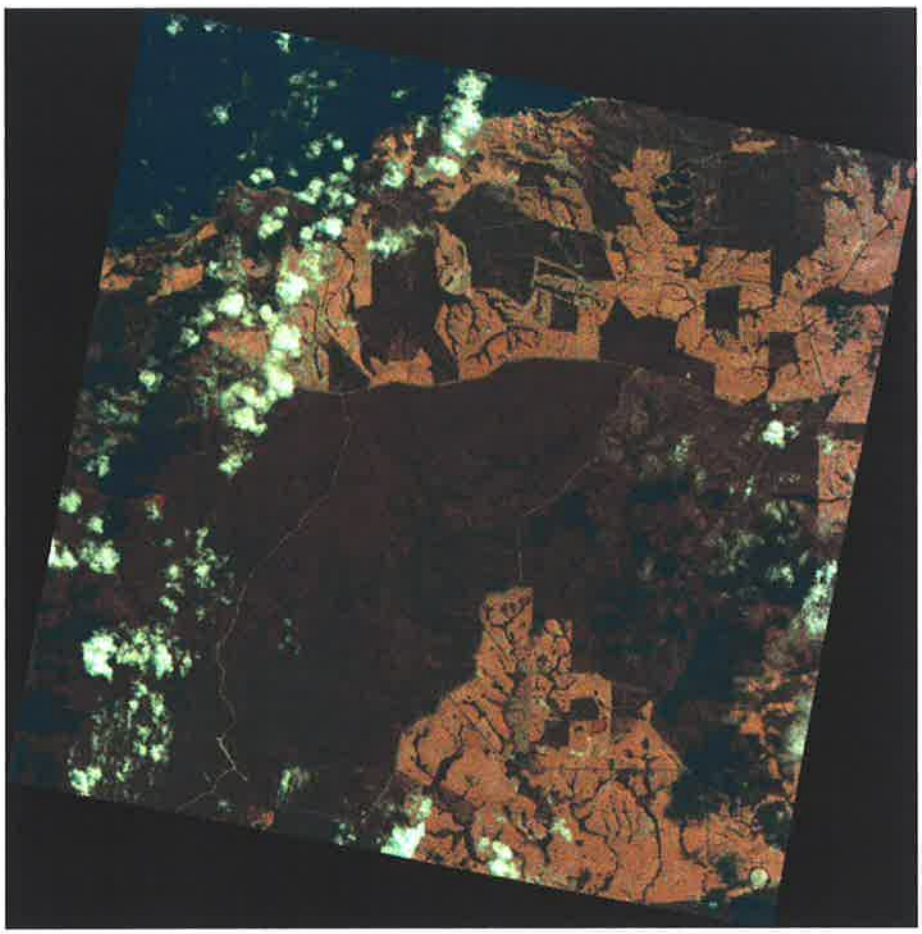

Figure 3.2: Rectified TM89 Image (Bands 5,3,2 as RGB)

3.4.2 TM93 Image

Again, 50 GCP points were manually located in both the images. The final transformation coefficients are shown in Table 3.3.

Information on the errors of the 10 independent test points after transformation is shown in Table 3.4.

Again, it is noted that every one of these independent test points has an error, in either x or y, of at least one pixel, whereas none of the 50 GCP used in determining

Table 3.3: Transformation Coefficients for TM93 Image

TM89 Transformation Coefficients		
	X'	Y'
constant	-308.462478824	-929.109078907
$\mathbf{X}$	-0.105348189	0.541069146
$\mathbf{Y}$	-0.000011904	0.000008910
$\mathbf{X^2}$	0.000005822	-0.000012048
$\mathbf{X \cdot Y}$	-0.000012003	0.000025712
$\mathbf{Y^2}$	0.000000000	-0.000000002
$\mathbf{X^3}$	0.000000004	0.000000001
$\mathbf{X^2 \cdot Y}$	-0.000000006	0.000000002
$\mathbf{X \cdot Y^2}$	0.000000005	-0.000000002
$\mathbf{Y^3}$	0.000000004	-0.000000005
RMS Residuals		
$\mathbf{X, Y}$ RMS	0.440055	0.5198876
Total RMS	0.681116	

Table 3.4: Errors of the Independent Test Points - TM93 Image

GCP Number		X error	Y error
GCP	#59	0.0310	-1.5760
GCP	#60	1.5840	0.5630
GCP	#61	-3.0980	1.5110
GCP	#62	0.0770	-1.1580
GCP	#63	1.3540	-0.5960
GCP	#64	2.7750	0.0580
GCP	#65	2.0430	0.4110
GCP	#66	-1.0760	0.7410
GCP	#67	-2.0320	-0.5820
GCP	#68	1.0510	0.6870
$\mathbf{X, Y}$ RMS		1.7948	0.9117
Total RMS			2.0131

the mapping polynomial had a residual of one pixel (or more), in either x or y.

The rectified TM93 image is shown at Figure 3.3.

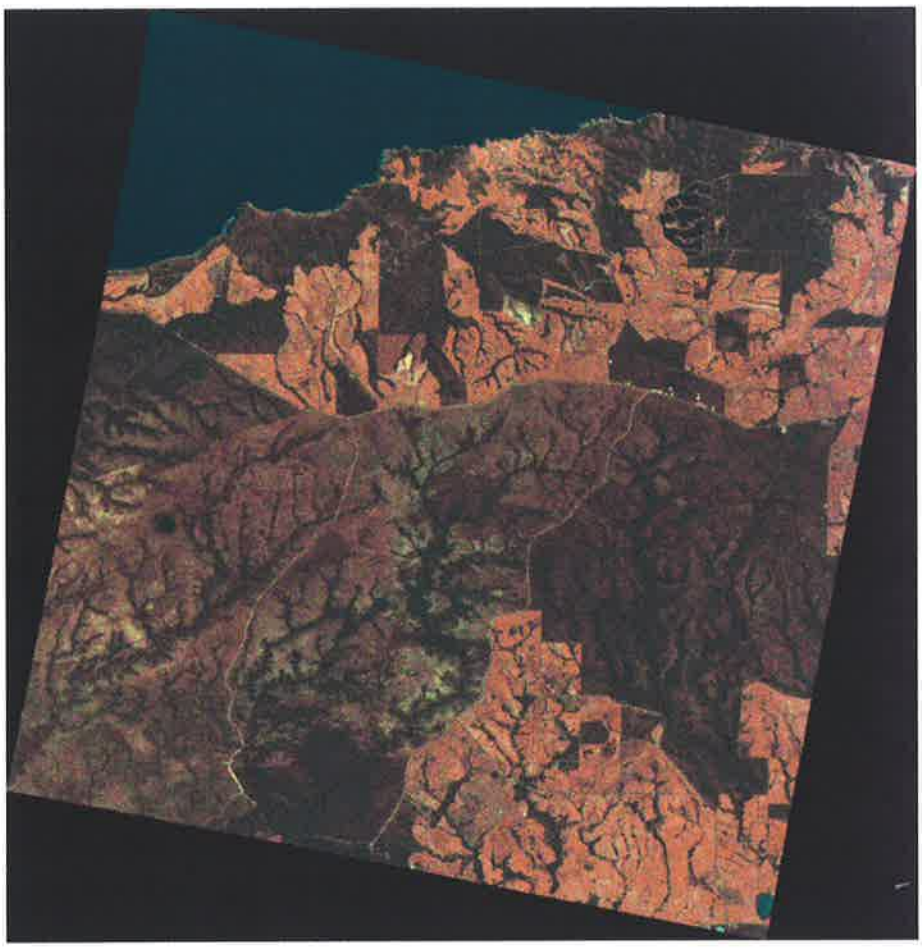

Figure 3.3: Rectified TM93 Image (Bands 5,3,2 as RGB)

3.5 Training Class Data

The same training class data were used for all supervised classifications (of any particular image). A number of regions were selected in each image for each training class by manual inspection of each image and consultation with the reference class map described earlier. The regions did not include boundary areas (as shown in the reference data): no attempt was made to deliberately include mixed pixels in the training data. No particular care was taken to ensure that the regions contained the

full range of spectral values for each of the desired classes. The data from the selected regions were grouped to form the training class data for each class. The classes of primary interest were forest, sea, pasture and pine. The clusters that resulted from unsupervised classifications were mapped, to the extent possible, onto these same classes.

3.5.1 TM89 Training Class Data

The presence of cloud in the 1989 image required the inclusion of an additional two classes: cloud and shadow.

The training class information for the TM89 image is shown in the following tables. Table 3.5 shows band means, and the number of training pixels, for each class. Tables 3.6, 3.7, 3.8, 3.9, 3.10 and 3.11 show the class covariance matrices.

Table 3.5: Band Means and Number of Training Pixels, For Each Class, TM89

	Band Means				
class	Band 2	Band 3	Band 4	Band 5	# training pixels
forest	24.6	27.1	52.6	47.7	39970
sea	18.2	15.3	5.9	6.0	24202
pasture	44.2	73.7	63.7	168.7	11270
pine	21.6	19.9	80.6	30.2	3070
cloud	147.6	211.5	154.8	224.2	5158
shadow	17.6	16.3	15.5	14.2	21689

Table 3.6: Class Covariance Matrix, Forest, TM89

	Band 2	Band 3	Band 4	Band 5
Band 2	1.287	1.533	0.935	3.997
Band 3	1.533	3.771	0.052	8.826
Band 4	0.935	0.052	26.519	-9.136
Band 5	3.997	8.826	-9.136	39.048

Table 3.7: Class Covariance Matrix, Sea, TM89

	Band 2	Band 3	Band 4	Band 5
Band 2	1.219	0.709	0.315	0.695
Band 3	0.709	1.558	0.414	0.847
Band 4	0.315	0.414	0.566	0.523
Band 5	0.695	0.847	0.523	2.382

Table 3.8: Class Covariance Matrix, Pasture, TM89

	Band 2	Band 3	Band 4	Band 5
Band 2	18.958	35.184	16.395	12.800
Band 3	35.184	73.222	41.473	43.220
Band 4	16.395	41.473	41.979	63.936
Band 5	12.800	43.220	63.936	174.629

Table 3.9: Class Covariance Matrix, Pine, TM89

	Band 2	Band 3	Band 4	Band 5
Band 2	0.937	0.967	-0.691	2.151
Band 3	0.967	2.736	-3.403	4.467
Band 4	-0.691	-3.403	30.531	-7.109
Band 5	2.151	4.467	-7.109	14.214

Table 3.10: Class Covariance Matrix, Cloud, TM89

	Band 2	Band 3	Band 4	Band 5
Band 2	1219.12	1399.70	1195.35	1022.85
Band 3	1399.70	1802.82	1401.59	1351.82
Band 4	1195.35	1401.59	1204.48	1085.58
Band 5	1022.85	1351.82	1085.58	1236.16

3.5.2 TM93 Training Class Data

The training class information for the TM93 image is shown in the following tables. Table 3.12 shows band means, and the number of training pixels, for each class.

Table 3.11: Class Covariance Matrix, Shadow, TM89

	Band 2	Band 3	Band 4	Band 5
Band 2	0.999	0.772	0.437	0.834
Band 3	0.772	2.191	0.729	1.760
Band 4	0.437	0.729	3.428	1.459
Band 5	0.834	1.760	1.459	5.624

Tables 3.13, 3.14, 3.15 and 3.16 show the class covariance matrices.

Table 3.12: Band Means and Number of Training Pixels, For Each Class, TM93

	Band Means				
class	Band 2	Band 3	Band 4	Band 5	# training pixels
forest	24.9	30.9	56.4	55.2	137803
sea	17.5	15.5	6.0	4.9	31590
pasture	36.5	61.3	62.8	163.0	5007
pine	20.1	19.6	78.4	28.5	7018

Table 3.13: Class Covariance Matrix, Forest, TM93

	Band 2	Band 3	Band 4	Band 5
Band 2	13.916	26.085	-1.517	55.725
Band 3	26.085	51.743	-4.840	108.989
Band 4	-1.517	-4.840	47.052	-28.628
Band 5	55.725	108.989	-28.628	280.803

Table 3.14: Class Covariance Matrix, Sea, TM93

	Band 2	Band 3	Band 4	Band 5
Band 2	0.798	0.087	0.07098	-0.00589
Band 3	0.087	1.008	0.21589	-0.04924
Band 4	0.071	0.216	1.030	-0.012
Band 5	-0.006	-0.049	-0.012	1.222

Table 3.15: Class Covariance Matrix, Pasture, TM93

	Band 2	Band 3	Band 4	Band 5
Band 2	8.509	19.192	5.878	10.012
Band 3	19.192	49.610	7.618	27.312
Band 4	5.878	7.618	33.590	-3.658
Band 5	10.012	27.312	-3.658	43.414

Table 3.16: Class Covariance Matrix, Pine, TM93

	Band 2	Band 3	Band 4	Band 5
Band 2	0.442	0.241	0.414	0.328
Band 3	0.241	1.733	-0.319	0.892
Band 4	0.414	-0.319	19.646	-1.029
Band 5	0.328	0.892	-1.029	5.272

Chapter 4

CRISP CLASSIFICATION

Summary: This chapter presents a quantitative comparison of supervised and unsupervised approaches to traditional classification. The purpose is twofold. First, we need to implement a crisp classifier in order to undertake crisp Post Classification Comparison. It is appropriate that we use a comparatively good classifier for this purpose. It was therefore considered necessary to investigate various candidate classifiers. Second, it was felt that an examination of crisp classifiers would provide some insight into both the form of a fuzzy classifier, and aspects of parameter inclusion and selection. Both supervised and unsupervised approaches to classification are examined. Not surprisingly, it is concluded that the supervised approaches are more suited to this type of application, when the classes of interest are known a priori. Of the supervised classification approaches, it was found that the full Maximum Likelihood classifier, with terms for a priori class probability and the determinant of the class covariance matrices, offered no appreciable improvement over a simple Mahalanobis Distance classifier. The nature of the logical process underlying classification is examined, and the use of spatial context is briefly discussed.

4.1 General

Approaches to classification may be divided into two broad methodologies: supervised and unsupervised (Townshend, 1981). In the supervised approach, the desired classes are decided upon *a priori*, and the feature space partitioned (in some manner) on the basis of labeled training data. In the unsupervised approach, patterns in the data are sought to define the classes, and then labels are applied *a posteriori*. Assumptions about the underlying data model may be necessary for particular algorithms in both approaches, and are almost universally required, in one form or another, in the supervised approaches. The main problems encountered in the implementation of supervised methods relate to the validity of the underlying data model, and that the data may simply not be separable, in the available feature space, into the desired classes. Unsupervised methods, on the other hand, are computationally complex, the number of clusters normally needs to be defined *a priori*, and the clusters determined may not align with 'meaningful' categorisations of the data i.e. categorisations which suit the purposes of the analysis. Hybrids of the two basic methodologies are, of course, possible.

The objective of this chapter is to compare some common approaches to supervised and unsupervised classification, using data from the Kangaroo Island study area. This will provide the best traditional classifier to use in change detection studies of this area. It will also provide guidance on approaches to fuzzy classification, and the importance of certain parameters in the fuzzy classification approaches described later in this thesis.

4.2 Supervised Classification

Some standard approaches to supervised classification include decision rules based on some minimum distance (e.g. Euclidean, diagonal or Mahalanobis) from training class means, or on maximum *a posteriori* probability. The relationship between these approaches can be seen, to some extent, from the development of the discriminant functions in Duda and Hart (1973), or Richards (1993). It is, in summary, as follows:

Let w_i represent the (spectral) classes in an image. Represent the column vector of brightness values for a particular pixel in an image as x. Classify the pixel to the class w_i according to

$$x \in w_i \text{ if } p(w_i|\mathbf{x}) > p(w_j|\mathbf{x}) \quad \forall i \neq j \tag{4.1}$$

(i.e. select the class of the greatest conditional probability, given the observed data).

Of course, we generally do not know $p(w_i|\mathbf{x})$: we are able, however, to estimate $p(\mathbf{x}|w_i)$ from training data (where $p(\mathbf{x}|w_i)$ is the probability of $\mathbf{x}$, given w_i). The term $p(\mathbf{x}|w_i)$ represents the probability of obtaining the particular (spectral) response, given that we are observing a pixel of known class w_i.

Using Bayes Rule, we can relate $p(\mathbf{x}|w_i)$ and $p(w_i|\mathbf{x})$

$$p(w_i|\mathbf{x}) = \frac{p(\mathbf{x}|w_i) \cdot p(w_i)}{\sum_{j=1}^{c}(p(\mathbf{x}|w_j) \cdot p(w_j))} \tag{4.2}$$

where $p(w_i)$ is the (unconditional) prior probability of the class w_i occurring, and $\sum_{j=1}^{c}$ represents summation over all classes.

Because we wish only to select the largest of these posterior probabilities, and the

divisor ($\sum_{j=1}^{c}(p(\mathbf{x}|w_j)\cdot p(w_j))$) is a common factor, we can ignore it. After taking logs for convenience, we see that the minimum error rate classifier (under the assumption that the 'cost' of all errors are identical) can be found from the discriminant function

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|w_i) + \ln p(w_i) \tag{4.3}$$

If the densities $p(\mathbf{x}|w_i)$ are multi-variate normally distributed (or if we choose to treat them as such), we can describe a discriminant function

$$g_i(\mathbf{x}) = -\frac{1}{2}\left\{(\mathbf{x}-\mu_i)^t\ \Sigma_i^{-1}(\mathbf{x}-\mu_i)\right\} - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\det\Sigma_i + \ln p(w_i) \tag{4.4}$$

where μ_i is the (vector) mean of class i, the superscript t defines vector (or matrix) transpose, Σ_i^{-1} is the inverse of the covariance matrix of class i, d is the number of spectral bands, ln is the natural logarithm, and $\det\Sigma_i$ is the determinant of the covariance matrix of class i.

Clearly, the second term ($\frac{d}{2}\ln 2\pi$) can be dropped (as it is common to all classes). This leaves

$$g_i(\mathbf{x}) = -\frac{1}{2}\left\{(\mathbf{x}-\mu_i)^t\ \Sigma_i^{-1}(\mathbf{x}-\mu_i)\right\} - \frac{1}{2}\ln\det\Sigma_i + \ln p(w_i) \tag{4.5}$$

We wish to select the class for which this expression is maximal. After noting that minimising $-g(\mathbf{x})$ gives the same result as maximising $g(\mathbf{x})$, some common supervised classifiers can be implemented by selecting the class according to the following rules:

The Minimum Euclidean Distance Classifier (which is appropriate for uncorrelated variables, with similar scales)

$$\min_c \left\{ (\mathbf{x} - \mu_c)^t \; I \; (\mathbf{x} - \mu_c) \right\} \tag{4.6}$$

where $\min_c$ represents the minimum over the classes, and I is the Identity Matrix. It is not, of course, strictly necessary to include I: it has been included only to show the similarity between the common supervised classifiers.

The Minimum Diagonal Distance Classifier (appropriate for uncorrelated variables, different scales)

$$\min_c \left\{ (\mathbf{x} - \mu_c)^t \; D^{-1} \; (\mathbf{x} - \mu_c) \right\} \tag{4.7}$$

(where D^{-1} is inverse of the diagonal matrix of class variances).

The Minimum Mahalanobis Distance Classifier (appropriate for correlated variables, different scales)

$$\min_c \left\{ (\mathbf{x} - \mu_c)^t \; \Sigma_c^{-1} \; (\mathbf{x} - \mu_c) \right\} \tag{4.8}$$

(where Σ_c^{-1} is the inverse of the covariance matrix of class c, as before).

The Maximum *a posteriori* Probability Classifier, without priors (i.e. not taking prior class probabilities into account)

$$\min_c \left\{ (\mathbf{x} - \mu_c)^t \; \Sigma_c^{-1} \; (\mathbf{x} - \mu_c) + \ln \det \Sigma_c \right\} \tag{4.9}$$

The Maximum *a posteriori* Probability Classifier, with (unequal) priors (i.e. taking global prior class probabilities into account)

$$\min_c \left\{ (\mathbf{x} - \mu_c)^t \; \Sigma_c^{-1} \; (\mathbf{x} - \mu_c) + \ln \det \Sigma_c - 2 \ln p(w_c) \right\} \tag{4.10}$$

We would expect to observe classifier accuracy increasing down this list, as each successive classifier uses one more piece of (seemingly) relevant information.

4.3 Unsupervised Classification

The standard approach to unsupervised classification is the use of the k-means clustering algorithm, or ISODATA (Iterative Self-Organizing Data Analysis technique) (Tou and Gonzalez, 1974). The ISODATA Algorithm (simplified) is:

1. Choose the number of and initial values for class means.

   Repeat Steps 2. and 3. until some stopping criterion is reached:
2. Assign all samples to the class of the 'closest' mean (where closest is determined by some distance metric).
3. Recompute means from the samples now in each class.

4.4 Materials

The Landsat TM 1993 image and ground truth data used in this chapter were described in Chapter 3.

4.5 Results

Classifications were undertaken using the Minimum Euclidean Distance Classifier, the Minimum Mahalanobis Distance Classifier, the Maximum *a posteriori* Probability (or Maximum Likelihood) Classifier without prior probabilities, the Maximum Likelihood Classifier with prior probabilities, and the unsupervised ISODATA Classifier. The prior probabilities used were determined by the actual proportion of that class in the total ground truth. This is clearly an ideal situation, and we would be unlikely to

be able to achieve this sort of accuracy in the estimation of global priors in practice. The Minimum Diagonal Distance Classifier was not investigated.

As observed earlier, the unsupervised k-means (ISODATA) classifier (Ball and Hall, 1967) requires the number of desired clusters to be specified. Options from four to twelve inclusive were tested, with nine returning the best results in terms of alignment of the resultant spectral clusters with the desired ground cover classes. ISODATA was unable to separate the class pine without introducing significant misclassification of forest as pine. The results shown below for ISODATA therefore do not include the class pine.

The ISODATA classifier produced a cluster that did not correspond closely to any of the desired ground cover classes: this cluster was labelled miscellaneous ("misc"). There are therefore two accuracy assessment tables for ISODATA: one with the class "misc" being shown as a separate class, and one with it being grouped with pasture (the class with which it most closely aligned).

4.5.1 Accuracy Assessment Tables

The accuracy assessment confusion matrices are shown in Tables 4.1 to 4.6. The classifier results are shown across the rows; the ground truth data are shown down the columns.

These tables reveal that the Minimum Euclidean Distance Classifier performs relatively poorly; it misclassifies a significant amount of forest as pine. It is not possible to draw statistically significant inferences about the relative accuracies of

Table 4.1: Confusion Matrix: Minimum Euclidean Distance Classifier

	forest	sea	roads	pasture	swamp	lake	pine	total	%
forest	3891116	6921	24378	213813	1492	4551	17189	4159460	93.55
sea	2179	663084	0	648	855	3914	0	670680	98.87
pasture	204063	542	14358	1227961	0	199	765	1447888	84.81
pine	196633	0	406	2338	0	52	17518	216947	8.07
total	4293991	670547	39142	1444760	2347	8716	35472	6494975	
%	90.62	98.89	0.00	84.99	0.00	0.00	49.39		89.29

Table 4.2: Confusion Matrix: Minimum Mahalanobis Distance Classifier

	forest	sea	roads	pasture	swamp	lake	pine	total	%
forest	4235779	13281	31150	337328	2347	8681	21498	4650064	91.09
sea	254	657243	0	81	0	0	0	657578	99.95
pasture	57896	23	7980	1107220	0	35	370	1173524	94.35
pine	62	0	12	131	0	0	13604	13809	98.52
total	4293991	670547	39142	1444760	2347	8716	35472	6494975	
%	98.64	98.02	0.00	76.64	0.00	0.00	38.35		92.59

Table 4.3: Confusion Matrix: Maximum Likelihood Classifier

	forest	sea	roads	pasture	swamp	lake	pine	total	%
forest	4234464	13032	31089	335047	2347	8669	20483	4645131	91.16
sea	267	657492	0	87	0	0	0	657846	99.95
pasture	58500	23	8034	1109446	0	47	370	1176420	94.31
pine	760	0	19	180	0	0	14619	15578	93.84
total	4293991	670547	39142	1444760	2347	8716	35472	6494975	
%	98.61	98.05	0.00	76.79	0.00	0.00	41.21		92.63

Table 4.4: Confusion Matrix: Maximum Likelihood Classifier (With Priors)

	forest	sea	roads	pasture	swamp	lake	pine	total	%
forest	4236770	13069	31211	340769	2347	8682	22181	4655029	91.01
sea	255	657455	0	87	0	0	0	657797	99.95
pasture	56953	23	7920	1103809	0	34	358	1169097	94.42
pine	13	0	11	95	0	0	12933	13052	99.09
total	4293991	670547	39142	1444760	2347	8716	35472	6494975	
%	98.67	98.05	0.00	76.40	0.00	0.00	36.46		92.55

Table 4.5: Confusion Matrix: Unsupervised (ISODATA) Classifier

	forest	sea	roads	pasture	swamp	lake	pine	total	%
forest	4000686	7667	18809	118892	1851	5020	33297	4186222	95.57
sea	1396	662417	0	485	496	3379	0	668173	99.14
pasture	186048	371	14317	1222971	0	181	785	1424673	85.84
misc	105861	92	6016	102412	0	136	1390	215907	0.00
total	4293991	670547	39142	1444760	2347	8716	35472	6494975	
%	93.17	98.79	0.00	84.65	0.00	0.00	0.00		90.63

Table 4.6: Confusion Matrix: Unsupervised (ISODATA) Classifier (Grouped Pasture and Miscellaneous Classes)

	forest	sea	roads	pasture	swamp	lake	pine	total	%
forest	4000686	7667	18809	118892	1851	5020	33297	4186222	95.57
sea	1396	662417	0	485	496	3379	0	668173	99.14
pasture	291909	463	20333	1325383	0	317	2175	1640580	80.79
total	4293991	670547	39142	1444760	2347	8716	35472	6494975	
%	93.17	98.79	0.00	91.74	0.00	0.00	0.00		92.20

the other supervised classifiers.

In comparison between the grouped ISODATA and the better supervised classifiers, the results are somewhat mixed. The overall classification accuracy is highly comparable. In terms of commission errors, ISODATA performed better on forest, comparably on sea, significantly poorer on pasture, and did not have a pine class at all. In terms of omission errors, ISODATA performed poorer on forest, comparably on sea, significantly better on pasture, and, again, did not have a pine class.

4.5.2 Classified Images

The Minimum Euclidean Distance classified image is shown in Figure 4.1. At the resolution that can be displayed in printing, the Minimum Mahalanobis Distance,

Maximum Likelihood, and Maximum Likelihood (with priors) classifiers produce indistinguishable images. Accordingly, only one is shown: Figure 4.2 is the Minimum Mahalanobis Distance classified image. The legend for Figure 4.2 is the same as for Figure 4.1.

The ISODATA classified image is shown in Figure 4.3. The legend for this figure is shown in Table 4.7. ISODATA took considerably more effort than the supervised classifiers (principally, in determining the optimal number of clusters), and was unable to satisfactorily discriminate one of the main classes on interest: pine. It is concluded that, for this application, the better supervised classifiers were superior. The remainder of this chapter with deals only with supervised classifiers.

4.6 Discussion of Supervised Classifier Performance

As noted earlier, the Minimum Euclidean Distance Classifier is clearly inferior to the other three supervised classifiers examined: the Minimum Mahalanobis Distance Classifier (MAH), the Maximum Likelihood Classifier (MAX), and the Maximum Likelihood Classifier (with priors) (PMAX).

The earlier tables reveal percentage error, overall, and by commission and omission. It is also of interest to see how the various classifiers compare for individual pixels, by examining the "per pixel" classifications. That is, in a pair-wise comparison, irrespective of whether either, both or neither classifier was correct, how often, and in what manner, did they differ? Results are shown for the pairwise comparisons: Minimum Mahalanobis Distance Classifier and the Maximum Likelihood Classifier (Table 4.8); the Maximum Likelihood Classifier and the Maximum Likelihood Classifier (with priors) (Table 4.9); and the Minimum Mahalanobis Distance Classifier

and the Maximum Likelihood Classifier (with priors) (Table 4.10).

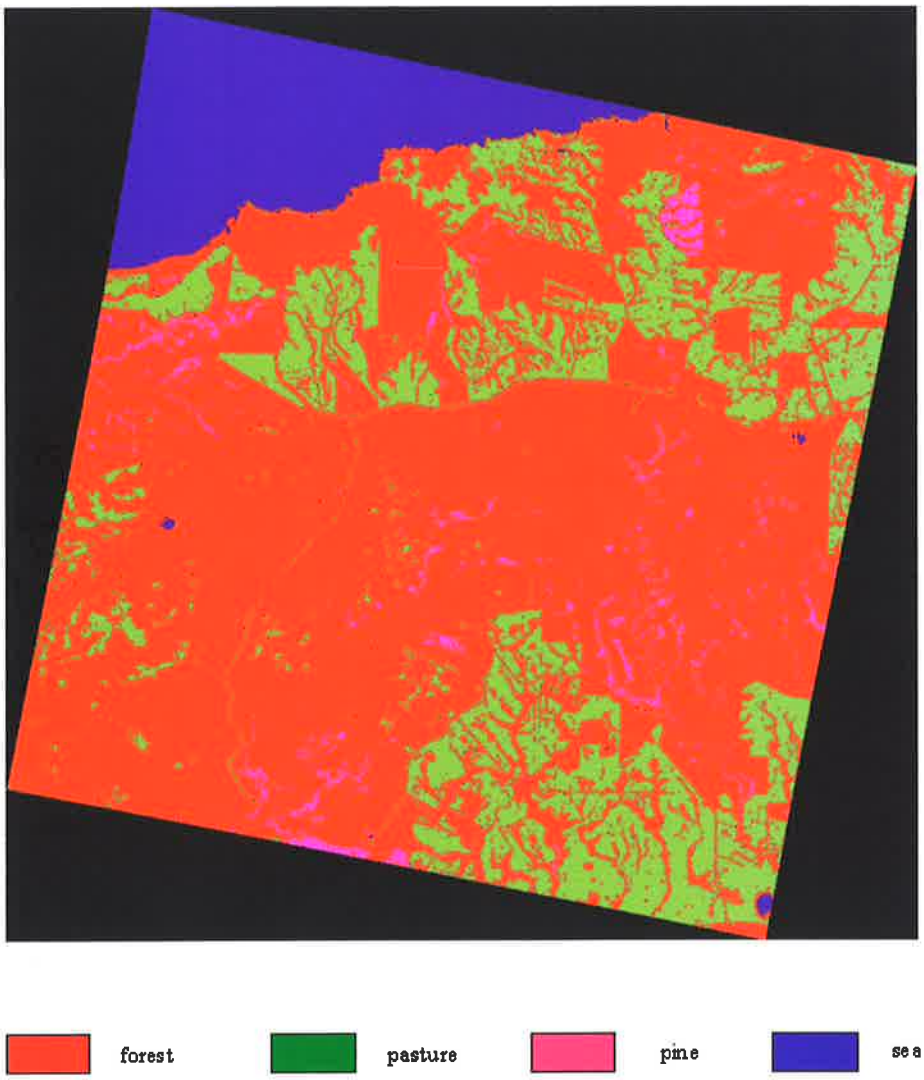

Figure 4.1: Minimum Euclidean Distance Classified TM93 Image.

The Minimum Mahalanobis Distance Classifier and the Maximum Likelihood Classifier differ only on 4,933 pixels in total (out of 6,494,975, or 0.076%). All of these differences were in the class forest, and all that were classified by the Minimum Mahalanobis Distance Classifier as forest (a subtotal of 4,650,064 pixels)

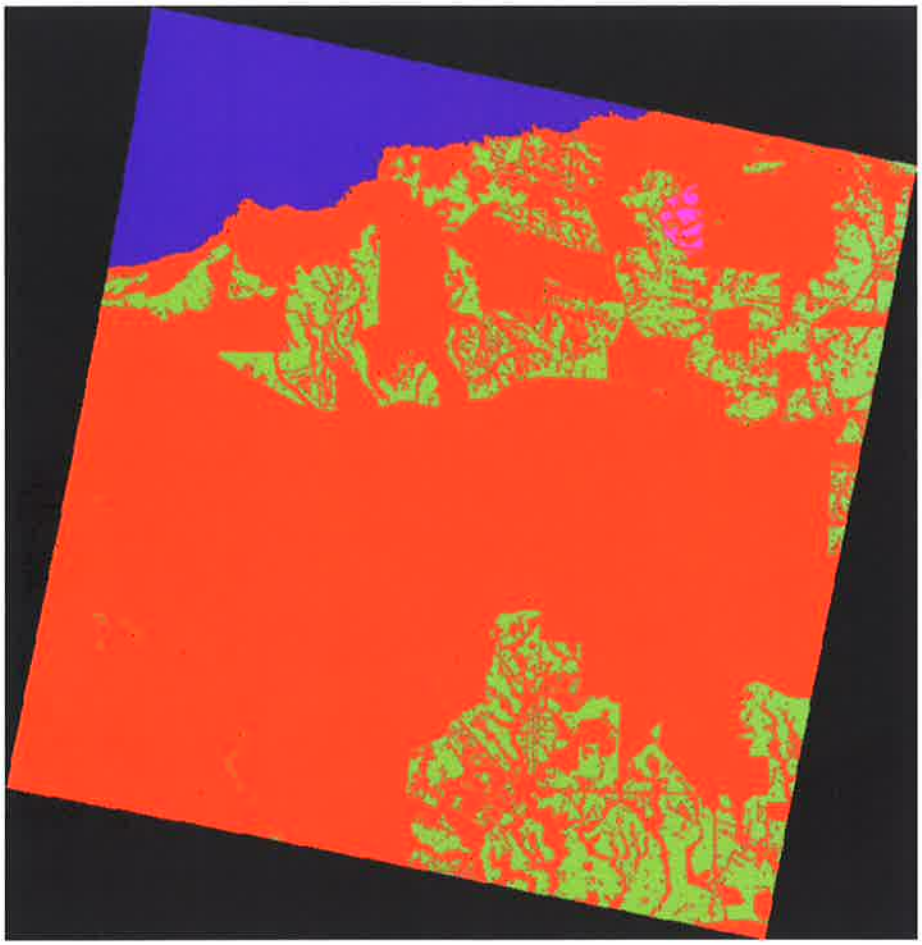

Figure 4.2: Minimum Mahalanobis Distance Classified TM93 Image.

Table 4.7: Legend for the Unsupervised (ISODATA) Classified TM93 Image.

Class	Class Name	Colour	Class after grouping
0	Unclassified	Black	unclassified
1	Sea	Dark Blue	sea
2	Forest/Native Vegetation 1	Dark Brown	forest
3	Forest/Native Vegetation 2	Burgundy	forest
4	Forest/Native Vegetation 3	Light Brown	forest
5	Forest/Native Vegetation 4	Red	forest
6	Forest/Native Vegetation 5	Pink	forest
7	Miscellaneous	Yellow	pasture
8	Pasture 1	Light Green	pasture
9	Pasture 2	Dark Green	pasture

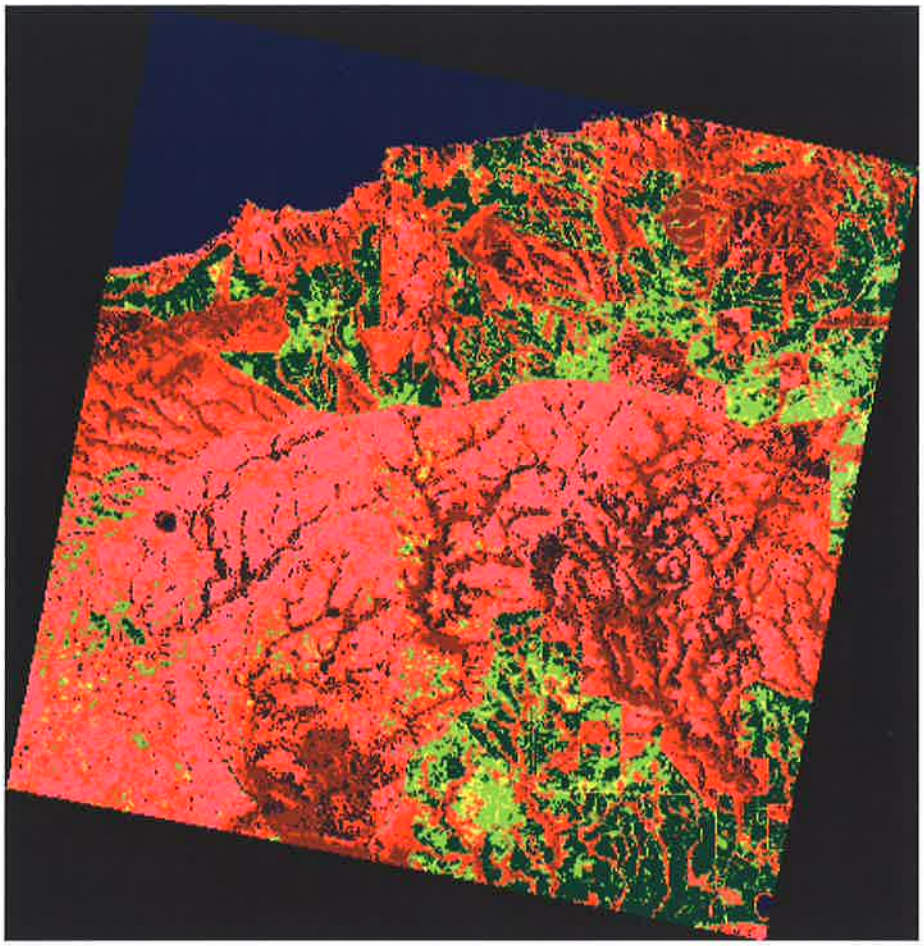

Figure 4.3: Unsupervised (ISODATA) Classified TM93 Image.

Table 4.8: Comparison of Assigned Class Labels - Mahalanobis Distance versus Maximum Likelihood

		MAX				
		forest	sea	pasture	pine	Total
MAH	forest	4645131	268	2896	1769	4650064
	sea	0	657578	0	0	657578
	pasture	0	0	1173524	0	1173524
	pine	0	0	0	13809	13809
	Total	4645131	657846	1176420	15578	6494975

were classified by the Maximum Likelihood Classifier as something else. This is a percentage 'error' of 0.106%.

Table 4.9: Comparison of Assigned Class Labels - Maximum Likelihood (With Priors) versus Maximum Likelihood

		MAX				
		forest	sea	pasture	pine	Total
PMAX	forest	4645131	49	7323	2526	4655029
	sea	0	657797	0	0	657797
	pasture	0	0	1169097	0	1169097
	pine	0	0	0	13052	13052
	Total	4645131	657846	1176420	15578	6494975

The Maximum Likelihood Classifier (with priors) and the Maximum Likelihood Classifier differ only on 9,898 pixels in total (out of 6,494,975, or 0.152%). All of these differences were in the class forest, and all that were classified by the Maximum Likelihood Classifier (with priors) as forest (a subtotal of 4,655,029 pixels) were classified by the Maximum Likelihood Classifier as something else. This is a percentage 'error' of 0.213%.

Table 4.10: Comparison of Assigned Class Labels - Maximum Likelihood (With Priors) versus Mahalanobis Distance

		MAH				
	forest	sea	pasture	pine	Total	
PMAX	forest	4649845	0	4427	757	4655029
	sea	219	657578	0	0	657797
	pasture	0	0	1169097	0	1169097
	pine	0	0	0	13052	13052
	Total	4650064	657578	1173524	13809	6494975

The Maximum Likelihood Classifier (with priors) and the Minimum Mahalanobis Distance Classifier differ only on 5,403 pixels in total (out of 6,494,975, or 0.083%). The majority of these differences were in the class forest (5,184 pixels). The percentage 'error' of the Maximum Likelihood Classifier (with priors) versus the Minimum Mahalanobis Distance Classifier for forest was 0.080%; the percentage

'error' of the Maximum Likelihood Classifier (with priors) versus the Minimum Mahalanobis Distance Classifier for sea was 0.033%.

Where these classifiers differ spatially is shown in the following pairwise comparison images. Where two classifiers differ is shown in white. Whether either is, in fact, correct, is not shown. Figure 4.4 shows where the Minimum Mahalanobis Distance Classifier and the Maximum Likelihood Classifier differ, Figure 4.5 shows where the Minimum Mahalanobis Distance Classifier and the Maximum Likelihood Classifier (with priors) differ, and Figure 4.6 shows where the Maximum Likelihood Classifier and the Maximum Likelihood Classifier (with priors) differ. It was not possible to draw any significant inferences from the spatial distribution of these differences.

4.6.1 Classification Errors

Of course, 'ground truth' is never perfect. There are often errors and inaccuracies in both the class assignment and the mapping accuracy, and there can be 'temporal misalignment' (i.e. the map is produced from data of various ages, but is usually interpreted to be correct as at its publication date). Also, there are invariably misregistrations between images and 'ground truth'. This is an apparent cause of 'error'.

It is therefore instructive to observe the spatial nature of the 'errors'. Using the Minimum Mahalanobis Distance classified image as representative of the better supervised classifiers, the errors of commission (class forest) are shown in Figure 4.7 and the errors of omission (class forest) are shown in Figure 4.8.

Figure 4.4: Where the Minimum Mahalanobis Distance and Maximum Likelihood Classifiers Differ Spatially (shown in white)

It is noted that there are some relatively large regions of error in these two images. It is suggested that the majority of these can be attributed to one of the following three causes. First, the ground truth has a number of classes (e.g. lakes) that the classified image does not have. These classes may therefore show as errors (of commission in the class forest), in that they were assigned to the class forest by the classifier, whereas they actually belonged to another class. Second, the 'ground truth' is wrong, or at least not temporally coincident with the imagery. Some regions show as errors (of omission from the class forest), because conversion to pasture had occurred by 1993,

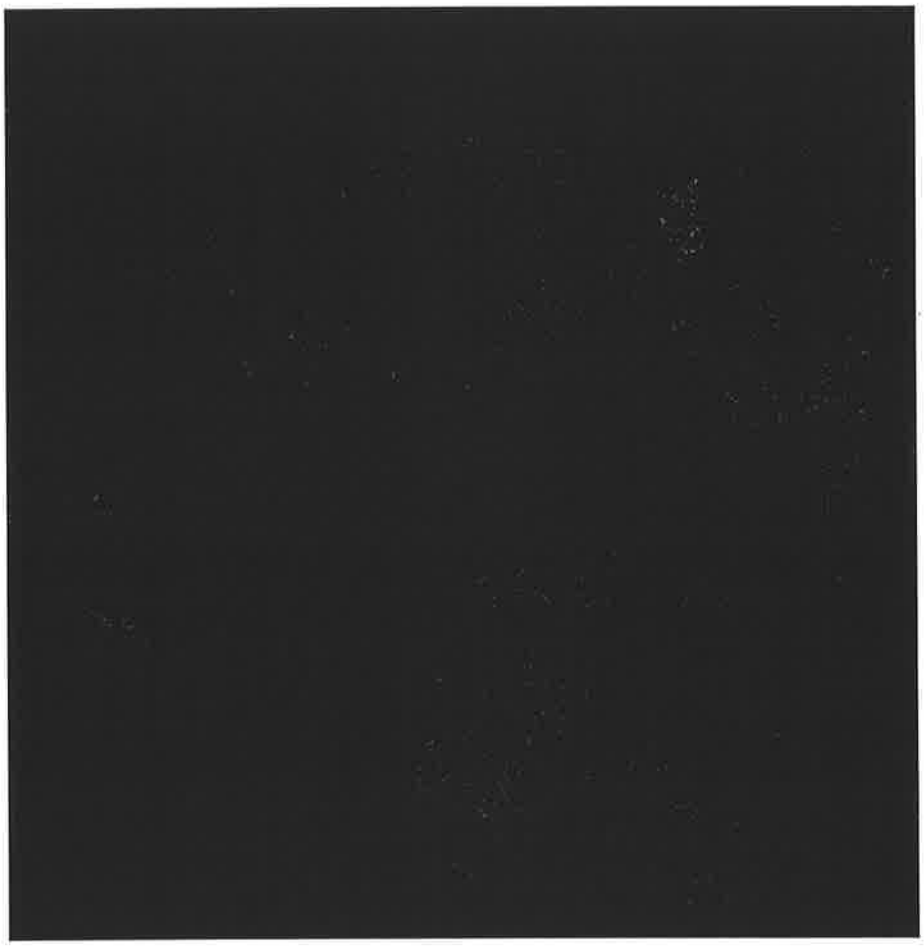

Figure 4.5: Where the Minimum Mahalanobis Distance and Maximum Likelihood (With Priors) Classifiers Differ Spatially (shown in white)

but the ground truth still showed them as forest. Third, there is a substantial area of immature pine. This appeared spectrally more similar to forest than to pine, as there was considerable regrowth of native vegetation in the area, and there was insufficient pine to dominate the spectral response. This cause of error is the only one discussed thus far which can be viewed as a failure, albeit minor, of the classifier.

It is anticipated that, in some cases, the classifier was actually wrong. No field work was undertaken to confirm this, or determine potential causes of error for these cases, but it is speculated that some at least could be attributed to the issue of class

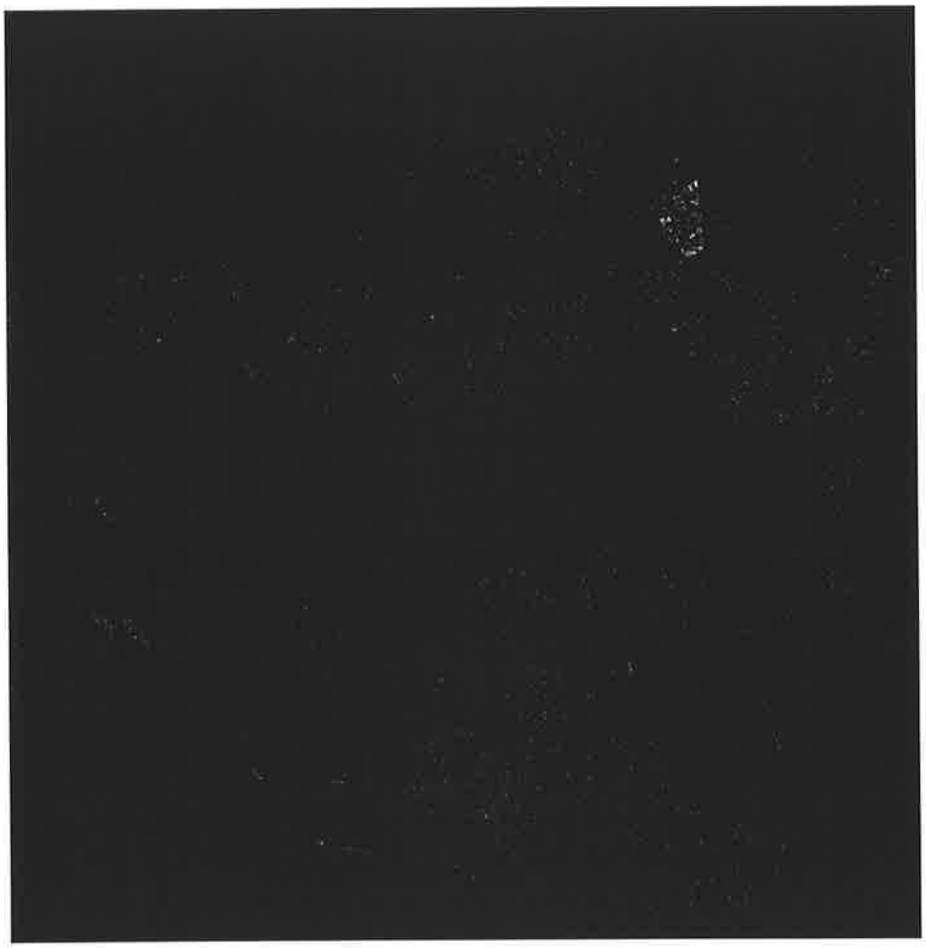

Figure 4.6: Where the Maximum Likelihood and the Maximum Likelihood (With Priors) Classifiers Differ Spatially (shown in white)

definition and the concept of 'fuzzy' classes. It is noted that a pixel with 20% tree canopy cover may be classified by a human expert in the field as forest (Anderson *et al.*, 1976, suggests 10% or greater canopy cover should be classified as forest), but there may be a substantial contribution to its spectral signature by the same plant species as comprise a pixel of the class pasture. Similarly, a pixel of pasture may well contain one or more trees. There may therefore be a spectral contribution from trees (and their shadows).

It is also noted that the figures reveal error artifacts caused by misregistration

Figure 4.7: Errors of Commission in the Minimum Mahalanobis Distance Classified TM93 Image, Class Forest

between the TM image and the ground truth image. These are apparent on or near the boundary of two classes in the ground truth image. We would like to adopt a filtering technique that removes these artifacts, but leaves the larger regions unchanged. A class of filters known as morphological filters offer potential to achieve this.

4.7 Morphological Filtering

As its name suggests, morphological filtering provides an approach to the processing of digital images that is based on shape. Morphological filters are non-linear filters,

Figure 4.8: Errors of Omission in the Minimum Mahalanobis Distance Classified TM93 Image, Class Forest

built by concatenations of primitive operations termed dilations and erosions. The primitive operations operate over some particular neighbourhood of a pixel in an image. The neighbourhood is defined by what is termed the structuring element. This can be of arbitrary shape. In a dilation, the pixel value in the new (processed) binary image is set to 1 if any pixel in the particular neighbourhood, as defined by the structuring element, is 1. Hence, this operation is sometimes termed a "hit" transform: the structuring element "hit" at least one pixel in the image being processed. In an erosion, the pixel value in the new (processed) binary image is set to

0 if any pixel in the particular neighbourhood, as defined by the structuring element, is 0. Hence, this operation is sometimes termed a "miss" transform: the structuring element "missed" at least one pixel in the image being processed.

Detailed early descriptions of morphological filtering of binary images can be found in Matheron (1975) and Serra (1982). Binary morphology can be extended to greyscale (or colour) morphology (Sternberg, 1986). A dilation (erosion) then essentially determines the pixel value in the new (processed) image as the largest (smallest) value of the specific neighbourhood of that pixel, as defined by the structuring element. It is noted that binary morphology can be defined in the same way, and, in practice, binary and greyscale (or colour) morphology can be computed by this same algorithm.

More complex morphological filters are then built by concatenations of dilations and erosions. An erosion followed by a dilation is termed an 'opening' (the opening of the image I with the structuring element B). This, generally speaking, has the effect of removing "small islands" of data (Serra and Vincent, 1992), while leaving larger regions relatively unaffected.

A detailed discussion of image analysis using morphological filters is in Haralick *et al.* (1987), and some important developments are described in Maragos and Schafer (1990). An comparatively recent overview of morphological filters can be found in Serra and Vincent (1992). Wang *et al.* (1992) compares the performance and properties of morphological, median and various mean filters.

The morphological filter chosen for this work was an 'opening'. It was expected that, in addition to removing small islands of data, such a filter, with an appropriate structuring element, would also remove the linear error features in a classified image caused by image to ground truth misregistration (and, for later work, would similarly

remove change errors caused by image to image misregistration). Various structuring elements were tested; the final choice was a 7 by 7 diamond shaped structuring element, shown in Figure 4.9.

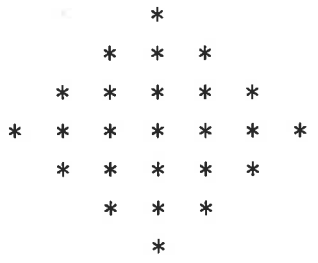

Figure 4.9: 7 by 7 Diamond Shaped Structuring Element

Visually, the effect of a morphological opening filter on the classification error (binary) images of the Mahalanobis Distance Classifier are shown in Figures 4.10 (Errors of Commission) and 4.11 (Errors of Omission).

The Minimum Mahalanobis Distance Classifier exhibited 414,285 errors of commission in the class forest (i.e. pixels that the classifier classed as forest, but were, according to the ground truth file, some other class). After applying a morphological opening filter with a 7 by 7 diamond shaped structuring element, the errors of commission (in the class forest) reduced to 189,527, approximately 46% of the original error rate.

There were 58,212 errors of omission in the class forest (i.e. pixels which were, according to the ground truth file, forest, but which the classifier classed as some other class). After applying a morphological opening filter with a 7 by 7 diamond shaped structuring element, the errors of omission (in the class forest) reduced to 15,083, approximately 26% of the original error rate.

Figure 4.10: Errors of Commission in the Minimum Mahalanobis Distance Classified TM93 Image, Class Forest, after Morphological Filtering

This shows the utility of the morphological opening filter to remove linear 'errors' in binary images while leaving the larger areas of interest unchanged. Work reported in Chapter 6 shows its application to fuzzy change images.

4.8 The Logical Basis of Supervised Classification

An examination of the logical form of supervised classification reveals that which we know through both experience and intuition: some errors are inevitable no matter how much care we take with the classification itself.

Figure 4.11: Errors of Omission in the Minimum Mahalanobis Distance Classified TM93 Image, Class Forest, after Morphological Filtering

With supervised classification, we determine some training sets. From these training sets, we essentially derive some knowledge about the spectral response that we expect to see from each identified class. This knowledge of the form "IF class THEN spectral response". We wish to use this knowledge to infer the class from any given observation. This inference of the form "IF spectral response THEN class". Alternatively, "with the knowledge that a particular class causes a particular spectral response, we conclude, from observing the particular spectral response, that the cause of this response was the particular class".

To simplify the following discussion, denote our knowledge $A \Rightarrow B$ ("A implies B"). For the purposes of this discussion, it does not matter whether the implication is material or logical (although the "cause and effect" nature of the discussion suggests the "stronger" logical implication).

If we possess the knowledge $A \Rightarrow B$, and if we are then given A, we may validly deduce B. This is an example of the form of inference known as deduction. More specifically, this particular form is known as *modus ponens*. It is commonly written

$$\frac{A \Rightarrow B, A}{B} \tag{4.11}$$

i.e., from $A \Rightarrow B$ and A, we conclude B.

The form of inference that we are employing in supervised classification is essentially

$$\frac{A \Rightarrow B, B}{A} \tag{4.12}$$

i.e., from $A \Rightarrow B$ and B, we conclude A.

The syllogism represented by (4.12) is known as "abduction" (see e.g. Fann, 1970; Charniak and McDermott, 1985; Paul, 1993; Josephson and Josephson, 1994). This is essentially the form of reasoning used in medical diagnosis, in which we observe the symptoms and attempt to infer the disease, or cause of those symptoms. Of course, this form of inference is not assured. There may be other causes that lead to the same observations or symptoms, such as a hitherto unknown cause, or a combination of known causes that collectively present similar or identical observations or symptoms.

To illustrate, let a medical diagnostician possess knowledge that disease D_1 causes the set of symptoms $\{S_1, S_2, S_3, S_4, S_5\}$. Let this particular set of symptoms be

observed in a particular patient. If the diagnostician possesses no knowledge of any other single disease causing the same set of symptoms, he/she will commonly infer the presence of disease D_1. This is, of course, done under the assumption that there is a single cause or disease. The more careful diagnostician would search the knowledge base for any other disease or set of diseases that would give rise to the same observed set of symptoms.

Let the relevant extracts of the knowledge base be

$$D_1 \Rightarrow \{S_1, S_2, S_3, S_4, S_5\}$$

$$D_2 \Rightarrow \{S_1, S_2\}$$

$$D_3 \Rightarrow \{S_3, S_4, S_5\}$$

$$D_4 \Rightarrow \{S_3, S_4\}$$

$$D_5 \Rightarrow \{S_5\}$$

That is, disease D_1 causes the set of symptoms $\{S_1, S_2, S_3, S_4, S_5\}$, disease D_2 causes the set of symptoms $\{S_1, S_2\}$, and so on.

On observing $\{S_1, S_2, S_3, S_4, S_5\}$, a more complete diagnosis would be that the patient had one of the following sets of diseases $\{D_1\}, \{D_2 \wedge D_3\}, \{D_2 \wedge D_4 \wedge D_5\}$.

It should not be overlooked that these represent the 'minimal' explanations of the observed symptoms: the patient could, in fact, actually have $\{D_1 \wedge D_2\}$, or $\{D_1 \wedge D_3\}$, or $\{D_1 \wedge D_4\}$, or $\{D_1 \wedge D_5\}$, or $\{D_1 \wedge D_2 \wedge D_3\}$, or ... or $\{D_1 \wedge D_2 \wedge D_3 \wedge D_4 \wedge D_5\}$. In the absence of any other diagnostic information (e.g. do any of these diseases cause other symptoms, have these symptoms been tested for, can they be tested for? etc.), conclusions may be guided by *a priori* probabilities or other beliefs, or some consideration of the relative "cost" of errors. For example, D_4 may be fatal if not treated in some particular manner, whereas the other diseases may have limited consequences, irrespective of treatment. In this case, it would be

prudent to either test for D_4 (if this is possible), or (otherwise) assume that the patient has D_4, and treat him or her accordingly. Of course, it remains possible that there is another disease, or set of diseases, that causes some or all of observed symptoms. If the diagnostician does not know of these, erroneous conclusions may be drawn from the observations. There may also be certain variability in patients with particular diseases presenting symptoms, and there may be some uncertainties relating to diagnostic tests actually detecting symptoms. These factors may also lead to erroneous conclusions.

In the case of remote sensing, the cause of a particular spectral response may be a particular class, or it may be a hitherto unknown class, or it may be a combination of known and unknown classes. It is noteworthy that the combination of known classes that gives rise to the particular spectral response need not necessarily include the class to which the response is 'closest'.

The key point about the preceding discussion is simply that the reasoning process that underlies classification is abduction, which, by its nature, seeks plausible explanations of observed facts: we are not assured that these explanations are in fact correct.

In addition to the uncertainty associated with the implication itself (i.e. the inference mechanism), there is, of course, variability in the observed data (the spectral responses), so the machinery of statistics is often brought to bear.

4.9 The Use of Spatial Context in Classification

The classifiers described in this chapter are what are sometimes referred to as "per-pixel" classifiers: they consider only the intensity values of the pixel in question, and make no use of information about the neighbourhood in which the pixel resides.

Prior beliefs are often used to aid classification. Prior beliefs could simply be of the likelihoods of particular classes over the entire imaged region (such as in Equation (4.2)). They could also incorporate information about the global likelihood of combinations of known classes (although no studies that do this are known to the author). There is, however, another form of prior belief that is very useful and powerful for remote sensing classification tasks; the belief, based on long observation, that like classes tend to co-occur spatially. This concept ("spatial autocorrelation") has been the subject of considerable study (see e.g. Student, 1914; Besag, 1974; Hubert and Golledge, 1982; Clifford *et al.*, 1989). It offers some assistance in the problem described in the last sub-section, when an observed spectral response could come from class 1, or a mixture of classes 2 and 3, or a mixture of all three classes.

To illustrate, let us assume that there are no unknown classes. Consider a pixel that, accordingly to whatever classifier we are using, is assessed as being either class 1, or a mixture of classes 2 and 3 in some particular proportion, or a mixture of classes 1, 2 and 3 in some (other) particular proportions. If we had some knowledge about the region surrounding the pixel, it seems reasonable that this be permitted to condition our assessment of which of the three classification possibilities was extant. For example, let us suppose that some of the pixels surrounding the pixel in question were unequivocally believed to be class 2, some were unequivocally believed to be class 3, and some were similar to the pixel in question. We may be inclined to discard the possibility that the pixel is class 1, or that it is a mixture of classes 1, 2 and 3. Being left only with the possibility that the pixel is a mixture of classes 2 and 3, we can proceed and determine the proportions through some method. A variant of this reasoning process that accounted for statistical variability in the data might calculate and utilise some form of local "prior probability" (Dunne and Campbell, 1995).

An alternative approach to the use of spatial context information in remote sensing classification is by incorporating some form of local texture information. A comprehensive early review of texture features for remote sensing image classification can be found in Haralick and Shanmugan (1974). The use of texture has been shown to improve classifier performance over the simple "per-pixel" classifiers (see e.g. Whitbread, 1992; or Bouzerdoum *et al.*, 1996).

Despite these promising indications, the work reported in this thesis makes no use of information about spatial context for fuzzy classification or change detection: it remains for future work.

4.10 Conclusion

For this application, and with this data, it has been shown that unsupervised classification offers no advantage over supervised classification in terms of accuracy, and was more costly and problematic to implement. Of the supervised classification approaches examined, the Minimum Euclidean Distance Classifier was found to perform relatively poorly, but the other classifiers (Minimum Mahalanobis Distance, Maximum Likelihood (without prior probabilities), and Maximum Likelihood (with prior probabilities)) produced highly comparable results. That is, the inclusion of terms addressing the relative 'size' (as indicated by the determinant) of the covariance matrices and prior probabilities of the classes had no appreciable effect. This result was mildly surprising.

The classifiers all performed better than a simple inspection of the accuracy assessment tables would indicate, as there were evidently a number of errors with the 'ground truth' data, and apparent error was introduced through imperfect registration of the image to the ground truth. Morphological filters were introduced to reduce

these error artifacts.

A consideration of the logical basis of supervised classification reinforces our understanding that errors appear to be inevitable, and that classification should take account of the possibility of mixed pixels in the data. Indications of the presence of mixed pixels may be found by examining the local neighbourhood of the pixel. This, and other approaches to the use of information about spatial context, offer promise to reduce classification error, but are not pursued in this thesis.

Chapter 5

FUZZY SETS AND LAND USE CLASSIFICATION

Summary. This chapter introduces the notion of fuzzy sets and their application to image processing and remote sensing. A supervised fuzzy classifier is described. This classifier is based on the reciprocals of the (squared) Mahalanobis Distances of each pixel from a number of training class means. A preliminary study on the use of this classifier for identifying general land use categories on Kangaroo Island is described. It is concluded that the Mahalanobis Distance fuzzy classifier offers a useful approach to land use classification. The TM89 and TM93 images are then classified using this approach, and the accuracy assessment of fuzzy classified images is discussed.

5.1 Introduction

The concept of fuzzy sets was formally introduced in Zadeh (1965). Fuzzy set theory extends classical set theory in that an object can have a degree of membership in a set that is not necessarily 0 or 1, or IN or OUT, but some real number in the closed interval $[0, 1]$. That is, some degree of partial membership in a set is allowed.

To illustrate how this differs from classical set theory, we could decide upon the membership of a person in the classical set TALL PEOPLE by whether their height was greater than or equal to 1.8 metres. If YES, they are in the set of set TALL PEOPLE; if NO, they are not in the set. This is the classical set approach. The fuzzy set approach would see each person being allocated a degree of membership in the fuzzy set of TALL PEOPLE, with, say, a person 2 metres tall having a degree of membership of 1, a person 1.5 metres tall having a degree of membership of 0, a person 1.8 metres tall having a degree of membership of 0.9, a person 1.75 metres tall having a degree of membership of 0.85, and so on. This approach permits us to reason that a person 1.75 metres tall possesses the attributes of a member of the set of TALL PEOPLE to almost the same degree as a person 1.8 metres tall.

It should be noted that a fuzzy set membership should not be likened to a probability of the object belonging to the set or not: the object does belong, to some degree (as defined by the membership of the object in the set). Discussions on the relationship between fuzzy set theory and probability can be found in Dubois and Prade (1993) and Bezdek (1994). Further, the classical set theory laws of the excluded middle and non-contradiction do not necessarily need to apply in the case of fuzzy sets. These laws require that the union of a set and its complement is the universal set, and that the intersection of a set and its complement is the empty set.

5.1.1 Definitions and Notation

A detailed treatment of definitions, notation and theory relating to fuzzy sets can be found in Klir and Folger (1988). The definitions and notation relevant to this thesis are included below.

Let $\mu_A(x)$ denote the degree of membership of x in the fuzzy set A, where $x \in X$,

and X is the universal set, or universe of discourse.

The support of a fuzzy set A in the universal set X is the crisp set that contains all the elements in X that have nonzero membership in A.

$$\text{supp } A = \{x \in X \mid \mu_A(x) > 0\} \tag{5.1}$$

Let $x_i \in \text{supp } A$ and μ_i be its grade of membership in A. The fuzzy set A can be denoted

$$A = \mu_1/x_1 + \mu_2/x_2 + \ldots + \mu_n/x_n \tag{5.2}$$

(where the "/" symbol links the grade of membership μ_i with the element x_i, and the "+" symbol shows that the various μ_i/x_i collectively comprise the set)

When X is an interval of real numbers, the fuzzy set A is often denoted

$$A = \int_X \mu_A(x)/x \tag{5.3}$$

(where the "$\int$" sign does not denote integration, but performs the same function as the "+" symbol in (5.2) above)

The *height* of A is the maximum μ_i

A is said to be normalised iff

$$\exists x \in X \quad \text{s.t.} \quad \mu_A(x) = 1 \tag{5.4}$$

This requirement is satisfied if there is at least one element with full membership in the set A, or, alternatively, if the *height* of A is 1.

An alpha-cut of A is the crisp set

$$A_\alpha = \{x \in X \mid \mu_A(x) \geq \alpha\} \tag{5.5}$$

A fuzzy set A is *convex* iff

$$\mu_A(r\lambda + (1-\lambda)s) \geq \min\ \{\mu_A(r), \mu_A(s)\}, \forall r, s \in R^n, \forall \lambda \in [0,1] \tag{5.6}$$

(where R denotes the set of real numbers)

A fuzzy number is a convex and normalised set defined on the set of real numbers with a piecewise continuous membership function.

$$\text{If} \quad \mu_A(x) \leq \mu_B(x) \quad \forall x \in X, \quad \text{then} \quad A \subseteq B \tag{5.7}$$

Similarly,

$$\text{if} \quad \mu_A(x) = \mu_B(x) \quad \forall x \in X, \quad \text{then} \quad A = B \tag{5.8}$$

The scalar cardinality of a fuzzy set is the sum of the membership grades of all the elements in the set.

$$\sum \text{count}(A) = \sum_{x \in X} \mu_A(x) \tag{5.9}$$

Scalar cardinality is also commonly denoted $|A|$. Similar notation is also often used to denote the determinant of a matrix, or the absolute value of a number. Both of these concepts are used later in this thesis. Therefore, to avoid possible confusion, this notation will not be used for any of these concepts.

It is common to denote and define complement, union and intersection as follows:

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \tag{5.10}$$

$$\mu_{A \cup B} = \max[\mu_A(x), \mu_B(x)] \tag{5.11}$$

$$\mu_{A \cap B} = \min[\mu_A(x), \mu_B(x)] \tag{5.12}$$

It is noted that these are not the only possible ways to define complement, union and intersection: we return to this matter in Chapter 6.

5.1.2 Fuzzy Sets in Image Processing and Remote Sensing

Traditional classification approaches seek to allocate each observation (e.g. pixel in an image) to one of a (hopefully exhaustive) set of mutually exclusive classes (or, more accurately, to associate a single class label with each observation). Approaches of this type are now sometimes referred to as 'hard' classifications. Unfortunately, classes commonly do not have distinct 'boundaries'. Even human experts in the field will often disagree on class assignment e.g. when distinguishing between 'stressed' and 'not-stressed' vegetation, or between pasture and thinly wooded areas.

In this context, 'boundaries' refer to the conceptual, or class definition, boundary: what defines a particular area as forest, and would two independent observers classify the particular area identically? To illustrate, we might consider the question "in a landscape in which the vegetation changes gradually from forest to grassland, where should the (geographic) boundary marking the separation of the two (or more) classes be drawn?". A common reference for guidance on such matters, particularly in remote sensing applications, is Anderson *et al.* (1976). This defines, e.g., forest as a region with a tree canopy cover of 10% or greater. Such a region may have significant proportions of other classes (understorey vegetation, grass, bare soil, rock, tree litter of various types, water or mud, shadow etc.). The selection of 10% canopy cover as

defining forest is essentially arbitrary, and judging the percentage of canopy cover would be somewhat subjective anyway (and prone to interpretation difficulties in the case of deciduous trees). Further, Anderson *et al.* (1976) is not the only source of definitions for ground cover classes. In this environment, we might reasonably expect human experts, either in the field or using imagery of various types, to differ somewhat in classification tasks, and the concept of classes is 'fuzzy'.

There are many papers in the literature arguing that crisp class assignment is often inappropriate in the geographical and remote sensing sciences (see e.g. Burrough, 1989; Key *et al.*, 1989; Wang, 1990a,b; Fisher and Pathirana, 1990; Foody, 1992; van der Meer, 1995). Moreover, a single pixel in a remotely sensed image may contain a number of classes (i.e. it may be a 'mixed pixel', or 'mixel').

Fuzzy set theory was introduced to address the issue of class or set 'vagueness'. Using fuzzy set theory, we can determine and reason with the grade of membership of a particular pixel in a number of classes. This provides an approach to (partial) class assignment in regions where there is a gradual transition from one class to another. It can also be used as an approach to the mixed pixel problem.

A basic premise of this thesis is that the use of fuzzy sets and related concepts is quite natural and intuitively pleasing in remote sensing applications, where class boundaries are 'fuzzy', and any particular observed pixel could contain a number of ground classes.

The application of fuzzy set theory to pattern recognition and image processing finds considerable support in the literature. Zadeh states (in the Foreword to Bezdek and Pal, 1992) "the initial development of the theory of fuzzy sets was motivated in large measure by problems in pattern classification and cluster analysis". Bezdek and Pal (1992, p. 8) observe that "feature vectors (and the objects they represent) can and

should be allowed to have degrees of membership in more than one class". Pal (1992) discusses the application of fuzzy sets to image processing in general. Discussion of its particular application to remotely sensed imagery can be found in Key *et al.* (1989), Wang (1990a), Blonda *et al.* (1991), and Foody (1992). The relevance of fuzzy set theory to the mixed pixel problem is discussed in Chapter 7.

5.1.3 Fuzzy Classification

Fuzzy classification itself is not new. As noted earlier, there have been numerous papers on fuzzy clustering (which is normally unsupervised) and fuzzy supervised classification.

The approach to fuzzy classification adopted in this thesis is based on the fuzzy *c*-means clustering algorithm (Bezdek *et al.*, 1984). This is an unsupervised fuzzy clustering, or classification, of the data. The fuzzy set memberships are determined by a distance measure (usually Euclidean, diagonal or Mahalanobis) from the class centroids, and the optimal clustering is determined by minimising an error function, subject to certain constraints. In this manner, the similarity (as measured by a distance metric) between an instance (i.e. the observation) and a prototypical class member (i.e. the class centroid) is used to determine the class memberships.

The fuzzy *c*-means clustering algorithm attempts to minimise an error function

$$J_m = \sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{ik}^{q} d_{ik}^{2} \tag{5.13}$$

where there are n observations (i.e. pixels), c classes, μ_{ik} is the membership of pixel x_k in class i, q is a parameter used to control the fuzziness of the partition, and

$$d_{ik}^{2} = (x_k - m_i)^t \mathbf{A} (x_k - m_i) \tag{5.14}$$

is the distance measure, or inner product norm, between pixel x_k and the mean of class i, induced by the positive definite matrix $\mathbf{A}$. The mean of class i, denoted m_i, is determined by

$$m_i = \frac{\sum_{k=1}^{n} \mu_{ik}^q x_k}{\sum_{k=1}^{n} \mu_{ik}^q} \tag{5.15}$$

The constraints on the partitioning are shown by

$$M_{fc} = \left\{ U : \mu_{ik} \in [0,1]; \sum_{k=1}^{n} \mu_{ik} > 0,\ i = 1, \ldots, c; \sum_{i=1}^{c} \mu_{ik} = 1,\ k = 1, \ldots, n \right\} \tag{5.16}$$

U is a real $c \times n$ matrix. Row i reveals the memberships of each pixel in the class i. Column j shows how the memberships of a given pixel (pixel j) are distributed between the classes.

The error function (5.13) can be shown to be minimised if the fuzzy memberships are of the following form (Bezdek *et al.*, 1984)

$$\mu_{ik} = \frac{(\frac{1}{d_{ik}^2})^{\frac{1}{q-1}}}{\sum_{j=1}^{c} (\frac{1}{d_{jk}^2})^{\frac{1}{q-1}}} \tag{5.17}$$

where

$$i = 1, \ldots, c \quad \text{and} \quad k = 1, \ldots, n$$

The algorithm proceeds iteratively, updating the cluster centres and the fuzzy memberships of each pixel, until some stopping criterion is reached.

The unsupervised fuzzy c-means approach was used in Key *et al.* (1989) to classify clouds. The authors noted that the number of classes could be specified and the class

means supplied "in a manner analogous to using training sites to provide spectral statistics for a supervised classification", thereby greatly decreasing the program execution time. That is, it could be adapted to operate in a supervised manner. In this approach, the fuzzy set memberships would still be determined by a distance measure (in their case, diagonal) from the class centroids, but the class centroids were supplied, rather than being determined by the algorithm. Foody (1992) used a similar approach to investigate the application of the fuzzy *c*-means algorithm to heathland classification and the relationship between fuzzy set memberships and canopy composition. Class spectral characteristics were determined from small training samples, and the Mahalanobis Distance measure was used.

5.1.4 A Fuzzy Classifier

A preliminary study was undertaken in which a fuzzy classifier was developed in the programming language C, using the software development environment and image management facilities provided by Khoros. This work is reported in detail in Deer *et al.* (1996a,b); an outline is presented below.

The software was developed and run on a SUN SPARCstation IPC (with 20 Mbytes of RAM and a colour display) operating under SunOS version 4.1.4. Khoros is a software development environment that is distributed through the Internet via anonymous ftp as Free Access Software (see http://www.khoros.unm.edu). Source code is available, so developers can write their own applications and 'link' them into Khoros. The Khoros environment provides a number of useful facilities (including image management and display), and a visual programming environment, Cantata (which was used in this work).

A supervised fuzzy classifier was developed to operate over arbitrary image sizes,

numbers of ground cover classes, and numbers of image bands. A suite of programs read the original image data and store it in a suitable format. The image data is radiometrically stretched for display, to aid training set selection by a (human) operator. Training class means and covariance matrices are computed, and used to determine the (squared) Mahalanobis Distance between each pixel and the mean of each class,

$$d_{ik}^2 = (\mathbf{x}_k - \mathbf{m}_i)^t \mathbf{\Sigma}_i^{-1} (\mathbf{x}_k - \mathbf{m}_i) \tag{5.18}$$

where d_{ik}^2 represents the (squared) Mahalanobis Distance between pixel $\mathbf{x}_k$ and the mean of the ith class, $\mathbf{m}_i$; superscript t denotes (vector) transpose; and $\mathbf{\Sigma}_i^{-1}$ is the inverse of the covariance matrix of the ith class. In the event of singular matrices, the identity matrix is chosen (this has the effect of computing Euclidean, rather than Mahalanobis, Distances).

The fuzzy class membership of each pixel in the image in each class is calculated according to

$$\mu_{ik} = \frac{(\frac{1}{d_{ik}^2})^m}{\sum_{j=1}^{c} (\frac{1}{d_{jk}^2})^m} \tag{5.19}$$

where

$$i = 1, \ldots, c \quad \text{and} \quad k = 1, \ldots, n$$

and where μ_{ik} is the fuzzy class membership of pixel x_k in class i, c is the number of classes, n is the number of pixels in the image, and m is a parameter used to control the 'degree of fuzziness' of the class membership allocation, related to q in earlier equations by

$$m = \frac{1}{q-1} \tag{5.20}$$

This substitution has been effected for subsequent notational and manipulative ease.

It is restated that determining fuzzy class memberships in this way ensures that

$$\sum_{i=1}^{c} \mu_{ik} = 1 \quad \forall k \tag{5.21}$$

That is, for each pixel in the image, the sum of memberships over all classes is unity. The result is effectively an $r \times p \times c$ array of fuzzy set memberships, where $r \times p$ is the image size, and c is the number of classes.

It was stated earlier that three common 'distance' measures used with the fuzzy c-means algorithm are the Euclidean, Diagonal and Mahalanobis norms. The Euclidean norm is appropriate for uncorrelated variables with similar scales; the Diagonal norm for uncorrelated variables with different scales; and the Mahalanobis norm for correlated variables with either the same or different scales (Key *et al.*, 1989).

The Mahalanobis norm was chosen for the work described in this thesis because there is known to be significant correlation between bands in remotely sensed MSS and TM data. Furthermore, the work in Chapter 4 clearly showed the Mahalanobis norm to produce superior results to the Euclidean norm in the case of traditional classification.

The Mahalanobis norm is used to determine the similarity between a particular pixel and the prototypical members (as represented by the means) of all classes, taking each class distribution into account by the use of the class covariances. The calculation uses the reciprocal of the distance to assign a high fuzzy set membership

for a pixel in a particular class when the pixel is close to the mean of that class, and a low membership otherwise. While there are other possibilities that can be applied to 'invert' the computed distances, the reciprocal approach is simple in concept and implementation (although the case $d_{ij} = 0$ has to be considered). Other approaches are investigated in Chapter 8 of this thesis.

The approach of using the Mahalanobis Distance aims at an objective and supportable measure of similarity, and is seen as more sound than other candidate methods of determining the membership functions of pixels in classes, such as by predefined (say, triangular) membership functions based on some approximate modeling or *a priori* beliefs.

It is noted that defining the class memberships in this manner will ensure that they sum to one for each pixel, irrespective of whether or not the pixel is 'close' to any class mean at all. This approach does not distinguish between actual closeness, merely relative closeness: the issue of typicality (discussed in Campbell, 1984) is not considered. Of course, should it be desired, it is a relatively minor matter to post-process the results of the fuzzy classification, setting all memberships to zero (i.e. leaving the pixel unclassified) if some typicality threshold is not reached. The Mahalanobis Distance itself commonly serves in this regard, because, with Gaussian class distributions, the Mahalanobis Distances are chi-squared distributed. Typicality thresholding was not, however, applied in this study.

Training set selection in this software was by grouping an arbitrary number of regions of known common class, with one such group for each class of interest. To examine the effect on the classifier of extreme values in the training data, the facility was provided to remove them from the training sets by specification of a percentage discard. The class signatures (means and covariance matrices) were computed using

the training data remaining after this discard.

Preliminary Study

Study Area and Materials The study area was on Kangaroo Island, South Australia, as described in Chapter 3. The image used was 500 by 500 pixels, extracted from a Landsat MSS image (Path 98 Row 85, acquisition date 8-3-89). The (unrectified) image is shown at Figure 5.1.

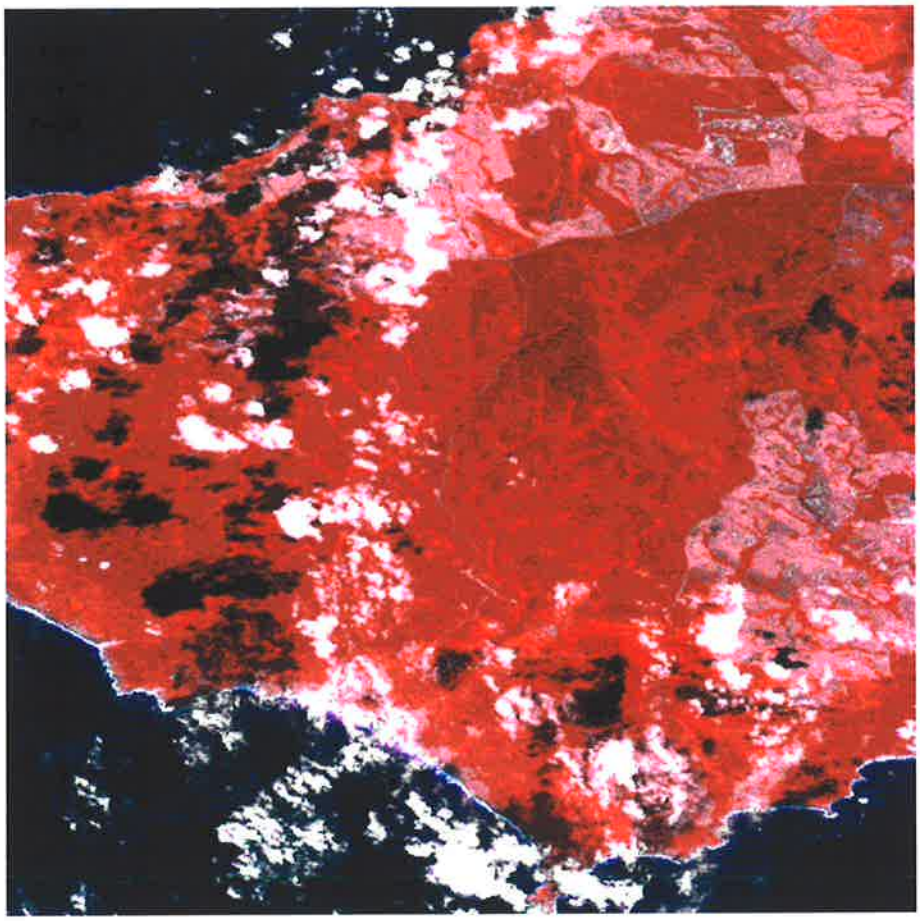

Figure 5.1: MSS 1989 Image (Bands 3,2,1 as RGB)

Training sets were chosen by grouping a number of regions of interest for each of six classes: forest, cleared pasture, sea, cloud, cloud shadow on forest and cloud

Table 5.1: Sample Fuzzy Set Class Memberships (Row 370)

	Column									
	61	62	63	64	65	66	67	68	69	70
forest	0.04	0.35	0.08	0.08	0.37	0.69	0.79	0.93	0.69	0.65
pasture	0.18	0.56	0.86	0.88	0.55	0.36	0.15	0.06	0.26	0.29
sea	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
cloud	0.10	0.06	0.04	0.03	0.06	0.03	0.03	0.01	0.03	0.05
shadow (forest)	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01
shadow (sea)	0.68	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01

shadow on sea. The numbers of pixels in the training sets of the various classes were: forest - 371; cleared pasture - 211; sea - 202; cloud - 86; cloud shadow on forest - 75; and cloud shadow on sea - 225. These exceed the minimum number of pixels for training classes recommended in Richards (1993) (i.e. ten times the number of classes).

Results and Discussion Various trials revealed that percentage discard had no appreciable effect on the results, so it was set to zero. The classifier was run a number of times with different values of m (in the range of "useful values" suggested in Bezdek *et al.*, 1984). The value $m = 1.25$ (equivalent to $q = 1.8$) was qualitatively assessed to give the best results. The issue of assessment of fuzzy classifications is addressed in Section 5.4.

To illustrate the transition of memberships across a class boundary with this fuzzy classification, an excerpt of the fuzzy classified output is shown at Table 5.1. The membership of forest shows a significant increase from pixel 64 to 68, with the membership of the class pasture showing a corresponding decrease. The membership of the other classes in this range is negligible.

The normal method of displaying a (traditional) classified image is to display each class as a unique colour in a single image. This approach is unsuitable to fuzzy classification (as each pixel is allowed a degree of membership in each class). The output of the fuzzy classifier may be 'hardened' in some way (e.g. by assigning each pixel to the class in which it has maximum membership, subject, perhaps, to some threshold value). A conventional classified image could then be displayed, but this largely defeats the purpose of fuzzy classification.

The fuzzy set memberships of each pixel in each class may be shown in a number of greyscale images, one for each class of interest. This is done in Figures 5.2 to 5.7, respectively showing: forest, pasture, sea, cloud, cloud shadow on forest and cloud shadow on sea. The images show high membership in the respective classes as high brightness values.

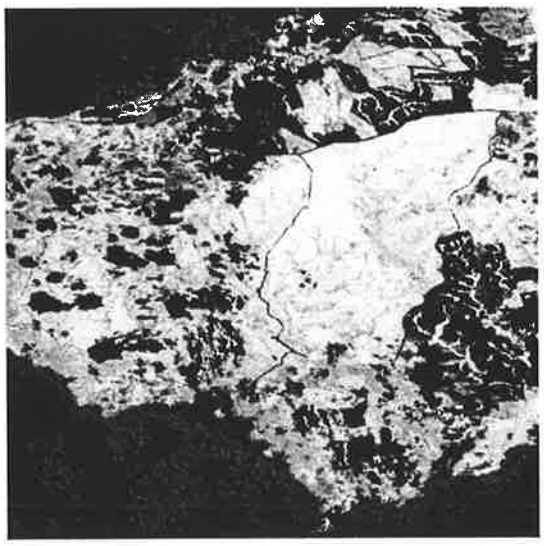

Figure 5.2: Fuzzy Membership of the Class forest, MSS89

It is possible, however, to display a single combined fuzzy classified image for a limited number of classes by assigning membership of particular classes to the

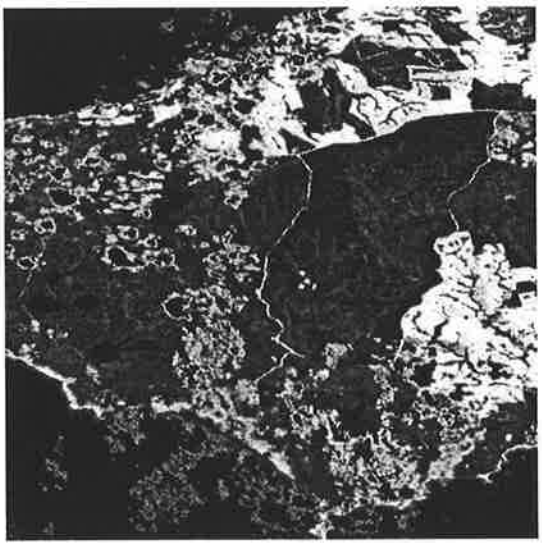

Figure 5.3: Fuzzy Membership of the Class pasture, MSS89

Figure 5.4: Fuzzy Membership of the Class sea, MSS89

red, green and blue components of a RGB display. This would, at first, seem only to allow the degree of membership in three classes to be displayed. The degree of membership of a fourth class can, however, be shown by allocating its membership

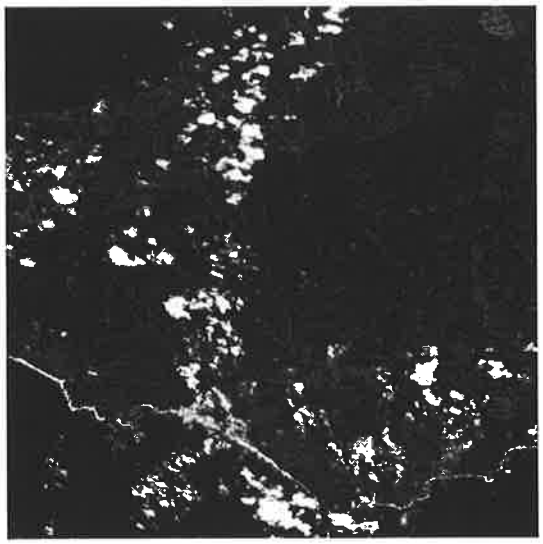

Figure 5.5: Fuzzy Membership of the Class cloud, MSS89

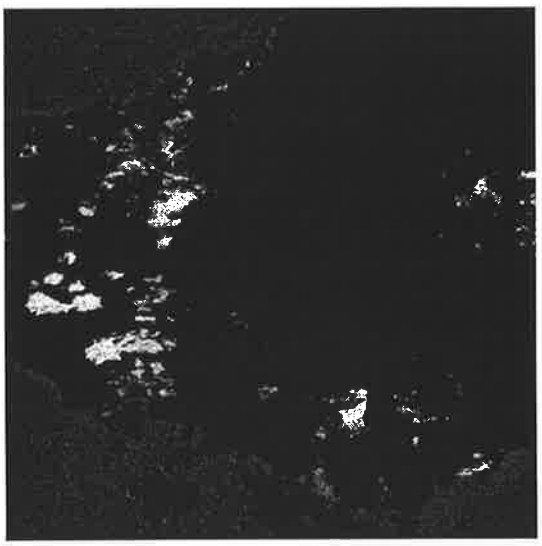

Figure 5.6: Fuzzy Membership of the Class cloud shadow on forest, MSS89

to each of red, green and blue equally (in addition to the allocations made by each of the three preceding classes). This has the effect of showing this class as shades of light grey/white, but it is noted that there may then be confusion between this fourth

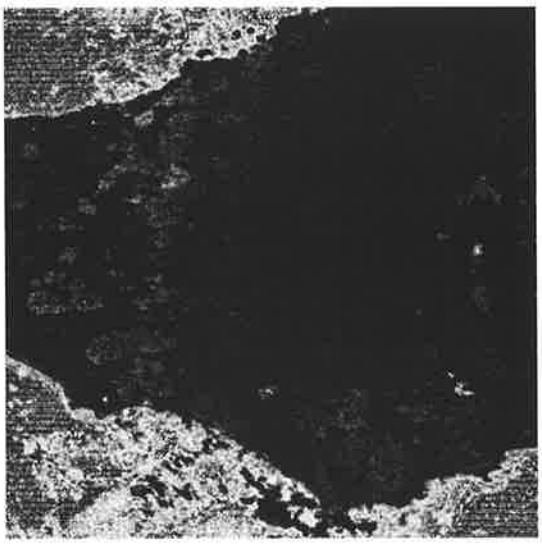

Figure 5.7: Fuzzy Membership of the Class cloud shadow on sea, MSS89

class, and pixels exhibiting equal memberships of the first three classes. All pixels having a low degree of membership in the chosen four classes will show as dark in the resultant image. Because the memberships for each pixel in all classes sum to one, we can say that any pixel showing dark in this combined image has memberships in the remaining classes summing to near one.

Of the six classes represented in this work, two were cloud shadow classes. It was therefore decided to assign the degree of membership of each pixel in the classes forest, pasture and sea to the colours green, red and blue respectively. The membership of each pixel in the class cloud was added to each colour equally (cloud therefore shows as white/light grey). The result is shown in Figure 5.8. All dark areas may be interpreted as cloud shadow.

The results achieved appear qualitatively reasonable and useful, despite the obvious errors (such as the borders of clouds/shadows showing a high membership

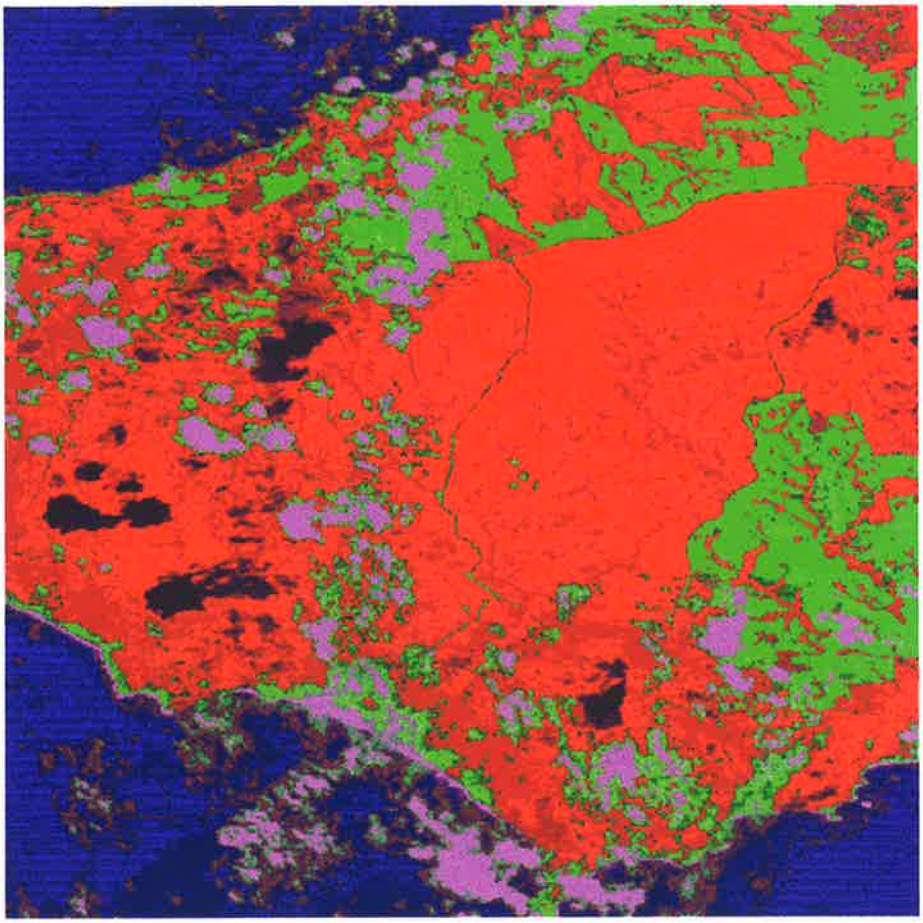

Figure 5.8: Combined Fuzzy Classified Image (forest-red; pasture-green; sea-blue; cloud-white/light grey; cloud shadow-black/dark grey), MSS89

of pasture). It was therefore considered that further work on the development and use of fuzzy classifiers was warranted, including: the efficacy of fuzzy classification in PCC change detection; the effect of the value of the weighting exponent m in the fuzzy classifier; other forms of fuzzy classification; measures of fuzziness; approaches to accuracy assessment and the nature of the required ground truth data; and the 'mixel' interpretation of the fuzzy classification.

Classification of Landsat TM 1989 and 1993 Images For the next (and subsequent) stages of the work reported in this thesis, the fuzzy classifier was

re-implemented in the ERDAS IMAGINE (version 8.2/8.3) image processing environment, running on both Silicon Graphics (under IRIX 5.3/6.2) and PC (under Windows NT 4.0) machines.

The fuzzy classification of the rectified Landsat TM 1989 and 1993 images was undertaken using the training sets described in Chapter 3, and with a fuzzy exponent value $m = 1.25$. Figures 5.9 to 5.14 show the fuzzy memberships of the TM 1989 image in the classes forest, sea, pasture, pine, cloud and shadow. The class pine was introduced because, although small in areal extent, any conversions from forest to pine or from pasture to pine were of interest. The class cloud shadow on sea was dropped, as it and the class sea were shown to be not practically separable in the preliminary study. The TM 1993 image does not have instances of cloud or cloud shadow; memberships in the other classes are shown in Figures 5.15 to 5.18.

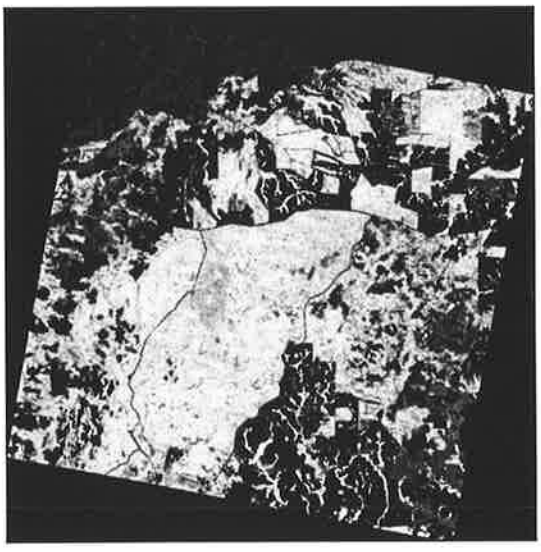

Figure 5.9: Fuzzy Membership of the Class forest, TM89

Colour images combining the memberships in the classes forest (red), sea (blue), and pasture (green), are shown for 1989 and 1993 as Figures 5.19 and 5.20 respectively.

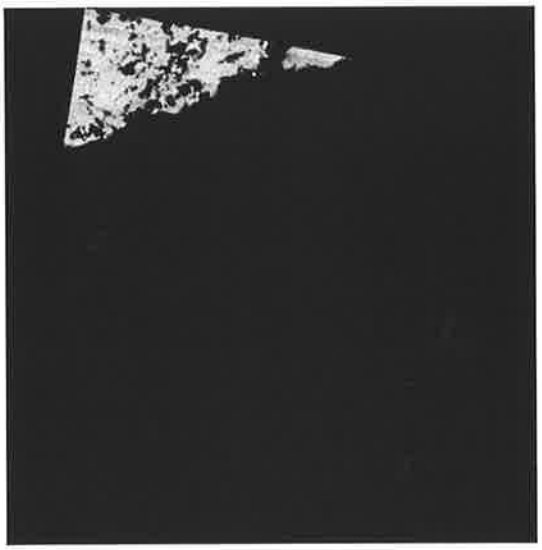

Figure 5.10: Fuzzy Membership of the Class sea, TM89

Figure 5.11: Fuzzy Membership of the Class pasture, TM89

5.2 Accuracy Assessment with Fuzzy Sets in Image Processing

The traditional method of accuracy assessment is discussed in Chapter 2, and example accuracy assessments for traditional classifications are given in Chapter 4. As noted, it

Figure 5.12: Fuzzy Membership of the Class pine, TM89

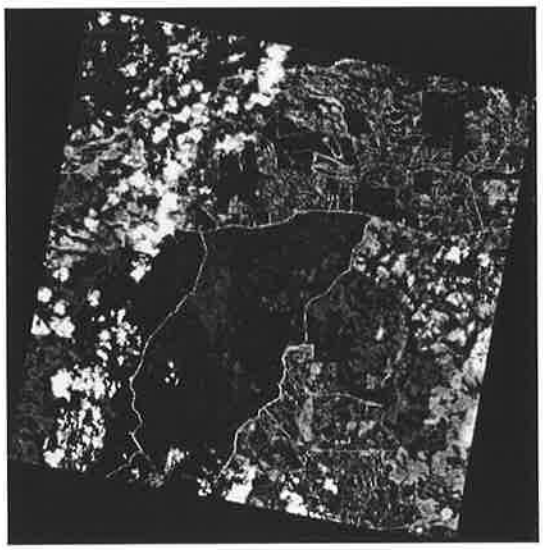

Figure 5.13: Fuzzy Membership of the Class cloud, TM89

is normally expected that traditional classifications and corresponding reference data for accuracy assessment will be crisp, and that there will be 'hard' class assignments, for both the image classification and for the ground truth data. We may therefore

Figure 5.14: Fuzzy Membership of the Class shadow, TM89

Figure 5.15: Fuzzy Membership of the Class forest, TM93

expect to find, for a particular point on the ground, or for a particular pixel, either no label, or a single, unambiguous, unequivocal label. Clearly, this approach is unsuitable for assessing the accuracy of a fuzzy classification which assigns a number of 'labels'

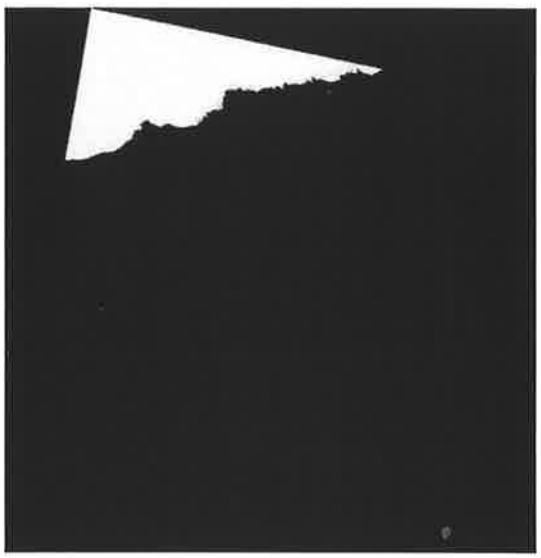

Figure 5.16: Fuzzy Membership of the Class sea, TM93

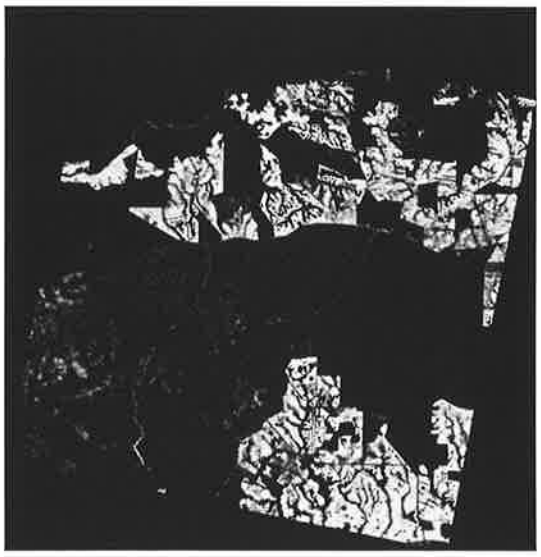

Figure 5.17: Fuzzy Membership of the Class pasture, TM93

to each pixel representing the computed degree of membership of the pixel in each class.

The quantitative accuracy assessment of a fuzzy classifier requires an appropriate

Figure 5.18: Fuzzy Membership of the Class pine, TM93

form of 'ground truth' (Wang, 1990a; and Foody, 1992). Some papers have used the approach of assessing the accuracy of a fuzzy classifier by 'hardening' the output by assigning each pixel to the class of greatest fuzzy set membership and comparing this label to coincident 'hard' ground truth (Foody and Trodd, 1993). This approach is considered inappropriate, and Foody (1995) notes that "The resulting accuracy statement is not, however, a good measure of the accuracy of a fuzzy classification".

Foody (1996) notes that "relatively little attention has addressed the problems of assessing the accuracy of classifications which include mixed pixels". Issues related to fuzzy classification and the mixed pixel problem are addressed in Foody (1995). Foody (1995) examines the accuracy assessment of fuzzy classifications with fuzzy ground data and the use of a cross-entropy measure between the two, under the assumption that fuzzy memberships should reflect class proportions in mixed pixels. This assumption is examined in Chapter 7, as some of the subsequent work is explicitly predicated on it.

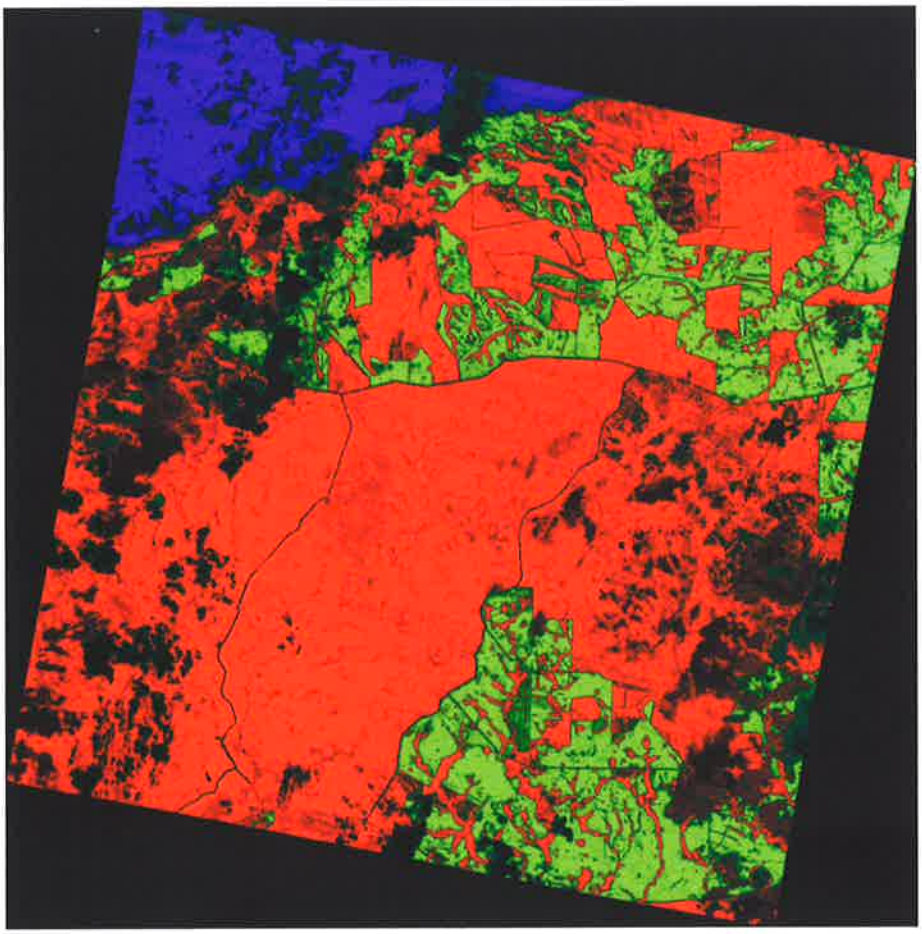

Figure 5.19: Combined Fuzzy Classified Image (forest-red; pasture-green; sea-blue), TM89

5.3 Conclusion

In this chapter, the notion of fuzzy sets and their application to image processing and remote sensing was introduced. A supervised fuzzy classifier drawn from the literature was described. This classifier is based on the reciprocals of the (squared) Mahalanobis Distances of each pixel from a number of training class means, weighted by an exponent. This provides the means to vary the 'fuzziness' of the classifier. A preliminary study (using Landsat MSS data) on the use of this classifier for identifying general land use categories on Kangaroo Island was described. It is

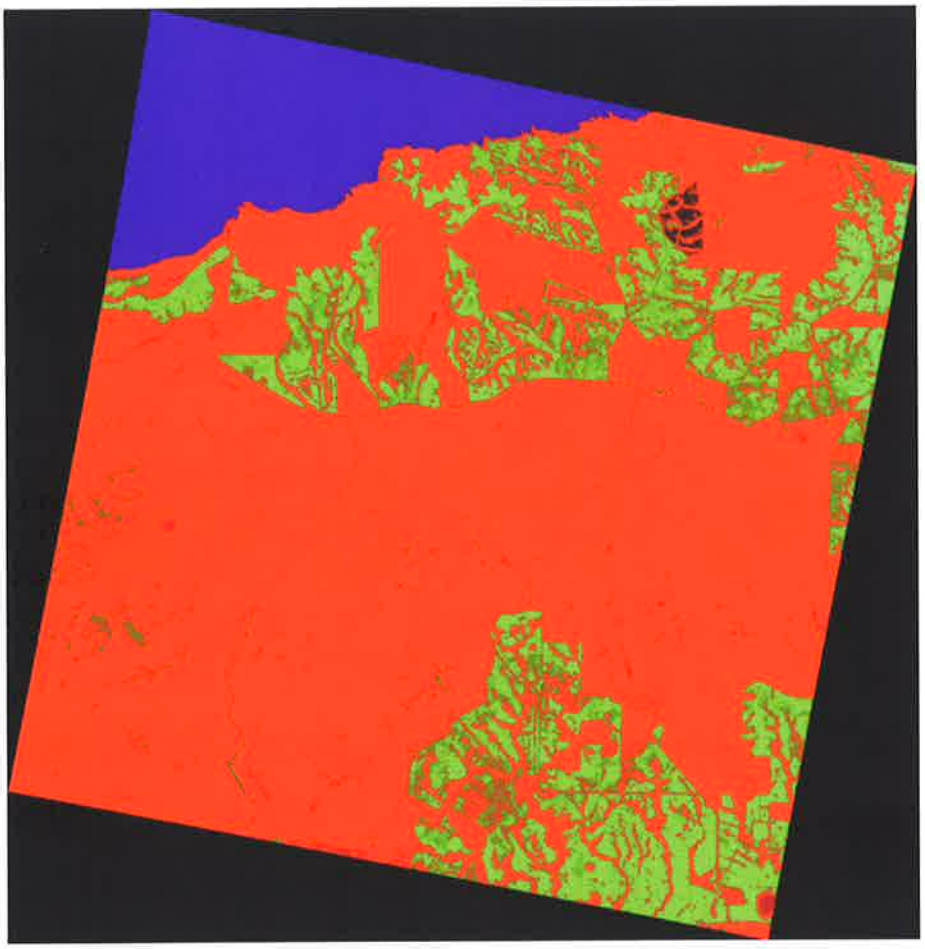

Figure 5.20: Combined Fuzzy Classified Image (forest-red; pasture-green; sea-blue), TM93

concluded that the Mahalanobis Distance fuzzy classifier offers a useful approach to land use classification. The higher spatial resolution TM89 and TM93 images were then classified using this approach. Finally, the accuracy assessment of fuzzy classified images was discussed.

Chapter 6

CHANGE DETECTION WITH CLASSIFIED IMAGES

Summary. This chapter introduces the notion that traditional Post Classification Comparison (PCC) change detection can be posed as a problem in Boolean logic, and shows some traditional PCC change detection results for the study region. It then introduces and discusses two candidate approaches to fuzzy PCC change detection: simple arithmetic operations and a fuzzy logic approach. The fuzzy logic approach is shown to be satisfying and highly functional; it represents a major contribution of this thesis.

6.1 Introduction

As noted earlier, a simple and intuitively attractive approach to the detection of changes in imagery is to independently classify each of the images, and then compare their class labels. This will not only detect that changes have occurred, and where, but will also identify the precise nature of the changes. Unfortunately, traditional 'hard' classifiers commonly suffer quite poor accuracy. This problem compounds in Post Classification Comparison (PCC) change detection. Pixels in which no change

has occurred, but were classified incorrectly on the first date and not on the second, and vice versa, will show up as false change detections. Pixels incorrectly classified on both dates may also show falsely as changed. The change detection approach adopted in this thesis involves the use of fuzzy classifications: rather than force the allocation of a single class label to each pixel, we consider that each pixel belongs, to some degree, to each class. It is expected that such an approach will show superior results to traditional PCC.

6.2 'Hard' Post Classification Comparison

The traditional approach for PCC (i.e. with 'hard' classification) is to simply compare class labels for the two (or more) separate dates. If they are different, it is concluded that a change has occurred. It makes no particular sense to undertake the comparison using arithmetic operations: even if the class labels are numbers, they are essentially arbitrary, and the same end result could derive from a number of situations. For example, an arithmetic difference of 3 could result from class 6 on t_1 and class 3 on t_2, or class 4 on t_1 and class 1 on t_2, and so on. An arithmetic operation on the class labels will therefore detect change, but will lose, unnecessarily, information about the nature of that change.

The comparison between class labels on two or more dates is therefore, with 'hard' PCC, essentially logical. We seek those pixels in which the class label on date 2 is not the same as the class label on date 1. Alternatively, we might seek those pixels in which the class label on date 1 is, say, class 7, and the class label on date 2 is, say, class 4. If we use the latter approach, not only is change detected, but the nature of the change is also known (i.e. the change is identified).

Logical operations of this nature (i.e. 'two-valued') are termed Boolean. The logical operator AND is normally termed conjunction. Expressions in Boolean logic involving only conjunctions return TRUE if and only if all antecedent conditions are TRUE, and they return FALSE otherwise. Of course, we can also construct queries involving other logical operators, such as disjunction (OR) and negation (NOT) if this suits our problem. In Boolean logic, a disjunction returns TRUE if any (or all) of the antecedent conditions are TRUE, and FALSE otherwise. Negation simply changes TRUE to FALSE and FALSE to TRUE.

Traditional PCC, posed in a logical form, consists of statements such as: in which pixels is the class label on date 2 different from the class label on date 1, or which pixels have a class label of 3 on date 2 AND a class label of 2 on date 1? Figure 6.1 shows the result of a 'hard' PCC of this form. Pixels that were assessed to be forest in 1989 AND pasture in 1993 are shown in white. Pixels that do not satisfy this particular condition are shown in black. Of course, this does not imply that we have displayed all changes; only the specified change class in particular.

Errors attributable to misregistration between the two images can be reduced using a morphological opening filter. Figure 6.2 shows the traditional PCC image, after filtering with a 7 by 7 diamond shaped morphological opening filter.

6.3 Change Detection with Fuzzy Classified Images

In the case of change detection with fuzzy classified imagery, we do not have single class labels to compare. Instead, we have the degree of membership of each pixel, in all classes. We wish to compare the corresponding class memberships to determine if

Figure 6.1: Traditional Post Classification Comparison, Forest 1989 AND Pasture 1993

there has been change and, if so, what it is, and to what extent do we believe that change has occurred.

6.3.1 Fuzzy Image Differencing

Because the fuzzy set/class memberships are numbers, we can compare them, and determine the extent of change that has occurred, by simple arithmetic operations, on a "per pixel per class" basis. The two most obvious candidates are subtraction and division. Using subtraction (differencing), results near zero indicate that little change

Figure 6.2: Traditional Post Classification Comparison, Forest 1989 AND Pasture 1993, After Morphological Filtering

has occurred (i.e. the pixel remains approximately as similar to the prototypical class member at the second date as it was at the first date), negative numbers indicate that the pixel is now less like the prototypical class member, and positive numbers indicate that the pixel is now more like the prototypical class member. The magnitude of the number reveals the extent of the change. In the case of division (ratioing), the discussion is similar, but now we look to numbers near unity, less than unity, and greater than unity, as indicating no change, less like and more like respectively.

Ratioing holds no apparent advantage over differencing for comparing fuzzy

set/class memberships. It has a disadvantage in that there is a significant compression of values less than unity relative to values greater then unity. This will not cause a problem for processing by machines, but it may for human interpretation and a non-linear contrast stretch may need to be applied. Subtraction of fuzzy set/class memberships therefore seems more suitable than ratioing for an arithmetic method of determining whether change has occurred, and, if so, the extent of that change, and what the nature of that change is.

If we wish to output a binary YES/NO result, we continue to face the problem of threshold selection. But we do not need to output a binary result. With a fuzzy classification, we essentially have an $m \times n \times c$ array for one date (where $m \times n$ is the number of pixels (i.e. m rows and n columns in the original image), c is the number of classes, and the elements of the array reveal the fuzzy set membership of the particular pixel x_{pq} in the class c_k). Denote this array $\mathbf{F}_i$. Suppose we also have an $m \times n \times c$ array for another date, denoted $\mathbf{F}_j$. We can easily compute the $m \times n \times c$ array $\mathbf{F}_i - \mathbf{F}_j$, denoting this $\mathbf{F}_{i-j}$. The elements of this array contain the change information.

The individual elements of $\mathbf{F}_i$ and $\mathbf{F}_j$ are in the range $[0, 1]$, and

$$\sum_{k=1}^{c} (f_{p,q,k})_i = 1 \text{ where } (f_{p,q,k})_i \in \mathbf{F}_i \tag{6.1}$$

$$\sum_{k=1}^{c} (f_{p,q,k})_{i-j} = 0 \text{ where } (f_{p,q,k})_{i-j} \in \mathbf{F}_{i-j} \tag{6.2}$$

The individual elements of $\mathbf{F}_{i-j}$ are in the range $[-1, 1]$, and are therefore not, in this form, interpretable as fuzzy set memberships (a fuzzy set membership must be

in the range $[0, 1]$). We could, of course, perform a simple arithmetic operation to ensure that all values were in this range e.g. add one to each value, then divide by two. But this does not really give us a result that can be interpreted as membership in some meaningful fuzzy set.

However, although the result cannot meaningfully be interpreted as a fuzzy set membership, it is never-the-less meaningful. For a particular pixel x_{pq}, the elements of $\mathbf{F}_{i-j}$ reveal the extent to which the pixel has become more like, or less like, each class.

Let us informally define the function $\text{INDEXMAX}_k(A)$ to take as input an array A (comprising the elements $a_{ij\ldots}$), and return as output the value of k such that $a_{ij\ldots}$ is maximal in the 'row' k: $INDEXMAX$ tells us the index address of the element of the array in which the maximum element of the particular 'row' is found. If desired, a more formal definition of $\text{INDEXMAX}_k(A)$ could be posed in terms of the established function $ARGMAX_k(h\{i, j, \ldots\})$ where $h\{i, j, \ldots\} = a_{ij\ldots}$. Similarly (informally) define a function INDEXMIN_k to reveal the index of the minimal element in any particular 'row'.

INDEXMAX_k $(\mathbf{F}_{i-j})$ for a particular pixel x_{pq} then reveals the class which the pixel has become maximally more like, and INDEXMIN_k $(\mathbf{F}_{i-j})$ reveals the class which the pixel has become maximally more unlike. Note that this reveals changes only: it does not (necessarily) say which class the pixel is now most like, and which class the pixel is now least like (although these classes may well coincide in many cases).

To illustrate, let us suppose that the fuzzy set memberships, for a particular pixel, at times t_1 and t_2 are $(0.1, 0.1, 0.1, 0.1, 0.6)$ and $(0.6, 0.1, 0.1, 0.1, 0.1)$ respectively. The corresponding element of $\mathbf{F}_{i-j}$ will be $(0.5, 0, 0, 0, -0.5)$ (i.e. we calculate the values

by simple subtraction).

INDEXMAX$_k$ ($\mathbf{F}_{i-j}$) will return class 1, and INDEXMIN$_k$ ($\mathbf{F}_{i-j}$) will return class 5. The pixel has become maximally more like class 1, and maximally more unlike class 5. The pixel has also become, coincidentally, most like class 1 (INDEXMAX$_k$ ($\mathbf{F}_i$) is class 1), and (equally) least like class 5 (INDEXMIN$_k$ ($\mathbf{F}_i$) returns classes 2,3,4 and 5). In this case, the two sets of classes coincide.

Now consider a different pixel, with fuzzy set memberships at times t_1 and t_2 given by $(0.2, 0, 0, 0, 0.8)$ and $(0.4, 0, 0, 0, 0.6)$ respectively. The corresponding element of $\mathbf{F}_{i-j}$ will be $(0.2, 0, 0, 0, -0.2)$.

The pixel is still most like class 5 (INDEXMAX$_k$ ($\mathbf{F}_i$) is class 5), but is also maximally more unlike class 5 (INDEXMIN$_k$ ($\mathbf{F}_{i-j}$) is also class 5). The pixel has become more like class 1 (the corresponding element of $\mathbf{F}_{i-j}$ is positive), and indeed has become maximally more like class 1 (the corresponding element of $\mathbf{F}_{i-j}$ is the row maximum), but class 1 is not the most like class (class 5 is). In this case, the two sets of classes do not coincide.

If it is required to display the change information, a number of monochrome images (one for each class) could be created. The elements of $\mathbf{F}_{i-j}$ could be translated and scaled to suit the display device, say by adding one and dividing by two (to bring them into the range $[0, 1]$), then multiplying the result by $2^l - 1$, where l is the number of bits in the display device. For $l = 8$ (i.e. an 8-bit display), $2^l - 1 = 255$. A mid grey level pixel could then be interpreted as having not changed its degree of likeness to the particular class, between the two dates. A dark pixel could be interpreted as now being less like (a prototypical member of) that class, and a bright pixel could be interpreted as now being more like (a prototypical member of) that class. The display might be further improved by applying a contrast stretch over the maximum

and minimum values achieved after this translation and scaling operations. If the contrast stretch was linear, the only concern to be faced in interpretation would be that 'zero change' areas may no longer be mid grey. If the contrast stretch was non-linear (or even piece-wise linear), an additional concern to be faced in interpretation would be that the pixel brightness values would not be linearly related to the degree of change in the pixel.

We could create a variety of colour displays to aid in the interpretation of change information, relating to one class at a time. To illustrate, we could multiply the value of each element of $\mathbf{F}_{i-j}$ by $2^l - 1$. If the value is negative, assign the absolute value to, say, red. If the value is positive, assign the value to, say, green. Such images, while arguably being more pleasing to the eye, contain no more information than the monochrome images described earlier. The use of colour does, however, permit us to include information about the degree of class membership on either, or both dates, as well as the change information. For example, we could add to the red-green image described above, a blue value determined by the degree of likeness of the pixel to the class on the second date. Alternately, we could take the monochrome image values described above and assign the values to, say, red (i.e. with both negative and positive change information displayed in the one colour, with the mid range colour representing no change). We could then plot a green value determined by the degree of likeness of the pixel to the class on the first date, and a blue value determined by the degree of likeness of the pixel to the class on the second date. The resultant (highly colourful) image contains a lot of information, but is difficult to interpret. We could also create some images, along similar lines to above, but using a value only if it was MAX or MIN over the classes. The change information for a small number of classes could be displayed simultaneously on a colour image, but interpreting such

images was found to be difficult.

After creating various images, it was concluded that the best approach was to view and interpret the monochrome images, and, if detailed information was required, to inquire about pixel values in the bright and dark areas (i.e. those areas in which change is believed to have occurred). Such an inquiry can return the pixel class vector for both dates, as well as the pixel change vector.

Figure 6.3 shows the membership difference image for the class forest. For each pixel, the degree of membership in the class forest on the first date (1989) is subtracted from the degree of membership in the class forest on the second date (1993). This image reveals the extent with which we believe that membership in the class forest has increased or decreased over the period. The light areas represent an increase in membership in the class forest between the two dates: the dark areas represent a decrease. The area around the border of the true images may be ignored: light and dark areas in this region are a consequence of the two images not being precisely coincident.

Many of the bright areas can be explained by the fact that there were two unwanted classes in the 1989 image: cloud and shadow. These were not present in the 1993 image. This causes a number of bright areas to appear in the difference image. All bright areas may be interpreted as areas of increased membership in the class forest, but, without considering what their class was at the earlier date, we cannot separate areas undergoing reforestation from areas that were simply obscured by cloud or covered by shadow in the earlier image. These latter effectively represent "errors of commission" in the cover class change set indicating an increased membership of forest.

The dark areas of the difference image may be interpreted as areas in which

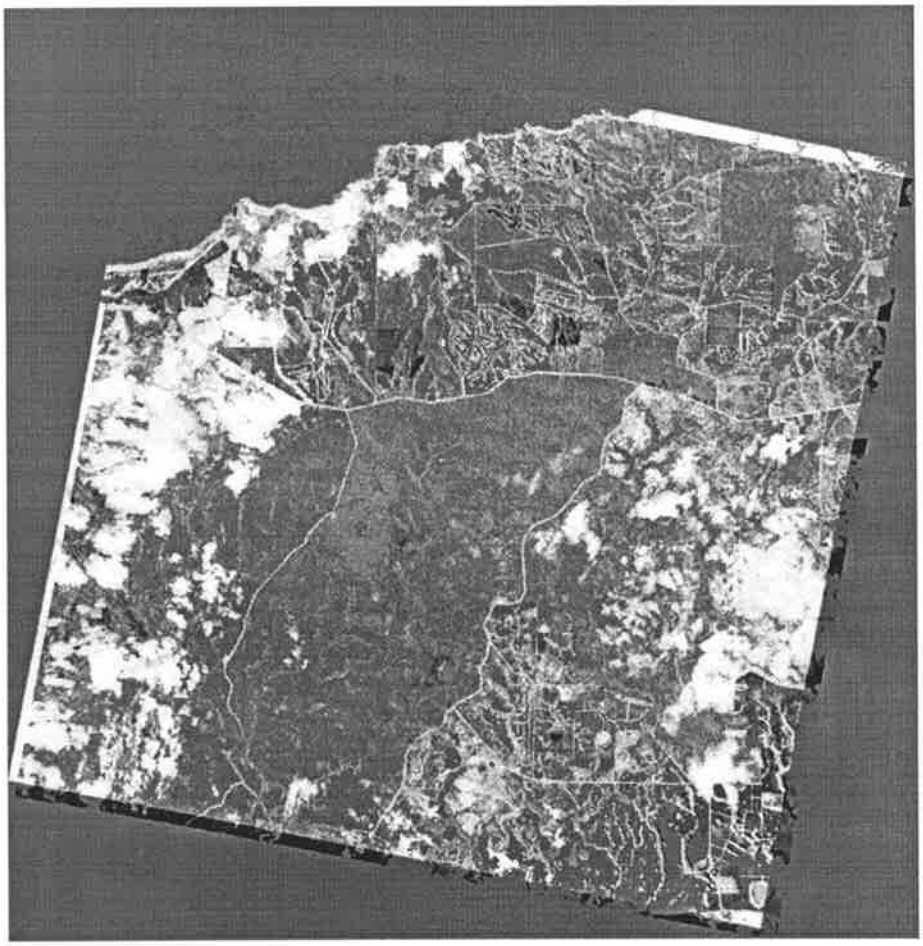

Figure 6.3: Fuzzy Difference Image, Forest 93 minus Forest 89

membership in the class forest has decreased. Because there were no cloud or shadow areas in the 1993 image, these areas may be directly interpreted as areas in which forest decline (thinning or clearing) has occurred. Areas that were obscured by cloud or covered by shadow in 1989 may, in fact, have been forest, and may, in fact, have been subsequently thinned or cleared. These will not, of course, show in the fuzzy difference image: information on such areas is, as expected, lost. These effectively represent "errors of omission".

Let us suppose that there had been cloud and shadow on the second date (1993),

and not the first (1989). Areas that were forest in 1989 but cloud and shadow in 1993 would show as dark in the forest fuzzy difference image, signifying a decreased membership in the class forest, irrespective of whether the actual ground cover class had changed or not. Unfortunately, such areas would tend to obscure information on areas in which the class had been forest in 1989, but had been converted to, say, pasture in 1993. They would be "errors of omission" in the change class "decreased membership in forest".

The situation is even more complex if various parts of both images had cloud and shadow. A particular pixel could be obscured on one date, or the other, or both, or neither, and it could either have actually undergone change, or not. We can therefore identify a number of problems with using a differencing approach to change detection with fuzzy classifications, particularly in the presence of unwanted classes such as cloud or shadow.

6.3.2 A Logical Approach to Change Detection with Fuzzy Classification

Fisher and Pathirana (1993) report on the use of fuzzy classification in change detection. Several methods are discussed, including visual inspection of corresponding fuzzy class images, and an approach that might be described as the application of Boolean logic to ranked fuzzy class memberships. Change class labels were assigned for conditions:

1. the highest ranked cover class at time t_1 (i.e. the cover class with highest membership at time t_1) was the same as the highest ranked cover class at time t_2;

2. the highest ranked cover class at time t_1 was the second ranked cover class at time t_2;
3. the highest ranked cover class at time t_1 was the third ranked cover class at time t_2; and so on.

This is, as stated, essentially a Boolean logic approach: conjunctions of antecedent conditions were tested, in a nested IF THEN ELSE structure, to assign a change code. A particular code was assigned if the first test returned TRUE. Otherwise, some other test of antecedent conditions was applied. Again, some code was assigned if the test returned TRUE. This continued until all the antecedent conditions of interest (i.e. the sets of antecedent conditions are non-exhaustive) had been tested (except, as implied, that the testing stopped as soon as one test returned TRUE). If all tests returned FALSE, some other code was assigned. It is noted that exactly one change code was applied to each pixel. Although a number of tests may have returned TRUE, the tests were performed in a particular order, and as soon as one test returned TRUE, a code was assigned and the testing stopped.

In this approach, the actual fuzzy membership values were not used; only their rankings (although it is noted that other approaches discussed in the paper did use the actual fuzzy membership values, or at least their ranges). To illustrate, consider a hypothetical pixel with class memberships of (0.95,0.05,0,0) at time t_1, and (0.05,0.95,0,0) at time t_2. Because the second ranked class on the second date is the same as the highest ranked class on the first date, this approach would determine that the likelihood that this pixel had changed between the two dates would be assessed as "not very likely". This result is not completely satisfactory.

This approach also loses information about what class a pixel was or is. For example, a pixel whose highest membership at time t_1 was forest, and whose second

highest membership at time t_2 was also forest, would be assigned to the same change class as a pixel whose highest membership at time t_1 was sea, and whose second highest membership at time t_2 was also sea. This is so irrespective of what the highest membership class was on the second date, in either case.

It is therefore concluded that, although this is a valid and useful approach for some applications, it provides no information about the change classes. The approximate measure of the 'likelihood' of change that is provided is also based only upon the rankings, not on the actual memberships involved, and we can easily hypothesise cases where this produces unsatisfactory results.

It is, as previously noted, essentially a classical Boolean approach to reasoning with fuzzy membership values. Despite the noted weaknesses (which, of course, may not be of concern in some applications), Fisher and Pathirana (1993) is important work in that it demonstrates an approach to identifying change through logical queries on fuzzy classifications. The value of fuzzy classifications has been previously discussed. Posing problems as logical queries is relatively natural, and analysts are familiar with it in a computing environment through text searches on the World Wide Web and database queries.

However, we seek a more general and powerful approach to logical querying with fuzzy classifications than that offered in Fisher and Pathirana (1993). Fortunately, there is a rich body of literature on fuzzy logic that could be applied to this problem. The next section contains theory drawn from this literature. At its conclusion, we will see that there are easily computed measures for fuzzy logic inference, with a sound theoretical basis.

6.3.3 Fuzzy Logic

Introduction

To effect inference (i.e. to draw conclusions from premises and evidence) we need only a small set of axioms, and the inference rule *modus ponens* (Thornber, 1993). *Modus Ponens* ("the mode of putting", sometimes referred to in the classical literature as "affirming the antecedent") is a common rule of valid deductive inference. It is so embedded in our method of reasoning that it is rarely explicitly stated. Simply put, it says that if we know p, and we know that p implies q, then we know q. In classical logic, the proposition "p and q" is TRUE if and only if p is TRUE and q is TRUE. In fuzzy logic, the concept "p is TRUE" can have a numeric value in the range $[0, 1]$. To effect inference with fuzzy sets, we clearly need a different approach to that of classical logic.

Such an approach was proposed by Bellman and Giertz (1973), and Fung and Fu (1975). It is developed in further detail in Dubois and Prade (1985) and Yager (1991). As noted in Dubois and Prade (1985), "the idea is to write down intuitively reasonable axioms which translate into functional equations, and then solve these equations in order to provide a mathematical representation of the class of connectives".

This will give us the means of aggregating or combining information about membership in fuzzy sets, using 'fusion' or 'aggregation' operators such as conjunction (AND) and disjunction (OR).

Axioms

Following Yager (1991), we denote the "AND-like" operator as $T(.,.)$, and the "OR-like" operator as $S(.,.)$. In either case, the desirable characteristics of the fusion operator are:

1. commutativity and associativity to ensure that the order of combining features is unimportant, and new features may be added without re-computing across all previous features.
2. the results of combining stronger evidence should be greater than that of combining weaker evidence. Further, it should only be necessary for the stronger evidence to be pair-wise stronger.

If the values of features are numeric, they may, for convenience, be scaled to the real interval $[0, 1]$, and the results of combining features also scaled to $[0, 1]$. If we also impose the (intuitively reasonable) boundary conditions $T(a, 1) = a$ and $S(a, 0) = a$, we have defined the operators t-norm and t-conorm respectively.

Mathematically:

$$
\begin{array}{lll}
T : [0,1] \times [0,1] \to [0,1] & \text{is called a } t\text{-norm operator if:} & \\
\text{a.} & T(a,b) = T(b,a) & \text{— commutativity;} \\
\text{b.} & T(a,T(b,c)) = T(T(a,b),c) & \text{— associativity;} \\
\text{c.} & T(c,d) \geq T(a,b) \text{ if } c \geq a \text{ and } d \geq b & \text{— monotonicity;} \\
\text{d.} & T(a,1) = a. & (6.3)
\end{array}
$$

$$
\begin{array}{lll}
S : [0,1] \times [0,1] \to [0,1] & \text{is called a } t\text{-conorm operator if:} & \\
\text{a.} & S(a,b) = S(b,a) & \text{— commutativity;} \\
\text{b.} & S(a,S(b,c)) = S(S(a,b),c) & \text{— associativity;} \\
\text{c.} & S(c,d) \geq S(a,b) \text{ if } c \geq a \text{ and } d \geq b & \text{— monotonicity;} \\
\text{d'.} & S(a,0) = a. & (6.4)
\end{array}
$$

It is worth noting that these differ only in the boundary conditions given by d and d'.

Operators

Parameterised families of operators satisfying these desirable characteristics are given in Dubois and Prade (1985), Bonissone and Decker (1986), and Klir and Folger (1988).

The better known operators are:

$$\begin{aligned} T_0(a,b) &= MIN(a,b) \text{ if } MAX(a,b) = 1 \\ &= 0 \text{ otherwise} \end{aligned} \tag{6.5}$$

$$T_1(a,b) = MAX(0, a+b-1) \tag{6.6}$$

$$T_2(a,b) = a \times b \tag{6.7}$$

$$T_3(a,b) = MIN(a,b) \tag{6.8}$$

$$\begin{aligned} S_0(a,b) &= MAX(a,b) \text{ if } MIN(a,b) = 0 \\ &= 1 \text{ otherwise} \end{aligned} \tag{6.9}$$

$$S_1(a,b) = MIN(1, a+b) \tag{6.10}$$

$$S_2(a,b) = a + b - (a \times b) \tag{6.11}$$

$$S_3(a,b) = MAX(a,b) \tag{6.12}$$

There are both theoretical and empirical studies that compare these operators. They are ordered:

$$T_0 \leq T_1 \leq T_2 \leq T_3 \qquad S_3 \leq S_2 \leq S_1 \leq S_0 \tag{6.13}$$

with T_0, T_3 forming the greatest lower and least upper bounds of all t-norms respectively, and S_0, S_3 forming the least upper and greatest lower bounds of all

t-conorms respectively (Bonissone and Decker, 1986).

Empirical studies comparing these and/or other operators include Tong and Shapiro (1985), Whalen and Schott (1985), Bonissone and Decker (1986), and Cross and Sudcamp (1991).

It is interesting to note that, with the inclusion of just one additional desirable characteristic, idempotence, all choice about fusion operators is removed. Idempotence ensures that if we have certain evidence or support for something, and we receive further evidence of the same value, then we neither increase nor decrease our belief in it. That is,

$$T(a,a) = a \text{ and } S(a,a) = a \tag{6.14}$$

If we impose a requirement for idempotence, the only conjunction and disjunction operators satisfying the desirable characteristics are MIN and MAX respectively. The same effect is achieved if we alternatively introduce the requirement for mutual distributivity between T and S (Dubois and Prade, 1985). Mutual distributivity requires that

$$T(a,S(b,c)) = S(T(a,b),T(a,c)) \text{ (and that } S(a,T(b,c) = T(S(a,b),S(a,c))) \tag{6.15}$$

To illustrate the use of the MIN and MAX operators for conjunction and disjunction,

> from "*feature1* AND *feature2* AND ..... AND *featureN*" (where *featureN* represents the fuzzy membership in the Nth class), we would compute

MIN(*feature1*, *feature2*, ... , *featureN)*,

and from "*feature1* OR *feature2* OR ..... OR *featureN*", we would compute MAX(*feature1*, *feature2*, ... , *featureN*).

If we wish to pose a query of the nature "*feature1* AND *feature2* BUT NOT *feature3*", we need to introduce a negation operator.

$N : [0,1] \rightarrow [0,1]$ is called the negation operator:

a. $N(1) = 0; \qquad N(0) = 1;$

b. $N(N(a)) = a \qquad$ — involution;

c. if $a \geq b$ then $N(a) \leq N(b)$. (6.16)

These axioms do not uniquely determine a negation operator (Bellman and Giertz, 1973) and, again, there are parameterised families of negation operators that satisfy them. E.g,

$$N_\lambda(a) = (1-a)/(1+\lambda a),\ \lambda > -1 \text{ (Sugeno, 1977), and} \tag{6.17}$$

$$N_w(a) = (1-a^w)^{1/w},\ w > 0 \tag{6.18}$$

(Klir and Folger, 1988, terms (6.18) the Yager class of fuzzy complements).

It is common to select $N(a) = 1-a$ as the negation operator. This is a special case of both Equations 6.17 and 6.18. If either of two additional (sometimes questioned) axioms (see Gaines, 1976, and Bellman and Giertz, 1973, for details) are added to the basic set of negation axioms, all choice is again removed: $N(a) = 1 - a$ becomes the only possible negation operator.

For negation, we would therefore compute NOT *feature3* = 1 - *feature3*,

The example "*feature1* AND *feature2*, but NOT *feature3*" would be computed as MIN(*feature1*, *feature2*, 1 - *feature3*).

It is noted that there is another class of aggregation operators: the "mean" operators. These are appropriate in situations where there is some degree of compromise or 'compensation' between features: situations when we say AND but really mean something a little different. For example, I want to buy a car that is RED AND FAST AND INEXPENSIVE-TO-BUY AND EYE-CATCHING AND INEXPENSIVE-TO-RUN AND EASY-TO-MAINTAIN AND ... . Neither AND nor OR accurately captures this situation, for we may choose a car that is a little less FAST (say), in order that it is a little more INEXPENSIVE-TO-BUY. Thus, AND does not accurately reflect this decision process (recall that AND-like operators are upper bounded by MIN). Similarly, an OR-like operator is not appropriate, for there is a limit to how much less FAST we will choose in order to increase the feature INEXPENSIVE-TO-BUY (recall that OR-like operators are lower bounded by MAX).

Mean operators are normally defined in such a way that they return values in the closed interval [MIN, MAX]. They are commutative and idempotent, but seldom associative (the only associative means are medians defined by $med_\alpha(a,b) = median(\alpha, a, b)$). For a more detailed treatment of mean operators, see Dubois and Prade (1985), Klir and Folger (1988), or Yager (1991). A study using an aggregation operator of this type for image feature fusion may be found in Abdulghafour (1992), but no application is identified for change detection in this thesis.

Discussion

A number of authors offer a treatment of the relationship between t-norms, t-conorms and negation operators under a generalised De Morgan's Law (which requires that $N(T(a,b)) = S(N(a),N(b))$ and $N(S(a,b)) = T(N(a),N(b))$), and note that the three cannot be independently defined (see Bonissone and Decker, 1986; Klir and

Folger, 1988; or Yager, 1991).

Zadeh's original paper on fuzzy sets (Zadeh, 1965) proposed MIN, MAX and $1 - a$ for fuzzy set intersection, union and complementation respectively. Bonissone and Decker (1986) observed that these were a common selection for most expert systems. Klir and Folger (1988) identify a highly desirable feature of these operators: their inherent prevention of error compounding. This feature is lacking in most other candidate fuzzy set operators. The literature shows that MIN, MAX and $1-a$ remain valid choices, from an axiomatic viewpoint, for aggregation operators, but it is noted that they may not be a unique choice.

The arguments in support of their use are solid, but by no means overwhelming. We could, under certain conditions, argue for the use of $S_1(a,b) = MIN(1, a+b)$ for disjunction. If we consider the fuzzy classification of a particular pixel, the unity sum condition for memberships means that the membership in class a OR class b would reduce to the sum of the memberships in the two classes ($S_1(a,b) = MIN(1, a+b) = a + b$, because $a + b \leq 1$). This is intuitively quite reasonable, even desirable. Computing disjunction over all possible classes would return a membership of unity. S_1 would thus appear to be superior to computing the membership in two or more classes as the maximum of the memberships in the separate classes.

Unfortunately, if we wish to be able to construct and simplify logical queries involving conjunction, disjunction and negation in accordance with the conventions of logic, we cannot define them all separately: conjunction and disjunction must be "De Morgan duals". The dual of $S_1(a,b) = MIN(1, a+b)$ is $T_1(a,b) = MAX(0, a+b-1)$. Because $a + b \leq 1$ (for any two classes for a particular pixel on a particular date), $T_1(a,b)$ will be identically equal to 0. That is, any conjunction of two classes would have zero membership. While this is satisfactory in traditional classification, it does

not accord well with the philosophy of fuzzy classification, in which a particular pixel can have a (non-zero) degree of membership in two or more classes simultaneously, and any conjunction of two (non-zero) class memberships should be non-zero.

Arguments can be formulated for the use of other possible operators. These arguments are largely based on an appeal to intuition to have disjunction return a value higher than the maximum of the input values e.g. to have $S(0.8, 0.9)$ to return a higher value than $S(0.1, 0.9)$. Palubinskas *et al.* (1995) used this line of reasoning in choosing to investigate the use of Yager functions for combining the outputs of fuzzy classifiers, with an aim of increased classifier accuracy.

The Yager function for disjunction and conjunction are, respectively

$$u_w(a,b) = MIN\left[1, (a^w + b^w)^{1/w}\right] \tag{6.19}$$

$$i_w(a,b) = 1 - MIN\left[1, (\{1-a\}^w + \{1-b\}^w)^{1/w}\right] \tag{6.20}$$

(the Yager function for negation was given earlier as Equation 6.18)

Klir and Folger (1988) show that, in the limit as $w \to \infty$, $u_w(a,b) \to MAX(a,b)$ and $i_w(a,b) \to MIN(a,b)$. It can also be shown that, for $w = 1$, the Yager functions reduce to $u_1(a,b) = S_1(a,b) = MIN(1, a+b)$, $i_1(a,b) = T_1(a,b) = MAX(0, a+b-1)$.

Not withstanding the intuitive appeal of these functions, Palubinskas *et al.* (1995) found that MIN and MAX performed comparatively with the Yager functions with $w = 2$. No other values of w were tested. We note that there would seem little point in investigating values of $w > 2$, as they would return values increasing close to MIN and MAX.

As stated earlier, the penalty for choosing some intuitively appealing disjunction operator (other than MAX) is that we should accept an obligation to choose the De

Morgan dual as the conjunction operator, and this will return a value less than the minimum of the two values.

The above discussion on choosing aggregation operators has been based upon values being numeric. In some cases, it may be more appropriate to select (non-continuous) values from an ordered linguistic set. In this case, Yager (1979) has shown that MIN and MAX are the only acceptable operators for intersection and union respectively. The selection of a negation operator may, however, be problematic. This matter is, however, merely academic: in this thesis, all memberships are defined numerically.

Choice of Aggregation Operators

It has been shown that general class of the aggregation operator is determined by the axioms that it is required to satisfy. We have stated, following the literature, a reasonable set of axioms for such operators. We have noted that the operator may not be uniquely specified by the chosen axioms. We have, however, chosen to use MIN, MAX and $1 - a$ exclusively in this thesis. They are sound, easily computed, and do not compound errors. As far as is known, there has been no previous use of fuzzy logic operators for change detection to influence this decision.

The relative merits of various fuzzy logic operators for change detection in imagery are not investigated. It is considered that determining the optimal choice of fusion operators for change detection requires further study, preferably with detailed 'ground truth' to allow quantitative evaluations.

6.3.4 Fuzzy Post Classification Comparison

We now have the machinery to compute any set of compound conditions, in a multi-valued fuzzy logic, that we can pose in a logical form. The statement "pixel x is in class A at time t_1" is, in such a fuzzy logic, TRUE, to the extent $\mu_A(x, t_1)$. We can therefore compute the extent to which the statement "pixel x is in class A at time t_1 AND in class B at time t_2" is TRUE: $MIN(\mu_A(x, t_1), (\mu_B(x, t_2))$. This is the membership of the pixel x in the specific change class (A, B). Because, as has been noted, MIN forms the least upper bound on the t-norm conjunction operators, it will provide the highest reasonable (i.e. in accordance with our stated desirable axioms) estimate of membership in such a change class.

Similarly, we can compute any other change class of interest. For example, "pixel x is in class A OR in class B at time t_1, AND in class C at time t_2" and so on (including the ability to compute negations, such as "AND NOT in class D"). We are thereby able to aggregate classes into superclasses (using disjunction), and to allow (to some extent) for the presence of unwanted classes such as cloud and shadow.

This approach lets us concentrate on particular change classes or superclasses, whereas:

1. the differencing (subtraction) approach only lets us consider whether a pixel has become more or less like a particular class.
2. the approach of Fisher and Pathirana (1993) aggregates change classes and deals only with rankings, and thereby loses valuable information on change class, and 'likelihood' of change.

We now use these results to compute changes between our two Kangaroo Island images.

We can reason with the extent to which pixels have changed from one particular class to another. To illustrate, "to what extent has change occurred from forest to pasture?" This is shown in Figure 6.4.

Figure 6.4: Fuzzy Post Classification Comparison, Forest 1989 AND Pasture 1993

Figure 6.5 shows the result of morphologically filtering this image to remove error effects of isolated 'misclassified' pixels, and misregistration between the images.

Figures 6.4 and 6.5 represent a major contribution of this thesis. The conversion of forest to pasture or pine was of primary concern in this study: the figures show

Figure 6.5: Fuzzy Post Classification Comparison, Forest 1989 AND Pasture 1993, After Morphological Filtering

regions that changed from forest to pasture over the image acquisition times. Regions that were not clearly defined, or not shown at all in the traditional PCC image (Figure 6.1) are clearly apparent in these figures, thus establishing the efficacy of the fuzzy PCC approach. The large regions in the upper centre of the figures are of particular interest. The straight edges of the boundaries of these regions draws attention to them, suggesting human-induced change. The regions at the far left centre, although displaying similar memberships in the change class, do not exhibit straight boundaries. A computer algorithm may well treat regions in these two areas of the figures equally,

but a human analyst, on visually inspecting the change images, will undoubtedly (if seeking human-induced change) concentrate on the upper centre regions. Of course, the far left centre regions may well remain of interest in that they indicate change, albeit more likely of natural causes. It is noted that these regions were not apparent in Figure 6.1. The effect of morphological filtering is, as expected, to filter out the majority of the effects of image misregistration.

The extent to which pixels are believed to have changed from forest to pine is shown in Figure 6.6.

Figure 6.6: Fuzzy Post Classification Comparison, Forest 1989 AND Pine 1993

The result of the morphological filtering of this image is shown in Figure 6.7, which reveals there were no significant regions that exhibited the spectral characteristics of forest in 1989 and mature pine in 1993.

Figure 6.7: Fuzzy Post Classification Comparison, Forest 1989 AND Pine 1993, After Morphological Filtering

We can also reason with the extent to which a pixel was and is some particular class e.g. to what extent was and is some particular pixel of class forest? This is shown in Figure 6.8.

Queries can be extended to be of the form "was some particular class, but is

Figure 6.8: Fuzzy Post Classification Comparison, Forest 1989 AND Forest 1993

now NOT that class?" i.e. the extent to which membership in a particular class has diminished. This is illustrated in Figure 6.9, which shows areas that have become 'less like' forest i.e. forest in 1989, but not forest in 1993. Both of these concepts (forest in 1989, and not forest in 1993) are fuzzy: each pixel possesses them to a certain degree. Membership in 'not forest' in 1993 is computed as one minus membership in forest in 1993; conjunction (AND) is computed by the MIN function.

The logical approach provides results that are in some senses comparable with

Figure 6.9: Fuzzy Post Classification Comparison, Forest 1989 AND NOT Forest 1993

the difference approach discussed earlier. They are, however, superior in that we can pose our query in a manner to focus on particular classes and changes, and reduce or remove error due to unwanted classes such as cloud or shadow. Let us suppose that a pixel was forest at time t_1, and obscured by cloud at time t_2. The simple difference of the class memberships at these two times will show a significantly diminished membership of forest, thereby leading to some measure of (erroneous) belief that change has occurred. The logical query "forest at time t_1 AND (forest OR cloud) at time t_2" would not cause this 'error'. Furthermore, the results can now

be interpreted as memberships of a meaningful fuzzy set (it is a fuzzy set because the values are $[0, 1]$; it is meaningful because the fuzzy set can be interpreted as the change class "less like (some particular class)"). Such results could therefore be used in some further compound analysis using fuzzy logic.

6.4 Conclusion

This chapter posed traditional PCC change detection as a problem in Boolean logic, and then showed, through some formal fuzzy logic theory, that a similar approach could be taken to PCC with fuzzy classifications. It is suggested that such an approach is considerably less prone to the compounding error problem experienced by traditional approaches, and reveals more subtle changes. Change detection problems posed in this manner provide a general, powerful and intuitive approach to formulating and computing any query relating to fuzzy class memberships on two (or more) dates. It would be a simple matter to construct an interface that is familiar to any computer user with experience in formulating queries for WWW search engines or databases. Such an interface need only display the classes of the two (or more) date fuzzy classified images, and have buttons for AND, OR and NOT. It is noted that some or all of the classified images do not even strictly need to be fuzzy, and that some or all of the inputs do not even strictly need to be images: they could be any other form of geospatial data, fuzzy or not. This matter is not pursued further in this thesis, but it does seem a potentially fruitful area for investigation of the fusion and analysis of disparate geospatial information types, irrespective of the fuzzy or 'hard' nature of the data.

Chapter 7

THE VALUE OF THE FUZZY EXPONENT

Summary. The Mahalanobis Distance supervised fuzzy classifier requires the a priori selection of two important parameters: the number of classes, and a fuzzy weighting parameter, or fuzzy exponent, m. The number of classes is essentially determined by the problem and the data: this is not the case with the fuzzy weighting parameter. Earlier literature has noted that there is no theoretical and only limited empirical guidance on the selection of m. It has also noted a strong correlation between fuzzy memberships and proportions of mixed classes. This provides our approach to an examination of the value of m: we require that the fuzzy memberships reflect the proportions of contributing classes for the two class mixed pixel case, but exhibit a high membership of the appropriate single class in the case of pure, typical pixels. Empirical results are presented which show a conflict between the requirements of these separate cases. A theoretical investigation of the value of m is also shown, again subject to the condition that the fuzzy memberships reflect the proportions of contributing classes for the two class mixed pixel case. This investigation reveals that the theoretically optimum value for m is a function of both the ratio of the class

distance metrics (in this work, the Mahalanobis norm), and the ratio of proportions of contributing classes (the very thing which we would often seek to estimate with the fuzzy classification approach). A special case is shown to occur if the ratio of class distance metrics is unity (as is the case, for example, with hyperspherical class distributions): in this case, $m = 0.5$ is shown to give the required result, for all class mixture proportions. This special case not-with-standing, the sensitivity of the classifier to the selection of particular values of m is investigated and shown for a range of distance and proportion ratios.

7.1 Introduction

It is noted that in both the unsupervised fuzzy c-means (FCM) classifier (clusterer) and the supervised Mahalanobis Distance fuzzy classifier, we have to make some *a priori* choices about the number of classes c, and the fuzzy exponent m. For the work reported in this thesis, using the supervised Mahalanobis Distance classifier for a specific application and set of data, the number of classes c is, to some extent, determined by the application. We have an interest in the native forests, and in the two classes that they could change to: pasture/cropland, and pine plantations. The other major class apparent in the region covered by the images is sea. While there are small areas of a number of other classes, such as swamps, sand dunes, lakes, and roads, the principal classes of interest are the four classes mentioned earlier.

The value of the fuzzy exponent m remains, however, to be determined.

The fuzzy exponent, denoted in this thesis (and in Choe and Jordan, 1992) by q, has previously been defined in a different manner (Bezdek *et al.* 1984; Foody, 1992; Choe and Jordan, 1992). The relationship between q and m (the fuzzy exponent used in this thesis) is

$$m = \frac{1}{(q-1)} \text{ or, equivalently, } q = \frac{(m+1)}{m} \tag{7.1}$$

This substitution is used because it permits simpler mathematical development later in this thesis.

Bezdek *et al.* (1984) suggested that, for the unsupervised fuzzy *c*-means clustering algorithm, q should be in the range 1 to 30, with the range 1.5 to 3 seeming "to give good results" (these values for q correspond to values of m in the range positive infinity to 0.03, with "good" results in the range 2.0 to 0.5). It was noted, however, that there was no strong theoretical justification or empirical evidence for these choices. Cannon *et al.* (1986) also noted that "no theoretical basis for choosing a good value ... is available", and suggested $1.1 \leq q \leq 5$ were "typically reported as the most useful range of values".

McBratney and Moore (1985) investigated the choice of q for the unsupervised FCM clustering algorithm, and found that a value of approximately 2 (corresponding to $m = 1$) was optimal (but this was somewhat dependent on the number of classes).

Choe and Jordan (1992) also addressed the matter of a good choice of q for the unsupervised FCM clustering algorithm. This work used the concepts of fuzzy decision theory, defining a goal of maximising the number of points in a cluster, and a fuzzy constraint of minimising the sum of the squared errors within a cluster. Two normally distributed classes of different means and variances were artificially generated. Using this approach, they found that the unsupervised FCM algorithm was relatively insensitive to the value chosen for q in the range 1.1 to 30. In terms of a quantity termed the "miss" classification (which is unfortunately not described or defined in the paper), they note that the results are the same with q ranging from 8

to 30. They suggest that the value $q = 12$ (approximately $m = 0.09$) is "optimal", but note that the algorithm is not sensitive to the chosen value. It is suggested that the nature of the examination, and the dataset used, provides only limited guidance.

Cannon and Jacobs (1984) found that, for image applications, $1.1 \leq q \leq 2.5$ "has proved adequate for all practical purposes". Foody (1992) used the value $q = 1.25$ (corresponding to $m = 4$) for the supervised Mahalanobis Distance fuzzy classifier based on the FCM clustering algorithm (as described in Chapter 5). It was stated that this value was qualitatively assessed to give good results for the application at hand. We note that this application involved particular species of the general class "heathland", and we might therefore expect a strong spectral similarity, and perhaps even overlap, between these species. Therefore the reported results do not necessarily provide definitive guidance for the choice of m for the work in this thesis, where pure examples of the classes of interest are spectrally distinct.

Foody (1996) reports on the FCM classification of three relatively distinct classes: trees, grass and asphalt. Spectral overlap existed only between two of the classes, and only in one band. A coarse resolution image was synthesised by spatially degrading a finer resolution image. The correspondence between fuzzy memberships and the known proportions of 35 synthetic pixels were examined, for various values of the fuzzy exponent q. It was found that $q = 2$ ($m = 1$) performed best. It is, however, noted that the correlation coefficients for the relation between q and the class proportions differed only in the second or third decimal place between $q = 2$ and $q = 1, 1.5, 2$ and 2.5.

In this chapter, choices for the value of m are examined both empirically (with respect to the remotely sensed image data used) and analytically. The sensitivity of the classifier of the fuzzy supervised Mahalanobis Distance classifier to the value of

m is also investigated.

7.2 Approaches to the Mixed Pixel Problem

7.2.1 The Mixed Pixel Problem

The mixed pixel problem occurs when, at the scale of observation, we have a number of classes contributing to the observed spectral response of the pixel. There are a number of ways in which mixed pixels can arise in remotely sensed imagery: they are illustrated in Figure 7.1.

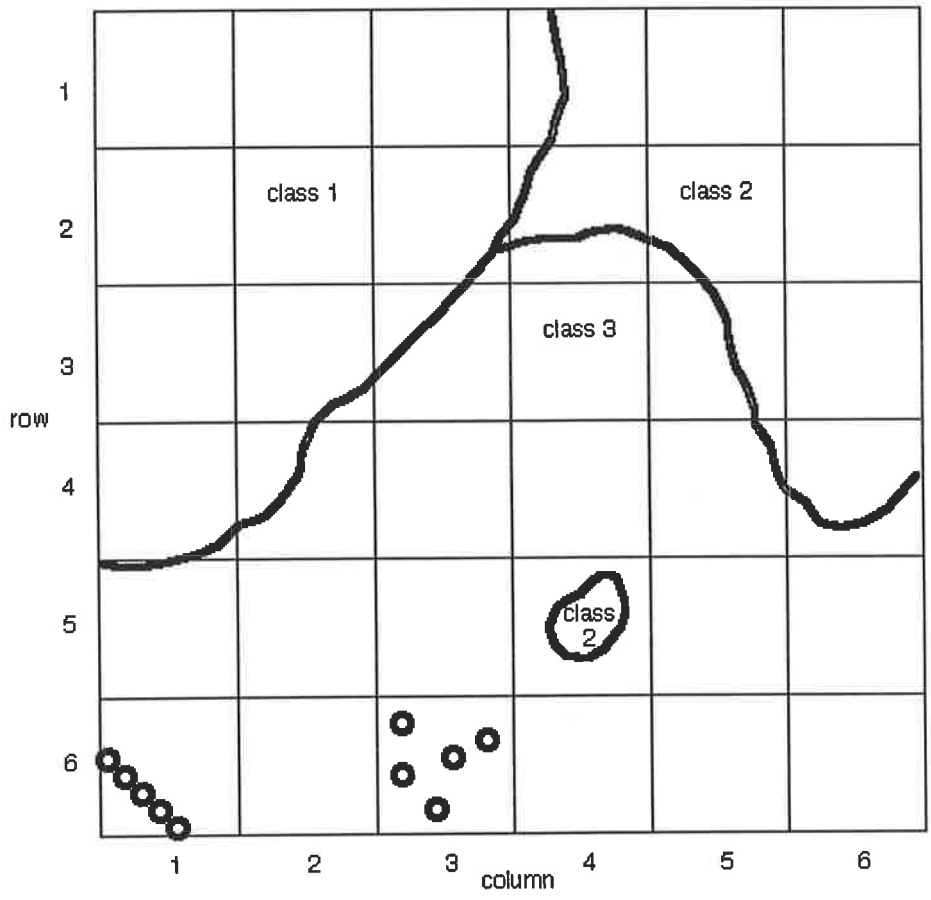

Figure 7.1: Illustration of the Mixed Pixel Problem in Remote Sensing

Pixels (row, column) (3,5) and (5,4) have the same proportions of classes 2 and 3. Given relatively homogeneous classes, and ignoring the Point Spread Function (PSF) of the imaging device (i.e. ignoring the effect of surrounding pixels on the imaged values of a particular pixel), we expect to see similar spectral responses in these two cases. Pixel (3,5) represents a distinct (physical) boundary between two (homogeneous) classes, where each of the classes are significantly larger (in area) than a pixel dimension. Pixel (5,4) represents a case where there is an object in the Instantaneous Field of View of the sensor that is smaller than a pixel dimension. Pixels (6,1) and (6,3) show a number of singular objects (say, trees) that collectively comprise a significant areal proportion of the pixels.

We observe that the notion of homogeneity or heterogeneity is essentially scale dependant. For example, an observation of a small area of a gibber desert may be considered, by a human observer, to be homogeneous. The observation will, however, contain sub areas of gibbers (rocks) of various types, sand (of various compositions/moisture levels), shadows, ground litter, the occasional animal (such as a small lizard), etc. All of these sub areas contribute something (spectrally) to the observation. Therefore even areas considered to be homogeneous will normally contain class mixtures, albeit classes that we may not wish to consider as discrete classes for our purposes, at that time. It may suit our purposes to consider an area to be homogeneous, even though we know it to contain a number of disparate spectral and ground classes. An area classified as native forest may contain a number of individual species, and we will normally be able to see through the canopy to some ground litter, shadow, understorey species, grasses, etc. Such is the case with the study undertaken for this thesis. We wish only to reason with relatively high level geographic and vegetation abstractions, such as forest, pine, pasture and sea. We do

not wish to be concerned with the hundreds of species and other spectral classes (such as ground litter, sand, soil, plant shadow etc.) which are undoubtedly represented in our study area, nor with variations within all of these classes caused by different conditions, density, stages of growth, or numerous other factors.

7.2.2 Classical Approach to Spectral Unmixing

Consider a perfect sensing environment, one in which all photons collected by the sensor at a single time 'instant' emanated from the single ground imaged area. Consider also that we possessed a library of spectral signatures for each class that is actually represented in the area of interest. If each class exhibited no variability, and there was linear independence of the class signatures, we could, in principle, 'unmix' each observation (i.e. the recorded response for a particular pixel), provided that there were no unknown classes, and that the number of observed 'dimensions' (e.g. bands in the case of remotely sensed imagery) was at least as large as the number of classes.

Of course, these ideal conditions are rarely, if ever, satisfied. First, sensors record photons that do not actually emanate from the ground imaged area. In reality, we also have photons scattered into the IFOV of the sensor by atmospheric particles, and these photons may have 'originated' from adjacent ground areas, or from the original incoming radiation. Second, classes will exist for which we do not hold spectral signatures. Third, each class does actually exhibit some (spectral) variability. Fourth, it is often the case that class signatures are not linearly independent. That is, it may be possible to construct a particular class signature from a linear combination of two or more other class signatures. Fifth, the number of actual classes often exceeds the number of observed 'dimensions'.

Techniques have been developed to address some of these problems, but this classical approach to class unmixing (sometimes called "end-member analysis") remains an active area of research in pattern classification and recognition.

7.2.3 Interpretation of Fuzzy Memberships as Mixel Proportions

Warner and Shank (1997) note that an alternative approach to class unmixing is to use classifier outputs as estimates of class proportions. For example, Marsh *et al.* (1980) used class Mahalanobis Distances for this purpose. Their approach is somewhat related to the fuzzy classification approach adopted in this thesis, but could unfortunately only be applied to mixtures of two classes.

The adoption of fuzzy classification approaches has been motivated, in many studies, by the presence of the mixed pixel problem. Some earlier work on the interpretation of fuzzy set memberships of pixels in remotely sensed images as class proportions in mixed pixels has produced good results. Wang (1989, 1990a, 1990b), Fisher and Pathirana (1990), Pathirana (1990, 1993), Foody (1992, 1994, 1996), Nishida *et al.* (1993), Foody and Cox (1994), Maselli *et al.* (1996), Warner and Shank (1997) and Atkinson *et al.* (1997) have all suggested, with varying degrees of empirical support, that there is a strong relationship between fuzzy memberships (derived from various approaches) and proportions of ground cover.

It has been observed that fuzzy memberships can be derived from many of the hard, statistical pattern classification approaches (Schowengerdt, 1996). Schowengerdt (1996) also notes that it is tempting to interpret classifier likelihood measures as measures of class mixing, and addresses the question "can these likelihood indicators also measure class mixing proportions?". The results were at least partially

equivocal. With two analyses, "One shows that such information can be extracted from the likelihoods, if the classes involved are separable, while the other indicates that it may not be possible to use classifier likelihood as a global detector of mixing". Campbell and Hashim (1992) also question the interpretation of fuzzy class memberships as proportional representation, noting that there is no theoretical justification for it. These papers not-with-standing, the interpretation of fuzzy class memberships as proportional representation has considerable support in the literature, and offers an approach to determining a suitable value for the fuzzy exponent m in the Mahalanobis Distance fuzzy classifier used in this thesis.

7.3 Empirical Investigation of the Value of m

7.3.1 Evaluation Criterion and Data

Maselli *et al.* (1996) notes that "quantitative investigations on the actual relationships existing between fuzzy membership grades and cover proportions would be valuable". In this section, this problem is approached in the opposite direction. The criterion that the computed fuzzy memberships of a pixel should reflect the true class proportions of that pixel is used in determining a suitable value for the fuzzy exponent m in the Mahalanobis Distance fuzzy classifier. This includes, of course, the special case where the pixel actually comprises only one class.

There are a number of approaches to an empirical determination of a suitable value for the fuzzy exponent m, subject to this criterion. If sufficient ground truth data reflecting true class proportions is held, and if there is confidence that the image has been registered to that data sufficiently accurately, the optimal value of m could be found via some simple algorithmic approach (e.g. the minimisation of the sum of squared errors). This approach was not adopted for this study: ground truth data

with class proportions was not held, and it was considered impractical to collect it in sufficient quantity. Furthermore, the accuracy of image registration was of concern.

Another possible approach is to synthesise a coarse spatial resolution image by averaging the raw pixel values of some higher resolution image for which ground truth data is held. If the higher resolution image is believed to comprise pure pixels, we again have data to use some simple algorithmic approach to determining the optimal value of m, as before. In our case, we had no confidence that the high resolution image (our original TM image) did indeed comprise pure pixels.

The approach adopted in this thesis required some preliminary processing. It was noted earlier that there were errors in both the crisp classified TM93 image and the ground truth data. It was considered, however, that where both datasets agreed on the class of a particular pixel, we could have a high level of confidence that the pixel actually was of the agreed class. The two datasets were examined, and, where they agreed, the pixel data from the original (registered, but unclassified) TM93 image were noted (as was the agreed class). We were therefore able to calculate expected raw pixel values for each class that were obviously influenced by the raw values of the chosen class training pixels (through the classification algorithm), but were more representative of pixel values of that class across the image. Using these, and the training class means and covariance matrices, we were able to calculate the expected Mahalanobis Distances of a typical member of each class, from each of the class means. These are shown in Table 7.1.

Using the data in this table, if the class was actually forest, we would expect a pixel to exhibit a Mahalanobis Distance of 6.2 from the mean of the class forest (hereafter abbreviated to a forest Mahalanobis Distance), a sea Mahalanobis Distance

Table 7.1: Expected Mahalanobis Distances (MahDist) of a Typical Member of Each Class from Each Class Mean

Mean MahDist, given actual class				
	Actual Class			
Class MahDist	forest	sea	pasture	pine
forest	6.2	107.2	100.9	14.6
sea	5769.0	5.1	24094.0	5802.0
pasture	304.2	937.3	16.0	476.8
pine	403.6	417.3	3406.0	3.8

of 5,769, a pasture Mahalanobis Distance of 304.2, and a pine Mahalanobis Distance of 403.6. Similarly, if the actual surface cover class imaged in the pixel was some other class, we would expect it to exhibit Mahalanobis Distances as indicated in the appropriate column.

A more complete table (Table 7.2) of relevant data of distances, given that the pixel is actually a member of some particular class, is shown for completeness.

Table 7.2: Statistical Data of Mahalanobis Distances (MahDist) of a Typical Member of Each Class from Each Class Mean

Actual class	Class MahDist	Min	Max	Mean	Median	Mode	Std Dev.
forest	forest	0.0	2742.0	6.2	0.0	0.0	9.5
forest	sea	89.8	62440.0	5769.0	4717.0	3987.0	2770.0
forest	pasture	34.8	3686.0	304.2	305.8	348.6	92.6
forest	pine	6.5	16548.0	403.6	200.3	71.1	515.4
sea	forest	79.3	357.0	107.2	105.3	103.2	8.3
sea	sea	0.0	340.1	5.1	3.2	0.5	6.5
sea	pasture	821.5	1189.0	937.3	934.9	932.0	17.4
sea	pine	313.9	890.3	417.3	415.2	419.7	17.0
pasture	forest	34.9	1005.6	100.9	95.6	95.6	30.9
pasture	sea	11377.0	62626.0	24094.0	23789.0	24.9	4483.0
pasture	pasture	0.0	701.5	16.0	8.3	2.8	18.5
pasture	pine	1032.0	16839.0	3406.0	3378.6	3317.0	946.8
pine	forest	5.4	29.3	14.6	14.4	13.2	4.1
pine	sea	4258.0	7648.0	5802.0	5807.0	5953.0	542.3
pine	pasture	412.7	547.4	476.8	477.0	473.0	17.5
pine	pine	0.0	21.2	3.8	3.0	0.3	3.2

For comparison, the table of Mahalanobis Distances, independent of the actual class of each pixel, is shown in Table 7.3.

Table 7.3: Expected Mahalanobis Distances (MahDist) from Each Class Mean, Independent of the Actual Class

	MahDist,Independent of Actual Classes					
Class Mahdist	Min	Max	Mean	Median	Mode	Standard deviation
forest	0.0	3722.0	35.0	0.0	0.0	48.5
sea	0.0	64089.0	8804.0	5235.0	0.5	8133.0
pasture	0.0	4088.0	309.0	286.0	0.0	253.0
pine	0.0	25302.0	993.0	395.0	99.0	1306.0

Now that we have determined the expected Mahalanobis Distances, given the actual class of each pixel, we can proceed with our empirical investigation.

Denote the forest (squared) Mahalanobis Distance of a pixel by M_f, the sea (squared) Mahalanobis Distance of a pixel by M_s, the pasture (squared) Mahalanobis Distance of a pixel by M_p, and the pine (squared) Mahalanobis Distance of a pixel by M_π. The general form of the equation for determining the fuzzy class membership of a particular pixel is given by

$$y = \frac{\left(\frac{1}{M_f}\right)^x}{\left(\frac{1}{M_f}\right)^x + \left(\frac{1}{M_s}\right)^x + \left(\frac{1}{M_p}\right)^x + \left(\frac{1}{M_\pi}\right)^x} \tag{7.2}$$

where y represents the fuzzy class membership, and x represents the fuzzy exponent m. This temporary change in notation is motivated by a temporary view of the fuzzy exponent m as a variable, rather than a fixed, but as yet undetermined value.

7.3.2 Pure Pixels

Consider the specific case where the pixel is actually a member of class forest. Clearly, we would like such a pixel to exhibit a high degree of membership (say, > 0.9) in the class forest, and a low degree of membership (say, < 0.1) in all other classes. Because of the unity sum condition imposed in our method of calculating the fuzzy class memberships, the latter requirement follows from the former.

Selecting the appropriate numbers from Table 7.1 above, we have the membership of the pixel in the class forest given by

$$y = \frac{\left(\frac{1}{6.2}\right)^x}{\left(\frac{1}{6.2}\right)^x + \left(\frac{1}{5769}\right)^x + \left(\frac{1}{304.2}\right)^x + \left(\frac{1}{403.6}\right)^x}$$

(where y, x are the fuzzy membership and exponent respectively, as before)

The fuzzy class membership is plotted as a function of the exponent x in Figure 7.2.

The desired membership in the class forest of > 0.9 occurs at (and above) $x = 0.8$.

It is interesting to calculate the memberships of this hypothetical pixel, actually of class forest, in the classes sea, pasture and pine. We know from the manner in which the fuzzy memberships are calculated that the sum of the other memberships will be one minus the value of, in this case, forest. Therefore, at a value of the fuzzy exponent at or above 0.8, we know that no membership other than forest can exceed 0.1 (because the membership of forest is ≥ 0.9. It is, however, instructive to observe what happens to the memberships as the fuzzy exponent increases.

The membership in the class sea is given by

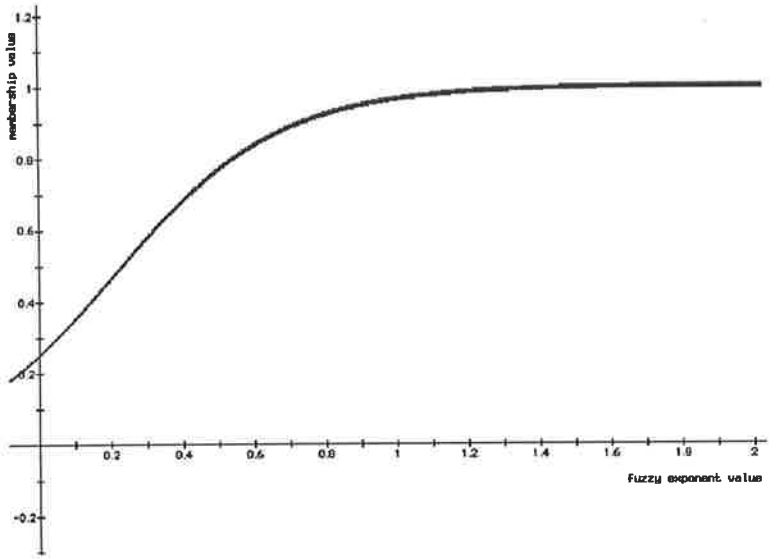

Figure 7.2: Fuzzy Membership of a Pure Pixel of Class Forest, in the Class Forest, as a Function of the Fuzzy Exponent Value

$$y = \frac{\left(\frac{1}{5769}\right)^x}{\left(\frac{1}{6.2}\right)^x + \left(\frac{1}{5769}\right)^x + \left(\frac{1}{304.2}\right)^x + \left(\frac{1}{403.6}\right)^x}$$

and shown in Figure 7.3.

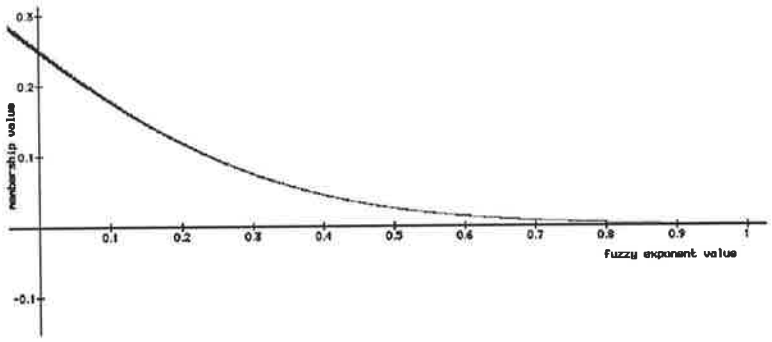

Figure 7.3: Fuzzy Membership of a Pure Pixel of Class Forest, in the Class Sea, as a Function of the Fuzzy Exponent Value

The membership in the class pasture is given by

$$y = \frac{\left(\frac{1}{304.2}\right)^x}{\left(\frac{1}{6.2}\right)^x + \left(\frac{1}{5769}\right)^x + \left(\frac{1}{304.2}\right)^x + \left(\frac{1}{403.6}\right)^x}$$

and shown in Figure 7.4.

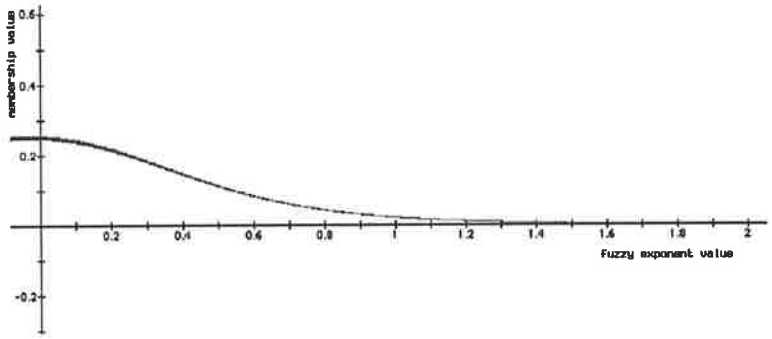

Figure 7.4: Fuzzy Membership of a Pure Pixel of Class Forest, in the Class Pasture, as a Function of the Fuzzy Exponent Value

The membership in the class pine is highly similar to that of pasture (due to the similarity in the Mahalanobis Distance values), and is therefore not plotted.

Consider now the specific case of where a pixel is actually a member of the class sea. Again, we use Mahalanobis Distances from Table 7.1.

The membership of the pixel in the class sea is given by

$$y = \frac{\left(\frac{1}{5.1}\right)^x}{\left(\frac{1}{107.2}\right)^x + \left(\frac{1}{5.1}\right)^x + \left(\frac{1}{937.3}\right)^x + \left(\frac{1}{417.3}\right)^x}$$

The fuzzy class membership is plotted as a function of the exponent x in Figure 7.5.

The membership in the class forest is given by

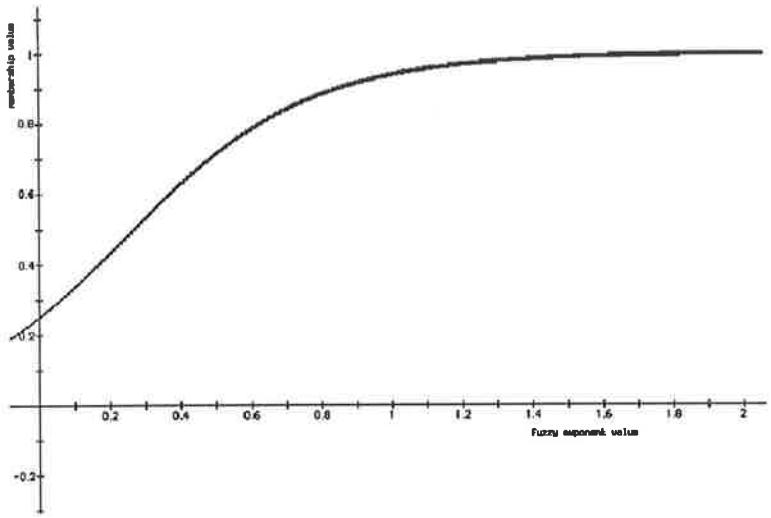

Figure 7.5: Fuzzy Membership of a Pure Pixel of Class Sea, in the Class Sea, as a Function of the Fuzzy Exponent Value

$$y = \frac{\left(\frac{1}{107.2}\right)^x}{\left(\frac{1}{107.2}\right)^x + \left(\frac{1}{5.1}\right)^x + \left(\frac{1}{937.3}\right)^x + \left(\frac{1}{417.3}\right)^x}$$

It is shown in Figure 7.6.

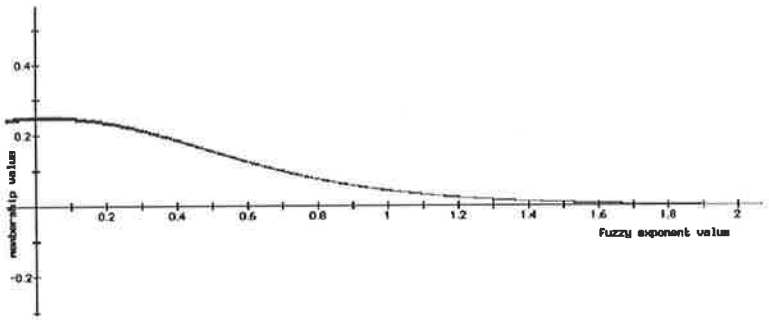

Figure 7.6: Fuzzy Membership of a Pure Pixel of Class Sea, in the Class Forest, as a Function of the Fuzzy Exponent Value

In this case, we have achieved the desirable results at a value of x at or above 0.9.

The other classes are not instructive, and are omitted.

We now address the case of the membership of a hypothetical pixel of class pasture.

The membership of the pixel in the class pasture is given by

$$y = \frac{\left(\frac{1}{16}\right)^x}{\left(\frac{1}{100.9}\right)^x + \left(\frac{1}{24094}\right)^x + \left(\frac{1}{16}\right)^x + \left(\frac{1}{3406}\right)^x}$$

The fuzzy class membership is plotted as a function of the exponent x in Figure 7.7, and indicates a required value of x at or above 1.3.

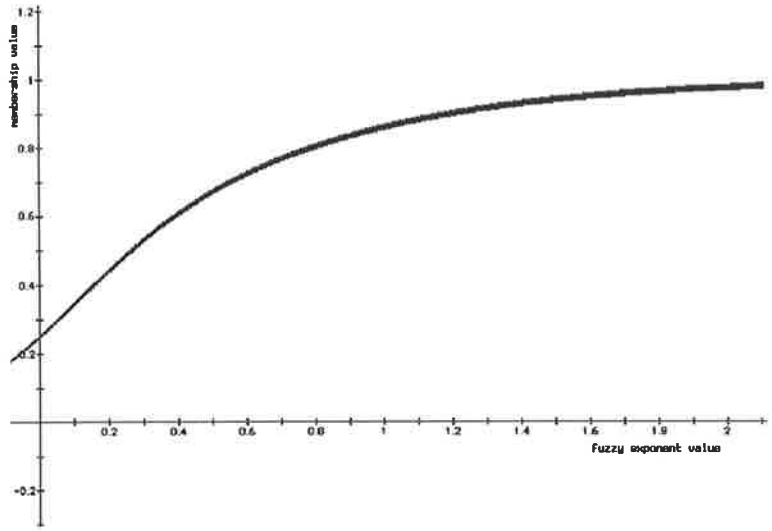

Figure 7.7: Fuzzy Membership of a Pure Pixel of Class Pasture, in the Class Pasture, as a Function of the Fuzzy Exponent Value

The membership of forest, as a function of the weighting exponent, is given by

$$y = \frac{\left(\frac{1}{100.9}\right)^x}{\left(\frac{1}{100.9}\right)^x + \left(\frac{1}{24094}\right)^x + \left(\frac{1}{16}\right)^x + \left(\frac{1}{3406}\right)^x}$$

It is shown in Figure 7.8.

This indicates a required x value of at least 1.2.

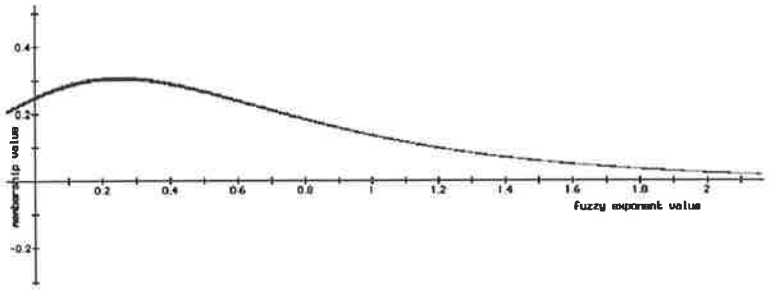

Figure 7.8: Fuzzy Membership of a Pure Pixel of Class Pasture, in the Class Forest, as a Function of the Fuzzy Exponent Value

The membership functions for both sea (Figure 7.9) and pine (Figure 7.10) quickly approach the desired value of zero.

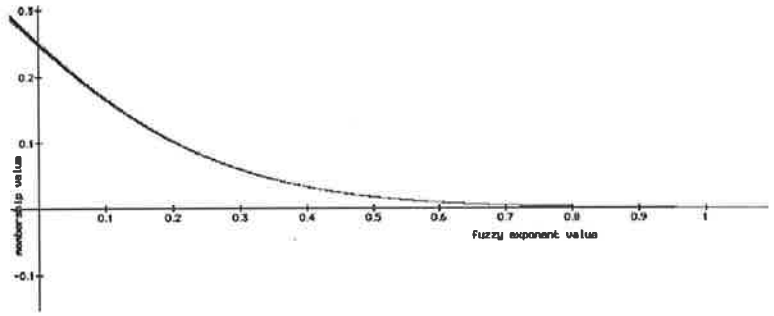

Figure 7.9: Fuzzy Membership of a Pure Pixel of Class Pasture, in the Class Sea, as a Function of the Fuzzy Exponent Value

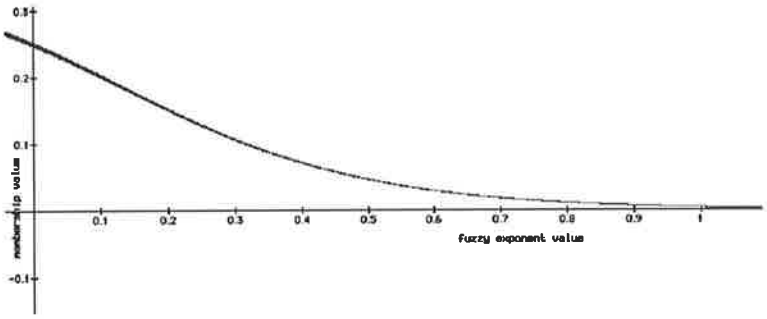

Figure 7.10: Fuzzy Membership of a Pure Pixel of Class Pasture, in the Class Pine, as a Function of the Fuzzy Exponent Value

Finally, we consider the case where the pixel is actually a member of the class pine.

Its membership in the class pine is determined by

$$y = \frac{\left(\frac{1}{3.8}\right)^x}{\left(\frac{1}{14.6}\right)^x + \left(\frac{1}{5802}\right)^x + \left(\frac{1}{476.8}\right)^x + \left(\frac{1}{3.8}\right)^x}$$

The fuzzy class membership is plotted as a function of the exponent x in Figure 7.11.

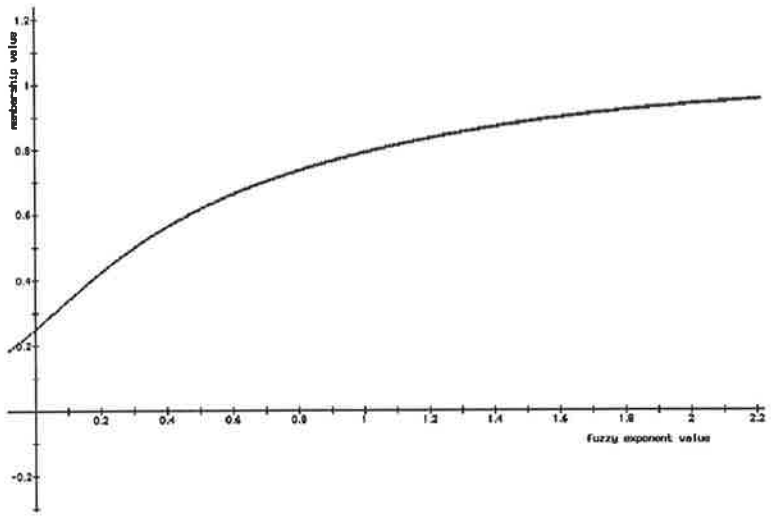

Figure 7.11: Fuzzy Membership of a Pure Pixel of Class Pine, in the Class Pine, as a Function of the Fuzzy Exponent Value

This graph indicates a requirement for an x value of 1.7 or greater.

The membership of forest, as a function of the weighting coefficient, is given by

$$y = \frac{\left(\frac{1}{14.6}\right)^x}{\left(\frac{1}{14.6}\right)^x + \left(\frac{1}{5802}\right)^x + \left(\frac{1}{476.8}\right)^x + \left(\frac{1}{3.8}\right)^x}$$

It is shown in Figure 7.12.

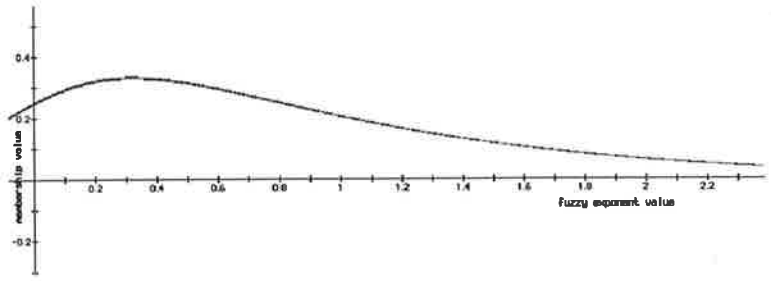

Figure 7.12: Fuzzy Membership of a Pure Pixel of Class Pine, in the Class Forest, as a Function of the Fuzzy Exponent Value

This indicates a x value of at least 1.6. The other classes are not instructive, and are omitted.

The empirical results indicate that the value of x (i.e. the fuzzy exponent) should be > 0.8 (indicated by forest), > 0.9 (sea), > 1.3 (pasture) and > 1.7 (pine). These are all satisfied if the exponent x is greater than or equal to 1.7. Recalling that this is the minimum value required to achieve the desired result, $m = 2$ is shown to be a good value to achieve the desired memberships in the case of pure pixels.

7.3.3 Mixed Pixels

Clearly, we cannot limit our investigation of required fuzzy exponent values to the case of pure pixels: if pixels were only going to display spectral behaviour that was highly typical of only one class, it might reasonably be viewed that fuzzy representations were not required.

Consider the case of a pixel comprising a mixture of two classes. Using a linear mixing model (see e.g. van der Meer, 1995), a synthetic pixel of arbitrary proportions can be constructed. This situation is illustrated in Figure 7.13, where the class distributions are represented by ellipses. The classes have means (x_1, y_1) and (x_2, y_2). Denote the band values of the synthetic pixel by (x_p, y_p).

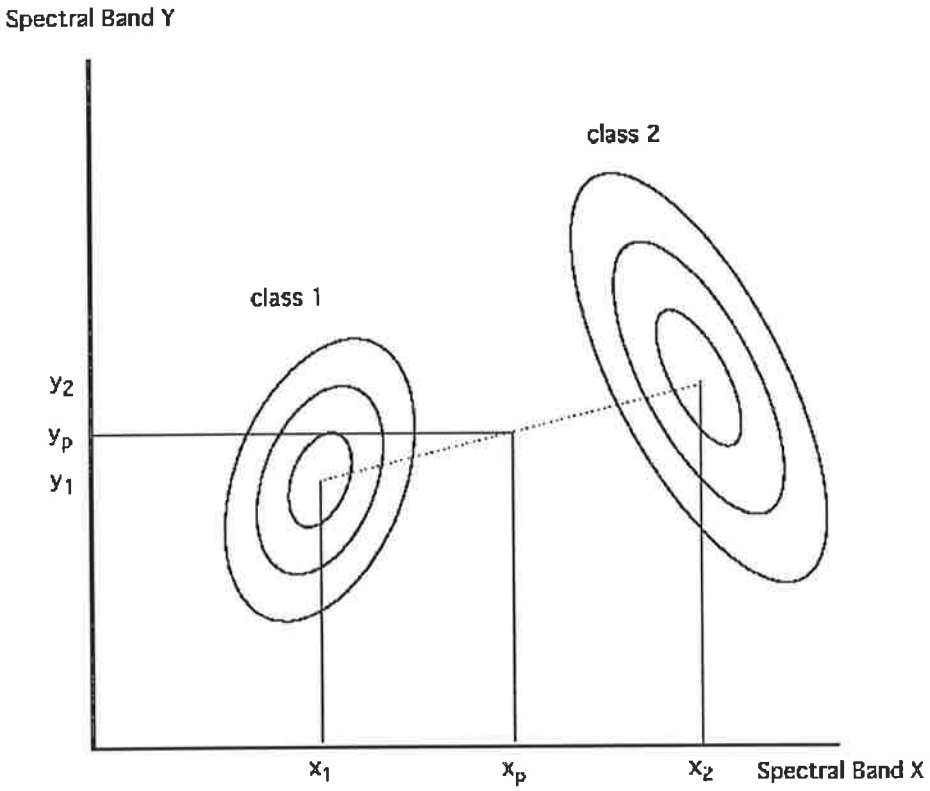

Figure 7.13: A Simple Two-Class Mixed Pixel Model

Using this model, and the class means and class covariances of the training data from the TM93 image, the Mahalanobis Distances from all the class means of a synthetic mixed pixel can be calculated for the actual images held. We can then pursue a graphical analysis similar to that undertaken for pure pixels, but now requiring that the memberships reflect the mixel proportions.

Table 7.4 shows the expected Mahalanobis Distances from all the class means of a synthetic pixel comprising equal two class mixtures of forest/pasture, forest/pine and pasture/pine. Other mixture classes were not investigated, as they hold little interest.

Table 7.4 indicates that, for a mixed pixel comprising equal proportions of

Table 7.4: Expected Mahalanobis Distances (MahDist) for a Two Class Mixed Synthetic Pixel (Equal Proportions) Based on TM93 Image Data

	Mixed pixel Comprising Equal Proportions of		
Mahdist Class	forest/pasture	forest/pine	pasture/pine
forest	26.9	3.2	30.6
sea	12435.1	4826.3	11190.1
pasture	83.9	398.6	119.8
pine	1380.7	49.3	935.9

forest and pasture, we would expect a forest Mahalanobis Distance of 26.9 (i.e. a Mahalanobis Distance computed using the mean and covariance matrix of the class forest), a sea Mahalanobis Distance of 12,435.1, a pasture Mahalanobis Distance of 83.9, and a pine Mahalanobis Distance of 1,380.7. Similarly, if the actual (surface cover) class imaged in the pixel was an equal mixture of forest and pine, we would expect a forest Mahalanobis Distance of 3.2, and so on.

Using these data, the fuzzy set membership in the class forest of a pixel comprising 50% forest and 50% pasture is given by

$$y = \frac{\left(\frac{1}{26.9}\right)^x}{\left(\frac{1}{26.9}\right)^x + \left(\frac{1}{12435.1}\right)^x + \left(\frac{1}{83.9}\right)^x + \left(\frac{1}{1380.7}\right)^x}$$

(where y, x are the fuzzy membership and exponent respectively, as before)

The fuzzy class membership is plotted as a function of the exponent x in Figure 7.14.

The desired membership in the class forest is 0.5 (reflecting its true proportion in the class forest). This is achieved at a value of x of about 0.4. We can see from this graph that if we selected an x value of 2 as suggested by our analysis of the requirement for pure pixels, this mixed pixel would have a forest membership of

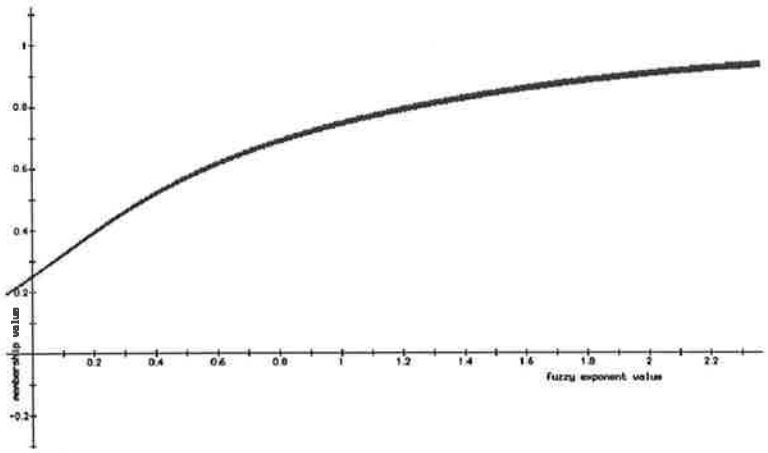

Figure 7.14: Fuzzy Membership of a Synthetic Mixed Pixel (Equal Proportions of Forest and Pasture) in the Class Forest as a Function of the Fuzzy Exponent Value

approximately 0.9.

The membership of the pixel in the class pasture is given by

$$y = \frac{\left(\frac{1}{83.9}\right)^x}{\left(\frac{1}{26.9}\right)^x + \left(\frac{1}{12435.1}\right)^x + \left(\frac{1}{83.9}\right)^x + \left(\frac{1}{1380.7}\right)^x}$$

This is depicted graphically in Figure 7.15.

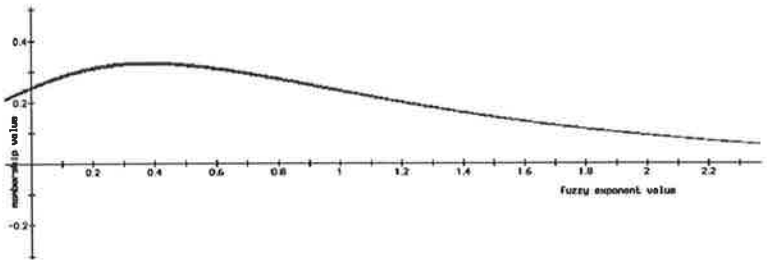

Figure 7.15: Fuzzy Membership of a Synthetic Mixed Pixel (Equal Proportions of Forest and Pasture) in the Class Pasture as a Function of the Fuzzy Exponent Value

This function does not achieve the desired value of 0.5: it achieves its maximum value (approximately 0.3) in the range $0.3 < x < 0.5$. The membership of the classes sea and pine decay to insignificant levels by $x = 0.8$. At $x = 0.4$, they are 0.05 and

0.1 respectively.

Considering now a mixed pixel of equal proportions of forest and pine, we have its membership of the class forest given by

$$y = \frac{\left(\frac{1}{3.2}\right)^x}{\left(\frac{1}{3.2}\right)^x + \left(\frac{1}{4826.3}\right)^x + \left(\frac{1}{398.6}\right)^x + \left(\frac{1}{49.3}\right)^x}$$

This is shown in Figure 7.16.

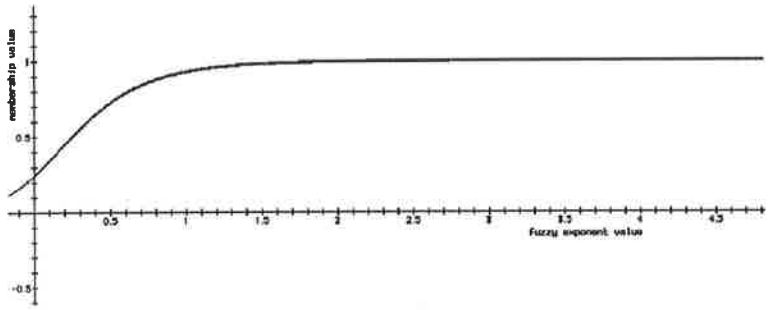

Figure 7.16: Fuzzy Membership of a Synthetic Mixed Pixel (Equal Proportions of Forest and Pine) in the Class Forest as a Function of the Fuzzy Exponent Value

The desired value is achieved at $m = 0.25$.

Its membership in the class pine is given by

$$y = \frac{\left(\frac{1}{49.3}\right)^x}{\left(\frac{1}{3.2}\right)^x + \left(\frac{1}{4826.3}\right)^x + \left(\frac{1}{398.6}\right)^x + \left(\frac{1}{49.3}\right)^x}$$

This is shown in Figure 7.17.

This achieves its maximum value (approximately 0.25) at about $x = 0.15$. Again, the membership of the other classes decay to insignificant levels by $x = 0.8$.

The case of the pixel comprising mixed pasture and pine is quite disturbing: its fuzzy set memberships reveal a high value (monotonically increasing with the

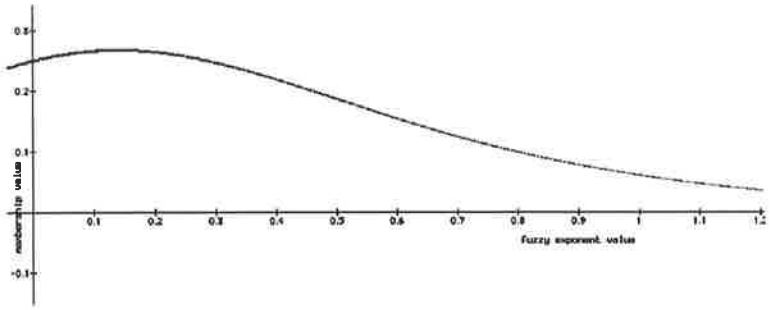

Figure 7.17: Fuzzy Membership of a Synthetic Mixed Pixel (Equal Proportions of Forest and Pine) in the Class Pine as a Function of the Fuzzy Exponent Value

exponent x, and approaching 1 by $x = 2.5$) in the fuzzy set forest, even though forest is not one of the constituent classes! This is because a mixture of pasture and pine can be 'closer' (as measured by its Mahalanobis Distance) to the class forest than to either pasture or pine. This is explained by the observation that the expected band brightness values for forest lie between those of pasture and pine in a significant number of bands (three out of four). This shows that a fuzzy classification approach is not quarantined from the problem observed in Campbell (1987) and Foody (1996), that a conventional classification will force the allocation of a mixed pixel to a single class, with this class not necessarily even being present in the pixel.

This suggests that, before using a fuzzy set membership function of this type, the data should be examined to determine if similar conditions to these exist. That is, does any class 'lie between' any two other classes, in the spectral space, and are pixels likely to exist, in reasonable number, as mixtures of these two classes? If so, it must be understood that, with fuzzy set membership functions of this form, any pixel assessed as a pure pixel of the middle class could be a mixture of the two (or more) outer classes, and vice versa. Of course, it is noted that the traditional crisp classifiers will also perform poorly in a case such as this. None-the-less, we have identified a

clear potential for classification error. It is possible that the introduction of a term taking account of the relative likelihoods of the pure and mixed pixels might mitigate this error somewhat, but this matter is not pursued in this thesis.

7.3.4 Summary

An empirical investigation has revealed conflicting requirements for the selection of the fuzzy exponent m (represented above by x) in a fuzzy classifier based on the ratios of the reciprocals of the class Mahalanobis Distances. For pure pixels we want the value of m to be about 2 (or higher), giving a high membership in the pure class that the pixel is actually a member of. For mixed pixels, we want it to be about 0.5 (or lower), thereby distributing the memberships between the two classes that actually comprise the (synthetic) pixel. The work shown in Chapters 5 and 6 of this thesis has been undertaken with an exponent value of 1.25, midway between these two values.

It has also illustrated that errors may be expected where the band values of the mean of one class can be expressed as a linear combination of the band mean values of two or more other classes.

7.4 Analytic Investigation of the Value of m

Of course, mixed pixels need not necessarily comprise equal proportions, and need not necessarily comprise only two classes. It would obviously be extremely tedious to undertake a graphical analysis of a wide range of possible cases. It is, however, possible to investigate the two class mixed pixel case analytically (and allowing for arbitrary proportions).

Consider a simple situation in which there are two classes and two image bands. Let the mean of class 1 be $\tilde{\mu}_1 = (x_1, y_1)^t$ and the mean of class 2 be $\tilde{\mu}_2 = (x_2, y_2)^t$.

Consider a mixed pixel in areal proportion $\lambda_1 : \lambda_2$. Without loss of generality, let us require that

$$\lambda_1 + \lambda_2 = 1 \tag{7.3}$$

Under the linear mixing assumption, we would expect such a mixed pixel to exhibit a spectral vector $\tilde{x}_p = (x_p, y_p)^t$ given by

$$\tilde{x}_p = \left((\lambda_1 x_1 + \lambda_2 x_2), (\lambda_1 y_1 + \lambda_1 y_2)\right)^t \tag{7.4}$$

Denote the class 1 (squared) Mahalanobis Distance of $\tilde{x}_p$ by $(D_{1p})^2$.

Now,

$$\begin{aligned}
(D_{1p})^2 &= (\tilde{x}_p - \tilde{\mu}_1)^t \Sigma_1^{-1} (\tilde{x}_p - \tilde{\mu}_1) \\
&= ((\lambda_1 x_1 + \lambda_2 x_2) - x_1, (\lambda_1 y_1 + \lambda_2 y_2) - y_1)^t \\
&\quad \Sigma_1^{-1} ((\lambda_1 x_1 + \lambda_2 x_2) - x_1, (\lambda_1 y_1 + \lambda_2 y_2) - y_1) \\
&= (\lambda_2 x_2 - (1 - \lambda_1) x_1, \lambda_2 y_2 - (1 - \lambda_1) y_1)^t \\
&\quad \Sigma_1^{-1} (\lambda_2 x_2 - (1 - \lambda_1) x_1, \lambda_2 y_2 - (1 - \lambda_1) y_1)
\end{aligned}$$

but $(1 - \lambda_1) = \lambda_2$ therefore

$$\begin{aligned}
(D_{1p})^2 &= (\lambda_2 x_2 - \lambda_2 x_1, \lambda_2 y_2 - \lambda_2 y_1)^t \Sigma_1^{-1} (\lambda_2 x_2 - \lambda_2 x_1, \lambda_2 y_2 - \lambda_2 y_1) \\
&= \lambda_2^2 ((x_2 - x_1), (y_2 - y_1))^t \Sigma_1^{-1} ((x_2 - x_1), (y_2 - y_1)) \\
&= \lambda_2^2 (D_{12})^2 \left(= (1 - \lambda_1)^2 (D_{12})^2\right)
\end{aligned} \tag{7.5}$$

Similarly,

$$(D_{2p})^2 = \lambda_1^2 (D_{21})^2 \left(= (1 - \lambda_2)^2 (D_{21})^2\right) \tag{7.6}$$

where $(D_{12})^2$ is the class 1 (squared) Mahalanobis Distance of $\tilde{\mu}_2$ (the mean of class 2), and $(D_{21})^2$ is the class 2 (squared) Mahalanobis Distance of $\tilde{\mu}_1$ (the mean of class 1).

This development permits us to derive analytically a value for the exponent m in the Mahalanobis Distance fuzzy classifier.

In the simple case where there are only two classes, we denote the fuzzy set membership of the pixel p in the class j

$$\mu_{jp} = \frac{\left(\frac{1}{D_{jp}^2}\right)^m}{\left(\frac{1}{D_{1p}^2}\right)^m + \left(\frac{1}{D_{2p}^2}\right)^m} \quad j = 1, 2. \tag{7.7}$$

A desire to interpret fuzzy set memberships as proportions of mixed pixels leads to the requirement that

$$\mu_1 : \mu_2 = \lambda_1 : \lambda_2 \tag{7.8}$$

That is, that the ratio of memberships should equal the ratio of proportions.

Therefore, we require that

$$\frac{\mu_{1p}}{\mu_{2p}} = \frac{\left(\frac{1}{D_{1p}^2}\right)^m}{\left(\frac{1}{D_{2p}^2}\right)^m} = \frac{\lambda_1}{\lambda_2} \tag{7.9}$$

$$\frac{\left(\frac{1}{\lambda_2^2 (D_{12})^2}\right)^m}{\left(\frac{1}{\lambda_1^2 (D_{21})^2}\right)^m} = \frac{\lambda_1}{\lambda_2}$$

$$\left(\frac{\lambda_1}{\lambda_2}\right)^{2m} \left(\frac{D_{21}}{D_{12}}\right)^{2m} = \frac{\lambda_1}{\lambda_2}$$

$$\left(\frac{D_{12}}{D_{21}}\right)^{2m} - \left(\frac{\lambda_1}{\lambda_2}\right)^{2m-1} = 0 \tag{7.10}$$

For convenience, denote the ratios $\left(\frac{D_{12}}{D_{21}}\right)$ and $\left(\frac{\lambda_1}{\lambda_2}\right)$ as K_1 and K_2 respectively.

If $K_1 = 1$ (i.e. when $D_{21} = D_{12}$), this equation is satisfied with $m = 0.5$ (irrespective of the value of K_2). It is noted that this case holds for hyperspherical distributions.

If $K_1 = K_2 = 1$ (i.e. when $D_{21} = D_{12}$, and with equal contributing proportions of each class), this equation is satisfied for all values of m. Moreover, if $K_2 = 1$ either:

1. $K_1 = 1$ (in which case, any value of m will suffice);
2. $m = 0$ (which is uninteresting).

More generally, we require

$$(K_1)^{2m} = (K_2)^{2m-1}$$

$$2m \ln K_1 = (2m-1) \ln K_2$$

$$2m \ln K_1 = 2m \ln K_2 - \ln K_2$$

$$\ln K_2 = 2m(\ln K_2 - \ln K_1)$$

$$m = \frac{\ln K_2}{2(\ln K_2 - \ln K_1)} \tag{7.11}$$

We note that:

1. this derivation is not valid for $K_1 = K_2$ (we would have performed a 'divide by zero' operation); and
2. if $K_1 = 1$, then $\ln K_1 = 0$, and $m = 0.5$, irrespective of the value of K_2 (as before).

More generally, the ideal value for the fuzzy exponent m in the Mahalanobis Distance Fuzzy Classifier (under the requirement that the fuzzy memberships equal proportions in the mixed pixel case), is a function of both the ratio of the Mahalanobis Distances, and the ratio of proportions.

We cannot, of course, know in advance what the proportions of mixtures are (this is what we hope to estimate via the fuzzy class memberships). Thus we cannot determine an optimal m (optimal in the sense of returning fuzzy class memberships in proportions of mixtures).

We have, however, derived a useful formula, which we can examine further.

If $K_2 < 1$, then $\ln K_2 < 0$.

Recalling that we require $m \geq 0$, then $(\ln K_2 - \ln K_1)$ must be < 0. Therefore $\ln K_1 > \ln K_2$. Therefore $K_1 > K_2$. Therefore $K_1 < K_2 < 1$ implies that no solution is possible.

If $K_2 > 1$, then $\ln K_2 > 0$.

For $m \geq 0$, then $(\ln K_2 - \ln K_1)$ must be > 0. Therefore $\ln K_2 > \ln K_1$. Therefore $K_2 > K_1$. Therefore $K_1 > K_2 > 1$ also implies that no solution is possible.

All other cases lead to analytic solutions.

As noted before, the cases for equal Mahalanobis Distances (i.e. $K_1 = 1$), and $K_1 = K_2 = 1$, are trivial. The case $K_1 = K_2$ is generally precluded. There are six remaining, more general cases:

$K_1 > 1$, then

$$K_2 > K_1 > 1 \tag{7.12}$$

$$K_1 > K_2 \geq 1 \tag{7.13}$$

$$K_1 > 1 \geq K_2 \tag{7.14}$$

$K_1 < 1$, then

$$K_2 < K_1 < 1 \tag{7.15}$$

$$K_1 < K_2 \leq 1 \tag{7.16}$$

$$K_1 < 1 \leq K_2 \tag{7.17}$$

To investigate these empirically, we require some typical values for K_1 (for K_2, the ratio of class proportions, it suffices to simply select the values 0.1 to 0.9, in increments of 0.1). We can derive typical values for K_1 from the TM93 training data.

Denote the class i (squared) Mahalanobis Distance to the mean of class j as D^2_{ij}, the forest Mahalanobis Distance of the mean of class sea as D_{fs}, ... , the pine Mahalanobis Distance of the mean of class forest as $D_{\pi f}$, ... , and the pine Mahalanobis Distance of the mean of class pasture as $D_{\pi p}$. Values for ratios of Mahalanobis Distances (i.e. K_1) from the TM93 training data are shown in Table 7.5.

Table 7.5: Empirical Values for the Ratios of Class Mahalanobis Distances to the Mean of Each Class

	D^2_{ij}/D^2_{ji}	D^2_{ji}/D^2_{ij}	D_{ij}/D_{ji}	D_{ji}/D_{ij}
D_{fs}/D_{sf}	0.02	44.83	0.15	6.70
D_{fp}/D_{pf}	0.32	3.12	0.57	1.77
$D_{f\pi}/D_{\pi f}$	0.06	15.53	0.25	3.94
D_{sp}/D_{fs}	27.49	0.04	5.24	0.19
$D_{s\pi}/D_{\pi s}$	13.79	0.07	3.71	0.27
$D_{p\pi}/D_{\pi p}$	0.13	7.81	0.36	2.79

Table 7.5 indicates that the forest Mahalanobis Distance to the mean of class sea divided by the sea Mahalanobis Distance the mean of class forest in 0.15, and so on.

Table 7.6 shows the ratios of mean Mahalanobis Distances (whereas the previous table showed Mahalanobis Distances to the means) for comparison.

It is noted that Tables 7.5 and 7.6 are quite similar. They reveal that ratios of Mahalanobis Distances can typically be found in the range 0.15 to 7. Using values

Table 7.6: Empirical Values for the Ratios of Mean Class Mahalanobis Distances

	D_{ij}^2/D_{ji}^2	D_{ji}^2/D_{ij}^2	D_{ij}/D_{ji}	D_{ji}/D_{ij}
D_{fs}/D_{sf}	0.02	53.82	0.14	7.34
D_{fp}/D_{pf}	0.33	3.01	0.58	1.74
$D_{f\pi}/D_{\pi f}$	0.04	27.64	0.19	5.26
D_{sp}/D_{fs}	25.71	0.04	5.07	0.20
$D_{s\pi}/D_{\pi s}$	13.90	0.07	3.73	0.27
$D_{p\pi}/D_{\pi p}$	0.14	7.14	0.37	2.67

in this range, we can investigate the various possibilities given by Equations 7.12 to 7.17.

Recall that the requirement is actually to solve (7.10). Consider the function:

$$y = (K_1)^{2x} - (K_2)^{2x-1} \tag{7.18}$$

Solving for $y = 0$, with values for K_1 and K_2 selected appropriately, will enable us to find the correct value for the fuzzy exponent value (represented by x). It will also permit us to examine the sensitivity of our solution.

Case (7.12) ($K_2 > K_1 > 1$)

$$y = \left(\frac{3}{2}\right)^{2x} - 3^{2x-1}$$

This is shown graphically in Figure 7.18, and has a (positive) root at 0.8.

Case (7.13) ($K_1 > K_2 \geq 1$) is plotted in Figure 7.19.

$$y = 2^{2x} - \left(\frac{4}{3}\right)^{2x-1}$$

This has no positive root (it is one of the cases identified earlier as having no solution).

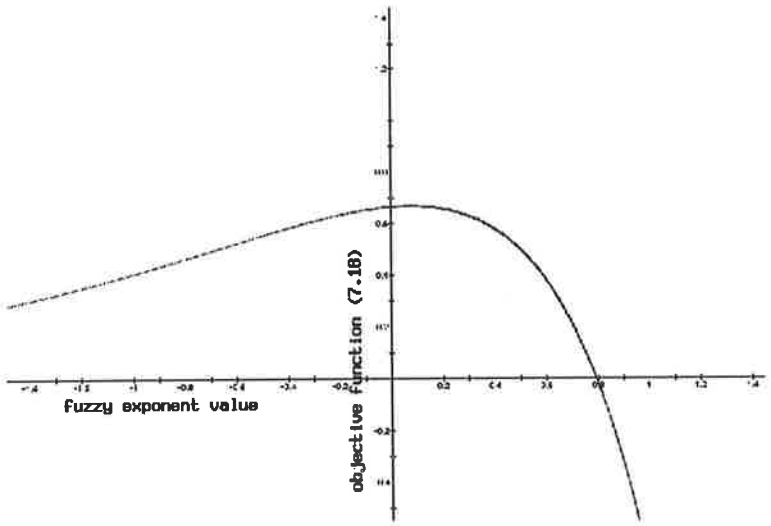

Figure 7.18: Solving for Fuzzy Exponent m when $K_2 > K_1 > 1$

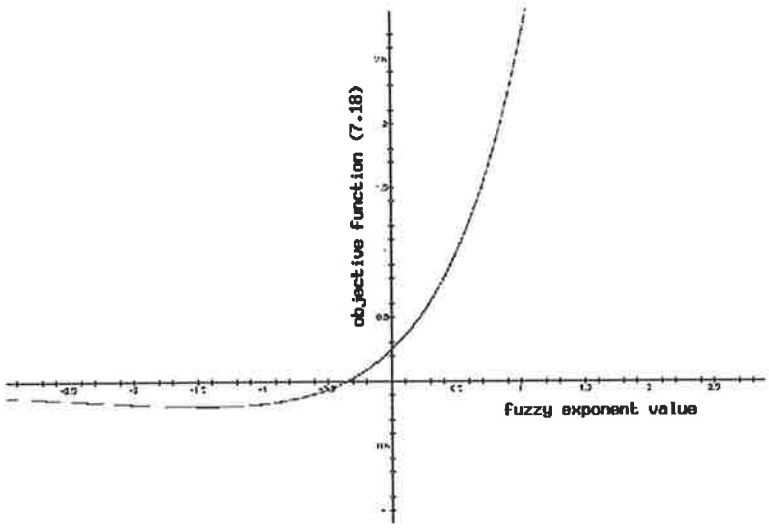

Figure 7.19: Solving for Fuzzy Exponent m when $K_1 > K_2 \geq 1$

Case (7.14) ($K_1 > 1 \geq K_2$) is plotted in Figure 7.20, and has a root at about 0.4.

$$y = \left(\frac{5}{4}\right)^{2x} - \left(\frac{1}{4}\right)^{2x-1}$$

Case (7.15) ($K_2 < K_1 < 1$) is plotted in Figure 7.21, and has a root at 0.6.

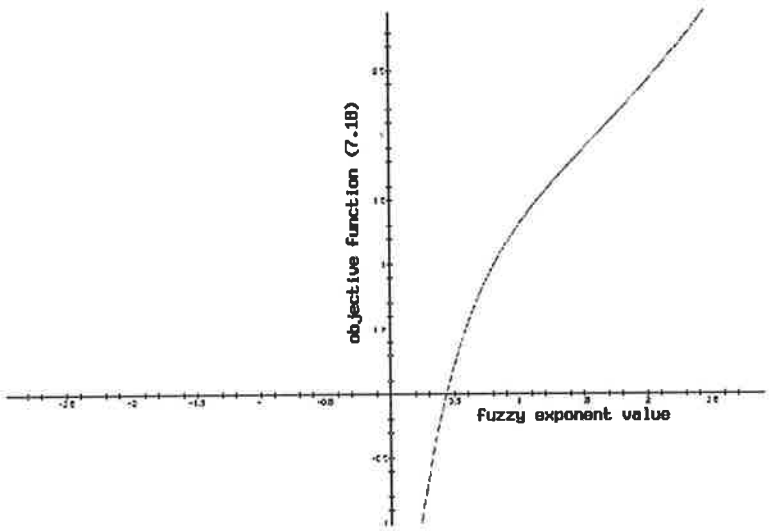

Figure 7.20: Solving for Fuzzy Exponent m when $K_1 > 1 \geq K_2$

$$y = (0.8)^{2x} - (0.3)^{2x-1}$$

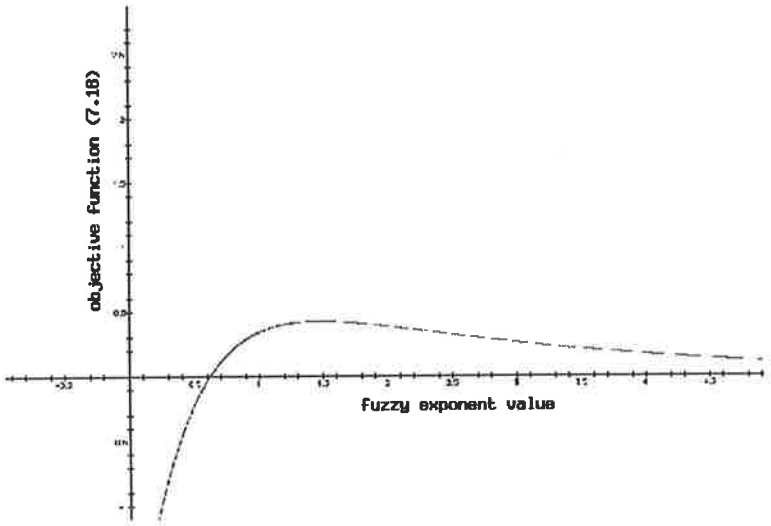

Figure 7.21: Solving for Fuzzy Exponent m when $K_2 < K_1 < 1$

Case (7.16) ($K_1 < K_2 \leq 1$) is plotted in Figure 7.22.

$$y = (0.25)^{2x} - (0.5)^{2x-1}$$

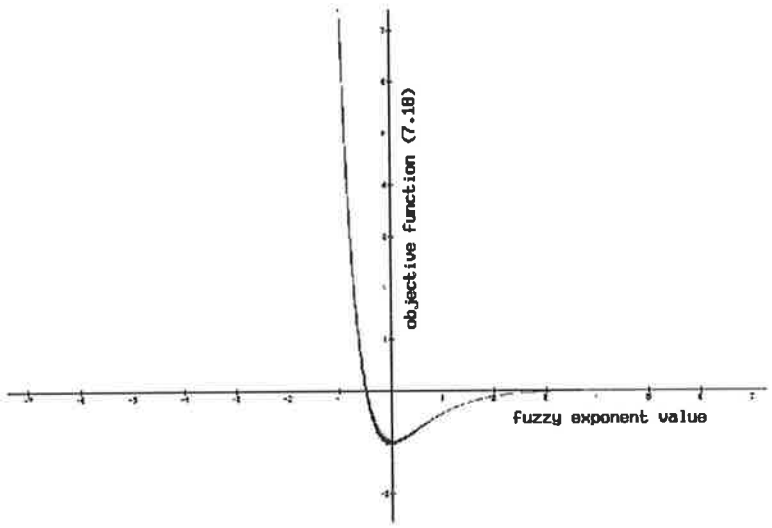

Figure 7.22: Solving for Fuzzy Exponent m when $K_1 < K_2 \leq 1$

This has no (positive finite) root (but is satisfactory for practical purposes at about $x = 3$).

Case (7.17) ($K_1 < 1 \leq K_2$) is plotted in Figure 7.23, showing a root at $x = 0.4$.

$$y = (0.7)^{2x} - 4^{2x-1}$$

Therefore we can say, based upon this limited empirical investigation, that where the equation for the analytic value for the fuzzy exponent (for the two class mixed pixel case) has positive (and finite) roots, they are found in the range 0.25 to 1.5 (over a range of 'reasonable' values for ratios of distances and ratios of proportions); figures which accord reasonably well with our earlier empirical investigations. In several of the cases plotted, the slope of the graph around the root ($y = 0$) suggests, however, a high degree of sensitivity to the value of the fuzzy exponent.

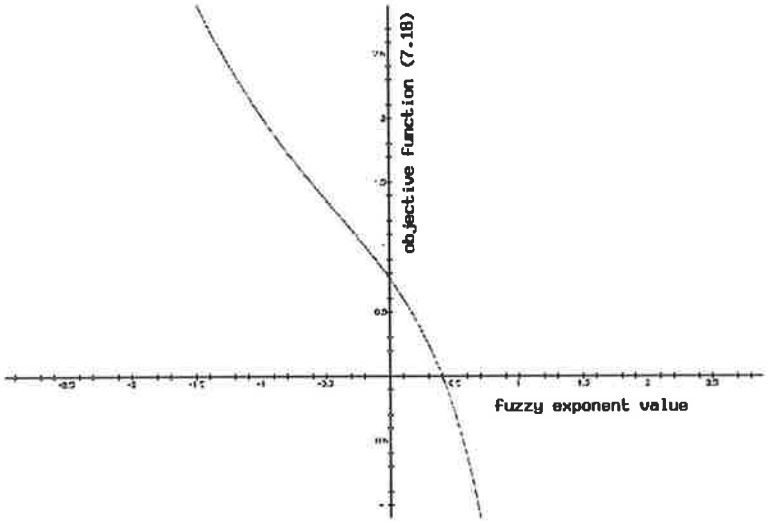

Figure 7.23: Solving for Fuzzy Exponent m when $K_1 < 1 \leq K_2$

7.5 Sensitivity of the Fuzzy Classifier to Selection of m

Tables 7.7 to 7.12 examine the sensitivity of the computed fuzzy class memberships, for various proportions in the two class mixed pixel case, for various values of the fuzzy exponent m. The input data for Mahalanobis Distances and class proportions are synthetic, although guided by the Mahalanobis Distance values found in our image data, and by the class proportions that we wish to be able to discriminate. The table columns indicate the proportion of class 1 (λ_1), from 0.1 to 0.9 in increments of 0.1. The values for K_2 are computed by $\lambda_1/(1-\lambda_1)$. The table rows indicate the ratios of class Mahalanobis Distances (K_1). The values selected are 0.2 to 1.0, in increments of 0.2, and their reciprocals. The values in the tables show the difference between the computed fuzzy memberships in class 1 and the desired value (which is identical to the proportion of class 1 in the mixel, λ_1). There is a separate table for various values of m ranging from 0.25 to 4.0. Data are also included to show the row average

(i.e. the average difference between the computed fuzzy membership and its desired value, for a given ratio of class Mahalanobis Distances) (ave), the maximum absolute value in each row (max abs), column average (ave), the maximum absolute value in each column (max abs), and the table sum of squared errors (SSE), mean squared error (MSE) and root mean squared error (RMSE).

Table 7.7: Difference Between Desired Value of Fuzzy Membership in Class 1 (equals λ_1) and Computed Value, for Various Ratios of Distances (K_1) and Proportions (K_2), $m = 0.25$

λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		
K_2	0.11	0.25	0.43	0.67	1.00	1.50	2.33	4.00	9.00		
K_1										ave	max abs
0.20	0.33	0.33	0.29	0.25	0.19	0.13	0.07	0.02	-0.03	0.18	0.33
0.40	0.25	0.24	0.21	0.16	0.11	0.06	0.01	-0.04	-0.07	0.10	0.25
0.60	0.20	0.19	0.16	0.11	0.06	0.01	-0.04	-0.08	-0.11	0.06	0.20
0.80	0.17	0.16	0.12	0.08	0.03	-0.02	-0.07	-0.11	-0.13	0.03	0.17
1.00	0.15	0.13	0.10	0.05	0.00	-0.05	-0.10	-0.13	-0.15	0.00	0.15
1.25	0.13	0.11	0.07	0.02	-0.03	-0.08	-0.12	-0.16	-0.17	-0.03	0.17
1.67	0.11	0.08	0.04	-0.01	-0.06	-0.11	-0.16	-0.19	-0.20	-0.06	0.20
2.50	0.07	0.04	-0.01	-0.06	-0.11	-0.16	-0.21	-0.24	-0.25	-0.10	0.25
5.00	0.03	-0.02	-0.07	-0.13	-0.19	-0.25	-0.29	-0.33	-0.33	-0.18	0.33
ave	0.16	0.14	0.10	0.05	0.00	-0.05	-0.10	-0.14	-0.16		
max abs	0.33	0.33	0.29	0.25	0.19	0.25	0.29	0.33	0.33		
	SSE	1.87		MSE	0.02		RMSE	0.15			

Some interesting observations can be drawn from these tables.

1. The tables exhibit, of course, a rotational 'anti-symmetry' (i.e. if we denote the table element in row i, column j, as a_{ij}, then $a_{ij} = -a_{ji}$).
2. There is a symmetry (or anti-symmetry) of row and column average and maximum absolute errors about $K_1 = 1$ and $K_2 = 1$ respectively.
3. SSE (MSE, RMSE) strictly increase with increasing m.

Table 7.8: Difference Between Desired Value of Fuzzy Membership in Class 1 (equals λ_1) and Computed Value, for Various Ratios of Distances (K_1) and Proportions (K_2), $m = 0.5$

λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		
K_2	0.11	0.25	0.43	0.67	1.00	1.50	2.33	4.00	9.00		
K_1										ave	max abs
0.20	0.26	0.36	0.38	0.37	0.33	0.28	0.22	0.15	0.08	0.27	0.38
0.40	0.12	0.18	0.22	0.23	0.21	0.19	0.15	0.11	0.06	0.16	0.23
0.60	0.06	0.09	0.12	0.13	0.13	0.11	0.10	0.07	0.04	0.09	0.13
0.80	0.02	0.04	0.05	0.05	0.06	0.05	0.04	0.03	0.02	0.04	0.06
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.25	-0.02	-0.03	-0.04	-0.05	-0.06	-0.05	-0.05	-0.04	-0.02	-0.04	0.06
1.67	-0.04	-0.07	-0.10	-0.11	-0.13	-0.13	-0.12	-0.09	-0.06	-0.09	0.13
2.50	-0.06	-0.11	-0.15	-0.19	-0.21	-0.23	-0.22	-0.18	-0.12	-0.16	0.23
5.00	-0.08	-0.15	-0.22	-0.28	-0.33	-0.37	-0.38	-0.36	-0.26	-0.27	0.38
ave	0.03	0.03	0.03	0.02	0.00	-0.02	-0.03	-0.03	-0.03		
max abs	0.26	0.36	0.38	0.37	0.33	0.37	0.38	0.36	0.26		
	SSE	2.22		MSE	0.03		RMSE	0.17			

Table 7.9: Difference Between Desired Value of Fuzzy Membership in Class 1 (equals λ_1) and Computed Value, for Various Ratios of Distances (K_1) and Proportions (K_2), $m = 1.0$

λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		
K_2	0.11	0.25	0.43	0.67	1.00	1.50	2.33	4.00	9.00		
K_1										ave	max abs
0.20	0.14	0.41	0.52	0.52	0.46	0.38	0.29	0.20	0.10	0.34	0.52
0.40	-0.03	0.08	0.23	0.34	0.36	0.33	0.27	0.19	0.10	0.21	0.36
0.60	-0.07	-0.05	0.04	0.15	0.24	0.26	0.24	0.18	0.10	0.12	0.26
0.80	-0.08	-0.11	-0.08	0.01	0.11	0.18	0.19	0.16	0.09	0.05	0.19
1.00	-0.09	-0.14	-0.14	-0.09	0.00	0.09	0.14	0.14	0.09	0.00	0.14
1.25	-0.09	-0.16	-0.19	-0.18	-0.11	-0.01	0.08	0.11	0.08	-0.05	0.19
1.67	-0.10	-0.18	-0.24	-0.26	-0.24	-0.15	-0.04	0.05	0.07	-0.12	0.26
2.50	-0.10	-0.19	-0.27	-0.33	-0.36	-0.34	-0.23	-0.08	0.03	-0.21	0.36
5.00	-0.10	-0.20	-0.29	-0.38	-0.46	-0.52	-0.52	-0.41	-0.14	-0.34	0.52
ave	-0.06	-0.06	-0.05	-0.03	0.00	0.03	0.05	0.06	0.06		
max abs	0.14	0.41	0.52	0.52	0.46	0.52	0.52	0.41	0.14		
	SSE	4.22		MSE	0.05		RMSE	0.23			

4. For a given K_1, the absolute value of the row average error strictly increases with m, and the row maximum absolute error generally increases with m (strictly increases for $m > 0.5$).

Table 7.10: Difference Between Desired Value of Fuzzy Membership in Class 1 (equals λ_1) and Computed Value, for Various Ratios of Distances (K_1) and Proportions (K_2), $m = 1.25$

λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		
K_2	0.11	0.25	0.43	0.67	1.00	1.50	2.33	4.00	9.00		
K_1										ave	max abs
0.20	0.09	0.44	0.57	0.55	0.48	0.39	0.30	0.20	0.10	0.35	0.57
0.40	-0.06	0.04	0.24	0.38	0.41	0.36	0.29	0.20	0.10	0.22	0.41
0.60	-0.09	-0.10	0.00	0.17	0.28	0.31	0.27	0.19	0.10	0.13	0.31
0.80	-0.09	-0.15	-0.13	-0.01	0.14	0.23	0.24	0.18	0.10	0.06	0.24
1.00	-0.10	-0.17	-0.19	-0.13	0.00	0.13	0.19	0.17	0.10	0.00	0.19
1.25	-0.10	-0.18	-0.24	-0.23	-0.14	0.01	0.13	0.15	0.09	-0.06	0.24
1.67	-0.10	-0.19	-0.27	-0.31	-0.28	-0.17	0.00	0.10	0.09	-0.13	0.31
2.50	-0.10	-0.20	-0.29	-0.36	-0.41	-0.38	-0.24	-0.04	0.06	-0.22	0.41
5.00	-0.10	-0.20	-0.30	-0.39	-0.48	-0.55	-0.57	-0.44	-0.09	-0.35	0.57
ave	-0.07	-0.08	-0.07	-0.04	0.00	0.04	0.07	0.08	0.07		
max abs	0.10	0.44	0.57	0.55	0.48	0.55	0.57	0.44	0.10		
	SSE	5.28		MSE	0.07		RMSE	0.26			

Table 7.11: Difference Between Desired Value of Fuzzy Membership in Class 1 (equals λ_1) and Computed Value, for Various Ratios of Distances (K_1) and Proportions (K_2), $m = 2.0$

λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		
K_2	0.11	0.25	0.43	0.67	1.00	1.50	2.33	4.00	9.00		
K_1										ave	max abs
0.20	-0.01	0.51	0.65	0.59	0.50	0.40	0.30	0.20	0.10	0.36	0.65
0.40	-0.09	-0.07	0.27	0.49	0.48	0.39	0.30	0.20	0.10	0.23	0.49
0.60	-0.10	-0.17	-0.09	0.20	0.39	0.38	0.30	0.20	0.10	0.13	0.39
0.80	-0.10	-0.19	-0.22	-0.07	0.21	0.33	0.29	0.20	0.10	0.06	0.33
1.00	-0.10	-0.20	-0.27	-0.24	0.00	0.24	0.27	0.20	0.10	0.00	0.27
1.25	-0.10	-0.20	-0.29	-0.33	-0.21	0.07	0.22	0.19	0.10	-0.06	0.33
1.67	-0.10	-0.20	-0.30	-0.38	-0.39	-0.20	0.09	0.17	0.10	-0.13	0.39
2.50	-0.10	-0.20	-0.30	-0.39	-0.48	-0.49	-0.27	0.07	0.09	-0.23	0.49
5.00	-0.10	-0.20	-0.30	-0.40	-0.50	-0.59	-0.65	-0.51	0.01	-0.36	0.65
ave	-0.09	-0.10	-0.09	-0.06	0.00	0.06	0.09	0.10	0.09		
max abs	0.10	0.51	0.65	0.59	0.50	0.59	0.65	0.51	0.10		
	SSE	7.01		MSE	0.09		RMSE	0.29			

5. The minimum absolute value of the row average error occurs at $K_1 = 1$, and is strictly increasing as the value of K_1 varies from 1. All errors are, as predicted by our analysis, 0 for $K_1 = 1$ and $m = 0.5$. The minimum row maximum absolute value also occurs at $K_1 = 1$, and is strictly increasing as the value

Table 7.12: Difference Between Desired Value of Fuzzy Membership in Class 1 (equals λ_1) and Computed Value, for Various Ratios of Distances (K_1) and Proportions (K_2), $m = 4.0$

λ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		
K_2	0.11	0.25	0.43	0.67	1.00	1.50	2.33	4.00	9.00		
K_1										ave	max abs
0.20	-0.09	0.66	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.37	0.70
0.40	-0.10	-0.18	0.33	0.58	0.50	0.40	0.30	0.20	0.10	0.24	0.58
0.60	-0.10	-0.20	-0.24	0.30	0.48	0.40	0.30	0.20	0.10	0.14	0.48
0.80	-0.10	-0.20	-0.29	-0.21	0.36	0.39	0.30	0.20	0.10	0.06	0.39
1.00	-0.10	-0.20	-0.30	-0.36	0.00	0.36	0.30	0.20	0.10	0.00	0.36
1.25	-0.10	-0.20	-0.30	-0.39	-0.36	0.21	0.29	0.20	0.10	-0.06	0.39
1.67	-0.10	-0.20	-0.30	-0.40	-0.48	-0.30	0.24	0.20	0.10	-0.14	0.48
2.50	-0.10	-0.20	-0.30	-0.40	-0.50	-0.58	-0.33	0.18	0.10	-0.24	0.58
5.00	-0.10	-0.20	-0.30	-0.40	-0.50	-0.60	-0.70	-0.66	0.09	-0.37	0.70
ave	-0.10	-0.10	-0.11	-0.08	0.00	0.08	0.11	0.10	0.10		
max abs	0.10	0.66	0.70	0.60	0.50	0.60	0.70	0.66	0.10		
	SSE	8.96		MSE	0.11		RMSE	0.33			

of K_1 varies from 1. Therefore the maximum absolute row average and row maximum absolute errors occur at extremes of distance ratios (K_1).

6. The situation with respect to column errors is more complex. The minimum absolute column average error (0) always occurs at $K_2 = 1$ (i.e. equal proportions of the two classes). It generally increases as K_2 varies from 1, but, for $m > 1$, it peaks at mid-low and mid-high values of K_2. A similar situation is observed with the column maximum absolute errors (although these are not 0 at $K_2 = 1$).

7.6 Classified Image Results

The supervised fuzzy Mahalanobis Distance classified TM93 image with $m = 1.25$ was shown in Chapter 5 (Figure 5.20). The results and discussion earlier in this chapter justify the earlier use of the value $m = 1.25$. The cases of $m = 0.5$ and $m = 2.0$ are, however, shown to be of interest in that they are seemingly ideal values of m to ensure

that the fuzzy classifier returns values close to contributing class proportions for the mixel and pure pixel cases respectively. Figures 7.24 and 7.25 show the membership in the class forest calculated by using the supervised fuzzy Mahalanobis Distance classifier with these values of m.

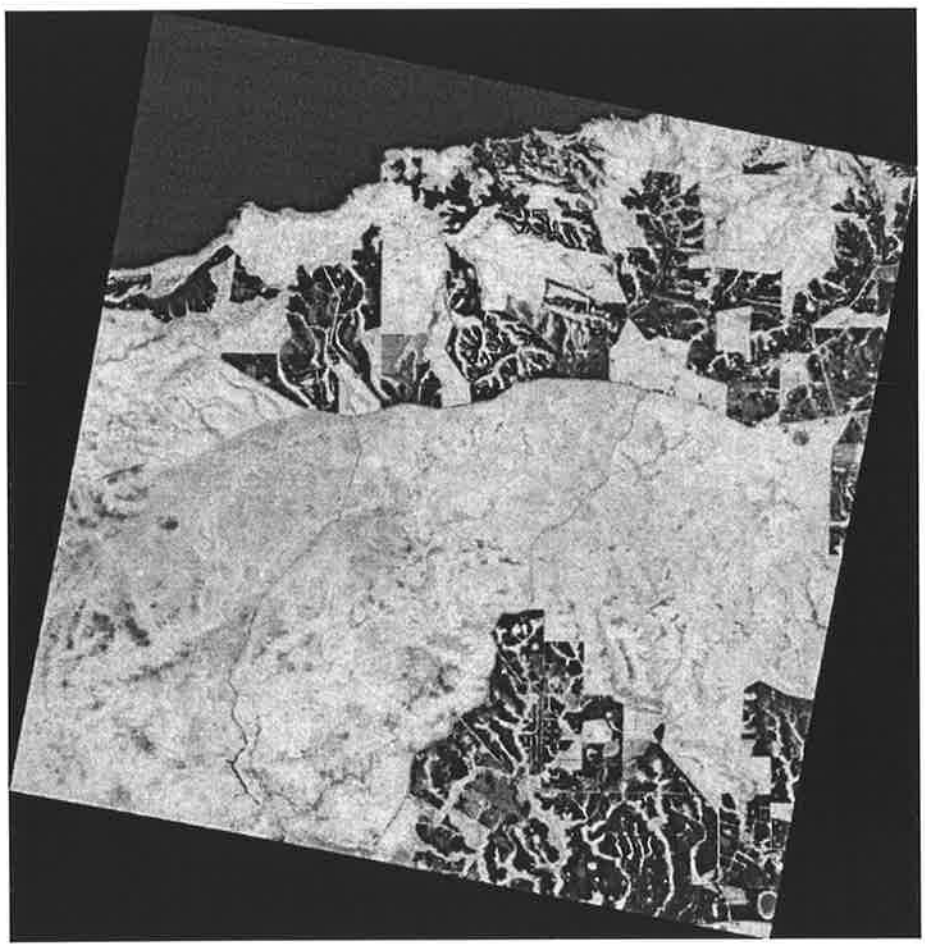

Figure 7.24: Fuzzy Supervised Mahalanobis Distance Classified TM93 Image, Membership in the Class Forest, $m = 0.5$

Again, a detailed assessment of the relative accuracies and merits of these images would require reference data that was not available for this work. Qualitatively, we may say, however, that the image with $m = 2.0$ appears to have lost much of the

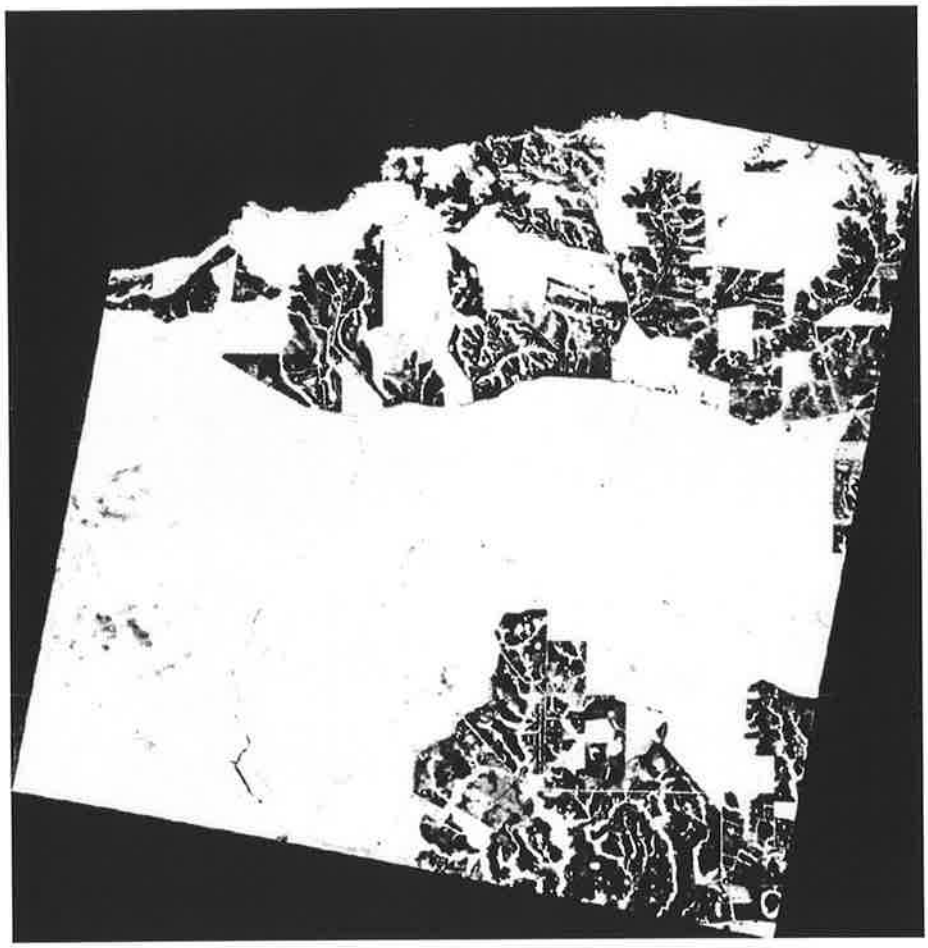

Figure 7.25: Fuzzy Supervised Mahalanobis Distance Classified TM93 Image, Membership in the Class Forest, $m = 2.0$

variability and detail of the other images, and is arguably 'too hard'. This is as expected.

Of course, this (and the analysis earlier in this chapter) refers only to values of m for use in classification: it does not directly address the question of the 'best' value for m in change detection studies. A change image showing the condition "forest in 1989 AND pasture in 1993", computed with a value $m = 0.5$, is shown in Figure 7.26. Qualitatively, this appears to be at least the equal of the change image with $m = 1.25$ (Figure 6.4). It is tentatively concluded that the 'fuzzier' classification

appears superior for change detection purposes. Because of the absence of suitable reference data, quantitative comparisons were not possible.

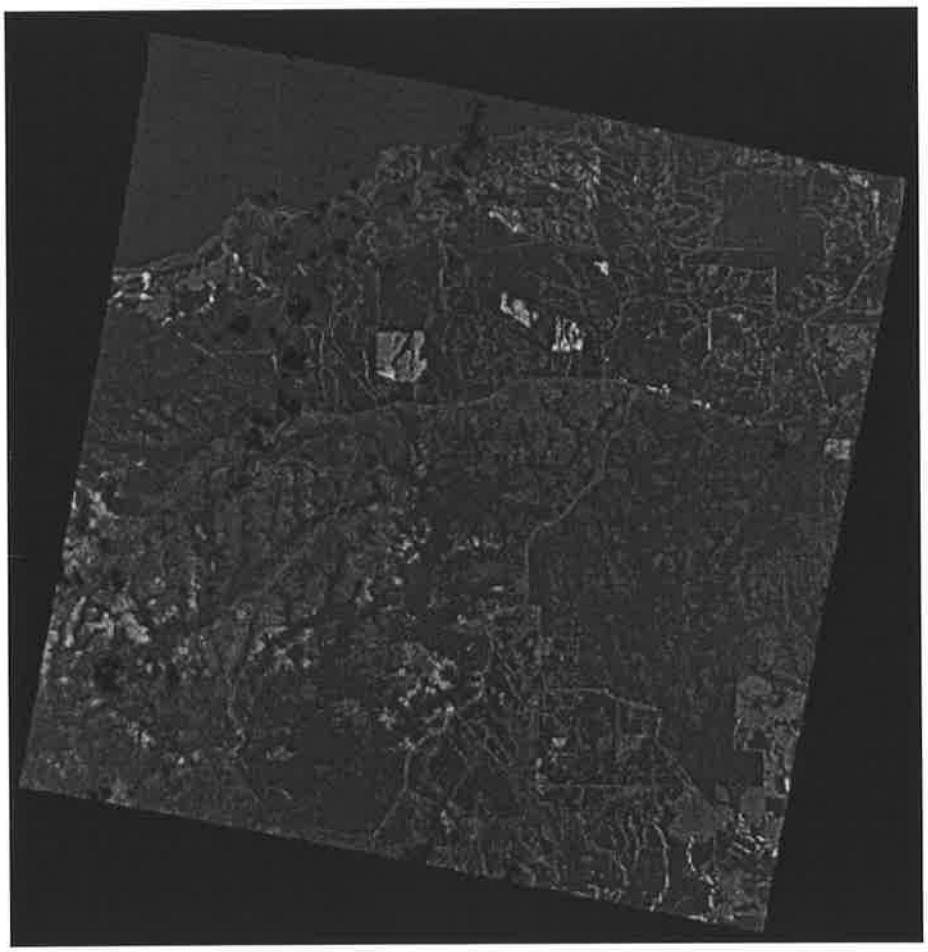

Figure 7.26: Fuzzy Supervised Mahalanobis Distance Change Image, TM93, Forest 1989 AND Pasture 1993, $m = 0.5$

7.7 Conclusion

This chapter has investigated the value of the fuzzy exponent m in the supervised fuzzy Mahalanobis Distance classifier. The underlying philosophy required that the computed fuzzy memberships reflect the proportions of contributing classes for the two class mixed pixel case, but exhibit a high membership of the appropriate single

class in the case of pure, typical pixels. An empirical investigation using the data from this study revealed a conflict between the requirements of these separate cases: the former requires a value of $m = 0.5$; the latter requires a value of about $m = 2$. This accords well with the intuition expressed in Bezdek *et al.* (1984), and arguably provides some explanation of why fuzzy exponent values in this range were said to give "good results". This empirical investigation justifies the use of the value $m = 1.25$ in the change detection results shown earlier in this thesis.

A theoretical investigation of the value of m is also shown, again subject to the condition that the fuzzy memberships reflect the proportions of contributing classes for the two class mixed pixel case. An expression is derived for the theoretically optimum value of m. This is shown, in general, to be a function of both the ratio of the class distance metrics and the ratio of proportions of contributing classes (the very thing which we would often seek to estimate with the fuzzy classification approach). Some special cases are described; most notably, that $m = 0.5$ returns the ideal fuzzy memberships for all class proportions in the case of identical distributions, and that, for some combinations of ratios of Mahalanobis Distances and class proportions, there is no correct choice for m (i.e. no value which will return fuzzy memberships equaling class proportions).

Despite the fact that the value of m is shown to be a function of both the ratio of the class distance metrics and the ratio of proportions of contributing classes, some particular value of m must be chosen *a priori* to implement the classifier. The error sensitivity of the classifier is examined for various values of m, and for various ratios of both Mahalanobis Distances and class proportions. Some trends are apparent: these are discussed in detail in the text.

Chapter 8

POSTERIOR PROBABILITY FUZZY CLASSIFIERS AND FUZZINESS MEASURES

Summary. In this chapter, the notion of calculating fuzzy memberships from the probabilities determined by a traditional Posterior Probability (or Maximum Likelihood) Classifier is examined. It is shown that fuzzy memberships calculated in such a way are 'hard', and unsuited to our purposes. Various measures of fuzziness of this classifier, and the supervised fuzzy Mahalanobis Distance Classifier, are shown. Two approaches to making the output of the Maximum Likelihood Classifier more fuzzy are then developed. The relationship between these approaches and the supervised fuzzy Mahalanobis Distance Classifier is examined, and it is concluded that these approaches both give highly similar results to the supervised fuzzy Mahalanobis Distance Classifier, for any particular chosen pair of classes.

8.1 The Maximum Likelihood Classifier

In Chapter 4, the Maximum Likelihood Classifier was described using the discriminant function approach. In this section, it is necessary to consider it in a more complete

form.

Under the assumption that the conditional probability densities $p(\mathbf{x}|w_i)$ are multivariate normally distributed, the class i probability density function is given by:

$$p(\mathbf{x}|w_i) = \frac{\exp\left\{-d_{ij}^2/2\right\}}{\sqrt{(2\pi)^d \det \Sigma_i}} \tag{8.1}$$

where, as before, d_{ij}^2 is the class i (squared) Mahalanobis Distance.

Using Bayes Rules,

$$p(w_i|\mathbf{x}_j) = \frac{p(\mathbf{x}_j|w_i)p(w_i)}{\sum_{i=1}^{c} p(\mathbf{x}_j|w_i)p(w_i)} \tag{8.2}$$

We can drop the term $(2\pi)^d$ from all subsequent calculations, as it is common to all classes, and is therefore a common factor of both the numerator and denominator in Equation (8.2).

For our purposes, it is convenient to define a function

$$f_i(\mathbf{x}_j) = \exp\left\{k_i - d_{ij}^2/2\right\} \tag{8.3}$$

where

$$k_i = \ln\left(\frac{p(w_i)}{\sqrt{\det \Sigma_i}}\right) \tag{8.4}$$

The posterior probability a particular pixel being from a particular class w_i, given the observation x_j (i.e. the spectral values of the pixel), can be written

$$p(w_i|\mathbf{x}_j) = \frac{f_i(\mathbf{x}_j)}{\sum_{g=1}^{c} f_g(\mathbf{x}_j)} \tag{8.5}$$

8.2 Fuzzy Classification Using Posterior Probabilities

The traditional Maximum Likelihood Classifier simply assigns the class label of the maximum posterior probability: the actual probabilities are discarded (in practice, they are not normally even computed, as the discriminant function approach described in Chapter 4 returns the same result, at lower computational cost).

Foody *et al.* (1992) proposed using the posterior probabilities (and the "typicality" of the pixel to the class, computed from the Mahalanobis Distance under assumptions about the distribution of such distances) as a measure of the uncertainty of the classification, to guide subsequent investigation, validation, recording and use of the information.

Some authors (e.g. Palubinskas, 1995; Schowengerdt, 1996) have suggested that fuzzy memberships can or should be derived from hard, statistical pattern classification approaches, such as the posterior probabilities determined by a Maximum Likelihood classifier.

A Fuzzy Posterior Probability Classifier in which the fuzzy membership values are determined by the posterior probabilities computed according to Equation (8.5) was implemented. The resultant image is shown as Figure 8.1.

Apart from the different colour used to display pine, this figure is indistinguishable from the traditional, 'hard' Maximum Likelihood classified image (which is indistinguishable from the traditional Mahalanobis Distance classified image, shown as Figure 4.2). It is noted that there is no apparent 'fuzziness' in this image.

In the next section, the 'degree of fuzziness' of the Fuzzy Posterior Probability classified image and some Fuzzy Mahalanobis Distance classified images are examined.

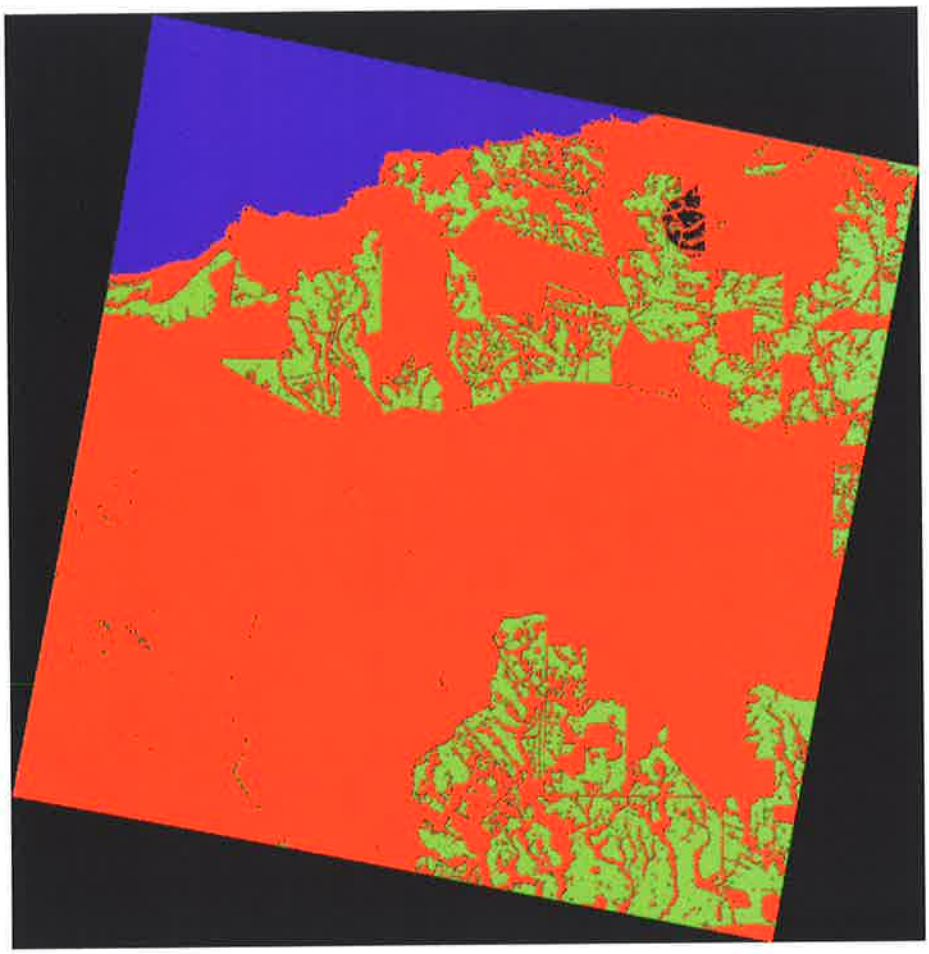

Figure 8.1: Fuzzy Posterior Probability Classified TM93 Image

8.3 Measures of Fuzziness

In investigating fuzzy classifications of images, it is useful to have measures of the 'degree of fuzziness'. An intuitive way of assessing this is to examine the percentage of pixels with membership values falling into various ranges. This can be done by plotting a histogram of the values, grouped into ranges, or bins. The images under examination in this thesis have a large number of pixels, and the fuzzy class memberships of these pixels have a tendency to group at either end of the interval $[0, 1]$. Plotting these memberships on a histogram with equally divided range bins

does not adequately display information about values not near the extremes, as there are relatively few of these in any given range bin. The scale required to plot the large number of pixels in ranges near the extreme values effectively causes all other values to appear to have near zero numbers of pixels.

Fuzzy membership value range information is therefore shown in tables. Table 8.1 shows the percentage of pixels in the fuzzy Mahalanobis Distance classified image, with $m = 1.25$ (equivalently $q = 1.8$), that fall outside the ranges 0.1 to 0.9, 0.01 to 0.99, and 0.001 to 0.999.

Table 8.1: Percentage of Fuzzy Memberships Outside Selected Value Ranges, Mahalanobis Distance Fuzzy Classified TM93 Image, m=1.25

	% memberships outside range		
class	.1 to .9	.01 to .99	.001 to .999
forest	86.01	29.83	1.26
sea	99.63	90.12	89.14
pasture	87.87	60.35	14.07
pine	98.79	79.33	14.60
average	93.07	64.91	29.77

Table 8.1 shows that 86.01% of all pixels in the image have a membership in the class forest less than or equal to 0.1, or greater than or equal to 0.9, and so on. It can be seen, for example, that the classes forest and pasture are more "fuzzy" than the classes sea and pine.

These data are illustrated in Figure 8.2.

Table 8.2 shows the percentage of pixels in the fuzzy Posterior Probability classified image that fall outside the ranges 0.1 to 0.9, 0.01 to 0.99, and 0.001 to 0.999.

These data are illustrated in Figure 8.3.

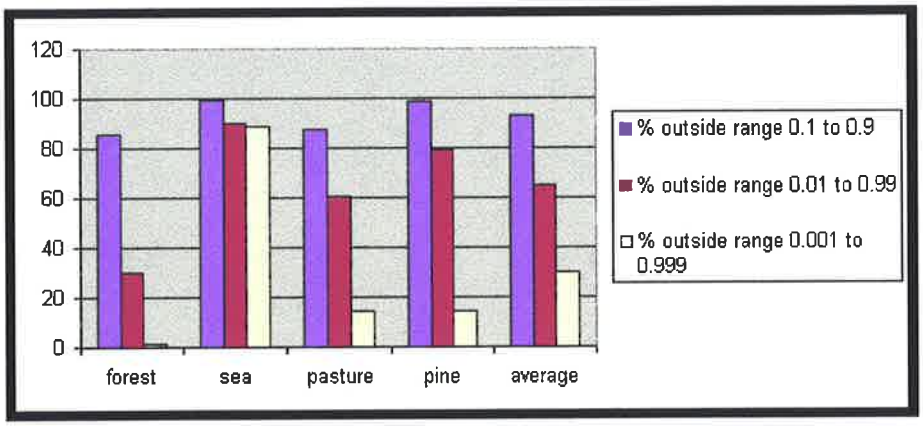

Figure 8.2: Percentage of Pixels With Memberships Outside Various Ranges, Mahalanobis Distance Fuzzy Classifier, TM93 Image

Table 8.2: Percentage of Fuzzy Memberships Outside Selected Value Ranges, Posterior Probability Fuzzy Classified TM93 Image

	% memberships outside range		
class	.1 to .9	.01 to .99	.001 to .999
forest	99.42	98.75	98.10
sea	100.00	99.99	99.99
pasture	99.47	98.90	98.36
pine	99.95	99.85	99.75
average	99.71	99.38	99.05

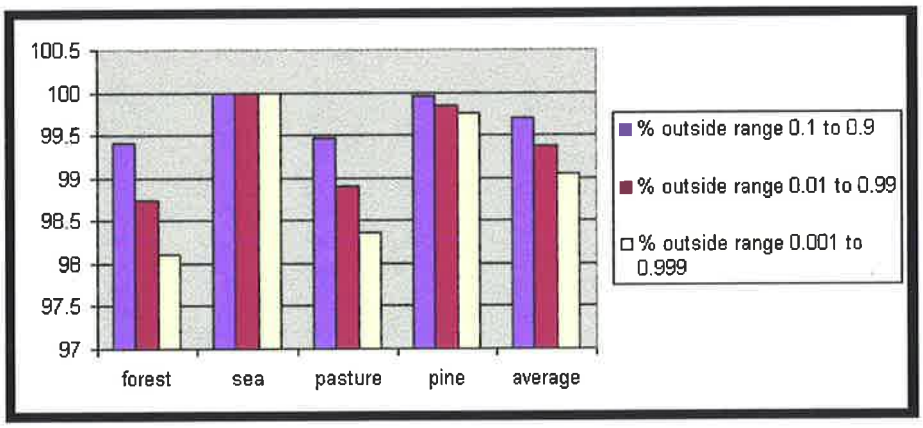

Figure 8.3: Percentage of Pixels With Memberships Outside Various Ranges, Posterior Probability Fuzzy Classifier, TM93 Image

Clearly, the fuzzy Mahalanobis Distance classified image is a much 'fuzzier' image than the fuzzy Posterior Probability classified image.

There has been considerable work reported in the literature on measures of fuzziness (see Pal and Bezdek, 1994, for a comprehensive review of measures of fuzziness and their properties). Bezdek (1981) develops and discusses two: the Partition Coefficient and entropy.

The Partition Coefficient is defined:

$$F = \frac{1}{n} \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^2 \tag{8.6}$$

This ranges between $1/c$ (the most fuzzy) and 1 (hard).

Entropy was developed in analogy to the Shannon entropy (Shannon, 1948), and is defined:

$$H = -\frac{1}{n} \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik} \log_a \mu_{ik} \tag{8.7}$$

This ranges between $\log_a c$ (the most fuzzy) and 0 (hard).

A number of measures have been developed from these, normally by relatively simple arithmetic operations. It is claimed that these developments (and other subsequent measures) more accurately reflect the degree of fuzziness of sets. See e.g. Xie and Beni (1991) or Chi *et al.* (1996). It is, however, noted that much of the cited work is concerned with cluster validity (i.e. what is the 'optimal' number of clusters?). Commonly, the minimum entropy clustering is sought, under the assumption that this will represent a 'good' clustering of the data. Foody (1995) observes that the assumption that a 'good' clustering should exhibit low entropy is, in turn, based on the assumption that the data are 'hard'. This will often be a poor assumption in the

case of remotely sensed data of land cover.

Therefore, in this thesis, we are not interested in the minimum entropy classification. In fact, it is suggested that fuzzy classifications will often more accurately represent the actual land cover conditions. Moreover, change detection using fuzzy classifications may well be enhanced by the use of a classifier more 'fuzzy' than we might have chosen if single date classification were our sole objective. Our interest in fuzziness measures is therefore merely to indicate the relative degree of fuzziness between a number of classifiers, or classes within a classification, not to examine the cluster validity, nor to provide the most accurate or appropriate absolute measure of 'fuzziness'. Therefore, Bezdek's original measures described above will be used.

The natural logarithm base e $(2.718\ldots)$ has been chosen as the logarithmic base for the calculation of entropy.

Table 8.3 compares the Partition Coefficients and entropies of fuzzy Mahalanobis Distance classified images (with $m = 0.5, 1.25$ and 2), and the fuzzy Posterior Probability classified image.

Table 8.3: Fuzziness Measures for Mahalanobis Distance (fuzmah) (Various Exponents) and Posterior Probability Fuzzy Classified TM93 Images

Fuzzy Classifier	Partition coefficient F	Bezdek entropy H
fuzmah ($m = 0.5$)	0.605421	0.750916
fuzmah ($m = 1.25$)	0.912506	0.172919
fuzmah ($m = 2.0$)	0.960164	0.068415
posterior probability	0.997341	0.004396

Another family of useful measures of fuzziness is provided in Kosko (1986). A simple, and for our purposes, adequate measure is given by:

$$R_1(A) = \frac{\sum \text{count}(A \cap A^c)}{\sum \text{count}(A \cup A^c)} \tag{8.8}$$

where $(A \cap A^c)$ denotes the intersection of the fuzzy set A with its complement A^c, $(A \cup A^c)$ denotes the union of the fuzzy set A with its complement, and $\sum \text{count}(.)$ denotes the cardinality of a fuzzy set, and is computed, in this case, by simple term by term addition of all the members of the set.

$R_1(A)$ ranges between 1 (the most fuzzy) and 0 (hard).

It is noted that $R_1(A)$ provides a measure of the fuzziness of each class, whereas the Partitition Coefficient and entropy are normally used to provide a measure of the fuzziness of the entire image.

To compute $R_1(A)$, the following (reasonably standard) conventions are adopted. Denote, as before, the membership of pixel x_j in the class i by μ_{ij}, then determine:

complement by $1 - \mu_{ij}$

intersection by the MIN operator

union by the MAX operator

Kosko (1986) observes that computing $R_1(A)$ is simplified by noting that

$$\sum \text{count}(A \cap A^c) = n - \sum \text{count}(A \cup A^c) \tag{8.9}$$

where n is the number of observations (pixels).

Kosko's fuzzy entropies are shown in Table 8.4. "Fuzmah" denotes the Mahalanobis Distance fuzzy classifier.

As expected, for each class, the fuzzy Mahalanobis Distance classified image, $m = 0.5$, is 'most fuzzy', and the fuzzy Posterior Probability classified image is 'most hard' (in fact, it is 'almost hard').

Table 8.4: Kosko's Entropy Measure $R_1(A)$ for Mahalanobis Distance (Various Exponents) and Posterior Probability Fuzzy Classified TM93 Images

Fuzzy Classifier	forest	sea	pasture	pine
fuzmah ($m = 0.5$)	0.301594	0.047186	0.164522	0.099363
fuzmah ($m = 1.25$)	0.057422	0.003851	0.044875	0.009637
fuzmah ($m = 2.0$)	0.028021	0.002452	0.025659	0.003678
posterior probability	0.001840	0.000008	0.001661	0.000170

Thus we have shown empirically that, for our purposes, a 'fuzzy' classifier based on the posterior probabilities from a Maximum Likelihood classifier is not an appropriate classifier. In fact, it may reasonably be stated that such a classifier appears to be scarcely a fuzzy classifier at all. Empirically, it reveals an almost identical result to the 'crisp' or 'hard' Maximum Likelihood classified image. An analysis of the distribution of its membership values also shows it to be a very 'hard' classifier. This is further supported by various measures of fuzziness.

8.4 Fuzzy Posterior Probability Based Classifiers

The empirical examination of a fuzzy classifier using the posterior probabilities from a Maximum Likelihood classifier reveals that it produces very close to a 'hard' classification. This is due to the nature of the function used to compute these probabilities.

The fuzzy class memberships derived from the *a posteriori* probabilities can be expressed in the form

$$\mu_{ij} = \frac{\exp\left\{k_i - d_{ij}^2/2\right\}}{\sum_{g=1}^{c} \exp\left\{k_g - d_{gj}^2/2\right\}} \quad i = 1, \ldots, c; \; j = 1, \ldots, n \tag{8.10}$$

where μ_{ij} is the membership of the pixel x_j in class i; there are c classes; exp(...) means raise e to this power; d_{ij} is the pixel's i-th class Mahalanobis Distance; and k_i

is a class dependant constant given by

$$k_i = -\frac{d}{2} \ln 2\pi - \frac{1}{2} \ln \det \Sigma_i + \ln p(w_i) \tag{8.11}$$

where d is the number of spectral bands; ln is the natural logarithm; $(\det \Sigma_i)$ is the determinant of the covariance matrix of the class i; and $p(w_i)$ is the *a priori* probability of a pixel being a member of class i.

The k_i for the data used in this thesis are: $k_f = -9.8$ (forest), $k_s = -5.4$ (sea), $k_p = -10.4$ (pasture), $k_\pi = -11.3$ (pine).

They are, in general, much smaller than the class Mahalanobis Distances, and, for simplicity, are ignored in this illustration.

In order to illustrate the 'hard' nature of the 'fuzzy' Posterior Probability membership function, consider a synthetic pixel of arbitrary proportions in the two class mixed pixel case, constructed under a linear mixing model assumption. Consider a single measurement 'dimension', with the two classes having means of 0 and 100 respectively, and unit variance. The synthetic pixel comprises a mixture of these two classes, constructed under the linear mixing model. Because the class variances are equal (and the means are 0 and 100), a pixel mixed in the proportion 0.75:0.25 would be expected to be found at a distance of 25 units from the mean of class 1 (0), and 75 units from the mean of class 2 (100), and so on.

Figure 8.4 shows fuzzy memberships in class 1 as a function of distance. Both the Fuzzy Mahalanobis Distance Classifier (with m = 1.25) and the Fuzzy Posterior Probability Classifier are shown on the same graph, with the former plotted in black, and the latter in green.

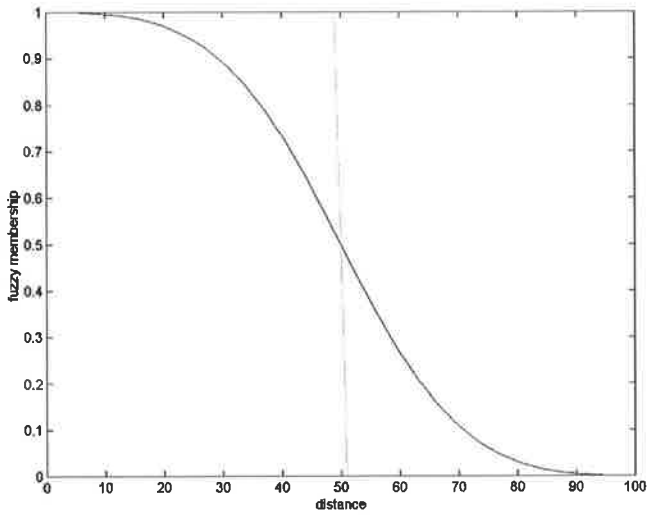

Figure 8.4: Fuzzy Membership of Class 1 versus Distance from Class 1: Fuzzy Mahalanobis Distance Classifier, $m = 1.25$ (in black) and Fuzzy Posterior Probability Classifier (in green)

Note the near step function in the Fuzzy Posterior Probability Classifier. In fact, the step is actually steeper than shown: it was necessary to degrade the step so that it showed in the plot. It was also necessary to move the plotted values of the function slightly away from the grid lines, so that they could be seen. Essentially, the membership in class 1 is unity until the distance is very close to 50 (halfway), when it abruptly changes to a membership of 0.5 in each class. As distance increases (slightly), it abruptly changes to a membership of 0 in class 1 (and therefore unity in class 2).

Thus it is illustrated that a 'fuzzy' membership derived from the posterior probabilities of a Maximum Likelihood classifier will almost invariably return values of 0 or 1 (due to this 'step function' characteristic).

8.4.1 Scaling by Exponentiating

We next pursue an investigation of an alternate fuzzy classifier of the same form as that used for the Mahalanobis Distance Classifier, but now choosing an 'exponential distance' measure.

$$\delta_{ij}^2 = \exp\left(-k_i + d_{ij}^2/2\right) \tag{8.12}$$

(with k_i and d_{ij} as previously defined)

Construct a fuzzy membership function as before

$$v_{ij} = \frac{\left(1/\delta_{ij}^2\right)^{\alpha}}{\sum_{g=1}^{c}\left(1/\delta_{gj}^2\right)^{\alpha}} \quad i = 1, \ldots, c; \quad j = 1, \ldots, n \tag{8.13}$$

This leads to

$$v_{ij} = \frac{\left(\exp\left(k_i - d_{ij}^2/2\right)\right)^{\alpha}}{\sum_{g=1}^{c}\left(\exp\left(k_g - d_{gj}^2/2\right)\right)^{\alpha}} \quad i = 1, \ldots, c; \quad j = 1, \ldots, n; \quad \alpha \geq 0 \tag{8.14}$$

With the exception of the introduction in this thesis of the exponent α (and a slightly different, though equivalent, mathematical form for determining the fuzzy memberships), this was essentially the approach adopted in an adaptation of the fuzzy c-means unsupervised clustering algorithm by Gath and Geva (1989). This adaptation was found to confer some advantages over the Euclidean Distance fuzzy c-means algorithm, but it was noted that it converged to a local optimum in a narrow region, and therefore required "good seed points". The advantages related to a better ability to accommodate variability in cluster shapes and densities: Gath and Geva

(1989) did not, however, examine the Mahalanobis Distance variant of the fuzzy c-means algorithm (which we might reasonably expect to also exhibit an ability to accommodate such variability).

Equation (8.14) has an obvious similarity to the classical 'hard' Posterior Probability Classifier.

If we write the classical 'hard' Posterior Probability Classifier in the form

$$\frac{f(x, \mu_i, \Sigma_i)}{\sum_{g=1}^{c} f(x, \mu_g, \Sigma_g)} \quad i = 1, \ldots, c \tag{8.15}$$

where $f(x, \mu_i, \Sigma_i)$ denotes the (assumed) multi-variate Gaussian probability density function for class i, then the fuzzy memberships determined by (8.14) are essentially of the form

$$\frac{(f(x, \mu_i, \Sigma_i))^{\alpha}}{\sum_{g=1}^{c} (f(x, \mu_g, \Sigma_g))^{\alpha}} \quad i = 1, \ldots, c \tag{8.16}$$

Each density function has been raised to the power α, essentially to influence the 'fuzziness' of the classification (in the same way that m influences the 'fuzziness' of the Mahalanobis Distance fuzzy classifier used previously).

Now, it might be observed that such an approach suffers from the same objection that may be raised against the Mahalanobis Distance fuzzy classifier: the essentially arbitrary nature of the final membership values through the arguably arbitrary selection of m. We have, however, taken some effort earlier to show that the selection of m is not entirely arbitrary, nor based only on aesthetic criteria (such as, "it looks good"): m can be selected in such a way as to provide a reasonably good correspondence between fuzzy memberships and class proportions.

We now compare the two approaches (the fuzzy Mahalanobis Distance Classifier

used previously and a Posterior Probability Classifier made 'more fuzzy').

We adopt the simplifying assumption that the k_i are approximately equal (as is the case with the classes of primary interest in this research): this permits us to drop these terms (as they become a common factor, and cancel out). It is also noted that they are small relative to the Mahalanobis Distance terms. They will be re-introduced when some subsequent analysis reaches a stage where they may prove relevant.

Consider, for simplicity, the two class case. In order for the two fuzzy classifiers to be identical, we require that $\mu_{ij} = v_{ij}, j = 1, ..., n; i = 1, 2$ (i.e. n pixels, two classes).

Examining the memberships in class 1, we require that

$$\frac{\left(1/d_{1j}^2\right)^m}{\left(1/d_{1j}^2\right)^m + \left(1/d_{2j}^2\right)^m} = \frac{\left(\exp\left(-d_{1j}^2/2\right)\right)^\alpha}{\sum_{g=1}^{2}\left(\exp\left(-d_{gj}^2/2\right)\right)^\alpha} = \frac{\exp\left(-\alpha \cdot d_{1j}^2/2\right)}{\exp\left(-\alpha \cdot d_{1j}^2/2\right) + \exp\left(-\alpha \cdot d_{2j}^2/2\right)} \tag{8.17}$$

$$\frac{\exp\left(-\alpha \cdot d_{1j}^2/2\right) + \exp\left(-\alpha \cdot d_{2j}^2/2\right)}{\exp\left(-\alpha \cdot d_{1j}^2/2\right)} = \frac{\left(1/d_{1j}^2\right)^m + \left(1/d_{2j}^2\right)^m}{\left(1/d_{1j}^2\right)^m}$$

$$1 + \frac{\exp\left(-\alpha \cdot d_{2j}^2/2\right)}{\exp\left(-\alpha \cdot d_{1j}^2/2\right)} = 1 + \frac{\left(1/d_{2j}^2\right)^m}{\left(1/d_{1j}^2\right)^m} \tag{8.18}$$

$$\frac{\exp\left(-\alpha \cdot d_{2j}^2/2\right)}{\exp\left(-\alpha \cdot d_{1j}^2/2\right)} = \left(\frac{d_{2j}^2}{d_{1j}^2}\right)^{-m} \tag{8.19}$$

Taking (natural) logs,

$$\alpha \cdot d_{1j}^2/2 - \alpha \cdot d_{2j}^2/2 \;=\; -m\left(\ln d_{2j}^2 - \ln d_{1j}^2\right)$$

$$\alpha\left(d_{1j}^2 - d_{2j}^2\right) = -2m\left(\ln d_{2j}^2 - \ln d_{1j}^2\right)$$

$$\alpha = 2m\frac{\ln d_{1j}^2 - \ln d_{2j}^2}{d_{1j}^2 - d_{2j}^2} \qquad \left(= 2m\frac{\ln d_{2j}^2 - \ln d_{1j}^2}{d_{2j}^2 - d_{1j}^2}\right) \tag{8.20}$$

There is therefore no value of α that can be selected, for a given m, to have $\mu_{ij} = \nu_{ij}$, independent of the d_{ij}, for all pixels in the image We can, however, select a value for α that, for a given m, returns a surprisingly close correspondence between the fuzzy membership values (at least for a certain range of Mahalanobis Distance values).

To illustrate, Figure 8.4 earlier showed the two class case, equal (unit) variances, class means at 0 and 100, $m = 1.25$, $\alpha = 1$ (i.e. the traditional Maximum Likelihood Classifier). Figure 8.5 shows the two class case, equal (unit) variances, class means at 0 and 100, $m = 1.25$, $\alpha = 0.0011$.

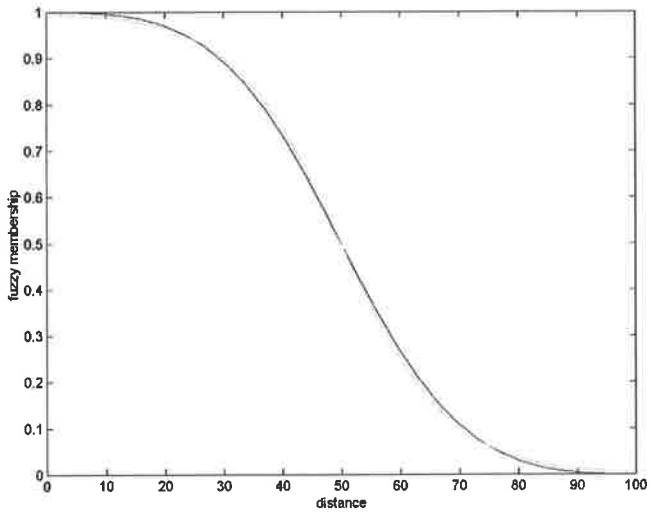

Figure 8.5: Fuzzy Membership of Class 1 versus Distance from Class 1: Fuzzy Mahalanobis Distance Classifier, $m = 1.25$ (in black) and Fuzzy Posterior Probability Classifier (Scaled by Exponentiating), $\alpha = 0.0011$ (in green)

The maximum (absolute value of the) discrepancy between these two functions (in the range $d_{ij} = [0, 100]$) is 0.0172 (this occurs at $d_{ij} = 19$ and 81, where $\mu_{ij} = 0.974$). As a relative value, this is quite small (less than 1.8%).

The value $\alpha = 0.0011$ was determined using Equation (8.20) and the Mahalanobis Distances 75 and 25 (representing a pixel with class proportions 0.25:0.75 respectively). Other class proportions also return values for α close to 0.0011. It is noted that Equation (8.20) cannot be used if the class proportions are, in this example, 0.5:0.5, because this gives equal Mahalanobis Distances, and causes a 'divide-by-zero' error. In the case of equal Mahalanobis Distances, any value of α will suffice. It is also noted that Equation (8.20) is inappropriate in the case of pixels having zero Mahalanobis Distances (as we might expect to find with 'pure', highly typical pixels), because then we would be attempting to calculate $\ln 0$.

The correspondence between the values of the two classifiers is even greater at higher values of m. Figure 8.6 shows the case with $m = 2$, $\alpha = 0.00168$, Figure 8.7 shows the case with $m = 4$, $\alpha = 0.0032$, and Figure 8.8 shows the case with $m = 8$, $\alpha = 0.0065$. Conversely, the correspondence between the two classifiers is less at lower values of m. Figure 8.9 shows the case with $m = 1$, $\alpha = 0.00092$, Figure 8.10 shows the case with $m = 0.5, \alpha = 0.00055$, and Figure 8.11 shows the case with $m = 0.25$, $\alpha = 0.00028$. Again, all of these values for α were computed using Equation (8.20), with class proportions of 0.25:0.75.

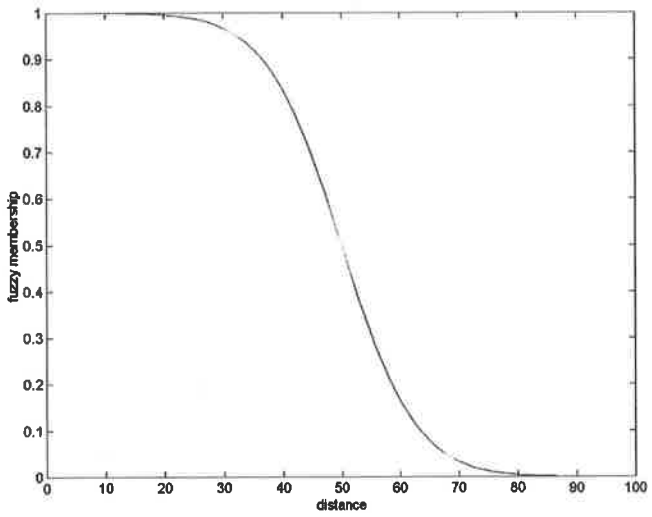

Figure 8.6: Fuzzy Membership of Class 1 versus Distance from Class 1: Fuzzy Mahalanobis Distance Classifier, $m = 2$ (in black) and Fuzzy Posterior Probability Classifier (Scaled by Exponentiating), $\alpha = 0.00168$ (in green)

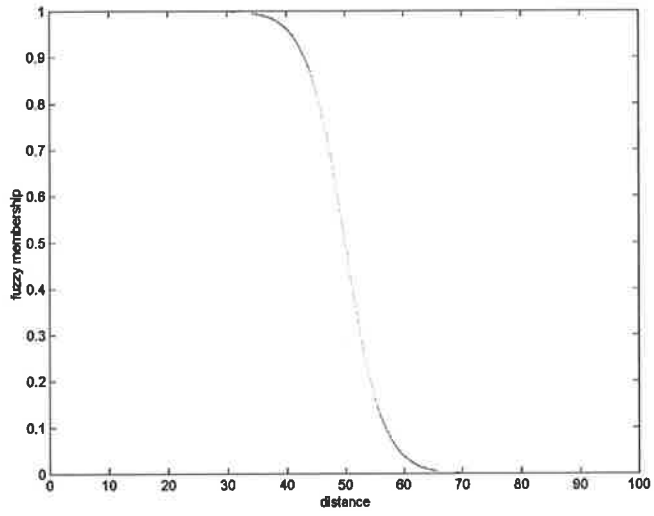

Figure 8.7: Fuzzy Membership of Class 1 versus Distance from Class 1: Fuzzy Mahalanobis Distance Classifier, $m = 4$ (in black) and Fuzzy Posterior Probability Classifier (Scaled by Exponentiating), $\alpha = 0.0032$ (in green)

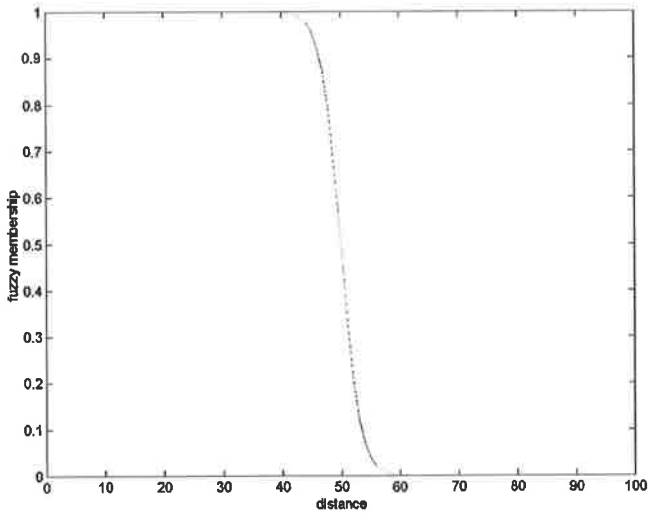

Figure 8.8: Fuzzy Membership of Class 1 versus Distance from Class 1: Fuzzy Mahalanobis Distance Classifier, $m = 8$ (in black) and Fuzzy Posterior Probability Classifier (Scaled by Exponentiating), $\alpha = 0.0065$ (in green)

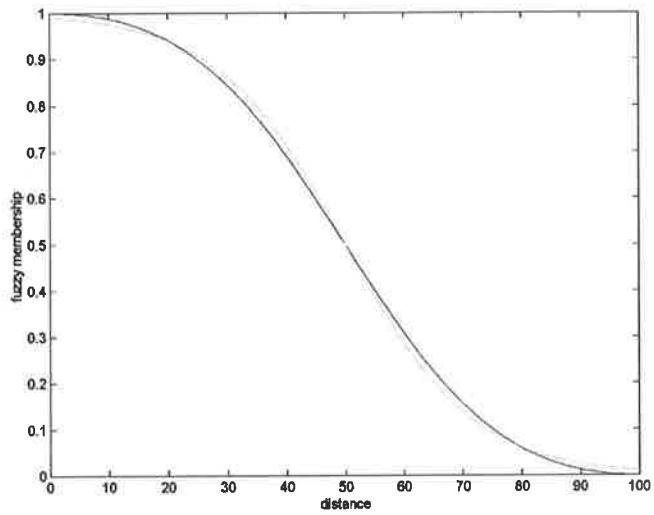

Figure 8.9: Fuzzy Membership of Class 1 versus Distance from Class 1: Fuzzy Mahalanobis Distance Classifier, $m = 1$ (in black) and Fuzzy Posterior Probability Classifier (Scaled by Exponentiating), $\alpha = 0.00092$ (in green)

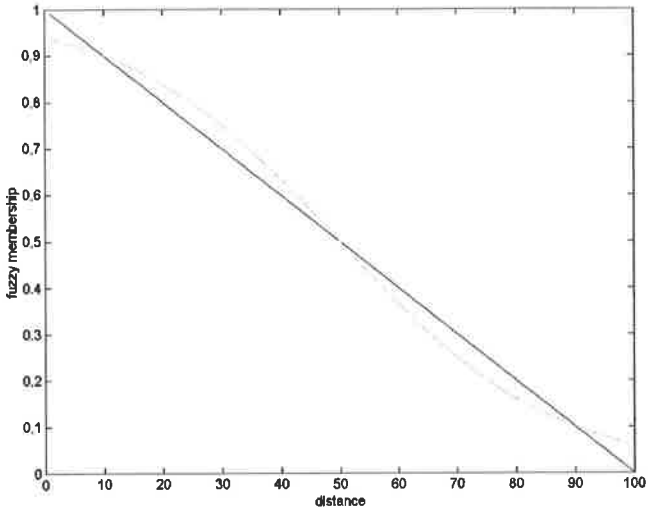

Figure 8.10: Fuzzy Membership of Class 1 versus Distance from Class 1: Fuzzy Mahalanobis Distance Classifier, $m = 0.5$ (in black) and Fuzzy Posterior Probability Classifier (Scaled by Exponentiating), $\alpha = 0.00055$ (in green)

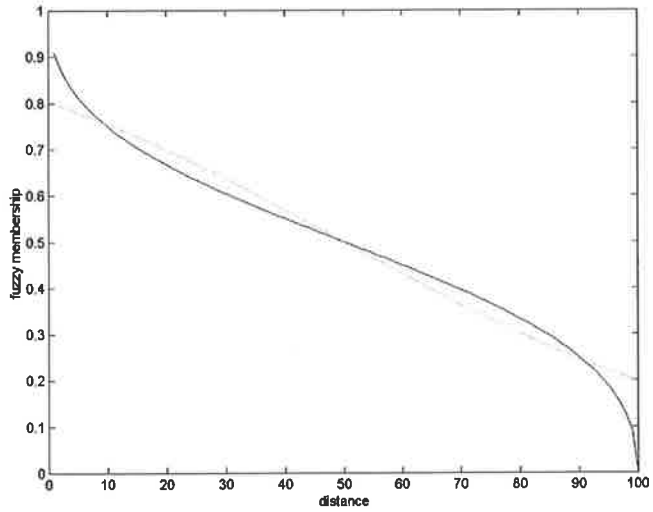

Figure 8.11: Fuzzy Membership of Class 1 versus Distance from Class 1: Fuzzy Mahalanobis Distance Classifier, $m = 0.25$ (in black) and Fuzzy Posterior Probability Classifier (Scaled by Exponentiating), $\alpha = 0.00028$ (in green)

This illustrates that a Posterior Probability based fuzzy classifier can return, under certain conditions and with appropriate selections of m and α, very similar results to those of a Mahalanobis Distance fuzzy classifier.

Rather than continue this analysis with non-equal variances, or including the previously omitted k_i terms in this illustration, we return our attention to our real data.

To calculate the value of α required so that the posterior probability based fuzzy classifier will give similar results to the Mahalanobis Distance fuzzy classifier with a value of $m = 1.25$, we need some empirical values for Mahalanobis Distances.

Clearly, we would like the results to be similar for all mixture values. The principal mixture of interest in this study is forest/pasture. Earlier work produced a table of expected (squared) Mahalanobis Distances for pixels comprising mixtures in the ratio 0.5:0.5. The relevant values for a 0.5:0.5 mixture of forest and pasture were 26.9 (expected (squared) forest Mahalanobis Distance) and 83.9 (expected (squared) pasture Mahalanobis Distance). With $m = 1.25$, and again using Equation (8.20), this implies $\alpha = 0.05$.

Calculations for a more complete range of mixture proportions are shown in Table 8.5.

Figure 8.12 shows the Posterior Probability based fuzzy classified TM93 image (prior probability and determinant terms omitted), with $\alpha = 0.05$.

We note that this image reveals highly similar values to the Mahalanobis Distance fuzzy image, $m = 1.25$, shown earlier (Figure 5.20) for those pixels with memberships distributed largely between the classes forest and pasture. However, those pixels that previously had memberships distributed between the classes forest and pine have

Table 8.5: Required Value of α to (approximately) Equal Results of the Mahalanobis Distance Fuzzy Classifier, $m = 1.25$, TM93 Image

Proportion of forest and pasture		Required α
forest	pasture	
0.9	0.1	0.05
0.8	0.2	0.05
0.7	0.3	0.05
0.6	0.4	0.05
0.5	0.5	0.05
0.4	0.6	0.05
0.3	0.7	0.06
0.2	0.8	0.07
0.1	0.9	0.10
	average	0.06
	median	0.05

become 'more fuzzy'. Pixels with high forest membership have had that membership reduced, with a corresponding increase in membership of pine. Similarly, pixels with high pine membership have had that membership reduced, with a corresponding increase in membership of forest.

It is attractive to postulate that, with $\alpha = 0.05$, pixels lying along the line joining the means of the training classes forest and pasture (as we would expect to be the case where pixels comprised a mixture of these two classes) return similar values to the Mahalanobis Distance fuzzy image ($m = 1.25$). However, pixels lying along the line joining the means of the training classes forest and pine have been made too 'fuzzy' by a selection of $\alpha = 0.05$ (the median value for α for various proportions of these two classes is 0.15, therefore $\alpha = 0.05$ makes these classes more fuzzy). Although it is hard to detect by visual inspection of the images, pixels that previously had their memberships distributed between the classes pasture and pine have now become much 'harder' (the median value for α for various proportions of these two classes is 0.01, therefore $\alpha = 0.05$ makes these classes more 'hard').

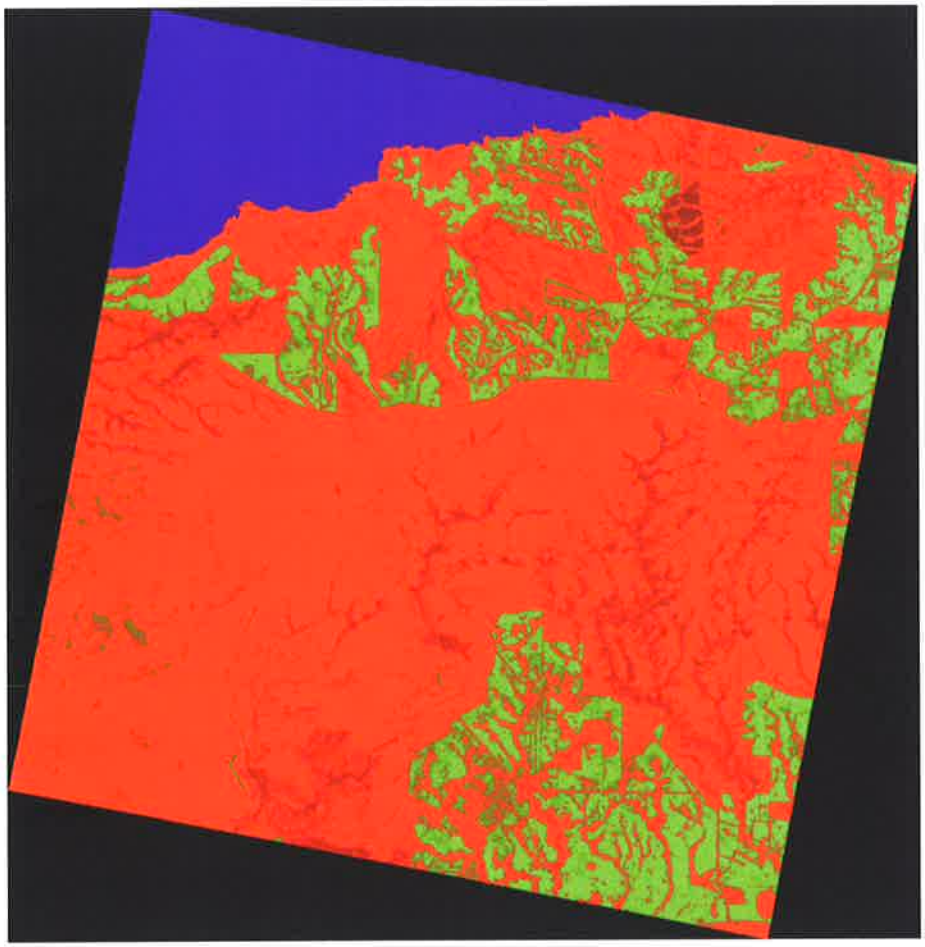

Figure 8.12: Fuzzy Posterior Probability Classified TM93 Image, Scaled by Exponentiating, $\alpha = 0.05$ (Prior Probability and Determinant Terms Omitted)

8.4.2 Direct Scaling of Covariance Matrix

Finally, we compare the Posterior Probability based fuzzy classifier approach described above, with a Posterior Probability based fuzzy classifier approach in which we simply scale the covariance matrices by an appropriate factor.

Using the notation introduced earlier, we denote a posterior probability based fuzzy classifier with a scaled covariance matrix by

$$\frac{f(x, \mu_i, \frac{1}{\alpha}\Sigma_i)}{\sum_{g=1}^{c} f(x, \mu_g, \frac{1}{\alpha}\Sigma_g)} \tag{8.21}$$

If we continue to omit the k_i terms,

$$f(x, \mu_i, \frac{1}{\alpha}\Sigma_i) = \exp\left\{-\left(d'_{ij}\right)^2/2\right\} \tag{8.22}$$

where $(d'_{ij})^2$ is the (squared) Mahalanobis Distance computed with the inverse of the scaled covariance matrix $\frac{1}{\alpha}\Sigma_i$.

$$\text{The inverse of the matrix } \frac{1}{\alpha}\Sigma_i \text{ denoted } \left(\frac{1}{\alpha}\Sigma_i\right)^{-1} \text{ is } \alpha\Sigma_i^{-1}. \tag{8.23}$$

$$\text{Proof: } \left(\frac{1}{\alpha}\Sigma_i\right)\left(\alpha\Sigma_i^{-1}\right) = \frac{\alpha}{\alpha}\Sigma_i\Sigma_i^{-1} = I$$

Therefore,

$$\begin{aligned} f(x, \mu_i, \frac{1}{\alpha}\Sigma_i) &= \exp\left\{-(x-\mu_i)^t \alpha\Sigma_i^{-1}(x-\mu_i)/2\right\} \\ &= \left(\exp\left\{-(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)/2\right\}\right)^{\alpha} \\ &= (f(x, \mu_i, \Sigma_i))^{\alpha} \end{aligned} \tag{8.24}$$

It is therefore shown that, with the omission of the k_i terms, these two forms of Posterior Probability based fuzzy classifier (i.e. scaling by exponentiation and direct scaling of the covariance matrix) are identical. It has been previously shown that a posterior probability based fuzzy classifier with scaling by exponentiation can,

with appropriate selection of the scaling parameter, produce highly similar results to a Mahalanobis Distance fuzzy classifier. Figure 8.12, which showed the Posterior Probability classified image with scaling by exponentiating, but with the prior probability and determinant terms omitted, therefore also represents the Posterior Probability classified image with direct scaling of the covariance matrix (again with the prior probability and determinant terms omitted).

Let us now consider the effect of including these terms.

$f(x, \mu_i, \frac{1}{\alpha}\Sigma_i)$ is now strictly a function of the form $f(x, \mu_i, \frac{1}{\alpha}\Sigma_i, p(w_i))$ (and, indeed, is also a function of the number of bands), but we will continue to use the notation $f(x, \mu_i, \frac{1}{\alpha}\Sigma_i)$.

$$
\begin{aligned}
f(x, \mu_i, \frac{1}{\alpha}\Sigma_i) &= \frac{p(w_i)}{(2\pi)^{d/2}\left(\det(\frac{1}{\alpha}\Sigma_i)\right)^{1/2}} \exp\left\{-(x-\mu_i)^t \left(\frac{1}{\alpha}\Sigma_i\right)^{-1} (x-\mu_i)/2\right\} \\
&= \frac{p(w_i)}{(2\pi)^{d/2}\left(\det(\frac{1}{\alpha}\Sigma_i)\right)^{1/2}} \exp\left\{-(x-\mu_i)^t \alpha\Sigma_i^{-1}(x-\mu_i)/2\right\}
\end{aligned}
$$

Now $\det\left(\frac{1}{\alpha}\Sigma_i\right) = \left(\frac{1}{\alpha}\right)^d \det\Sigma_i$ (because Σ_i is an $d \times d$ matrix).

Therefore

$$
\begin{aligned}
f(x, \mu_i, \frac{1}{\alpha}\Sigma_i) &= \left(\frac{p(w_i)}{(2\pi)^{d/2}\left(\frac{1}{\alpha}\right)^{d/2}(\det\Sigma_i)^{1/2}}\right)^{N/2} \exp\left\{-\alpha(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i)/2\right\} \\
&= \left(\frac{\alpha}{2\pi}\right)^{d/2} \frac{p(w_i)}{(\det\Sigma_i)^{1/2}} \exp\left\{-\alpha(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i)/2\right\} \qquad (8.25)
\end{aligned}
$$

Therefore, a fuzzy classifier of the form (8.21) is

$$\frac{\frac{p(w_i)}{(\det \Sigma_i)^{1/2}} \exp\left\{-\alpha(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)/2\right\}}{\sum_{g=1}^{c} \frac{p(w_g)}{(\det \Sigma_g)^{1/2}} \exp\left\{-\alpha(x-\mu_g)^t \Sigma_g^{-1}(x-\mu_g)/2\right\}} \tag{8.26}$$

with the $\left(\frac{\alpha}{2\pi}\right)^{d/2}$ terms cancelling out.

Considering now a classifier with scaling by exponentiation,

$$(f(x,\mu_i,\Sigma_i))^\alpha = \left\{\frac{p(w_i)}{\left(\frac{\alpha}{2\pi}\right)^{d/2}(\det \Sigma_i)^{1/2}} \exp\left\{-(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)/2\right\}\right\}^\alpha$$

$$= \left(\frac{1}{2\pi}\right)^{\alpha d/2} \left(\frac{p(w_i)}{(\det \Sigma_i)^{1/2}}\right)^\alpha \exp\left\{-\alpha(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)/2\right\} \tag{8.27}$$

Therefore, a classifier of the form (8.16) is

$$\frac{\left(\frac{p(w_i)}{(\det \Sigma_i)^{1/2}}\right)^\alpha \exp\left\{-\alpha(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)/2\right\}}{\sum_{g=1}^{c} \left(\frac{p(w_g)}{(\det \Sigma_g)^{1/2}}\right)^\alpha \exp\left\{-\alpha(x-\mu_g)^t \Sigma_g^{-1}(x-\mu_g)/2\right\}} \tag{8.28}$$

with the terms $\left(\frac{1}{2\pi}\right)^{\alpha d/2}$ terms cancelling out.

Now, α is required to be greater than zero. Further, we expect that $\alpha << 1$ (where $<<$ denotes "much less than"). We may therefore expect that the term

$$\left(\frac{p(w_i)}{(\det \Sigma_i)^{1/2}}\right)^\alpha$$

will have relatively little impact because it will be similar for all classes. This is illustrated in Table 8.6 with figures from our study. Of course, as α tends to zero, this term tends to unity, regardless of the values of the prior probability and the

determinant of the covariance matrix (provided, of course, that the determinant is non-zero).

Table 8.6: Values for $p(w_i)/\sqrt{\det \Sigma_i}$ (Unscaled and Scaled), TM93 Image

Class	$p(w_i)/\sqrt{\det \Sigma_i}$	$\left(p(w_i)/\sqrt{\det \Sigma_i}\right)^{0.05}$
forest	0.0021	0.74
sea	0.1864	0.92
pasture	0.0012	0.71
pine	0.0005	0.68
Std/mean	1.9462	0.1391

The term "Std/mean" represents the standard deviation divided by the mean. The significantly larger value of this term for the first column of the table shows that the unscaled term $p(w_i)/\sqrt{\det \Sigma_i}$ has a significantly greater impact than the scaled term $\left(p(w_i)/\sqrt{\det \Sigma_i}\right)^{0.05}$.

Therefore, a Posterior Probability fuzzy classifier of the exponentiated form effectively scales the covariance matrix by a factor of $1/\alpha$, effectively making it 'bigger', while rendering the prior probability and determinant terms of relatively little importance. A Posterior Probability fuzzy classifier with a scaled covariance matrix (with the same factor of $1/\alpha$) continues to note a contribution from both of these terms.

It is worth noting that while we might expect an improvement in the classifier's performance by taking these terms into account, we saw no such improvement in the case of the crisp classifier.

The images produced by these two approaches are highly similar. Let us once again consider the two classes of greatest interest, forest and pasture, with $\alpha = 0.05$ for the Posterior Probability fuzzy classifier (of both forms). Figure 8.13 shows the

image with direct scaling of the covariance matrix, and including the prior probability and determinant terms.

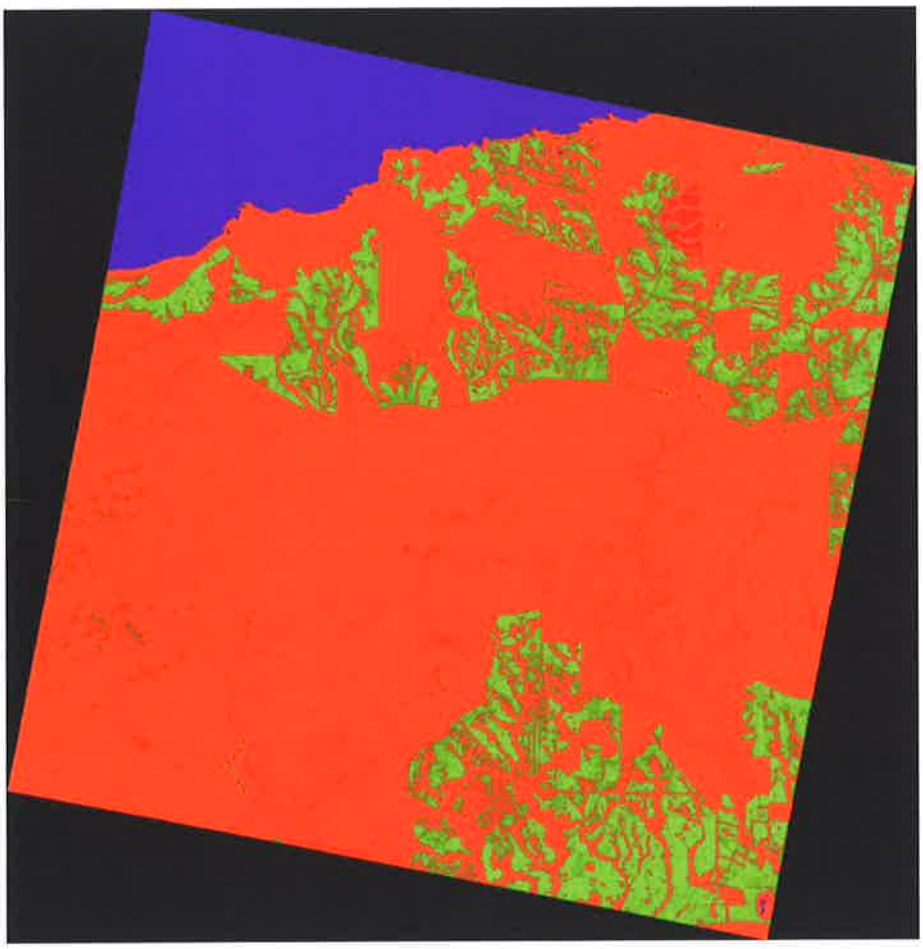

Figure 8.13: Fuzzy Posterior Probability Classified TM93 Image, Scaled Covariance Matrix, $\alpha = 0.05$ (Prior Probability and Determinant Terms Included)

This is compared with the image with scaling by exponentiating (Figure 8.14), with the prior probability and determinant terms also now included. We note that it appears to produce highly similar results for pixels with their memberships distributed partially to forest and partially to pasture.

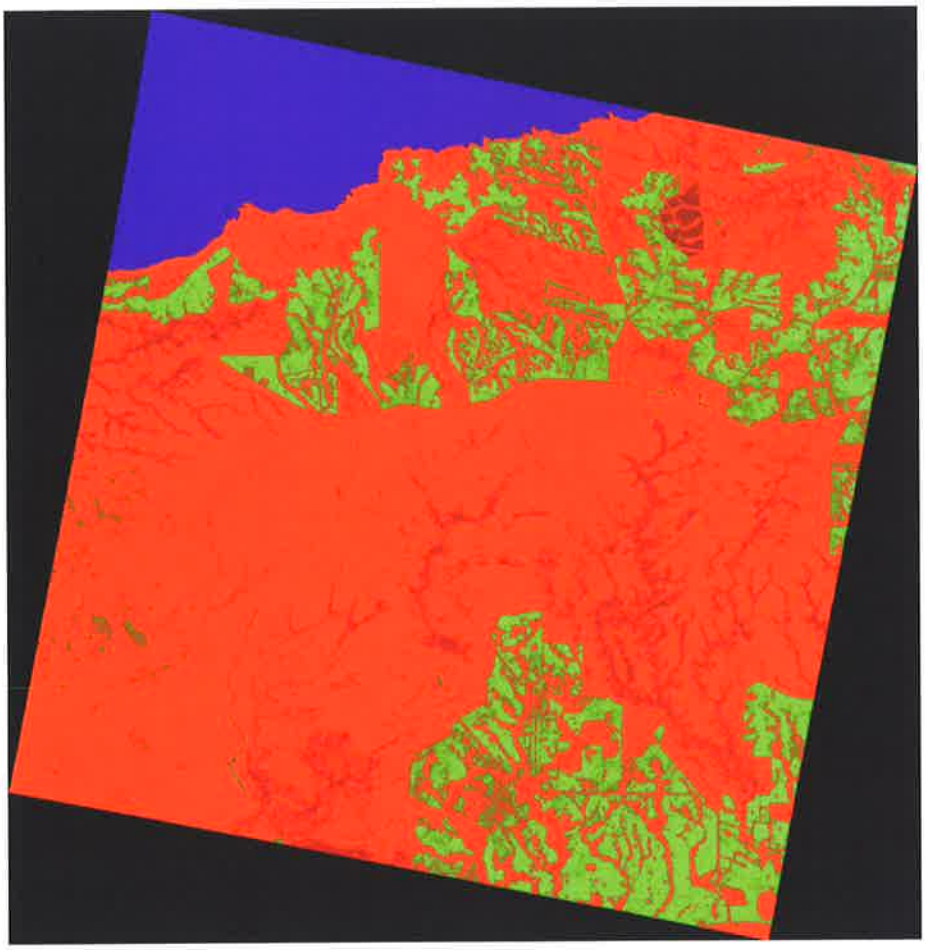

Figure 8.14: Fuzzy Posterior Probability Classified TM93 Image, Scaled by Exponentiating, $\alpha = 0.05$ (Prior Probability and Determinant Terms Included)

It was shown earlier that the two approaches (i.e. scaling by exponentiating and directly scaling the covariance matrix) give identical results if we ignore the terms relating to prior probability and the determinant of the class covariance matrix.

Figure 8.12 showed the image produced if we ignore these terms. As expected, Figure 8.14 is almost identical to Figure 8.12 (we had concluded that the inclusion of the prior probability and determinant terms should have relatively little effect). It is also noted that Figure 8.13 is highly similar to Figure 8.12, particularly in respect of pixels with their memberships distributed partially to forest and partially

to pasture. Also, as expected, these images are highly similar, for these particular circumstances, to the Mahalanobis Distance fuzzy classified image with $m = 1.25$. There are, however, noticeable differences between the images where pixels do not satisfy this condition (because $\alpha = 0.05$ is not the correct scaling factor for pixels with memberships distributed between other classes).

To allow some numeric comparison of the 'degree of fuzziness' of the Posterior Probability classified image (with the prior probability and determinant terms omitted) with images shown earlier, the Partition Coefficient is 0.871490, and the entropy is 0.223090. These values are, as expected, relatively close to those of the Mahalanobis Distance fuzzy classified image with $m = 1.25$ (with the scaled probability fuzzy classified image being, overall, slightly 'more fuzzy'). Information about the 'fuzziness' of each class can be found in the Kosko fuzzy entropy, given in Table 8.7.

Table 8.7: Kosko's Entropy Measure for Scaled ($\alpha = 0.05$) Posterior Probability Fuzzy Classified TM93 Image (Prior Probability and Determinant Terms Omitted)

Kosko's fuzzy entropy $R_1(A)$				
Fuzzy Classifier	forest	sea	pasture	pine
fuzprob ($\alpha = 0.05$)	0.088361	0.007593	0.044311	0.032231

Each class is more fuzzy than those of the Mahalanobis Distance fuzzy classified image with $m = 1.25$, with the exception of pasture, which is very slightly less fuzzy. It is noted that even an overall class 'fuzziness' measure obscures information on how the fuzziness has been distributed. As stated earlier, pixels with memberships distributed between forest and pasture have remained approximately 'equally fuzzy' (a consequence of the method of choosing the scaling exponent α). It is also noted that the contribution that each class makes to the Partition Coefficient and the entropy

could also be separately shown, as each of these would also serve as a measure of the individual class 'fuzziness'.

8.5 Conclusion

In conclusion, it is noted that:

1. A Posterior Probability fuzzy classifier approach, without scaling, does not produce useful results for our purposes. Indeed, such a classifier is scarcely fuzzy at all, with a very high proportion of the pixel memberships in each class being very close to 0 or 1.
2. A Posterior Probability based fuzzy classifier approach, with direct scaling of the covariance matrix, is shown analytically to be highly similar to the Posterior Probability based fuzzy classifier approach with scaling by exponentiating (with the scaling parameter of one being the reciprocal of the scaling parameter of the other). In fact, these approaches are shown to be identical if we ignore the terms relating to prior probabilities and the determinants of the class covariance matrices.
3. A posterior probability based fuzzy classifier approach, with scaling by an appropriately selected exponent, produces a highly similar result to the Mahalanobis Distance fuzzy classifier for any particular choice of two classes. Other classes may be more or less fuzzy than those of the Mahalanobis Distance fuzzy classifier. This is dependent on their class Mahalanobis distances to the means of the other classes.
4. The Mahalanobis Distance fuzzy classifier shown earlier gives good results, and is simple to implement. It is of the same general order of computational complexity as the Posterior Probability based fuzzy classifiers, but noticeably

more efficient.

It is recognised that these conclusions are drawn from limited study, and, in the absence of suitable ground truth with which to evaluate the classifiers, only qualitative analysis has been possible. Never-the-less, it is suggested that there seems no obvious advantage in using a Posterior Probability based fuzzy classifier approach over the Mahalanobis Distance fuzzy classifier used earlier in this thesis.

Chapter 9

CONCLUSION

9.1 Summary

The objective of this thesis was to examine the use of fuzzy classification in the Post Classification Comparison approach to change detection using remotely sensed imagery. The Post Classification Comparison approach was chosen because it addresses the relevant and important aspects of change: detection, location and identification. The use of fuzzy classification was motivated by observations about the inappropriateness of traditional crisp approaches to land cover classification, their relatively poor accuracy, and the tendency for errors in traditional classification to compound in change detection. Research using fuzzy approaches to classification of remotely sensed imagery had shown very good results: it addressed the issue of the inappropriateness of traditional crisp approaches to land cover classification, and showed promise in addressing the mixed pixel problem. It was hypothesised that the issues of poor classification accuracy with traditional classifiers, and compounding errors in traditional Post Classification Comparison change detection, would be treated advantageously by using a fuzzy approach to classification.

Chapter 2 reviewed the literature on change detection, and described and

compared the major change detection techniques. Some observations were made about the essential similarity of a number of techniques that appear, *prima facie*, to be quite different. The notion of pixel, feature and object level analysis was introduced. It was argued that an object level, or classification-based, approach to change detection was highly desirable. Chapter 2 also contains a discussion of a number of issues highly relevant to change detection: registration of imagery, radiometric correction, threshold value selection, and accuracy assessment.

The study area, data and image processing environment used during this research are described in Chapter 3.

Chapter 4 presented a quantitative comparison of supervised and unsupervised approaches to traditional classification. Both supervised and unsupervised approaches to classification were examined. It was concluded, not surprisingly, that supervised approaches are superior when the classes of interest are known *a priori*, and training class information is available or can easily be determined from the data.

Of the supervised classification approaches, it was found that the full Maximum Likelihood classifier (under the multi-variate Gaussian distribution assumption), with terms for *a priori* class probability and determinant of the class covariance matrices, offered no appreciable improvement over the simpler Mahalanobis Distance classifier. The investigation of traditional classification approaches provided insight into choices for fuzzy classification approaches, and whether or not parameters such as class prior probabilities and covariance matrix 'size' need be included in the analysis. Furthermore, the traditional classified images were required for traditional Post Classification Comparison, to allow a comparison with fuzzy Post Classification Comparison.

Various filtering operations were examined as a means of reducing classification

'error' attributable to misregistration between images and the 'ground truth' data. It was found that morphological "opening" filters worked well, removing boundary region error while having no appreciable impact on larger homogeneous 'error' regions. An examination of these latter regions revealed that many were attributable to errors in the 'ground truth' data (or, at least, temporal misalignment between the image and the 'ground truth' data).

It was appreciated that the nature and cause of classification errors were highly important issues to Post Classification Comparison change detection. Accordingly, the nature of the logical process underlying classification was examined. It was suggested that this is essentially abductive inference, seeking a plausible explanation (i.e. membership of some particular cover class) for the observed spectra. This form of inference is not, of course, logically 'valid', and there may be other plausible explanations for the observed spectra. For our purposes, the most important of these is the possible presence of mixed pixels.

The use of spatial context was briefly discussed. It was noted that its use in classification gave promising results (see e.g. Whitbread, 1992; or Bouzerdoum *et al.*, 1996). It was, however, decided not to include work on spatial context in Post Classification Comparison in this thesis.

Chapter 5 introduced the notion and notation of fuzzy sets. Their application to image processing and remote sensing was briefly surveyed. A supervised fuzzy classifier drawn from the literature is described. This classifier is based on the unsupervised fuzzy c-means clustering algorithm, and computes fuzzy memberships from the reciprocals of the (squared) Mahalanobis Distances of each pixel from each of the training class means. A preliminary study on the use of this classifier for identifying general land use categories on Kangaroo Island was described. The

purpose of this study was to investigate the feasibility of the proposed approach, and to gain insight into choices for one of its key parameters. It is concluded that the Mahalanobis Distance fuzzy classifier offers a useful approach to land use classification. The TM89 and TM93 images are then classified using this approach. The accuracy assessment of fuzzy classified images is also discussed in this chapter.

Chapter 6 commences the detailed investigation of the key issue of this thesis: change detection using Post Classification Comparison. It introduces the notion that traditional Post Classification Comparison (PCC) change detection can be posed as a problem in Boolean logic, and shows some traditional PCC change detection results for the study region. It then introduces and discusses two candidate approaches to fuzzy PCC change detection: simple arithmetic operations and a fuzzy logic approach. The fuzzy logic approach requires the development of theory drawn from the formal fuzzy logic literature. Some results from each approach are shown. It is concluded that the fuzzy logic approach more satisfying and functional, and produces superior results to traditional Post Classification Comparison.

In Chapter 7, it is noted that the Mahalanobis Distance supervised fuzzy classifier requires the *a priori* selection of two important parameters: the number of classes, and a fuzzy weighting parameter, or fuzzy exponent, m. The number of classes is essentially determined by the problem and the data: this is not the case with the fuzzy weighting parameter. Earlier literature had noted that there is no theoretical and only limited empirical guidance on the selection of m. Other research had noted a strong correlation between fuzzy memberships and proportions of mixed classes. This provided our approach to investigating suitable values for the fuzzy exponent m. We imposed the condition that the fuzzy memberships exhibit a high membership of the appropriate single class in the case of pure, typical pixels, and reflect the proportions

of contributing classes for the two class mixed pixel case. Empirical results are presented which show a conflict between the requirements of these separate cases: the former suggests a value of m of approximately 2 or greater, whereas the latter indicates a value of approximately 0.5 or less. Earlier literature had expressed some intuitions and findings regarding the suggested range of values of the fuzzy exponent (using a slightly different arithmetic form). It is suggested that the results presented in this thesis provide a physical interpretation for these earlier intuitions and findings.

A theoretical investigation of the value of m is then shown, again subject to the condition that the fuzzy memberships reflect the proportions of contributing classes for the two class mixed pixel case. It is shown that the theoretically optimum value for m is a function of both the ratio of the class distance metrics (in this work, the Mahalanobis norm), and the ratio of proportions of contributing classes. It is noted that the ratio of proportions is actually what we seek to estimate with the fuzzy classification approach. A special case is shown to occur if the ratio of class distance metrics is unity (as is the case, for example, with hyperspherical class distributions): in this case, $m = 0.5$ is shown to give the required result, irrespective of the class mixture proportions.

The sensitivity of the Mahalanobis Distance fuzzy classifier to the selection of particular values of m is investigated and shown for a range of distance and proportion ratios. In general, it may be stated that the classifier error (as measured by the difference between the assigned memberships, and the 'true' proportions in a synthetic mixed pixel constructed under the linear mixing assumption) increases with increasing m.

Chapter 8 examines the notion of determining fuzzy memberships from the posterior probabilities determined by a Maximum Likelihood classifier. It is shown

that such a classifier is 'hard', and unsuited to our purposes. Various measures of fuzziness of this classifier, and the Mahalanobis Distance supervised fuzzy classifier, are shown. Two approaches to 'fuzzifying' the output of the Maximum Likelihood classifier are then examined. These involve either: directly scaling the covariance matrices of each training class, or (effectively) raising each class probability density function to an exponent between 0 and 1. It is shown that these two approaches are identical if the terms relating to prior class probabilities and class covariance matrix determinants are omitted.

The relationship between these approaches and the Mahalanobis Distance supervised fuzzy classifier is also examined, both empirically and analytically. It is shown that the major parameter of these approaches can be chosen to give highly similar (empirical) results to those of the Mahalanobis Distance supervised fuzzy classifier, for any particular chosen pair of classes: pixels that were largely assigned to one class or the other, or had their memberships distributed largely between these two classes, would show highly similar results. All other pixels would be more or less fuzzy, dependent on the ratio of the class Mahalanobis Distances of each class mean from the other. In the absence of 'fuzzy' ground truth data, it is not possible to draw quantitative conclusions about the relative accuracies of the approaches. It is, however, noted that the Mahalanobis Distance approach is simple to implement, and computationally efficient.

9.2 Future Work

9.2.1 Quantitative Comparisons of Change Detection Techniques

In the investigation of fuzzy classification algorithms and their application to change detection, this thesis contains, in general, only qualitative analysis of 'real' images, or synthetic, hypothesised pixels and structures. There remains a need to quantitatively investigate the relative accuracies of classification and change detection using fuzzy approaches with real data. As noted in the literature (and cited in this thesis), such data needs to be of a particular form (i.e. inherently 'fuzzy'). It has not historically been collected, and was not available for this study.

In light of its current lack of availability, the use of real data requires the collection of appropriate (i.e. 'fuzzy') data, a period of waiting (for change to occur), and then a further collection of appropriate data. An expedient interim solution might be to undertake comparisons using synthetic data (but it is emphasised that this is seen as an interim solution only). It is considered desirable that such a synthetic dataset is populated by 'real' pixel values (i.e. values that were actually recorded by an imaging system).

Furthermore, the effect of imperfect registration on change detection accuracies has been discussed, and is believed to be imperfectly understood. We would like to undertake our studies with perfectly registered images in the first instance, and then examine the effect of misregistration by 'moving' one image relative to the other.

A proposed approach to the generation of a synthetic dataset that uses 'real' pixel values (and hence avoids assumptions about the nature of data distributions), provides 'ground truth' for two (or more) dates, and ensures perfect registration and

'temporal alignment' between datasets, is as follows:

1. Select an image of a date for which reasonably reliable coincident ground truth data is held. Denote this as image A.
2. Acquire a second image of the same area, but of a different date. Denote this as image B.
3. Classify or segment image A. Denote this labeled image A'.
4. Identify regions of image B known or reasonably believed to correspond to the ground cover classes in the segmented image A. Identify as many such regions as possible. Collect (a large number of) raw pixel values for each class.
5. Create a new labeled image A" by making some (class label) changes to the labeled image A'. These changes should desirably be in regions of at least several pixels by several pixels.
6. Synthesise an image B' (the same size and coordinate locations as A) by writing each pixel value by a random selection from the collections of pixel values following step 4 above, with the collection to select from being identified by the class label for the pixel in the changed labeled image A".

Following this process, we will have two perfectly registered images, A and B'. The pixel values contained in each will be values actually recorded by an imaging device, on separate dates. Ground truth data will exist for both (we originally held it for image A, and B' is formed from the changed labeled image A"). We will, of course, know in advance what changes we should hope to find.

An important characteristic of this approach is the potential to have perfect registration and 'temporal alignment'. Approaches and algorithms could then be compared without the influence of errors from these causes. Using the reasonably sound and widely accepted assumption about linearity in spectral mixing, it would

be possible, with such a synthetic dataset, to investigate the sensitivity of change detection techniques to (increasingly) imperfect registration, and the quantitative effect of (the inevitable) misregistration on change detection results.

We must, of course, carefully consider to what extent the results of an investigation with this synthetic data could be carried forward to the real world situation of imperfectly registered images.

9.2.2 Investigation of Fuzzy Operators for Change Detection

In this thesis, the fuzzy logic operators for conjunction, disjunction and negation (AND, OR and NOT) were implemented as MIN, MAX, and 1-μ respectively. As noted in the body of the thesis, there were numerous other operators that could have been used. Some appear, *prima facie*, to offer advantages over those actually selected, certainly as far as the disjunction operator is concerned. It was, however, noted that the desire to be able to use the standard rules for simplifying logical expressions leads to a requirement to select operators as duals under De Morgan's Laws. This not-with-standing, it seems appropriate to investigate the choice of fuzzy logic operators for Fuzzy Post Classification Comparison change detection. In view of the uncertainties from a theoretical viewpoint, it is suggested that this matter be best investigated with quantitative, 'real' data. With such data, various operators for Fuzzy Post Classification Comparison could be meaningfully compared. The suggested approach is to define the operators in a parameterised form (e.g. the Yager class), and investigate the accuracy versus the parameter value.

9.2.3 The Fusion of Change Detection Techniques

An approach to improving change detection results is to combine, or 'fuse', the results of a number of different change detection techniques, in the hope that such a combined result would be superior, in some manner, to any of the individual results. This approach is analogous to applications in 'data fusion', in which the output of a number of individual sensors (say, tracks of aircraft) are combined to produce a result of higher accuracy, and possibly with more attributes, than possible with any single sensor. There are a number of approaches used in this domain, most notably Kalman filtering, Bayes Rule, Dempster-Shafer's Method, and fuzzy logic. Abdulghafour (1992) considers the application of the last three approaches to 'fusing' image segmentation results, and Deer and Eklund (1996) provides a theoretical discussion of the use of fuzzy logic in combining image features and results from traditional change detection algorithms.

9.2.4 Incorporation of Spatial Features

The research reported in this thesis implements change detection by comparison of fuzzy classification 'labels'. Despite promising research showing that improved classification accuracy can be achieved through incorporation of information on spatial context, operational classification of remotely sensed imagery generally uses spectral features only. There has, to our knowledge, been no use of spatial context for fuzzy classification. The approach to fuzzy classification adopted in this thesis is essentially to determine fuzzy memberships based on a measure of similarity, or distance, of each pixel from each of a number of pre-determined class means, using spectral values only. The distances are 'weighted' by the class covariances.

It is possible to determine the similarity of the 'texture' of a neighbourhood of a pixel to the 'texture' of a number of reference neighbourhoods. If these reference neighbourhoods represented classes of interest, it is easy to see how a similar approach to that of the Mahalanobis Distance fuzzy classifier could be devised, but now using only class covariances, not class means, and pixel neighbourhood textures, not pixel spectral values.

Denote the covariance matrix of class i by Σ_i and the covariance matrix of some particular neighbourhood of the pixel x_j by Λ_j. Define a function Δ_{ij}^2 representing the 'distance' (where close 'distance' connotes high similarity) between Λ_j and Σ_i. As before, calculate memberships in each of a number of fuzzy classes by

$$\nu_{ik} = \frac{\left(\frac{1}{\Delta_{ik}^2}\right)^{\beta}}{\sum_{j=1}^{c}\left(\frac{1}{\Delta_{jk}^2}\right)^{\beta}}, i = 1, \ldots, c; k = 1, \ldots, n \tag{9.1}$$

There are numerous candidates for the function Δ_{ij}^2. The work on speaker identification and analysis by Bogner (1981), as refined for remote sensing classification by Whitbread (1992), is of particular interest.

However, the literature shows that better results can be achieved by using both spectral and spatial information, rather than using either alone. The following three candidate approaches to using both in fuzzy classification appear, *prima facie*, to have some merit:

1. Augment the spectral vector for each pixel with one or more texture features (a number can be found in Haralick and Shanmugan, 1974, and Haralick, 1979), then undertake a fuzzy classification of the augmented vectors.
2. Compute fuzzy memberships of the form

$$\mu_{ik} = \frac{\left(\frac{1}{d_{ik}^2}\right)^m + \left(\frac{1}{\Delta_{ik}^2}\right)^\beta}{\sum_{j=1}^{c}\left(\left(\frac{1}{d_{jk}^2}\right)^m + \left(\frac{1}{\Delta_{jk}^2}\right)^\beta\right)}, i = 1, \ldots, c; k = 1, \ldots, n \qquad (9.2)$$

where, as before, d_{ij}^2 and Δ_{ij}^2 represent spectral distance from prototypical class means and 'distance' from prototypical class 'textures' respectively. We note that we might more generally define fuzzy memberships of the form

$$\mu_{ik} = \frac{g\left(d_{ik}^2\right) + h\left(\Delta_{ik}^2\right)}{\sum_{j=1}^{c}\left(g\left(d_{jk}^2\right) + h\left(\Delta_{jk}^2\right)\right)}, i = 1, \ldots, c; k = 1, \ldots, n \qquad (9.3)$$

where g(.) and h(.) are functions that 'invert' the distances d_{ij}^2 and Δ_{ij}^2 in some manner.

3. Calculate the fuzzy memberships defined by Equations (5.18) and (9.1) independently, then combine the results in some way using, for example, a fuzzy logic theoretic approach.

Each of these candidate approaches also appears to have some disadvantages from a theoretical viewpoint. Their implementation and analysis would require some effort and the investigation of a number of issues: this is left to future work.

REFERENCES

Abdulghafour M.B. (1992). *Data Fusion Through Fuzzy Reasoning Applied to Feature Extraction from Multi-Sensory Images.* Unpublished PhD dissertation, University of Tennessee, Knoxville.

Adams J.B., D.E. Sabol, V. Kapos, R.A. Filho, D.A. Roberts, M.O. Smith and A.R. Gillespie (1995). Classification of Multispectral Images Based on Fractions of Endmembers: Application to Land-Cover Change in the Brazilian Amazon. *Remote Sensing of Environment*, 52, pp. 137-154.

Ahern F.J. (1985). *Radiometric Calibration and Correction of Landsat 1, 2 and 3 MSS Data.* Canadian Centre for Remote Sensing Research Report 78-4.

Al-Sultan K.S. and S.Z. Selim (1993). A Global Algorithm for the Fuzzy Clustering Problem. *Pattern Recognition*, 26, pp. 1357-1361.

Allison D., M.J.A. Zemerly and J-P. Muller (1991). Automatic Seed Point Generation for Stereo Matching and Multi-Image Registration. *Proceedings of the 1991 International Geoscience and Remote Sensing Symposium (IGARSS 91)*, pp. 2417-2421.

Anderson J.R., E.E. Hardy, J.T. Roach and R.E. Witmer (1976). *A Land Use and Land Cover Classification System for use with Remote Sensor Data.* U.S.G.S. Paper 964.

Atkinson P.M., M.E.J. Cutler and H. Lewis (1997). Mapping Sub-Pixel Proportional Land Cover With AVHRR Imagery. *International Journal of Remote Sensing*, 18, pp. 917-935.

Ball G.H. and D.J. Hall (1967). A Clustering Technique for Summarizing Multi-Variate Data. *Behavioural Science*, 12, pp. 153-155.

Banner A.V. and T. Lynham (1981). Multitemporal Analysis of Landsat Data for Forest Cut Over Mapping - a Trial of Two Procedures. *Proceedings of the 7th Canadian Symposium on Remote Sensing*, Winnipeg, pp. 233-240.

Bartl R., M. Petrou, W.J. Christmas and P.L. Palmer (1996). On the Automatic Registration of Cadastral Maps and Landsat TM Images. *SPIE Vol. 2955*, pp. 9-20.

Bellman R. and M. Giertz (1973). On the Analytic Formalism of the Theory of Fuzzy Sets. *Information Sciences*, 5, pp. 149-156.

Besag J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society*, Series B 40, pp. 192-236.

Bezdek J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum.

Bezdek J.C. (1994). Fuzziness vs. Probability - Again (!?). *IEEE Transactions on Fuzzy Systems*, 2, pp. 1-3.

Bezdek J.C. and S.K. Pal (eds.) (1992). *Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data.* IEEE Press.

Bezdek J.C., R. Ehrlich and W. Full (1984). FCM: The Fuzzy c-Means Clustering Algorithm. *Computers and Geoscience*, 10, pp. 191 - 203.

Blonda P.N., G. Pasquariello, S. Losito, A. Mori, F. Posa and D. Ragno (1991). An Experiment for the Interpretation of Multitemporal Remotely Sensed Images Based

on a Fuzzy Logic Approach. *International Journal of Remote Sensing*, 12, pp. 463-476.

Bogner R.E. (1981). Pattern Recognition via Observation Correlations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3, pp. 128-133.

Bonissone P.P. and K.S. Decker (1986). Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-Off Precision and Complexity. *Uncertainty in Artificial Intelligence*, Kanal L.N. and J.F. Lemmer (eds.), North-Holland, pp. 217-247.

Borne F. (1994). New Features for Texture Analysis. *Proceedings of the 1994 International Geoscience and Remote Sensing Symposium (IGARSS 94)*, pp. 1160-1162.

Bouzerdoum A., R.J. Cole and P. Deer (1996). Classification of Satellite Imagery Based on Texture Features Using Neural Networks. *Proceedings of the 4th International Conference on Control, Automation, Robotics and Vision (ICARCV'96)*, Singapore, pp. 2257-2261.

Brera A.M. and F. Shahrokhi (1978). Application of Landsat Data to Monitor Desert Spreading in the Sahara Region. *Proceedings of the 12th International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan.

Brivio P.A., A.D. Ventura, A. Rampani and R. Schettini (1992). Automatic Selection of Control Points from Shadow Structures. *International Journal of Remote Sensing*, 13, pp. 1853-1860.

Brown L.G. (1992). A Survey of Image Registration Techniques. *Computing Surveys*, 24, pp. 325-376.

Burrough P.A. (1989). Fuzzy Mathematical Methods for Soil Survey and Land Evaluation. *Journal of Soil Science*, 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*, 10, pp. 175-184.

Campbell N.A. (1984). Mixture Models and Atypical Values. *Mathematical Geology*, 16, pp. 465-477.

Campbell J.B. (1987). *Introduction to Remote Sensing.* Guilford Press, New York.

Campbell J.G. and A.A. Hashim (1992). Fuzzy Sets, Pattern Recognition, Linear Estimation, and Neural Networks - a Unification of the Theory with Relevance to Remote Sensing. *Proceedings of the 18th Annual Conference of the Remote Sensing Society.* Cracknell A.P. and R.A. Vaughan (eds.), University of Dundee, pp. 508-517.

Cannon R.L., J.V. Dave and J.C. Bezdek (1986). Efficient Implementation of the Fuzzy *c*-Means Clustering Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, pp. 248-255.

Cannon R.L. and C. Jacobs (1984). *Multispectral Pixel Classification with Fuzzy Objective Functions.* Center for Automation Research, Technical Report CAR-TR-51, University of Maryland, College Park.

Charniak E. and D. McDermott (1985). *Introduction to Artificial Intelligence.* Addison-Wesley, Reading, Massachusetts.

Chellapa R. (1997). Model Supported Exploitation of Aerial Images. *SPIE Vol. 3024*, pp. 766-777.

Chi Z., H. Yan and T. Pham (1996). *Fuzzy Algorithms.* World Scientific.

Choe H. and J.B. Jordan (1992). On the Optimal Choice of Parameters in a Fuzzy *c*-Means Algorithm. *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 349-354.

Choo A., B. Pham, and A.J. Maeder (1989). Change Detection in an Image Sequence Using Shape Analysis. *Proceedings of the Conference on Image Processing and the Impact of New Technologies*, Canberra, pp. 111-114.

Clement V.and M. Thonnet (1993). A Knowledge-Based Approach to Integration of Image Processing Procedures. *CVGIP: Image Understanding*, 57, pp. 166-184.

Clifford P., S. Richardson and D. Hemon (1989). Assessing the Significance of the Correlation Between Two Spatial Processes. *Biometrics*, 45, pp. 123-134.

Coiner J.C. (1980). Using Landsat to Monitor Changes in Vegetation Cover Induced by Desertification Processes. *Proceedings of the 14th International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan.

Collins J.B. and C.E. Woodcock (1994). Change Detection Using the Gramm-Schmidt Transformation Applied to Mapping Forest Mortality. *Remote Sensing of Environment*, 50, pp. 267-279.

Colwell J. E. and F.P. Weber (1981). Forest Change Detection. *Proceedings of the 15th International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, pp. 839-852.

Congalton R., R. Oderwald and R. Mead (1983). Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques. *Photogrammetric Engineering and Remote Sensing*, 49, pp. 1671-1678.

Congalton R.G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, 37, pp. 35-46.

Cross A.M., J. Settle, N.A. Drake and R.T.M. Paivinen (1991). Sub-Pixel Measurement of Tropical Forest Cover Using AVHRR Data. *International Journal of Remote Sensing*, 12, pp. 1119-1129.

Cross V. and T. Sudcamp (1991). An Empirical Study of Fuzzy Compatibility Measures and Aggregation Operators. *SPIE Vol. 1607*, pp. 415-428.

Dai X. and S. Khorram (1997a). Development of New Automated Land Cover Change Detection System from Remotely Sensed Imagery Based on Artificial Neural Networks. *Proceedings of the 1997 International Geoscience and Remote Sensing Symposium (IGARSS 97)*, pp. 1029-1031.

Dai X. and S. Khorram (1997b). Quantification of the Impact of Misregistration on the Accuracy of Remotely Sensed Change Detection. *Proceedings of the 1997 International Geoscience and Remote Sensing Symposium (IGARSS 97)*, pp. 1763-1765.

Deer P.J. (1995). Digital Change Detection Techniques: Civilian and Military Applications. *Proceedings of the International Symposium for Spectral Sensing Research (ISSSR'95)*, Melbourne, Australia.

Deer P.J. and M.E. Longmore (1994). The Application of Principal Components Analysis to Monitoring the Clearance of Native Forest Stands on Kangaroo Island, South Australia. *Proceedings of the 7th Australasian Remote Sensing Conference*, Melbourne, Australia.

Deer P.J. and P.W. Eklund (1996). On the Fusion of Image Features. *Proceedings of the International Discourse on Fuzzy Logic and the Management of Complexity (FLAMOC'96)*, Sydney, Australia, pp. 283-287.

Deer P.J., P.W. Eklund and B. Norman (1996a). An Application of Fuzzy Set Theory to Image Processing and Remote Sensing. *Proceedings of the 8th Australasian Remote Sensing Conference*, Canberra, Australia.

Deer P.J., P.W. Eklund and B. Norman (1996b). A Mahalanobis Distance Fuzzy Classifier. *Proceedings of the 4th Australian and New Zealand Intelligent Information*

Systems Conference (ANZIIS 96), Adelaide, Australia, pp. 220-223.

Dreschler-Fischer L., C. Drewniok, H. Lange and C. Schroder (1993). A Knowledge-Based Approach to the Detection and Interpretation of Changes in Aerial Images. *Proceedings of the 1993 International Geoscience and Remote Sensing Symposium (IGARSS 93)*, pp. 159-161.

Duane G. (1988). Pixel-Level Sensor Fusion for Improved Object Recognition. *SPIE Vol. 931*, pp. 180-185.

Dubois D. and H. Prade (1985). A Review of Fuzzy Set Aggregation Connectives. *Information Sciences*, 36, pp. 85-121.

Dubois D. and H. Prade (1993). Fuzzy Sets and Probability: Misunderstandings, Bridges and Gaps. *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 1059-1068.

Duda R.O. and P.E. Hart (1973). *Pattern Classification and Scene Analysis.* Wiley.

Dunne R.A. and N.A. Campbell (1995). Neighbour-Based MLPs. *Proceedings of the IEEE International Conference on Neural Networks*, pp. 270-274.

Eghbali H. J. (1979). K-S Test for Detecting Changes from Landsat Imagery Data. *IEEE Transactions on Systems, Man and Cybernetics*, 9, pp. 17-23.

El-Raey M., S.M. Nasr and M.M. El-Hattab (1995). Change Detection of Rosetta Promontory Over the Last Forty Years. *International Journal of Remote Sensing, 16*, pp. 825-834.

Elvidge C.D. and Z. Chen (1995). Comparison of Broad-Band and Narrow-Band Red Versus Near Infrared Vegetation Indices. *Remote Sensing of Environment*, 54, pp. 38-48.

Fann K.T. (1970). *Peirce's Theory of Abduction.* Martinus Nijhoff, The Hague, Holland.

Fisher P.F. and S. Pathirana (1990). The Evaluation of Fuzzy Membership of Land Cover Classes in the Suburban Zone. *Remote Sensing of Environment*, 34, pp. 121-132.

Fisher P.F. and S. Pathirana (1993). The Ordering of Fuzzy Land-Cover Information Derived from Landsat MSS Data. *Geocarto International*, 3, pp. 5-14.

Fitzgerald R.W. and B.G. Lees (1994). Assessing the Classification Accuracy of Multisource Remote Sensing Data. *Remote Sensing of Environment*, 47, pp. 362-368.

Flemons P. (1996). Monitoring Vegetation Change in NSW: - A Comparison of Change Detection Techniques. *Proceedings of the 8th Australasian Remote Sensing Conference*, Canberra, Australia, pp. 270-278.

Fonseca L.M.G. and B.S. Manjunath (1996). Registration Techniques for Multisensor Remotely Sensed Imagery. *Photogrammetric Engineering and Remote Sensing*, 62, pp. 1049-1056.

Foody G.M. (1992). A Fuzzy Sets Approach to the Representation of Vegetation Continua from Remotely Sensed Data: An Example from Lowland Heath. *Photogrammetric Engineering and Remote Sensing*, 58, pp. 221 - 225.

Foody G.M. (1994). Ordinal-Level Classification of Sub-Pixel Tropical Forest Cover. *Photogrammetric Engineering and Remote Sensing*, 60, pp. 61-65.

Foody G.M. (1995). Cross-entropy for the Evaluation of the Accuracy of a Fuzzy Land Cover Classification with Fuzzy Ground Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50, pp. 2-12.

Foody G.M. (1996). Approaches for the Production and Evaluation of Fuzzy Land Cover Classifications from Remotely-Sensed Data. *International Journal of Remote Sensing*, 17, pp. 1317-1340.

Foody G.M., N.A. Campbell, N.M. Trodd and T.F. Wood (1992). Derivation and Applications of Probabilistic Measures of Class Membership from the Maximum-Likelihood Classification. *Photogrammetric Engineering and Remote Sensing*, 58, pp. 1335-1341.

Foody G.M. and D.P. Cox (1994). Sub-Pixel Land Cover Composition Estimation Using a Linear Mixture Model and Fuzzy Set Membership. *International Journal of Remote Sensing*, 15, pp. 619-631.

Foody G.M. and N.M. Trodd (1993). Non-Classificatory Analysis and Representation of Heathland Vegetation from Remotely Sensed Imagery. *GeoJournal*, 29, pp. 343-350.

Ford G.E. and C.I. Zanelli (1985). Analysis and Quantification of Errors in the Geometric Correction of Satellite Images. *Photogrammetric Engineering and Remote Sensing*, 51, pp. 1725-1734.

Fung L.W. and K.S. Fu (1975). An Axiomatic Approach to Rational Decision-Making in a Fuzzy Environment. *Fuzzy Sets and their Applications to Cognitive and Decision Processes.* Zadeh L.A., K.S. Fu, K. Tanaka and M. Shimura (eds.). Academic, New York, pp. 227-256.

Fung T. and E. LeDrew (1987). The Application of Principal Component Analysis to Change Detection Using Various Accuracy Indices. *Photogrammetric Engineering and Remote Sensing*, 53, pp. 1649-1658.

Fung T. and E. LeDrew (1988). The Determination of Optimum Threshold Levels for Change Detection. *Photogrammetric Engineering and Remote Sensing*, 54, pp. 1449-1454.

Gaines B.R. (1976). Foundations of Fuzzy Reasoning. *International Journal of Man-Machine Studies*, 8, pp. 623-668.

Gath I. and A.B. Geva (1989). Unsupervised Optimal Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, pp. 773-781.

Gong P., E.F. LeDrew and J.R. Miller (1992). Registration-Noise Reduction in Difference Images for Change Detection. *International Journal of Remote Sensing*, 13, pp. 773-779.

Gopal S. and C. Woodcock (1994). Theory and Methods for Accuracy Assessment of Thematic Maps using Fuzzy Sets. *Photogrammetric Engineering and Remote Sensing*, 60, pp. 181-188.

Gordon S. (1980). Utilizing Landsat Imagery to Monitor Land Use Change. *Remote Sensing of Environment*, 9, pp. 189-196.

Graetz R.D. and M.R. Gentle (1982). A Study of the Relationships Between Reflectance Characteristics in the Landsat Wavebands and the Composition and Structure of an Australian Semi-arid Shrubland. *Photogrammetric Engineering and Remote Sensing*, 48, pp. 1721-1730.

Guoling T. (1989). *Spectral Signatures and Vegetation Indices of Crops.* Divisional Report 89/4, CSIRO Division of Water Resources, Canberra.

Hanaizumi H., H. Okumura and S. Fujimura (1991). Change Detection from Remotely Sensed Multi-Temporal Images using Spatial Segmentation. *Proceedings of the 1991 International Geoscience and Remote Sensing Symposium (IGARSS 91)*, pp. 1079-1081.

Hanaizumi H. and Y. Kanemoto (1996). Automated Registration Method with Disparity Detection for Remotely Sensed Multispectral Images. *SPIE Vol. 2955*, pp. 2-8.

Haralick R.M. (1979). Statistical and Structural Approaches to Texture. *Proceedings of the IEEE*, 67, pp. 786-804.

Haralick R.M. and K.S. Shanmugan (1974). Combined Spectral and Spatial Processing of ERTS Imagery Data. *Remote Sensing of Environment*, 3, pp. 3-13.

Haralick R.M., S.R. Sternberg and X. Zhuang (1987). Image Analysis Using Mathematical Morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9, pp. 532-550.

Harrison B.A. and D.L.B. Jupp (1989). *Introduction to Remotely Sensed Data.* MicroBRIAN Resource Manual, Part 1, CSIRO.

Hoffer R.M. and K.S. Lee (1989). Forest Change Classification Using SEASAT and SIR-B Satellite SAR Data. *Proceedings of the 1989 International Geoscience and Remote Sensing Symposium (IGARSS 89)*, pp. 1372-1375.

Holm M. (1991). Towards Automatic Rectification of Satellite Images Using Feature Based Matching. *Proceedings of the 1991 International Geoscience and Remote Sensing Symposium (IGARSS 91)*, pp. 2439-2442.

Horler D.N.H. and F.J. Ahern (1986). Forestry Information Content of Thematic Mapper Data. *International Journal of Remote Sensing*, 7, pp. 405-428.

Howarth P.J. and G. M. Wickware (1981). Procedures for Change Detection Using Landsat Digital Data. *International Journal of Remote Sensing*, 2, pp. 277-291.

Hubert L.J. and R.G. Golledge (1982). Measuring Association Between Spatially Defined Variables: Tjostheim's Index and Some Extensions. *Geographical Analysis*, 14, pp. 273-278.

Hudson W.D. and C.W. Ramm (1987). Correct Formulation of the Kappa Coefficient of Agreement. *Photogrammetric Engineering and Remote Sensing*, 53, pp. 421-422.

Ingebritsen S.E. and R.P.J. Lyon (1985). Principal Component Analysis of Multitemporal Image Pairs. *International Journal of Remote Sensing*, 12, pp. 687-696.

Izraelivitz D. and J.A. Cochand (1990). Multisource Fusion for Target Detection. *SPIE Vol. 1301*, pp. 58-67.

Jensen J.R. (1986). *Introductory Digital Image Processing: A Remote Sensing Perspective.* Prentice-Hall.

Jensen J.R., E.W. Ramsay, H.E. Mackey, E.J. Christensen and R.R. Sharitz (1987). Inland Wetland Change Detection Using Aircraft MSS Data. *Photogrammetric Engineering and Remote Sensing*, 53, pp. 521-529.

Jensen J.R., K. Rutchey, M.S. Koch and S. Narumalani (1995). Inland Wetland Change Detection in the Everglades Water Conservation Area 2A Using a Time Series of Normalized Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing*, 61, pp.199-209.

Jha C.S. and N.V.M. Unni (1994). Digital Change Detection of Forest Conversion of a Dry Tropical Forest Region. *International Journal of Remote Sensing*, 15, pp. 2543-2552.

Jiaju L. (1988). Development of Principal Components Analysis Applied to Multitemporal Landsat TM Data. *International Journal of Remote Sensing*, 9, pp. 1895-1907.

Josephson J.R. and S.G. Josephson (1994). *Abductive Inference.* Cambridge University Press.

Joyce A.T., J.H. Ivey and G.S. Burns (1980). The Use of Landsat MSS Data for Detecting Land Use Changes in Forestland. *Proceedings of the 14th International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, pp. 979-988.

Kent J.T. and K.V. Mardia (1988). Spatial Classification Using Fuzzy Membership Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10, pp. 659-671.

Key J.R., J.A. Maslanik and R.G. Barry (1989). Cloud Classification From Satellite Data Using a Fuzzy Sets Algorithm: A Polar Example. *International Journal of Remote Sensing*, 10, pp. 1823-1842.

Klir G.J. and T.A Folger (1988). *Fuzzy Sets, Uncertainty, and Information.* Prentice Hall.

Kosko B. (1986). Fuzzy Entropy and Conditioning. *Information Sciences*, 40, pp. 165-174.

Lambin E.F. and A.H. Strahler (1994). Change Vector Analysis in Multitemporal Space: a Tool to Detect and Categorise Land-Cover Change Processes Using High Temporal Resolution Satellite Data. *Remote Sensing of Environment*, 48, pp. 231-244.

Lazaroff M.B.and M.W. Brennan (1992). Multi-Temporal Texture Analysis of Features Computed from Remotely Sensed Imagery. *SPIE Vol. 1819*, pp. 166-175.

Lee A.J., N.H. Carender, D.J. Knowlton, D.M. Bell and J.K. Bryan (1993). Fast Autonomous Registration of Landsat, SPOT, and Digital Map Imagery. *SPIE Vol. 1944*, pp. 68-79.

Lillesand T.M.and R.W. Kiefer (1987). *Remote Sensing and Image Interpretation.* Wiley.

Lodwick G.D. (1979). Measuring Ecological Changes in Multitemporal Landsat Data Using Principal Components. *Proceedings of the 13th International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, pp. 1131-1141.

Lyon J.G., D. Yuan, R.S. Lunetta and C.D. Elvidge (1998). A Change Detection Experiment Using Vegetation Indices. *Photogrammetric Engineering and Remote Sensing*, 64, pp. 143-150.

McBratney A.B. and A.W. Moore (1985). Application of Fuzzy Sets to Climatic Classification. *Agricultural and Forest Meteorology*, 35, pp. 165-185.

Malila W.A. (1980). Change Vector Analysis: An Approach for Detecting Forest Change with Landsat. *Proceedings of the Annual Symposium on Machine Processing of Remotely Sensed Data*, IEEE, pp. 326-336.

Manavalan P., K. Kesavasamy and S. Adiga (1995). Irrigated Crops Monitoring Through Seasons Using Digital Change Detection Analysis of IRD-LISS 2 Data. *International Journal of Remote Sensing*, 16, pp. 633-640.

Maragos P. and R.W. Schafer (1990). Morphological Systems for Multidimensional Signal Processing. *Proceedings of the IEEE*, 78, pp. 690-710.

Marsh S.E., P. Switzer, W.S. Kowalik and R.P.J. Lyon (1980). Resolving the Percentage of Component Terrains Within Single Resolution Elements. *Photogrammetric Engineering and Remote Sensing*, 46, pp. 1079-1086.

Martin L.R.G. (1989). Accuracy Assessment of Landsat-Based Visual Change Detection Methods Applied to the Rural-Urban Fringe. *Photogrammetric Engineering and Remote Sensing*, 55, pp. 209-215.

Maselli F., A. Rodolfi and C. Conese (1996). Fuzzy Classification of Spatially Degraded Thematic Mapper Data for Estimation of Sub-Pixel Components. *International Journal of Remote Sensing*, 17, pp. 537-551.

Mather P.M. (1995). Map-Image Registration Accuracy using Least-Squares Polynomials. *International Journal of Geographic Information Systems*, 9, pp. 543-554.

Matheron G. (1975). *Random Sets and Integral Geometry.* Wiley, New York.

Matsuyama T. (1987). Knowledge-Based Aerial Image Understanding Systems and Expert Systems for Image Processing. *IEEE Transactions on Geoscience and Remote*

Sensing, 25, pp. 305-316.

Michener W.K. and P.F. Houhoulis (1997). Detection of Vegetation Changes Associated with Extensive Flooding in a Forested Ecosystem. *Photogrammetric Engineering and Remote Sensing*, 63, pp. 1363-1374.

Miller L.D., K. Nualchawee and C. Tom (1978). *Analysis of the Dynamics of Shifting Cultivation in the Tropical Forests of Northern Thailand Using Landscape Modelling and Classification of Landsat Imagery.* NASA Technical Memorandum 79545, Goddard Space Flight Center, Maryland.

Milne A.K. (1987). Land Cover Change Detection in Semi-Arid Environments Using Landsat Imagery. *Proceedings of the 4th Australasian Remote Sensing Conference, Adelaide, Australia*, pp. 225-238.

Milne A.K. (1988). Change Detection Analysis Using Landsat Imagery: A Review of Methodology. *Proceedings of the 1988 International Geoscience and Remote Sensing Symposium (IGARSS 88)*, pp. 541-544.

Muchoney D.M. and B.N. Haack (1994). Change Detection for Monitoring Forest Defoliation. *Photogrammetric Engineering and Remote Sensing*, 60, pp. 1243-1251.

Nelson R.F. (1983). Detecting Forest Canopy Change Due to Insect Activity Using Landsat MSS. *Photogrammetric Engineering and Remote Sensing*, 49, pp. 1303-1314.

Nezry E., E. Moughin, A. Lopes, J.P. Gastellu-Etchegorry and Y. Laumonier (1993). Tropical Vegetation Mapping with Combined Visible and SAR Spaceborne Data. *International Journal of Remote Sensing*, 14, pp. 2165-2184.

Nielsen A.A. (1996). Change Detection in Multispectral, Bi-Temporal Data Using Orthogonal Transformations. *Proceedings of the 8th Australasian Remote Sensing Conference*, Canberra, Australia, pp. 286-295.

Nishida M., K. Hayashi, N. Yoshimura and K. Ohtsuka (1993). Application of Fuzzy Reasoning to Estimation of Class Mixture Proportion of Mixed Pixel on Remote Sensing Images. *Proceedings of the 1993 International Geoscience and Remote Sensing Symposium (IGARSS 93)*, pp. 1965-1968.

Novak K. (1992). Rectification of Digital Imagery. *Photogrammetric Engineering and Remote Sensing*, 58, pp. 339-344.

Olsson H. (1995). Reflectance Calibration of Thematic Mapper Data for Forest Change Detection. *International Journal of Remote Sensing*, 16, pp. 81-96.

Orti F. (1981). Optimal Distribution of Ground Control Points to Minimize Landsat Image Registration Errors. *Photogrammetric Engineering and Remote Sensing*, 47, pp. 101-110.

Pal N.R. and J.C. Bezdek (1994). Measuring Fuzzy Uncertainty. *IEEE Transactions on Fuzzy Systems*, 2, pp. 107-118.

Pal S.K. (1992). Fuzzy Sets in Image Processing and Recognition. *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 119-126.

Palubinskas G., R.M. Lucas, G.M. Foody and P.J. Curran (1995). An Evaluation of Fuzzy and Texture-Based Classification Approaches For Mapping Regenerating Tropical Forest Classes From Landsat-TM Data. *International Journal of Remote Sensing*, 16, pp. 747-759.

Pathirana S. (1990). *Fuzzy Membership Approach to the Mixed Pixel Problem of Remotely Sensed Data: An Application to a Sub-Urban Fringe Zone of Northeast Ohio.* Unpublished PhD dissertation, Kent State University.

Pathirana S. (1993). Detection of Linear and Sub-Pixel Phenomena Using the Fuzzy Membership Approach. *Proceedings of the 6th Australasian Remote Sensing Conference*, Melbourne, Vol. 2, pp (2-)424 - (2-)433.

Paul G. (1993). Approaches to Abductive Reasoning: An Overview. *Artificial Intelligence Review*, 7, pp. 109-152

Pilon P.G., P.J. Howarth, R.A. Bullock and P.O. Adeniyi (1988). An Enhanced Classification Approach to Change Detection in Semi-Arid Environments. *Photogrammetric Engineering and Remote Sensing*, 54, pp. 1709-1716.

Pons X. and L. Sole-Sugranes (1994). A Simple Radiometric Correction Model to Improve Automatic Mapping of Vegetation from Multispectral Satellite Data. *Remote Sensing of Environment*, 48, pp. 191-204.

Pratt W.K. (1991). *Digital Image Processing* (second edition). Wiley.

Qi J., A.R. Huete, M.S. Moran, A. Chehbouni and R.D. Jackson (1993). Interpretation of Vegetation Indices Derived from Multi-Temporal SPOT Images. *Remote Sensing of Environment*, 44, pp. 89-101.

Quarmby N.A. and J.L. Cushnie (1989). Monitoring Urban Land Cover Changes at the Urban Fringe from SPOT HRV Imagery in South-East England. *International Journal of Remote Sensing*, 10, pp. 953-963.

Richards J.A. (1984). Thematic Mapping from Multitemporal Image Data Using the Principal Components Transformation. *Remote Sensing of Environment*, 16, pp. 35-46.

Richards J.A. (1986). *Remote Sensing Digital Image Analysis.* Springer-Verlag, Berlin.

Richards J.A. (1993). *Remote Sensing Digital Image Analysis* (second edition). Springer-Verlag.

Richards J.A. (1996). Classifier Performance and Map Accuracy. *Remote Sensing of Environment*, 57, pp. 161-166.

Richardus P. and R.K. Adler (1972). *Map Projections.* North-Holland.

Rignot E.J.M., R. Kowk, J.C. Curlander and S.S. Pang (1991). Automated Multisensor Registration: Requirements and Techniques. *Photogrammetric Engineering and Remote Sensing*, 57, pp. 1029-1038.

Riordan C.J. (1980). *Non-urban to Urban Land Cover Change Using Landsat Data.* Summary Report of the Colorado Agricultural Research Experiment Station. Fort Collins, Colorado.

Robinove C.J. (1982). Computation of Physical Values for Landsat Digital Data. *Photogrammetric Engineering and Remote Sensing*, 48, pp. 781-784.

Robinson J.W. (1979). *A Critical Review of the Change Detection and Urban Classification Literature.* Technical Memorandum CSC/TM-79/6235, Computer Sciences Corporation, Maryland, U.S.A.

Rosenfield G. (1981). Analysis of Variance of Thematic Mapping Experiment Data. *Photogrammetric Engineering and Remote Sensing*, 47, pp. 1685-1692.

Rosenfield G. and K. Fitzpatrick-Lins (1986). A Coefficient of Agreement as a Measure of Thematic Classification Accuracy. *Photogrammetric Engineering and Remote Sensing*, 52, pp. 223-227.

Schistad Solberg A.H., A.K. Jain and T. Taxt (1994). Multisource Classification of Remotely Sensed Data: Fusion of Landsat TM and SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 32, pp. 768-778.

Schlien S. (1979). Geometric Correction Registration and Resampling of Landsat Imagery. *Canadian Journal of Remote Sensing*, 5, pp. 74-87.

Schowengerdt R.A. (1996). On the Estimation of Spatial-Spectral Mixing with Classifier Likelihood Functions. *Pattern Recognition Letters*, 17, pp. 1379-1387.

Serra J. (1982). *Image Analysis and Mathematical Morphology.* Academic Press, London.

Serra J. and L. Vincent (1992). An Overview of Morphological Filtering. *IEEE Transactions on Circuits, Systems and Signal Processing*, 11, pp. 47-108.

Shannon C.E. (1948). A Mathematical Theory of Communication. *Bell Systems Technical Journal*, XXVII-3, pp. 379-423.

Singh A. (1986). Change Detection in the Tropical Forest Environment of Northeastern India Using Landsat. *Remote Sensing and Tropical Land Management*, (Eden M.J. and J.T. Parry eds.), John Wiley & Son, Chichester, pp. 237-254.

Singh A. (1989). Digital Change Detection Techniques Using Remotely-Sensed Data. *International Journal of Remote Sensing*, 10, pp. 989-1003.

Singh A. and A. Harrison (1985). Standardized Principal Components. *International Journal of Remote Sensing*, 6, pp. 883-896.

Sloggett D., C. Gurney, W. Newton, R. Gooding and I. Dowman (1994). An Automated Change Detection System. *Proceedings of the 1994 International Geoscience and Remote Sensing Symposium (IGARSS 94)*, pp. 875-877.

Soares V.P. and R.M. Hoffer (1994). Eucalyptus Forest Change Classification Using Multi-Date Landsat TM Data. *SPIE Vol. 2314*, pp. 281-291.

Sternberg S.R. (1986). Grayscale Morphology. *Computer Vision, Graphics, and Image Processing*, 35, pp. 333-355.

Stow D.A., D. Collins and D. McKinsey (1990). Land Use Change Detection Based on Multi-date Imagery from Different Satellite Sensor Systems. *Geocarto International*, 5, pp. 3-12.

Stow D.A., L.R. Tinney and J.E. Estes (1980). Deriving Land Use/Land Cover Change Statistics From Landsat: a Study of Prime Agricultural Land. *Proceedings of the 14th International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, pp. 1227-1237.

Student (Gosset W.S.) (1914). The Elimination of Spurious Correlation Due to Position in Time or Space. *Biometrika*, 10, pp. 179-181.

Sugeno M. (1977). Fuzzy Measures and Fuzzy Integrals: A Survey. *Fuzzy Automata and Decision Processes*, Gupta M.M., G.N. Sardis and B.R.Gaines (eds.), North-Holland, Amsterdam, pp. 89-102.

Swain P.H. and S.M. Davis (1978). *Remote Sensing: The Quantitative Approach.* McGraw-Hill.

Tanaka S., T. Sugimura and K. Kameda (1992). Spatial Resolution and Frequency of Satellite Data Acquisition for Multi-Temporal Analysis of Environment. *Advances in Space Research*, 12, pp. 333-432.

Thornber K.K. (1993). A Key to Fuzzy-Logic Inference. *International Journal of Approximate Reasoning*, 8, pp. 105-121.

Todd W.J. (1977). Urban and Regional Land Use Change Detected by Using Landsat Data. *Journal of Research by the United States Geological Survey*, 5, pp. 527-534.

Toll D.L., J.A. Royal and J.B. Davis (1980). Urban Area Up-Date Procedures Using Landsat Data. *Proceedings of the Fall Technical Meeting of the American Society of Photogrammetry.*

Tong R.M. and D.G. Shapiro (1985). Experimental Investigations of Uncertainty in a Rule-based System for Information Retrieval. *International Journal of Man-Machine Studies*, 22, pp. 265-282.

Tou J.T. and R.C. Gonzalez (1974). *Pattern Recognition and Principles.* Addison-Wesley.

Townshend J.R.G. (1981). Image Analysis and Interpretation for Land Resources Survey. *Terrain Analysis and Remote Sensing*, Townshend J.R.G. (ed.), Allen and Unwin, London.

Townshend J.R.G., C.O. Justice, C. Gurney and J. McManus (1992). The Impact of Misregistration on Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 30, pp. 1054-1060.

Tucker C.J. (1979). Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, 8, pp. 127-150.

van Deusen P.C. (1994). Correcting Bias in Change Estimates from Thematic Maps. *Remote Sensing of Environment*, 50, pp. 67-73.

van der Meer F. (1995). Spectral Unmixing of Landsat Thematic Mapper Data. *International Journal of Remote Sensing*, 16, pp. 3189-3194.

Virag L.A. and J.E. Colwell (1987). An Improved Procedure for Analysis of Change in Thematic Mapper Image-Pairs. *Proceedings of the 21st International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, pp. 1101-1110.

Vogelmann J.E. (1988). Detection of Forest Change in the Green Mountains of Vermont Using Multispectral Scanner Data. *International Journal of Remote Sensing*, 9, pp. 1187-1200.

Wang D., J. Ronsin and V. Haese-Coat (1992). Compared Performances of Morphological, Median Type and Running Mean Filters. *Visual Communications and Image Processing '92, SPIE Vol. 1818*, pp. 384-391.

Wang F. (1989). A Fuzzy Expert System For Remote Sensing Image Analysis. *Proceedings of the 1989 International Geoscience and Remote Sensing Symposium (IGARSS 89)*, vol 2, pp. 848-851.

Wang F. (1990a). Fuzzy Supervised Classification of Remotely Sensed Images. *IEEE Transactions on Geoscience and Remote Sensing*, 28, pp. 194-201.

Wang F. (1990b). Improving Remote Sensing Image Analysis Through Fuzzy Information Representation. *Photogrammetric Engineering and Remote Sensing*, 56,

pp. 1163-1169.

Warner T.A. and M. Shank (1997). An Evaluation of the Potential for Fuzzy Classification of Multispectral Data Using Artificial Neural Networks. *Photogrammetric Engineering and Remote Sensing*, 63, pp. 1285-1294.

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*, 43, pp. 1533-1539.

Weydahl D.J. (1991). Change Detection in SAR Images. *Proceedings of the 1991 International Geoscience and Remote Sensing Symposium (IGARSS 91)*, pp. 1421-1424.

Whalen T. and B. Schott (1985). Alternative Logics for Approximate Reasoning in Expert Systems: A Comparative Study. *International Journal of Man-Machine Studies*, 22, pp. 327-346.

Whitbread P.J. (1992). *Multi-Spectral Texture: Improving Classification of Multi-Spectral Images by the Integration of Spatial Information.* Unpublished PhD Thesis, The University of Adelaide, Australia.

White R.G. (1991). Change Detection in SAR Imagery. *International Journal of Remote Sensing*, 12, pp. 339-360.

Williams D.L. and M.L. Stauffer (1978). Monitoring Gypsy Moth Defoliation by Applying Change Detection Techniques to Landsat Imagery. *Proceedings of the Symposium for Vegetation Damage Assessment*, American Society for Photogrammetry, pp. 221-229.

Xie X.L. and G. Beni (1991). A Validity Measure for Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, pp. 841-847.

Yager R.R. (1979). A Measurement Informational Discussion on Fuzzy Union and Intersection. *International Journal of Man-Machine Studies*, 11, pp. 189-200.

Yager R.R. (1991). Connectives and Quantifiers in Fuzzy Sets. *Fuzzy Sets and Systems*, 40, pp. 39-76.

Yasuoka Y., Y. Yamagata, S. Otoma, T. Miyazaki and S. Takeuchi (1993). Vegetation Mapping and Change Analysis in South-East Asia from NOAA AVHRR LAC Imageries. *Proceedings of the 1993 International Geoscience and Remote Sensing Symposium (IGARSS 93)*, pp. 1683-1684.

Zadeh L.A. (1965). Fuzzy Sets. *Information and Control*, 8, pp. 338-353.

Zadeh L.A. (1983). A Computational Approach to Fuzzy Quantifiers in Natural Languages. *Computers and Mathematics*, 9, pp. 149-184.

Zheng Q. and R. Chellapa (1993). A Computational Vision Approach to Image Registration. *IEEE Transactions on Image Processing*, 3, pp. 311-326.

Zheng Q. and R. Chellapa (1994). Automatic Registration of Oblique Aerial Images. *Proceedings of the IEEE International Conference on Image Processing*, pp. 218-222.

Zhou Y.T. (1994). Multi-Sensor Image Fusion. *Proceedings of the IEEE International Conference on Image Processing*, pp. 193-197.

Zhuang X., B.A. Engel, X. Xiong and C.J. Johannsen (1995). Analysis of Classification Results of Remotely Sensed Data and Evaluation of Classification Algorithms. *Photogrammetric Engineering and Remote Sensing*, 61, pp. 427-433.