UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

School of Geography

Soft classification and land cover mapping from remotely sensed imagery

By

Huong Thi Xuan Doan

Thesis for the degree of Doctor of Philosophy

March 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF GEOGRAPHY

Doctor of Philosophy

SOFT CLASSIFICATION AND LAND COVER MAPPING FROM REMOTELY SENSED IMAGERY

By Huong Thi Xuan Doan

Mixed pixels are often abundant in remotely sensed imagery and cannot be appropriately or accurately classified by conventional hard classification. Soft classification has become a popular approach to reduce the negative impact of mixed pixels. Although soft classification analyses can reduce problems such as those associated with mixed pixels their accuracy is often low. The key aim of this research is to investigate the ways to increase the accuracy of soft classification, the factors that impacts on soft classification and its implications for the real world applications.

Firstly, four possible methods for combining soft classifications to increase classification accuracy were assessed. All four ensemble approaches were found to increase classification accuracy. Relative to the most accurate individual classification, the increases in overall accuracy derived ranged from 2.20% to 4.45%, increases that were statistically significant at 95% level of confidence. The results highlighted that ensemble approaches may be used to significantly increase classification accuracy.

The interpretation and use of sub-pixel scale information typically places great confidence on the estimated class cover proportions. This reliance on the single set of class proportion predictions derived from a soft classification may be inappropriate. In particular, this reliance often seems to be based on an implicit assumption that a class can be represented by a single spectral end-member. This may be unrealistic since classes typically display a degree of spectral variability. The impact of intra-class spectral variation on the estimation of sub-pixel class composition was, therefore, investigated. Results from the analyses showed that the nature of intra-class variation has a negative impact on the accuracy of sub-pixel estimation as it opposed the assumption that a class can be represented by a single spectral end-member. It was suggested that a distribution of possible class compositions could be derived from pixels instead of a single class composition prediction. This distribution provided a richer indication of possible sub-pixel class composition estimates and may be used to derive different scenarios of land cover change through the use of a post-classification comparison technique approach.

The impacts of intra-class spectral variation on the use of soft classification for super-resolution mapping were assessed. It was apparent that the accuracy of the soft classification and super-resolution mapping declined as the degree of intra-class variation increased. As a distribution of possible class composition predictions may be derived for each pixel, it may be preferable to be aware of the range of possible class boundary positions.

Finally, a possible method to reduce the impacts of intra-class spectral variation on sub-pixel classification was investigated. This was achieved by defining spectral sub-classes for use in the sub-pixel classification. Results from the analysis highlighted that reducing intra-class spectral variation may be used to reduce the impact of intra-class variation so that to increase the accuracy of soft classification and super-resolution mapping.

Contents

List of Figures

List of Tables

Acknowledgement

A journey is easier and enjoyable when you travel together. This thesis is the result of three years of work whereby I have been accompanied and supported by many people. It is a wonderful moment that I now have the opportunity to express my gratitude to all of them.

I would like to thank everyone who helped me to successfully complete this thesis. I sincerely express my deepest gratitude to my supervisor Prof. G.M. Foody for his guidance, support and his ever encouraging words all through the research and produce the work in its present form. He guided me successfully to build my career in remote sensing. His advice and suggestions I will value forever.

I am grateful to department of Surveying, faculty of Geodesy and Cartography, Hanoi University of Mining and Geology for giving me permission to study for a Ph.D. degree at the University of Southampton. I am also thankful to the Vietnamese Government for the provision of a scholarship and financial support towards my research.

I wish to thank the NASA - USGS Distributed Active Archive Center at the Goddard Space Flight Center, Greenbelt for producing the NOAA AVHRR and IGBP DISCover data in their present form and distributing them. The original data products were produced under the NOAA/NASA Pathfinder AVHRR Land (PAL) program. I am also grateful to the GLCF facility at the University of Maryland for the Landsat sensor imagery. The artificial neural networks used were based on Trajan software. Hopfield Neural Network was based on Fliers software developed by Dr Hugh Lewis. I would like to thank Prof. Peter Atkinson and Dr Hugh Lewis for allowing me to use the Hopfield Neural Network software.

I have great appreciation for all members of School of Geography, University of Southampton, especially to Prof. Peter Atkinson, Dr Ajay Mathur, Dr Jadunandan Dash, Dr Charles Kerr, Jeff Neal, Nick Odoni.

Finally, I would like to thank my big family in Vietnam for their constant support and encouragement during this period. I feel a deep sense of gratitude for my late father and mother who formed part of my vision and taught me the good things that really matter in life. The happy memory of my late father and my late brother still provides a persistent inspiration for my journey in this life. I am deeply thoughtful to my husband Dung Dang for his endless emotional support and encouragement in my work and particularly for my little girl Huong Trang Dang sent from heaven to fill our hearts with love. I am also grateful to my two sisters and two brothers in laws for their helps and encouragements to my small family in Vietnam during my leave from home.

Abbreviations

ANN	Artificial neural network
AVHRR	Advanced Very High Resolution Radiometer
DN	Digital number
D-S	Dempster-Shafer
FCM	Fuzzy *c*-means
GCP	Ground control points
GRE	Ground resolution element
HNN	Hopfield neural network
IFOV	Instantaneous field of view
LDA	Linear discriminant analysis
LMM	Linear mixture model
ME	Mean error
MLC	Maximum likelihood classifier
MLP	Multi layer perceptron
OSH	Optimal separating hyperplane
PNN	Probabilistic neural network
QDA	Quadratic discriminant analysis
RBF	Radial basis function
RMSE	Root mean square error
SVM	Support vector machine

Chapter 1

Introduction

1.1 Introduction

Land cover and land cover change mapping through the means of image classifications is one of the most common applications of remote sensing. However, the full potential of remote sensing as a source of both land cover mapping and monitoring activities is often unrealized due mainly to a set of technical problems. One of the most important problems limiting classification accuracy is that of mixed pixels (Fisher, 1997; Cracknell, 1998). A mixed pixel occurs when the area represented by the pixel contains more than one land cover class (Fisher, 1997). The proportion of mixed pixels in an image is often large. The number and proportion of mixed pixels in an image, for example, generally increases with a coarsening of the spatial resolution of the sensor and/or increasing fragmentation of the landscape to be represented (Campbell, 2002; Foody, 2002a). Thus, the problem of mixed pixels occurs most seriously in coarse spatial resolution remotely sensed imagery. Unfortunately, the applications of remote sensing using these coarse spatial resolution data sets are widely used in the mapping and monitoring of large areas and impacting on studies attempting to study major land cover transformations such as deforestation.

Since a mixed pixel represents an area of more than one land cover class, the image which contains mixed pixels cannot be appropriately or accurately classified by a conventional hard classification. The mixed pixel problem may lead to substantial error in land cover mapping from remotely sensed data. This error may impact on studies of land cover change when basing on post-classification comparison technique. The mixed pixels problem may lead to very large error, with, for example, study from Skole and Tucker (1993) showing that deforestation may be over-estimated by about 50% if coarse spatial resolution data sets used to derive classifications. The mixed pixel problem, therefore, should be addressed in order to derive accurate land cover information. Using fine spatial resolution remotely sensed imagery is one approach that may reduce the proportion of mixed pixels. This can be a very effective approach

for the studies of small area but not for large area where a large number of images may be required. It may costly to acquire the imagery and substantially pre-processing could be needed to inter-calibrate and mosaic the images. An alternative approach to avoid these problems is to derive sub-pixel scale information from the coarse spatial resolution imagery.

Sub-pixel scale land cover information is typically derived by unmixing the spectral responses of image pixels to identify their class compositions. There are a variety of methods for estimating the class composition of mixed pixels which have been applied to remotely sensed data, such as the linear mixture model (Foody and Cox, 1994, Atkinson *et al.*, 1997; Van De Meer and De Jong, 2000; Theseira *et al.*, 2003; Small, 2004) and soft or fuzzy classifications (e.g. Foody, 1996a; Bateson *et al.*, 2000;). Soft classification approaches are those based on the maximum likelihood (Wang, 1990a and 1990b; Maselli *et al.,* 1995; Ibrahim *et al.,* 2005), fuzzy *c*-means (Zhang and Foody, 2001a; Ibrahim *et al.,* 2005), *k*-nearest neighbour (Schowengerdt, 1996), artificial neural network (Foody, 1996b; Zhang and Foody, 2001) and support vector machine classifications (Brown *et al.,* 2000) ect. The output of these analyses is typically a set of fraction images, one per-class, that show the predicted coverage of each thematic class in the area represented by each pixel. These fraction images can achieve accurate estimation of class composition. They also support to the representation of environmental continua and the detection of land cover changes when used in post-classification analyses (Foody, 2001b; Haertel *et al.,* 2004). The sub-pixel information also forms the basis of super-resolution mapping, in which the geographical location of the estimated class composition is located within each pixel's area to yield a thematic map at a finer resolution than the imagery used in its derivation (Tatem *et al.*, 2002; Mertens *et al.*, 2006; Muslim *et al.*, 2006; Boucher and Kyriakidis, 2006).

Although the fraction images derived from a soft classification can provide a more accurate and appropriate representation of themes such as land cover than a conventional hard classification there are many concerns. In particular, the accuracy of soft classifications is often low. The literature, for example, contains many examples of soft classification analyses in which the degree of correlation between the actual and predicted class composition was low (r<0.5) (e.g., Foody and Arora, 1996, Zhang and

Foody, 2001, Shalan *et al.,* 2003, Ibrahim *et al.*, 2003). This may likely to be a major limitation to studies seeking to estimate class fractional cover and its change over time or the sub-pixel distribution of classes as in super-resolution mapping.

In reality the land cover classes of interest are rarely fully separable spectrally. Consequently, it is impossible to expect a perfect classifier which gives 100% accuracy. Different classifiers will commonly vary in their allocations, with some yielding high accuracies for some parts of the data while others may achieve high accuracies in other parts. A variety of different classification outputs can be derived from the application of a suite of classifiers to the same data set. The derived classifications may differ greatly in accuracy, on both a per-class and overall basis. Classifiers that have, largely, uncorrelated errors may sometimes usefully be combined to form an ensemble approach to classification (Giacinto and Roli, 2001a ; 2001b). By combining the outputs of a set of classifiers it is possible to derive a classification that is often more accurate than the individual classifications used. The use of such an ensemble of classifiers has been shown to be able to increase the accuracy of hard classification analyses (Lam and Suen, 1997; Ji and Ma, 1997; Briem *et al.,* 2002; Steele, 2000; Giacinto *et al.,* 2000; Liu *et al.,* 2002) but its potential for soft classification has rarely been investigated (Huang and Lees, 2004).

The interpretation and use of sub-pixel classification output often rely on the estimated class cover proportions. By using the outputs derived from two soft classifications for land cover change detection the estimated class proportions are typically compared directly (Foody, 2001b; Haertel *et al.,* 2004). In super-resolution mapping some approaches try to maintain the class proportion information output from soft classification (Tatem *et al.*, 2002; Muslim *et al.*, 2006). That is, if the soft classification indicates a pixel contains 30% forest the super-resolution mapping will map the pixel 30% forest. This reliance on the single set of class proportion estimates may be unwise. This seems to be based on the assumption that a class can be represented by a single spectral end-member. This is likely to be unrealistic as classes typically display a degree of spectral variability since the spectral signatures of land cover classes vary from pixels to pixels due to changes in biophysical (e.g., leaves, stems and bark) and biochemical (e.g. chlorophyll content) composition (Bateson *et al.*, 2000; Song, 2005). In fact, it is known that the accuracy of soft classification is

negatively related to the degree of intra-class variation (Petrou and Foschini, 1999) and approaches to refine basis unmixing methods to accommodate for this have been investigated (Bateson *et al.*, 2000; Song, 2005). However, it is still common to see basic approaches to unmixing being used and the sub-pixel estimates used in manner that places great confidence in their accuracy.

The research program reported in this thesis aims to investigate the use of soft classification for the extraction of land cover information from remotely sensed data. The research considers the issues including the ways to increase soft classification accuracy, the impacts of intra-class spectral variability on sub-pixel class composition estimates and its implications for the real world applications. There were three specific objectives:

i. To investigate some of the possible approaches to increase the accuracy of soft land cover classification though the use of an ensemble of classifiers;

ii. To investigate the impacts of intra-class spectral variability on soft classification prediction and highlight its implications for analyses based on soft classification such as the detection of land cover change;

iii. To explore the impacts of intra-class spectral variability on super-resolution mapping and investigate a possible approach to reduce these impacts through the reduction of the intra-class spectral variation.

1.2 Thesis outline

Chapter 2 reviewed how remotely sensed imagery may be used to provide land cover maps, with special regard to land cover classification. The chapter focused on the techniques to extract land cover information from remotely sensed imagery such as supervised classification. The problem of mixed pixels was then presented and its solution through the use of soft classification and super-resolution mapping was discussed.

Chapter 3 reported on the possible approaches to increase the accuracy of land cover classification though the use of an ensemble of classifiers. Initially, the approaches which have been used in the literature to combine hard classifications were explored. Following this, the methods to combine soft classifications were provided. The potential of the approaches to combine soft classifications as well as the approaches to combine hard classifications were assessed through a range of experiments.

Chapter 4 investigated one possible factor that may affect the accuracy of soft classification, the intra-class spectral variability. Attentions focussed on the impacts of intra-class spectral variability on soft classification prediction and highlighted its implications for analyses based on soft classification such as the detection of land cover change.

Although soft classifications estimate the proportion of land cover classes for each image pixel they do not show where within pixels the land cover classes are spatially located. The class proportions estimated in each image pixel may be geographically located using super-resolution mapping approach. Chapter 5, therefore, explored the impacts of intra-class spectral variability on super-resolution mapping and investigates a possible approach to reduce these impacts through the reduction of the intra-class spectral variation.

Finally, chapter 6 summarised the main results of the thesis and discussed the conclusions that arise from the research findings in the whole thesis.

Chapter 2

Literature review

2.1 Introduction

This chapter reviews how remote sensing may be used to provide land cover maps. Initially, the topic of land cover is introduced, followed by a review of the principles of remote sensing. The techniques to extract land cover information from remotely sensed imagery are then presented, including techniques for image pre-processing, and some methods of hard classification. Finally, the problem of mixed pixels is presented and its possible resolution through the use of soft classification is discussed in detail.

2.2 Land cover

2.2.1 Definition of land cover

The term land cover relates to the surface cover of the Earth (Lillesand and Kiefer, 2000). Land cover corresponds to the physical condition of the ground surface, for example, forest, grassland, concrete, pavement *etc.*, either natural or human-made, on the Earth's surface at a specific time of observation (Campbell, 2002). Land cover mapping is one of the most important and typical applications of remote sensing data. Information on land cover is required to help our understanding and management of the environment. Land cover maps are presently being developed from local to global scales.

2.2.2 The need for land cover information

Information on land cover is central to all scientific studies that aim to understand terrestrial dynamics and is required from local to global scales to guide land use planning. Land cover information can be of benefit to planners in some ways. First land cover data inventories have the potential to provide knowledge which is valuable for determining land management practices (Townshend, 1992, Veitch *et al.*, 1995). Second, land cover data can be used to provide information on land cover dynamics for

monitoring changes in land cover (Mather, 1999). Finally, the analysis of land cover data can help planers understand the interaction between land use practices and environmental processes (Dimyati *et al.,* 1996). This puts planners in a better position to implement land use policies that minimize harm to the environment.

Although land cover plays a very important role in society, data on land cover are often out-of-date, of poor quality or inappropriate for a specific application (Foody, 1996a). However, land cover data are not easy to collect, especially if data are required for large areas or if frequent up-dating is required. A common way to obtain land cover information is remote sensing. Other alternative methods such as field survey can be relatively expensive and time consuming, especially for large areas (e.g., national or global scales).

2.3 Introduction to remote sensing

2.3.1 Definition of remote sensing

The field of remote sensing is very broad (Campbell, 2002). According to Lillesand and Kiefer (2000, p.1), "remote sensing is the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation". The remote sensors acquire data on variations in electromagnetic energy. An electromagnetic remote sensing system is illustrated in Figure 2.1.

Based on this definition, there are two main basic processes involved in the remote sensing system that are data acquisition and data analysis. The elements of the data acquisition process are energy sources, propagation and retransmission of energy through the atmosphere, energy interactions with Earth surface, and the image or digital image for data collection. The data analysis process involves examining and interpreting the image data collected by the sensors to collect information and provide the information for users.

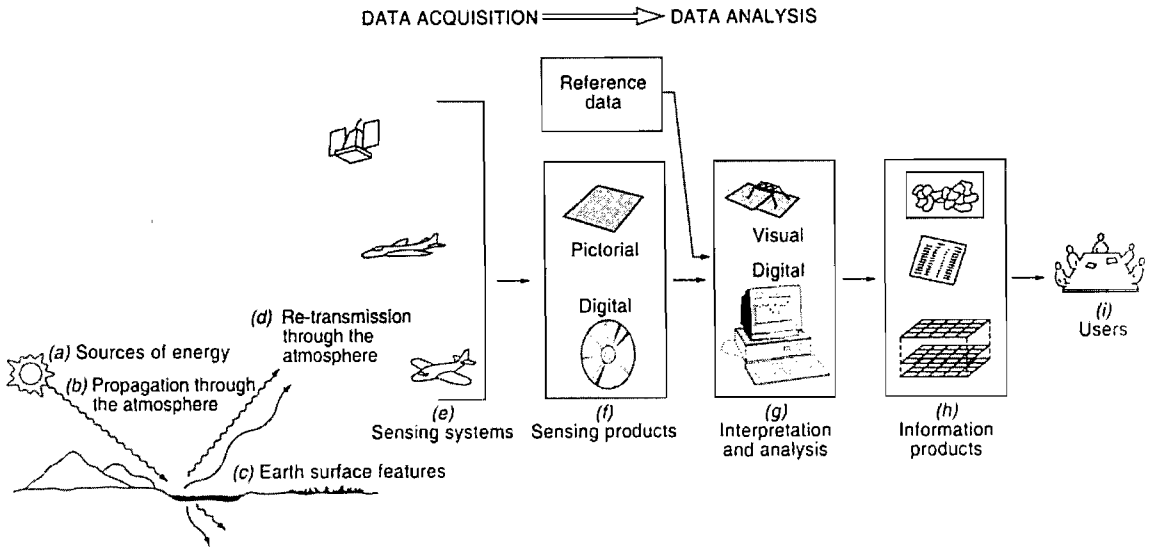


Figure 2.1: Overview of a remote sensing system (Lillesand and Kiefer, 2000).

2.3.2 Physical principles of remote sensing

2.3.2.1 The electromagnetic spectrum

Remote sensing makes use of electromagnetic radiation. In fact, all objects emit electromagnetic radiation and some objects also reflect radiation that has been emitted by the other objects (Campbell, 2002). The main source of electromagnetic radiation used in remote sensing is our Sun, which emits radiation over the entire electromagnetic spectrum (Lillesand and Kiefer, 2000, Campbell, 2002) (Figure 2.2).

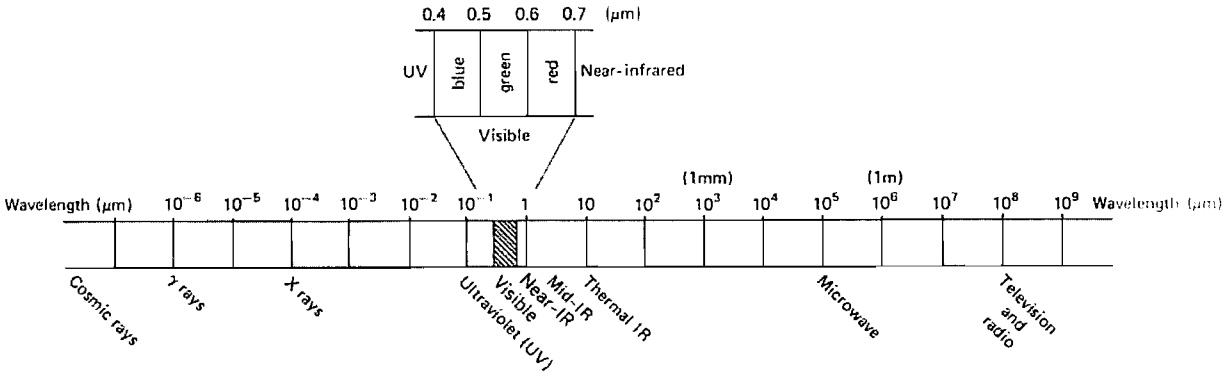


Figure 2.2: Principle division of the electromagnetic spectrum (Lillesand and Kiefer, 2000).

2.3.2.2 Interaction with the atmosphere

All radiation detected by remote sensing must pass through the Earth's atmosphere (Lillesand and Kiefer, 2000, Campbell, 2002). As a result, the quality of images and data that the sensors create are affected by the atmospheric conditions. These effects are caused mainly through several physical processes, including: scattering, absorption, and refraction. All of these effects are, in general, dependent on wavelength (Rees, 1990).

2.3.2.3 Interaction with Earth surface features

When electromagnetic energy reaches the Earth's surface, it interacts with the surface in three fundamental ways such as reflected, absorbed, and/or transmitted. The proportions of light accounted for each process depend on the nature of the surface, the energy's wavelength and the angle of illumination (Campbell, 2002). The interrelationship among these three energy interactions can be expressed as follows (Lillesand and Kiefer, 2000):

$$E_I(\lambda) = E_R(\lambda) + E_A(\lambda) + E_T(\lambda) \tag{2.1}$$

where $E_T(\lambda)$ represents for the incident energy, $E_R(\lambda)$ is the reflected energy, $E_A(\lambda)$ is the absorbed energy, $E_I(\lambda)$ is the transmitted energy, and λ denotes the wavelength.

2.4 Image processing and analysis

2.4.1 Pixel and digital data

Pixels are the smallest picture elements or cells of an image, and represent the ground resolution element - GRE of a sensor. The ground area represented by each pixel is defined by the characteristics of the sensor that produced the image. Remote sensing images are recorded in digital form and then processed by computer to produce images for interpretation purposes. A digital image is composed of many thousands of discrete pixels. Each pixel represents the remotely sensed response of a small region on the Earth's surface, recorded digitally as a numeric value represented as DN (Digital

Number). Within a GRE, the properties of Earth surface are considered to be homogeneous although this is rarely the case in reality.

2.4.2 Image resolution

Image resolution can be defined as the ability of an imaging system to record fine details in a distinguishable manner (Campbell, 2002). In remote sensing, it is important to distinguish between 4 resolutions: spatial resolution, radiometric resolution, spectral resolution, and temporal resolution.

2.4.2.1 Spatial Resolution

Spatial resolution determines the fineness of spatial detail visible in an image. The finer the spatial resolution the smaller the ground objects that can be identified. For example, fine spatial resolution imagery such as IKONOS is created at 1 m in the panchromatic band and 4 m in the multispectral bands and a moderate resolution image such as Landsat ETM image is produced at 30 m. In contrast, coarse spatial resolution imagery like National Oceanic and Atmospheric Administration's Advanced Very High Resolution Radiometer (AVHRR) is produced at 1.1 km spatial resolution. The spatial resolution is related to the instantaneous field of view (IFOV) of the sensor, which is the angular subtend from which the sensor receives the energy at the given instant in time (Schowegerdt, 1997, Tso and Mather, 2001).

2.4.2.2 Radiometric Resolution

Radiometric resolution is defined as the number of digital levels used to express the data collected by the sensor. It is commonly expressed as the number of bits (binary digits) needs to store the maximum level. For example Landsat TM data are displayed to 256 levels (0 to 255) and this equivalent to 8 bits.

2.4.2.3 Spectral resolution

Spectral resolution refers to the width of spectral bands and expresses the ability of a sensor to define the fine wavelength intervals. Different features on the Earth surface

obtain different spectral reflectances and emissivities and these spectral characteristics specify the spectral position and spectral sensitivity to distinguish features.

2.4.2.4 Temporal resolution

Temporal resolution refers to the frequency with which remotely sensed imagery of a defined geographic area can be acquired. The temporal resolution is determined by orbital characteristics and swath width, the width of the imagery in the area. This variable is important where imagery is being used to monitor land cover change over time.

2.4.3 Overview of land cover mapping

Digital image processing comprises a range of techniques for the manipulation of digital images by computers. The raw data acquired from the imaging sensors on the satellite platforms contains flaws and deficiencies. To get the originality of the data, it needs to undergo several steps of processing to reject these effects. This varies from image to image depending on the type of image format, initial condition of the image and the information of interest. Generally, digital image processing consists of three main steps: (i) Pre-processing, (ii) display and enhancement, (iii) information exaction. It can be illustrated by the flowchart in Figure 2.3.

2.5 Pre-processing

In the raw form, prior to any form of processing, remotely-sensed data generally contain distortion and deficiencies due to the influence of external factors which alter the spectral signal from surface features. These factors include atmospheric interference, variation in illumination geometry, geometric distortion and technical problems with the sensor. Therefore, to facilitate an accurate representation of land cover from remotely sensed imagery, it is first necessary to remove the influence of these factors. Preprocessing includes a wide range of operations. Typically preprocessing operations include: (1) radiometric corrections, (2) atmospheric correction, and (3) geometric corrections (Mather, 2004, Lillesand and Kiefer, 2000, Campbell, 2002), though all these corrections might not be necessarily be applied in all cases. The analyst should

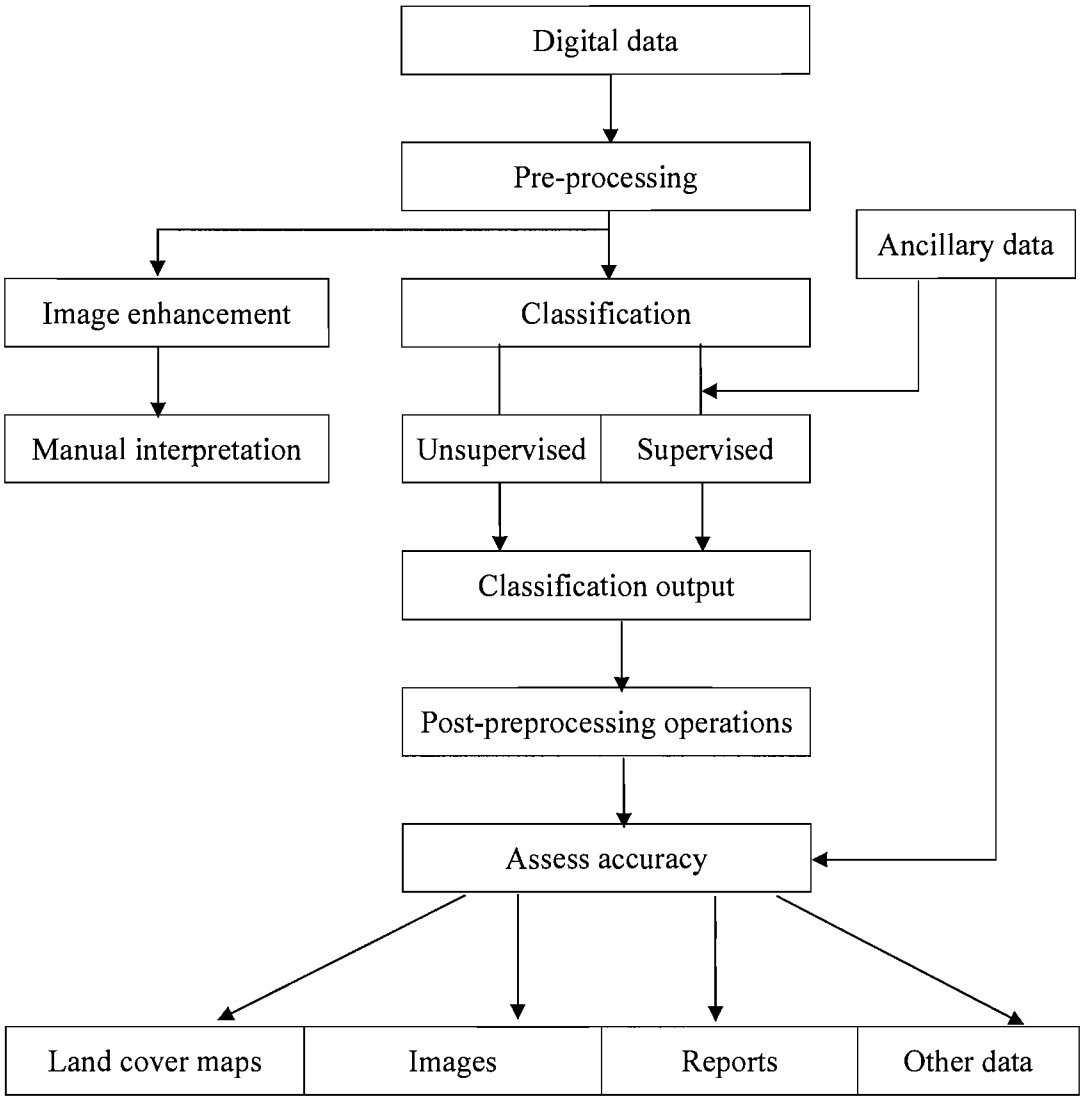


Figure 2.3: Overview of land cover mapping.

decide which pre-processing techniques are needed based on the quality of image data and the information to be extracted from those image data. Of the three techniques mentioned above, the last one is generally required.

2.5.1 Radiometric correction

When an image is recorded by a sensor the image contains errors in the measured brightness values of the pixels. These errors are referred as radiometric errors. Radiometric correction affects the brightness values of an image to correct for sensor malfunctions or to adjust the values to make up for atmospheric degradation (Campbell, 2002). It is the process of converting the spectral values of remotely sensed imagery

from DN values - arbitrary values based on the radiometric resolution of the sensor, to radiance values (Curran, 1985). This process, which changes the DN values following the gain and offset parameters of the sensor, is important for some research where absolute values of radiation are required (Gu *et al.*, 1999).

2.5.2 Atmospheric correction

Electromagnetic energy detected by the remote sensing sensors contains energy reflected and emitted by the ground surface and energy that has been scattered within or emitted by the atmosphere. Thus, the objective of atmospheric correction is to recover the surface reflectance (that characterizes the surface properties) from remotely sensed imagery by removing the atmospheric effects. This is the process of converting values of radiance to values of reflectance. This is implemented by accounting for the contribution of atmospheric interference to the spectral values of an image. Atmospheric correction algorithms mainly include two major steps. First, the optical characteristics of the atmosphere are estimated either by using special features of the ground surface or by direct measurements of the atmospheric compositions or by using theoretical models. Second, the remotely sensed imagery can be corrected by inversion procedures that derive the surface reflectance. The procedure necessary for atmospheric correction depends on the purposes of the analysis. For instance, land cover classifications based on single-date images do not need atmospheric correction if it can be assumed that all pixels in the image are affected equally by the atmosphere since the pixels are being compared with other pixels within the image. However, atmospheric correction is necessary when multidate images are being classified, or for identification of land cover change over time, to make sure that pixel values are comparable from one image to the next (Tso and Mather, 2001).

2.5.3 Geometric correction

Geometric correction is carried out to reduce geometric distortions from a distorted image, and is achieved by projecting image data onto a plane and making it conform to a map projection and co-ordinate system.

The geometric correction process is carried out by two steps. First, systematic image distortions such as those arsing from Earth curvature, panoramic distortion and non-

linear scan lines are well understood and corrected by using formulas derived by modelling the sources of the distortions mathematically. Second, random image distortions such as those arising from variations in platform movement and relief can be corrected by transforming distorted image to regular coordinate systems. To do this, ground control points (GCPs – features of known ground location on a coordinate system in terms of UTM coordinates or latitude and longitude) are identified on both the image and map coordinate system and the two sets of coordinates (image and map) are compared (Lillesand and Kiefer, 2000).

After producing the transformation function, a resampling approach is used to determine the spectral value of the pixels in the transformed image, based on the values of the original image. Several different resampling schemes may be applied including nearest neighbour, bilinear interpolation and cubic convolution (Lillesand and Kiefer, 2000, Campbell, 2002). Among them, nearest neighbour is the simplest method and widely used since it uses the value of the 'nearest' original pixel prior to transformation as the new pixel value and thus, had the benefit of maintaining the original pixel values, although this can cause a disjointed appearance in the output image product.

2.6 Hard classification

A common way to extract land cover information from remotely sensed imagery is to classify it. Image classification is the process of assigning pixels to classes (Campbell, 2002). Land cover classes can be identified by their spectral, spatial, temporal and geometric characteristics. The simplest form of automatic (computer-based) image classification is to consider each pixel individually, assigning it to the most similar spectral class according to its spectral values in separate wavebands (Mather, 2004). For convenience per-pixel classification will be assumed throughout this thesis although the research may be applied to classification based on other spatial units (e.g., per-field, per-polygon classification *etc*).

Classification techniques may be performed using either of two main approaches, namely unsupervised or supervised classification. Unsupervised classification can be defined as the process of automatically aggregating pixels within multispectral data into a number of classes based on the natural groupings or clusters present in the image values (Campbell, 2002). Supervised classification can be defined as the process of

using pixels representing areas of known class membership to classify pixels of unknown membership. Pixels of known membership are located within training areas (section 2.6.2.1), which are identified by the operators according to the regions that matched to the areas of known land cover classes on the image. The main benefit of unsupervised classification is that it has a low requirement for manual processing. Besides, the pixels are grouped automatically by their spectral reflectance values within the imagery. Thus, unsupervised classification may be useful as a means to identify class separability in imagery prior to subsequent analysis. Conversely, supervised classification needs more efforts from the analysis, but has the benefit that the classes generated meet the analysis's requirement and may, therefore, relate directly to the problem. This is the reason why supervised classification has been more popular than unsupervised classification in remote sensing.

In addition to the distinction between unsupervised and supervised classification, the classification techniques can also grouped on the basis of how classifiers represent land cover. Generally, they can be divided into two main types: hard classification and soft classification. Hard classification or conventional classification is a process in which each pixel is allocated to the single land cover class with which it has the greatest similarity (Rees, 1994; Foody, 2002a; Mather, 2004; Lillesand and Kiefer, 2000). In contrast, soft (fuzzy) classification represents land cover as a mixture of classes, and for each pixel, the probability of membership to each class is identified (Wang, 1990; Foody, 1996a; Atkinson, 1997; Campbell, 2002; Mather, 2004). The main advantage of hard classification is that hard classification outputs are quite simple and, thus, hard classified images require much less computer storage than soft classified images. However, soft classifiers are much more realistic than hard classifiers in terms of the representation of land cover. While hard classification operates under the hypothesis that each pixel is perfectly pure, fuzzy classification recognizes that, in fact, classes are not mutually exclusive and a single pixel may contain more than one land cover class. The latter technique may solve the problem of mixed pixels (section 2.7). As a result, there is a growing interest in, and application of, soft classification (e.g., Fisher, 1990; Foody and Cox; 1994; Atkinson *et al.*, 1997; Embashi, 1998; Ghosh *et al.,* 2003; Gopal, 1998; Zhang and Foody, 2001; Foody, 2002b). This section provides a brief overview of hard classification, and soft classification is described in the section 2.8.

The majority of conventional classifiers are hard, representing land cover as a series of discrete units, whereby each pixel is associated with a single land cover class. This means that hard classification operates under the assumption that each pixel is perfectly pure, i.e. the ground area represented by a pixel is occupied by a single information class. The result of this process is a classified image in which each pixel is assigned to a land cover class.

2.6.1 Stages of hard classification

After appropriately pre-processing an image (section 2.5), the image can be used to classify in order to derive a thematic map from imagery. This process has three stages: training stage, class allocation stage, and accuracy assessment stage.

2.6.1.1 Training stage

Supervised classification methods require prior knowledge of the number of information classes. The information classes are expressed in training samples. The analyst defines training areas by locating areas on the image that can be matched to areas of known land cover class. This is an important procedure which has considerable impacts on classification accuracy. Training areas should be homogeneous in respect to the classes to be classified. In general, the class selection depends on the content of the thematic map needed or on each project and should be implemented using some knowledge of the study area. In practice, training areas can be normally identified by fieldwork, from aerial photographs, or from map interpretation, and their positions on the image can be located by visual observations directly or by carrying out a geometric correction on the image to be classified (Mather, 2004).

When identifying training areas in an image, some important factors of concern are the overall number of training pixels in the image, size, location of the training area, and the number of training areas in the image (Campbell, 2002). The two most important factors are the number of training pixels and the size of training area. According to Campbell (2002), the overall number of training pixels in an image should be at least 100 pixels for each category. Mather (2004) stated that the number of training pixels per class should be at least $30p$ where p is the number of wavebands. Lillesand, and Kiefer (2000) suggested this figure should be at least $10p$. For training size, some researchers

suggested that single-pixel training (or randomly selected training pixels) resulted in higher classification accuracies than contiguous blocks of training pixels. According to Campbell (1981), positive spatial autocorrelation exists among pixels which are spatially contiguous and the randomly selected pixels may reduce the effect of positive interrelationships between adjacent pixels upon the subsequent estimates. Gong and Howarth (1990) stated that the block training method violates the independent sampling requirement and therefore makes the training spectral responses for each class less representative.

After all training areas are selected, they will be used in the training process (or training stage). The training stage is the process in which areas of known class membership in the image are used to derive a statistical description of each class (Foody, 2002b). The statistical characteristics of the classes that are to be estimated from training pixels depend on which method of supervised classification is used. For example, the parallelepiped algorithm requires estimates of the extreme values on each feature for each class, while maximum likelihood requires estimates of the mean vector and variance-covariance matrix of each class. Thus, the training data sets help the classifier to recognize the spectral properties that characterise or distinguish each class.

2.6.1.2 Class allocation stage

The derived training statistics from the training stage are then used in the second stage, the class allocation stage (or pixel labelling), to assign pixels representing an area of unknown class to the class with which it has greatest spectral similarity. There are various methods of classification which are used to distinguish the spectral similarity such as parallelepiped, minimum distance, maximum likelihood, decision tree, discriminant analysis, artificial neural network, support vector machine, *etc*. These techniques are described in section 2.6.3. The spectral similarity is assessed differently among different classification methods. For example, the minimum distance classifier labels pixels with the classes of which they are closest to in spectral space, but maximum likelihood assigns pixels to the class with which it has the highest a *posterior* probability of membership. The class allocation stage has the effect of converting the remotely sensed image displaying the spectral response of the Earth's surface into a map.

2.6.1.3 Accuracy assessment stage

The final stage of the image classification problem is the accuracy assessment stage (or the testing stage) in which the accuracy of the classification is evaluated. This is very important stage because without it, the user does not know the accuracy of the classification and, therefore, does not know the utility of the classification (Nishii and Tanaka, 1999). The term accuracy in image classification means the level of agreement between labels assigned by the classifier and class allocations based on ground data collected by the user, known as reference data. In thematic mapping from remotely sensed data, accuracy defines the degree of 'correctness' of a classified image (Campbell, 2002; Foody, 2002c). In general, accuracy assessment is implemented by comparing the classified image with the known reference land cover data of the study. This can be a simple visual examination of the two data sets, involving the comparison of a statistical sample of points in study area (Gong and Howarth, 1990). Errors present in a classification are caused by misidentification of pixels, excessive generalization, errors in registration, variation in detail of interpretation, and other factors (Campbell, 2002). The simplest error is the assignment of a pixel belonging to one category to another category through the classification process.

Many methods of accuracy assessment have been discussed in the remote sensing literature (Rosenfield and Fitzpatrick-Lins, 1986; Lillesand and Kiefer, 2000; Campbell, 2002). The most commonly used, however, derived from a confusion or error matrix (Figure 2.4). This indicates the number of sample pixels assigned to each class relative to the actual land cover class as identified using reference data. Such a matrix is square, with the number of rows and columns equal to the number of categories whose classification accuracy is being assessed. The confusion matrix identifies not only overall accuracies for each category but also misclassifications by category. The columns represent the sample elements assigned to corresponding actual classes (reference data), while the rows show the sample elements assigned to corresponding classified classes (predicted class). The meaning of columns and rows can be reversed in some applications since the convention is not universal. The diagonal elements of this matrix indicate the number of sample elements which have been classified correctly, whereas off-the-diagonal elements show misclassification.

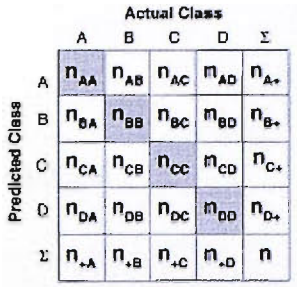


Figure 2.4: The confusion matrix (Foody, 2002c).

Many statistical measures of classification accuracy can be derived from an error matrix. Alternative formulas exist for matrices based on different sampling designs.

(i) Overall classification accuracy

This is one of the most popular measures of accuracy, and it is computed by dividing the total number of correctly classified pixels (i.e., the sum of elements along the major diagonal) by the total number of reference pixels.

$$A_0 = \frac{\sum_{i=1}^{c} n_{ii}}{n} \tag{2.2}$$

where c is the number of classes, n_{ii} is the number of correctly classified pixels in class i, and n is the total number of testing (reference) pixels for all classes.

(ii) Errors of omission and errors of commission

All non-diagonal elements of the matrix represent errors of omission and commission. Omission errors correspond to non-diagonal column elements. Commission errors correspond to non-diagonal row elements.

(iii) Producer's accuracy

The producer accuracy is the proportion of the number of correctly classified pixels in each class to the total pixels of that class in the reference data (Campbell, 2002). It can be computed as:

$$A_u = \frac{n_{ii}}{n_{+i}} \tag{2.3}$$

where n_{i+} is the marginal total of column i.

(iv) User's accuracy

The user accuracy is the proportion of the number of the correctly classified pixels in each class to the total pixels of that class to be classified (Campbell, 2002). It can be computed as:

$$A_p = \frac{n_{ii}}{n_{i+}} \tag{2.4}$$

where n_{+i} is the marginal total of row i.

(v) Kappa coefficient

Kappa ($\hat{k}$) is a measure to accommodate for the effects of chance agreement. In some circumstances, it can be used as a measure of classification accuracy (Smits, *et al.*, 1999; Foody, 2002c). It can be calculated as follows,

$$\hat{k} = \frac{A_0 - \sum_{i=1}^{c} n_{i+} n_{+i}}{1 - \sum_{i=1}^{c} n_{i+} n_{+i}} \tag{2.5}$$

The kappa coefficient has the benefit that all elements in the error matrix participate in its computation, rather than simply the main diagonal. Overall, the confusion matrix is a valuable tool of representing the results of classification accuracy assessment, from which various measures can be calculated to indicate levels of accuracy as required.

2.6.2 Supervised classification algorithms

Supervised classification algorithms may be divided into two groups: parametric or nonparametric. Parametric algorithms assume a class statistical distribution and require estimates of the distribution parameters, namely the mean vector and variance-

covariance matrices for each of classes, for classification (Schowengerdt, 1997; Kuncheva, 2000). The common type of statistical distribution for parametric classification algorithm applied to remote sensing classification is the normal (or Gausian) distribution. Parametric algorithms depend on the form of the probability distribution for each class. An example is maximum likelihood algorithm, which uses a probability model to identify the decision boundaries. The necessary parameters for the classification model are derived from training data. The limitations of the parametric classifiers are the dependence on statistical distribution, inability to handle data from different sources.

Nonparametric algorithms make no assumptions about the probability distribution of the data and are often considered robust since they may implement for many class distributions, provided that the class signatures are reasonably distinct (Schowengerdt, 1997; Kuncheva, 2000). They use information directly from training pixels. An example of nonparametric classifier is artificial neural network. The advantages of nonparametric algorithms are that they are the distribution free, and their ability to handle multi-source data efficiently.

Some of the main methods of parametric and non-parametric classifiers are described on the next following sections.

2.6.3 Parametric classifiers

2.6.3.1 Parallelepiped classification

In this method, the lower and upper limits of reflectances (pixel values) in each waveband, defined using training data, form the dimensions for each side of a parallelepiped. The unclassified pixels are projected into data space; those pixels that fall within the areas defined by the training data are assigned to the appropriate land cover classes. The parallelepiped method is fast and simple, but errors may occur, especially when a pixel lies outside all parallelepipeds or inside more than one parallelepiped.

2.6.3.2 Minimum distance classification

Minimum distance classification uses the mean values of the spectral data in the training data to allocate unknown pixels to land cover classes. Each training class is represented by its class centroid. Thus, the multidimensional distances from unknown pixel to each class centres are calculated and this unknown pixel will be assigned to the closest class. The distance between a pixel and a class centre is commonly measured either by the Euclidean distance or Mahalanobis distance (Tso and Mather, 2001). The Euclidean distance can be expressed as:

$$D_E = \sqrt{(x_i - \mu_i)^2} \tag{2.6}$$

where D_E is the Euclidean distance, x_i is the observed vector of the ith pixel and μ_i is the current mean vector of the jth class (or cluster). The dimension of vector x_i is equal to the number of bands being used as input.

The Mahalanobis distance can be expressed as:

$$D_M = (x_i - \mu_j)^{\mathrm{T}} C_j^{-1} (x_i - \mu_j) \tag{2.7}$$

where D_M is the Mahalanobis distance, $^{\mathrm{T}}$ denotes the matrix transpose, and C_j^{-1} is the inverse of the variance-covariance matrix for cluster j.

The minimum distance classifier is simple in concept and in implementation, but it is not widely used in remotely sensed image classification. It is sometimes not accurate in its simplest form because it does not support for the differences in variability of classes and therefore, some classes may overlap at their boundaries (Campbell, 2002).

2.6.3.3 Maximum likelihood classification

The maximum likelihood classifier (MLC), a supervised statistical approach to pattern recognition, is one of the most widely used methods of hard classification. It operates by using the training data as a means of estimating means and variances of the classes, which are then used to estimate the probabilities. ML classification considers not only

the mean, or average, values in labelling classification but also the variability of brightness values in each class. The probability density functions are calculated on the assumption that both training data and the classes themselves display a multivariate normal (Gaussian) distribution (Campbell, 2002; Mather, 2004). In this method, the conditional probabilities of an unknown pixel belonging to each of the predefined classes are calculated, and then the pixel is allocated to the land cover class which has the highest posterior probability of membership. MLC is based on the Bayesian probability formula (Tso and Mather, 2001):

$$P(x,\omega) = P(\omega \mid x)P(x) = P(x \mid \omega)P(\omega) \tag{2.8}$$

where $P(x, \omega)$ is the probability of pixel x belonging to class ω, $P(x)$ and $P(\omega)$ are the prior probabilities of pixel x and class ω, and $P(\omega|x)$ is the conditional probability of pixel x given to class ω. If x_i is the i^{th} pattern vector and ω_j is class j then the probability that x_i belongs to class ω_j is calculated by:

$$P(\omega_j \mid x_i) = \frac{P(x_i \mid \omega_j)P(\omega_j)}{P(x_i)} \tag{2.9}$$

The conditional probability is normally calculated from a Gaussian distribution as follows:

$$P(x_i \mid \omega_j) = \frac{1}{\sqrt{2\pi^{\rho}}\sqrt{|C_j|}} \exp\left(-\frac{1}{2}(x_i - \mu_j)^{\mathrm{T}} C_j^{-1}(x_i - \mu_j)\right) \tag{2.10}$$

where C_j is the covariance matrix of class ω_j with dimension ρ, μ_j is the mean vector of class ω_j, and $|.|$ is the determinant. The value in mean vector, μ_j, and the covariance matrix, C_j, are estimated from the training data.

An extension of the MLC is the Bayesian classifier. This method uses two weighting factors to estimate probabilities. Fist, the analyst identifies the prior probability for each class in the given scene. Second, a weight related to the "cost" of misclassification is used for each class. Together, these two weighting factors operate to minimize the "cost" of misclassification, resulting in theoretically optimum classification. In practice, most MLC is operated assuming equal prior probability of occurrence and cost of

misclassification for all classes. If the data required are suitable for these factors, the Bayesian classifier is preferable.

2.6.3.4 Discriminant analysis

Method of discriminant analysis (Klecka, 1980; Hastie *et al.*, 2001) is widely used in the classification of remotely sensed data (Tom and Miller, 1984; Lark, 1994; Foody and Mathur, 2004). Discriminant analysis is a conventional statistical classifier which generally assigns each case to the class with which it has the highest posterior probability of membership. The posterior probability $L(i \mid X)$ of case X belonging to class i can be calculated as follow (Foody, 1996a):

$$L(i \mid X) = P_i p(X \mid i) / \sum_{j=1}^{c} P_j p(X \mid j) \tag{2.11}$$

where $p(X \mid i)$ is the typicality probability (the probability that case X would be a member of class i given the distance it is from the centroid of class i), P_i is the *prior* probability for class i, and c is the number of classes. The posterior probabilities lie on 0-1 scale and sum to 1.0 for each pixel. The class centroids can be calculated from the training data by using variance-covariance matrices, which can be a single pooled matrix or separate matrices for each class (Lark, 1994).

There are two widely used methods of discriminant analysis namely linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) (Lark, 1994; Hastie *et al.*, 2001). In LDA a pixel is assigned to a land cover class which it has the minimum Mahalanobis distance, computed using a pooled variance-covariance matrix. Whereas, in QDA a pixel is assigned to a land cover class which has the smallest generalised square distance measured using separate estimates of the variance-covariance matrix for each class. Lark (1994) proposed the choice between LDA and QDA in remote sensing classification. He suggested QDA is preferred when the variance-covariance matrices of different classes are heterogeneous or LDA when the sampling error of the separate class variance-covariance matrices is high.

2.6.4 Nonparametric classifiers

2.6.4.1 Decision tree

Decision tree classifier (Friedl and Brodley, 1997) is a hierarchically based classifier which can perform automatic feature selection and complexity reduction. The tree structure is easily understandable and interpretable information (Figure 2.5). A decision tree divides a data set into smaller subdivisions on the basis of a set of tests defined at each branch (or node) in the tree. This procedure acquires three steps such as splitting nodes, determining which nodes are terminal nodes, and assigning class label to terminal nodes. It is very straightforward to assign class labels to terminal nodes in which labels are located following a majority vote or a weighted vote when it assumed that certain classes are more likely than others (Pal and Mather, 2001).

A tree consists of a root node, a set of internal nodes (splits), and a set of terminal nodes (leaves). Each node in a decision tree has only one parent node and two or more descendant nodes. A data set is classified by moving down the tree and sequentially subdividing it according to the decision framework defined by the tree, and a class label is assigned to each observation following the leaf node into which the observation falls (Friedl and Brodley, 1997).

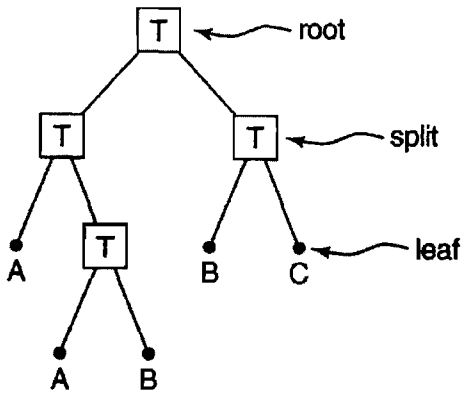


Figure 2.5: A decision tree classifier. Each box is a node at which test (T) are applied to recursively split the data into small groups. The label (A, B, C) at each leaf node refer to the class label assigned to each observation (Friedl and Broley, 1997).

Friedl and Brodley (1997) introduced details about these classification algorithms including a univariate decision tree, a multivariate decision tree, and a hybrid decision tree within a single decision tree structure. Results from their analysis also showed that decision tree algorithms were more accurate than maximum likelihood and discriminant classifiers.

Decision tree classifiers have been commonly applied to remote sensing data (Suarez and Lutsko, 1999; Simard *et al*, 2000; Pal and Mather, 2001; McIver and Friedl, 2002). The decision tree approach does not depend on any a prior statistical assumption and its rules are explicit and allow for identification of features which are relevant to distinguish specific classes (Simard *et al*, 2000). Especially, decision trees are nonparametric and do not require assumptions regarding the distribution of the input data. They use non-linear relations between features and classes, allow for missing values, and can deal with both numeric and categorical input in a natural way. However the disadvantage of this classifier is that the accuracy depends on the design of the decision tree and the selected features (Friedl and Brodley, 1997).

2.6.4.2 Artificial neural network

Artificial neural networks (ANNs) are computer programs designed to simulate human learning processes through establishment and reinforcement of connections between input data and output data (Campbell, 2002). These connections (or pathways) form the analogy with the human learning process in the brain, in that repeated relationships between input and output in the training process consolidate linkages, which are then employed to connect input and output, in the absence of training data. ANNs have been studied for many years in many fields such as speech and hand writing recognition and in pattern recognition, *etc*. In recent years, the use of ANNs in pattern recognition applied to remotely sensed images has significantly increased due to some important advantages over conventional classifiers, especially the freedom from distribution assumption and ability to integrate ancillary data acquired at a low level of measurement precision (Zhang and Foody, 2001; Kavzoglu and Mather, 2003).

There are many types of ANNs namely feed-forward back-propagation multi-layer perceptron (MLP), radial basis function (RBF) neural network, ARTMAP, probabilistic neural network (PNN) introduced in literature (Aleksander and Morton, 1990; Bishop,

1995; Petterson, 1996; Atkinson and Tatnall 1997; Tso and Mather, 2001) but the most commonly ones used in remotely sensed image classification are the feed-forward back-propagation MLP and RBF. The characteristics and applications of these two types are presented in the following sections.

2.6.4.2.1 Multi-layer perceptron

The MLP using the back-propagation learning algorithm is one of the most widely used neural network models. A neural network contains a number of interconnected nodes (equivalent to biological neurons). Each node is a simple processing unit that responds to the weighted inputs it receives from other nodes (Atkinson and Tatnall, 1997; Zhang and Foody, 2001). The architecture of this network used for image classification can be simply illustrated in Figure 2.6.

The feed-forward back-propagation MLP is generally composed of three types of layer. Each layer consists of processing nodes that are fully connected to each other, except that there are no interconnections between nodes within the same layer. These layers are the input, hidden and output layers, respectively. An MLP neural network architecture usually contains one input layer, one or two hidden layers and one output layer. The input layer comprises a set of nodes corresponding to individual data sources

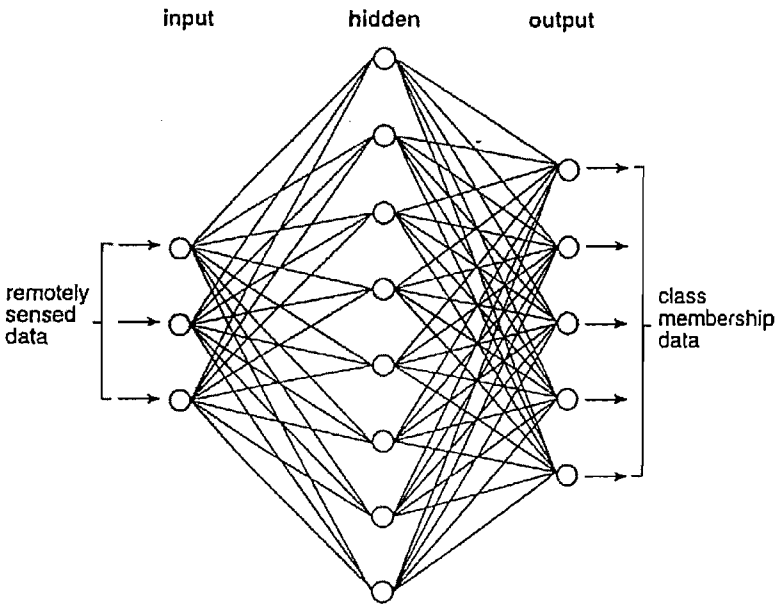


Figure 2.6: An MLP neural network architecture for image classification (Zhang and Foody, 2001).

(e.g. different wavebands of remotely sensed data). There is no computation performed in the input layer, unlike the other layers. Hidden layers are used for computations, and the values associated with each node are calculated from the sum of the multiplications between input node values and the weighted links connected to that node. The number of hidden layers is usually determined by the analyst (there may be more than one hidden layer in complex networks). The output layer produces the results of the classified image. Note that the number of nodes in the input layer and the output layer is determined by the number of discriminating variables (e.g. wavebands) and classes, respectively. The number of hidden nodes is usually determined subjectively on the basic of trial runs (Zhang and Foody, 2001; Kavzoglu and Mather, 2003).

Once the remote sensing data are put into the input layer as input signals, the input signals are passed to the nodes in the next layer in a feed-forward manner. As the signal passes from node to node, it is modified by the weights associated with the connection. The receiving node sums the weighted signals from all nodes to which it is connected in the preceding layer. The input that a node receives is weighted from (Atkinson and Tatnall, 1997; Zhang and Foody, 2001),

$$net_j = \sum \omega_{ji} o_i \tag{2.12}$$

where ω_{ji} represents the weight for the connection between node i and node j, and o_i is the output from the node i,. The output from a given node j is then computed by the node 'activation function' (usually a sigmoid activation function or hyperbolic tangent) such as (Kanellopoulos and Wilkinson, 1997),

$$o_j = f(net_j) = \frac{1}{1+\exp(-net_j + \theta_j)} \qquad \text{[sigmoid function]} \tag{2.13}$$

$$o_j = f(net_j) = m\tanh(k(net_j)) \qquad \text{[hyperbolic tangent]} \tag{2.14}$$

where θ_j, m, and k are constants, f stands for an activation function that is used to the weighted sum of inputs before the signal passes to the next layer.

A common learning algorithm used for ANN classification is back-propagation (Tso and Mather, 2001; Zhang and Foody, 2001). With this, training pixels are presented to the network via the input layer and the signals are fed-forwards through the network

using equations (2.12), and (2.13) or (2.14). In the output layer, network outputs are compared with the target outputs (training pixels) and the error computed (Atkinson and Tatnall, 1997; Zhang and Foody, 2001). This error is then back-propagated through the network and, generally weights are updated during training with the generalized delta rule (Kanellopoulos and Wilkinson, 1997; Atkinson and Tatnall, 1997),

$$\Delta\omega_{ji}(n+1) = \eta(\delta_j o_i) + \alpha\Delta\omega_{ji}(n) \tag{2.15}$$

where $\Delta\omega_{ji}(n+1)$ is the change of a weight connecting nodes *i* and *j*, in two successive layers, at the *(n+1)th* interaction, δ_j is the rate of change of error with respect to the output from node *j*, η is the learning rate parameter, and α is the momentum parameter.

This process of feeding forward signals and back-propagating the error is repeated until the total error of the network is minimized or reaches an acceptable magnitude. After training, the network can be used for classification of unknown class membership. The outputs from an ANN exist as activation levels ranging from 0 to 1. Thus, a pixel is allocated to the class with which it has the highest activation level.

The use of ANNs is more complicated than statistical classifiers, basically due to problems met in their design and implementation. Training a neural network requires that the users defines the network structure and sets the learning parameters. For the network's design, the identification of the number and size of the hidden layer(s) is important for the network capability to learn and generalize. In addition, the choice of appropriate values for network parameters also has a major influence on the performance of the learning algorithm. Significant parameters to be specified are: the range of the initial weights, the learning rate, the value of the momentum term, and the number of training iterations, all of which relate to the issue of when and how to stop the training process (Kavzoglu and Mather, 2003).

Kavzoglu and Mather (2003) have investigated the major issues in the design and training of ANNs for image classification. Attention was paid to the components of the network structure (input, hidden and output layers) and to the learning parameters (initial weight range, value of learning rate and momentum terms. The results showed that classification accuracies achieved through the use of ANNs can differ significantly,

depending on the selection of the network structure and parameter values. The authors suggested a number of useful guidelines for the effective design and use of ANNs in the classification of remotely sensed image data. They also stated that ANNs can perform well for small training samples and in the same small training sample ANNs perform much better than other statistical classifiers such as maximum likelihood in terms of classification accuracy.

2.6.4.2.2 Radial basis function networks

The MLP is just one of a set of networks that may be used for image classification; another is the radial basis function (RBF) network (Bishop, 1995; Bastos *et al.*, 1999; Pasika, *et al.*, 1999; Bruzzone and Prieto, 1999; Foody, 2001). The RBF differs from the MLP in many ways but mainly in its ability to divide feature space locally. This characteristic makes the RBF network less likely to make untenable extrapolations and indicates the potential to identify atypical cases by post-classification analysis of the network outputs (Foody, 2001).

The architecture of RBF network classifier is composed of three layers (Figure 2.7): one input layer, one hidden layer, and one output layer. The input layer relies on as many neurons as input features. Input neurons just create input features to the next layer. Each neuron in the hidden layer is related to a kernel function $\Phi_j(.)$ (commonly Gaussian function), created by a mean μ_j and a variance σ_j. The output layer is consisted of as many neurons as classes to be classified. Each output neuron o_l calculates a simple weighted summation over the responses of the hidden neurons for a given input pattern $\underline{x}_i = (x_1, \cdots, x_n)$, which is to be labelled to one of c different land-cover classes $\Omega = \{\omega_1, \omega_2, \cdots, \omega_c\}$ (Bruzzone and Prieto, 1999; Bruzzone *et al*, 1999):

$$o_l(\underline{x}_i) = \sum_{j=1}^{s} \omega_{lj} \Phi_j(\underline{x}_i) + \omega_{bias,l} \tag{2.16}$$

where s represents the number of hidden neurons, ω_{lj} is the weight associated with the connection between the kernel function $\Phi_j(.)$ and the output neurons o_l, and $\varpi_{bais,l}$ is the bias of the output neuron o_l.

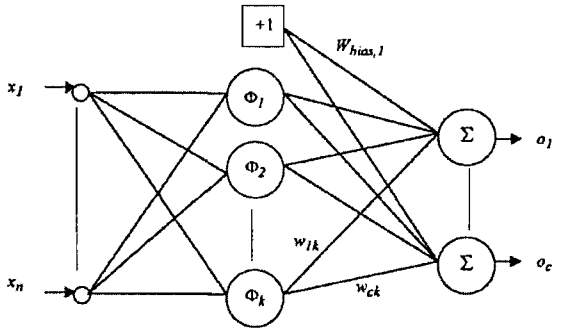


Figure 2.7: An RBF neural network architecture (Bruzzone and Prieto, 1999).

The training of the RBF neural classifiers is usually implemented in two stages with each stage relatively independent of the other. The first stage is training the hidden layer by selecting the centres and widths of the kernel functions related to the hidden units (Bishop, 1995). This selection is mainly implemented by applying a clustering technique such as the *k*-means clustering algorithm. When the centres of the kernel functions have been chosen, the widths of these functions are selected. The values of widths control the generalization abilities of the network (Bruzzone and Prieto, 1999). The width of a given kernel function can be selected as the standard deviation measured over all training samples in the cluster concerned with the kernel function considered. The second step of the training process is calculating the weights corresponding to the connection between the hidden units and the output units. This stage is basically done by minimizing a *sum-of-squares* error function (Bishop, 1995; Bruzzone and Prieto, 1999). In fact, a *sum-of-squares* error function is a quadratic function of the weights and its minimization can be identified in terms of the solution of a set of linear equations calculated by a pseudo-inverse matrix.

2.6.4.3 Support vector machine

Support vector machines (SVMs) (Vapnik, 1995; Hermes *et al*, 1999; Perkin *et al*, 2001; Hastie *et al.*, 2001; Huang *et al*, 2002; Pal and Mather, 2003; Foody and Mathur, 2004) have considerable potential as classifiers of remotely sensed imagery. They use a method to construct a learning machine based on statistical learning theory and produce non-linear boundaries by constructing a linear boundary in a large, transformed version of the feature space (Hastie *et al.*, 2001).

In the two linearly separable classes, a SVM classifier identifies a hyperplane that maximises the distance from the numbers of each class to the optimal hyperplane (Figure 2.8). A separating hyperplane refers to a plane in a multi-dimensional space that separates the data samples of two classes. The optimal separating hyperplane is a separating hyperplane that maximizes the margin from closest data points to the plane (Huang *et al*, 2002).

Let the training data with k samples be represented by (x_1, y_1), ..., (x_k, y_k) where $x \in R^n$ is an n-dimensional space, and $y \in \{-1,+1\}$ is the class label. A hyperplane can be defined by the equation $w.x + b = 0$, where x is a point lying on the hyperplane, w is normal to the hyperplane, b is the bias, and $(|b|)/(\|w\|)$ is the perpendicular distance from the hyperplane to the origin, with $\|w\|$ as the Euclidean norm of w (Foody and Mathur, 2004). Suppose the two classes can be separated by two hyperplanes parallel to the optimal hyperplane (Figure 2.8 a) (Huang *et al*, 2002):

$$w.x_i + b \geq 1 \quad \text{for } y_i=1,\ i=1,2,\ldots,k \tag{2.17a}$$

$$w.x_i + b \leq -1 \quad \text{for } y_i=-1 \tag{2.17b}$$

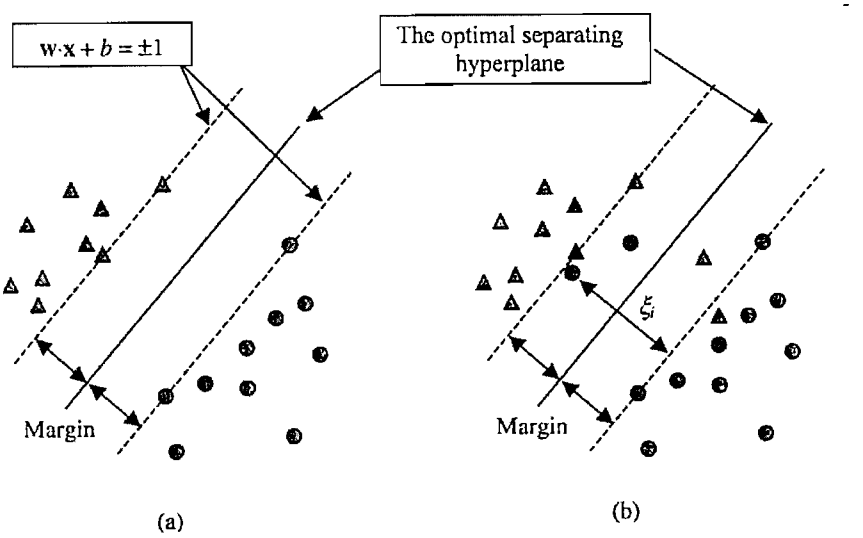


Figure 2.8: The optimal separating hyperplane between (a) separate data samples and (b) non-separate data samples (Huang *et al.*, 2002).

The purpose is to find a hyperplane that divides the data so that all the points with the same label lie on the same side of the hyperplane. This can be used to define w and b so that

$$y_i(w.x_i + b) > 0 \tag{2.18}$$

If a hyperplane exists that satisfies (2.18), the two classes are supposed to be linearly separable. Thus, it is possible to rescale w and b as

$$\min_{1 \le i \le k} y_i(w.x_i + b) \ge 1 \tag{2.19}$$

That is, the distance from the closest point to the hyperplane is $1/\|w\|$. Then (2.18) can be written as,

$$y_i(w.x_i + b) \ge 1 \tag{2.20}$$

The training data points on these two hyperplanes are called support vectors and are central to the optimal separating hyperplane (OSH). Since the distance to the closest point is $1/\|w\|$, the OSH can be identified by minimizing $\|w\|^2$ under constraint (2.20). The minimisation procedure uses Lagrange multipliers and Quadratic Programming (QP) optimisation methods. If λ_i, $i = 1, k$ are the non-negative Lagrange multipliers related to constraint (2.20), the optimisation problem becomes one of maximising (Pal and Mather, 2003):

$$L(\lambda) = \sum_i \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j (x_i . x_j) \tag{2.21}$$

with $\lambda_i \ge 0$, $i = 1, \ldots, k$.

If $\lambda^a = (\lambda_1^a, \ldots, \lambda_k^a)$ is an optimal resolution of maximisation problem (2.21) then the OSH can be computed as:

$$w^a = \sum_i y_i \lambda_i^a x_i \tag{2.22}$$

The support vectors are the points for which $\lambda_i^a > 0$ when the equality in (2.20) holds.

An important assumption to the above solution is that the data are linearly separable in the feature space. Commonly, the classes are not linearly separable, and the constraint (2.20) cannot be satisfied. To solve this problem, a slack variable ξ_i, $i=1, k$ is introduced to indicate the distance the case is from the optimal hyperplane (Figure 2.8b) with $\xi_i \geq 0$ such that (2.20) becomes

$$y_i(w.x_i + b) \geq 1 - \xi_i \tag{2.23}$$

And the solution to find a generalised OSH, also called a soft margin hyperplane, can be found using the following conditions (Pal and Mather, 2003):

$$\min_{w,b,\xi_1,\ldots\xi_k} \left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{k} \xi_i \right] \tag{2.24}$$

$$y_i(w.x_i + b) \geq 1 - \xi_i \tag{2.25}$$

$$\xi_i \geq 0 \qquad i = 1,\ldots,k \tag{2.26}$$

The first term in (2.24) is similar to the linearly separable case, and controls the learning capacity, while the second term controls the number of misclassified points. The parameter C is chosen by the user. Larger values of C indicate the assignment of higher penalty to misclassification errors.

Where it is not possible to have a hyperplane defined by linear equations on the training data, the basic approach outlined above for linearly separable data can be extended to allow for non-linear decision surfaces (see e.g. Huang *et al.*, 2002; Pal and Mather, 2003; Foody and Mathur, 2004).

In the above theoretical development, the SVM was designed for binary classification (i.e. one SVM can only separate two classes), whereas most remote sensing applications involve multiple classes. Thus, integration strategies are needed to extend them to classifying multiple classes (Gualtieri and Cromp, 1998; Huang *et al*, 2002; Foody and Mathur, 2004).

The SVM classifiers were more accurate than other comparable classifiers such as disciminant analysis, decision tree, and neural network (Foody and Mathur, 2004). In a

comparative study, Huang *et al.* (2002) also indicated that SVM has the highest accuracies, followed by ANN, decision tree, and MLC.

2.6.4.4 Evidential reasoning

The mathematical theory of evidence is a field in which a number of data sources can be combined to generate a joint inference concerning pixel labelling (Tso and Mather, 2001). The theory was first introduced by Dempster in the 1960s and later extended by Shafer (1976), who provided details of the development of evidential theory. Thus, this theory has become known as the Dempster-Shafer (D-S) theory of evidence. D-S theory is a valuable tool for dealing with the problems associated with analysis, integration, and classification (Peddle, 1995a; Tso and Mather, 2001). Moreover, D-S theory is also a means of distinguishing between uncertainty and ignorance (Lein, 2003). Several researchers have applied this theory for remotely sensed image classification in the literature. Lee *et al.* (1987) applied general methods of evidential calculus for multisource classification. Peddle (1995a;1995b) using D-S theory of evidence for supervised evidential classification by deriving evidence from histogram bin transformation of supervised training data frequency distributions. The algorithms of evidential reasoning approaches are based on D-S theory of evidence. Thus, the basic concepts of D-S theory of evidence are briefly described with the applications to evidential reasoning classification as following.

For a given pixel, the task of the classification is to assign the pixel to one member within a set of classes. In theory of evidence, the set of all possible classes is referred to as the frame of discernment which is denoted by the symbol Θ. Suppose that an analyst is trying to classify an image, which involves labelling pixels as belonging to one of three classes $\{A, F, U\}$, where A is agriculture, F is forest, ands U is urban. In D-S theory, this set $\{A, F, U\}$ is called a frame of discernment. The number of all possible subsets of the frame of discernment Θ is equal to $2^{|\Theta|}$, where $|\Theta|$ denotes the number of one-element subsets (called 'singletons'). In our case, there are three singletons in $|\Theta|$ so the total number of subsets of Θ is $2^3 = 8$, as shown in Figure 2.9. Note that the empty set { } is one of these subsets and it does not display in Figure 2.9 (Tso and Mather, 2001).

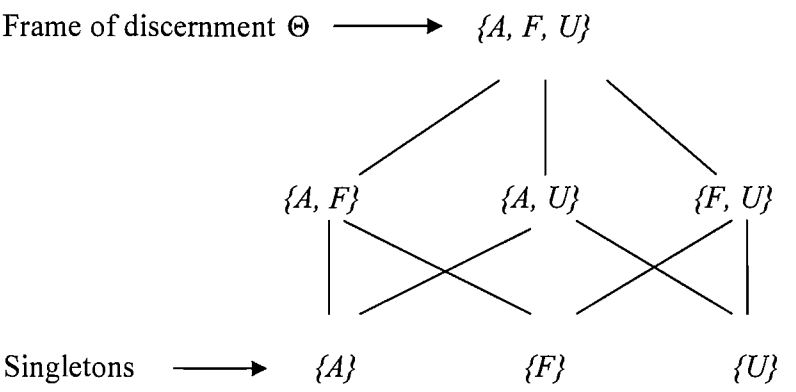


Figure 2.9: A representation of subsets of frame of discernment {A, F, U} created by agriculture, forest, and urban.

One important aspect of evidence theory is the basic probability assignment written as *m*. This basic probability assignment expresses the degree to which all available and relevant evidence supports the claim that a particular element (e.g., a pixel) of the universal set belongs to a particular set such as a land cover class (a singleton) or a subset of the frame of discernment. Those subsets of the universal set are called focal elements of the basic probability assignment, and equate to the classes being mapped. The basic probability assignment assigns value in [0, 1] to every element of $2^{|\Theta|}$ such that the numeric values sum up to 1. Usually, $m(A) \geq 0$ is used to denote the value assigned to subset A. For a basic probability assignment of a focal element A, there will be a belief function $Bel(A)$ and plausibility function $Pl(A)$. They can be calculated as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{2.27}$$

$$Pl(A) = \sum_{B \cap A \neq \phi} m(B) \tag{2.28}$$

with the special cases:

1) When $A = A_i$ is a singleton, then $Bel(A) = Bel(A_i) = m(A_i)$.

2) When $A = \Theta$, $Bel(\Theta) = 1$

where $\emptyset$ is the empty set, *Bel(A)* is the sum of the mass of evidence which is committed to class A as well as various subsets of A, and *Pl(A)* is not only the equivalence of *Bel(A)* but also the additional evidence related to sets overlapping with A (Klir and Folger, 1998; Zhang and Goodchild, 2002; Foody, 2002a).

Belief measures communicate a level of confidence in the range [0.0, 1.0], where 0.0 indicates no confidence and 1.0 explain absolute confidence in a given conclusion. Thus, belief measures can be summed to produce overall expression of certainty (Lein, 2003). When a set of evidence has been assembled for each pixel of the image classified, the evidence needs to be combined (Peddle, 1995a). Suppose the evidence obtained from two independent sources and expressed by two basic probability assignment m_1 and m_2. The two frames of evidence can be combined using Dempster's orthogonal sum, or rule of combination to get a joint basic assignment $m_{1,2}$ by:

$$m_{1,2}(A) = m_1 \oplus m_2(A) = k \sum_{X \cap Y = A, A \neq \emptyset} m_1(X).m_2(Y) \tag{2.29}$$

For all $A \neq \phi$, X and Y are any subsets of the universal set, and $m_{1,2}(\phi) = 0$, where k is a normalizing constant:

$$k^{-1} = 1 - \sum_{X \cap Y = \emptyset} m_1(X).m_2(Y) \tag{2.30}$$

To apply evidential reasoning to classification, a measure of evidence for subsets of class labels is computed. The requirements for a supervised evidential classifier include (1) the representative information for each class to base classification making is acquired. This requirement is implemented by using standard training data identified for each class; (2) converting this information into measure of evidence by individual class proportion (Peddle, 1995a). Assuming that focal elements consist of class label singletons, the procedures to combine two sources of data are quite simple.

However, remote sensing image data do not provide direct measures of evidence for input to the mathematical theory of evidence and therefore a separate process to derive these measures is required (Peddle, 1995a; 1995b). They can be computed with respect to the frequency of occurrence of individual pixel values from training samples over the entire set of classes or frame of discernment. Training data are read from each data feature in sequence, and a frequency distribution of training values is interpreted for

each class. Thus, for c classes and k sources, there will be a total of $c \times k$ frequency distributions (Peddle, 1995a). When a set of evidential measures has been assembled for a pixel in a multisource data set, the evidence from all sources is combined by repeating application of Dempster's orthogonal sum into a manageable, one-dimensional format containing one measure of support per class. Then, a decision rule is applied to the mass function to classify the pixel into one of the classes within the frame of discernment. The decision rule may be based on maximum belief function or maximum plausibility function, where the class with the greatest support is assigned as the pixel label.

Peddle (1995b) proposed a MERCURY software package as a multisource evidential reasoning classification software system based on the D-S theory of evidence. In this software, measures of both support and plausibility are included in the decision rule. Class allocation is based on the sum of support and plausibility for each class, with the pixel assigned as the class with the greatest sum.

2.7 The mixed pixel problem

Most remotely sensed image classification methods are implemented under the hypothesis that each pixel is perfectly pure (e.g. consists of one land cover class). If the spatial resolution of the image pixels is coarse in comparison with the variability of land cover objects then a single pixel may contain more than one land cover class in the area that it represents. Such pixels are called mixed pixels (Cracknell, 1997; Fisher, 1997; Tso and Mather, 2001). There are four main reasons which may lead to mixed pixel (figure 2.10): (1) Boundaries between two or more mapping units (e.g. field-woodland boundary), (2) The intergrade between central concepts of mappable phenomena (ecotone), (3) Linear sub-pixel objects (e.g. a road), or (4) Small sub-pixel objects (e.g. a house or tree).

The DN value recorded for a pixel may be derived in a number of different ways. It may be the majority of area of the pixel; represent a value at some systematic location within the area of the pixel; or be considered representative for some other reason (Fisher, 1997). Manslow (2000) suggested that the reflectance information in a pixel can be described by the point spread function of the sensor. This means that the pixel's reflectance information tends to be most similar to the reflectance of land cover located towards the centre of pixel's ground area, and least similar to cover towards its edge.

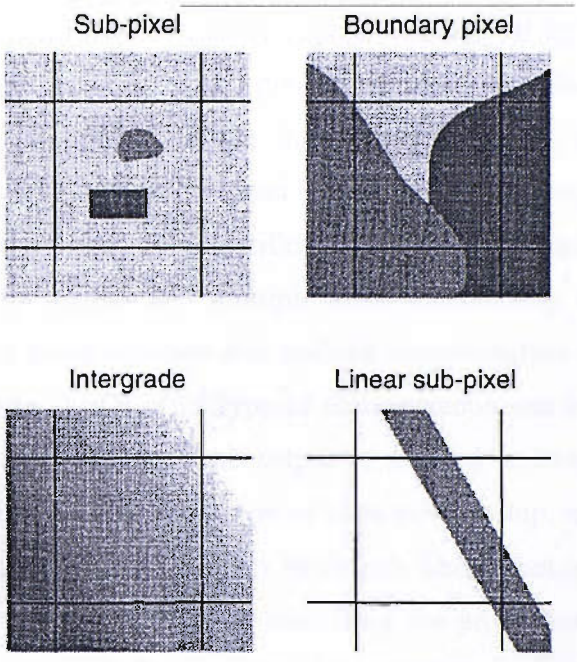


Figure 2.10: Four causes of mixed pixels (Fisher, 1997).

Thus, when a pixel area consists of two or more areas that differ greatly from brightness, its value may consist of several very different values, and the single digital value which represents the pixel may not accurately represent any of the classes present.

As in Figure 2.10, mixed pixels often occur at the edges of large parcels, or along linear features, such as rivers or roads, where contrasting brightnesses are immediately adjacent to one another. The edge, or border, pixels then cause errors in image classification. Especially the difficult situation can be created by landscape composed of many parcels that are small relative to the spatial resolution of the sensor. A mosaic of such parcels will create an image, all formed by mixed pixels. The number and proportion of mixed pixels in an image typically increases with a coarsening of the spatial resolution of the sensor and/or increasing fragmentation of the landscape to be represented (Campbell, 2002, Foody, 2002b).

2.8 Soft classification

The techniques described in section 2.6 are concerned with 'hard' pixel labelling. All of the above classification methods require that each individual pixel is given a single, unambiguous label. They were developed for the classification of classes that may be

considered to be discrete and mutually exclusive, and assume that each pixel is perfectly pure (i.e. the ground area represented by a pixel is occupied by a single class). Such approaches are inappropriate for mixed pixels problem, which discussed in section 2.7. The fact discussed in section 2.7 was that mixed pixels are the main problem in mapping land cover from remotely sensed imagery and such image may be dominated by mixed pixels. Thus, a fuzzy/soft classification (or sub-pixel classification) is required and should allow for partial and multiple class membership of mixed pixels and therefore can give a more accurate and realistic representation of many land covers (Foody, 1996a; Bastin, 1997). This type of representation can be derived by using a fuzzy classification or 'softening' the output of a 'hard' conventional classification, which means that the measures of strength of class membership, rather than just code of the most likely class of membership may be output. The output of soft classification is typically a set of fraction images each describing the proportion of a particular land cover class within each pixel. The number of fraction images is equal to the number of target classes defined to map the land cover. These images are more informative and appropriate depicting land cover than those derived from the conventional hard classification.

There are many methods of soft classification such as spectral mixture analysis, fuzzy set theory, artificial neural network, evidential reasoning, *etc.* Focusing on remotely sensed image classification based approaches, the use of fuzzy classifiers and softening of 'hard' classifications are the most commonly used approaches for the derivation of a soft thematic map (Foody, 2002a). Thus, this research will introduce and apply some methods of these two main approaches.

2.8.1 Stages of soft classification

As with their hard counterparts that were described in section 2.6.1, a soft classification also has three basic stages such as training stage, class allocation stage, and accuracy assessment stage after the application of any necessary pre-processing operations.

2.8.1.1 Training stage

As stated in the hard classification section, training stage is the stage in which pixels of known class membership are identified and used to characterise the classes in the image.

The purposes and the procedures of training stage in soft classification are similar to hard classification. However, they may only be different in selecting the type of training pixels. For hard classification, only pure pixels should be selected for training. This typically involves selecting training sites from only very large homogeneous regions of each class to avoid contamination of training sites by other classes. However, training sets for soft classification can be performed either in pure or mixed pixels. Some approaches to fuzzy training are to feed the network of ANN classifier with the proportional set of each class in the training pixels (Foody, 1997; Bernard *et al.*, 1997; Zhang and Foody, 2001). An empirical study is presented by Bernard *et al.*, (1997) to test training procedures with neural network for soft classification suggested that land cover mixtures are best recognized by training with two-component mixed pixels (pure and mixed pixels). Foody and Arora (1996) used mixed pixels in all three stages of MLC and ANNs. The results show that by accommodating for mixed pixels in the classification, more accurate, appropriate and useful outputs may be derived.

2.8.1.2 Class allocation stage

In the class allocation stage of a conventional supervised classification, the characteristics of the classes that are defined by training stage are used to allocate pixels of unknown class membership to the class with which they have the greatest similarity. For example, in MLC, each pixel is assigned to the class with which it has the greatest posterior probabilities of membership. For ANN, each pixel is assigned to the class related to the unit in the output layer with the greatest activation level. Such class assignments are "crisp" and it leads to the waste of information on the relative strength of class membership and its partition between classes.

A soft classification allocation may be produced by outputting the measures of the strengths of class membership that are created in the classification. These measures can be used to represent for the proportional composition of the relevant classes and the spatial distribution of the classes represented by fraction images which display the class proportions in each pixel. For instance, in fuzzy MLC, the derived posterior probabilities of membership are considered to represent the proportional cover of the classes in each image pixel. For ANN, the measures of strength of class membership in each pixel are the activation levels of a neural network output unit (Foody and Arora, 1996). The details of these two soft classifiers will be described in later sections.

The output of soft classification is a set of fraction images each describing the proportion of a particular land cover class within a pixel. The number of fraction images is equal to the number of target classes defined to map the land cover map. For example, if there are four land cover classes to be used to classify an image by soft classification, the results will be four fraction images. Each fraction image represents the strengths of class membership of that pure class in each pixel.

2.8.1.3 Accuracy assessment

As proposed in section 2.6.2.3, the measures derived from traditional confusion matrix are the most widely used ones to assess the accuracy of image classification. Unfortunately, in their present form they are only appropriate for use with a 'hard' classification since these measures of classification accuracy may be used when each pixel is assigned to only one class in the classification and only one class in the ground data (Congalton, 1991; Foody, 1996a). Consequently, a class labelling is judged exactly right, or exactly wrong. Therefore, the measures of classification accuracy derived from the confusion matrix are inappropriate for the evaluation of soft classification. The accuracy of the representation provided by a fuzzy classification is, however, difficult to evaluate (Foody, 1996a; Binaghi *et al.*, 1999). A number of approaches have been proposed with emphasis on fuzzy measures (Foody, 1995; Foody, 1996a; Arora and Foody, 1997; Binaghi *et al.,* 1999; Atkinson, 1999; Woodcock and Gopal, 2000; Lewis and Brown, 2001; Oki *et al.*, 2004). They are generally implemented by comparing the actual and predicted class of membership for a set of pixels not used in training the classification. The following sections represent some of the most common use approaches for the evaluation of soft classification.

2.8.2 Some approaches of accuracy assessment for soft classification

2.8.2.1 Root mean square error

The mean error (ME), root mean square error (RMSE) are statistics that are used commonly to calculate the level of agreement between a set of known fraction (reference data) y and a set of predicted fraction (classification data) z. They are the two simplest measures of agreement between y and z and computed as:

$$ME_j = \frac{1}{n}\sum_{i=1}^{n}(y_{ij} - z_{ij}) \tag{2.31}$$

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^{n}(y_{ij} - z_{ij})^2}{n}} \tag{2.32}$$

where, j is the j^{th} class and n denotes the total number of pixels in the whole testing data set, y_{ij} is the proportion of class j in pixel i from the fuzzy reference data, and z_{ij} is the proportion of class j in pixel i from the fuzzy classification. Thus, the overall mean error and RMSE are calculated as:

$$ME = \frac{1}{c.n}\sum_{j=1}^{c}\sum_{i=1}^{n}(y_{ij} - z_{ij}) \tag{2.33}$$

$$RMSE = \frac{1}{c}\sum_{j=1}^{c}\sqrt{\frac{\sum_{i=1}^{n}(y_{ij} - z_{ij})^2}{n}} \tag{2.34}$$

where, c is the number of classes.

The ME shows bias and RMSE shows accuracy (bias and precision). For the evaluation of soft classification, a low relative ME or RMSE indicates well classified and a high relative ME or RMSE indicates poor classified. The ME and RMSE for each class between fuzzy classification output and fuzzy reference data may also be used to indicate the accuracy of individual classes.

However, there are some disadvantages with RMSE. The first disadvantage with RMSE is that it confounds two separate information dimensions into a single value. The second problem with RMSE is that it is not standardized by any measure of variance; therefore, if the variance of the estimated data is large, the RMSE may be large also, and *vice versa*. If no correlation between the actual and predicted proportions, the RMSE can be small. Consequently, comparison between RMSE is problematic, even for different classes of the same data (Atkinson, 1999).

2.8.2.2 Correlation coefficient

The correlation coefficient per class (r) is a measure of the correlation between the proportions of corresponding memberships of reference and classification data by means of correlation. It is can computed from,

$$r_j = \frac{c_{yj\,zj}}{s_{yj} . s_{zj}} \tag{2.35}$$

where $c_{rj\,cj}$ is the covariance between y and z for class j calculated as:

$$c_{rjcj} = \frac{\sum_{i=1}^{n} (\bar{y}_j - y_{ij})(\bar{z}_j - z_{ij})}{n-1} \tag{2.36}$$

and s_{yj} and s_{zj} are standard deviations of y and z for class j calculated as:

$$s_{yj} = \sqrt{\frac{\sum_{i=1}^{n} (\bar{y}_j - y_{ij})^2}{n-1}} \tag{2.37}$$

$$s_{zj} = \sqrt{\frac{\sum_{i=1}^{n} (\bar{z}_j - z_{ij})^2}{n-1}} \tag{2.38}$$

When coefficient r is large ($r \rightarrow 1$), it indicates close association between a target and predicted data set. The r for each class between fuzzy classification output and fuzzy reference data may also be used to indicate the accuracy of individual classes. The larger the r the more accurate the classification of the specific class considered.

One problem with r is that it could be large showing close association between a target and predicted data set while the RMSE is large; the error of the estimation is large. Especially, r could be large when the predictions are biased, with z_j being some fraction of y_j; therefore, r shows precision, and it is normalized by the two variances, but it is not sensitive to bias. In addition, the predictions could be accurate with r=0 if the variances in the two distributions are small (Atkinson, 1999).

2.8.2.3 Entropy (*H*)

Entropy is a measure of uncertainty and information formulated in terms of probability theory, which expresses the relative support associated with mutually exclusive alternative classes (Foody, 1996a). It may be used to express the variations in class membership probabilities associated with each pixel. Especially, it is an indicator of classification quality in case where ambiguity exists since it shows the degree to which the class membership probabilities are occupied between the defined classes. The entropy H_i of ith pixel can be calculated from class membership probabilities within that pixel as following:

$$H_i = -\sum_{j=1}^{c} (z_{ij}) \log_2 (z_{ij}) \tag{2.39}$$

The choice of logarithm base is customized, but the logarithm base 2 is commonly used. With the logarithm base 2 the entropy is measured in bits (Klir and Folger 1988). Entropy may be generated only for situations in which all the possibilities of class membership are >0, which is likely in a truly fuzzy classification; if appropriate the data may be re-scaled to ensure this requirement is met.

Entropy is maximized when the probability of class membership is equal between all defined classes in the classification and minimized when it consists of one class. It may be used to indicate the confidence that may be associated with classification outputs – with pixels showing a low relative entropy assumed to be well classified and those with a high relative entropy poorly classified. Its value as an indicator of classification accuracy is, therefore, based on the hypothesis that in an accurate classification each pixel will have a high probability of membership with only one class. Thus, Foody (1995) suggested that the interpretation of entropy calculated for each pixel is more difficult when a mixed land cover composition is possible. Entropy, therefore, does not provide an assessment of how closely the fuzzy classification models the continuous classes. Further, entropy is only appropriate for situations in which the output of a classification is soft (i.e., the probabilities of membership to all defined classes are output for each pixel) and the reference data are crisp (i.e., the code of the single class of membership) (Foody, 1996a; Binaghi *et al*, 1999).

2.8.2.4 Cross-entropy (*d*)

Cross entropy (or directed divergence) is the measure of closeness of fuzzy classification where multiple and partial class membership is a feature of both remotely sensed data and ground data (Foody 1995). It is identified from the probability distributions of class membership of the ground data (y) and that of the fuzzy classification output (z). Cross entropy of a pixel can be computed from,

$$d_i(y,z) = -\sum_{j=1}^{c}(y_{ij})\log_2(z_{ij}) + \sum_{j=1}^{c}(y_{ij})\log_2(y_{ij}) \tag{2.40}$$

This provides a measure of closeness of the classification to the ground data. A small value of entropy could shows that the classification was an accurate representation of the land cover. Cross-entropy, therefore, may be used as an indicator of the quality of the classification output and was most apparent for pixels which consisted of significant proportions of two or more classes.

2.8.2.5 Fuzzy error matrix

Binaghi *et al.* (1999) proposed a method called fuzzy error matrix that uses the fuzzy set theory based on the Minimum operator to extend the applicability of the traditional error matrix method to the accuracy assessment of soft classifiers. It is designed to deal with those situations in which classification and/or reference data are represented in multi membership form and the strengths of membership show different levels of approximation ambiguous classes. The method assumes that membership values in classes are known for the reference data set. According to the basic concepts of the traditional error matrix (as described in detail in section 2.6.2.3), Binaghi *et al.* (1999) formalize these concepts and the matrix building procedure in terms of classical set theory and derived set operations. The layout of fuzzy error matrix is similar to the traditional error matrix with the exception that elements of a fuzzy error matrix can be any non-negative real numbers instead of non-negative integer number. The elements of the fuzzy error matrix represent class proportion corresponding to reference data (i.e., fuzzy reference data) and classified outputs (i.e., fuzzy classified image) respectively.

Let R_n and C_m be considered fuzzy sets of reference data assigned to class n and of classification data assigned to class m, respectively, with $1 \le n \le c$, $1 \le m \le c$ and c as the number of classes. $\{R_n\}$ and $\{C_m\}$ form two fuzzy partitions of the sample data set X. R_n and C_m have the membership function

$$\begin{aligned} \mu_{R_n} &: X \to [0,1] \\ \mu_{C_m} &: X \to [0,1] \end{aligned} \tag{2.41}$$

where [0,1] represents the interval of real numbers from 0 to 1 inclusive.

$\mu_{R_n}(x)$ and $\mu_{C_m}(x)$ indicate the gradual membership of the same element x in class n and m as presented in the reference and classification data, respectively. Since, in the context of soft classification, these membership functions also represent the proportion of a class in the testing sample, which for fuzzy reference data may be represented as,

$$\sum_{n=1}^{c} \mu_{R_n}(x) = 1 \tag{2.42}$$

The fuzzy error matrix M will be provided by using fuzzy set operators within the error matrix. The value of the element $M(m,n)$ involves the calculation of the degree of membership in the fuzzy intersection set $C_m \cap R_n$. There are several different functions for the intersection operation proposed in the literature, in this algorithm the authors use the "min" operator introduced in the original formulation of the theory of fuzzy sets (Zadeh, 1977) as follows,

$$\mu_{C_m \cap R_n}(x) = \min(\mu_{C_m}(x), \mu_{R_n}(x)). \tag{2.43}$$

Thus, the value of the element $M(m,n)$ in row m and column n computed on the overall data set:

$$M(m,n) = |C_m \cap R_n| = \sum_{x \in X} \mu_{C_m \cap R_n}(x). \tag{2.44}$$

The fuzzy error matrix is shown in Table 2.1, in which $p_{m,n}$ represents the cardinality of the intersection set $C_m \cap R_n$ computed according to (2.43), p_{i+} and p_{+i} are the total assignment to the i^{th} class for classification and reference data, respectively.

Predicted data \ Actual data	1	2	....	c	Σ
1	p_{11}	p_{12}	⋯	p_{1q}	p_{1+}
2	p_{21}	p_{22}	⋯	p_{2q}	p_{2+}
⋮	⋮	⋮	⋯	⋮	⋮
c	p_{q1}	p_{q2}	⋯	p_{qq}	p_{q+}
Σ	p_{+1}	p_{+2}	⋯	p_{+q}	

Table 2.1: Fuzzy error matrix.

The fuzzy error matrix can be used as a foundation to derive some measures that can be used to evaluate the accuracy of soft classifiers, in the same way as the conventional error matrix. The simplest element in both cases is overall accuracy (OA). In the fuzzy case, it is computed by dividing the sum of the major diagonal by the total grades of membership found in reference data, considering OA as a measure of the total match between reference and classification data. The accuracies of the individual classes is computed by dividing the corresponding element of the major diagonal by the total grades of membership found in reference and classification data in either the corresponding column, or corresponding row. For each category we have the producer's accuracy (PA) and user's accuracy (UA), respectively. All these measures, OA, PA and UA, are limited in the range [0,1] and assume the value of 1 is the perfect match between the gradual membership of reference and classification data. To fully understand this method, an example is illustrated in the table 2.2.

Binaghi *et al.* (1999) also represented some hypothetical examples and the results show that the accuracy information the proposed procedure provides, consistently reflects how correctly the strength of class membership is partitioned among classes. Due to the correspondence of the fuzzy error matrix and the traditional error matrix, the use of fuzzy error matrix to evaluate fuzzy classification may therefore be more appropriate than other methods such as entropy and correlation coefficients. Moreover, the formulation of fuzzy error matrix can also be used to assess the accuracy of hard

classification given the crisp reference data. Thus, from the view of standardizing the accuracy assessment procedures for both hard and soft classification, fuzzy error matrix and the associated measures seem to be more suitable in assessing the quality of remotely sensed derived classifications.

	Actual data				
Predicted data	Water	Wetland	Other	Σ	User's accuracy
Complete matching (**OA=1**)					
Water	**111.58**	74.36	0	111.58	1
Wetland	74.36	**128.53**	0	128.53	1
Other	0	0	**0**	0	#
Σ	111.58	128.53	0		
Producer's accuracy	1	1	#		
ANN (**OA=0.74**)					
Water	**90.78**	86.59	0	116.25	0.78
Wetland	71.48	**87.16**	0	97.12	0.90
Other	26.34	26.21	**0**	26.75	0
Σ	111.58	128.53	0		
Producer's accuracy	0.81	0.68	#		
Fuzzy statistical classification (**OA=0.59**)					
Water	**24.34**	20.88	0	35.34	0.69
Wetland	97.12	**117.53**	0	204.66	0.57
Other	0	0	**0**	0	#
Σ	111.58	128.53	0		
Producer's accuracy	0.22	0.91	#		

Table 2.2: Three fuzzy error matrices with the first one is the complete matching case which designed to be easy to compare with the results obtain from other two fuzzy classifiers such as neural network classifier and fuzzy statistical classifier (Binaghi *et al*, 1999).

2.8.2.6 Generalized confusion matrix

As concerned about evaluating the accuracy of soft classification, Lewis and Brown (2001) proposed a method called a generalized area-based confusion matrix. They use a multiplication operator to form a generalized confusion matrix for assessing area estimates from remotely sensed data. The generalized confusion matrix is appropriate for both traditional and soft classification algorithms. In addition, a corresponding error matrix, derived from the generalized confusion matrix, may be used to assess

quantitatively the accuracy of the maps generated using soft classification algorithms and to compare with the accuracy of the traditional classification techniques.

For soft classification algorithms, the memberships in a target reference dataset containing n pixels and c classes can be represented in an $n \times c$ matrix, R as follows,

$$R = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1c} \\ R_{21} & R_{22} & \cdots & R_{2c} \\ \cdots & & \cdots & \cdots \\ R_{n1} & R_{n2} & \cdots & R_{nc} \end{bmatrix} \tag{2.45}$$

and $\sum_{j=1}^{c} R_{kj} = 1$ for $k = 1, \ldots, n$. Similarly, the membership estimated for the n pixels by the soft classification algorithms can be presented in an $n \times c$ matrix, C as follows,

$$C = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1c} \\ C_{21} & C_{22} & \cdots & C_{2c} \\ \cdots & & \cdots & \cdots \\ C_{n1} & C_{n2} & \cdots & C_{nc} \end{bmatrix} \tag{2.46}$$

and $\sum_{j=1}^{c} C_{kj} = 1$ for $k = 1, \ldots, n$.

The $c \times c$ area-based confusion matrix, P, can be derived from matrices C and R as follows,

$$P = C^T R = \begin{bmatrix} \sum_{k=1}^{n} C_{k1}R_{k1} & \sum_{k=1}^{n} C_{k1}R_{k2} & \cdots & \sum_{k=1}^{n} C_{k1}R_{kc} \\ \sum_{k=1}^{n} C_{k2}R_{k1} & \sum_{k=1}^{n} C_{k2}R_{k2} & \cdots & \sum_{k=1}^{n} C_{k2}R_{kc} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{k=1}^{n} C_{kc}R_{k1} & \sum_{k=1}^{n} C_{kc}R_{k2} & \cdots & \sum_{k=1}^{n} C_{kc}R_{kc} \end{bmatrix} \Rightarrow \begin{bmatrix} \sum_{k=1}^{n} C_{k1} \\ \sum_{k=1}^{n} C_{k2} \\ \cdots \\ \sum_{k=1}^{n} C_{kc} \end{bmatrix}$$

$$\Downarrow \tag{2.47}$$

$$\begin{bmatrix} \sum_{k=1}^{n} R_{k1} & \sum_{k=1}^{n} R_{k2} & \cdots & \sum_{k=1}^{n} R_{kc} \end{bmatrix}$$

where the row and column sum of the matrix P are also shown.

While the individual elements of this area-based confusion matrix cannot be interpreted as area, the row and column shown in (2.49) can be interpreted in this mode. The sum of the element in row i of the confusion matrix, P, is

$$\sum_{j=1}^{c} P_{ij} = \sum_{k=1}^{m}\sum_{j=1}^{c} C_{ki} R_{kj} = \sum_{k=1}^{m}\left(C_{ki} \sum_{j=1}^{c} R_{kj} \right) = \sum_{k=1}^{m} C_{ki} \tag{2.48}$$

This is the total area covered by class i in the target estimated data. Similarly, the sum of the element in column j represents the total area covered by class j in the target reference data.

Here, the individual elements of the confusion matrix cannot be interpreted straightforwardly since there is no concept of correct or incorrect classification. However, the row sums and column sums represent the estimated and the referenced areas of the classes, respectively, and theses values can be used to assess the accuracy of the area estimates.

For the ideal soft classification algorithm the referenced and estimated class areas are the same and the confusion matrix is symmetrical. This ideal confusion matrix is equivalent to the $c \times c$ reference matrix, T, calculated from the target reference data,

$$T = R^T R \tag{2.49}$$

In general, soft classification algorithms are not ideal and errors are apparent in a confusion matrix that is not symmetrical and differs from the ideal, symmetrical confusion matrix. The differences between the reference matrix, T, and the confusion matrix, P, therefore can form an error matrix, E derived from,

$$E = T - P = \begin{bmatrix} \sum_{k=1}^{m} R_{k1}(R_{k1} - C_{k1}) & \sum_{k=1}^{m} R_{k2}(R_{k1} - C_{k1}) & \cdots & \sum_{k=1}^{m} R_{kc}(R_{k1} - C_{k1}) \\ \sum_{k=1}^{m} R_{k1}(R_{k2} - C_{k2}) & \sum_{k=1}^{m} R_{k2}(R_{k2} - C_{k2}) & \cdots & \sum_{k=1}^{m} R_{kc}(R_{k2} - C_{k2}) \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{k=1}^{m} R_{k1}(R_{kc} - C_{kc}) & \sum_{k=1}^{m} R_{k2}(R_{kc} - C_{kc}) & \cdots & \sum_{k=1}^{m} R_{kc}(R_{kc} - C_{kc}) \end{bmatrix} \Rightarrow \begin{bmatrix} \sum_{k=1}^{m} (R_{k1} - C_{k1}) R_{k1} \\ \sum_{k=1}^{m} (R_{k2} - C_{k2}) R_{k1} \\ \cdots \\ \sum_{k=1}^{m} (R_{kc} - C_{kc}) R_{k1} \end{bmatrix}$$

$$\Downarrow \tag{2.50}$$

$$\begin{bmatrix} \sum_{i=1}^{c} E_{i1} = 0 & \sum_{i=1}^{c} E_{i2} = 0 & \cdots & \sum_{i=1}^{c} E_{ic} = 0 \end{bmatrix}$$

where the row and column sums are represented in the vectors that are also shown.

Values near to zero in the error matrix (for which C_{kj} similar to R_{kj}) indicate accurate area estimation and high (positive and negative) values indicate inaccurate area estimation by the classification algorithm. The ideal case, where the values of all

elements of the error matrix are equal to zero, is the case where the measures of class membership of estimated data are equal to that of the reference data for every pixel and every class.

Assuming $\sum_{j=1}^{c} R_{kj} = 1$ and $\sum_{j=1}^{c} C_{kj} = 1$ for $k = 1, \ldots, n.$, the sum of elements in column j of the error matrix is zero and the sum of the elements in row i of the error matrix is

$$\sum_{j=1}^{c} E_{ij} = \sum_{j=1}^{c}\left(\sum_{k=1}^{m} R_{kj}\left(R_{ki} - C_{ki}\right)\right) = \sum_{k=1}^{m}\left(R_{ki} - C_{ki}\right) \tag{2.51}$$

The error matrix can be used as a means to derive some measures of accuracy describing the proportion of area in error that can be used to evaluate the accuracy of soft classifiers. They are the overall proportion of area in error calculated all over classes is

$$OE = \frac{\sum_{i=1}^{c}\left|\sum_{j=1}^{c} E_{ij}\right|}{n} \tag{2.52}$$

and the proportion of area in error calculated for class i:

$$OE_i = \frac{\sum_{j=1}^{c} E_{ij}}{\sum_{k=1}^{m} R_{ki}} = \frac{\sum_{k=1}^{m}\left(R_{ki} - C_{ki}\right)}{\sum_{k=1}^{m} R_{ki}} \tag{2.53}$$

The sign of OE_i shows whether the measure of class membership was underestimated (positive) or overestimated (negative). These summary measures may be used to assess the accuracy of the classification algorithm.

One problem of the generalized confusion matrix using multiplication operator is that it is possible to achieve a counter result that the agreement between a pixel and itself is less than the agreement between the pixel and a different pixel. For instance, supposing that there are two pixels, X and Y, and they have membership in two land cover classes called a and b. Assume that pixel X has class proportions for class a and b are 0.6 and 0.4, and pixel Y has class proportions for class a and b are 0.8 and 0.2, respectively. According to the generalized confusion matrix, the agreement between pixel X and itself is 52%, while the agreement between pixel X and pixel Y is 58%.

2.8.2.7 Generalized cross-tabulation matrix

Pontius and Cheuk (2005) propose a multiple-resolution approach that uses a Composite operator to compute the cross-tabulation matrix to compare soft classified maps at multiple resolutions. In this method, the pixel in both the classified and reference data has membership in a given class according to the proportion of the pixel that the class constitutes; therefore, the sum of class membership within a pixel is 1.

Let C_{ni} is the membership of pixel n to class i in the target estimated data and R_{nj} is the membership of pixel n to class j in the target reference data. Both $0 \le C_{ni} \le 1$ and $0 \le R_{nj} \le 1$. In addition, they follow the constraint:

$$\sum_{i=1}^{c} C_{ni} = \sum_{j=1}^{c} R_{nj} = 1 \tag{2.54}$$

Assume the rows of the cross-tabulation matrix are classes of the classified data and the columns are classes of the reference data. Commonly, the cross-tabulation matrices are standardized by expressing all the entries of the matrix as a percent of the study area. This can be done dividing all of the matrix's entries by the total number of the pixels in the study area, and then expressing the result as a percent.

Using Composite operator, there are two steps to compute the elements of the cross-tabulation matrix. The first step calculates the agreement (diagonal entries) of each class according to the minimum operator by apply the following equation for the case $i = j$:

$$P_{nij} = \min(C_{ni}, R_{nj}) \tag{2.55}$$

After the agreement is computed for each pixel n, the remaining membership for each class is the disagreement. Therefore, the disagreement in each pixel for each class will be computed in the second step. The disagreement in pixel n of estimated data for class i is C_{ni} minus P_{nii} because the total membership is C_{ni} and the agreement is P_{nii}. Similarly, the disagreement in pixel n of reference data for class j is R_{nj} minus P_{njj}. The disagreement for class i is then distributed among the off-diagonal entries in proportion to the distribution of the disagreement for class j. In this case, the sums of the diagonal

and off-diagonal entries for each pixel make a total contribution of one unit to the cross-tabulation matrix. The disagreement in each pixel for each class calculated as follows,

$$P_{nij} = \left(C_{ni} - P_{nii}\right)\left[\frac{\left(R_{nj} - P_{njj}\right)}{\sum_{j=1}^{c}\left(R_{nj} - P_{njj}\right)}\right] \quad \text{for } i \neq j \tag{2.56}$$

For this Composite operator, the interpretation of the standardized matrix entry in position *ij* depends on its position in the matrix. If the entry is on the diagonal ($i = j$), entry *jj* is the agreement between the classified data and reference data with respect to class *j* that means entry *jj* is the proportion of landscape for which class *j* in the classified map exits at the same location as class *j* in the reference map. When the entry is off the diagonal ($i \neq j$), entry *ij* is the disagreement between class *i* in the classified data and class *j* in the reference data. Specifically, the entry *ij* is the proportion of landscape for which class *i* in the classified map exits at the same location as class *j* in the reference map. The concept of same location in this case means within the boundary of the same pixel.

In comparison with the generalized cross-tabulation matrices computed from other operators such as the Multiplication operator (Lewis and Brown, 2001) and Minimum operator (Binaghi, *et al.*, 1999), Pontius and Cheuk (2005) show that Composite operator solve some problems associated with these two operators. First, the Composite operator guarantees that the matrix's entries sum to 100%, which the Minimum operator fails to do. Second, the Composite operator gives the identity matrix when it compares a map to itself, which both the Multiplication and Minimum operators fail to do. Third, the Composite operator maintains the row totals and columns total as the resolution changes, which the minimum operator fail to do.

Pontius and Cheuk (2005) also represent some hypothetical examples and indicate that the Composite operator is also useful to present how well two maps agree in terms of how classes are clustered spatially. Furthermore, for the Composite operator, agreement tends to increase when the resolution becomes coarser, since the disagreement by the location will be resolved when the resolution becomes coarser.

2.8.3 Spectral mixture analysis

Spectral mixture analysis includes a number of techniques in which mixed spectral signatures are compared to a set of pure reference spectra. The reference spectra can be measured in the laboratory, in the field, or from the image. The fundamental assumption is that the spectral variation in an image is caused by mixtures of a limited number of surface materials. The results of these approaches are the estimations of proportions of the ground area of each pixel that are occupied by each of the reference classes (Lillesand and Kiefer, 2000). Spectral mixture analysis is a physical based model in which a mixed spectrum is modeled as a combination of pure spectra, called end-members. It supplies useful information at the sub-pixel level because multiple land cover classes can be identified within a single pixel.

Many approaches of spectral mixture analysis apply linear mixture models, in which the observed spectral response from an area on the ground is considered to be a linear mixture of the individual spectral signatures of the land cover classes apparent within the area. The basic linear mixture model is described by two basic conditions for each pixel. First, the total of the proportion of all end-members in a pixel must equal to 1. It expresses as follows,

$$\sum_{j=1}^{c} f_j = 1 \text{ with the constraints } 0 \le f_j \le 1 \tag{2.57}$$

The second condition is that:

$$r_i = \sum a_{ij} f_j + e_i \tag{2.58}$$

where r_i is the reflectance (or digital number) of a given pixel actually observed in band i^{th} of m spectral bands, c is the number of mixture components or number of end-members; f_j represents the j^{th} fractional component (proportion of end-member j) in the make-up of r_i; a_{ij} is the reflectance of end-member j in spectral band i; and e_i is the error term expressing the difference between the observed pixel reflectance r_i and the reflectance for that pixel calculated from the model.

With multispectral data with m spectral bands, there would be m+1 equations available to solve for the various end-members. Therefore, in order for the above equations to be computable, the number of end-members c must be equal or less than the number of spectral bands m ($c \leq m$), which means that one cannot have more classes than wavebands. If $c = m+1$, the set of the equations can only be computed to have exact end-member fractions (f_j) without any error term. If $c \leq m$, it is possible to estimate end-member fractions and error terms by least squares principle.

A number of groups of researchers have applied linear mixture modeling remotely sensed data in some different fields of study. For example, linear mixture modeling has been applied to estimate sub-pixel land cover composition (Foody and Cox, 1994; Van De Meer and De Jong, 2000), to generate fraction images of forest covers proportion and crop area proportions (Townshend *et al.* 2000), to predict vegetation abundance for urban land (Small, 2001). Besides, this method is also used to compare with other soft classifiers. Examples are Bastin (1997), who compare fuzzy *c*-means, linear mixture modeling and MLC as a tools for unmixing coarse pixels, Atkinson e*t al.* (1997), who compare linear mixture modelling for sub-pixel land cover mapping with other alternative approaches namely fuzzy *c*-means and artificial neural networks. However, the approach has some drawbacks such as the neglect of multiple reflections, which can result in complex non-linearities in the spectral mixing process and the limit of number of end-members (or number of land cover classes) for unmixing (Lillesand and Kiefer, 2000). In this case, a more sophisticated non-linear spectral mixture model may be required (Borel and Gerstl, 1994; Carlotto *et al.*, 1995).

2.8.4 Methods based on fuzzy set theory

Fuzzy classification is designed to solve the mixed-pixel problem by applying the fuzzy set theory, in which a given pixel may have partial membership in more than one land cover class (Lillesand and Kiefer, 2000). The concept of fuzzy sets was first proposed by Zadeh (1965, 1973) as a means of describing and quantifying imprecision (Bastin, 1997). The fuzzy concept has been applied in different fields such as fuzzy logic control (Yamakawa, 1993), fuzzy neural networks (Pal and Mitra, 1992), and fuzzy rule base (Ishibuchi *et al.*, 1995). The fuzzy concept can also be applied for classification problem. In remote sensing classification, fuzzy-based classifiers are very popular these years (Wang, 1990; Tzeng and Chen, 1998; Tso and Mather, 2001; Bardossy and

Samaniego, 2002). This section firstly presents brief introduction to fuzzy set theory which provides the basis for soft classification problem in remote sensing and secondly the main fuzzy-base classifiers such as fuzzy *c*-means classification, fuzzy maximum likelihood classification, fuzzy rule base classification, fuzzy neural networks are described.

2.8.4.1 Introduction to fuzzy set theory

The operations of fuzzy sets are extensions of those used for traditional crisp sets. Traditional crisp sets use probability theory to explain if an event is expected to occur, but fuzzy sets measure the degree to which an event occurs (Tzeng and Chen, 1998). In fuzzy set theory, let S represent a universal of generic elements expressed by s. A fuzzy subset G of S is identified by a membership function μ_G, which assigns a membership grade within the interval [0, 1] to each element s. The membership grade can be denoted by:

$$\mu_G : s \rightarrow [0,1] \tag{2.59}$$

In traditional crisp set approach, the membership grade must be either 0 or 1. Thus, the fuzzy set is more flexible for solving the problems of ambiguous boundaries that are common in nature world. Let $\mu_G(s)$ represents the grade of membership of s in a fuzzy subset G that can be computed as (Tso and Mather, 2001):

$$G = \sum_i \mu_G(s_i)/s_i \tag{2.60}$$

The equation (2.60) represents the discrete case. In the continuous case, G becomes:

$$G = \int \mu_G(s)/s \, ds \tag{2.61}$$

The symbol '/' in equation (2.61) does not represent for the division operator but the link between the values of s and its corresponding membership grade $\mu_G(s)$ in the fuzzy subset G. The membership grade $\mu_G(s)$ can be defined as the partial memberships of a pixel for a set of land cover classes. Thus, these memberships can be used for the representation of land cover proportions within a pixel. Based on fuzzy set theory, there are several approaches such as fuzzy maximum likelihood classifier, fuzzy *c*-means

classifier, fuzzy rules base and fuzzy neural networks proposed for sub-pixel land cover proportion estimation.

2.8.4.2 Maximum likelihood classification

The MLC is represented in section 2.6.3.3 as for hard classification. But, this algorithm can also be applied for soft classification through two common approaches namely 'soften' maximum likelihood and fuzzy maximum likelihood.

'Hard' MLC operates by using the training data as of the way to estimate the means and variances of the classes, which are then used to estimate the posterior probabilities. The conditional probabilities of an unknown pixel belonging to each of the predefined classes are calculated, and then the pixel is allocated to the land cover class which has the highest posterior probability of membership. However, the posterior probability provides information on the relative strength of class membership. As evident from equation 2.9 the posterior probability indicates the membership of a pixel to a class from the set of defined classes and is a measure of the uncertainty in making a hard class allocation. Thus, it is possible to 'soften' a maximum likelihood classification by using the posterior membership probability values as indices of class membership and derive a soft classification (Bastin, 1997; Foody, 2002b). This method can be called softened MLC.

However, many researchers (i.e Wang, 1990a and 1990b; Maselli *et al.*, 1995; Bastin, 1997; Tso and Mather, 2001; Foody, 2002a) suggested that fuzzy set theory may also be extended to the maximum likelihood algorithm to calculate the membership grade of the pixels. In 'hard' maximum likelihood algorithm, pixels are assigned to the class for which they have the highest membership probability with the use of the 'hard' mean and 'hard' covariance matrix for each class. Thus for fuzzy maximum likelihood approach, the fuzzy means v_i and fuzzy covariance matrix F_i for class i can also be computed as (Tso and Mather, 2001):

$$v_i = \frac{\sum_j \mu_i(x_j) x_j}{\sum_j \mu_i(x_j)} \tag{2.62}$$

and

$$F_i = \frac{\sum_j \mu_i(x_j)(x_j - v_i)(x_j - v_i)^T}{\sum_j \mu_i(x_j)} \tag{2.63}$$

where x_j is the feature vector for pixel j.

The membership grade for each land cover class of each pixel can be calculated from the conditional probabilities as follows (Maselli *et al.*, 1995; Tso and Mather, 2001):

$$\mu_k(x_j) = \frac{P_k(x_j)}{\sum_i P_i(x_j)} \tag{2.64}$$

where k is the land cover class and probability $P_i(x_j)$ represents the class conditional probability for class i given the observation x_j and calculated by equation (2.10) , except that the crisp mean and covariance matrix in equation (2.10) are replaced by fuzzy mean and fuzzy covariance matrix. The accuracy of fuzzy maximum likelihood to mixed pixel classification depends on how well fuzzy mean and fuzzy covariance matrix are estimated.

In application of the maximum likelihood algorithm, Wang (1990) proposed that the classification results from 'soft' maximum likelihood algorithm are better than those from the 'hard' maximum likelihood algorithm in terms of accuracy.

2.8.4.3 Fuzzy *c*-means classification

To derive fuzzy membership, which may be used to create a soft classifier, the fuzzy classifier such as the fuzzy *c*-means (FCM) clustering algorithm (Bezdek *et al.*, 1984) or its possibilistic counterpart the possibilistic *c*-means (PCM) algorithm (Krishnapuram and Keller, 1996) may be used. The FCM has widely applied for deriving of soft thematic maps in remote sensing (Foody, 1996a and 2002; Atkinson et al., 1997; Bastin, 1997). The FCM algorithm supplies a means of solving the false clustering problem. This algorithm separates data clusters with the algorithm is presented as follows,

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a data set containing n observations (pixels) in an s-dimensional Euclidean space (e.g. with s wavebands). The integer c is the number of

clusters or classes in X. Thus, a fuzzy c partition of X can be represented by a real $(c \times n)$ matrix U consists of elements denoted by u_{ik}, satisfies the following constraints (Bezdek *et al.*, 1984; Tso and Mather, 2001; Foody, 2001),

$$u_{ik} \in [0,1] \tag{2.65}$$

$$\sum_{k=1}^{n} u_{ik} > 0, \qquad i = 1..c \tag{2.66}$$

$$\sum_{i=1}^{c} u_{ik} = 1, \qquad k = 1..n \tag{2.67}$$

u_{ik} is the fuzzy membership value of an observation x_k to the i^{th} class and is positively related to the strength of membership of a pixel to a specified class.

There are several algorithms to derive an optimal FCM clustering. One widely used algorithm is associated with the generalized least-squared error function J_m (Bezdek *et al.*, 1984),

$$J_m(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m \|x_k - v_i\|_A^2 \tag{2.68}$$

where m is the membership weighting exponent that lies within the range $1 \leq m < \infty$ which controls the degree of fuzziness (increasing m tends to increase fuzziness). When m=1 a hard classification is acquired in which each is related to just one class. There is no optimal value of m and for most data, 1.5< m <3.0 are commonly chosen (Bezdek *et al.*, 1984; Foody, 1996a; Zhang and Foody, 2001). $V = (v_1, v_2, ..., v_c)$ are the vector centres. $v_i = (v_{i1}, v_{i2}, ..., v_{in})$ is the centre of cluster i (i.e. the mean of the cluster). The squared distance between x_k and v_i (the Mahalanobis distance) is computed in the A-norm as,

$$\|x_k - v_i\|_A^2 = (x_k - v_i)^T A (x_k - v_i) \tag{2.69}$$

There are infinitely many A-norms available for use in equation (2.68). In practice, however, only a few of these norms such as Euclidean Norm, Diagonal Norm, and Mahalonobis Norm are widely used (Bezdek *et al.*, 1984). For m>1 and $x_k \neq v_i$, for all i,

k, a local minimum of J_m may be calculated under the circumstance (Bezdek *et al.*, 1984, Tso and Mather, 2001),

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{|x_k - v_i|}{|x_k - v_j|}\right)^{2/(m-1)}} \quad \text{for all } k \tag{2.70}$$

and the i^{th} cluster mean is calculated from,

$$v_i = \frac{\sum_{k=1}^{n}(u_{ik})^m . x_k}{\sum_{k=1}^{n}(u_{ik})^m} \quad \text{for all } i \tag{2.71}$$

Thus the FCM clustering is performed by iteratively using equations (2.70) and (2.71).

Although originally proposed as a clustering (unsupervised) technique, the algorithm may be modified so that it can be applied for supervised classification. The ability of FCM applying for unmixing coarse remote sensing image has examined by many analysis (Foody, 1996a; Atkinson, *et al.*, 1997; Bastin, 1997; Zhang and Foody, 2001; and Foody, 2002,). In comparison of FCM, linear mixture modeling and maximum likelihood classifier as tools for unmixing mixed pixels, Bastin (1997) presented the differences between these three classifiers and also pointed out that the FCM gave the best prediction of sub-pixel land cover, followed by the linear mixture modeling and then fuzzy maximum likelihood classifier.

2.8.4.4 Fuzzy rule-based classification

This section introduces briefly a soft image classification using fuzzy rule-based methodology, as proposed by Ishibuchi, *et al.* (1992), Tso and Mather (2001).

A fuzzy rule base (or fuzzy system) applied for soft classification in remote sensing generally consists of three main steps. In the first step, the analyst has to define membership functions to calculate the membership grade for the input pixels which will be used in the fuzzification (the second step) of the data. The types of membership functions are varieties (but the value of the function must be on the range [0, 1]) and

there are four kinds of them namely monotonic, triangular, trapezoidal, and bell-shaped (Figure 2.14), are used widely in practice (Tso and Mather, 20001).

The second step, fuzzification, divide the input feature space into fuzzy subspaces, each specified by a fuzzy membership function which defined in the first step. When the construction of fuzzy subspace is finished, fuzzy rules are then generated for each fuzzy subspace.

The third step, namely inference, generates a set of fuzzy rules in order to derive the information class membership grades from the subspace membership grade vector. The rule strengths w_{ij} are then computed from the training data. These rule strengths can then be used to calculate the membership grade values of each land cover class in each classified pixel. Thus, they can reflect the mixture information between the information classes.

To illustrate the above steps, take an example. For simplicity, a two-dimensional input case is considered and the input features are on the range [0, 1]. Each dimension in the input space is then divided into c fuzzy subspaces (c is also the number of land cover classes) represented by $\{A_1, A_2, \ldots, A_c\}$, where A_i is the i^{th} fuzzy subspace. A symmetric triangular membership function is applied, and the fuzzy subspace for A_i is then identified by:

$$\mu_i = \max\left\{1 - \frac{|s - a_i|}{\lambda}, 0\right\} \tag{2.72}$$

where s is the input pixel value, a_i is the triangular centre for fuzzy subspace i and λ is the membership function width. Both a_i and λ are measured as:

$$a_i = \frac{i-1}{c-1} \text{ and } \lambda = \frac{1}{c-1} \quad (i = 1, 2, \ldots, c) \tag{2.73}$$

Thus, each fuzzy rule in the fuzzy subspace can take the flowing form:

IF s_1 is in A_i and s_2 is in A_j
THEN pixel s assigns to the class c_{ij} with strength w_{ij}.

where s_1 and s_2 are the input features for pixel s, and A_i and A_j $(1 \leq i, j \leq c)$ are fuzzy subspace on both dimensions, respectively. The class c_{ij} and rule strength w_{ij} depended on the fuzzy rule are calculated from the training data.

Further detail on fuzzy rule-based methodology can be seen in the text of Ishibuchi, *et al.* (1992); Nozaki, *et al.* (1996); Tso and Mather (2001); Bardossy and Samaniego (2002); Ghosh, *et al.* (2003).

There are two issues relating to the use of the fuzzy rule base for solving the mixed pixel problem (Tso and Mather, 2001). Firstly, the membership grades for a given pixel over all classes must sum to one. When the number of data dimensions is large, the number of rules may encounter a problem for computation. This issue can be solved by data dimension reduction techniques.

The application of fuzzy rule base methods is still at an early stage, especially in soft classification of remotely sensed data (Tso and Mather, 2001). Further investigation is needed before an adequate method of dealing with mixed pixels is produced. Bardossy and Samaniego (2002) investigated the application of fuzzy rule-based modelling to classify Landsat TM imagery and the results were a land cover map with four different categories of land cover and an image depicting the degree of ambiguity of classification for each pixel. Ghosh *et al.* (2003) used fuzzy rule based approaches for cloud cover estimation. The technique automatically shows some fuzzy rules for estimating the proportion of each class (three classes) in each pixel and then is hardened by the defuzzification to assign each pixel to one of three possible classes.

2.8.5 Artificial neural networks

ANNs have received attention for use in both hard and soft classification of remotely sensed imagery (e.g. Foody, 1996b; Atkinson *et al.*, 1997; Kanellopoulos *et al.*, 1997; Bernard *et al.*, 1997; Carpenter *et al.*, 1999; Zhang and Foody, 2001; Foody, 2002a). The use of artificial neural network classifiers for traditional hard classification is described in section 2.6.4.2. In this section, ANNs are considered to resolve the mixed pixel problem.

The architecture of MLP and RBF networks are described in section 2.6.4.2 as they use for hard classification. For soft classification, their architectures are similar but it may be different from the network training and output membership. As for hard classification, training samples for an ANN classifier should be pure and perform by a crisp method. However, training for an ANN can be performed either a crisp or a fuzzy method with the training samples containing either pure pixels or both pure and mixed pixels. In crisp training, each training pixel is associated fully with a single class and thus discrete classes are represented to the network. An approach to fuzzy training is to feed the network with the proportional coverage of each class in the training pixels (Foody, 1997; Zhang and Foody, 2001). After training, the networks can be used for classification of unknown class membership. The output from an ANN exists as activation levels ranging from 0 to 1. Traditionally, the results of the network are used for deriving a hard classification, in which a pixel is allocated to the class with which it has the highest activation level. However, several researchers suggested that the activation level of each output unit was correlated with the proportion of the area represented by a pixel which was covered with the class associated with the unit. Thus, the activation levels can also be treated as fuzzy membership values to indicate the strength of class membership in each pixel and forming the concept of soft classification. Since the output activations from an ANN classifier generally do not sum to 1, the output values must be rescaled in order to have the sum of the proportions of all classes in one pixel equal to 1.0.

The potential of ANNs for mixed pixel problems is recognized by several authors (e.g. Foody, 1996b and 1997; Atkinson, *et al.*, 1997; Warner and Shank, 1997; Bernard *et al.*, 1997; Gopal *et al.*, 1998; Lin *et al.*, 2000; Foody, 2002; Zhang and Foody, 2001). Some of them approved the superiority of the ANN to fuzzy maximum likelihood and fuzzy *c*-means classifiers for soft classification in terms of both information provided and the accuracy of classification (Atkinson *et al.* 1997; Zhang and Foody, 2001).

2.8.6 Evidential reasoning classifier

In section 2.6.4.4, evidential reasoning classifier is described as a hard classification, however, the underlying theory of evidence is closely related to fuzzy set theory, and consequently evidential classifiers could be adapted for fuzzy classification (Zhang and Goodchild, 2002; Foody, 2002b) As stated in section 2.6.4.4, to apply evidential

reasoning to classification, a measure of evidence for subsets of class labels is computed. Evidential support is computed with respect to the frequency of occurrence of values within training samples. Class membership may be estimated from these frequency distributions or histograms of training samples over the entire set of classes, or frame of discernment. Training data are read from each data feature in sequence, and a frequency distribution of training values is interpreted for each class. Frequency distributions may be interpreted as approximating probability densities or as general fuzzy measures. So, for a pixel with value x, it probabilistic and fuzzy membership values of belonging to class i, denoted by $p_i(x)$ and $\mu_i(x)$ respectively, can be measured as

$$p_i(x) = \frac{f_x}{T}, \tag{2.74}$$

$$\mu_i(x) = \frac{f_x}{f_m} \tag{2.75}$$

where f_x, f_m and T represent the frequency of value x occurring, the maximum frequency and the total frequencies count in the histogram (Peddle, 1995a; Foody, 2002b). Using the frequency distributions compiled from the training data for all the classes to be mapped, a soft classification may be derived for a pixel by estimating the frequencies of occurrence (Zhang and Goodchild; 2002, Foody, 2002b). Implementing the analysis for each pixel, a soft classification may be generated by the spatial distribution of the fuzzy membership derived.

In recent years, some researchers have applied evidential reasoning soft classification of remotely sensed image. Foody (2002b) proposed the use of evidential reasoning to derive a soft thematic map. While Lein(2003) demonstrated an alternative soft-classification strategy based on D-S theory of Evidence and applied this technique to the problem of agriculture land cover mapping.

2.9 Super-resolution mapping

The output of a soft classification is typically a set of proportion images that show the predicted coverage of each thematic class in the area represented by each pixel. Soft classifications have been found to provide more informative and potentially more

accurate representations of land cover than conventional hard classifications. Although soft classifications predict the proportion of each land cover class within each pixel they do not indicate where the land cover classes are spatially located within the pixels. The sub-pixel class proportions may, however, be located geographically through super-resolution mapping.

Several techniques have been used for super-resolution mapping based on the output of a soft classification. They were briefly presented in Atkinson (2005). Atkinson (1997) mapped the location of the proportions output from a soft classification within each pixel based on an assumption of spatial dependence within and between pixels. This approach produces accurate land cover maps at a finer spatial resolution than the initial input images. However, the technique compared sub-pixels to pixel proportions and so the complex mixing in the data caused the simple technique to suffer from problems (Tatem *at al.*, 2002). Verhoeye and De Wulf (2002) built on earlier work of Atkinson (1997), but formulated as a linear optimization problem. The approach was evaluated using synthetic imagery and a SPOT image of Sahelian wetlands. Compared to the conventional hard classification techniques, this algorithm achieved a certain degree of success, but problems were once again due to the fact that sub-pixels were compared to pixels, rather than other sub-pixels (Tatem *at al.*, 2002).

A solution to the super-resolution mapping problem may be achieved by comparing sub-pixels to sub-pixels (Atkinson, 2005). Recently, Tatem *et al.* (2001, 2002) developed a Hopfield neural network (HNN) technique (Hopfield and Tank, 1995) as an energy optimisation tool constrained by output from the soft-classification technique for super-resolution target mapping. The HNN technique is proposed as simple and robust technique for predicting the location of class proportions within each pixel. More details on the technique will be discussed in the next section.

2.9.1 Hopfield Neural Networks

Hopfield neural network (HNN) is a fully connected recurrent network which is mostly used for auto-association and optimisation. The HNN may be implemented physically by interconnecting a set of resistors and amplifiers with symmetrical outputs and external bias current sources (Tatem *et al.*, 2001, 2002). The HNN has been used in remote sensing for applications such as ice-mapping, cloud motion, and ocean current

tracking (Cote and Tatnall, 1997, Lewis, 1998). Recently, this approach has been demonstrated to be a successful tool for super-resolution land cover mapping based on the output of soft classification (Tatem *et al.*, 2002, Nguyen *et al.*, 2006). The HNN was used to map the spatial location of classes using information of sub-pixel class compositions estimated from soft classification. This was implemented by converting soft land cover proportions to hard (per-sub-pixel) land cover classes (e.g., at a finer spatial resolution). The procedure of the HNN was outlined in detail in Tatem *et al.*, (2002).

The HNN is used as an optimization tool. It is initialized randomly using the class composition estimates from a soft classification and run until it converges to a monotonic stable state (Tatem *et al.*, 2001 and 2002). The zoom factor, z, determined the increase in spatial resolution from the original remotely sensed imagery, which was used to derive soft classification output, to the new fine spatial resolution image. After convergence to a stable state, the output values of all neurons of the network were either 0 or 1, representing a binary classification of the land cover at the finer spatial resolution. The specific goals and constraints of the HNN energy function determined the final distribution of neuron output values. The energy function can be defined as,

$$E = -\sum_{i}\sum_{j}(k_1 G1_{ij} + k_2 G2_{ij} + k_3 P_{ij} + k_4 M_{ij}) \tag{2.76}$$

where k_1, k_2, k_3 and k_4 are weighting coefficients which define the effects of the corresponding two goal functions ($G1_{ij}$ and $G2_{ij}$), proportion constraint (P_{ij}) and multi-class constraint (M_{ij}).

Using the class proportion images derived from a soft classification as the input, the HNN is implemented using some parameters which should be carefully chosen by the user. Firstly, the optimum values of four weighting constants, k_1, k_2, k_3 and k_4, should be used because they mainly control the direction of the optimisation process of the network. Typically, identifying optimum weighting constraint values is a difficult and tedious task, so an estimation of optimal values was made via certain assumptions and multiple network trial runs (Tatem *et al.*, 2002). In addition to the constant values, a zoom factor, z should also be used for the system. Furthermore, the number of iterations for the performance of the network should be set up by the user. Like the zoom factor,

the higher the number of iterations the more accurate the output of the network produced but this also increases the computational time.

The output of the HNN approach is a set of binary images with a spatial resolution that is finer than that of the input class proportional images derived from soft classification (Figure 2.11). The number of the binary images is equal to the number of land cover classes to be mapped with each image is shown the location of a defined class.

2.10 Summary

Information on land cover is required to help our understanding and management of the environment. Accurate mapping of land cover, as well as identifying and extracting the target land cover features, are extremely important procedures for this. Commonly, remote sensing is the most appropriate means of providing this information, particularly for large area studies. Because remotely sensed images provide a complete, relatively consistent and inexpensive representation of the Earth's surface, remote sensing is a huge potential source for land cover mapping based on the surface's reflectance response.

In the raw form, prior to any form of processing, remotely-sensed data generally contain distortion and deficiencies due to the influence of external factors which alter the spectral signal from surface features. These factors include atmospheric interference, variation in illumination geometry, geometric distortion and technical problems with sensor. Therefore, to acquire an accurate representation of land cover from remotely sensed imagery, pre-processing is often necessary to reject the influence of these factors. After pre-processing, the remotely sensed imagery can be used for required image processing such as image classification.

A common way to extract land cover information from remotely sensed imagery is via image classification analysis. Image classification is the process of assigning pixels to classes. Land cover classes can be identified by their spectral, spatial, temporal and geometric characteristics. The classification process generally consists of three basic steps: class training, pixel labeling and accuracy assessment.

In addition to the distinction between unsupervised and supervised classification, the classification techniques can also be grouped on the basis of how classifiers represent land cover. Generally, they can be divided into two main types such as hard classification and soft classification. Traditionally 'hard' classification techniques, which are among the most commonly undertaken analyses of remotely sensed data, assign each pixel to a single land cover class with which it has the greatest similarity. They were developed for the classification of classes that may be considered to be discrete and mutually exclusive, and assume that each pixel is perfectly pure (i.e. the ground area represented by a pixel is occupied by a single information class). However, if the pixel resolution is coarse in comparison with the variability of land cover objects then a single pixel in such imagery may be of more than one land cover class. Thus, hard classification techniques are inappropriate for mixed pixels. This led to the development of 'soft' classification techniques.

Soft classification allows image pixels to have partial and multiple class membership and therefore can give a more accurate and realistic representation of many land covers. A fuzzy classification may be derived through the use of fuzzy classifiers or by 'softening the output of a 'hard' classification. For example, measures the strength of class membership, rather than just assign a single land cover class with which it has the greatest similarity, to each pixel. Soft classification approaches include a number of methods such as spectral mixture analysis, MLC, FCM, fuzzy rule-based, ANN, evidential reasoning, and *k*-nearest neighbour. The output of such approaches generally takes the form of a set of proportion images, each display the proportion of a certain class within each pixel. Although soft classifications can provide a more accurate and appropriate representation of themes such as land cover than a conventional hard classification, the application and development of such techniques require further research.

Chapter 3

Increasing classification accuracy through the use of an ensemble of classifiers

3.1 Introduction

Land cover mapping through the means of an image classification is one of the most common applications of remote sensing. However, the full potential of remote sensing as a source of land cover information is often unrealized due mainly to a set of technical problems. As indicated in section 2.7, one of the most important problems limiting classification accuracy is that of mixed pixels (Fisher, 1997; Cracknell, 1998; Campbell, 2002), which can make the single class allocation derived from conventional hard classification inappropriate and inaccurate. Soft or fuzzy classification techniques allow for the partial and multiple class membership within each mixed pixel, and, therefore, may be used to refine the standard mapping process and to increase the accuracy of land cover mapping from remote sensing (Foody and Cox, 1994; Atkinson *et al,.* 1997; Tso and Mather, 2001).

The output of a soft classification is typically a set of fraction images that show the predicted coverage of each thematic class in the area represented by a pixel. There are a range of approaches that can be used to derive a soft classification such as those based on the maximum likelihood (Wang, 1990; Maselli *et al.*, 1995; Ibrahim *et al.*, 2005), fuzzy *c*-means (Zhang and Foody, 2001; Ibrahim *et al*, 2005), *k*-nearest neighbour (Schowengerdt, 1996), artificial neural network (Foody, 1996; Zhang and Foody, 2001) and support vector machine classifications (Brown *et al.*, 2000). These soft classification approaches were presented in section 2.8. Different classifiers will commonly vary in their allocations, with some yielding high accuracies for some parts of the data while others may achieve high accuracies in other parts. Although soft classifications can provide a more accurate and appropriate representation of themes such as land cover than a conventional hard classification there are many concerns. In particular, the accuracy of soft classifications is still often low. The literature, for example, contains many examples of soft classification analyses in which the degree of

correlation between the actual and predicted class composition was low (r<0.5) (e.g., Foody and Arora, 1996; Zhang and Foody, 2001; Shalan *et al.*, 2003; Ibrahim *et al.* 2003).

A variety of different classification outputs can be derived from the application of a suite of classifiers to the same data set. The derived classifications may differ greatly in accuracy, on both a per-class and overall basis. Classifiers that have, largely, uncorrelated errors may sometimes usefully be combined to form an ensemble approach to classification. By combining the outputs of a set of classifiers it is possible to derive a classification that is more accurate than any of the individual classifications used. The use of such an ensemble of classifiers has been shown to be able to increase the accuracy of hard classification (Rogova, 1994; Lam and Suen, 1997; Ji and Ma, 1997; Briem *et al.*, 2000; Steele, 2000; Giacinto *et al.*, 2000; Liu *et al.*, 2002) but its potential for soft classification has rarely been investigated (Huang and Lees, 2004). The main objective of this chapter, therefore, was to investigate some of the possible approaches to increase the accuracy of soft land cover classification though the use of an ensemble of classifiers.

The approaches which have been used in the literature to combine hard classifications were investigated with some brief descriptions in section 3.2.1. Following this, the methods to combine soft classifications are provided in section 3.2.2. The data used for the research and some procedures required to pre-process them are presented in section 3.3. The potential of the approaches to combine soft classifications as well as the approaches to combine hard classifications were assessed through a range of experiments with results in section 3.4. Finally, section 3.5 closes the chapter with some general conclusions.

3.2 Methods for combining land cover classifications

3.2.1 Methods for combining hard classifications

There have been many methods that may be applied to combining hard classifications such as voting principle, Bayesian average, Bayesian belief function (Xu *et al.*, 1992; Giacinto, 2000), boosting, bagging (Briem *et al.*, 2002; Kuncheva *et al.*, 2002)), and

neural network (Wanas and Kamel, 2001), *etc.* In the following sections, two commonly used methods are briefly summarized. They are voting principle and Bayesian belief function. These methods are used with assuming that classifiers in the combination process are different from each other in their methodologies and make uncorrelated classification errors.

3.2.1.1 Voting principle

The voting principle or majority vote is a simple and popular method for combining hard classification outputs. It has been investigated by some researchers (e.g., Xu *et al.,* 1992; Giacinto and Roli, 1997; Lam and Suen, 1997; Kuncheva, 2000; Giacinto *et al.*, 2000).

Let consider an ensemble made up of L different classifiers. Each classifier provides the outputs in terms of the land cover class labels assigned to the image pixels. A given input pixel X receives, therefore, L classification labels from the ensemble, each label corresponding to one of the c land cover classes. Voting principle is used to combine the results provided by L different classifiers by interpreting each classification's output as a 'vote' for one of the land cover classes. The final combined class label of an input pixel is the class which received the maximal number of votes among the individual classifications. Sometimes a majority vote is required. This means the class label that receives a vote greater than or equal to a prefixed threshold is taken as the final output of the input pattern X. The threshold is commonly equal to half of the number of the classifiers in the ensemble.

3.2.1.2 Bayesian belief function

In the voting principle, the combination of classification outputs is only based on the class label outputted for each input pixel by each classifier and each classifier is treated equally. However, the classifiers may differ in the classification accuracy which should be considered in the combination process. The belief function, therefore, will take the errors of the classifiers in the ensemble into consideration as the prior knowledge of the classifiers (Giacinto and Roli, 1997; Giacinto, *et al.*, 2000; Lam and Suen, 1997; Kuncheva, 2000).

The Bayesian belief function method considers the available prior knowledge of the error of each classifier as its confusion matrix. For each classifier e_l, $l = 1 \ldots L$, the confusion matrix can provide an estimate of the conditional probability that the true label is ω_j given that e_l assigns class label ω_i. The conditional probabilities can be calculated as follows,

$$P\left(X \in \omega_j \mid e_l(X) = \omega_i\right) = \frac{n_{ij}^{(l)}}{\sum_{j=1}^{c} n_{ij}^{(l)}} \qquad i = 1..c,\ j = 1..c,\ l = 1..L \tag{3.1}$$

where the element $n_{ij}^{(l)}$ denotes the number of samples whose true class label was ω_j, and were assigned by classifier e_l to class ω_i, c is the number of classes.

On the basis of the above conditional probabilities, the combination of the L classifiers can be obtained according to the following belief functions:

$$bel(j) = \eta \prod_{l=1}^{L} P\left(X \in \omega_j \mid e_l(X) = \omega_i\right) \qquad j = 1...c \tag{3.2}$$

with η is a constant that ensures that $\sum_{j=1}^{c} bel(j) = 1$. The final classification is then taken by assigning the input pattern X to the land cover class for which *bel(j)* has the maximum value.

3.2.2 Methods for combining soft classifications

As indicated in section 2.8, the output of a soft classification is different from that of a hard classification. The output of the former is typically a set of fraction images, each describing the proportion of a particular land cover class within each pixel, while that of the latter is an image in which each pixel is associated with a single land cover class label. As a result of the different nature of the classification outputs, the methods to combine classifications that were applied for hard classifications may not be suitable for soft classifications. However, it may still be possible to combine the outputs of different soft classifications in order to increase classification accuracy. Four methods

to combine soft classification outputs were suggested here. They were (i) an approach that selects the most accurate predictions on a class-specific basis to construct the final combined classification output, (ii) an approach in which the combined classification output for a case is the average of the outputs derived from the individual soft classifiers (iii) the direct combination of the individual soft classifications through the use of the Dempster-Shafer theory of evidential reasoning and (iv) an approach which degrades the soft classification outputs into a set of ordered classes and then combines these through the use of a conventional ensemble approach. Details of these approaches are described in the following sections.

3.2.2.1 Combination by selecting the most accurate predictions

The ensemble approach was based on a comparison of the individual class accuracies. Using a selected measure of classification accuracy (e.g. producer's accuracy), the individual classifiers were ranked in terms of the accuracy with which they classified a class. The outputs from the most accurate classifier were then selected to form the representation of that class in the final classification. This process was then repeated for each class in the classification (Figure 3.1).

Individual class accuracy may be expressed using class producer's accuracy, user's accuracy derived from fuzzy error matrix, and class correlation coefficient. Table 3.1 shows an example of comparing the accuracies (e.g. the class correlation coefficients) of a land cover class among three individual soft classifications.

Class	**Correlation coefficient**		
	Classification 1	Classification 2	Classification 3
Class 1	0.775	**0.802**	0.727
Class 2	0.577	0.615	**0.624**
Class 3	0.434	**0.507**	0.496
Class 4	**0.780**	0.766	0.727
Class 5	**0.527**	0.518	0.474

Table 3.1: An example using fictional data of the ensemble approach that selects the most accurate predictions of the individual classifications. The numbers in the Table are the class correlation coefficients and the ensemble outputs were highlighted in bold.

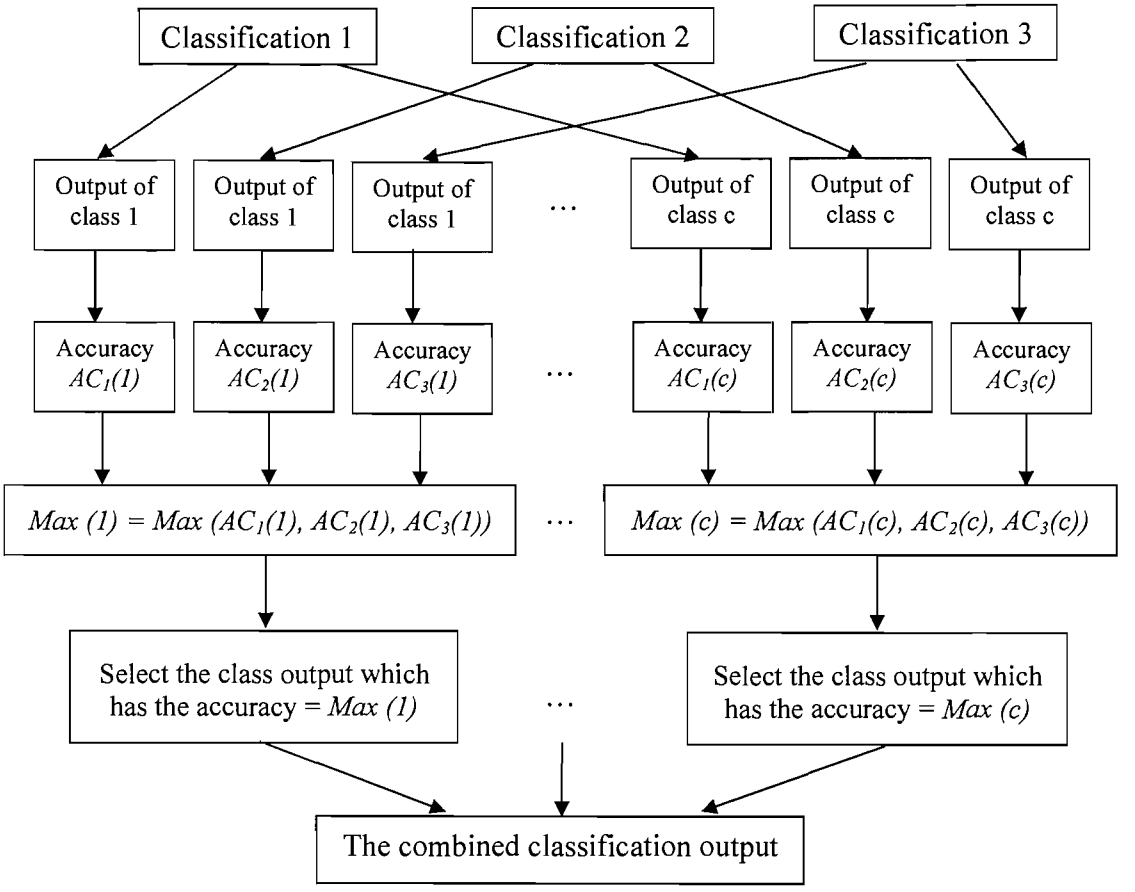


Figure 3.1: Flow diagram of the approach to combine soft classifications by selecting the most accurate predictions on a class-specific basis.

3.2.2.2 Combination based on average operator

In this approach, the outputs of the individual soft classifiers were averaged to derive the combined classification output. For an input pixel X, each classifier produces a real vector $[P_l(1), \ldots, P_l(c)]^T$, where $P_l(i)$ is the measure of class membership for class i within pixel X estimated by the classifier, and c is the number of land cover classes. To combine the estimates provided by L different classifiers, the average measure of class membership of each pixel X among L classifiers was calculated from:

$$P_{av}(i) = \frac{1}{L}\sum_{l=1}^{L} P_l(i) \qquad i = 1..c, l = 1..L \tag{3.3}$$

where $P_{av}(i)$ is the average proportion of class i within input pixel X, and c is the number of land cover classes.

The average values derived may be used as the final classification output for pixel X.

3.2.2.3 Combination based on evidential reasoning

The D-S theory of evidential reasoning (Shafer, 1976; Xu *et al.*, 1992; Rogova, 1994; Ahmadzadeh *et al.*, 2000; Tso and Mather, 2001; Foucher *et al.*, 2002) is a tool for representing and combining measures of evidence and also the means to accommodate uncertainty and ignorance that is often inherent in the data sets. In remote sensing, this theory has been applied to deal with the remotely sensed multisource classification problem as indicated in section 2.6.4.4 and therefore, may also be applied for combining land cover classifications. The main points of D-S theory were briefly described in section 2.6.4.4. This section only presents some more information about this theory in supporting for combining soft classifications.

In this approach, the soft classification outputs were directly combined through the use of the Dempster-Shafer theory of evidential reasoning (Shafer, 1976; Xu *et al.*, 1992; Ahmadzadeh *et al.*, 2000; Tso and Mather, 2001; Foucher *et al.*, 2002) to derive the final combined classification output.

The evidential reasoning approach may be illustrated for a classification of c land cover classes $\Theta = \{\omega_1, \cdots, \omega_c\}$, where Θ is the focal element and ω_i denotes land cover class i. The measure of class membership generated for each pixel by a soft classifier is treated as the degree to which the input pattern X belongs to the class. Thus, for an input pixel X, each individual classifier e_l produces a real vector $[P_l(1), \cdots, P_l(c)]^T$, where $P_l(i)$ is the proportion of class i within pixel X.

One concept that plays an important role in evidential reasoning is the basic probability assignment m. This indicates the degree to which the evidence advocates the claims that a particular element (e.g., a pixel) of the universal set belongs to a particular set such as a land cover class (Tso and Mather 2001). For an input pixel X, each classifier e_l produce a mass of evidence from its basic probability assignments as follows:

$$m_l = [m_l(1), \cdots, m_l(c), \theta] \qquad l = 1 \ldots L \tag{3.4}$$

$$m_l(i) = R_l . P_l(i) \tag{3.5}$$

$$\theta = 1 - R_l \tag{3.6}$$

where $m_l(i)$ is a basic probability assignment of classifier e_l of focal ω_i, R_l is the individual classifier's confidence, and θ is used to represent uncertainty (1 – classifier's confidence). An estimation of the individual classifier's confidence may be derived from the overall accuracy of that individual classifier.

For a basic probability assignment of a focal element A, there will be a belief function *Bel(A)*. This belief function has a numerical value in the range [0, 1] to indicate belief in a position (subset) $A \subset \Theta$ based on the occurrence of the evidence. The belief function can be calculated as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{3.7}$$

where *Bel(A)* is the sum of the mass of evidence which is committed to class A as well as various subsets of A.

When two or more masses of evidence $[m_1, \ldots, m_L]$ exist, here representing the opinions of a set of different soft classifiers, there will be two or more sets of basic probability assignments and belief functions given to the subset of the same focal elements. Dempster's combination rule could be used to combine them into a new basic probability assignment and belief function, which represents the combined impact of the evidences. If m_1 and m_2 denote two basic probability assignments derived from two classifiers to be combined, the Dempster-Shafer combination rule defines a new basic probability assignment $m = m_1 \oplus m_2$ which represents the combined effect of m_1 and m_2 by:

$$m_{1,2}(A) = m_1 \oplus m_2(A) = k \sum_{X \cap Y = A, A \neq \emptyset} m_1(X).m_2(Y) \tag{3.8}$$

where Ø is the empty set and $m_{1,2}(\phi) = 0$, X and Y are any subsets of the universal set, and k is a normalizing constant:

$$k^{-1} = 1 - \sum_{X \cap Y = \emptyset} m_1(X).m_2(Y) \tag{3.9}$$

The Dempster-Shafer combination rule is associative and commutative (Xu *et al.* 1992), thus it could be used to combine multiple evidences sequentially (and with an arbitrary order) using (3.8) and (3.9) to obtain:

$$m = m_1 \oplus m_2 \oplus \cdots \oplus m_L \tag{3.10}$$

After combining all of the masses of evidences of L classifiers, the final belief function *Bel(.)* can be computed by equation (3.7) and will be used as the final measures of class membership for input pixel X after combining the outputs of L individual classifiers. The process of combing soft classifications by the D-S theory of evidential reasoning can be illustrated in Figure 3.2.

3.2.2.4 Combination through a conventional ensemble approach.

As shown in section 3.2.1, there are a range of ensemble approaches for combining hard classifications such as those based on the voting principle (Xu *et al.*, 1992; Lam and Suen, 1997) and Bayesian belief function (Lam and Suen, 1997; Giacito *et al.*, 2000) as well as bagging and boosting algorithms (Briem *et al.*, 2002). To use the conventional ensemble methods to combine soft classifications, the soft classification output of each land cover class in the individual classifications could be degraded into a set of ordered classes. For example, the soft classification output in the range [0.0–0.19], [0.2–0.39], [0.4–0.59], [0.6–0.79] and [0.8–1.0] could be assigned to classes very small, small, medium, large and very large cover, respectively. Although this action results in the continuous soft classification being converted to a hard classification, the ordered nature of the classes preserves some of the main attractions of soft classification. The degraded output of each land cover class is, however, a hard classification and may then be combined through the use of a conventional ensemble method. The voting principle approach was selected to combine the degraded outputs here because it is simple and used widely (section 3.2.1.1). The process of the method

for combining soft classifications using a conventional ensemble method can be seen in Figure 3.3.

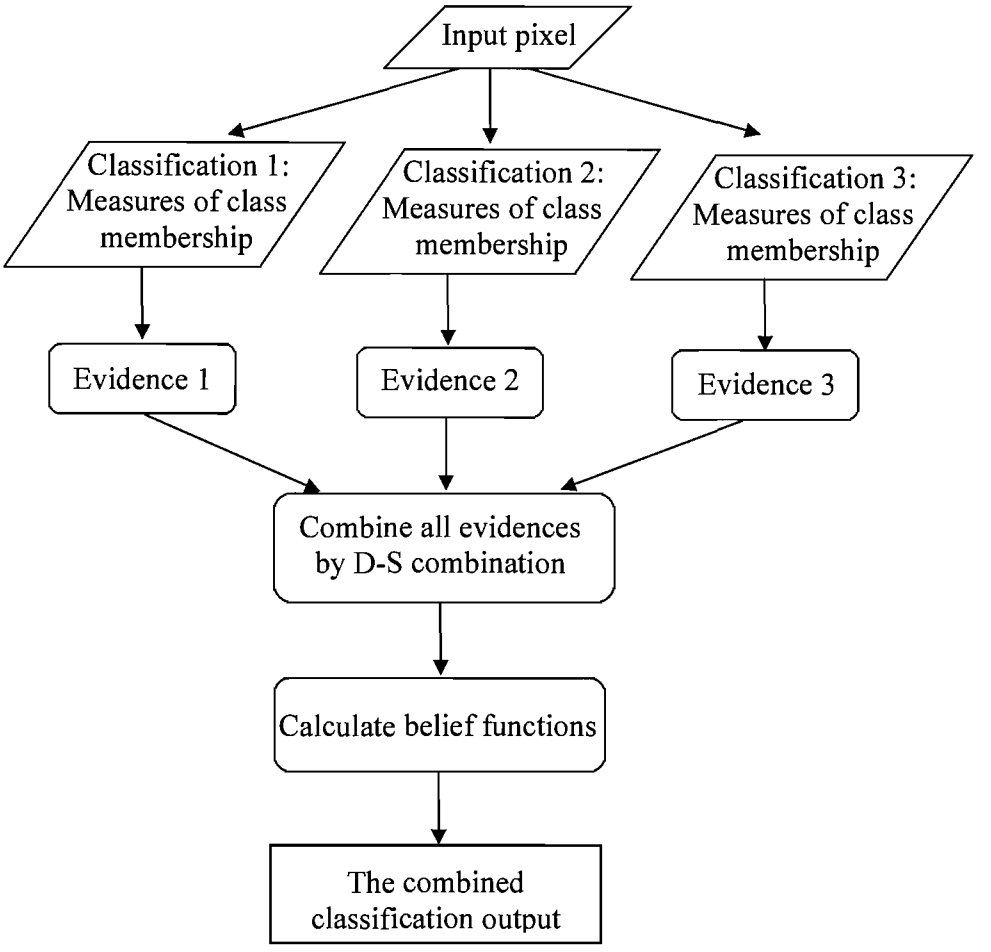


Figure 3.2: Flow diagram of the D-S theory of evidence to combine soft classifications.

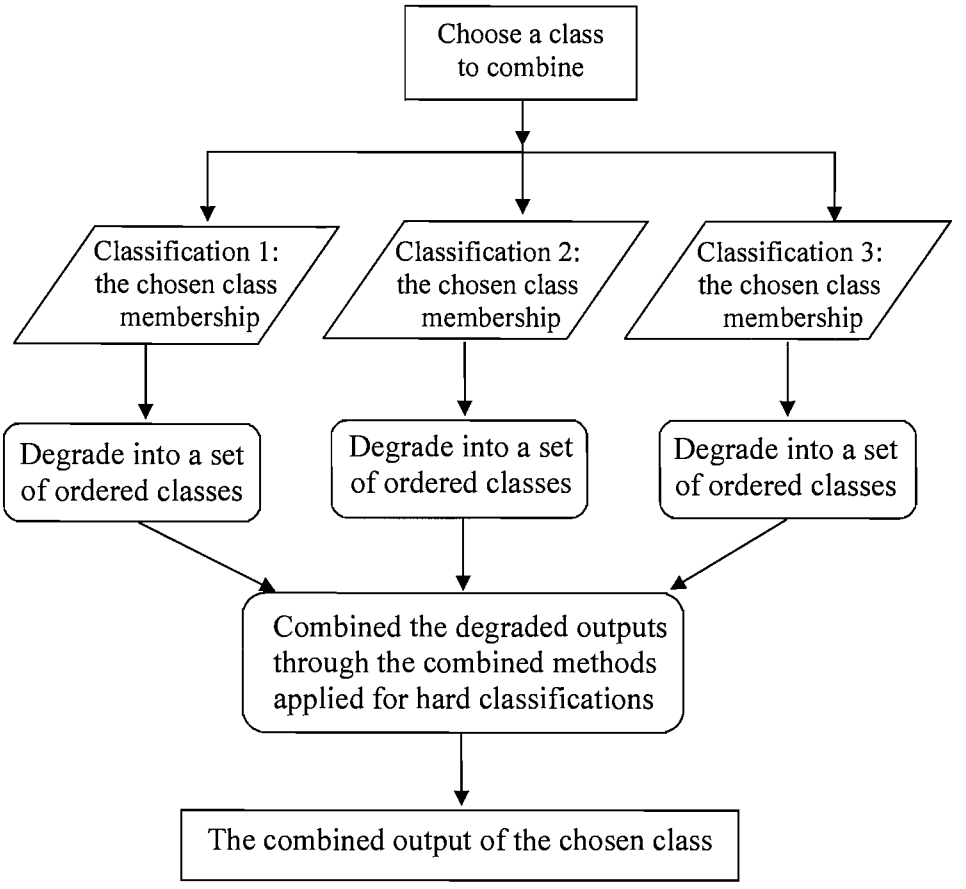


Figure 3.3: Flow diagram of the approach to combine soft classifications through a conventional ensemble approach.

3.3 Data

3.3.1 Study area and data used

For the purposes of this research it was necessary to have remotely sensed imagery to be classified together with a ground data set at a finer spatial resolution that would describe the class composition of image pixels. Furthermore, the analysis required data comprising both pure pixels and mixed pixels in a large area to provide a sufficiently large sample size for training and testing sets. Here, attention focused on the land cover of Australia. The image data, with 8 km spatial resolution, were derived from the NOAA/NASA Pathfinder AVHRR Land (PAL) archive. The reference data set used was the IGPB DISCover land cover map, with 1 km spatial resolution.

In terms of the remotely sensed image data to be classified, monthly maximum value composite NDVI which were derived from the NOAA AVHRR data of Australia with 8 km spatial resolution were used. These data were acquired in July 1992, December 1992, and March 1993, providing a temporal sample to help enhance class separability (Figure 3.4a-c). The NOAA AVHRR imagery provided the data to which the individual soft classifications were applied to estimate the sub-pixel proportions of different land cover classes. The data were registered in the Interrupted Goode Homolosine map projection.

The IGBP DISCover land cover map (Loveland and Belward 1997) of Australia with 1 km spatial resolution was used as the reference data. This land cover map was derived from NOAA AVHRR data spanning a 12-month period (April 1992 to March 1993) by an unsupervised clustering classification scheme using 12 monthly maximum NDVI composites as inputs and showed the distribution of 17 land cover classes (Table 3.2). For the purpose of the analysis, some of the classes which appeared to be similar in terms of their general characteristics, such as closed and open shrubland, were combined while the land cover classes covering only a small portion of the study area were discarded. The final set of classes defined for the study comprised five classes forest, shrubland, savanna, grassland, and cropland (Figure 3.4d). All the databases share the same map projection, Interrupted Goode Homolosine.

Class value	Description	Class value	Description
1	Evergreen Needleleaf Forest	9	Savannas
2	Evergreen Broadleaf Forest	10	Grasslands
3	Deciduous Needleleaf Forest	11	Permanent Wetlands
4	Deciduous Broadleaf Forest	12	Croplands
5	Mixed Forest	13	Urban and Built-Up
6	Closed Shrublands	14	Cropland/Natural Vegetation Mosaic
7	Open Shrublands	15	Snow and Ice
8	Woody Savannas	16	Barren or Sparsely Vegetated
		17	Water Bodies

Table 3.2: IGBP DISCover land cover legend.

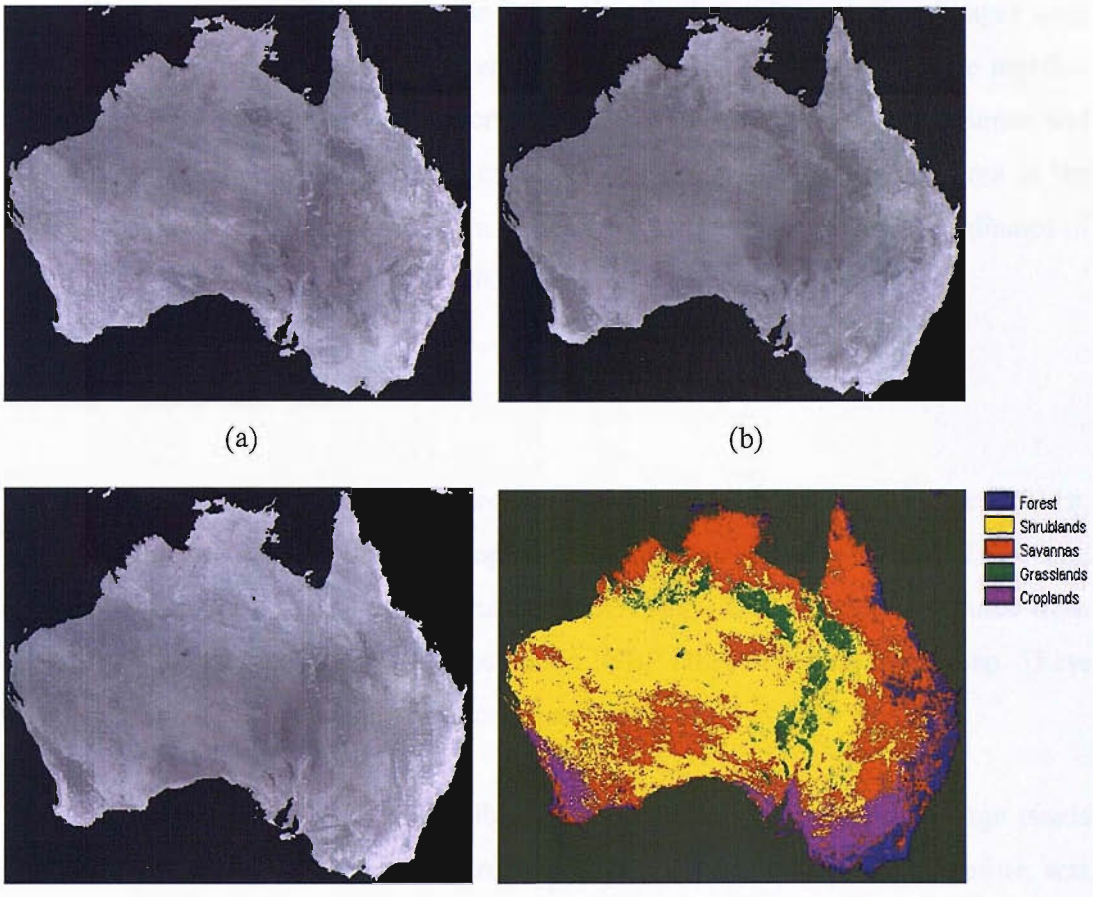


Figure 3.4. The data of the study area: (a) NDVI data in July 1992; (b) NDVI data in December 1992; (c) NDVI data in March 1993; (d) reference data derived from the IGBP DISCover land cover map.

3.3.2 Data processing

3.3.2.1 Co-registration of the IGBP DISCover land cover map and the NOAA AVHRR data

To enable the classification of the NOAA AVHRR data using the training and testing samples derived from the IGBP DISCover land cover map, it was necessary to co-register them. Since the IGBP land cover map and the NOAA AVHRR image had the same coordinate system, Interrupted Goode Homolosine map projection, the procedures to co-register them were as follows. Firstly, the IGBP DISCover land cover map with 1 km spatial resolution was degraded to the same spatial resolution with the NOAA AVHRR imagery (e.g. 8 km). Then, two identical points (e.g. one in the spatially degraded land cover map and the other in the remote sensing image) were identified to fit the degraded land cover map and the NOAA AVHRR image together by cutting the same studied areas from them to make the coordinates (columns and lines) of the two above identical points were equal. Finally, the studied area in the IGBP DISCover land cover map 1 km was cut following the geometric coordinates of the studied area in the NOAA AVHRR image.

3.3.2.2 Reference data

The set of land cover classes defined for the study comprised five classes: forest, shrubland, savanna, grassland, and cropland. The percentage cover of each of five land cover classes within each 8 km resolution NOAA AVHRR pixel was computed from the corresponding area of the 1 km resolution IGBP DISCover land cover map. These data were used as the reference data for the analysis.

As stated in section 2.8.2.3, the possible way to identify how mixed the image pixels was using the entropy values. The entropy of each AVHRR image pixel, therefore, was calculated based on Equation 2.39. The entropies of all NOAA AVHRR pixels were illustrated by the histogram in Figure 3.5. There were a number of mixed pixels in the AVHRR image (e.g. the pixels which have entropy larger than 0 are mixed pixels). This indicated that this data were suitable for soft classification.

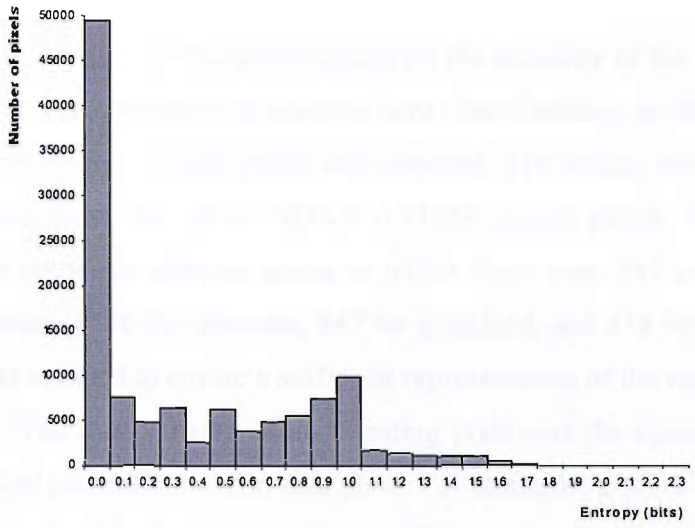


Figure 3.5: The entropies of NOAA AVHRR pixels.

3.4 Analysis for combining the outputs of hard classifications

3.4.1 Methodology

3.4.1.1 Training and testing data

The training data were selected manually from the study area such that they were spread over the image to capture the spectral variability in class response. Although the nature of the training set should depend on the classifier to be used, it is generally recommended that at least 10 training samples per class per-waveband be used (Congalton and Green 1999). Following this heuristic, a training set that contained a sufficiently large sample for each class was acquired. This training set contained 604 pure pixels, of which 70 pixels were forest, 179 were shrubland, 165 were savanna, 70 were grassland, and 120 were cropland.

A testing set was designed to assess accuracy of the individual classifications as well as the ensemble methods. The testing set comprised 250 pure pixels with 50 pixels for each class.

To understand the impacts of the mixed pixels on the accuracy of the individual hard classifications and the approaches to combine hard classifications, another independent testing set containing only mixed pixels was obtained. The testing mixed pixels were randomly selected from the set of NOAA AVHRR mixed pixels. This testing set comprised 5000 randomly selected pixels in which there were 397 pixels for forest, 2182 for shrublands, 1736 for savannas, 247 for grassland, and 438 for cropland. This large sample was selected to ensure a sufficient representation of the rare classes in the testing data set. The class labeled of each testing pixel was the class which has the highest land cover proportion within that pixel. For example, a pixel which contains 45% forest, 30% shrublands, 25% savannas, 0% grassland and 0% cropland was labeled as forest in the testing data set.

Two cases were analysed, hereafter referred to as case A and case B for simplicity (Table 3.3), generated from combinations of training and testing sets.

3.4.1.2 Methods

Four hard classifiers such as MLC, minimum distance, RBF, and MLP were trained by the selected training data to classify the NOAA AVHRR image. These classifiers were selected because they are different from each other in terms of the methodology and also commonly used classification algorithms. The first two methods were parametric classifiers while the others were non-parametric classifiers. The key concepts of these classification algorithms were introduced in section 2.6.3 and 2.6.4.

The classification outputs of the individual classifiers were assessed using the overall accuracy and individual class accuracy (e.g. user's accuracy and producer's accuracy) derived from the confusion matrix based approach (section 2.6.1.3). The outputs of four individual classifications were then combined by two approaches to combine hard classifications, the voting principle and Bayesian belief function, described in section 3.2.1. The classification outputs of combined algorithms were assessed using the same testing set as used for the individual classifiers in order to evaluate the potential of the approaches to combine hard classifications.

Case	Training set	Testing set
A	604 pure pixels with 70 pixels for forest, 179 for shrublands, 165 for savannas, 70 for grassland, and 120 for cropland.	250 pure pixels with 50 pixels for each class.
B	Same as in case A.	5000 mixed pixels with 397 pixels for forest, 2182 for shrublands, 1736 for savannas, 247 for grassland, and 438 for cropland.

Table 3.3: Training and testing sets for the analysis of combining hard classifications.

3.4.2 Results

Case A (testing set comprising only pure pixels)

Table 3.4 summarises the accuracies of four individual hard classifiers based on the testing set in case A. The MLP classifier produced the highest overall accuracy of 93.60%, followed by MLC with an overall accuracy of 93.20%, RBF of 91.60% and minimum distance of 89.20%. Table 3.5 shows the accuracies of two methods to combine the outputs of these four individual classifications.

	MLC		Minimum distance		RBF		MLP	
Class	Producer's (%)	User's (%)	Producer's (%)	User's (%)	Producer's (%)	User's (%)	Producer's (%)	User's (%)
Forest	96.00	97.96	94.00	95.92	88.00	97.78	94.00	95.92
Shrubland	88.00	91.67	88.00	77.19	88.00	86.27	92.00	90.20
Savanna	92.00	92.00	88.00	95.65	96.00	94.12	94.00	92.16
Grassland	96.00	97.96	82.00	95.35	92.00	100.00	96.00	100.00
Cropland	94.00	87.04	94.00	85.45	94.00	82.46	92.00	90.20
Overall accuracy (%)	93.20		89.20		91.60		93.60	

Table 3.4: Classification accuracies of individual classifications in case A.

	Voting principle		Bayesian belief function	
Class	Producer's (%)	User's (%)	Producer's (%)	User's (%)
Forest	96.00	96.00	96.00	97.96
Shrubland	92.00	90.20	92.00	93.88
Savanna	92.00	95.83	94.00	100.00
Grassland	96.00	100.00	100.00	100.00
Cropland	94.00	88.68	96.00	87.27
Overall accuracy (%)	94.00		95.60	

Table 3.5: Classification accuracy of two combined methods in case A.

The method called the Bayesian belief function to combine hard classifications produced the highest overall accuracy of 95.60% (e.g., 2.00% higher than MLP, the most accurate individual classification) and the highest individual class accuracy, producer's accuracy and user's accuracy when compared against the individual classifiers as well as the integrated voting principle. The voting principle yielded an overall accuracy of 94.00%, lower than the Bayesian belief function but still higher than the most accurate individual classifications (e.g., 0.4% higher than MLP). The producer's and user's accuracies of the voting principle were also higher than those of the individual classifications. The statistically significant difference between two proportions of the approach to combine classifications and the most accurate individual classification was evaluated using McNemar's test (Foody, 2004). However, the differences between the approaches to combine classifications (e.g., voting principle and Bayesian belief function) and the most accurate individual classification were insignificant (95% confidence, difference between proportions test derived from related sample).

Case B (testing set comprising only mixed pixels)

Table 3.6 summarises the accuracies derived from the confusion matrices of four individual classifications based on the testing set in case B. The results in this table show that MLP classifier produced the highest overall accuracy of 63.50%, followed by RBF with an overall accuracy of 60.08%, MLC of 57.80% and minimum distance of 57.54%.

	MLC		Minimum distance		RBF		MLP	
Class	Producer's (%)	User's (%)	Producer's (%)	User's (%)	Producer's (%)	User's (%)	Producer's (%)	User's (%)
Forest	79.85	55.71	78.59	55.71	75.82	56.58	78.09	54.39
Shrubland	59.40	70.59	64.80	67.49	55.59	76.43	63.20	79.71
Savanna	44.70	58.35	43.38	59.76	61.06	56.75	61.23	61.41
Grassland	80.97	51.55	57.89	40.51	65.59	55.86	65.18	58.33
Cropland	68.72	34.32	58.22	34.84	61.19	37.07	59.82	37.81
Overall accuracy (%)	57.80		57.54		60.08		63.50	

Table 3.6: Classification accuracies of individual classifications in case B.

Table 3.7 shows the same information about the accuracies of two methods to combine the outputs of four individual hard classifications. Two ensemble algorithms, the Bayesian belief function and the voting principle, produced the overall accuracies of 63.36% and 61.80%, respectively. In terms of the overall accuracy, they were slightly lower (about 0.14% and 1.70%, respectively) than the MLP, the most accurate individual classification, but still higher than other individual classifications. The overall accuracy of the combined output through the use of the Bayesian belief function was about 3% higher than RBF and about 6% higher than both MLC and Minimum distance. Moreover, there were increases in the individual class accuracies of the Bayesian belief function compared against those of the individual classifications. For example, the producer accuracies of class forest, shrublands, grassland and cropland of the Bayesian belief function were higher than those of the most accurate individual classification.

	Voting principle		Bayesian belief function	
Class	Producer's (%)	User's (%)	Producer's (%)	User's (%)
Forest	79.85	56.81	79.09	58.36
Shrubland	62.51	75.90	66.54	77.23
Savanna	55.59	61.00	55.01	64.31
Grassland	67.21	58.25	66.80	59.14
Cropland	63.47	35.73	64.38	34.47
Overall accuracy (%)	61.80		63.36	

Table 3.7: Classification accuracies of two combined methods in case B.

3.4.3 Discussion

Case A analysis

The overall accuracies of four individual classifications were very high (ranging from 89.20% to 93.60%) which implemented with only pure pixels in both the training and testing data sets. All two approaches (e.g. the voting principle and the Bayesian belief function) to combine the hard classification outputs increased accuracy. The increases in overall accuracy derived ranged from 0.04% to 2.00%, although the differences in accuracy between the ensemble approaches and the most individual classification were insignificant. Large increases in individual class accuracy were also observed. For example, the producer accuracies of class forest, grassland and cropland were 94%, 96%, and 92%, respectively in MLP (the most accurate individual classification); while these figures were reached to 96%, 100%, and 96%, respectively in the Bayesian belief function.

Case B analysis

In case B four individual classifiers were trained by the same training data set as for case A but the testing set contained only mixed pixels. The overall accuracies of them were much lower (e.g. about 30% lower) than those in case A. In addition, the overall accuracies of the combined outputs did not increase compared with the most accurate individual classification. These poor results of the individual classifications were

expected as hard classification is inappropriate for the mixed pixel problem. Hard classification assigned pixel to only one land cover class based on the class of which it has the most similarity. For example, a pixel with only 45% forest may be classified as forest.

The reason that led to the reduction of the overall accuracies of the two combined algorithms compared with the highest individual classifier may be because of the low accuracies of class savannas in MLC and minimum distance (its producer accuracies were 44.70% and 43.38%, respectively). In the combined results, these figures were reached to about 56%. Furthermore, the effect of mixed pixels may be the additional reason for this reduction in case B compared against the increase in the overall accuracy of combined methods in case A. There was about 6% different between the overall accuracies of the two lowest individual classifications (Minimum distance and MLC) and the highest individual classification (MLP) in case B while this number was only about 3% in case A.

Although the overall accuracies of combined outputs did not increase compared with the most accurate individual classification, they were still much higher than the rest of the individual classifications. Moreover, there were large increases in class accuracy of combined outputs compared against those in the individual classifications. For instance, the producer accuracies of class forest, shrublands, grassland and cropland were 78.09%, 63.20%, 65.18%, and 59.82%, respectively in the most accurate individual hard classification, these numbers in the Bayesian belief function, however, were reached to 79.09%, 66.54%, 66.80%, and 64.38%, respectively.

In summary, the results from the two above analyses highlighted that ensemble based approaches may be used to increase hard classification accuracy and the combined method, the Bayesian belief function, yielded higher classification accuracy than the voting principle.

3.5 Analysis for combining soft classifications

A problem with the NOAA AVHRR imagery is that the natural scale of spatial variation in land cover in Australia is usually finer than the scale of sampling imposed

by the image pixels. Consequently, many NOAA AVHRR image pixels contain a mixture of land cover types (Figure 3.4). The following analyses were designed to evaluate the approaches to combine the individual soft classifications of the NOAA AVHRR image.

3.5.1 Methodology

3.5.1.1 Training and testing data

The training data set was the same as the one used in the analysis of combining hard classifications described in Table 3.3.

Since the focus of this work was on soft classification, attention was focused on only the accuracy with which mixed pixels were classified. The testing set used to assess the accuracy of the classifications derived from each of the three soft classifiers contained 5000 mixed pixels drawn randomly from the NOAA AVHRR data. This large sample was selected to ensure a sufficient representation of the rare classes in the testing data set. This testing set was used to evaluate the accuracy of each individual classification derived and inform the production of the ensemble. To avoid bias in evaluating the accuracy of the classification derived from an ensemble approach, a further independent testing set was selected to assess the accuracy of the combined output. This second testing set also comprised 5000 mixed pixels which had been selected at random from the NOAA AVHRR data but excluding pixels contained in the first testing set. Hereafter, the testing set used to assess the accuracy of the individual classifications is referred to as testing set 1, while that used for the evaluation of the output from the ensemble approaches is referred to as testing set 2 (Table 3.8).

	Training	Testing set 1	Testing set 2
Composition	604 pure pixels with 70 pixels for forest, 179 for shrublands, 165 for savannas, 70 for grassland, and 120 for cropland	5000 random mixed pixels	5000 random mixed pixels
Purpose	To train the individual soft classifiers used to classify the AVHRR image by	To assess the accuracy of the individual soft classifications	To assess the accuracy of the combined classification output

Table 3.8: Training and testing sets for the analysis of combining soft classifications.

3.5.1.2 Algorithms used

3.5.1.2.1 Individual soft classifications

Three classifiers were applied to the NOAA AVHRR data. The set of classifiers used comprised two neural networks, MLP and RBF network, and a probabilistic classifier, the Bayesian classifier. These three soft classifiers were briefly described in section 2.8.4 and 2.8.5. They were selected as the candidates for use in an ensemble because they have been used in other studies to derive accurate classifications (e.g. Atkinson *et al.* 1997, Zhang and Foody 2001, Eastman and Laney 2002) and are different in terms of their methodology. Each individual classification was undertaken using the same training data set.

The Bayesian classifier (Eastman and Laney 2002) is similar to the maximum likelihood classification, the most popular parametric classifier used in remote sensing. A soft classification may be derived from the Bayesian classification by computing the posterior probabilities of membership to all classes for each pixel. The derived posterior probabilities of membership are considered here to represent the proportional cover of the classes in each 8 km resolution pixel.

The MLP (Atkinson and Tatnall 1997, Bernard *et al.* 1997) and RBF (Bruzzone and Prieto 1999, Bastos *et al.* 1999) are non-parametric classifiers. A soft classification

may be derived from these classifiers by outputting the activation levels of the network output units for each pixel. These activation levels range from 0 to 1, and may be used as the measures of class membership that reflect the class composition of the pixel (Zhang and Foody 2001).

The three individual soft classifiers were undertaken with the same training and testing data sets. The multi-layer perceptron, comprising 3 input, 16 hidden and 5 output nodes, was used to derive a fuzzy classification with the errors in training and testing of 0.095 and 0.272, respectively. Meanwhile, the radial basis function network, consisting of 3 input, 17 hidden and 5 output nodes, had errors of 0.183 and 0.274 in training and testing respectively.

3.5.1.2.2 Assessing the accuracy of soft classifications

The accuracy of the three individual soft classifications was assessed using testing set 1. As stated in section 2.8.2, there were a number of methods to assess the accuracy of soft classification. Although all of these methods have something to offer, none is at the moment universally applicable. Advantages and disadvantages coexist in each of them. Here, the accuracy of the soft classifications was evaluated using the generalized cross-tabulation matrix approach proposed by Pontius and Cheuk (2006). This approach was chosen because of the following reasons:

- The confusion matrix is the most popular way to assess the accuracy of classification and therefore, people are familiar with it. The structure of the generalized cross-tabulation matrix using composite operator is the same with that of the traditional confusion matrix which are used to assess the accuracy of hard classification.

- The measures such as producer's accuracy, user's accuracy, and the overall accuracy derived from the generalized cross-tabulation matrix may be used to assess the accuracy of the soft classification. The producer's and user's accuracy indicate the proportions of agreement of each land cover class and the overall accuracy indicates the total agreement between the predicted data and the reference data. The matrix also

informs about bias using the totals of column and totals of row which are corresponding to the total grades of reference and classification data, respectively.

- The statistical significant difference between two classifications can be evaluated using two overall proportions of agreement derived from the generalized cross-tabulation matrix.

The accuracy measures derived from the generalized cross-tabulation matrix were overall accuracy, individual class accuracy, producer's accuracy and user's accuracy. Accuracy was also expressed on a per-class basis using class correlation coefficient. The correlation coefficient per class is a measure of the correlation between the predicted and actual (ground data) coverage (Atkinson 1999).

3.5.1.2.3 Combining individual soft classifications and evaluate the ensemble approaches

Based on the accuracy of the individual soft classifications, the classification outputs were combined using the four ensemble approaches described in section 3.2.2. The potential of the ensemble approaches for increasing accuracy was assessed by comparing the classification accuracy of the ensemble classification with the individual classifications. The statistical significance of differences in classification accuracy between the combined outputs and the individual classifications was determined. In this research, testing set 1, which was used to assess the accuracy of the three individual classifications, was independent from testing set 2 which was used to assess the accuracy of the ensemble approaches. The statistical significance of differences between the overall accuracy of the most accurate individual classification and that derived from the application of an ensemble approach may, therefore, be estimated using a difference between proportions test (Foody, 2004). The statistical significance of a difference between two proportions in the independent samples may be estimated from:

$$z = \frac{\frac{G_1}{T_1} - \frac{G_2}{T_2}}{\sqrt{P(1-P)\left(\frac{1}{T_1} + \frac{1}{T_2}\right)}} \tag{3.11}$$

where G_1 and G_2 represent the total grades of membership to be correctly allocated in two independent samples with the total grades of membership in reference data T_1 and T_2, respectively; and $P = (G_1 + G_2)/(T_1 + T_2)$.

To assess whether the accuracy of the combined classification outputs was higher than that of the individual classifications, one-tailed tests for the significance of difference were undertaken. The significance of difference between the two proportions was evaluated by comparing the value of z calculated in the difference between proportions test against tabulated values. For example, with a one-tailed test, the null hypothesis, of no significant difference, would be rejected at the 95% level of confidence if $|z| > 1.64$ (Congalton *et al.*, 1983; Foody, 2004).

3.5.2 Results

3.5.2.1 Individual soft classification outputs

Soft classifications predict the proportional cover of each land cover class within each pixel (section 2.8). Ideally, each proportion is on the range [0, 1] and the total of all class proportions within each pixel is equal to 1. However, for some classifiers, the proportional results may have negative values and their sum in each pixel may not be equal to 1. This happened to the RBF and MLP in the experiment. For these cases the proportions of land cover classes within each pixel can be rescaled to get fractional values that are positive and for which the sums are equal to 1. This study used rescaled outputs of the individual soft classification for the analyses.

The accuracy of the three individual soft classifications was assessed using testing set 1. In terms of the overall accuracy, the classifications were broadly similar with an accuracy of about 61% (Table 3.9). On a per-class basis, however, differences in accuracy were more apparent. This was evident with the user's and producer's accuracy derived from the generalized cross-tabulation matrix (Table 3.9) and the correlation coefficient between the predicted and actual (ground data) coverage (Table 3.10).

	Actual data						
Predicted data	Forest	Shrubland	Savanna	Grassland	Cropland	Total	User's accuracy (%)
(a) Bayesian classification (Overall accuracy = 61.96%)							
Forest	6.49	0.13	3.88	0.00	0.75	11.26	57.66
Shrubland	0.01	26.22	6.93	0.73	0.19	34.08	76.93
Savanna	0.26	7.72	18.84	0.61	1.51	28.95	65.10
Grassland	0.01	2.90	0.16	4.31	0.00	7.38	58.39
Cropland	1.33	5.23	5.62	0.04	6.10	18.32	33.31
Total	8.10	42.19	35.44	5.69	8.56	99.98	
Producer's accuracy (%)	80.10	62.14	53.15	75.70	71.32		
(c) RBF (Overall accuracy = 61.27%)							
Forest	6.47	1.76	3.14	0.24	0.87	12.48	51.82
Shrubland	0.14	22.90	3.55	0.80	0.41	27.80	82.39
Savanna	0.24	8.83	22.56	0.56	1.54	33.73	66.89
Grassland	0.13	3.07	1.15	3.90	0.31	8.57	45.57
Cropland	1.12	5.63	5.06	0.18	5.43	17.43	31.19
Total	8.10	42.19	35.46	5.69	8.56	100.000	
Producer's accuracy (%)	79.82	54.28	63.63	68.62	63.50		
(e) MLP (Overall accuracy = 61.20%)							
Forest	5.91	0.07	4.38	0.00	1.00	11.36	52.06
Shrubland	0.00	25.98	5.69	1.90	0.32	33.89	76.65
Savanna	0.94	10.51	21.05	0.34	2.39	35.23	59.74
Grassland	0.00	1.85	0.27	3.42	0.00	5.54	61.75
Cropland	1.25	3.79	4.07	0.03	4.84	13.98	34.62
Total	8.10	42.19	35.46	5.69	8.56	100.000	
Producer's accuracy (%)	72.97	61.57	59.36	60.15	56.56		

Table 3.9: Accuracy of the individual soft classifications.

Class	Bayesian	RBF	MLP
Forest	0.782	0.810	0.723
Shrubland	0.573	0.610	0.624
Savanna	0.412	0.485	0.479
Grassland	0.777	0.768	0.729
Cropland	0.511	0.504	0.458
Average	0.611	0.636	0.603

Table 3.10: Correlation coefficients between the estimated and actual class coverage from the individual classifications; all correlation coefficients were significant at the 99.9% level of confidence.

3.5.2.2 Combination by selecting the most accurate predictions

In this ensemble approach, the class accuracies among the three individual soft classifiers were compared to derive the final combined classification outputs. Accuracy was expressed on a per-class basis using the producer's accuracy and user's accuracy derived from the generalized cross-tabulation matrix together with the correlation coefficients. The accuracies of the combined outputs derived from this ensemble approach are shown in Tables 3.11-3.13. Irrespective of the measure of accuracy used for the comparison, all of the ensemble approaches increased classification accuracy significantly.

Table 3.11 presents the classification accuracy of the ensemble approach that selects the most accurate predictions on a class-specific basis based on class correlation coefficient. The overall accuracy of this ensemble approach was about 2.2% higher than that of the Bayesian, the most accurate individual classification. The difference in accuracy between the ensemble approach and the most accurate individual classification was statistically significant at the 95% level of significance ($z = 2.28$).

Table 3.12 shows the classification accuracy of the ensemble approach that selects the most accurate predictions on a class-specific basis of the producer's accuracy. It was apparent that the overall accuracy of the combined output based on the comparison of class producer's accuracy was 2.73% higher than that of the Bayesian, the most accurate individual classification. The difference between two overall accuracies of

this ensemble approach and the most accurate individual classification was statistically significant at the 99% level of significance ($z = 2.83$).

From Table 3.13, it was apparent that there was a large difference in accuracy between the ensemble approach that selects the most accurate predictions based on class user's accuracy and the most accurate individual classification. Their difference in the overall accuracy was about 4.45% and the statistical significance of the difference was at the 99.9% level of significance ($z = 4.46$).

	Actual data						
Predicted data	Forest	Shrubland	Savanna	Grassland	Cropland	Total	User's accuracy (%)
Forest	5.74	1.96	3.73	0.23	0.80	12.45	46.10
Shrubland	0.00	22.96	2.74	0.83	0.22	26.75	85.83
Savanna	0.33	11.15	25.75	0.63	1.48	39.34	65.46
Grassland	0.00	1.89	0.10	4.00	0.00	5.99	66.80
Cropland	1.03	4.63	4.04	0.07	5.71	15.47	36.91
Total	7.10	42.58	36.36	5.75	8.21	100.00	
Producer's accuracy (%)	80.84	53.92	70.83	69.60	69.53		
Overall accuracy (%)	64.16						
z value	2.28						

Table 3.11: Accuracy of the ensemble approach that selects the most accurate predictions based on class correlation coefficient.

	Actual data						
Predicted data	Forest	Shrubland	Savanna	Grassland	Cropland	Total	User's accuracy (%)
Forest	5.46	0.11	3.17	0.00	0.61	9.35	58.37
Shrubland	0.00	25.47	5.32	0.68	0.20	31.67	80.42
Savanna	0.44	9.45	23.63	0.73	1.53	35.79	66.01
Grassland	0.00	2.53	0.10	4.26	0.00	6.90	61.81
Cropland	1.20	5.02	4.13	0.07	5.87	16.28	36.04
Total	7.10	42.58	36.36	5.75	8.21	100.00	
Producer's accuracy (%)	76.87	59.82	64.99	74.19	71.46		
Overall accuracy (%)	64.69						
z value	2.83						

Table 3.12: Accuracy of the ensemble approach that select the most accurate predictions based on class producer's accuracy.

	Actual data						
Predicted data	Forest	Shrubland	Savanna	Grassland	Cropland	Total	User's accuracy (%)
Forest	5.68	0.11	3.51	0.00	0.91	10.22	55.60
Shrubland	0.16	27.26	4.48	1.65	0.56	34.11	79.92
Savanna	0.55	11.24	25.77	0.92	2.16	40.63	63.42
Grassland	0.00	1.05	0.16	3.11	0.00	4.32	72.06
Cropland	0.72	2.93	2.44	0.06	4.59	10.72	42.76
Total	7.10	42.58	36.36	5.75	8.21	100.00	
Producer's accuracy (%)	79.97	64.02	70.88	54.19	55.84		
Overall accuracy (%)	66.41						
z value	4.46						

Table 3.13: Accuracy of the ensemble approach that select the most accurate predictions based on class user's accuracy.

3.5.2.3 Combination based on average operator

The outputs of each land cover class among three individual soft classifications were averaged to construct the final combined output. The accuracy derived from the generalized cross-tabulation matrix of this ensemble approach is shown in Table 3.14. The overall accuracy of the averaging combination approach was 3.95% higher than that of the most accurate individual classification. The increases in individual class accuracy were also observed. For example, the producer's accuracy of class forest, shrublands, and savannas increased compared with those of the most accurate individual classification, the Bayesian. The difference in overall accuracy between the averaging ensemble approach and the most accurate individual classification was statistically significant at the 99.9% level of significance ($z = 4.11$).

3.5.2.4 Combination based on D-S theory of evidential reasoning

The result of this ensemble approach is presented in Table 3.15. The overall accuracy of the combined output based on the D-S theory of evidential reasoning approach was 3.61% higher than that of the most accurate individual classification. Increases in individual class accuracy were also observed. For example, the user's accuracy of class

shrublands, savannas, grasslands and croplands increased compared with those of the most accurate individual classification. The difference in overall accuracy between this ensemble approach and the most accurate individual classification was statistically significant at the 99.9% level of significance ($z = 3.76$).

	Actual data						
Predicted data	F	SH	SV	G	C	Total	User's (%)
F	5.76	0.63	3.39	0.10	0.62	10.51	54.84
SH	0.04	27.03	4.50	0.90	0.31	32.78	82.47
SV	0.34	7.92	23.32	0.51	1.52	33.61	69.41
G	0.05	2.24	0.49	4.12	0.09	6.99	58.96
C	0.91	4.75	4.65	0.12	5.67	16.11	35.22
Total	7.10	42.57	36.36	5.75	8.21	99.99	
Producer's (%)	81.14	63.48	64.15	71.74	69.08		
Overall accuracy	65.91%						
z value	4.11						

Table 3.14: Accuracy of the ensemble approach that averages the predictions of the individual classifiers.

	Actual data						
Predicted data	F	SH	SV	G	C	Total	User's (%)
F	5.66	0.30	3.63	0.05	0.65	10.29	54.99
SH	0.02	27.26	4.94	1.12	0.25	33.60	81.13
SV	0.42	8.38	23.07	0.47	1.69	34.03	67.79
G	0.02	2.12	0.31	4.01	0.04	6.50	61.63
C	0.98	4.53	4.41	0.09	5.57	15.58	35.76
Total	7.10	42.58	36.36	5.75	8.21	100.00	
Producer's (%)	79.68	64.02	63.44	69.74	67.87		
Overall accuracy	65.57%						
z value	3.76						

Table 3.15: Accuracy of the ensemble approach using D-S theory of evidential reasoning.

3.5.2.5 Combining soft classifications through a conventional ensemble approach

In this ensemble approach, the classification outputs for each land cover class were combined separately. To illustrate the potential of this ensemble approach, attention focuses on the forest class as an example. The proportions of forest within each pixel of three individual soft classifications were degraded into a set of 5 ordered classes. The ordered classes were 'very small' if forest proportion lay in the range [0.0–0.19], 'small' [0.2–0.39], 'medium' [0.4–0.59], 'large' [0.6–0.79], or 'very large' [0.8–1.0] (Table 3.16). These degraded outputs were then combined by a conventional ensemble approach, the voting principle. The accuracies of the three individual classifications as well as of the ensemble output were estimated from the traditional confusion matrix using testing set 1 (Table 3.17).

The MLP was the most accurate individual classification with the overall accuracy of 87.42%, followed by Bayesian classification with an accuracy of 87.12% and RBF with an accuracy of 86.36%. The accuracy with which the ordered classes were classified varied markedly. A high accuracy was achieved for both the very small and very large classes but low accuracies for the other intermediate classes. The overall accuracy of the ensemble approach was 0.62% higher than that of MLP, although the difference was statistically insignificant (95% level of confidence). Although the difference in accuracy was insignificant the results indicate that conventional ensemble approaches may be used to combine soft classification output. The use of other methods that accommodate the ordered nature of the classes in the analysis may yield greater increases in accuracy.

Ordered class	Forest proportion
Very small	0.0 – 0.19
Small	0.2 – 0.39
Medium	0.4 – 0.59
Large	0.6 – 0.79
Very large	0.8 – 1.00

Table 3.16: The ranges used to degrade the soft classification output to ordered classes.

	Individual classifications						Combined method	
	Bayesian		RBF		MLP		Voting principle	
Class	Producer's accuracy (%)	User's accuracy (%)	Producer's accuracy (%)	User's accuracy (%)	Producer's accuracy (%)	User's accuracy (%)	Producer's accuracy (%)	User's accuracy (%)
Very small	95.76	97.08	94.35	98.03	96.57	95.78	96.36	96.34
Small	09.84	17.65	21.86	16.81	01.64	23.08	13.11	22.43
Medium	07.43	14.10	22.30	16.02	03.38	19.23	13.51	33.90
Large	09.63	15.85	23.70	17.68	01.48	10.00	08.15	52.38
Very large	78.92	30.54	55.42	53.80	86.14	26.63	83.13	31.08
Overall accuracy (%)	87.12		86.36		87.42		88.04	

Table 3.17: Classification accuracies of the degraded outputs from three individual soft classifications and of the combined output from these degraded outputs.

3.5.3 Discussion

Through the use of the ensemble approach that selects the most accurate predictions on a class-specific basis based on class user's accuracy, the overall accuracy was 4.45% higher than that of the most accurate individual soft classification, while based on class producer's accuracy, and correlation coefficient these figure were 2.73%, and 2.20%, respectively.

Using the approach that averages the outputs of soft classifications, the overall accuracy of the combined output was 3.95% higher than that of the most accurate individual soft classification and large increases in individual class accuracy were also observed.

The direct combination of soft classifications using D-S theory of evidence also increased classification accuracy. The overall accuracy of the combined output was 3.61% higher than the most accurate individual classification and the increases in individual class accuracy were also achieved.

Moreover, the differences in overall accuracy between the ensemble approaches and the most accurate individual classification were statistically significant at the 95% to 99.9% level of significance.

Focusing on per-class accuracy, the outputs of class forest in three individual soft classifications were degraded and then combined using a conventional ensemble approach, the voting principle. The accuracy of forest in the combined degraded output increased 0.62% compared with that of the most accurate individual classification. This ensemble approach lost some information of soft classification since it made the soft classification output degrade to a set of order classes. However, the spatial resolution of degraded output was still finer than that of the hard classification and this ensemble method may be more useful in case of focusing on one-class classification.

In summary, all four above approaches to combine the soft classification outputs increased accuracy. The D-S theory of evidential reasoning approach was the most reliable one. It can accommodate the uncertainty of the individual classifications based on their classification confidences. The classification confidences were then used for combined process, whereas in other ensemble approaches such as selecting the most accurate predictions on a class-specific basis, averaging individual soft classification outputs, each individual classification was treated equally.

3.6 Summary and conclusions

The possible approaches to increase the accuracy of land cover classification in general and of soft land cover classification in particular through the use of an ensemble of different classifications were investigated. For this purpose a coarse spatial resolution NOAA AVHRR imagery of Australia was used.

The AVHRR image was fist hard classified using two parametric classifiers (e.g. MLC and minimum distance) as well as two non-parametric neural network classifiers (e.g., RBF and MLP). They were all trained by the same training data and tested by the same testing data which comprised only pure pixels. Overall accuracy was assessed using confusion matrix based approach and showed that the accuracy of the individual classifications ranged from 89.2% to 93.6%. Both approaches to combining hard

classification outputs (e.g. the voting principle and Bayesian belief function) increased accuracy in comparison with the most accurate individual classifications. The increases in overall accuracy ranged from 0.04% to 2% and large increases in individual class accuracy were also observed although the differences between two overall accuracies of the ensemble approaches and the most accurate individual classification were insignificant (95% confidence, difference between proportions test derived from related sample).

To study the impacts of the mixed pixel problem on the accuracy of the individual hard classifications and the approaches to combine hard classifications, another testing set which contained only mixed pixels was used to assess the accuracy of the individual hard classifications and the combined outputs. The overall accuracy of the individual classifications ranged from 57.8% to 63.5% and that of the combined outputs ranged from 61.8% to 63.4%. The reason that caused the low accuracy of the individual classifications as well as the combined outputs may be the effect of mixed pixels in the testing set which was explained in section 3.4.1. Although the overall accuracies of combined outputs did not increase compared with the most accurate individual classification, they were still much higher than the rest of the individual classifications and increases in class accuracy were also achieved.

To evaluate the potential of the approaches to combine soft classifications, the NOAA AVHRR imagery of Australia was soft classified using two neural networks (RBF and MLP) as well as a probabilistic classifier (Bayesian). The classification accuracy of the three individual soft classifications was assessed using the testing set 1. Overall classification accuracy was assessed using the generalized cross-tabulation matrix based approach and showed that the overall accuracies of the individual soft classifications were similar at about 61%. Accuracy was also expressed on a per-class basis using class producer's accuracy, user's accuracy and correlation coefficient.

Four methods to combining soft classification outputs were assessed. These methods were (i) an approach that selects the most accurate predictions on a class-specific basis, (ii) an approach that averages the outputs of the individual classifications (iii) the direct combination of classifications using evidential reasoning and (iv) an approach which degrades the soft classification outputs into a set of ordered classes and then

combines these using a conventional ensemble approach. Based on the accuracy of the individual soft classifications, their classification outputs were combined using these four ensemble approaches.

The combined classification outputs were then assessed with the testing set 2 using the generalized cross-tabulation matrix approach. Apart from comparing the accuracy measures, the statistical significance of differences in classification accuracy between the ensemble approaches and the most accurate individual classification were also evaluated. This was undertaken by comparing the value of z calculated in the difference between proportions test against tabulated values. The results highlighted that all four approaches to combining the soft classification outputs have been shown to be able to increase soft classification accuracy.

The original outputs of the soft classifications were used in three of the four ensemble approaches. These three ensemble approaches were the selection of the most accurate prediction on a per-class basis, the use of the average of the outputs of the individual soft classifications, and the combination of the outputs through evidential reasoning. The increases in overall accuracy of these three ensemble approaches derived ranged from 2.20 to 4.45% and large increases in individual class accuracy were also observed. Moreover, the differences in overall accuracy between the ensemble approaches and the most accurate individual classification were statistically significant at the 95% to 99.9% level of significance.

The fourth ensemble approach in which the individual soft classification outputs were degraded into a set of ordered classes and then combined using a conventional ensemble approach was also shown to be able to increase classification accuracy, although the difference was insignificant at 95% level of confidence. This approach, however, lost some information of soft classification output in the transition from a continuous representation to one based on a set of order classes.

The results showed that the individual classifiers yielded classifications of similar overall accuracy (the range in accuracy was only 0.62%) but varied markedly in terms of the accuracy with which the individual classes were classified. This gave the potential to increase accuracy through an ensemble approach based on using the best feature (in terms of classification accuracy) of the individual classifiers in the

derivation of the final classification. The results highlighted that ensemble approaches of different classifications to increase classification accuracy may not only apply for hard classifications but also for soft classifications.

Although the ensemble of soft classifications can increase soft classification accuracy, the accuracy of land cover classification derived from soft classifications themselves as well as from their ensemble approaches was still low. The next research, therefore, plan to investigate the factors that may impact on the soft classification predictions and its implications for sub-pixel scale change detection and super-resolution mapping.

Chapter 4

Impacts of intra-class spectral variability on soft classification prediction and its implications for sub-pixel scale change detection

4.1 Introduction

Mixed pixels are one of the main problems limiting the accuracy of mapping land cover from remotely sensed imagery. Soft classification techniques allow for the partial and multiple class membership within each mixed pixel, and, therefore, may be used to refine the standard mapping process to increase the accuracy of land cover mapping from remote sensing (Foody and Cox 1994, Atkinson *et al.* 1997, Tso and Mather 2001). The output of a soft classification is typically a set of proportion images that show the predicted coverage of each thematic class in the area represented by each pixel. These proportion images can provide an accurate estimate of class composition. They also support to the representation of environmental continua and the detection of subtle land cover changes when used in post-classification analyses (Foody 2001, Haertel *et al.* 2004). Although such soft classifications can reduce some of the problems associated with mixed pixels there are still concerns. In particular, the accuracy of soft classification is often low (section 3.1). One possible solution to this problem was to combine the outputs of different soft classifiers. This was evaluated in section 3.5 and all the ensemble approaches investigated were found to increase classification accuracy. However, the accuracy of land cover classification after the combination of soft classifications was still low. This may be a major limitation to studies attempting to predict class proportional cover and its change over time.

In soft classification, each image pixel is provided with only one prediction which indicates the measures of class membership of each land cover class within the pixel. For example, if three land cover classes exist within one pixel then the prediction might be 25%, 35% and 40%, respectively for three land cover classes. The single prediction of soft classification has been used for a number of applications. For

example, by using a pair of soft classifications for land cover change detection the estimated class proportions are typically compared directly (Foody 2001, Haertel *et al.* 2004). This trust in the single set of class proportion estimates may be unwise. Specifically, this trust seems to be based on the assumption that a class can be represented by a single spectral end-member. This is unrealistic as classes typically display a degree of spectral variability since the spectral signatures of land cover classes vary from pixel to pixel, often due to changes in biophysical (e.g., leaves, stems and bark) and biochemical (e.g. chlorophyll content) composition (Bateson *et al.*, 2000, Song, 2005). In fact, it is known that the accuracy of soft classification is negatively related to the degree of intra-class spectral variation (Petrou and Foschi, 1999) and approaches to refine unmixing methods to accommodate for this have been investigated (Bateson *et al.*, 2000, Song, 2005). However, it is still common to see basic approaches to unmixing being employed and the sub-pixel estimates used in a manner that places great confidence in their accuracy. This chapter aims to investigate the impacts of intra-class spectral variability on soft classification prediction and highlight its implications for analyses based on soft classification such as the detection of land cover change.

The outline of the chapter is:

1. To introduce the linear mixture model applied to sub-pixel estimation.

2. To study the impacts of intra-class spectral variation on soft classification prediction.

3. To explore the impacts of intra-class spectral variation on change detection.

4. To investigate the ensemble of sub-pixel classifications.

5. Finally the summary and conclusions will close the chapter.

4.2 Linear mixture model

Linear mixture model (LMM) was briefly described in section 2.8.3. This section presents more details about this model in support of the estimation of sub-pixel class composition.

The LMM is an established technique which is applied to remotely sensed imagery to estimate sub-pixel proportions of ground cover classes. It is based on an assumption of linear mixing between the spectral responses of each land cover type. This means the spectral response of an individual pixel is the linear sum of the spectral response of the individual ground components (e.g., land cover classes), weighted by their relative proportions on the ground. Mathematically, the model may be expressed as,

$$\mathbf{x} = \mathbf{M}\mathbf{f} + \mathbf{e} \tag{4.1}$$

where $x = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{pmatrix}$ is the observation vector for a pixel (*i.e.*, the spectral response or digital values) of a given pixel observed in each of m spectral wave bands.

$f = \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_c \end{pmatrix}$ is the vector of ground cover proportions for each of c land cover classes.

e is the error term expressing the differences between the reflectance observation vector x and the reflectance for that pixel calculated from the model (Settle and Drake, 1993).

$M = \begin{pmatrix} M_{11} & M_{12} & \cdots & M_{1c} \\ M_{21} & M_{22} & \cdots & M_{2c} \\ \cdots & \cdots & \cdots & \cdots \\ M_{m1} & M_{m2} & \cdots & M_{mc} \end{pmatrix}$ is the $m \times c$ end-member spectra matrix with c is the number of land cover classes. In the end-member spectra matrix, each column is a

vector which represents the pure spectral signatures given by c land cover classes in the absence of noise.

There are a number of approaches used in the literature to identify the spectral response of end-members (Van Der Meer and De Jong, 2000; Theseira *et al.*, 2003). Although the specific techniques for selecting end-members may differ, the spectral responses of end-members are commonly derived from two sources: (1) reference end-members, whose spectral responses are derived from a set of laboratory or field spectra measured in units of reflectance (Adams *et al.*, 1995, Bryant, 1996), or (2) image end-members, whose spectral responses are identified from remotely sensed imagery to be classified (Foody and Cox, 1994, Atkinson *et al.*, 1997). Image end-members are often achieved from the average spectral response of pure pixels of each class in the image to be unmixed (Foody and Cox, 1994, Roberts *et al.*, 2002, Theseira *et al.,* 2003) or from a feature space (Theseira *et al.,* 2003, Wu and Murray, 2003). Identifying image end-member spectra can sometimes be difficult because the spectral signature of an end-member may vary significantly (Theseira *et al.,* 2003, Song, 2005). An approach for manual end-member selection from an image was investigated by Bateson and Curtis (1996). However, due to the class intra-spectral variability different users may provide different end-member spectral values for the same image. Variation in class spectral responses is a major problem in LMM since it has a negative effect on the accuracy of class composition estimates (Carpenter *et al.*, 1999, Petrou and Foschi, 1999, Theseira *et al.*, 2003).

Two more constraints are also generally added to the mixture model:

$$\sum_{j=1}^{c} f_j = 1 \text{ and } 0 \le f_j \le 1 \text{ for all } j \text{ classes} \tag{4.2}$$

The first constraint indicates that the total class proportions within one pixel is 1 and the second constraint indicates that all class proportions of mixed pixels must take values between 0 and 1.

Equations (4.1) and (4.2) form a linear system of equations. With multispectral data with m spectral bands, there would be m+1 equations available to solve for the various end-members.

In the LMM, generally x and M are known and the aim is to estimate the vector f which is the class proportions of pixels (Settle and Drake 1993). The solution of the system of equations is subject to the constraint that the number of estimated parameters is less than the number of equations, $c < m+1$, where c is the number of end-members (land cover classes). The vector land cover proportion f can be estimated using least-squares methods. That means that, with respect to equation 4.1, the sum of squares error e can be minimized by carrying out a standard least square fit:

$$Q(x, f) = (x - Mf)^{\mathrm{T}} (x - Mf) = e^{\mathrm{T}} e \tag{4.3}$$

Without two constraints of f in equation (4.2), in the case the error e is independent and normally distributed, the least squares solution is estimated by minimizing the quadratic Q by setting to zero its partial derivatives with respect to each of the f_i as follows:

$$\frac{\partial Q}{\partial f_1} = \frac{\partial Q}{\partial f_2} = \ldots = \frac{\partial Q}{\partial f_c} = 0 \tag{4.4}$$

where ∂ is the partial derivative.

The estimation of the proportion f is given by:

$$f = \left(M^T M\right)^{-1} M^{\mathrm{T}} x \tag{4.5}$$

The solution for the least square method in the case where the error e is correlated was discussed by Settle and Drake (1993).

4.3 Impacts of intra-class spectral variability on soft classification prediction

4.3.1 Data and methods

The impacts of intra-class spectral variability on soft classification prediction were explored using simulated data, so as to control the variables impacting on the remotely sensed response. As the major aim was to illustrate the effects of intra-class variation on soft classification prediction, a simple data set was generated. The data set comprised of three classes (classes A, B, and C) and four simulated spectral wavebands to accommodate for the dimensionality constraint in unmixing (Settle and Drake, 1993). For simplicity, the data for each class in each waveband were normally distributed. Keeping the class centroids constant, further data set with different level of intra-class variation were constructed to illustrate the impacts of differences in intra-class variation, including co-variation, in the spectral feature space on the accuracy of soft classification estimation. The key parameters describing the data are summarised in Tables 4.1 - 4.3.

Class	Mean	Variance-covariance matrix
A	[380 490 300 320]	$\begin{bmatrix} 64 & 50 & 40 & 50 \\ 50 & 100 & 45 & 64 \\ 40 & 45 & 49 & 45 \\ 50 & 64 & 45 & 100 \end{bmatrix}$
B	[310 335 235 260]	
C	[250 410 180 390]	

Table 4.1: Summary of simulated classes.

Class	Mean	Variance-covariance matrix		
		Small variability	Medium variability	Large variability
A	[380 490 300 320]	$\begin{bmatrix} 4 & 3 & 2 & 3 \\ 3 & 6 & 3 & 4 \\ 2 & 3 & 3 & 3 \\ 3 & 4 & 3 & 6 \end{bmatrix}$	$\begin{bmatrix} 64 & 50 & 40 & 50 \\ 50 & 100 & 45 & 64 \\ 40 & 45 & 49 & 45 \\ 50 & 64 & 45 & 100 \end{bmatrix}$	$\begin{bmatrix} 1024 & 800 & 640 & 800 \\ 800 & 1600 & 720 & 1024 \\ 640 & 720 & 784 & 720 \\ 800 & 1024 & 720 & 1600 \end{bmatrix}$
B	[310 335 235 260]			
C	[250 410 180 390]			

Table 4.2: Summary of simulated data with small, medium and large intra-class spectral variability.

Class	Mean				Variance-covariance matrix		
					Equal class variability	Increase class C variability	Change the direction of class C
A	380	490	300	320	$\begin{bmatrix} 64 & 50 & 40 & 50 \\ 50 & 100 & 45 & 64 \\ 40 & 45 & 49 & 45 \\ 50 & 64 & 45 & 100 \end{bmatrix}$	$\begin{bmatrix} 64 & 50 & 40 & 50 \\ 50 & 100 & 45 & 64 \\ 40 & 45 & 49 & 45 \\ 50 & 64 & 45 & 100 \end{bmatrix}$	$\begin{bmatrix} 64 & 50 & 40 & 50 \\ 50 & 100 & 45 & 64 \\ 40 & 45 & 49 & 45 \\ 50 & 64 & 45 & 100 \end{bmatrix}$
B	310	335	235	260			
C	250	410	180	390	$\begin{bmatrix} 64 & 50 & 40 & 50 \\ 50 & 100 & 45 & 64 \\ 40 & 45 & 49 & 45 \\ 50 & 64 & 45 & 100 \end{bmatrix}$	$\begin{bmatrix} 256 & 200 & 160 & 200 \\ 200 & 400 & 180 & 256 \\ 160 & 180 & 196 & 180 \\ 200 & 256 & 180 & 400 \end{bmatrix}$	$\begin{bmatrix} 64 & -50 & 40 & -50 \\ -50 & 100 & -45 & 64 \\ 40 & -45 & 49 & -45 \\ -50 & 64 & -45 & 100 \end{bmatrix}$

Table 4.3: Summary of simulated data with differing degree of variation and co-variation for class C.

For each simulated remote sensing data set, 3000 pure pixels with 1000 pixels for each class were generated using their relevant descriptive statistics in Tables 4.1 – 4.3. These data were used to train the classifier. A further sample of 400 pixels with known and variable class proportions were simulated for use as the testing data. These simulated testing data were obtained using a simulated land cover map and a corresponding mean and standard deviation of each land cover class as shown in Tables 4.1 – 4.3. This was implemented as follows.

i) A land cover map was simulated in an area of 200×200 pixels (Figure 4.1) that consists of three land cover classes.

ii) From the simulated land cover map and the corresponding statistical information of each land cover class in Tables 4.1 – 4.3, the remote sensing images were simulated. The simulated images were then degraded by a factor of 10 to produce coarser spatial resolution images with 20×20 image pixels for each (400 pixels in total). The class proportions of pixels in the degraded image were identified from the corresponding area of the land cover map. Figure 4.2 shows the original image and its degraded image with the statistical information of each land cover class in Table 4.1 as an example.

The LMM (section 4.2) was used as the convenient tool to estimate the class composition for the 400 testing pixels. For illustrative purposes, the data were subjected to a principal component analysis and the first two principal components that explained most of the variation in the data set are used to display the classes in the feature space.

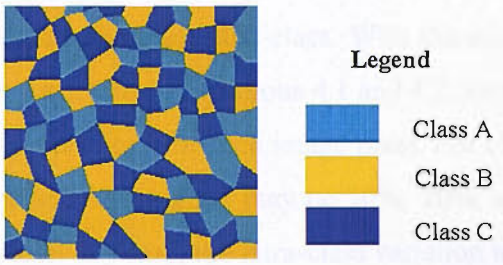


Figure 4.1: Simulated land cover map in an area of 200×200 pixels.

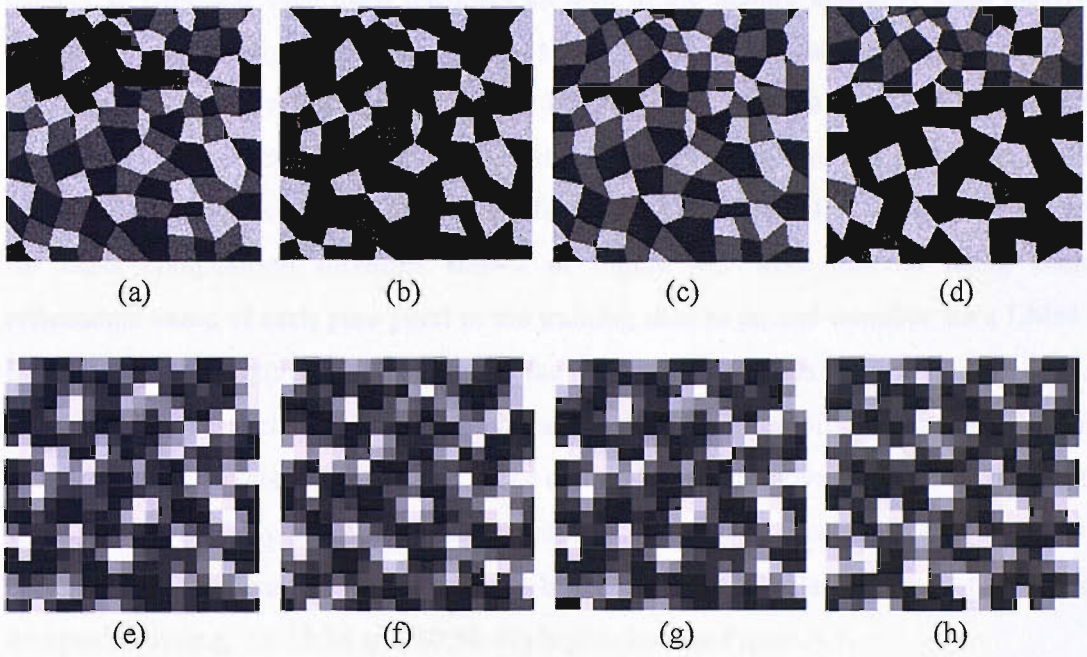


Figure 4.2: Four bands (a) band 1, (b) band 2, (c), band 3, and (d) band 4 of the original simulated image; and other four bands (e) band 1, (f) band 2, (g) band 3, and (h) band 4 of the degraded image. The classes were defined in Table 4.1.

4.3.2 Possible outputs of soft classification

One possible approach to determine end-members, *M,* is through the use of training data which consists of a set of pure pixels representing end-members, called image end-members. These training data can be extracted from the image. The reflectance values of the end-members are commonly the average spectral response of pure pixels of each class in the training data (Foody and Cox, 1994). This means that a single spectral response is taken to represent the class. With the constant values of the end-members, according to the LMM in Equations 4.1 and 4.2, there will be only one set of prediction of the class proportions for each image pixel. For example, the prediction of a soft classification for one image pixel may be 30%, 20%, and 50% respectively for three land cover classes. In this case, the intra-class variation is, therefore, ignored.

Figure 4.3 shows the location of classes defined by the parameters in table 4.1 in the feature space in the first two principal components derived from four spectral wave-bands. As expected each class occupies an area in the feature space. A class clearly cannot be adequately represented by a single end-member, an assumption in the traditional unmixing researches (Settle and Drake, 1993, Theseira *et. al.*, 2003, Liu and Wu, 2005). As a result of the intra-class variation, pixels with a particular class composition could occupy an area in the feature space (Figure 4.3). The distributions for class composition mixtures shown in Figure 4.3 were derived using each reflectance value of each pure pixel in the training data as an end-member for a LMM. For example the mixture 80:12:08 means the proportions of class A, B, and C are 80%, 12%, and 8%, respectively. Due to the intra-class spectral variability, there were many possible image pixels that have this class composition. Moreover, from Figure 4.3 it was apparent that any one point in the feature space could be associated with a variety of class compositions. It was illustrated by the area of overlap between two class compositions (e.g, 33:33:34 and 30:30:40) highlighted in Figure 4.3.

Assuming the spectral response of each pure pixel in the training data as an end-member reflectance value for the class representing that pixel, there would be a number of different sets of end-member values. By unmixing the spectral response of a pixel many times with different end-members a series of sub-pixel class proportion predictions could be derived for a pixel of any given spectral response. As a result of

this it was possible to form a distribution of class proportion estimates for each pixel. The nature of the distribution will depend on the position of the pixel in the feature space and the degree of intra-class variation and class co-variation. Figure 4.4 shows an example about the distributions of possible composition estimates of each land cover class within an image pixel (e.g. pixel *x* in shown in Figure 4.3).

Using LMM, the distributions of possible mixing predictions may not follow a normal distribution. The interquartile range of each distribution was, therefore, used to express the spread of its distribution.

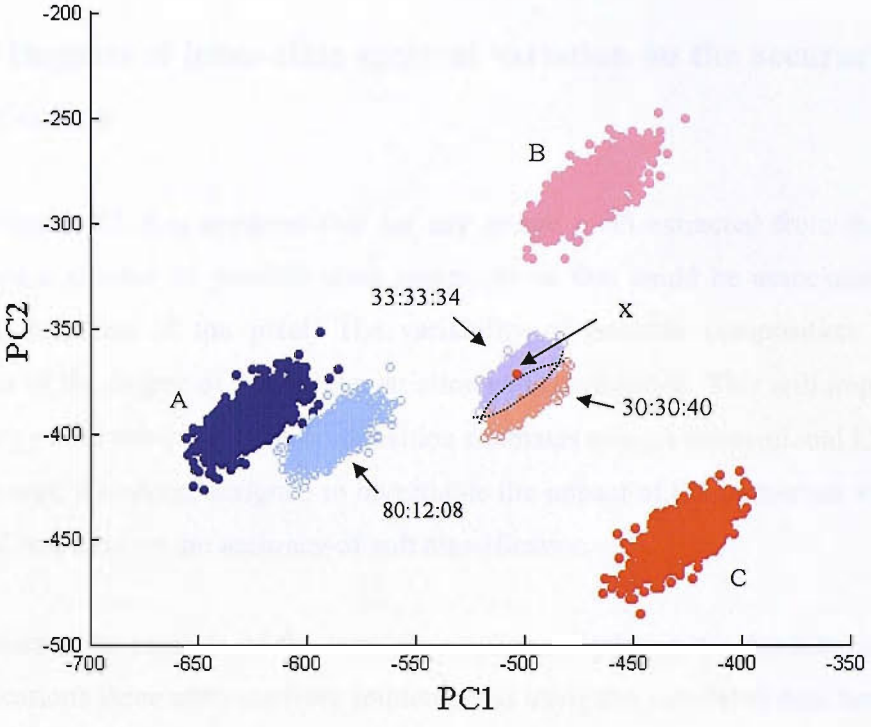
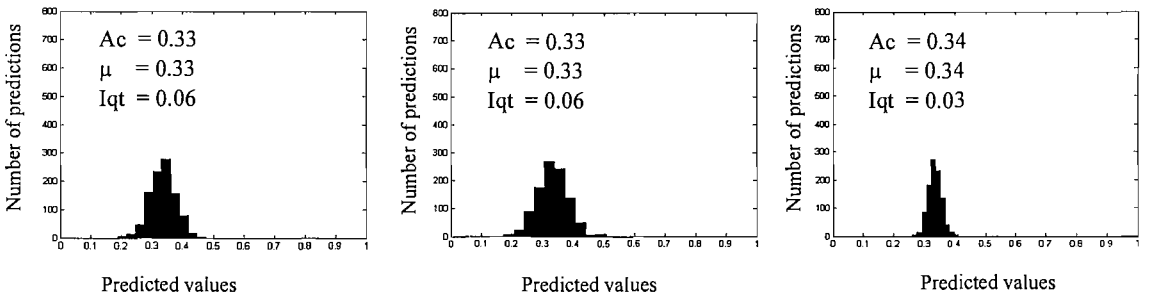


Figure 4.3 Location of the classes defined by Table 4.1 and three mixed compositions in the feature space defined by the first two principal components. The mixed compositions are defined in terms of the percentage cover of class A:B:C. Note the large area of overlap between the 33:33:34 and 30:30:40 mixtures.



(a) (b) (c)

Figure 4.4: Distributions of possible class composition predictions (a) class A's predictions, (b) class B's predictions, and (c) class C's predictions. (AC – actual value; μ: the mean value of the distribution; Iqt – Interquartile range of the distribution).

4.3.3 Impacts of intra-class spectral variation on the accuracy of soft classification

From Figure 4.3 it is apparent that for any image pixel extracted from the imagery there are a number of possible class compositions that could be associated with the spectral response of the pixel. The variability of possible composition will be a function of the degree of intra-class variation and co-variation. This will impact on the accuracy of the sub-pixel class composition estimates using a conventional LMM. This section was, therefore, designed to investigate the impact of the intra-class variation in spectral response on the accuracy of soft classification.

To illustrate the impacts of the intra-class spectral variation on the accuracy of soft classifications three analyses were implemented using the simulated data based on the statistical information summarised in Table 4.2. The spectral response of the classes was varied in the analyses in terms of both the intra-class variation and co-variation in the spectral feature space. There were three simulated data sets which were correspondingly simulated from the small, medium, and large class variability (section 4.3.1).

For each simulated remote sensing data set, 3000 pure pixels with 1000 pixels for each class were generated using their relevant descriptive statistics in Table 4.2. These data were used to train the classifiers. The location of the classes in the training data of

three simulated datasets, which have the class variability from small to large, are illustrated in Figure 4.5a using the first two principal components. The sample of 400 pixels with known and variable class proportions in each degraded simulated data were used as their corresponding testing data (section 4.3.1).

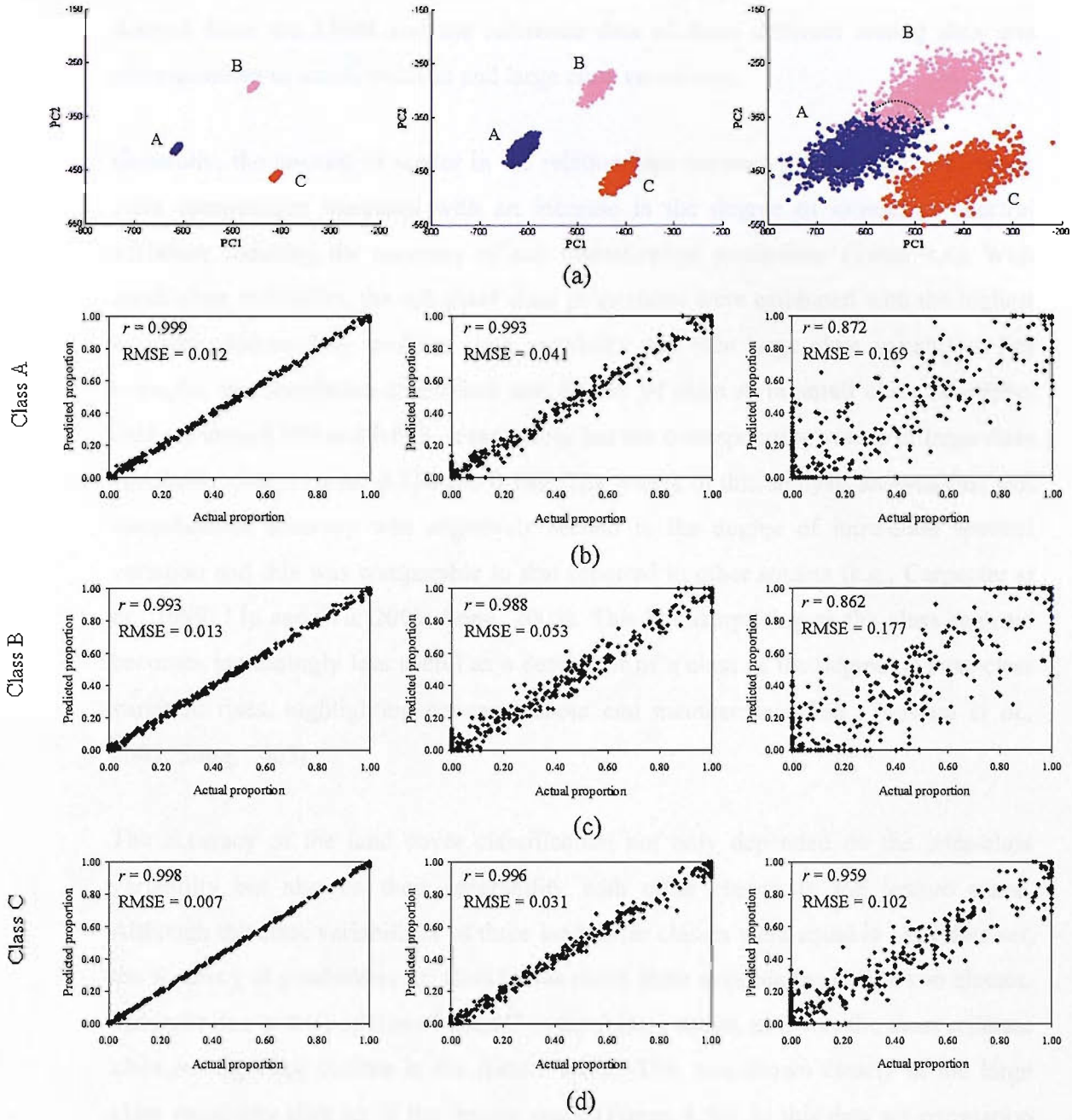


Figure 4.5: The impact of class variability on classification accuracy (a) the class variability; (b), (c), (d): relationship between the proportional coverage of a class derived from the LMM and the actual data. The classes were defined in table 4.2.

Using the class centroids defined in Table 4.2 as the class end-member spectra for the input in the LMM, the class composition for 400 testing pixels was estimated. Accuracy of the sub-pixel class composition predictions was evaluated using the correlation coefficient between the predicted and reference coverage and RMSE. Figures 4.5(b-d) show the relationship between the proportional coverage of a class derived from the LMM and the reference data of three different testing data sets corresponding to small, medium and large class variability.

Generally, the amount of scatter in the relationships between predicted and reference class composition increased with an increase in the degree of intra-class spectral variation, reducing the accuracy of soft classification predictions (Table 4.4). With small class variability, the sub-pixel class proportions were estimated with the highest accuracy, followed by medium class variability and then large class variability. For example, the correlation coefficient and RMSE of class A in small class variability data set were 0.999 and 0.012, respectively but the corresponding values in large class variability data set were 0.872 and 0.169. The results of this analysis showed that soft classification accuracy was negatively related to the degree of intra-class spectral variation and this was comparable to that reported in other studies (e.g., Carpenter *et al.*, 1999; Liu and Wu, 2005; Song, 2005). This is unsurprising as the class centroid becomes increasingly less useful as a descriptor of a class as the degree of intra-class variation rises, highlighting concerns about end member selection (Theseira *et al.*, 2003; Song, 2005).

The accuracy of the land cover classification not only depended on the intra-class variability but also on their separability with other classes in the feature space. Although the class variabilities of three land cover classes were equal in each data set, the accuracy of predictions for class C was much more accurate than other two classes. This was due to the position of class C in the feature space, as it was the most separate class among three classes in the feature space. This was shown clearly in the large class variability data set in the feature space (Figure 4.5a). In this data set correlation coefficient of class C was 0.959 while that of class A and class B were 0.872 and 0.862, respectively.

(a)

Class	*r*	RMSE
A	0.999	0.012
B	0.993	0.013
C	0.998	0.007

(b)

Class	*r*	RMSE
A	0.993	0.041
B	0.988	0.053
C	0.996	0.031

(c)

Class	*r*	RMSE
A	0.872	0.169
B	0.862	0.177
C	0.959	0.102

Table 4.4: Accuracy of predictions in Figure 4.5 (a) small, (b) medium and (c) large class variability.

4.3.4 Impacts of intra-class spectral variation on soft classification prediction

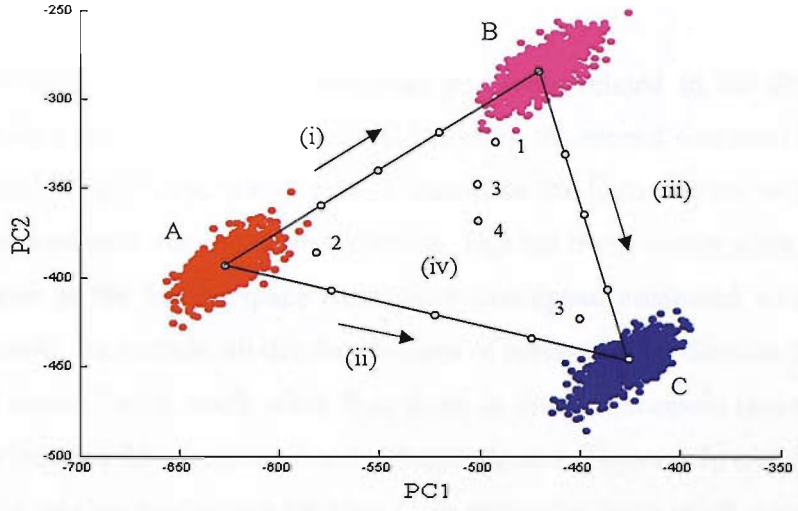
From Figure 4.5, it is shown that the classification accuracy was affected by the intra-class variation in spectral response in the feature space. The larger the intra-class spectral variability was, the less accurate the classification. This indicated that the class centroid becomes increasingly less useful as a representation of a class as the degree of intra-class variation rises, showing concerns about end-member selection (e.g. Song, 2005 and Bateson *et al.*, 2000). Mixing the distribution of class spectral response of the individual training pixels rather than the centroids of classes may be a more realistic basis for soft classification. This allows the prediction of a distribution of possible class compositions for each image pixel (Figure 4.4). The nature of the distribution of possible class composition estimates will depend on the location of the pixel in the feature space and the degree of intra-class variation and class co-variation. The objective of this section was, therefore, to study the impact of intra-class variation in spectral response on the shape of the distributions which are formed by the possible output of each class proportion for each image pixel.

Three simulated data sets were generated for three scenarios based on the class descriptions summarised in Table 4.3. In the first scenario, the classes had equal intra-class spectral variability. The second scenario was designed with a four-fold increase in the intra-class variability of class C and finally the third scenario was repeated but with the distribution of class C rotated 90^{o} in feature space. The locations of the classes in the training data of three scenarios are illustrated in Figures 4.6(a-c).

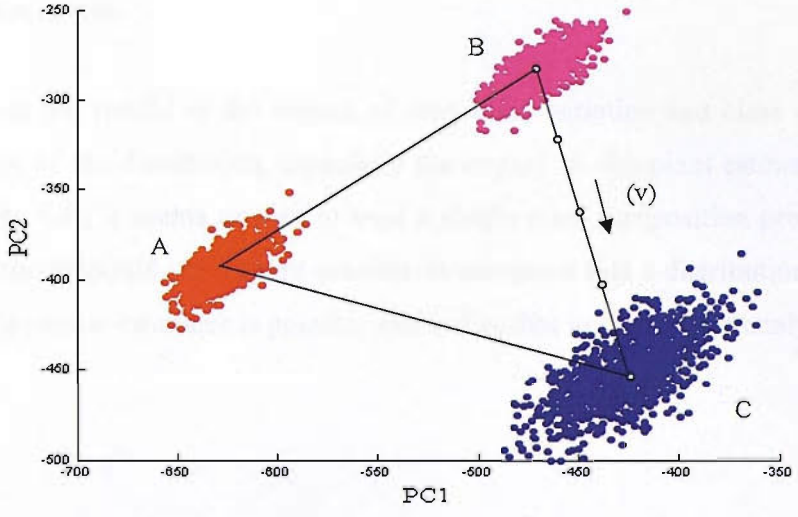
Using the spectral response of each pixel in the training data as an end-member value for the class representing in that pixel, there would be a number of different values for each end-member. By unmixing the spectral response of a pixel many times with different end-members a number of sub-pixel class composition estimates could be derived for a pixel of any given spectral response. As a result, it was possible to derive a distribution of sub-pixel class composition estimates for each pixel. The class proportions of 27 pixels, which were located in 6 different transects (e.g., from transect i to transect vi) in Figures 4.6(a-c), were predicted in this manner. The distributions of possible class composition predictions within each image pixel in the 6 transects are correspondingly illustrated in 6 figures (Figures 4.7(a-f)).

In general, the nature of the distribution of possible mixing predictions for an image pixel will depend on the location of the point in feature space and the degree of intra-class variation and class co-variation. The more overlap between the scatter plots of land cover classes in the feature space, the larger the distributions of the mixing predictions.

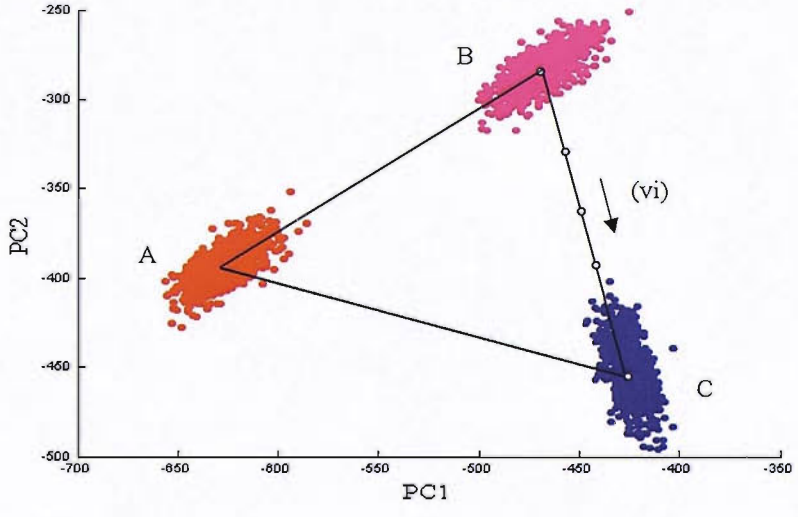
Although the intra-class spectral variability of three classes equalled in the first scenario (Figure 4.6a), the distributions of mixing predictions for class A and class B were much wider that that for class C (compare especially the distributions for transects i to iv in Figures 4.7(a-d)). This was due to the position of class C in the feature space as it was the most separate class among three classes in the feature space. Specifically, in the third scenario, the direction of class C was rotated 90^{o} in the feature space (Figure 4.6c) and this made the separability between B and class C reduced. Consequently, the distributions of possible class composition predictions for class B and class C were more spread than those in the first scenario (compare especially the distributions for transects iii and v highlighted in Figure 4.7c and e).



(a)



(b)



(c)

Figure 4.6: Location of the classes in three scenarios defined in Table 4.3 with the position of 27 testing points (°) in 6 transects (i, ii, iii, iv, v, and vi) highlighted.

The spread of the distributions was positively related to the degree of intra-class variation and co-variation. This was shown in the second scenario (Figure 4.6b). In the second scenario, the scatter plot of class C in the feature space was increased 4 times compared with that in the first scenario. This led to the scatter plots of three land cover classes in the feature space were more overlapped compared with those in the first scenario. As a result, all the distributions of proportion predictions for class A, class B, and class C were much wider than those in the first scenario (compare especially the distributions for transects iii and vi highlighted in Figure 4.7c and f). The distributions of the mixing predictions for class C, in particular, were much wider than those in the first scenario.

Given the results of the impact of intra-class variation and class co-variation on the shape of the distribution, especially the impact on sub-pixel estimation accuracy (e.g. Table 4.4), it seems unwise to trust a single class composition prediction for a pixel. Instead it would seem more sensible to recognise that a distribution of sub-pixel class composition estimates is possible and utilise that in subsequent analyses.

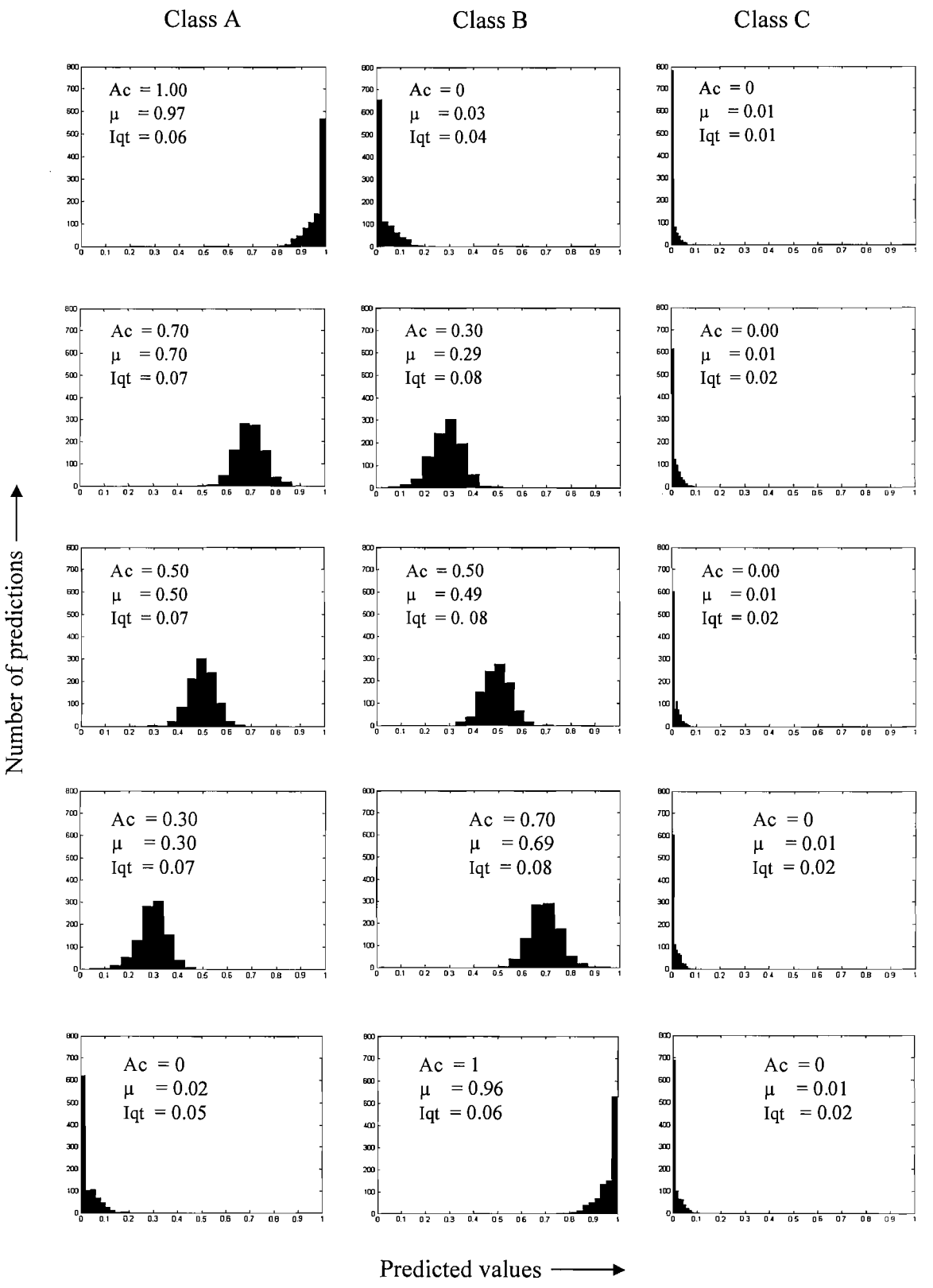


Figure 4.7a: Histograms show distribution of possible composition estimates for a class located in transect (i) defined in Table 4.6.

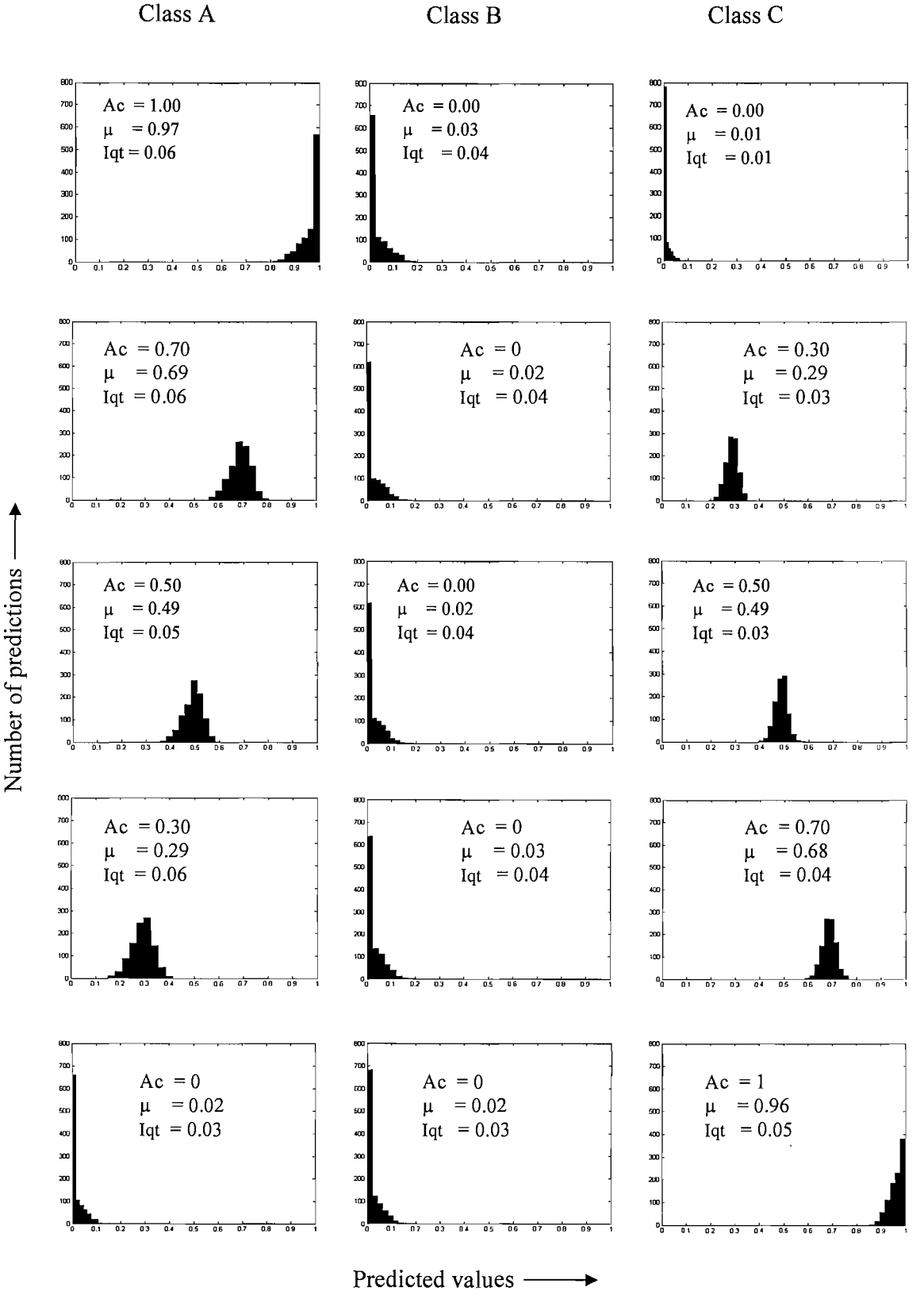


Figure 4.7b: Histograms show distribution of possible composition estimates for a class located in transect (ii) defined in Figure 4.6.

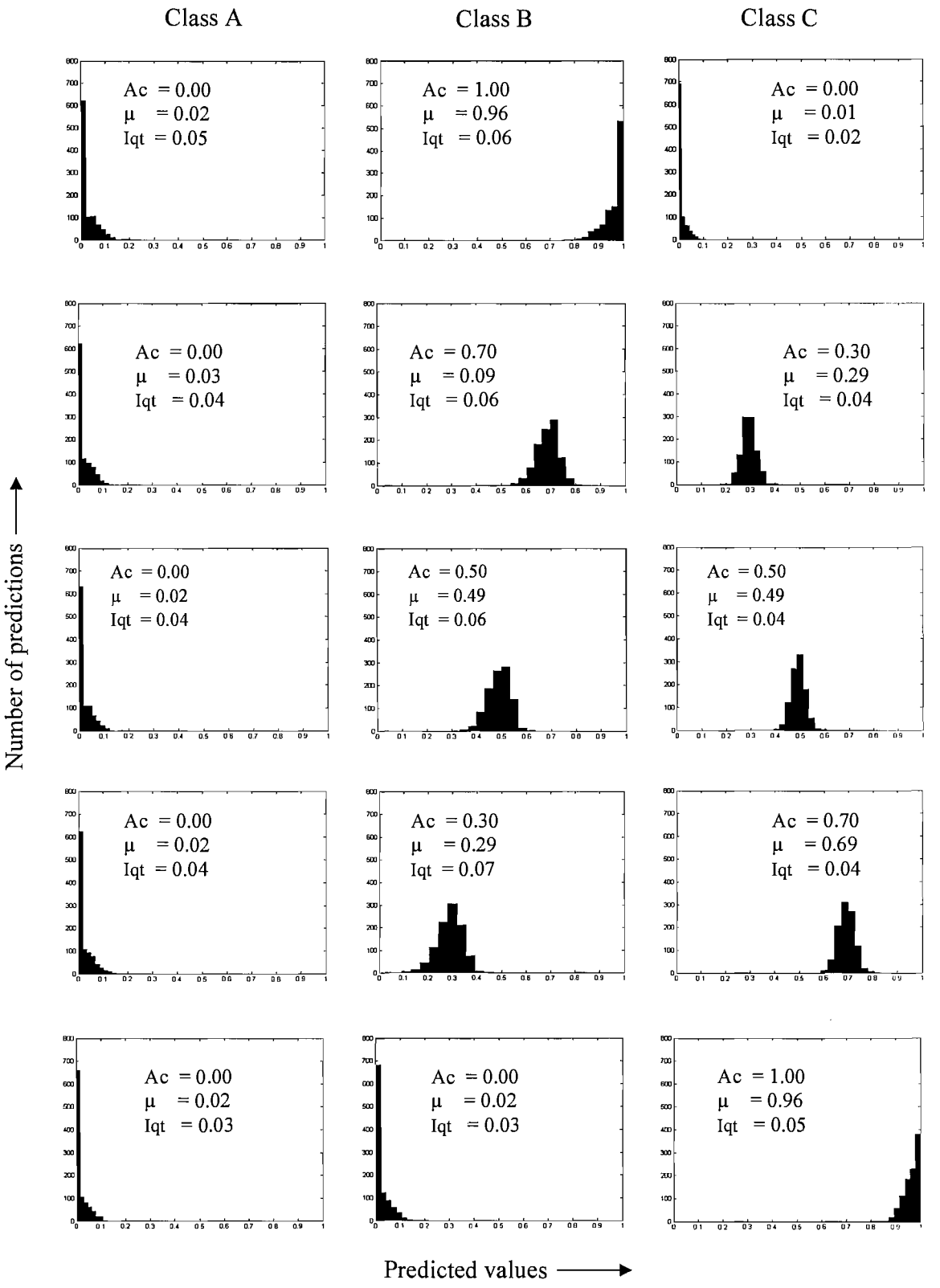


Figure 4.7c: Histograms show distribution of possible composition estimates for a class located in transect (iii) defined in Figure 4.6.

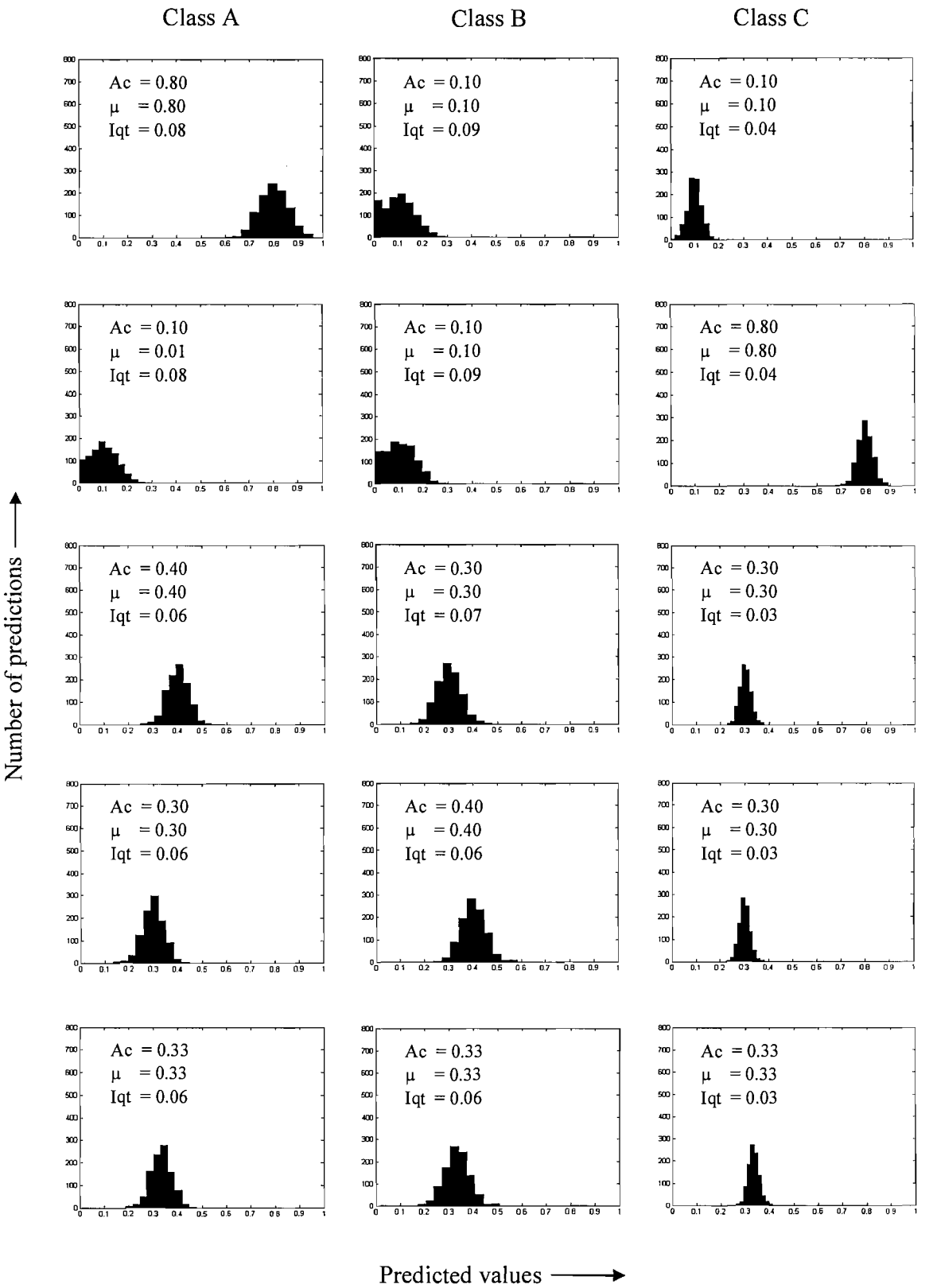


Figure 4.7d: Histograms show distribution of possible composition estimates for a class located in transect (iv) defined in Figure 4.6.

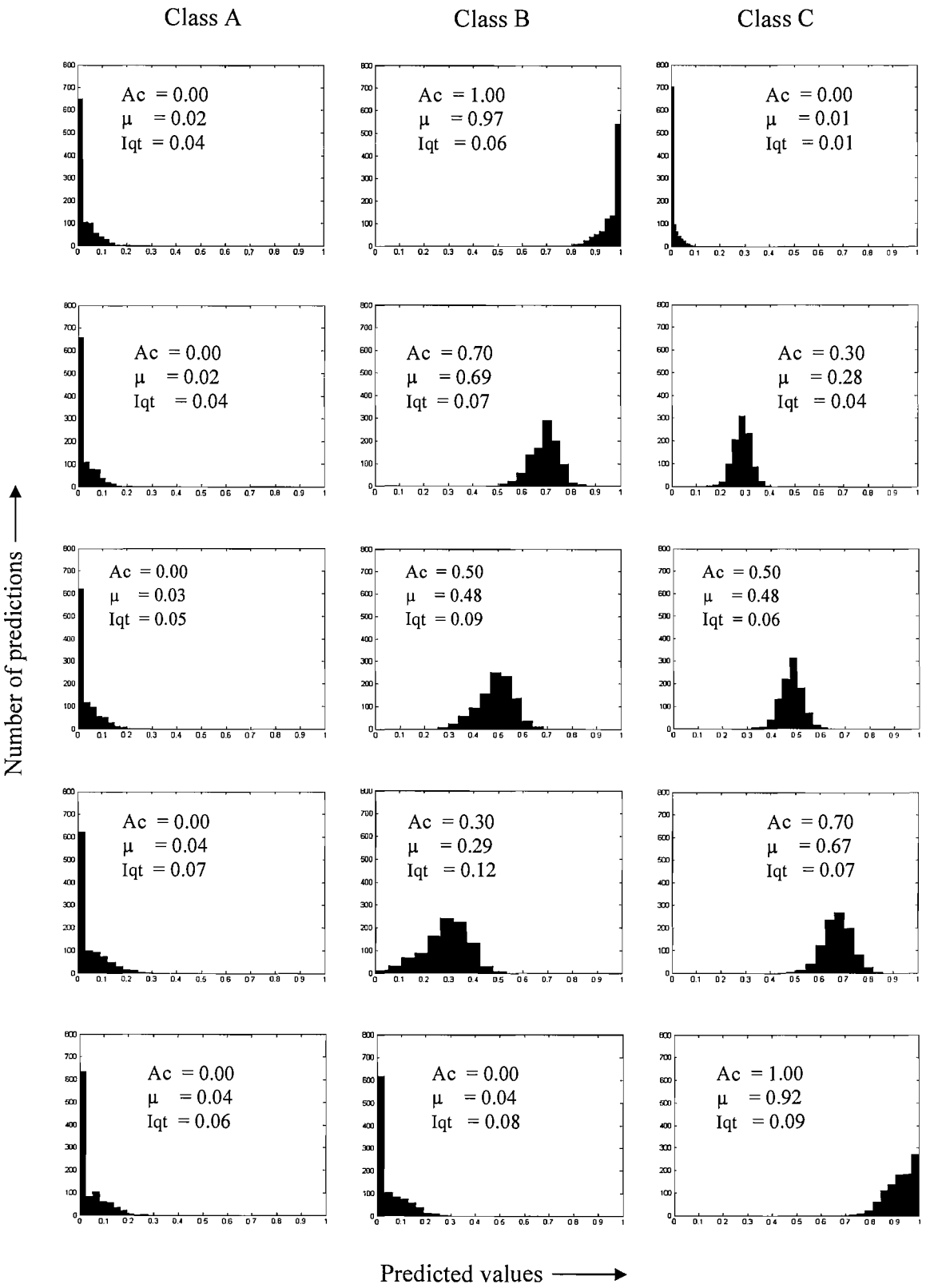


Figure 4.7e: Histograms show distribution of possible composition estimates for a class located in transect (v) defined in Figure 4.6.

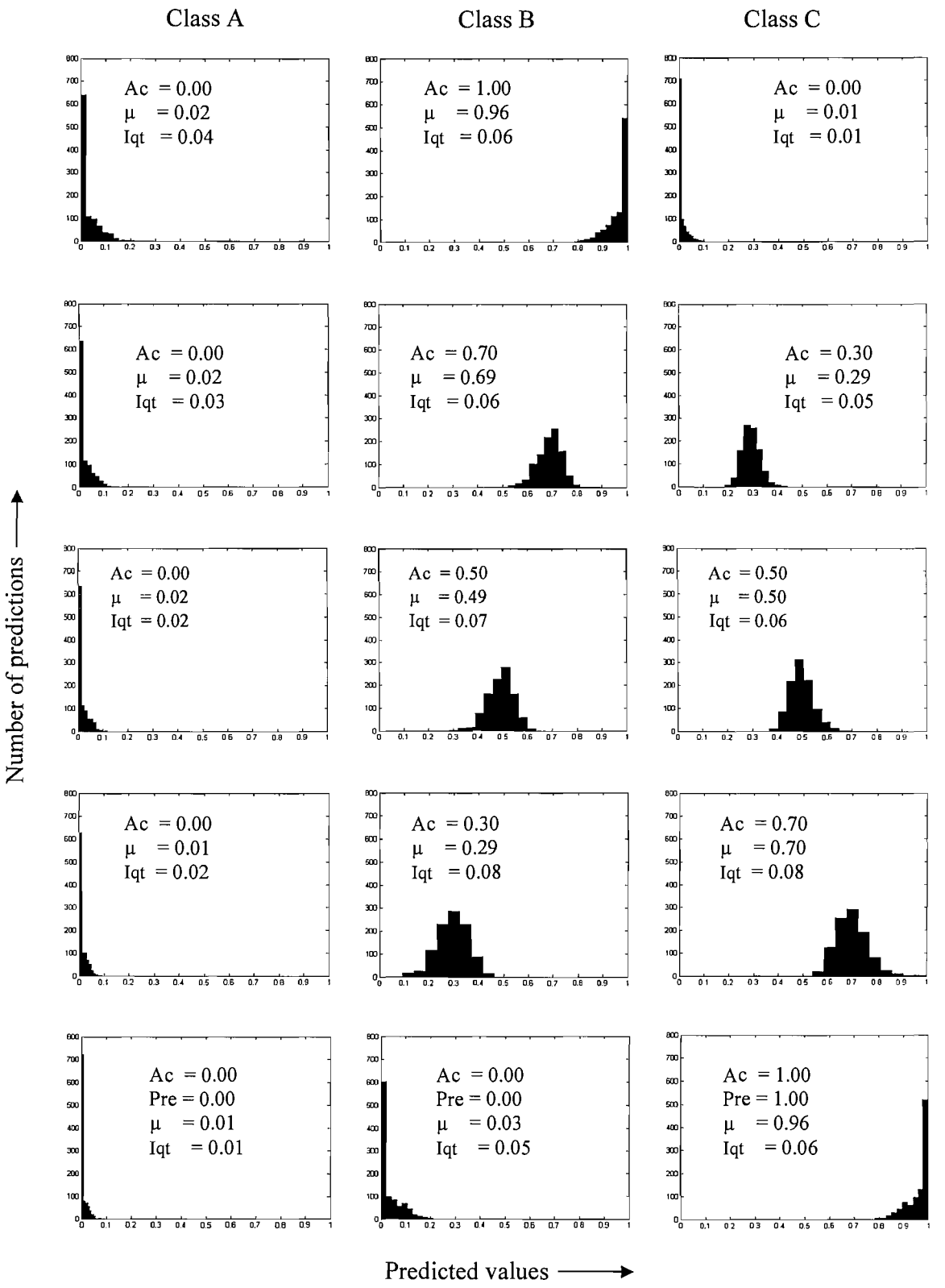


Figure 4.7f: Histograms show distribution of possible composition estimates for a class located in transect (vi) defined in Figure 4.6.

4.3.5 Output of the distributions of possible class composition estimates

Results of the analyses in section 4.3.3 and 4.3.4 showed that soft classification accuracy was negatively related to the degree of intra-class variation. It was suggested that a distribution of possible class compositions be derived from pixels instead of a single class composition prediction. The nature of the distribution depended on the position of the pixel in the feature space and the degree of intra-class variation and class co-variation. The larger the intra-class variation was, the more spread the distribution of possible class composition estimates and the less accurate the soft classification. In soft classification, proportion images are used as a means to display the output of soft classification. To illustrate the results derived from the distributions, some outputs may be derived.

For representational purposes an area of 8x8 image pixels (64 pixels in total) was simulated using the information summarised in Tables 4.1 - 4.3 (Figure 4.8a).

4.3.5.1 Variation images

Variation images were a set of fraction images in which the value of each pixel was the interquartile range of the corresponding distribution of possible class composition predictions for that pixel. The numbers of variation images were equal to the numbers of target classes defined to map the land cover map. Figure 4.8c shows three variation images for three land cover classes, A, B, and C. The data used to derive these variation images were 64 image pixels which were generated using the data summarised in Table 4.1, data with the same class variability for three land cover classes.

The information from the variation images may show the degree of the 'uncertainty' of the unmixing predictions in using the class centroids to define the class end-member spectra. The larger value of a pixel in the variation images (the more spread the distribution of possible class composition predictions) was, the more 'uncertainty' the prediction of class proportions for that pixel.

4.3.5.2 Closeness images

Closeness images were a set of fraction images where each shows the measures of the distance between the actual class proportion and the predicted class proportion based on the distributions of possible class composition estimates for a particular land cover in each pixel. The closeness of a class proportion in a pixel was defined as illustrations in Figure 4.9(a-c). The number of closeness images was equal to the numbers of target classes defined to map the land cover map. Figure 4.8d shows three closeness images for three land cover classes, A, B, and C.

The information from the closeness images could be used to evaluate the accuracy of the predicted data. The smaller the closeness was the less the difference between actual and predicted data and the more accurate the classification. In Figure 4.9b, the predicted class proportion was equal to actual class proportion and the closeness was 0.0, the most accurate predictions. Conversely, the actual class proportion was outside the distribution of mixing predictions in Figure 4.9c and the closeness was 1.0, the least accurate predictions.

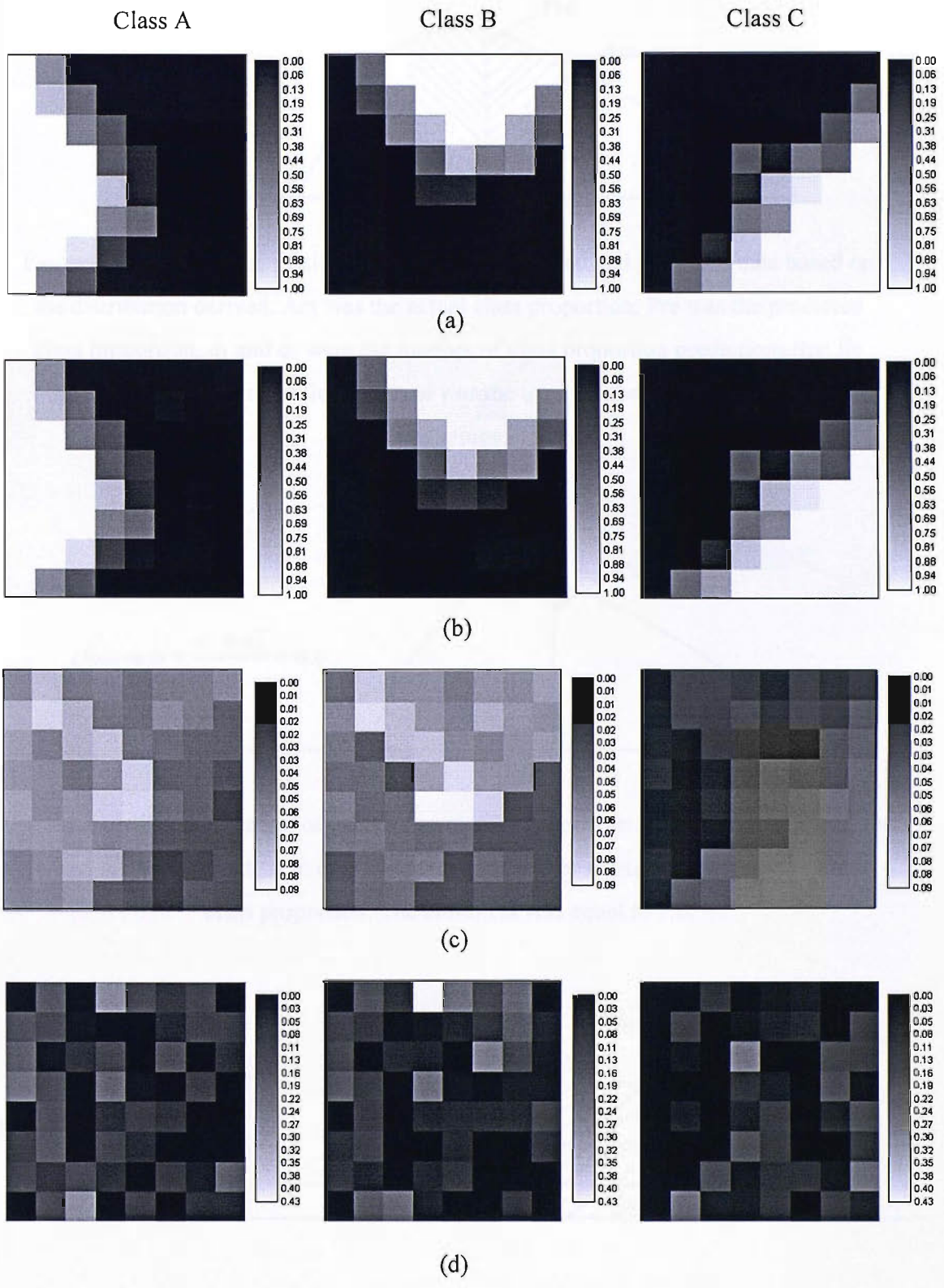


Figure 4.8: Results derived from soft classification using the data set with the classes defined in Table 4.1. (a): actual proportion images; (b): predicted proportion images; (c) variation images; (d): closeness images.

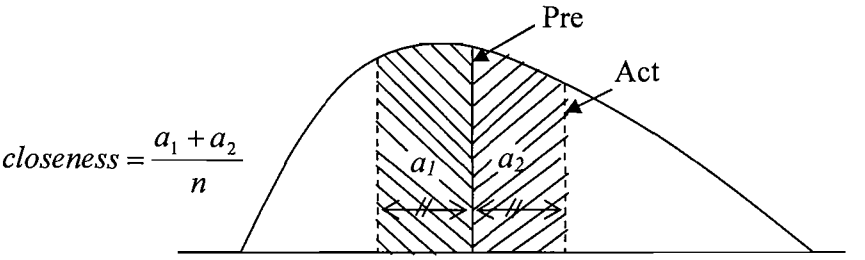


Figure 4.9a: Calculation of closeness between the actual and predicted data based on the distribution derived. Act was the actual class proportion, Pre was the predicted class proportion, a_1 and a_2 were the number of class proportion predictions that lie inside the actual and predicted data, n was the total number of class proportion predictions.

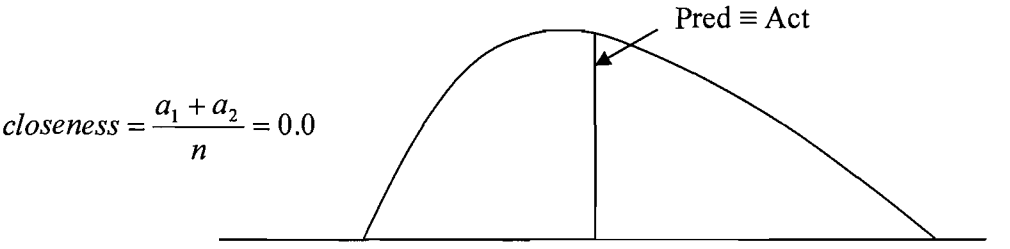


Figure 4.9b: The illustration of calculating closeness between actual and predicted data based on the distribution in case the predicted class proportion was equal to actual class proportion. The closeness was equal to 0.0.

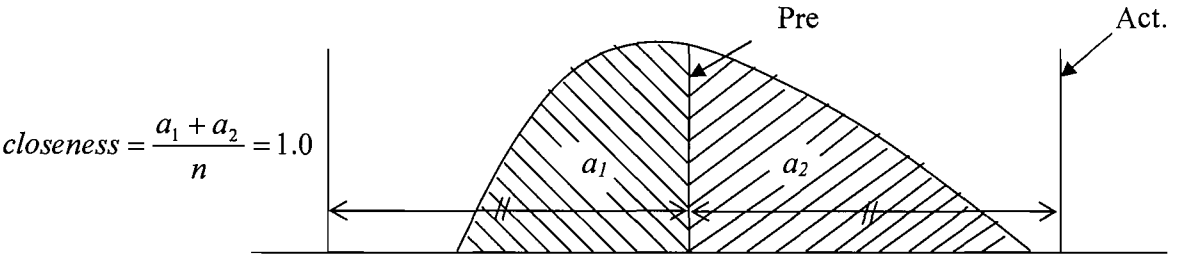


Figure 4.9c: Calculation of closeness between actual and predicted data based on the distribution derived in case the predicted class proportion was outside of the distribution. The closeness was equal to 1.0.

4.4 Impacts of intra-class spectral variability on change detection

Soft classification accuracy was negatively related to the degree of intra-class spectral variability. This was illustrated by some analyses using the simulated data sets in section 4.3. Here, some analyses of real remotely sensed data were undertaken to investigate the impacts of the intra-class spectral variation on soft classification and its implications. Attention focused on a common type of analysis in a major environmental science context; the assessment of tropical deforestation. Initially the data used for the analyses and the procedures needed to process the data are presented. The technique to extract spectral responses of end-members from mixed pixels is then proposed. Finally, the application of the distributions of possible sub-pixel class composition predictions to the assessment of deforestation is investigated.

4.4.1 Study area, data, and data processing

4.4.1.1 Study area and data used

Two Landsat TM images of the same site in part of Para, Brazil with the path of 223 and the row of 064 were used. These two images were acquired on 11 July 1984 and 22 July 1988 and extracted in the internet from the Global Land Cover Facility (GLCF) at the University of Maryland at http://glcf.umiacs.umd.edu. The images were acquired in 7 spectral wavebands with a spatial resolution of 30 m, except spectral band 6 with spatial resolution of 120 m. Due to the striping effects on three spectral bands: blue (TM band 1), green (TM band 2), and red (TM band 3) in both of the images, and the spatial resolution in band 6 of 120 m, only three infrared bands (TM band 4, band 5 and band 7) were used for the analyses.

4.4.1.2 Data processing

In terms of the estimation of deforestation from sub-pixel classifications in two time periods of the same region, two Landsat TM images were co-registered using the image-to-image registration approach. Nearest neighbour resampling was used with an

estimated RMSE of ± 0.5 pixel. After registration, an area of 400 × 400 pixels was selected from the two images for this study because of its rapid deforestation (see Figure 4.10).

The test-site was dominated by forest and the region cleared of forest cover was evident. For the purposes of this study, each pixel in the two original TM images was assumed to be pure and classified visually into two land cover classes, forest and non-forest, using a supervised classifier, the MLC. In the absence of accurate ground survey data, these classification data were considered as the reference data for the analyses.

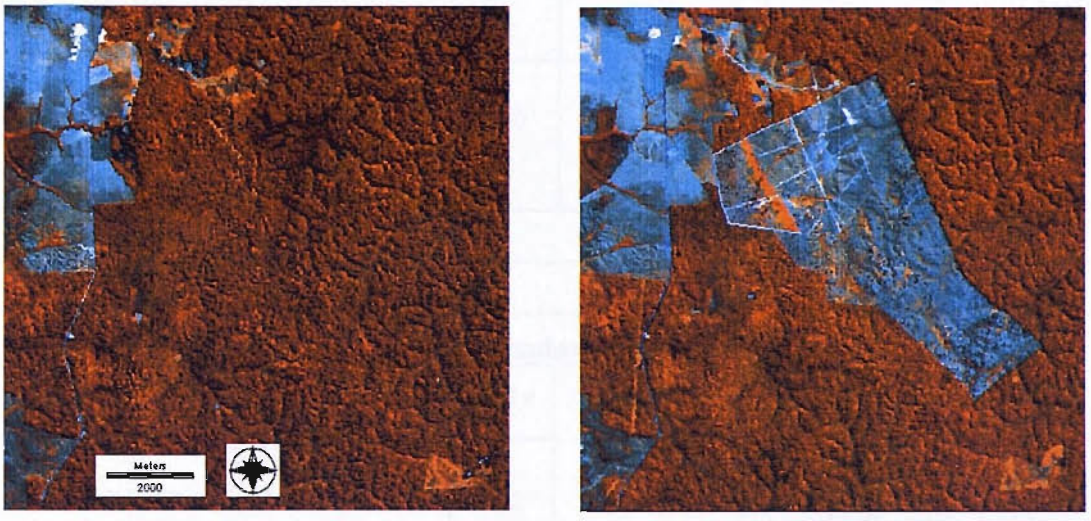
The two original Landsat TM images were spatially degraded by a factor of 10 to simulate data sets with a relatively coarse spatial resolution of 300 m. This coarse spatial resolution is comparable to the spatial resolution of medium spatial resolution system such as MODIS and MERIS data used to derive deforestation. The degraded imagery was obtained by aggregating pixels to the desired spatial resolution, with each degraded DN expressed as the mean DN of the original un-degraded pixels it comprised. The degraded data sets represented simulated coarse spatial resolution imagery and they were used to derive sub-pixel estimations and deforestation for the analyses (Figure 4.11).

The class composition of pixels in the coarse spatial resolution images were defined from the corresponding areas of the land cover maps derived from the classification of two original Landsat TM images. This was used as the reference data to derive the sub-pixel classification of the two simulated remotely sensed images to estimate deforestation.

4.4.1.3 Training and testing data

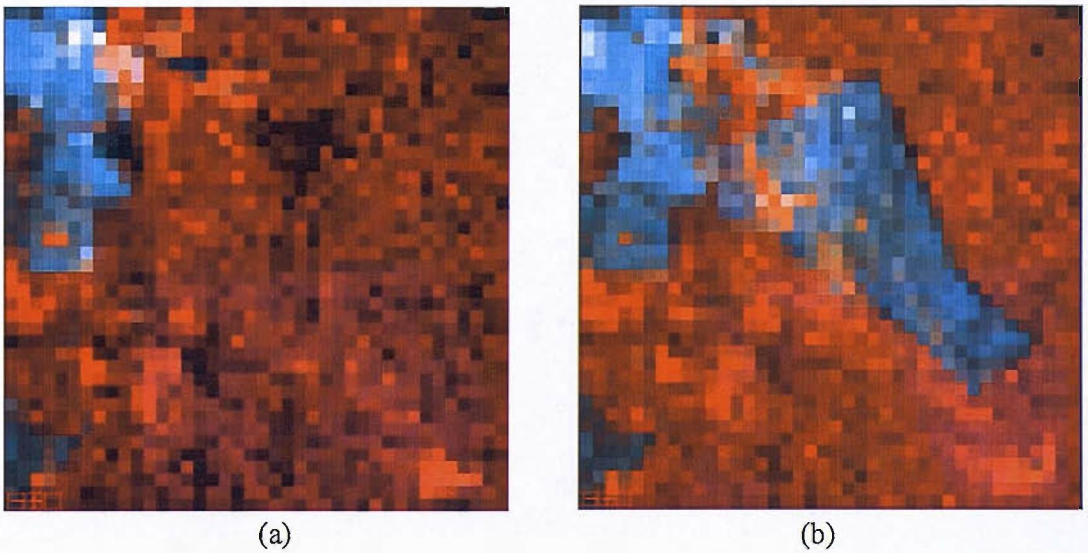
The training data consisted of 180 randomly selected pure pixels with 90 pixels for each land cover class (e.g., forest and non-forest). The statistical information of the two land cover classes is shown in Table 4.5. For illustrative purposes, the data were subjected to principal components analyses from three used wave-bands and the first two components (PC1 and PC2) that explained most of the variation in the data sets are

used to display classes in the feature space. The scatter plots of the training samples were displayed in this manner in Figure 4.12. The testing set contained 500 randomly selected pixels.



(a) (b)

Figure 4.10: Three-band composite Landsat TM images using bands 4, 5, and 7 mapped to red, green, and blue respectively of the study area in part of Para, Brazil, (a) July 1984; (b) July 1988.



(a) (b)

Figure 4.11: Three-band composite spatially degraded images using bands 4, 5, and 7 mapped to red, green, and blue respectively of the study area in part of Para, Brazil, (a) July 1984; (b) July 1988.

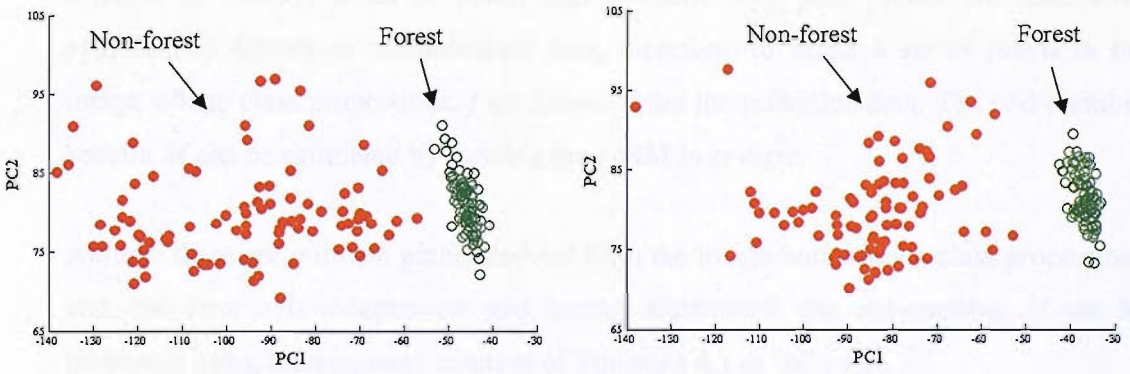
(a)

Class	Mean			Standard deviation			Variance-covariance matrix
	Band 4	Band 5	Band 7	Band 4	Band 5	Band 7	
Forest	76.83	51.25	12.72	3.39	2.43	0.60	$\begin{bmatrix} 11.50 & 6.03 & 1.25 \\ 6.03 & 5.88 & 1.41 \\ 1.25 & 1.41 & 0.36 \end{bmatrix}$
Non-forest	71.88	96.64	34.55	6.10	17.92	10.03	$\begin{bmatrix} 37.20 & -24.17 & -11.82 \\ -24.17 & 320.98 & 177.87 \\ -11.82 & 177.88 & 100.55 \end{bmatrix}$

(b)

Class	Mean			Standard deviation			Variance-covariance matrix
	Band 4	Band 5	Band 7	Band 4	Band 5	Band 7	
Forest	72.23	51.79	9.70	2.92	2.37	0.54	$\begin{bmatrix} 8.53 & 4.07 & 0.61 \\ 4.07 & 5.63 & 1.19 \\ 0.61 & 1.19 & 0.29 \end{bmatrix}$
Non-forest	62.00	94.04	29.14	6.506	12.62	6.49	$\begin{bmatrix} 42.33 & -32.49 & -17.51 \\ -32.49 & 159.13 & 80.58 \\ -17.51 & 80.58 & 42.13 \end{bmatrix}$

Table 4.5: Class description for the imagery, (a) July 1984 and (b) July 1988.



(a) (b)

Figure 4.12: Location of the classes in feature space, (a) July 1984 and (b) July 1988.

4.4.2 Incorporating mixed pixels in the end-member selection for LMM

Conventionally in LMM, the image end-members are identified from the training pixels assumed to be pure. Therefore, in an attempt to derive sub-pixel class composition within image pixels, mixed pixels are intentionally avoided in training a classifier. In many remote sensing images, especially for coarse spatial resolution imagery, a majority of image pixels are mixed. It may however be difficult to identify a training set of an appropriate size that contains only pure pixels. Therefore, mixed pixels need to be incorporated in the training set to define the spectral response of the end-members. The accommodation of mixed pixels in three stages (e.g. training, allocation, and testing) of fuzzy supervised classification has been investigated by Foody and Arora (1996), Foody (1997), and Zhang and Foody (2001). In these studies, mixed pixels were incorporated in the fuzzy classification of land cover from remotely sensed imagery with the MLC and ANN. In the research presented here an approach of identifying end-member spectra for the input of the conventional LMM from mixed pixels was explored.

4.4.2.1 Methods

One approach to derive end-members, M, is through the average spectral response of pure pixels in the image. In imagery dominated by mixed pixels it may, however, be difficult to identify a set of pixels that contains only pure pixels. An alternative approach to determine end-members was, therefore, to select a set of pixels in the image whose class proportions, f are known from the reference data. The end-member spectra M can be estimated by running the LMM in reverse.

Assume there are n mixed pixels derived from the image with known class proportions and, the error e is independent and normal distributed, the end-member M can be estimated using least-squares solution of Equation 4.1 as follows:

$$M = (X f^{\mathrm{T}})(f f^{\mathrm{T}})^{-1} \tag{4.6}$$

$$X = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ X_{1m} & X_{2m} & \cdots & X_{nm} \end{pmatrix}$$ is the $m \times n$ matrix of the observation vector for n pixels which are known as the class proportions. Each column is a vector which represents the observation vector of a pixel.

$$f = \begin{pmatrix} f_{11} & f_{21} & \cdots & f_{n1} \\ f_{12} & f_{22} & \cdots & f_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ f_{1c} & f_{23} & \cdots & f_{nc} \end{pmatrix}$$ is the $n \times c$ matrix of the ground cover proportions for each of c land cover classes. Each column is a vector which represents the class proportions of a pixel.

Once the end-member spectra are known, the class compositions of all the pixels in the image may be estimated using conventional LMM (equations 4.1 and 4.2).

4.4.2.2 Analysis

The degraded imagery in 1984 was used in this experiment as an example. Analysis was implemented to explore the approach of identifying end-member spectra from mixed pixels. Before presenting results of this approach, for the comparison purpose results of the LMM using the end-members derived from pure pixels is illustrated.

4.4.2.2.1 End-member selection from pure pixels

In this section, the spectral responses of the end-members were derived from the average spectral responses of the training pixels for each land cover class. The training sample contained 180 randomly selected pure pixels with 90 pixels for each land cover class as described in section 4.4.1.3. Class proportions within spatially degraded image pixels in the year of 1988 were estimated using LMM. The testing sets used to assess the accuracy of the sub-pixel composition estimates derived from LMM contained 500 randomly selected pixels as described in section 4.4.1.3. The accuracy of the sub-pixel class composition estimates was evaluated with a strong and significant relationship between actual and predicted class cover observed although there was a large degree of scatter (Figure 4.13). This was achieved with the correlation coefficient of 0.95 and RMSE of 0.15.

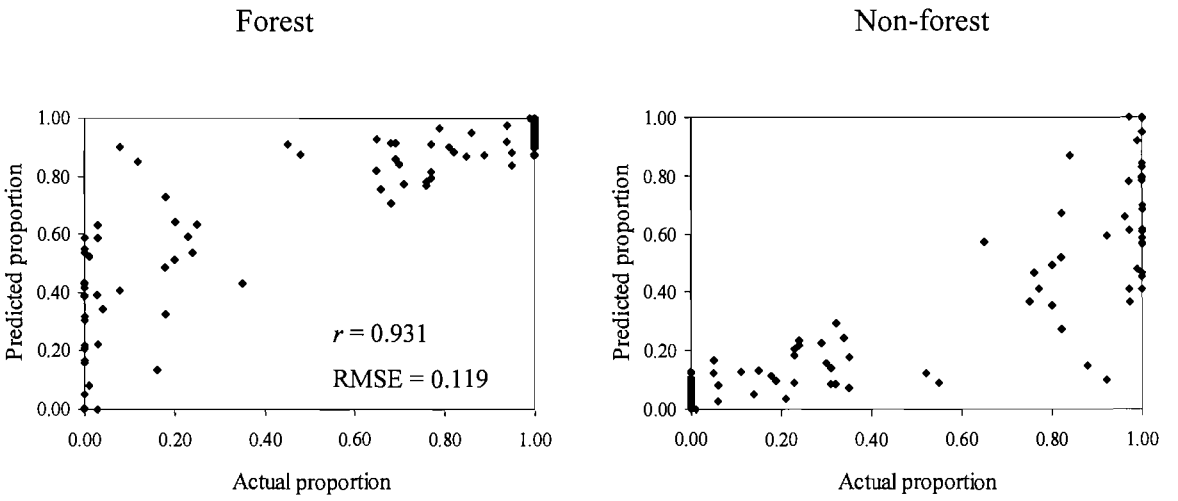


Figure 4.13: Relationships between predicted and actual cover for data in 1984.

4.4.2.2.2 End-member selection from mixed pixels

For the purpose of this analysis, all the training pixels had a mixed and known land cover composition. The training pixels were selected in such a way that each of two classes dominated the land cover in the training pixels. To assess the relative impacts of mixed pixels on the accuracy of sub-pixel estimatons using LMM, four training samples were used for the analysis with each training sample contained mixed pixels with differently dominated class proportions (Table 4.6). For example, the first training sample presented in Table 4.6 consisted of 30 mixed pixels, of which 15 pixels was dominated by forest contained between 90-99.99% forest cover and the other 15 pixels dominated by non-forest contained between 90-99.99% non-forest cover. The characteristics of the second and third training samples were the same as those of the first one except the dominant class proportion were between 80–89.9% and 70–79.9%, respectively. The last sample contained 100 randomly selected mixed pixels.

Two cases were analysed, hereafter referred to as the unadjusted case and adjusted case. In the unadjusted case, the spectral response of the end-member of a land cover class was derived from the average spectral response of the training pixels which have that class as the dominant class. These end-member values were then used for the input to the conventional LMM (equations 4.1 and 4.2) to estimate the class proportions within pixels of the whole image.

Dominant class composition	Number of pixels in the training data
90 – 99.99	30
80 – 89.99	30
70 – 79.99	30
Random mixed pixels	100

Table 4.6: Training samples used to estimate spectral end-member values.

In adjusted case, the pixels with known class proportions in the training sample were used to estimate the spectral end-member values using equation 4.6. The sub-pixel class compositions within each image pixel were then predicted using these estimated end-member values from the conventional LMM (equations 4.1 and 4.2).

The testing sets used to assess the accuracy of the sub-pixel composition estimates derived from LMM contained 500 randomly selected pixels as described in section 4.4.1.3. Figure 4.14 shows the relationship between the proportional coverage of a class derived from LMM and actual data of the testing dataset.

From Figure 4.14 it was apparent that the adjusted case was more accurate than the unadjusted case, although the difference was statistically insignificant (95% level of confidence). The accuracy of the LMM using pure pixels (Figure 4.13) to derive end-member was less accurate than those derived from mixed pixels in both cases, adjusted and unadjusted (Figure 4.14). This indicated that identifying end-members as the "purest" pixels may lead to systematic bias in end-member proportions if the "purest" pixels were not spectrally representative of the end-members. The end-members not only need to be pure but also to be representative.

The results highlighted the potential to derive end-members from mixed pixels for LMM in the estimation of sub-pixel proportions. This may help to derive the end-member spectra for the imagery which is mainly dominated by mixed pixels.

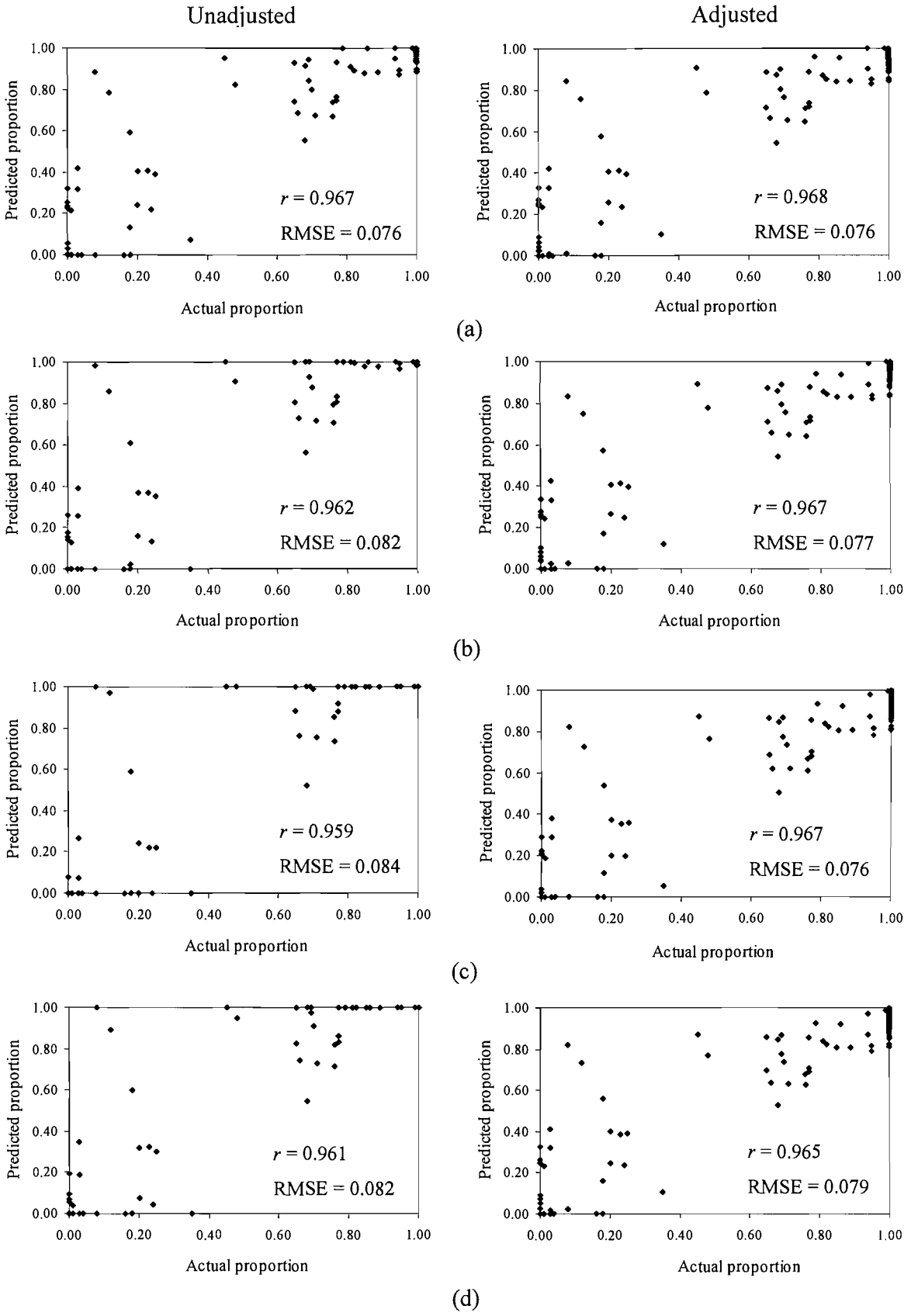


Figure 4.14: Relationships between predicted and actual cover of forest with different dominant class proportion within the training pixels used to derive end-member spectra, (a) 90 - 99.99%; (b) 80 – 89.99%; (c) 70 -79.99%; and (d) randomly selected mixed pixels.

4.4.3 Impacts of intra-class spectral variation on the estimation of deforestation

By using soft classifications, the sub-pixel class composition estimates derived from two time periods can be compared to derive the sub-pixel scale change in class composition in time. This has been undertaken with a single class composition prediction for each image pixel when the class centroids were used to define the end-members. The centroids, however, do not fully describe the classes spectrally (section 4.3). In this analysis, every pixel in the training set was used to provide the end-member spectra and a distribution of possible sub-pixel class compositions could be derived for each pixel in the simulated coarse spatial resolution imagery. The distribution was then used for the assessment of tropical deforestation when used in post-classification comparison approach.

4.4.3.1 Sub-pixel classification and the estimation of deforestation

The training data which included 90 pure pixels for each of forest and non-forest classes was used to derive the spectral response of the end-members. The spectral response of the classes in two images were summarised in Table 4.5. The forest and, in particular, the non-forest classes displayed a degree of variation in the feature space (Figure 4.12). For each pixel in the spatially degraded imagery, the sub-pixel class composition predictions were derived. Initially the end-members were defined as the class centroids for input to the conventional LMM. The accuracy of the sub-pixel class composition estimates was evaluated using correlation coefficient and RMSE. There was a strong and significant relationship between the actual and predicted class compositions, although with a large degree of scatter (Figure 4.15). The sub-pixel scale deforestation information can be constructed from the comparison of the sub-pixel class composition estimates derived from two time periods (Figure 4.17(a - c)).

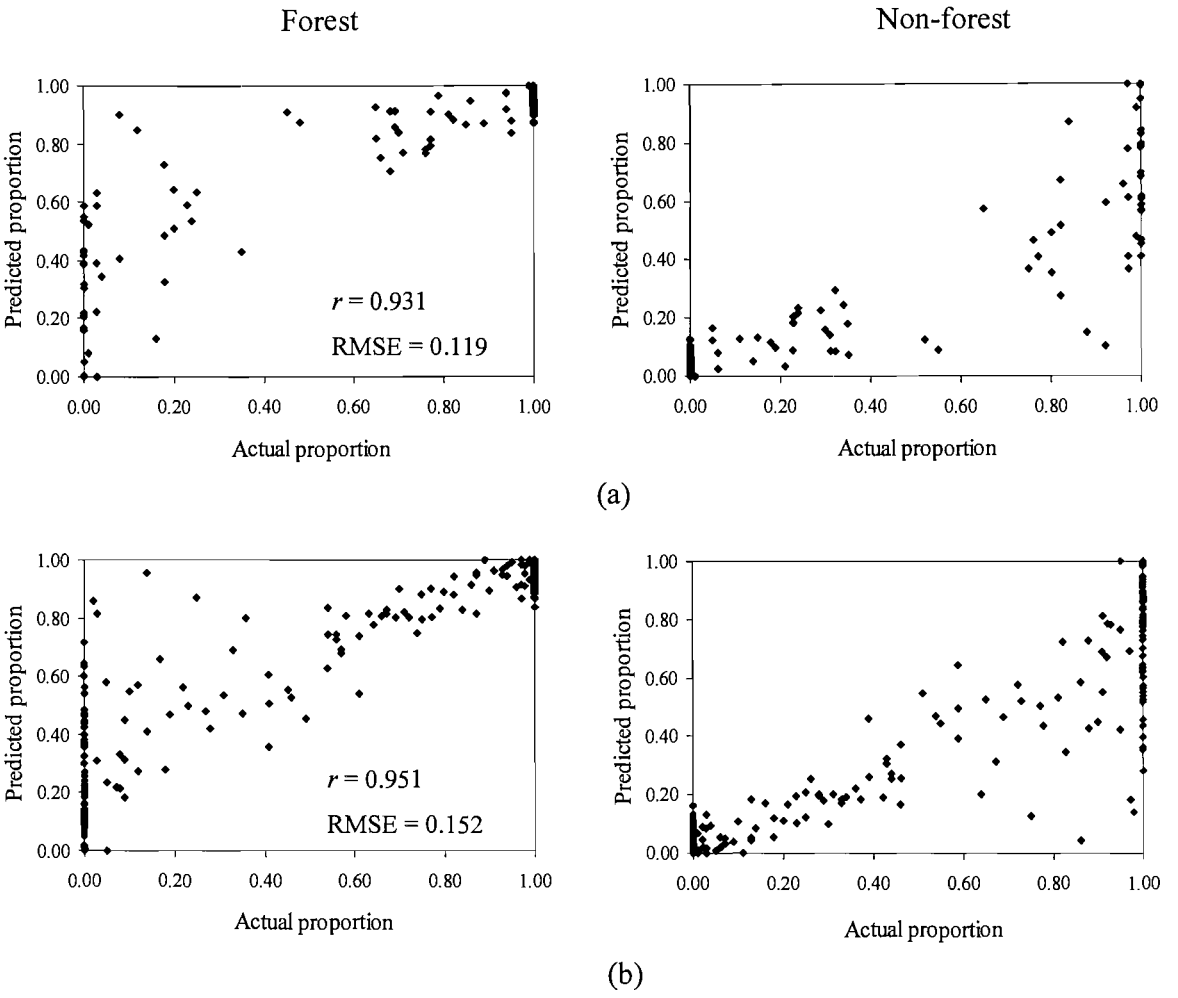


Figure 4.15: Relationships between predicted and actual cover (a) 1984 and (b) 1988.

4.4.3.2 Impacts of intra-class spectral variation on the estimation of deforestation

As with the simulated data set in section 4.3, the centroids do not fully describe the characteristics of the classes spectrally. The spectral responses of the individual training pixels were, therefore, used to define the end-member spectra to repeat the analyses many times. Consequently, a distribution of possible sub-pixel class composition could be derived for each pixel in the simulated coarse spatial resolution imagery. Comparison of the distributions derived at two time periods may provide a richer indication of deforestation information than the comparison of the single class composition estimate from standard sub-pixel classification. This was illustrated for six pixels in Figure 4.16.

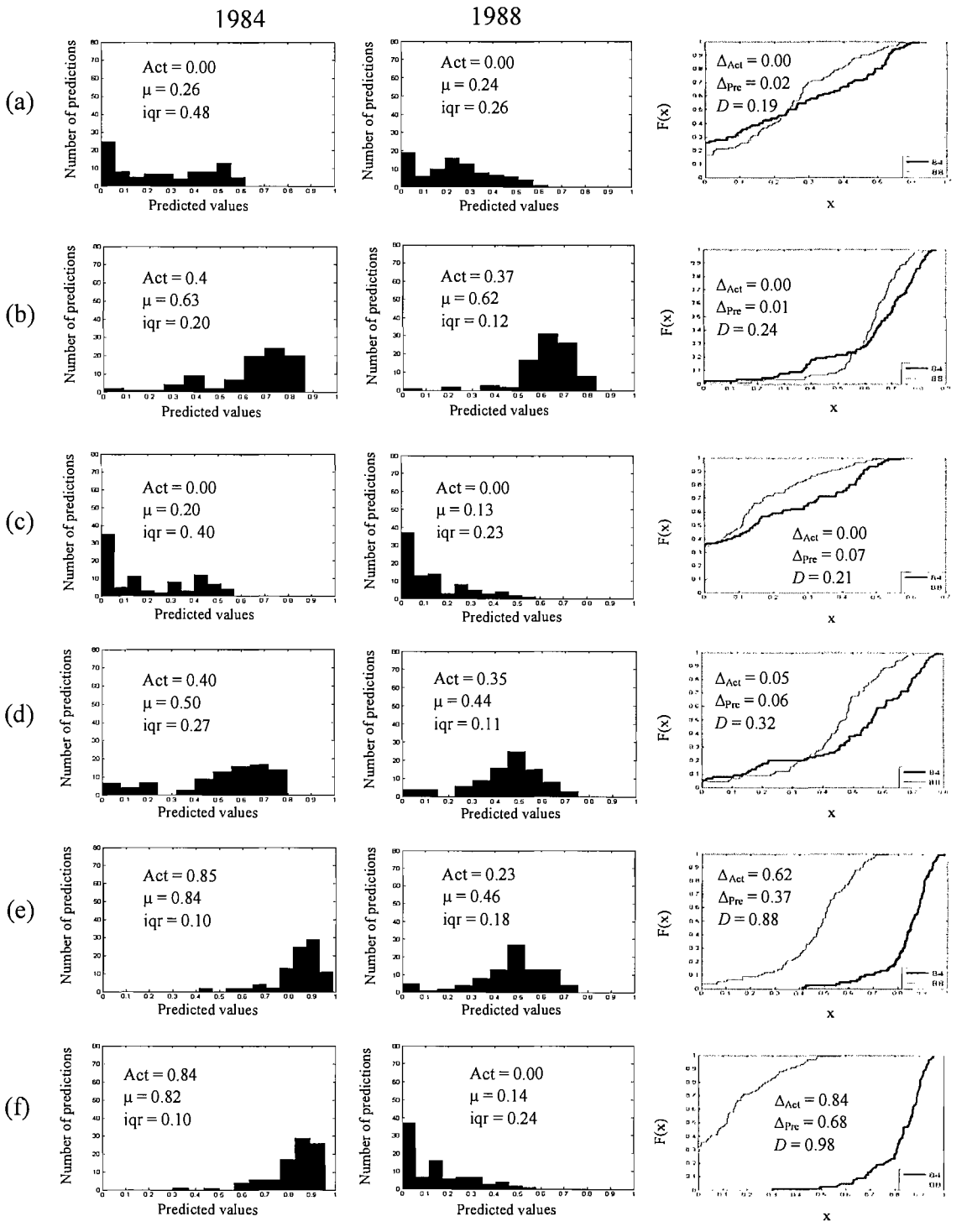


Figure 4.16: Sub-pixel class composition distributions for six pixels derived from 1984 and 1988 images together with a graphical comparison of their difference in cumulative distributions. Summary statistics are given for each case in Table 4.7 (Act: Actual class proportion; μ: mean of the distribution; iqt: interquartile range of the distribution; Δ_{Act}: deforestation from actual values; Δ_{Pre}: deforestation from single prediction from a LMM using class centroids as end-members; D: maximum absolute difference between two cumulative probability distributions).

One possible approach to compare the distributions of two samples is using the Kolmogorov-Smirnov (K-S) test (Daniel, 1978). The K-S statistic of two distributions is simply the maximum absolute difference between their cumulative probability distributions, written as D. This statistic can be used to test the (null) hypothesis that the population distributions are identical and, therefore, that both samples have been drawn from the same population. If the calculated value of D is greater than the tabled critical value at a specified significance level, the null hypothesis can be rejected at that level.

By comparing directly the single estimate of sub-pixel class composition for each image pixel from standard unmixing analysis, there was a single estimate of change detection (Table 4.7). The comparison based on the single set of forest proportion predictions may be unwise. For example, in Figure 4.16c, the change detection based on single prediction of sub-pixel class composition in 1984 and 1988 was 6.6% (with the pixel size of 300m, this figure was corresponding to an area of 594 m^2 at the ground surface). However, using D value ($D = 0.21$) in the K-S test two distributions of possible class composition predictions in these times did not differ significantly at 95% level of confidence (Table 4.7). The value of D from K-S test of two samples, therefore, may be used to test the significant difference of the change detection from soft classification, especially for small change detection. Figure 4.17e shows a proportion image in which each pixel's value was the value of D when comparing two distributions of possible mixing predictions for that pixel in 1984 and 1988 using K-S test. This proportion image may provide the additional information about the significant difference of change detection.

Case	Δ from actual (%)	Δ from prediction (%)	*D*
a	0.0	2.3	0.19
b	0.0	0.8	0.24
c	0.0	6.6	0.21
d	5.0	6.3	0.32
e	62.0	37.4	0.88
f	84.0	68.5	0.98

Table 4.7: Summary of changes based on actual and single sub-pixel estimation from a LMM using class centroids as end-members. The *D* value was derived from a Kolmogorov-Smirnov test, with a critical value of 0.21 at 95% level of confidence.

4.4.3.3 Different scenarios of deforestation

The direct comparison of the single predictions from standard unmixing analysis (e.g. from Figures 4.17a and 4.17b, the proportion images of forest in 1984 and 1988) achieved a single estimate of change detection (e.g. Figure 4.17c). In reality, since a distribution of possible class compositions may be derived for each pixel, it may allow to view the change in different perspectives. For example, the use of the single prediction from a standard application of the LMM may be unwise and the distributions used to indicate change. So rather than directly compare single predictions (Figure 4.17c), one could, for instance, focus on the upper and lower quartiles of the distributions to derive different scenarios of change detection. Figure 4.18a to 4.18d shows the four possible scenarios of deforestation between 1984 and 1988 using the distributions of possible forest proportion predictions. Assuming that deforestation was viewed negatively, some scenarios might be considered a relatively 'good' case (Figure 4.18b) and a relatively 'bad' case scenario of change detection (Figure 4.18c). This provided useful information and qualification to the standard use of single prediction of sub-pixel class composition.

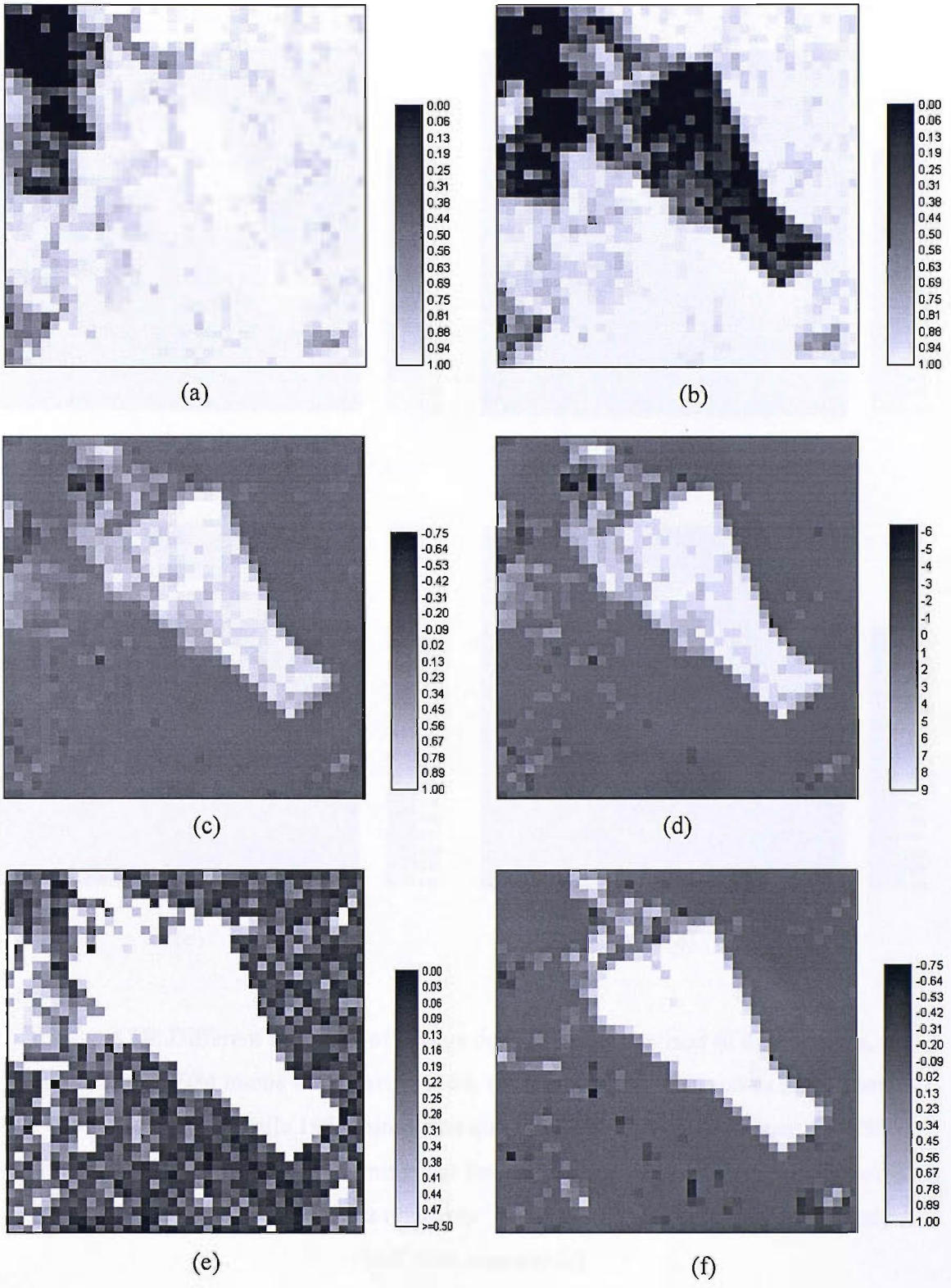


Figure 4.17: Sub-pixel estimates and change detection, (a) sub-pixel forest cover estimate 1984; (b) sub-pixel forest cover estimate 1988; (c) Difference (1984-1988) in forest cover estimates; (d) change assessed if sub-pixel estimates placed in classes of 0.1 (10%) cover; (e) *D* value from Kolmogorov-Smirnov test; and (f) actual change (1984-1988).

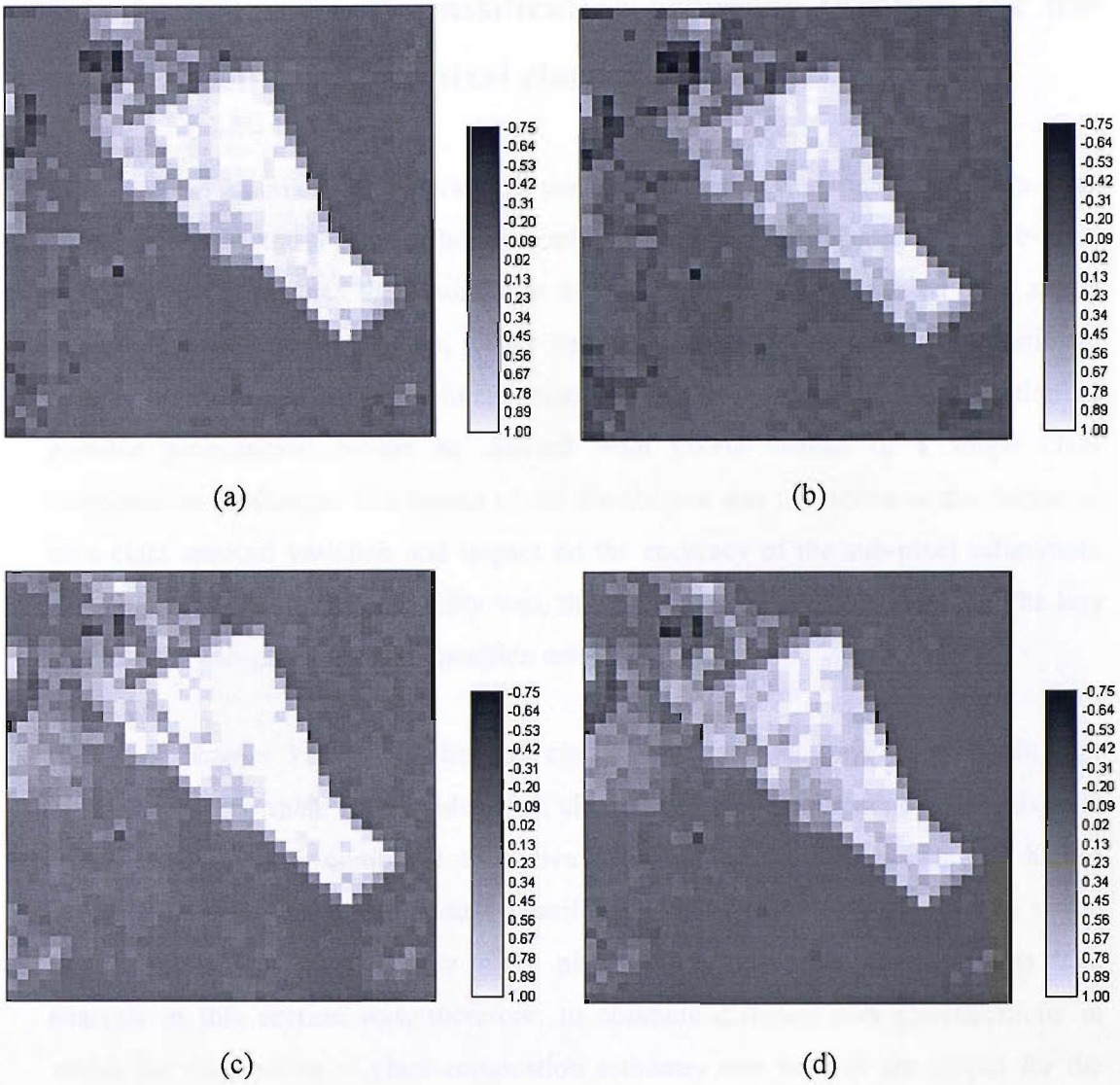


Figure 4.18: Different scenarios of change derived by comparison of distributions, (a) first quartile 1984 minus first quartile 1988; (b) first quartile 1984 minus third quartile 1988; (c) third quartile 1984 minus first quartile 1988; and (d) third quartile 1984 minus third quartile 1988. Assuming that forest clearance is viewed negatively, some scenarios might be considered a relatively 'good' case scenario (b) and a relatively 'bad' case scenario (c).

4.5 Increasing soft classification accuracy through the use of an ensemble of sub-pixel classifiers

The analyses detailed in section 4.3 demonstrated that the nature of intra-class variation in spectral response has a marked impact on the accuracy of sub-pixel estimation as it violates the assumption that a class can be represented by a single spectral endmember. As a result, it may be inappropriate to provide a single estimate of class composition for each image pixel. It was suggested that a distribution of possible proportional covers be derived from pixels instead of a single class compostition prediction. The spread of the distribution was a function of the degree of intra-class spectral variation and impact on the accuracy of the sub-pixel estimation. The larger the intra-class variability was, the more spread the distribution and the less accurate the sub-pixel class composition estimates.

Results in chapter 3 indicated that soft classification accuracy may increase through the use of an ensemble of individual soft classifications. The outputs of the individual soft classifiers were combined to derive the final output which obtained higher accuracy than that of the individual classifiers. This was implemented using a single class composition prediction for image pixels in the individual classifications. The analysis in this section was, therefore, to combine different soft classifications in which the distribution of class compostion estimates was used as the output for the individual classifications. The main aims of this analysis were two fold:

- To increase the accuracy of sub-pixel class composition estimates

- To reduce the variety of possible sub-pixel class composition estimates.

4.5.1 Data and methods

4.5.1.1 Data

The data set obtained in 1988 (section 4.4.1) was used for the analysis. It included an original Landsat TM image (Figure 4.10b) which was used as the reference data and the degraded coarse spatial resolution image of 300 m (Figure 4.11b) which was used to derive the sub-pixel classifications. The TM image was classified into forest and non-forest visually and used as the reference data for the analysis. Sub-pixel classifications estimated the forest and non-forest proportions within each image pixel.

4.5.1.2 Individual sub-pixel classifications

Different soft classifications may be derived from different classification algorithms or from the same classification algorithm using different approaches to identify training data or using different parameter settings. In this analysis, the LMM was used to derive sub-pixel class compositions. As with the analyses in sections 4.3 and 4.4, it was possible to derive a distribution of class composition estimates for each pixel. The distributions were derived by using each training pixel's spectrum as an endmember in the LMM. Different distributions of possible class composition estimates for each class within each image pixel may be derived from different sets of end-members used as the input for the LMM. Here, the end-members defined from pure pixels. Two approaches to derive 'pure' pixels which then used as the end-member spetra were applied. In the first approach, the pure pixels were derived from the reference data which were used as the training data in section 4.4.1.3. Pixel purity index (PPI) was the second approach to derive pure pixels (Plaza *et al.*, 2004, Chang and Plaza, 2006) and these were used as the second training data. Hereafter, these two approaches were called the reference data approach and PPI approach, respectively.

The first training data consisted of 180 randomly selected pure pixels derived from the reference data with 90 pixels for each land cover class (e.g., forest and non-forest). The second training data also consisted of 180 randomly selected pure pixels (e.g., 90 pixel for forest and 90 pixels for non-forest) derived from the PPI approach using the simulated coarse spatial resolution imagery. The statistical information of the two land

cover classes is shown in Table 4.8. For illustrative purposes, the data were subjected to principal components analyses from three used wave-bands and the first two components (PC1 and PC2) that explained most of the variation in the data sets were used to display classes in the feature space. The scatter plots of the training samples were displayed in this manner in Figure 4.19. The testing set contained 30 randomly selected mixed pixels.

Every pixel in each training data set was used to provide the end-member spectrum to estimate the sub-pixel class composition for each pixel in the spatially degraded imagery. By unmixing the spectral response of a pixel many times with different end-members derived from each pixel in two training data sets, a series of sub-pixel class composition estimates could be derived for a pixel of any given spectral response. As a result of this it was possible to form two different distributions of possible sub-pixel proportion estimates of a class within an image pixel. Two distributions of each class composition for each image pixel were then combined to get the final combined distribution.

4.5.1.3 Methods for combining distributions

Section 3.2.2 presented four possible approaches to combining individual soft classifications. These four ensemble approaches were implemented using a single class composition prediction for image pixels in the individual classifications. The four ensemble approaches may not be suitable for this analysis as the distribution of possible class composition estimates was used as the output for the individual classifications. One possible method to combine two different distributions of class composition estimates for each pixel which derived from two individual soft classifications was investigated. This combined method was a simplified form of the Bayesian integration updating implicit in the Ensemble Kalman Filter (EKF) based approach (Evensen, 2003).

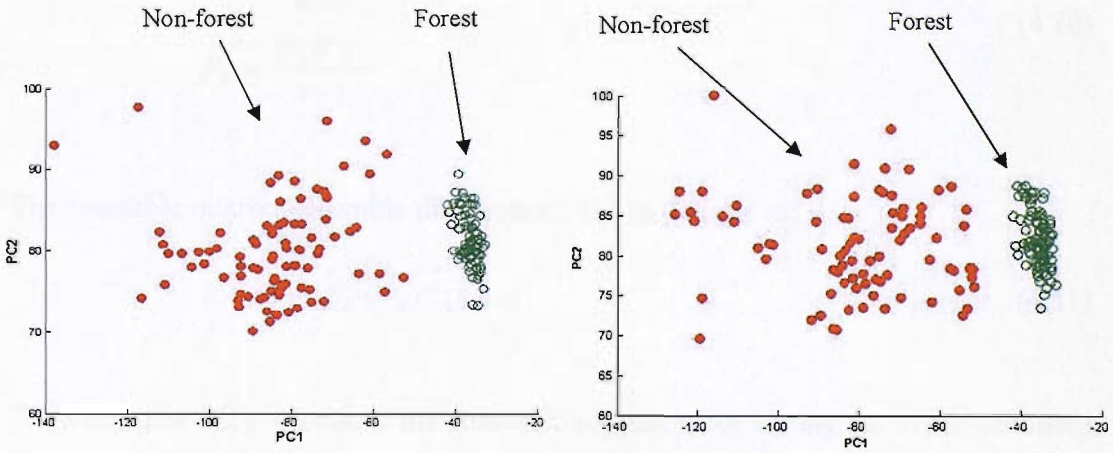
(a)

Class	Mean			Standard deviation			Variance-covariance matrix
	Band 4	Band 5	Band 7	Band 4	Band 5	Band 7	
Forest	72.23	51.79	9.70	2.92	2.37	0.54	$\begin{bmatrix} 8.53 & 4.07 & 0.61 \\ 4.07 & 5.63 & 1.19 \\ 0.61 & 1.19 & 0.29 \end{bmatrix}$
Non-forest	62.00	94.04	29.14	6.51	12.62	6.49	$\begin{bmatrix} 42.33 & -32.49 & -17.51 \\ -32.49 & 159.13 & 80.58 \\ -17.51 & 80.58 & 42.13 \end{bmatrix}$

(b)

Class	Mean			Standard deviation			Variance-covariance matrix
	Band 4	Band 5	Band 7	Band 4	Band 5	Band 7	
Forest	72.10	52.01	9.84	3.11	2.58	0.70	$\begin{bmatrix} 9.66 & 4.22 & 0.63 \\ 4.22 & 6.67 & 1.66 \\ 0.63 & 1.66 & 0.49 \end{bmatrix}$
Non-forest	62.19	91.88	26.70	6.33	16.96	8.35	$\begin{bmatrix} 40.04 & -42.66 & -26.09 \\ -42.66 & 287.75 & 138.39 \\ -26.09 & 138.39 & 69.67 \end{bmatrix}$

Table 4.8: Class description for the training data used (a) reference data approach; (b) PPI approach.



(a) (b)

Figure 4.19: Distribution of the training samples in feature space with PC1 and PC2 are the first two principal components derived from three spectral wave-bands (a) reference data approach; (b) PPI approach.

Assuming two matrices holding two distributions of possible class composition predictions derived from two sub-pixel classifications:

$$A = (a_1, a_2, \ldots, a_n) \in \Re^{1 \times n}$$
$$B = (b_1, b_2, \ldots, b_n) \in \Re^{1 \times n} \tag{4.7}$$

where n is the number of class composition predictions in the distributions.

The ensemble means are stored in $\overline{A}$ and $\overline{B}$ that can be defined as:

$$\overline{A} = A1_n$$
$$\overline{B} = B1_n \tag{4.8}$$

where $1_n \in \Re^{n \times n}$ is the matrix where each element is equal to $1/n$. We can define the ensemble perturbation matrix as

$$A' = A - \overline{A}$$
$$B' = B - \overline{B} \tag{4.9}$$

The ensemble variance can be defined as

$$P_A = \frac{A'(A')^T}{n-1}$$
$$P_B = \frac{B'(B')^T}{n-1} \tag{4.10}$$

The ensemble matrix (ensemble distribution) can be defined as:

$$E = A + P_A(P_A + P_B)^{-1}(B - A) \tag{4.11}$$

To investigate the potential of the ensemble approach, the testing set which consisted of 30 mixed pixels was used. The process of the analysis was illustrated in Figure 4.20.

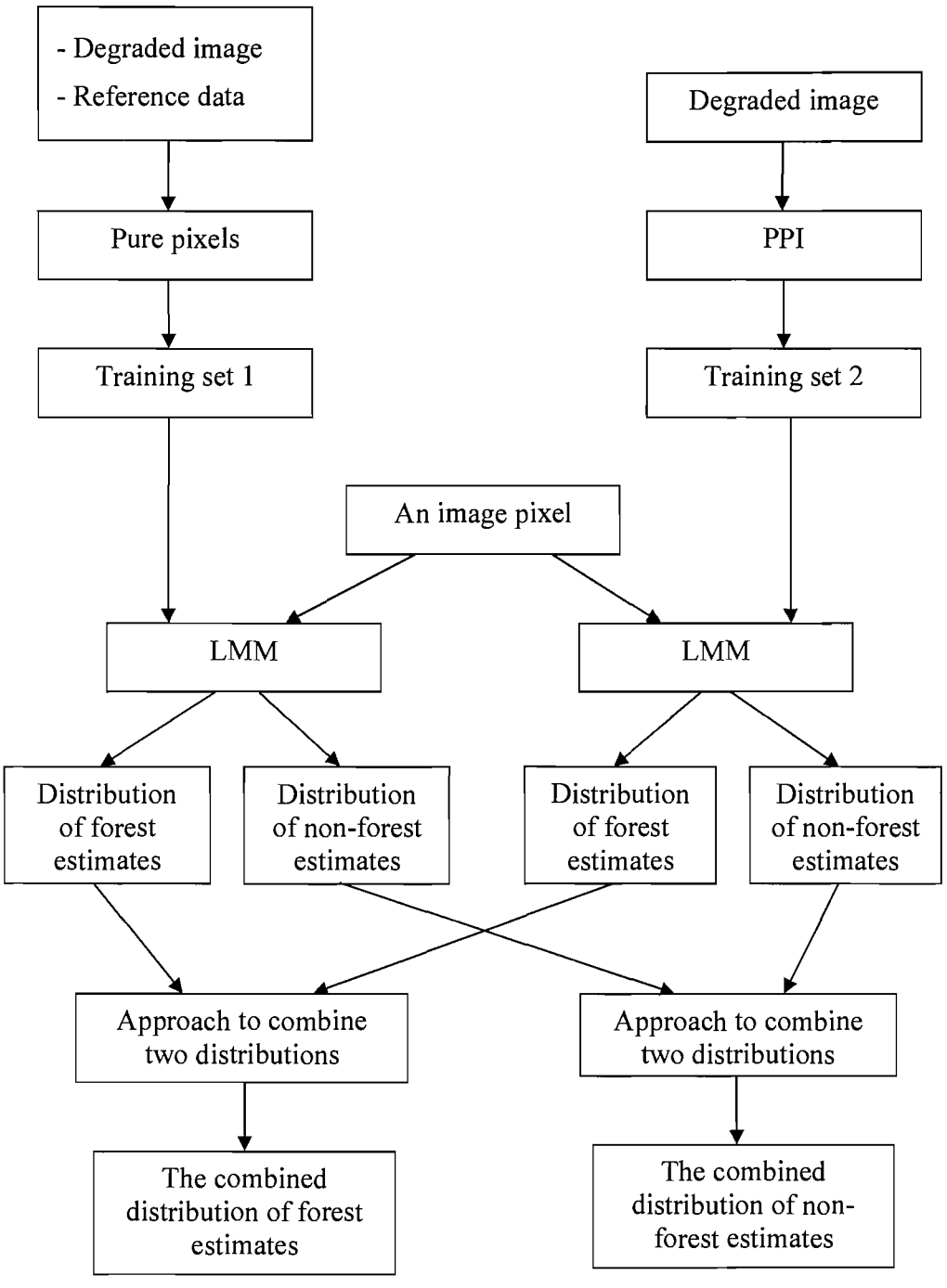


Figure 4.20: Flow diagram of the ensemble sub-pixel classifications using distribution of possible class composition estimates as the output.

4.5.2 Results

As with the analyses in sections 4.3 and 4.4, it was possible to derive a distribution of class composition estimates for each pixel in the simulated coarse spatial resolution imagery. Figures 4.21a and 4.21b present an example of two distributions of possible sub-pixel proportion estimates of forest within an image pixel. They were derived using the two training data sets derived from the reference data approach and the PPI approach, respectively as the end-member spectrum to input for the LMM. These two distributions were then combined using some equations of Bayesian integration (section 4.5.1.3) to derive the ensemble distributions (Figure 4.21c). From Figure 4.21, it was apparent that the ensemble distribution was derived with less uncertainty (e.g., less spread) than the two individual distributions.

To evaluate the potential of the ensemble approach, the outputs of the ensemble approach and of the individual sub-pixel classifications were assessed. They were evaluated using the variability of the distribution of possible sub-pixel class composition estimates derived.

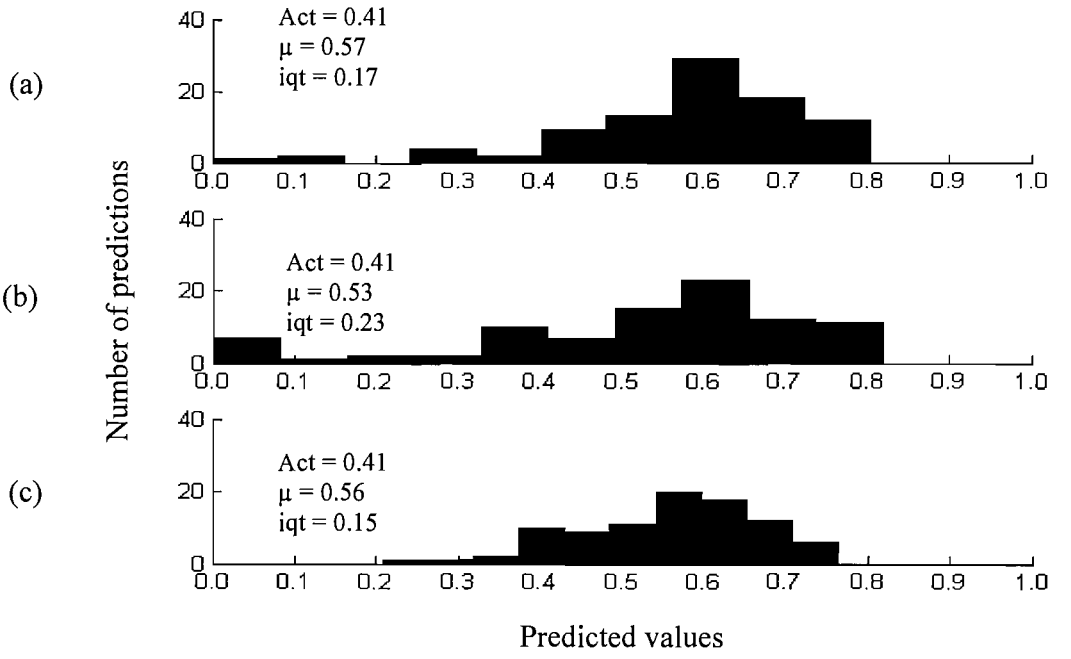


Figure 4.21: Distributions of possible sub-pixel proportion estimates of forest within an image pixel, (a) pure pixels from reference data; (b) PPI approach; and (c) ensemble output.

4.5.1.4 Evaluation of the accuracy of sub-pixel class composition predictions

The accuracy of the sub-pixel class composition estimates was evaluated based on a comparison of predicted class proportions with that derived from the reference data, the 30 m spatial resolution image using correlation coefficient and RMSE. Figure 4.22 shows the relationships between predicted and actual cover of two individual sub-pixel classifications and of the ensemble approach. From Figure 4.22, it was apparent that the accuracy of the sub-pixel classification using the end-members derived from pixels of the PPI approach (r = 0.95 and RMSE = 0.15) was more accurate than that from the reference data approach (r = 0.94 and RMSE = 0.18), although the difference between two correlation coefficients was statistically insignificant (95% level of confidence). The accuracy of the sub-pixel class composition estimates derived from the ensemble approach (r = 0.94 and RMSE = 0.17) was slightly lower than that from the PPI approach, but it increased compared with the accuracy derived from the reference data approach.

4.5.1.5 Evaluation of the width of the distribution of possible sub-pixel class composition estimates

Since a distribution of possible class compositions may be derived for each image pixel, it may be preferable to be aware of the spread of the distribution using the inter-quartile range. The variety of possible compositions will be a function of the degree of intra-class variation and impact on the accuracy of the sub-pixel estimation using a conventional LMM (section 4.3.3). To assess whether the variability of possible composition in the ensemble approach was reduced compared with that in two individual sub-pixel classifications, the average interquartile range of the class composition distributions generated were calculated and compared (Table 4.9).

According to Table 4.9, the distribution of sub-pixel class composition estimates in the ensemble approach was less spread than that from two individual sub-pixel classifications. This was highlighted that combining the output of different sub-pixel class composition estimates, the variety of possible sub-pixel class composition

estimates may reduce so that may increase the confidence of soft classification using one single prediction of class composition.

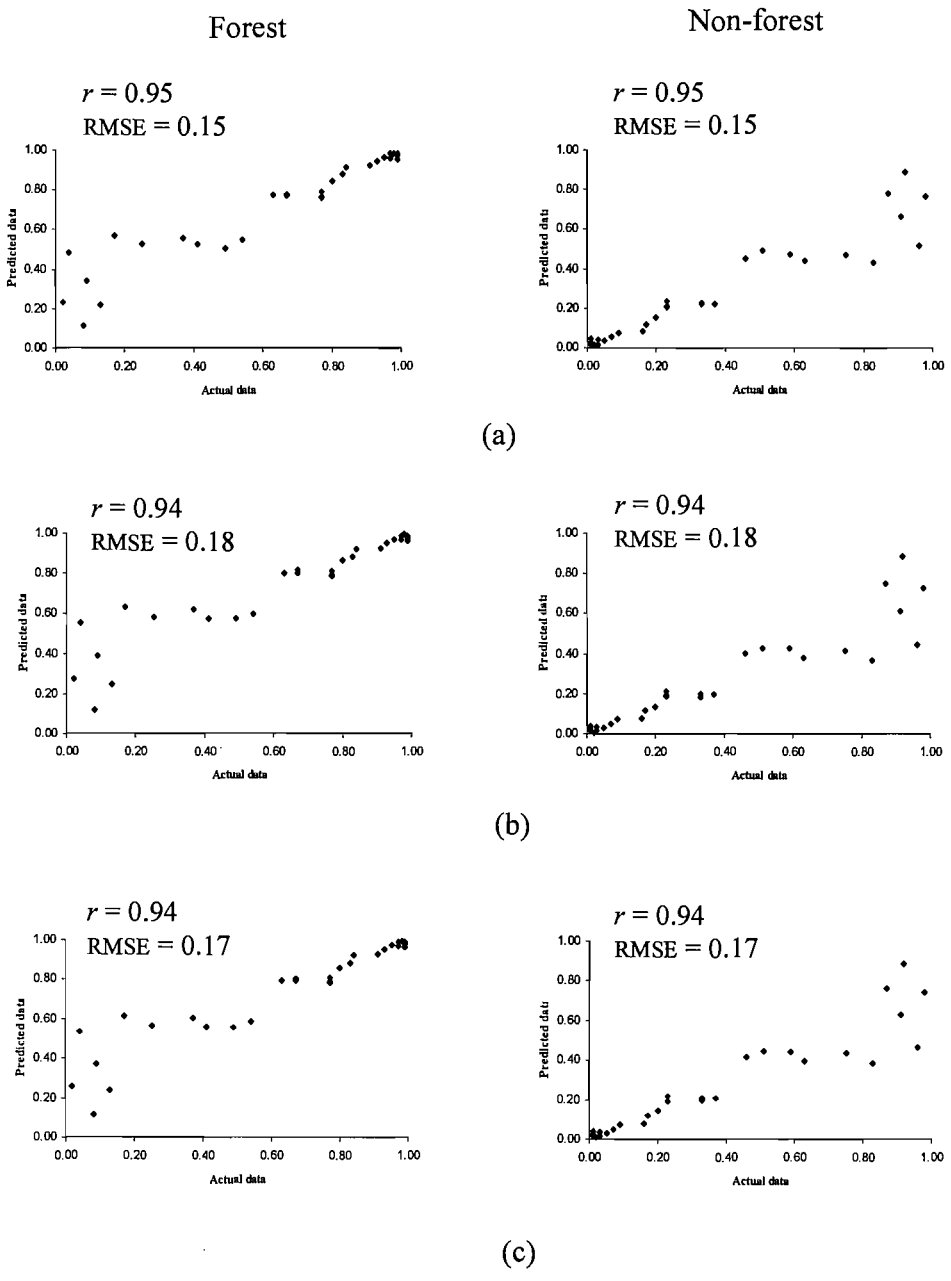


Figure 4.22: Relationships between predicted and actual cover (a) from PPI approach; and (b) from pure pixels of the reference data; and (c) from combined output.

Method	Reference data approach	PPI approach	Ensemble approach
Average interquartile range of the distribution	0.10	0.13	0.09

Table 4.9: With of the distribution of possible class composition estimates.

4.6 Summary and Conclusions

In many sub-pixel scale analyses in remote sensing an assumption made implicitly is that the classes may be described by a single endmember. Following from this, great trust is placed in the single class composition prediction made in sub-pixel classification. However, intra-class spectral variability means that no single spectrum can adequately describe a class. Instead of just one single sub-pixel class prediction, which derived from the constant values of spectral end-members, a number of sub-pixel predictions were derived for the estimation of class proportion within image pixel. A consequence of this was that a distribution of possible class compositions exists for each pixel.

The impacts of intra-class spectral variability on the accuracy of land cover classification as well as on the spread of the distributions of possible sub-pixel class composition predictions were investigated using simulated data sets. The results indicated that intra-class spectral variability was an important factor that affects the estimates of sub-pixel class proportions. It has been shown that the accuracy of sub-pixel class composition estimation is a function of the degree and nature of intra-class spectral variation (section 4.3.3). Similarly, the spread of the distribution of possible sub-pixel class composition predictions within image pixel also depended on the intra-class variability (section 4.3.4). The larger the intra-class variability was the more spread the distributions.

Some outputs were derived from the distribution of possible sub-pixel class composition predictions. These were variation images and closeness images. The variation images showed the variability (or spread) of the distribution while closeness

image showed the measures of the distance between the actual class proportions and the predicted class proportions based on the distribution of sub-pixel estimates.

The impacts of intra-class spectral variation on sub-pixel estimation and its implications on deforestation were undertaken using real remotely sensed data, the Landsat TM imagery. First of all, the selections of the spectral signatures for the end-members from mixed pixels with known class proportions were studied. This highlighted the potential of incorporating mixed pixels in the end-member selection for LMM in the process of sub-pixel class composition estimation.

Distributions of class composition estimates for each pixel were used for the assessment of tropical deforestation. Comparison of distributions derived at the two time periods may result in a different interpretation to that derived through comparison of the single class composition estimate derived from standard sub-pixel analysis. The direct comparison of the single predictions from a standard unmixing analysis yields a single estimate of land cover change. The uncertainty of the deforestation derived from these single predictions was assessed by comparing the distributions of each land cover class proportion for each image pixel in two time periods. A K-S test based approach was used to compare two distributions to test the statistical significance of differences of forest in each image pixel area in two periods of time.

The distributions of class composition estimates also provided a richer description of the class composition that may allow the change to be viewed from different perspectives. For example, the danger in using the single prediction from a standard application of the linear mixture model could be recognised and the distributions used to indicate change. This could be taken from a range of perspectives. So rather than directly compare single predictions, one could, for instance, focus on the upper and lower quartiles of the distributions to derive what could be considered by some to be a relatively good and bad case scenario of change. This provides a useful extension and qualification to the standard use of single prediction estimates.

The ensemble of the sub-pixel classifications was explored. This was implemented by combining the output of different sub-pixel classifications by using the distribution of possible class composition estimates as the output for each pixel in the individual classifications. Different distributions of possible class composition estimates for each

class within each image pixel were derived with different set of end-members using as the input for LMM. Pure pixels derived from reference data and from PPI approach were used as two different training data sets to derive end-members. As a result, two different distributions of possible sub-pixel proportion estimates of a class within each image pixel were derived. They were then combined to get the final combined distribution. It was suggested that although ensemble based approach may not increase sub-pixel classification accuracy compared with the most accurate individual classification, it increased accuracy compared with other individual classification of the ensemble. Furthermore, ensemble approach reduced the variety of possible sub-pixel class composition estimates so that may increase the confidence of standard sub-pixel analysis.

In conclusion, results from this chapter have shown that the accuracy of sub-pixel class composition estimation is a function of the degree and nature of class spectral variation as it violates the assumption that a class can be represented by a single spectral end-member. A distribution of possible class compositions could be derived from pixels instead of a single class composition prediction. This distribution provided a richer indication of possible sub-pixel class compositions and may be used to derive different scenarios of change when used in a post-classification comparison type approach to change detection. The application of this for other applications such as super-resolution mapping will be the topic for next research.

Chapter 5

Reducing the impacts of intra-class spectral variability on soft classification and its implications for super-resolution mapping

5.1 Introduction

Soft classifications predict the proportion of each land cover class within each image pixel. However, they do not indicate where the land cover classes are spatially located within the pixels. The sub-pixel class proportions may, however, be located geographically through super-resolution mapping.

Super-resolution mapping is a set of techniques to predict the location of land cover classes within each image pixel based on the proportion images derived from soft classification (Tatem *et al.*, 2001, 2002, Foody *et al.*, 2005, Nguyen *et al.*, 2006). In super-resolution mapping some approaches attempt to maintain the class proportion information output from soft classification (Tatem *et al.*, 2002, Muslim *et al.*, 2006). That is, if the soft classification indicates that a pixel contains 30% forest the super-resolution mapping will locate the pixel with 30% forest. This trust in the single set of class proportion estimates may be unwise. This seems to be based on the assumption that a class can be represented by a single spectral end-member. This is likely to be unrealistic as the results stated in section 4.3 that classes typically display a degree of spectral variability (Petrou and Foschi, 1999).

The analyses detailed in chapter 4 demonstrated that soft classification accuracy was negatively related to the degree of intra-class spectral variation and this against the assumption that a class can be represented by a single spectral end-member. It may be inappropriate to derive a single prediction for class proportion for each pixel. A consequence of this is that a distribution of possible class composition exists for each pixel. The distribution provided a richer indication of possible class composition estimates. It was then used to derive different scenarios of change when used in post-

classification comparison technique to change detection. Moreover, the distribution of possible class composition estimates may also be used to highlight a limitation of conventional soft classification output for super-resolution mapping.

This chapter aims to explore the impacts of intra-class spectral variability, which resulted in a distribution of possible class proportions for each image pixel, on super-resolution mapping and to investigate a possible approach to reduce these impacts through the reduction of the intra-class spectral variation. These issues are illustrated with an example of shoreline mapping in Isle of Wight, UK from Landsat ETM+ data.

The outline of the chapter is:

1. To analyse data used in the chapter. Attention focuses on a small region of Isle of Wight in UK.

2. To investigate methods used to derive shoreline mapping from the output of soft classification.

3. To explore the impacts of intra-class spectral variability on super-resolution mapping. Here attention focused on a common type of analysis in a major environmental science context; the assessment of shoreline mapping using soft classification output.

4. To investigate an approach to reduce the impacts of intra-class variation on soft classification predictions as well as on shoreline mapping. The approach investigated here is the reduction of the intra-class spectral variation of the land cover classes used for sub-pixel classification. In particular, attention focused on the division of a land cover class into sub-classes.

5. Finally the summary and conclusions will close the chapter

5.2 Study area and data used

To illustrate the impacts of the intra-class spectral variation on super-resolution mapping some analyses of real remotely sensed data were undertaken. Here, the assessment of shoreline mapping using soft classification as input was implemented as an example. Attention was focused on a small region on the Isle of Wight in the United Kingdom. This study area was extracted from a Landsat ETM+ image of the region with path of 202 and the row of 25 (Figure 5.1). It was apparent from Figure 5.1 that there was a large degree of intra-class spectral variation for land cover classes to be mapped. Specifically, there were many land cover classes and the water class varied in terms of the turbidity. It is, therefore, difficult to define the land and water boundary. The image was acquired on 12 May 2001 and extracted in the internet from the Global Land Cover Facility (GLCF) at the University of Maryland at http://glcf.umiacs.umd.edu. It was already system-corrected using UTM-30 and WGS-84 as the projection and reference ellipsoid respectively and acquired in 8 spectral wavebands with a spatial resolution of 30 m, except spectral band 6 with spatial resolution of 120 m and band 8 with spatial resolution of 15 m. Since the spectral response of water in band 4, 5, and 7 was similar and the purpose of the research was to assess the impact of class spectral variability on the shoreline mapping only three bands (ETM+ band 1, band 2 and band 3) were used for the analyses.

5.2.1 Data processing

To evaluate the error in shoreline mapping, a ground/ reference data was required. Due to the difficulty in accurately obtaining the actual shoreline, this ground dataset was generated from the original Landsat ETM+ image. Each pixel in this image was assumed to be pure and classified into two land cover classes, land and water, using a supervised classifier, the MLC. The classified image was then vectorised along the boundary between land and water classes to generate the reference shoreline (Figure 5.1).

Figure 5.1: Three-band composite Landsat TM images using bands 3, 2, and 1 mapped to red, green, and blue respectively of the study area in part of Isle of Wight in United Kingdom (——— : Reference shoreline).

The Landsat ETM+ image was spatially degraded by a factor of 10 to simulate data sets with a relatively coarse spatial resolution of 300 m. This coarse spatial resolution is comparable to the spatial resolution of medium spatial resolution system such as MODIS and MERIS. The degraded image was obtained by aggregating pixels to the desired spatial resolution, with each degraded DN expressed as the mean DN of the original un-degraded pixels it comprised. The degraded data set represented simulated coarse spatial resolution imagery (Figure 5.2). This data set was used in the analyses to predict the shoreline.

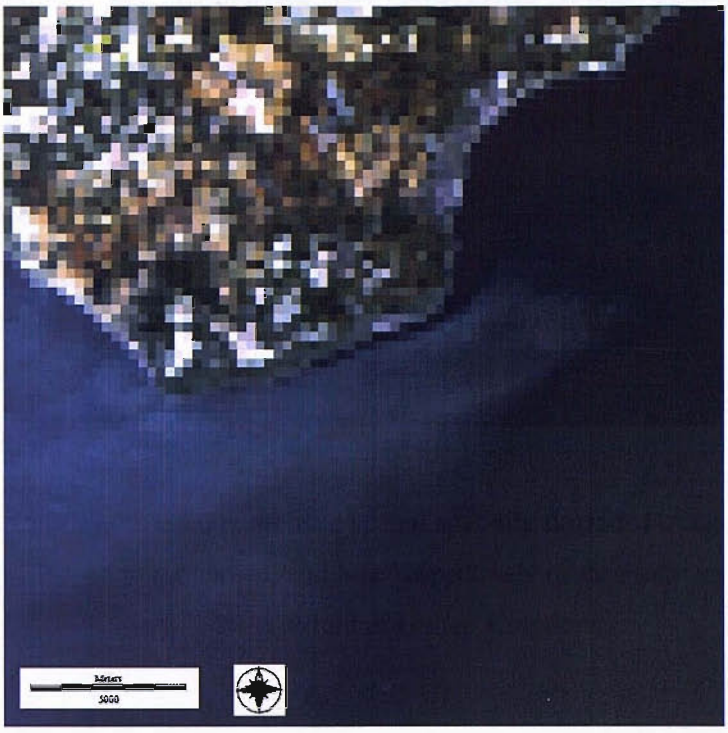


Figure 5.2: Three-band composite spatial spatially degraded images using bands 3, 2, and 1 mapped to red, green, and blue respectively of the study area in part of Isle of Wight in United Kingdom.

5.3 Methodology

Land cover mapping from remotely sensed imagery is commonly obtained through the conventional hard classification that located each pixel to the class with which it has the greatest similarity. Using the output of hard classification, boundary between classes is constrained to lie between pixels. However, with the pixels having a mixed class composition, the boundary between classes will run through the area represented by a pixel. Soft classification techniques allow for the partial and multiple class membership within each mixed pixel, and, therefore, may be used to refine the standard mapping process as well as increase the accuracy of shoreline mapping from remote sensing. This section, therefore, presents the methods used to derive the shoreline from soft classification and the approach to assess the accuracy of predicted shoreline.

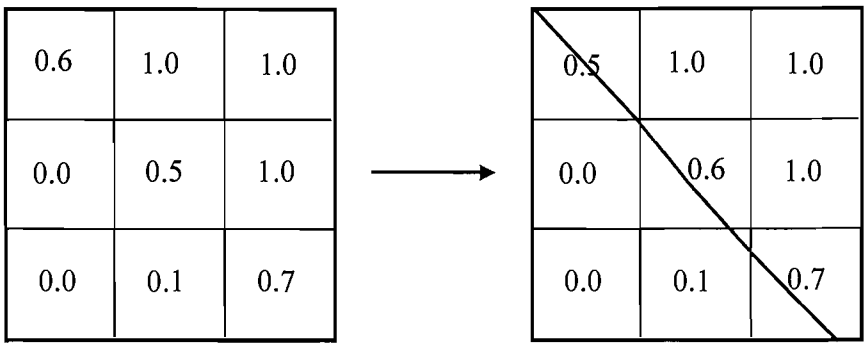
5.3.1 Methods for mapping the shoreline from soft classification output

The output from a soft classification is typically a set of proportion images with pixel values representing the proportion of a certain class within pixels. Here, a soft classification depicting the two classes, land and water, was derived using conventional LMM. The output of the soft classification for each pixel was an indication of the relative membership to the two classes. This output, however, does not indicate where within the pixels these classes are located, information that is required to fit a class boundary at the sub-pixel scale. To address this issue, the approaches for fitting the shoreline to a soft classification derived from remotely sensed imagery were investigated. In this study, two approaches, contouring and Hopfield Neural Network (HNN), were undertaken.

5.3.1.1 Contouring soft classification

Outputs derived from soft classification are class membership values representing the proportion of a class within each image pixel. Using these proportion values a shoreline is identified by fitting the output of the soft classification a contour of 0.5 membership to the land class, representing the 50% membership to land and 50% membership to water scenario. This provides an approach of positioning the shoreline location within a pixel. This method has been applied previously to allocating the waterline (Foody *et al.*, 2005) and coastal features (Foody, 2002) at a sub-pixel scale.

The shoreline located from the contour fitted to the soft classification output may be more accurate than that from hard classification and the implementation of the approach is simple and quick. A draw back of the contour-based method to shoreline mapping, however, is that the class proportional information provided by the soft classification is not maintained in fitting the contour. This means the shoreline derived from contouring soft classification using the class compositional information provided by the soft classification but the proportions either side of the shoreline derived may not match those shown in the soft classification. This is due to the generalization process related to fitting the contour. An example of the contouring process was shown in Figure 5.3.



(a) (b)

Figure 5.3: Contouring process (a) proportion image of soft classification; (b) contouring output.

5.3.1.2 Hopfield neural network

Hopfield neural network (HNN) was briefly discussed in section 2.9.1. Using the class proportion images derived from a soft classification as the input, the HNN is implemented using some parameters (Equation 2.76) which should be carefully chosen by the user. They are four weighting constants: k_1, k_2, k_3 and k_4, a zoom factor z and the number of iterations for the performance of the network.

The output of the HNN approach is a set of binary images with a spatial resolution that is z times finer than that of the input class proportional images derived from soft classification (Figure 5.4). In the analyses of this chapter the HNN was undertaken with a zoom factor of 10. With that zoom factor applied for the input proportion images with spatial resolution of 300 m, the HNN produced the output maps with a spatial resolution of 30 m which was equal to the spatial resolution of the reference image. The number of the binary images is equal to the number of land cover classes to be mapped with each image is shown the location of a defined class. In this study, with the purpose of shoreline mapping, the remotely sensed imagery used was mapped to two land cover classes water and land. The binary images derived from the HNN approach were then vectorised along the boundary between the land and water classes to generate the shoreline.

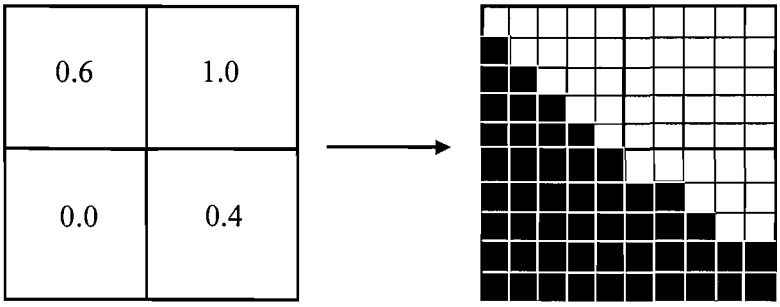


Figure 5.4: HNN process (a) a proportion image of soft classification; (b) a binary image derived from HNN.

The shoreline determined from the binary images of the HNN based approach may be more accurate than that from hard classification and the output of the HNN maintains the class proportions per pixel derived from soft classification. However, in comparison with the contouring based approach for shoreline mapping, HNN is more complicated since the user need to carefully choose the suited parameters for the network and it is a time consuming approach.

5.3.2 Accuracy assessment for shoreline mapping

The shorelines predicted from two methods (e.g., contouring and HNN) using the soft classification output derived from the degraded coarse spatial resolution imagery were compared with the reference shoreline derived from the classification of the original Landsat ETM+ imagery. To achieve an estimate of the accuracy of the shoreline, the perpendicular distance between the predicted and actual location of the shoreline was measured at each 10 metres point along the shoreline and its RMSE was calculated. The length of the shoreline in this study area was 41.68 km. As a result, the closeness of predicted shoreline to actual shoreline indicated by the average distance between them and by RMSE.

5.4 Impacts of class spectral variability on shoreline mapping

5.4.1 Training and testing data

For the purposes of this study, each pixel in the original Land ETM+ image was assumed to be pure and classified into two land cover classes, water and land, using a supervised classifier, the MLC. In the absence of accurate ground survey data, these classification data were considered as the ground/reference data.

The class proportion of pixels in the coarse spatial resolution image were defined from the corresponding area of the land cover map derived from the classification of original Landsat ETM+ image. The data extracted for each pixel included its DN value in three wavebands and the proportional coverage of each land cover class.

The training data consisted of 180 randomly selected pure pixels with 90 pixels for each land cover class (e.g., land and water). The descriptive statistics for the two land cover classes are shown in Table 5.1. For illustrative purposes, the data were subjected to principal components analyses from three used wave-bands and the first two components (PC1 and PC2) that explained most of the variation (98.9%) in the data sets were used to display the data in feature space. The scatter plots of the training samples were displayed in this manner in Figure 5.5. From this Figure, it was apparent that the water and, in particularly, the land classes exhibited a degree of variation in the feature space. The testing set for the analysis contained 5000 randomly selected pixels.

Class	Mean			Standard deviation			Variance-covariance matrix
	Band 1	Band 2	Band 3	Band 1	Band 2	Band 3	
Land	79.92	69.10	64.76	7.71	9.56	17.42	$\begin{bmatrix} 59.45 & 71.03 & 121.12 \\ 71.03 & 91.42 & 157.33 \\ 121.12 & 157.33 & 303.38 \end{bmatrix}$
Water	77.09	53.59	37.565	7.80	8.17	5.78	$\begin{bmatrix} 60.86 & 62.35 & 44.36 \\ 62.35 & 66.72 & 46.89 \\ 44.36 & 46.89 & 33.47 \end{bmatrix}$

Table 5.1: Class descriptions for the spatially degraded imagery.

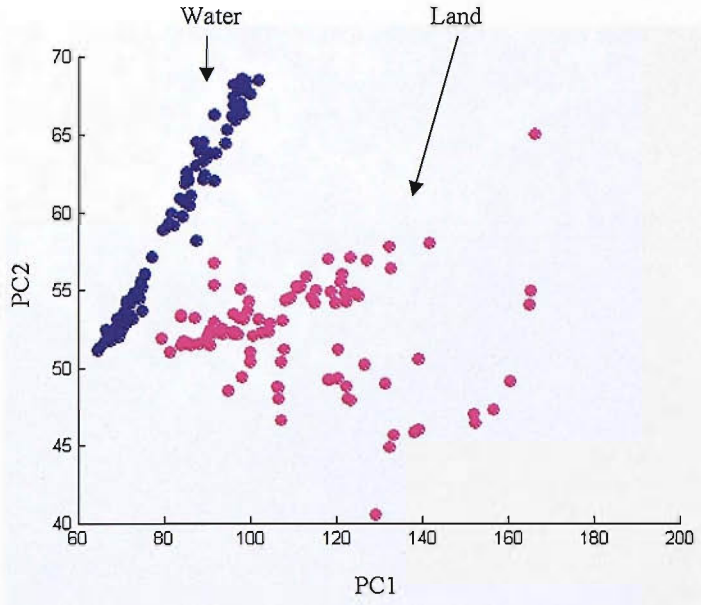


Figure 5.5: Location of the classes in feature space.

5.4.2 Sub-pixel prediction

For each pixel in the spatially degraded imagery, sub-pixel class composition estimates were derived. The training data, which included 90 pure pixels for each of water and land classes, were used to derive the spectral response of the end-members. Initially the end-members were defined as the class centroids for input to the conventional LMM. The LMM then classified each pixel in the simulated image to proportion values representing proportion of land and water within a pixel. This resulted in two proportion images with each one contained the class proportion of each class in each image pixel (Figure 5.6).

The accuracy of the sub-pixel class composition estimates was evaluated based on a comparison of predicted class proportions with that derived from the reference data, the 30 m spatial resolution image. The testing set consisted of 5000 randomly selected pixels. The accuracy was assessed using correlation coefficient and RMSE. The resulting soft classification output has a r value of 0.87 and RMSE of 0.26. The results showed that there was a strong and significant relationship between the actual and predicted class compositions (Figure 5.7) and the output of the sub-pixel class proportion estimates accurately represented the actual composition of classes within each image pixel. This was vitally important as further analysis in locating the shoreline depends on the accuracy of the sub-pixel class composition estimates.

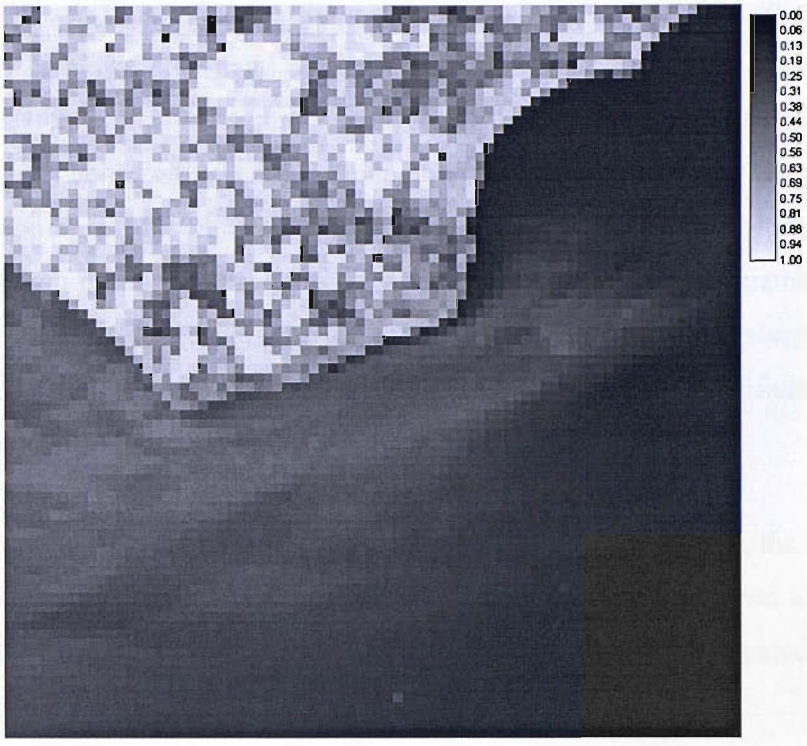


Figure 5.6: Output of soft classification. The grey level represents the proportion of land class in each image pixel.

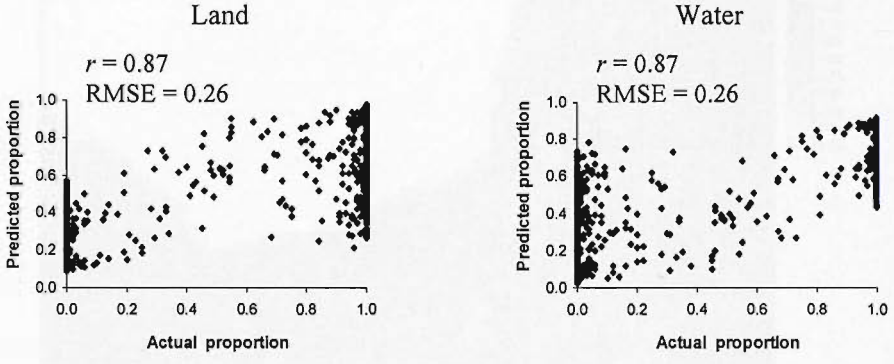


Figure 5.7: Relationships between predicted and actual cover.

5.4.3 Results of shoreline mapping

For the comparison purpose shoreline mapping from the output of both the hard and soft classifications applied to the coarse spatial simulated image was presented. Using the centroids of each class as end-members, the class composition of pixels was estimated using the LMM.

In the output of soft classification the proportion values at pure pixel (0 and 1) still contain membership of other land cover classes. This could be seen in Figure 5.5, for example, the pixels that were purely land but still have water membership in their output. This may confuse the shoreline mapping process. Therefore, for the purpose of shoreline mapping in this chapter, pure pixels of land or water were located to the value of 1 in their proportion image while maintaining proportion information at pixels adjacent to the shoreline (Figure 5.8). The accuracy of the predicted shorelines was assessed by comparing them with the shoreline derived from the classification of the original Landsat ETM+ image with 30 m spatial resolution.

To evaluate the potential of the soft classification for shoreline mapping, the traditional method of producing a shoreline map from a hard classification output was also carried out. Predicted shoreline derived from hard classification was then compared with that derived from soft classification.

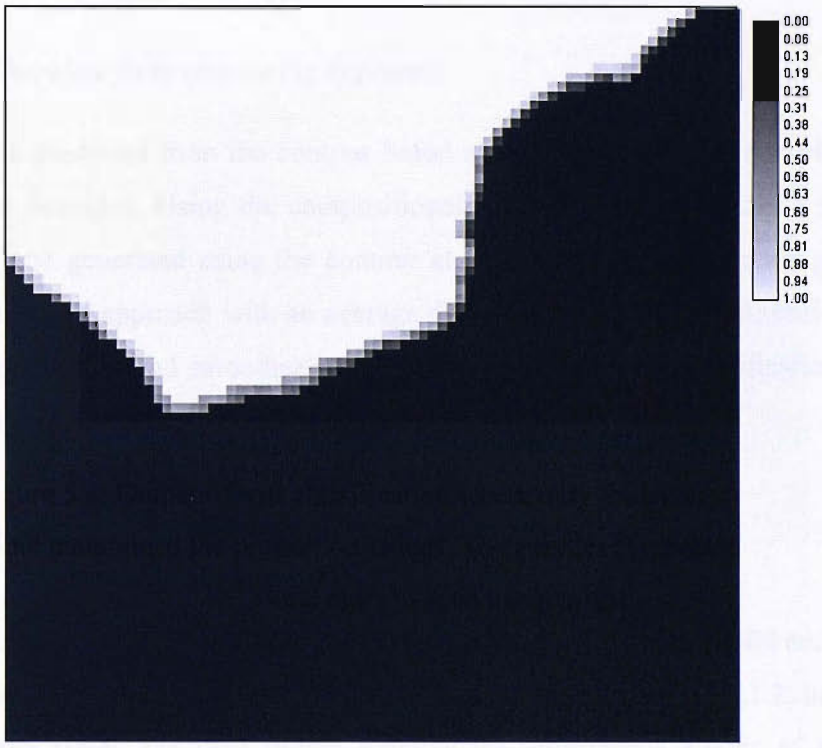


Figure 5.8: Output of soft classification where only the pixels located along the shoreline maintained the proportion values. The grey level represents the proportion of land class in each image pixel.

5.4.3.1 Shoreline from hard classification

The hard classification of the simulated image was derived by degrading the soft classification output of conventional LMM in which the class label for the pixel was the dominant class in that pixel. This was classified into two classes (land and water). Shoreline was fitted between pixels allocated to the different classes, resulting in an unrealistically jagged boundary (Figure 5.10a). The average distance between the actual shoreline and 'hard' shoreline derived from hard classification was 81.11 m and the RMSE calculated along the shoreline was 71.38 m (Table 5.3).

5.4.3.2 Shoreline from soft classification

The shorelines were derived using two approaches (e.g., contouring and HNN) based on the output of the sub-pixel class compositions. They were more realistic and accurate representation than that derived from hard classification (Table 5.3)

5.4.3.2.1 Shoreline from contouring approach

The shoreline predicted from the contour based approach fitted to the sub-pixel class compositions provided. Using the compositional values within each pixel a shoreline location may be generated using the contour at 0.5 interval. The shoreline generated from contour based approach with an average distance of 41.91 m and RMSE of 38.57 m was more accurate and smoother than that derived from hard classification (Table 5.3 and Figure 5.10(a-b)).

5.4.3.3 Shoreline from HNN approach

The sub-pixel class composition estimates were used to initialise the HNN and provide area estimates for the proportion constraint. As discussed in section 5.3.1.2, in order to implement the HNN, the user should identify the appropriate values of two goal function constraints, proportion constraint and multi-class weighting constraint, k_1, k_2, k_3 and k_4 respectively, the zoom factor z, and the number of iterations for the network. In this analysis, the network was run with the number of iterations equaled to 5000 and with a zoom factor of 10. With that zoom factor applied for the input

proportion images with spatial resolution of 300 m, the HNN was produced the output maps with a spatial resolution of 30 m which was equal to the spatial resolution of reference image.

The values of the goal and weighting constraints estimation were derived via certain assumptions and multiple network trial runs. According to Tatem *et al.* (2001, 2002), these weighting constants should be equal and the optimal values were found to be of 150. Several trial networks were run with different values of the weighting constants (Table 5.2), the zoom factor of 10, and number of iterations of 5000. The greatest accuracy of shoreline mapping was obtained with the weighting constants of $k_1 = k_2 = k_3 = k_4 = 70$ and they were used for the HNN in the analyses of this chapter.

After running the HNN using proportion images derived from soft classification and the above defined parameters, a super-resolution map was produced (Figure 5.9) with spatial resolution 10 times finer than that of the input class proportions in Figure 5.8. The shoreline boundary was derived from this super-resolution map using the raster to vector conversion approach (Figure 5.10c). This HNN based approach increased the shoreline mapping accuracy and provided a shoreline with an RMSE of 41.13 m.

Weighting constants				**Accuracy of shoreline mapping**	
k_1	k_2	k_3	k_4	Average distance (m)	RMSE (m)
70	70	70	70	45.00	41.13
100	100	100	100	53.10	44.92
120	120	120	120	47.39	43.08
150	150	150	150	47.50	43.37
170	170	170	170	47.06	42.90
200	200	200	200	47.31	43.12

Table 5.2: Accuracy of shoreline mapping derived from trial HNN (k_1, k_2, k_3, k_4 are the two goal functions, proportion, and multi- class weighting constants, respectively).

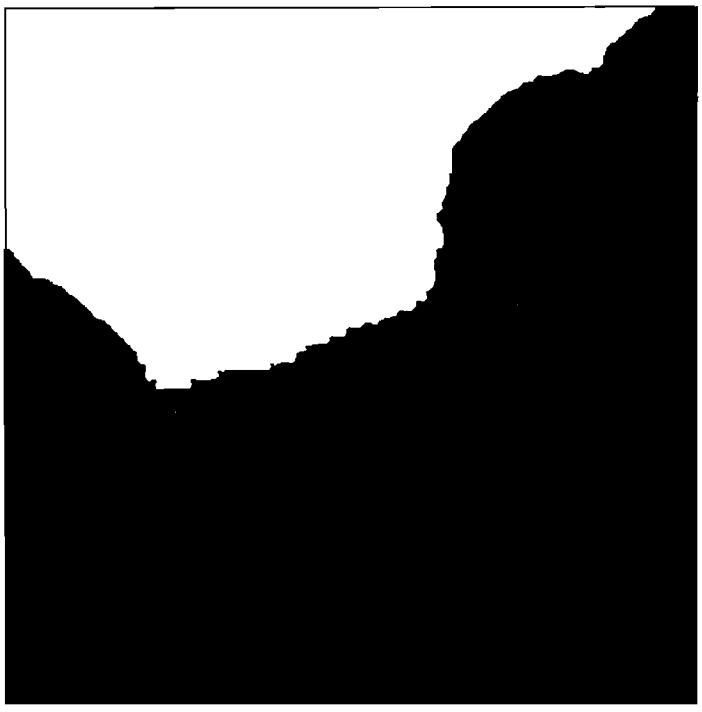


Figure 5.9: Super-resolution land cover map derived from HNN. White pixels represent land class and dark pixels represent water class.

5.4.3.4 Evaluation of methods for shoreline mapping

From Table 5.3, it was apparent that shoreline derived from contouring approach was the most accurate with the average distance of 41.91 m and RMSE of 38.57 m, followed by HNN approach derived shoreline with the average distance of 45.00 m and RMSE of 41.13 m. The shoreline derived from hard classification was the least accurate. This was produced with the average distance of 81.11 m and the RMSE of 71.38 m.

The shoreline generated from contouring approach was visually smoothest one among three shorelines derived from three methods used. A disadvantage of this method was that it modified the class proportion values when fitting the contour. The shoreline derived from HNN approach was less smooth than that derived from the contour based approach. Hard classification generated a shoreline which was jagged compared with those generated by soft classification. The advantage of the HNN against the contouring approach was that it was very dependent on the initial soft classification output as the class proportions in each pixel were maintained. This explained why in certain portion sharps peaks were visible (Figure 5.10c).

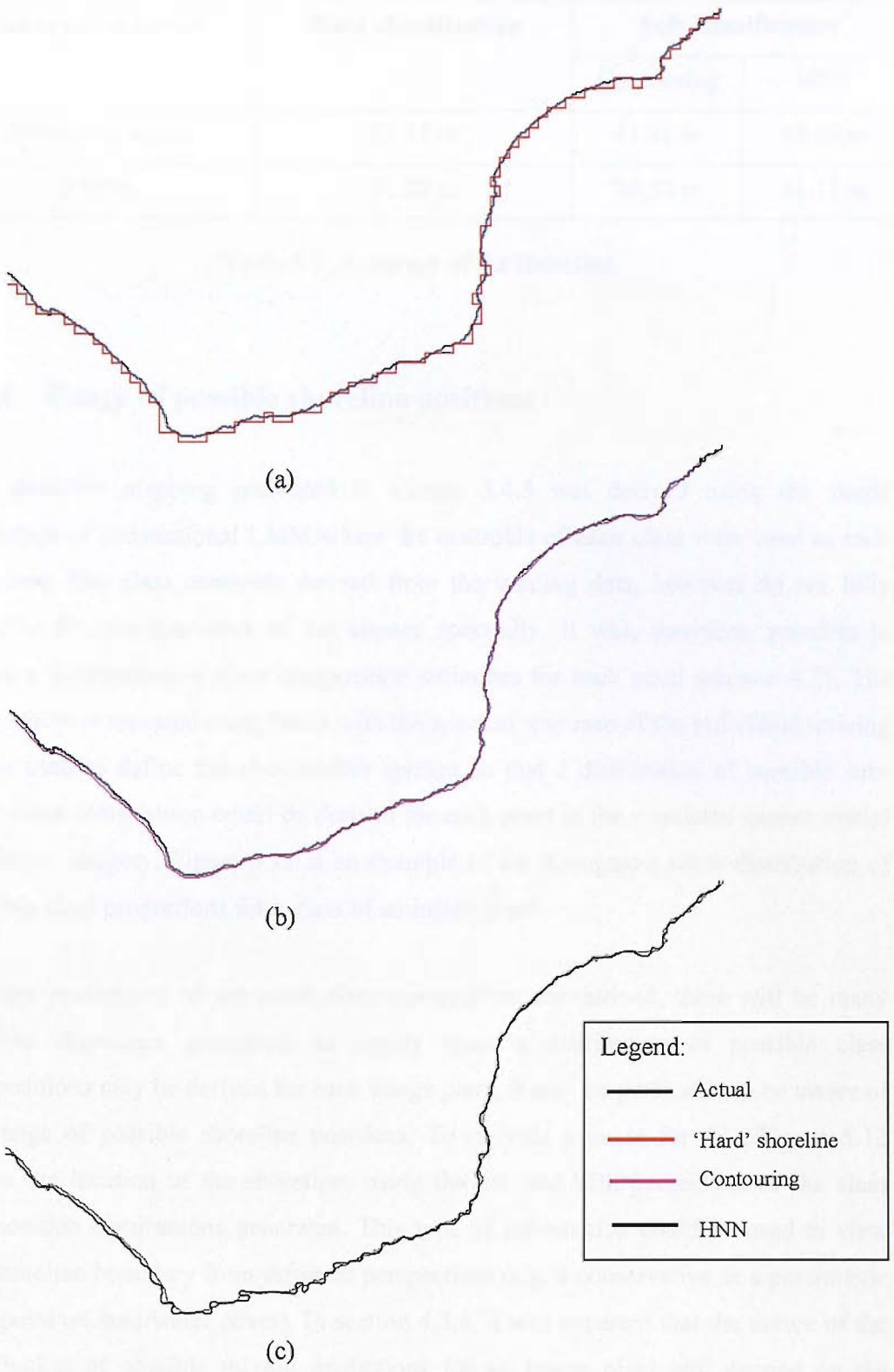


Figure 5.10: Results of shoreline mapping (a) Actual shoreline and 'hard' shoreline derived from hard classification; (b) Actual shoreline and 'soft' shoreline derived from soft classification using contouring approach; (c) Actual shoreline and 'soft' shoreline derived from super-resolution mapping using HNN.

Accuracy measure	Hard classification	Soft classification	
		Contouring	HNN
Average distance	81.11 m	41.91 m	45.00 m
RMSE	71.38 m	38.57 m	41.13 m

Table 5.3: Accuracy of the shoreline.

5.4.4 Range of possible shoreline positions

The shoreline mapping presented in section 5.4.3 was derived using the single prediction of conventional LMM where the centroids of each class were used as end-members. The class centroids derived from the training data, however do not fully describe the characteristics of the classes spectrally. It was, therefore, possible to derive a distribution of class composition estimates for each pixel (section 4.3). The analyses were repeated many times with the spectral response of the individual training pixels used to define the end-member spectra so that a distribution of possible sub-pixel class composition could be derived for each pixel in the simulated coarse spatial resolution imagery. Figure 5.11 is an example of the histograms show distribution of possible class proportions for a class of an image pixel.

If many predictions of sub-pixel class composition are derived, there will be many possible shorelines generated. In reality since a distribution of possible class compositions may be derived for each image pixel, it may be preferable to be aware of the range of possible shoreline positions. To provide a guide for this, Figure 5.12 shows the location of the shorelines using the 5th and 95th percentiles of the class composition distributions generated. This type of information could be used to view the shoreline boundary from different perspectives (e.g. a conservative or a pessimistic viewpoint on land/water cover). In section 4.3.4, it was apparent that the nature of the distribution of possible mixing predictions for an image pixel will depend on the location of the point in feature space and the degree of intra-class variation and class co-variation. This may impact on the width of the zone of possible shoreline locations, bounded by the 5th and 95th percentiles of land coverage.

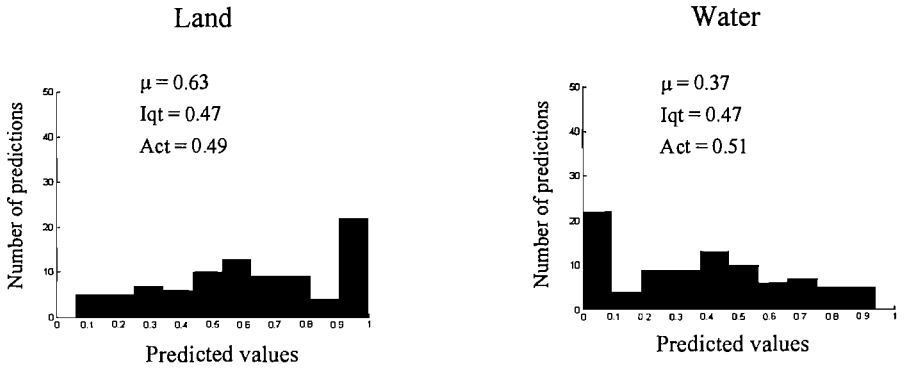


Figure 5.11: Histograms show distribution of possible class proportions within a pixel (AC – actual value; μ: the mean value of the distribution; Iqt – Interquartile range of the distribution).

The width of the possible shoreline positions varied along the study area. In particular, Figure 5.13 shows part of the study area where the width of the shoreline was much larger than the other part. The average distance between the zone of possible shoreline positions were shown in Table 5.4. This information may be used as the means to measure the effect of intra-class variation on shoreline mapping. This range of possible shoreline positions may show the confidence of the shoreline mapped from one single prediction soft classification. The larger the width of the zone of possible shoreline positions the less confidence the shoreline mapped from one single prediction of soft classification. It was suggested that the trust in the single set of class proportion predictions as input for super-resolution mapping may be unwise and the distribution of possible predictions may be used to provide a richer interpretation for this process. Furthermore, if the width of the uncertainty of the distribution is large, there may be a problem for the applications of using soft classification such as change detection (section 4.4) and super-resolution mapping here. The next section, therefore, was focused on the approach to reduce the impact of intra-class spectral variability on sub-pixel class composition estimate as well as on shoreline mapping.

Contouring	HNN
311.29 m	283.65 m

Table 5.4: Average distance between the shorelines using the 5th and 95th percentiles of the class composition distribution.

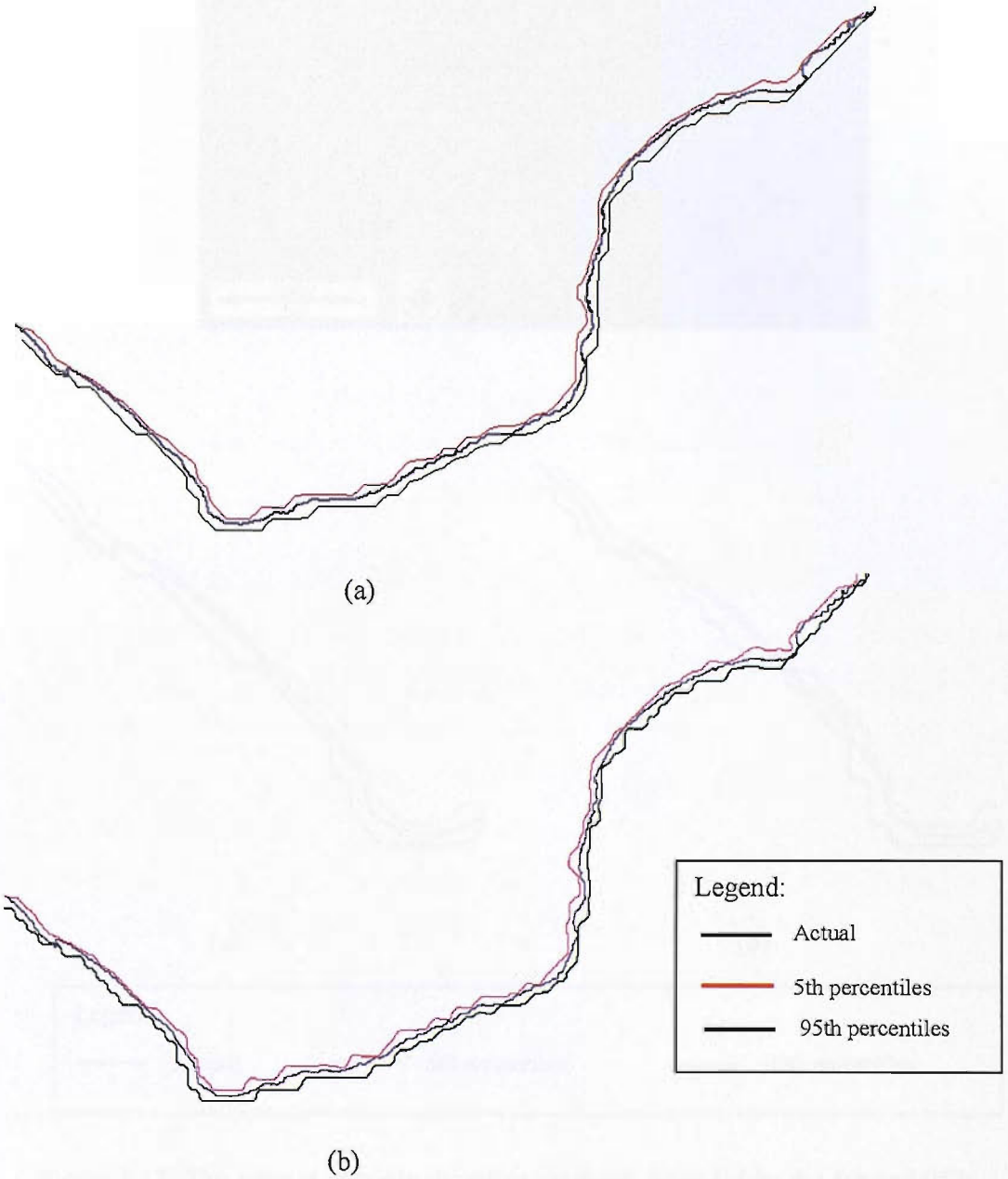
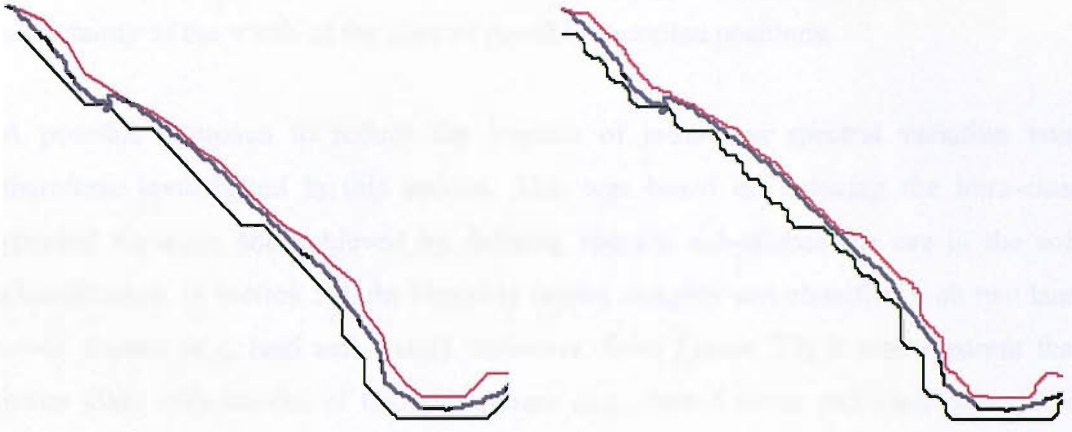


Figure 5.12: The zone of possible shoreline locations, bounded by the 5th and 95th percentiles (a) contouring approach; (b) HNN.



(a) (b)

Legend:

Actual; 5th percentiles; 95th percentiles

Figure 5.13: The zone of possible shoreline locations, bounded by the 5th and 95th percentiles in the area highlighted in the ETM+ imagery (a) contouring approach; (b) HNN.

5.5 Reducing the impacts of intra-class spectral variability on shoreline mapping.

The analyses detailed in section 4.3 and 5.4 demonstrated that the nature of the intra-class spectral variation has negative impacts on the accuracy of the sub-pixel class composition estimates as well as the accuracy of shoreline mapping. This violated the assumption that a class can be represented by a single spectrum end-member. It was suggested that a distribution of possible class compositions be derived from pixels instead of a single class composition prediction. The nature of this distribution will depend on the location of the point in feature space and the degree of intra-class variation and class co-variation (section 4.3). Using this distribution as the output of soft classification, a range of possible shoreline positions would be derived, highlighting a limitation for super-resolution mapping when using a single set of class proportion predictions of soft classification. It may suggest that reducing the degree of intra-class spectral variation and class co-variation may lead to the increase of soft classification accuracy as well as shoreline mapping accuracy and reduce the uncertainty of the width of the zone of possible shoreline positions.

A possible approach to reduce the impacts of intra-class spectral variation was, therefore, investigated in this section. This was based on reducing the intra-class spectral variation and achieved by defining spectral sub-classes for use in the soft classification. In section 5.4, the remotely sensed imagery was classified with two land cover classes (e.g, land and water). However, from Figure 5.1, it was apparent that water class may consist of two sub-classes (e.g., turbid water and clear water). To illustrate the method, the water class was divided into two sub-classes, based on relative turbidity. In this study, the LMM with real ETM+ data was performed with three end-members, land, turbid water and clear water as an example and these could readily identified from the images (Figure 5.1). Hereafter this analysis is called three-class analysis and the analysis in section 5.3 called two-class analysis.

To assess the potential of the approach that reducing intra-class spectral variation by increasing the number of classes in the analysis, the results derived from the three-class analysis and two-class analysis were compared. The study in this section aims:

1. To increase the accuracy of sub-pixel classification;

2. To increase the accuracy of shoreline mapping;

3. To reduce the width of the zone of possible shoreline positions.

5.5.1 Training and testing data

Each pixel in the original ETM+ image was assumed to be pure and classified visually into three land cover classes (e.g., land, turbid water and clear water) using a supervised classifier, the MLC (Figure 5.14). In the absence of accurate ground survey data, these classification data were considered as the ground/reference data.

Sub-pixel class compositions within pixels in the coarse spatial resolution imagery, which were used to derive training and testing data, were defined from the corresponding area of the land cover map derived from the classification of original Landsat ETM+ imagery. The data extracted for each pixel included its DN value in three wavebands and the proportional coverage of each land cover class.

The training data consisted of 270 randomly selected pure pixels with 90 pixels for each land cover class (e.g., land, turbid water and clear water). The description statistics of the three land cover classes was shown in Table 5.5 and displayed in Figure 5.15. It was apparent from Figures 5.4 and 5.14 that the spectral variation of turbid water and clear water classes were narrower than that of water class in the two-class analysis presented in section 5.4.

The testing set which contained 5000 randomly selected pixels was the same with that in the two-class analysis.

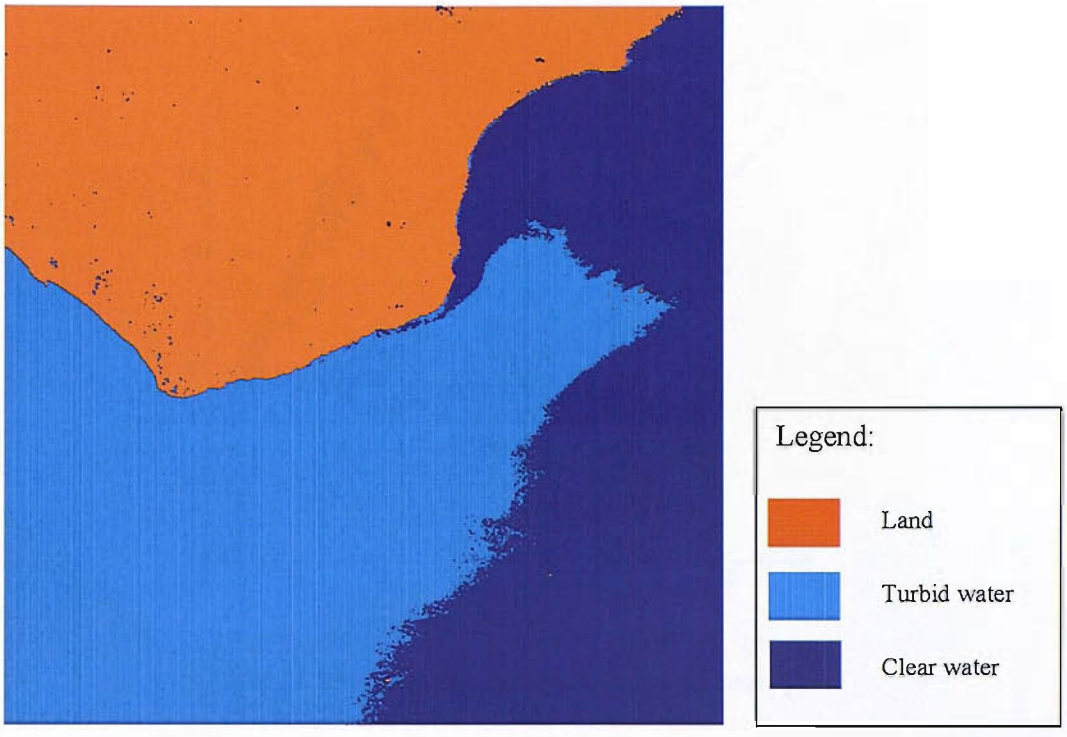


Figure 5.14: 30 m spatial resolution classified imagery with three classes, land, turbid water and clear water that was used as ground data in analysis.

Class	Mean			Standard deviation			Variance-covariance matrix
	Band 1	Band 2	Band 3	Band 1	Band 2	Band 3	
Land	81.56	70.81	67.52	9.13	10.93	19.22	$\begin{bmatrix} 83.29 & 97.28 & 165.57 \\ 97.28 & 119.56 & 203.92 \\ 165.57 & 203.92 & 369.50 \end{bmatrix}$
Turbid water	82.26	59.28	41.49	4.68	4.86	3.62	$\begin{bmatrix} 21.88 & 21.12 & 15.84 \\ 21.12 & 23.62 & 17.20 \\ 15.84 & 17.20 & 13.07 \end{bmatrix}$
Clear water	68.60	44.56	31.42	1.49	1.38	1.28	$\begin{bmatrix} 2.22 & 1.58 & 1.74 \\ 1.58 & 1.91 & 1.46 \\ 1.74 & 1.46 & 1.65 \end{bmatrix}$

Table 5.5: Class descriptions of the degraded imagery.

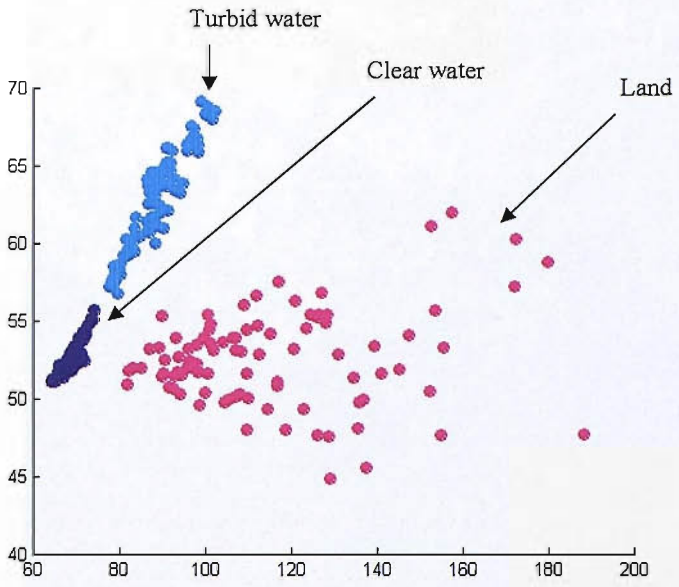
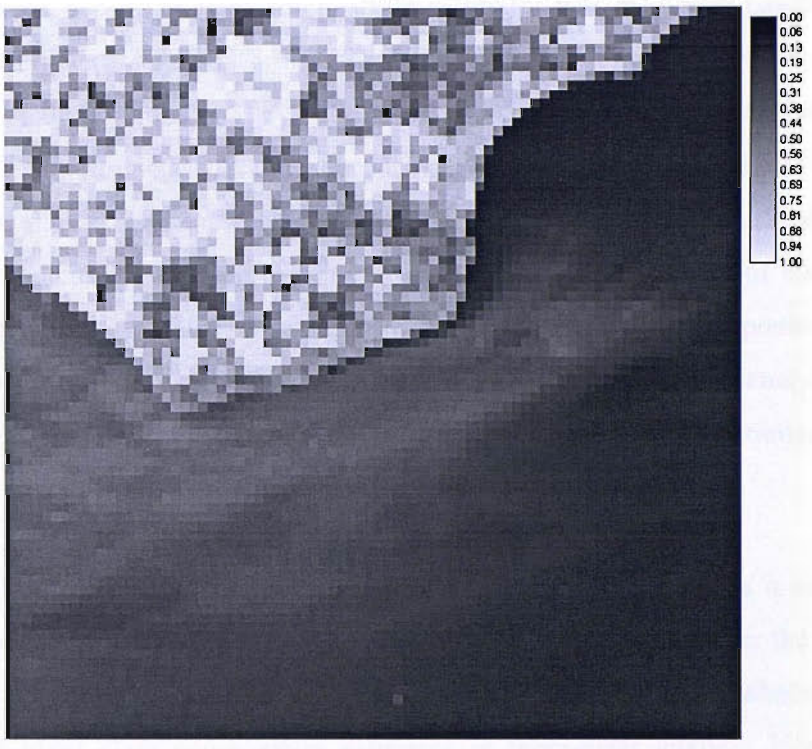


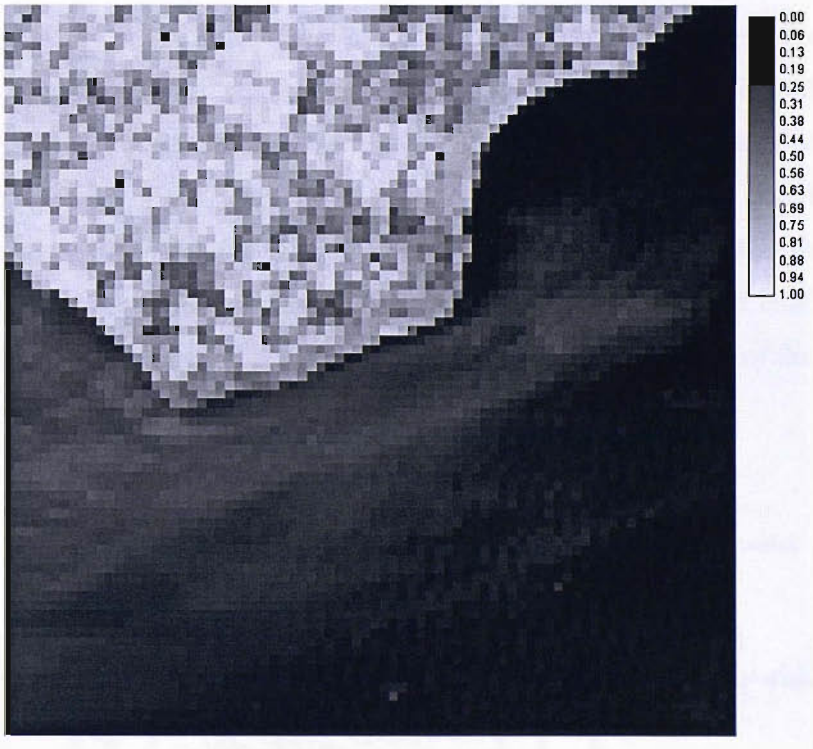
Figure 5.15: Location of the classes in feature space.

5.5.2 Sub-pixel class composition predictions

For each pixel in the spatially degraded imagery (Figure 5.2), sub-pixel class composition estimates were derived. The training data which included 90 pure pixels for each of land, turbid water and clear water classes was used to derive the spectral response of the end-members. The spectral response of the classes in the image are summarised in Table 5.5. Initially the end-members were defined as the class centroids for input to the conventional LMM. The LMM then classified each pixel in the simulated image to proportion values representing proportion of land, turbid water and clear water within a pixel. This resulted in three proportion images with each one contained the class proportion of each class in the image pixels. Figure 5.16 shows the proportion images of land class derived from the soft classification of both two-class analysis and three-class analysis.



(a)



(b)

Figure 5.16: Output of soft classification. The grey level represents the proportion of land class in each image pixel. (a) two-class analysis scenario; (b) three-class analysis-scenario.

The accuracy of the sub-pixel class composition estimates was evaluated based on a comparison of predicted class proportions with that derived from the reference data, the 30 m spatial resolution image. They were assessed using correlation coefficient and RMSE. The resulting soft classification output has a r value of 0.94 and RMSE of 0.20. The results showed that there was a strong and significant relationship between the actual and predicted class compositions (Figure 5.17) and the output of the sub-pixel class proportion estimates accurately represented the actual composition of classes within each image pixel. This was vitally important as further analysis in locating the shoreline depends on the accuracy of the sub-pixel class composition estimates.

In section 4.3.3, it was shown that the nature of intra-class variation has a marked impact on the accuracy of sub-pixel class composition estimation; the larger the intra-class variability, the less accurate the sub-pixel classification. To assess whether the accuracy of sub-pixel class composition estimates in three-class analysis increased compared with two-class analysis, Table 5.6 illustrates the accuracy (e.g., correlation coefficient and RMSE) of the sub-pixel class composition predictions from the two-class and three-class analyses. It was apparent that the accuracy of the predictions from three-class analysis (e.g., r = 0.94 and RMSE = 0.20) was much higher than that from two-class analysis (e.g., r = 0.87 and RMSE = 0.27). The two correlation coefficients were statistically significant at 99% level of confidence. It was suggested that the reduction of the intra-class spectral variability by increasing the number of land cover classes increased soft classification accuracy so that it reduced the impacts of the intra-class variation on sub-pixel class composition estimates.

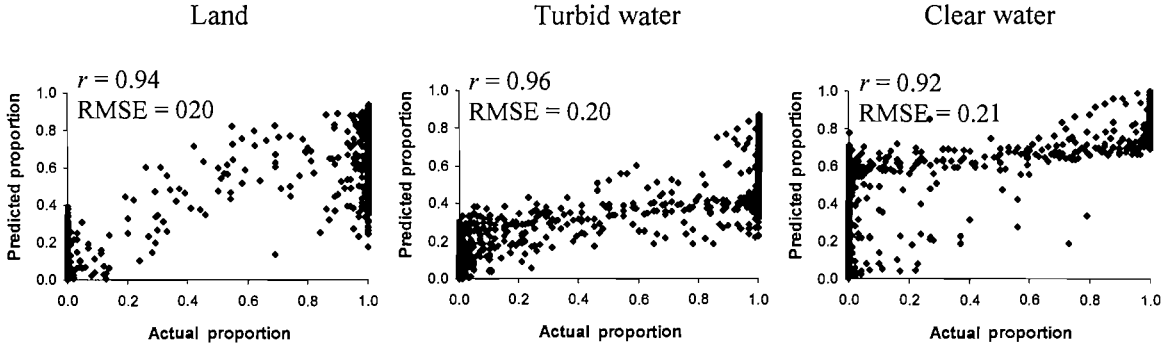


Figure 5.17: Relationships between predicted and actual cover.

Methodology	Class	r	RMSE
Two-class analysis	Land	0.87	0.27
	Water	0.87	0.27
	Average	0.87	0.27
Three-class analysis	Land	0.94	0.20
	Turbid water	0.96	0.20
	Clear water	0.92	0.21
	Average	0.94	0.20

Table 5.6: Accuracy of the sub-pixel class composition estimates.

5.5.3 Shoreline mapping

The assessments of shoreline mapping from two-class and three-class analyses were presented to evaluate the potential of the reducing intra-class spectral variation approach. The output of soft classification used to generate shoreline mapping here was derived from conventional LMM in which the centroids of each class were used as the end-members. The shoreline was generated based on two approaches (e.g., contouring and HNN) using the output of the sub-pixel class composition estimates.

The reduction of the intra-class variation increased the accuracy of soft classification (section 5.5.2). In this study, shoreline position was determined from the output of soft classification so that its accuracy may depend on the accuracy of the soft classification.

Figure 5.18 and 5.19 show the shoreline positions derived from the output of soft classification of both the two-class and three-class analyses using contouring and HNN based approaches. For the comparison purpose, Table 5.7 presents the accuracy of the shoreline mapping from these two analyses. The accuracy of the shoreline predictions was assessed by comparing them with the shoreline predicted from the classification of the original Landsat ETM+ image with 30 m spatial resolution.

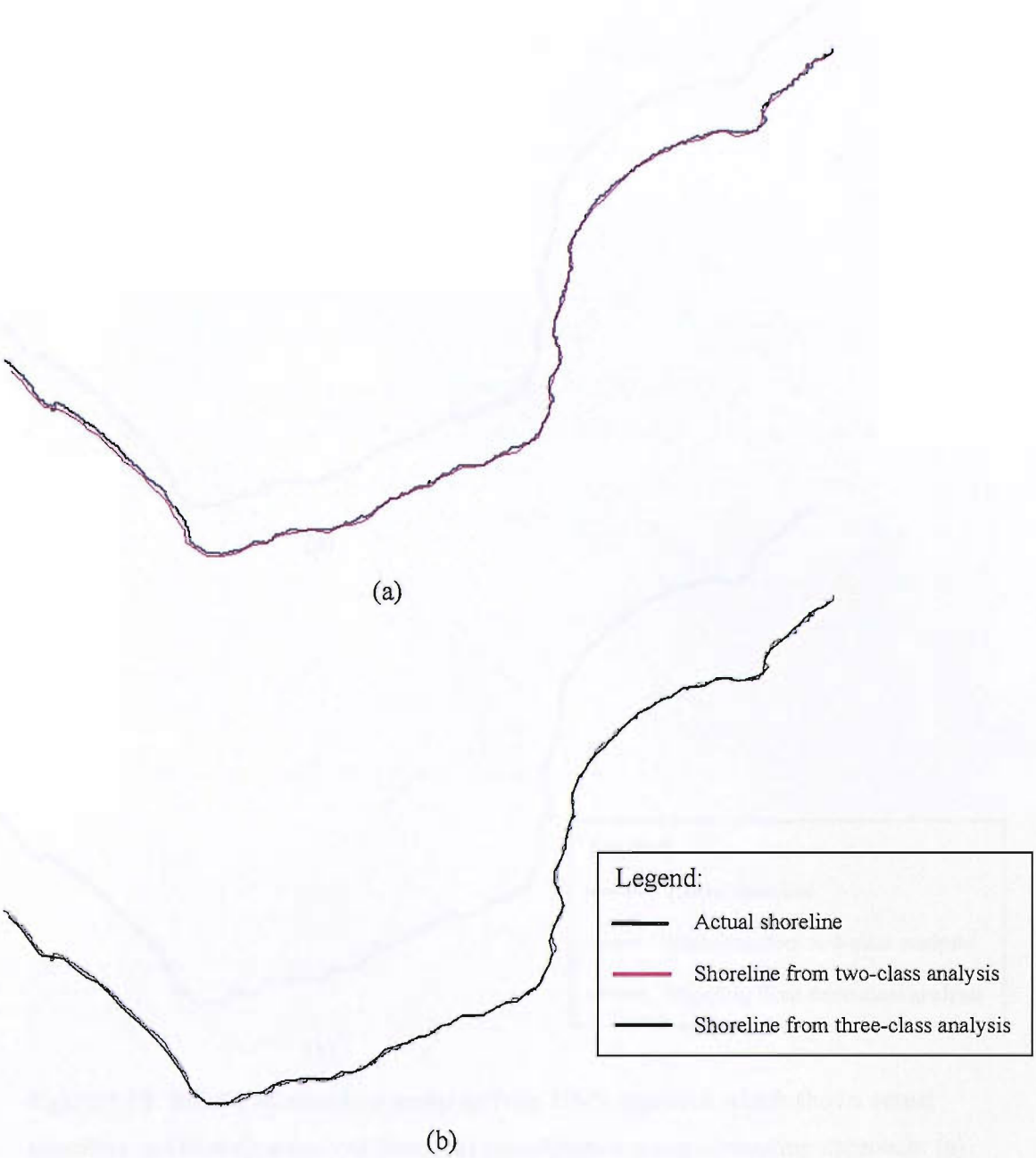


Figure 5.18: Results of shoreline mapping from contouring approach which shows actual shoreline and shoreline derived from soft classification using contouring approach: (a) two-class analysis; (b) three-class analysis.

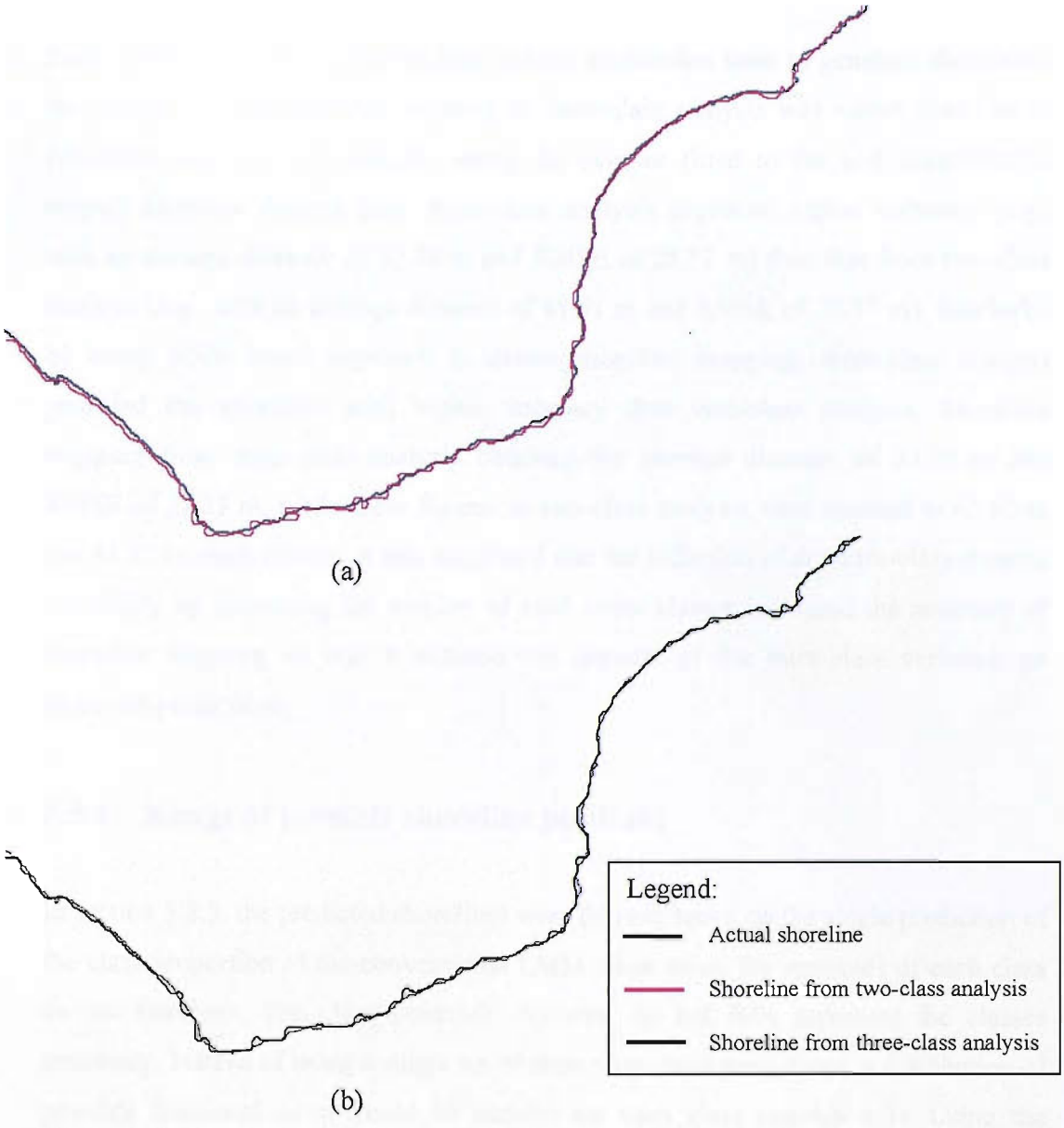


Figure 5.19: Results of shoreline mapping from HNN approach which shows actual shoreline and shoreline derived from soft classification using contouring approach: (a) two-class analysis; (b) three-class analysis.

Methodology	**Contouring**		**HNN**	
	Average distance (m)	RMSE (m)	Average distance (m)	RMSE (m)
Two class analysis	41.91	38.57	45.00	41.13
Three class analysis	32.76	28.12	39.76	37.21

Table 5.7: Accuracy of the shoreline predictions.

From Table 5.7, it was apparent that in both approaches used to generate shorelines, the accuracy of the shoreline mapping in three-class analysis was higher than that in two-class analysis. For example, using the contour fitted to the soft classification output, shoreline derived from three-class analysis produced higher accuracy (e.g., with an average distance of 32.76 m and RMSE of 28.12 m) than that from two-class analysis (e.g., with an average distance of 41.91 m and RMSE of 38.57 m). Similarly, by using HNN based approach to derive shoreline mapping, three-class analysis provided the shoreline with higher accuracy than two-class analysis. Shoreline mapping from three-class analysis obtained the average distance of 39.76 m and RMSE of 37.21 m, while these figures in two-class analysis were reached to 45.00 m and 41.13 m, respectively. It was suggested that the reduction of the intra-class spectral variability by increasing the number of land cover classes increased the accuracy of shoreline mapping so that it reduced the impacts of the intra-class variation on shoreline predictions.

5.5.4 Range of possible shoreline positions

In section 5.5.3, the predicted shorelines were derived based on the single prediction of the class proportion of the conventional LMM when using the centroids of each class as end-members. The class centroids, however do not fully represent the classes spectrally. Instead of using a single set of class proportion predictions, a distribution of possible fractional cover could be derived for each class (section 4.3). Using the spectral response of the individual training pixels to define the end-member spectra as the input for the LMM, a distribution of possible sub-pixel class composition could be derived for each pixel in the simulated coarse spatial resolution imagery. Figure 5.20 is an example of the histograms which show distributions of possible class proportions within a defined image pixel of two analyses, two-class and three-class. From Figure 5.20, it was apparent that within the same image pixel the widths of the distributions of possible class proportions derived from three-class analysis were correspondingly smaller than those from two-class analysis.

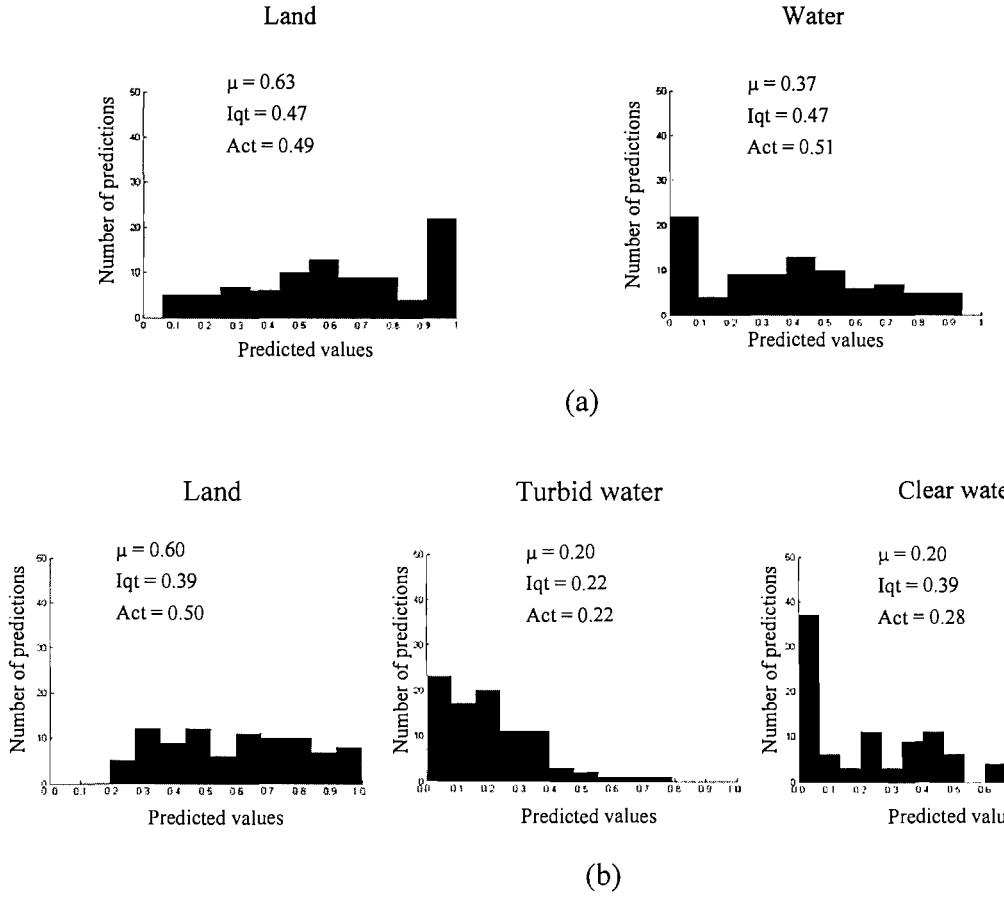


Figure 5.20: Histograms show distribution of possible class proportions within a pixel: (a) two-class analysis; (b) three-class analysis (AC – actual value; μ: the mean value of the distribution; Iqt – Interquartile range of the distribution).

In reality since a distribution of possible class compositions may be derived for each image pixel, it may be preferable to be aware of the range of possible shoreline positions (section 5.4.4). To provide a guide for this, the location of the shorelines using the 5th and 95th percentiles of the class composition distribution generated may be derived (Figure 5.21 and 5.22). The width of the zone of possible shoreline locations was bounded by the 5th and 95th percentiles of land coverage.

To assess whether the width of the zone of possible shoreline positions in three-class analysis was reduced compared with that in two-class analysis, the average distance between the locations of the shorelines using the 5th and 95th percentiles of the class composition distribution generated were calculated (Table 5.8).

According to Table 5.8, the average ranges of possible shoreline positions derived from both contouring and HNN based approaches in three-class analysis was smaller than that from two-class analysis. For example, using HNN based approach the width of the zone of possible shoreline locations, bounded by the 5th and 95th percentiles of land coverage, was narrower, 215.82m, from the three-class analysis than from two-class analysis, 283.65m.

The results highlighted that reducing the intra-class class spectral variability may reduce the width of the possible shoreline positions in shoreline mapping so that this may increase the confidence of shoreline mapping using one single prediction of soft classification.

Methodology	**Contouring**	**HNN**
Two class analysis	311.29 m	283.65 m
Three class analysis	250.13 m	215.82 m

Table 5.8: Average distance between the shorelines using the 5th and 95th percentiles of the class composition distribution.

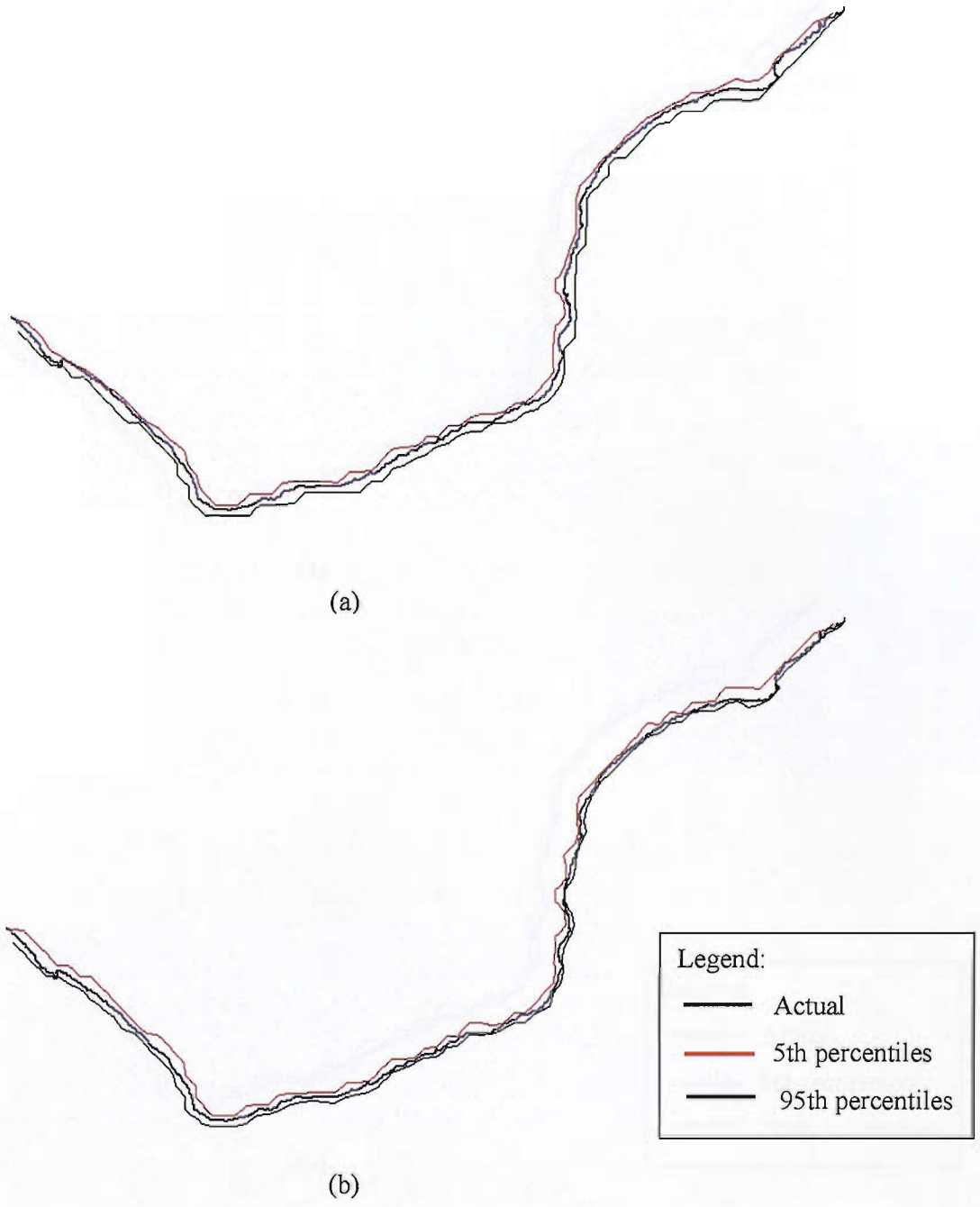


Figure 5.21: The zone of possible shoreline locations, derived based on the contouring approach, bounded by the 5th and 95th percentiles of the distributions (a) two-class analysis; (b) three-class analysis.

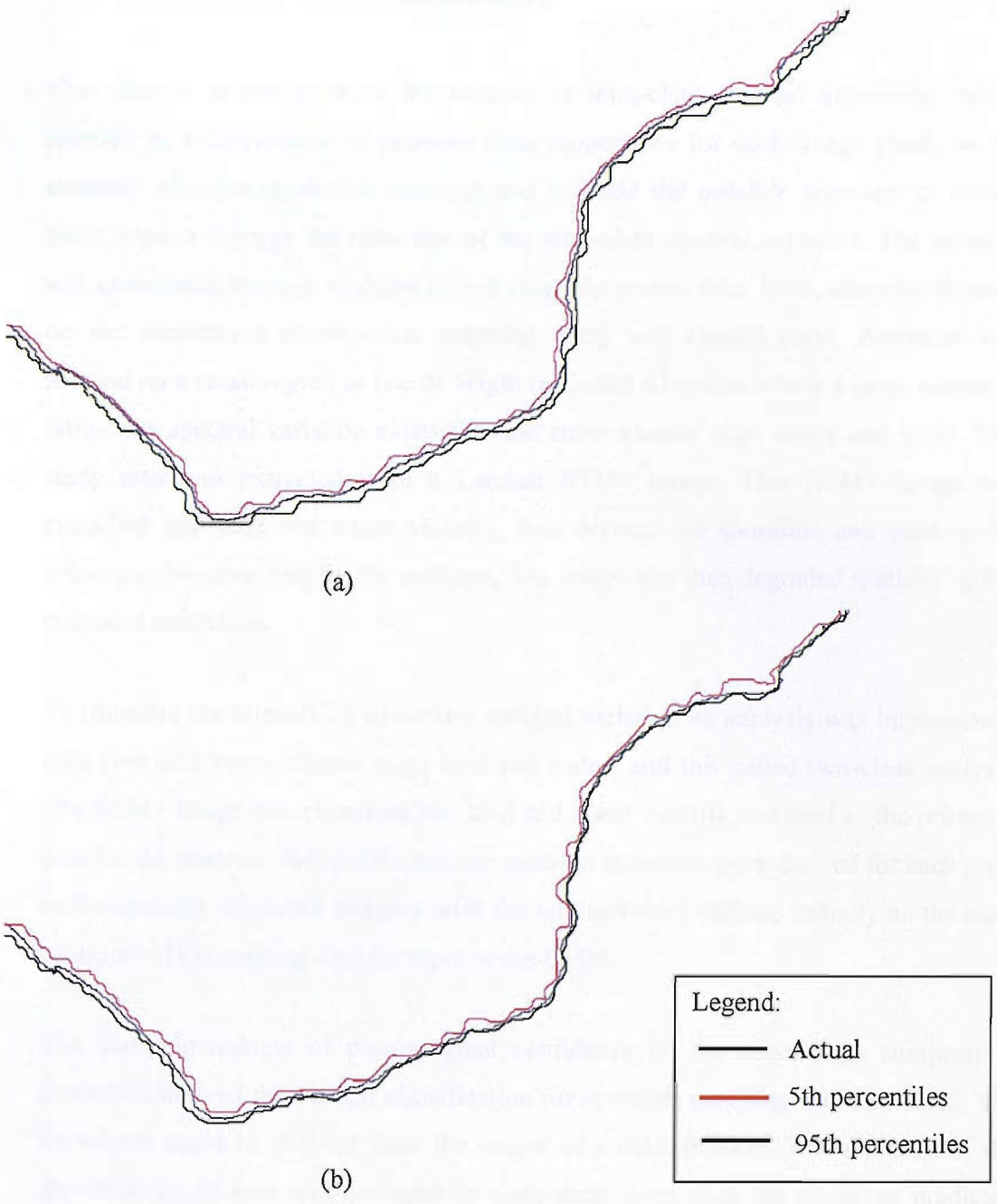


Figure 5.22: The zone of possible shoreline locations, derived based on the HNN approach, bounded by the 5th and 95th percentiles of the distributions (a) two-class analysis; (b) three-class analysis.

5.6 Summary and conclusions

This chapter aimed to study the impacts of intra-class spectral variability, which resulted in a distribution of possible class proportions for each image pixel, on the accuracy of super-resolution mapping and to study the possible approach to reduce these impacts through the reduction of the intra-class spectral variation. The research was undertaken through analyses of real remotely sensed data. Here, attention focused on the assessment of shoreline mapping using soft classification. Attention was focused on a small region in Isle of Wight in United Kingdom where a large degree of intra-class spectral variation exists for land cover classes (e.g., water and land). This study area was extracted from a Landsat ETM+ image. This ETM+ image was classified into land and water visually, then derived the shoreline and used as the reference shoreline data for the analyses. The image was then degraded spatially to 300 m spatial resolution.

To illustrate the impacts of intra-class spectral variation an analysis was implemented with two land cover classes (e.g., land and water) and this called two-class analysis. The ETM+ image was classified into land and water visually and used as the reference data for the analysis. Sub-pixel class composition estimates were derived for each pixel in the spatially degraded imagery with the end-members defined initially as the class centroids of the training data for input to the LMM.

The inappropriateness of placing great confidence on the singe class composition prediction derived from a soft classification for shoreline mapping was illustrated. The shorelines could be defined from the output of a classification. The accuracy of the shoreline predictions was assessed by comparing them with the shoreline predicted from the classification of the original Landsat ETM+ image with 30 m spatial resolution. For a hard classification, the shoreline was resulted in an unrealistically jagged boundary and achieved with the least accuracy (e.g., average distance of 81.11 m and the RMSE of 71.38 m). A refinement on this representation was to produce a super-resolution map based on the output of the soft classification. Here, contouring and HNN approaches were used to derive the super-resolution map and these were used to derive shoreline. Shoreline derived from contouring approach was visually the smoothest shoreline and the most accurate with the average distance of 41.91 m and

RMSE of 38.57, followed by HNN approach derived shoreline with the average distance of 45.00 m and RMSE of 41.13.

The shoreline mapping presented was derived using the single prediction of conventional LMM where the centroids of each class were used as end-members. The class centroids however do not fully describe the characteristics of the classes spectrally as classes typically display a degree of spectral variability. It was, therefore, possible to derive a distribution of class composition estimates for each pixel. Since there were a range of predictions of sub-pixel class composition derived, there would be a range of shoreline locations generated. To provide a guide for this, a range of possible shoreline positions determined using the 5th and 95th percentiles of distribution generated was derived. The width of the possible shoreline positions was varied along the study area. This type of information could be used to view the effect of intra-class variation on shoreline mapping highlighting a limitation for super-resolution mapping as the trust in the single set of class proportion predictions.

Reducing the intra-class spectral variation may increase soft classification accuracy as well as shoreline prediction accuracy and reduce the range of possible positions of shoreline was the purpose of the second analysis in the chapter. This second analysis was implemented with three land cover classes and called three-class analysis in which water class was divided into two classes (e.g., turbid water and clear water classes) and this led to the reduction of the intra-class variation of water class compared with two-class analysis. The potential of the approach that reducing intra-class spectral variation was assessed by comparing the outputs of the sub-pixel estimation, shoreline mapping and range of possible shoreline positions from three-class analysis with those from two-class analysis.

In terms of the sub-pixel class composition estimates, the accuracy of the predictions from three-class analysis (e.g., $r = 0.94$ and RMSE = 0.20) was higher than that from two-class analysis (e.g., $r = 0.87$ and RMSE = 0.26). The two correlation coefficients derived from these two analyses were statistically significant at 99% level of confidence. In terms of shoreline mapping, with both approaches used to generate shorelines, the accuracy of the shoreline in three-class analysis was higher than that in two-class analysis. For example, using the contour fitted to the soft classification

output, shoreline derived from three-class analysis produced RMSE of 28.12 while that from two-class analysis achieved the RMSE of 38.57. Similarly, by using HNN based approach to derive shoreline mapping, three-class analysis provided the shoreline with RMSE of 37.21 m, while these figures in two-class analysis were reached to 41.13 m.

Furthermore the width of the zone of possible shoreline locations, bounded by the 5th and 95th percentiles of the distributions of sub-pixel class composition estimates, was narrower from the three-class analysis than from two-class analysis. For example, using the contouring approach to derived shorelines, the average distance between the possible shorelines in three-class analysis was 250.13 m, while this number in two-class analysis reached to 311.29 m. Similarly, using HNN based approach it this average distance was 283.65 m in three-class analysis and 215.82 m in two-class analysis.

In conclusion, this chapter has shown that intra-class spectral variation has negative impacts on the accuracy of sub-pixel class composition estimates and super-resolution mapping. This based on the assumption that the classes may be described by a single end-member. However, class spectral variability means that single spectrum may not adequately describe a class and therefore a distribution of possible class composition estimates should be derived. This resulted in a range of possible shoreline positions instead of just one single shoreline. Reducing the intra-class class spectral variability was a possible approach to reduce the negative impacts of intra-class spectral variability on these procedures. This increased soft classification and shoreline mapping accuracy and reduce the range of possible shoreline positions.

Chapter 6

Summary and conclusions

In this final chapter, an attempt is made to summarise the main results of the thesis and to discuss the conclusions that arise from the research findings. Possible future directions for research are suggested finally.

6.1 Summary

As presented in chapter 1, remote sensing has been used commonly as a source of information on land cover and land cover change. Unfortunately, the accuracy of both land cover mapping and monitoring activities is often limited by the presence of mixed pixels, which can make the single class allocation derived from a conventional hard classification inappropriate and inaccurate. Sub-pixel land cover classification (or soft classification) techniques allow for the partial and multiple class membership within each mixed pixel, and, therefore, may be used to refine the standard mapping process as well as increase the accuracy of land cover mapping from remote sensing (Wang 1990, Foody and Cox 1994, Tso and Mather 2001). The output of a soft classification is typically a set of fraction images that show the predicted coverage of each thematic class in the area represented by each pixel. Although soft classifications can provide a more accurate and appropriate representation of themes such as land cover than a conventional hard classification there are many concerns. In particular, the accuracy of soft classifications is often low. The main aims of the thesis were, therefore, to investigate the possible approaches to increase the accuracy of soft land cover classification and study one of the factors that may impact on soft classification prediction, the intra-class spectral variability, and highlight its implications for analyses based on soft classification such as the detection of land cover change and super-resolution mapping. Finally an approach to reduce the impacts of intra-class spectral variability on soft classification and its implications for analyses based on soft classification was explored.

As a first step, the literature was studied to explore the potential of remote sensing to land cover mapping. Chapter 2 reviewed how remote sensing and related techniques

may be used to provide the means to map land cover. Consequently, a brief review of the issues involved in land cover mapping from remotely sensed imagery was provided. The problem of mixed pixels was presented and its solution through the use of soft classification was discussed. Specifically, a number of methods for sub-pixel class composition estimates were presented with some discussion of the approaches of the accuracy assessment. Although soft classifications estimate the sub-pixel class compositions within each pixel they do not indicate where the land cover classes are spatially located within the pixels. The sub-pixel class fractions may, however, be located geographically through super-resolution mapping. A brief discussion of super-resolution mapping using soft classification output was, therefore, presented. The chapter closed with the conclusion that although soft classifications can provide a more accurate and appropriate representation of themes such as land cover than a conventional hard classification, the application and development of such techniques require further research including the ways to increase soft classification accuracy, the impacts of intra-class spectral variability on sub-pixel class composition estimates and its implications to the real world applications from remotely sensed data.

The possible approaches to increasing the accuracy of soft classification through the use of an ensemble of classifiers were investigated in chapter 3. A variety of different outputs can be derived from the application of a suite of classifiers applied to the same data set. The classification output derived may differ greatly in accuracy, on both a per-class and overall basis. By combining the outputs of a set of classifiers it is possible to derive a classification that is more accurate than any of the individual classifications used. The use of such an ensemble of classifiers has been shown to be able to increase the accuracy of hard classification but its potential for soft classification has rarely been demonstrated. Here, four methods for combining soft classifications to increase soft classification accuracy were assessed. These methods were based on (i) the selection of the most accurate predictions on a class-specific basis, (ii) the average of the outputs of the individual classifications for each case, (iii) the direct combination of classifications using evidential reasoning and (iv) the adaptation of the outputs to enable the use of a conventional (hard classification) ensemble approach. The potential of these approaches was assessed using coarse spatial resolution NOAA AVHRR imagery of Australia.

NOAA AVHRR data were classified using two neural networks (a multi-layer perceptron and a radial basis function network) as well as a probabilistic classifier.

Overall classification accuracy was assessed using a generalized cross-tabulation matrix approach and showed that in terms of the overall accuracy, the classifications were broadly similar with an accuracy of about 61% (section 3.5.2.1). Accuracy was also expressed on a per-class basis using the correlation coefficient between the predicted and actual (ground data) coverage. Based on the accuracy of the individual soft classifications, the classification outputs were combined using the four ensemble approaches proposed. The results highlighted that all four approaches to combining the soft classification outputs have been shown to be able to significantly increase classification accuracy. The increases in overall accuracy derived from these ensemble approaches ranged from 2.20% to 4.45% and large increases in individual class accuracy were also observed. Moreover, the differences in overall accuracy between the ensemble approaches and the most accurate individual classification were statistically significant at at least the 95% level of confidence.

Although the ensemble of soft classifications can increase soft classification accuracy, the accuracy of land cover classification derived from soft classifications themselves as well as from their ensemble was still often low. This was clearly a major limitation to studies seeking to estimate class fractional cover and its change over time or the sub-pixel distribution of classes as in super-resolution mapping. Chapter 4, therefore, aimed to evaluate one major cause of error, intra-class spectral variability, and explore its impacts on the accuracy of soft classification prediction and highlight its implications for analyses based on soft classification such as the detection of land cover change.

The impacts of intra-class spectral variability on the accuracy of soft classification were assessed using simulated data, so as to control variables impacting on the remotely sensed response. The data were classified using LMM. A series of analyses were undertaken in which the degree of intra-class variation was altered. Initially, the class centroids were taken to define the class end-member spectra and the intra-class variation was, therefore, ignored. In later analyses, a multitude of LMMs were applied. In this, every pixel in the training set was used to provide the end-member spectrum for the analysis. By unmixing the spectral response of a pixel many times with different end-members a series of sub-pixel class composition estimates could be derived for a pixel of any given spectral response. As a result of this it was possible to form a distribution of sub-pixel estimates for each pixel. The nature of this distribution depended on the location of the point in feature space and the degree of intra-class variation and class co-

variation. The results indicated that intra-class spectral variability was an important factor that affects the estimates of sub-pixel class proportions. It has been shown that the accuracy of sub-pixel class composition estimation was a function of the degree and nature of intra-class spectral variation (section 4.3).

The impacts of intra-class variation on the use of soft classification for change detection were assessed. This analysis focussed on the detection of deforestation in part of Para, Brazil from Landsat TM data. Specifically, two images separated by a four year period in which major forest clearance activities had been undertaken were used. The forest and, in particular the more heterogeneous, the non-forest classes exhibited a degree of variation in feature space. The class composition of pixels in each image was estimated using the LMM. Comparison of the estimates derived from the two time periods could then be derived in order to an estimate the sub-pixel scale change in class composition in time.

Distributions of class composition estimates for each pixel in two time periods were derived. Comparison of these distributions may result in a different interpretation to that derived through comparison of the single class composition estimate derived from a standard sub-pixel analysis. This was illustrated with reference to several different examples drawn from the study area. The direct comparison of the single predictions from a standard soft classification analysis yielded a single estimate of deforestation. The apparent precision of the estimated amount of change could be dangerous (section 4.4.3.2). Moreover, the distributions of class composition estimates provided a richer description of the class composition that may allow the change to be viewed from different perspectives. For example, the danger in using the single prediction from a standard application of the LMM could be recognised and the distributions used to indicate change. This could be taken from a range of perspectives. So rather than directly compare single predictions, one could, for instance, focus on the upper and lower quartiles of the distributions to derive what could be considered by some to be a relatively 'good' and 'bad' case scenario of change. This provided a useful extension and qualification to the standard use of single prediction estimates.

Results in chapter 3 showed that soft classification accuracy may increase through the use of an ensemble of individual soft classifications. This was implemented using a single class composition prediction for image pixels in the individual classifications.

The study in chapter 4 (section 4.5) was, therefore, to combine different soft classifications in which the distribution of class composition estimates was used as the output for the individual classifications. Different distributions of possible class composition estimates for each class within each image pixel were derived from different set of end-members using as the input for LMM. Pure pixels derived from reference data and from PPI approaches were used as two different training data sets to derive end-members. As a result, two different distributions of possible sub-pixel proportion estimates of a class within each image pixel were derived. They were then combined to get the final combined distribution. The accuracy of the sub-pixel class composition estimates derived from the ensemble approach (r = 0.94 and RMSE = 0.17) was slightly lower than that from the PPI approach (r = 0.95 and RMSE = 0.15), but it slightly increased compared with the accuracy derived from the reference data approach (r = 0.94 and RMSE = 0.18). However, the distribution of sub-pixel class composition estimates in the ensemble approach was less spread than that from the two individual sub-pixel classifications. It was suggested that although ensemble based approach may not increase sub-pixel classification accuracy compared with the most accurate individual classification, it reduced the variety of possible sub-pixel class composition estimates so that may increase the confidence of standard sub-pixel analysis.

Due to the impacts of intra-class spectral variability on the soft classification prediction, chapter 5 aimed to study the impacts of intra-class spectral variability on super-resolution mapping using soft classification as the input and investigate the possible approach to reduce these impacts through the reduction of the intra-class spectral variation. The impacts of intra-class spectral variation on the use of soft classification outputs in super-resolution mapping were assessed. This work focussed on mapping part of the shoreline of the Isle of Wight, UK, from Landsat ETM+ data. There was a large degree of intra-class spectral variation for the land cover classes to be mapped. A soft classification of the remotely sensed imagery was derived with two land cover classes, water and land, using a LMM. This soft classification was then used to derive a super-resolution map of the shoreline through the use of a Hopfield neural network and contouring approaches. It was evident that the accuracy of the soft classification as well as of the shoreline mapping declined as the degree of intra-class variation increased. Since a distribution of possible class compositions may be derived for each pixel, it may be preferable to be aware of the range of possible shoreline positions. To provide a guide to this, a range of the possible shoreline positions was determined and was

illustrated in the analysis using the 5th and 95th percentiles of the class composition distributions generated.

A possible approach to reduce the impacts of intra-class spectral variation was investigated. This was based on the reduction of the intra-class spectral variation of the land cover classes used for sub-pixel classification and achieved by defining spectral sub-classes for use in the soft classification. To illustrate the method, the water class was divided into two sub-classes, based on relative turbidity (section 5.5). A three-class soft classification of the Landsat ETM+ image was derived with three land cover classes, turbid water, clear water and land, using a LMM. The accuracy of the soft classification was assessed using correlation coefficient between the predicted and actual (ground data) coverage. The accuracy of the shoreline was evaluated using RMSE based on the predicted and actual locations of the shorelines. The range of possible shoreline positions was indicated by the distance between the shoreline represented by the 5th and 95th percentiles of the class composition generated. The three-class analysis provided higher accuracies of the soft classification and shoreline mapping and narrower range of possible shoreline positions than those from the original two-class analysis. Soft classification from three-class analysis was achieved with the correlation coefficient of 0.94, while that from two-class analysis was 0.87. These two correlation coefficients were statistically significant at 99% level of confidence. Similarly, shoreline mapping from three-class analysis was derived with the RMSE of 37.21m, while this figure in two-class analysis was reached to 41.13m. Furthermore, the width of the zone of possible shoreline locations, bounded by the 5th and 95th percentiles of land coverage, was narrower, 215.81m, from the three-class analysis than from two-class analysis, 283.65m.

The results from chapter 5 highlighted that reducing intra-class spectral variation may be used to increase soft classification accuracy so that to increase shoreline mapping accuracy and reduce the range of possible shoreline positions.

6.2 Conclusions

A number of different soft classification outputs may be delivered from soft classifiers, which are different in terms of their methodology, applied to the same data set. These classification outputs may differ greatly in accuracy, on both a per-class and overall

basis. Classifiers that make different errors may sometimes usefully be combined to form an ensemble approach to classification. It was possible to result in a final classification that was often more accurate than the most accurate individual classification from the combination of the outputs of a group of different classifiers. Results from the analyses highlighted that ensemble based approaches may be used to significantly increase classification accuracy.

The reliance on the single set of class proportion predictions derived from a standard soft classification for the analyses such as change detection and super-resolution mapping may be unwise. This trust was assumed that a single spectral end-member can be used to represent a land cover class to be mapped. This is likely to be unrealistic as classes typically display a degree of spectral variability. The degree of intra-class spectral variation impacts negatively on the accuracy of sub-pixel class composition estimates as it opposes to the assumption that a class can be represented by a single spectral end-member. A consequence of this is that, a distribution of possible class composition estimates should be derived from pixels instead of a single class composition prediction. The width of the distribution depends on the location of the pixel in the feature space and the degree of intra-class variation and class co-variation. This distribution provided a richer indication of sub-pixel class composition estimates and may be used to derive different scenarios of deforestation through the use of a post-classification comparison technique.

Furthermore, using soft classification outputs as the input for super-resolution mapping process to derive shoreline map, the accuracy of shoreline mapping reduced when the degree of intra-class variation increased. Since a distribution of possible class compositions may be derived for each pixel, it may be preferable to be aware of the range of possible shoreline positions. This range of possible shoreline positions may show the confidence of the shoreline mapped from one single prediction of standard soft classification. The larger the width of the zone of possible shoreline positions the less confidence the shoreline mapped from one single prediction of soft classification. It was suggested that the trust in the single set of class proportion predictions as input for super-resolution mapping may be unwise and the distribution of possible predictions may be used to provide a richer interpretation for this process.

One possible approach of reducing the negative impacts of the intra-class spectral variation on the accuracy of soft classification was to reduce the degree of intra-class spectral variation. Specifically, this increased the accuracy of shoreline mapping and reduced the width of the zone of possible shoreline positions.

6.3 Future work

In chapter 3, some possible approaches to increase the accuracy of soft land cover classification though the use of an ensemble of classifiers have been investigated. Although the ensemble of soft classifications can increase soft classification accuracy, the accuracy of land cover classification derived from soft classifications themselves as well as from their ensemble was still often low. Further research addressing this issue need to be undertaken to be able to increase soft classification accracy. This may be the optimisation of the training data characteristics for soft classification to increase classification accuracy. Different training data sets with different class components will be evaluated to find the one that may achieve the highest accuracy for the soft classification algorithm used.

In chapters 4 and 5, LMM was the only convenient tool used to derive the sub-pixel scale information on class composition. By unmixing the spectral response of a pixel many times with different end-members a series of sub-pixel class composition estimates could be derived for a pixel of any given spectral response. As a result of this it was possible to form a distribution of sub-pixel estimates for each pixel. The implications of the distributions derived for the analyses such as change detection and super-resolution have been illustrated. It was suggested that the distribution of possible predictions could be used to provide a richer interpretation of sub-pixel cover and change. It, therefore, would be worth to explore other soft classifiion approaches which can derive distribution information of the sub-pixel predictions for each pixel.

In chapter 5, shoreline mapping from soft classification output was studied. The results highlighted that by using a soft classification to a coarse spatial resolution imagery to estimate the class composition within a pixel and applying super-resolution mapping to locate classes geographically, an accurate prediction of the shoreline position may be derived. However, the accuracy of the shoreline was varied along the shoreline in the study area. A possible approach may be a new research direction is to derive a soft

classification based on local training data to increase soft classification accuracy and ultimately the accuracy of shoreline predictions (Muslim *et al.*, 2006).

In chapter 5, the impacts of the intra-class spectral variability on soft classification in general and on shoreline mapping in particular were reduced by decreasing the intra-class spectral variation of the land cover classes used. In particular, attention focused on the division of a land cover class into sub-classes used for soft classification. However, due to the limited number of spectral wave-bands used in the remotely sensed imagery, the number of end-members is also limited, it may be impossible to increase the number of end-members. Further research may be worth to explore other possible approaches that can reduce the impacts of intra-class spectral variability on the accuracy of sub-pixel classification.

REFERENCES

Adams, J. B., Sabol, D. E., Kapos, V., Filho, R. A., Roberts, D. A., Smith, M. O., and Gillespie, A. R., 1995, Classification of multispectral images based on fractions of end-members: Application to land-cover change in Brazilian Amazon, *Remote Sensing of Environment*, **52**, 137-154.

Aleksander, I. and Morton, H., 1990, *An introduction to neural computing*, Chapman and Hall, London.

Arora, M. K. and. Foody, G. M, 1997, Log-linear modelling for the evaluation of the variables affecting the accuracy of probabilistic, fuzzy and neural network classifications, *International Journal of Remote Sensing*, **18**, 785-798.

Atkinson, P. M. and Tatnall, A. R. L., 1997, Introduction: Neural networks in remote sensing, *International Journal of Remote Sensing*, **18**, 699-709.

Atkinson, P. M., 1997, Mapping sub-pixel boundaries from remotely sensed images, *Innovation in GIS 4*, 166-180.

Atkinson, P.M., 1999, Assessing uncertainty in fuzzy land cover maps obtained by remote sensing, *Proceedings of the 4th International Conference on GeoComputation.* http://www.geocomputation.org/1999/071/gc_071.htm.

Atkinson, P. M., 2005, Super-resolution target mapping from soft-classified remotely sensed imagery, *Photogrammetric Engineering and Remote Sensing*, **71:77**, 839-846.

Atkinson, P.M., Cutler M.E.J. and Lewis H., 1997, Mapping sub-pixel proportional land cover with AVHRR imagery, *International Journal of Remote Sensing*, **18**, 917-935.

Bardossy, A., and Samaniego, L., 2002, Fuzzy rule-based classification of remotely sensed imagery, *IEEE Transactions on Geoscience and Remote Sensing*, **40**, 362-374.

Bastin, L., 1997, Comparison of fuzzy c-means classification, linear mixture model, and MLC probabilities as tools for unmixing coarse pixels, *International Journal of Remote Sensing*, **18**, 3629-3648.

Bastos, L.C., Bastos, R.C., and Nishida, W., 1999, Radial basis function for classification of remote sensing images, *Neural Networks. IJCNN '99. International Joint Conference*, **3**,

1959 – 1962.

Bateson, C.A., Asner, G.P., and Wessman, C.A., 2000, Endmember bundles: a new approach to incorporating endmember variability into spectral mixture analysis, *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 1083-1094.

Benediktsson, J.A., Swain, P.H. and Ersoy, O.K., 1990, Neural network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 540-552.

Benediktsson, J.A., Swain, P.H. and Ersoy, O.K., 1993, Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data, *International Journal of Remote Sensing*, **14**, 2883-2903.

Bernard A.C., Wilkinson G.G. and Kanellopoulos I., 1997, Training strategies for neural network soft classification of remotely-sensed imagery, *International Journal of Remote Sensing*, **18**, 1851-1856.

Bezdek J.C., and Pal S.K., 1992, *Fuzzy Models for Pattern Recognition*, the Institute of Electrical and Electronics Enginners, *Inc.*, New York.

Bezdek, J.C., Ehrlich, R. and Full, W., 1984, FCM: the fuzzy c-means clustering algorithm, *Computers and Geosciences*, **10**, 191-203.

Bezdek, J.C., Hathaway, R.J., Sabin, M.J. and Tucker, W.T., 1987, Convergence theory for fuzzy c-means counter-examples and repairs, *IEEE transactions on Systems, Man and Cybernetics*, **17**, 873-877.

Binaghi, E., Brivio, P.A., Ghezzi, P. and Rampini, A., 1999, A fuzzy set-based accuracy assessment of soft classification, *Pattern Recognition Letters*, **20**, 935-948.

Bishop, C.M., 1994, Mixture density networks, *Technical Report* NCRG94/004. Neural Computing Research Group, Aston University.

Bishop, C.M., 1995, *Neural Networks for Pattern Recognition*, Oxford, U.K.: Clarendon.

Borel, C.C. and Gerstl, S. A. W., 1994, Nonlinear spectral mixing models for vegetative and soil surfaces, *Remote Sensing of Environment,* **47**, 403-416.

Boucher, A. and Dick, W.F.L., 1951, Super-resolution land cover mapping with indicator

geostatistics, *Remote Sensing of Environment*, **104**, 264-282.

Boucher, A. and Kyriakidis, P. C., 2006, Super-resolution land cover mapping with indicator geostatistics, *Remote Sensing of Environment*, **104**, 264-282.

Breiman, L., 1996, Bagging predictors, *Machine Learning*, **24**, 123-140.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., 1984, *Classification and Regression Trees*, New York:Chapman & Hall.

Briem, G.J., Benediktsson, A., Sveinsson, J.R., 2002, Multiple classifiers applied to multisource remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, **40**, 2291-2299.

Brookes, B.C. and Dick, W.F.L., 1951, *Introduction to Statistical Method*, Heimann Educational Books LTD, London.

Brown, M., Lewis, H.G., and Gunn, S.R., 2000, Linear spectral mixture models and support vector machines for remote sensing, *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 2346-2360.

Bruzzone, L. and Prieto, D.F., 1999, A technique for the selection of Kernel-Function parameters in RBF neural networks for classification of remote-sensing images, *IEEE Transactions and Geoscience and Remote Sensing*, **37**, 1179-1184.

Bryll, R., Gutierrez-Osuna, R., and Quek, F., 2003, Attribute bagging: improving accuracy of classifiers ensembles by using random feature subsets, *Pattern Recognition*, **36**, 1291-1302.

Burt, J.E. and Barrer, G.M., 1996, *Elementary Statistics for Geographers*, Second edition, the Guilford Press.

Burzzone, L., Fernandez Prieto, D., and Cossu, R., 1999, An incremental learning classifier for remote-sensing images, *Geoscience and Remote Sensing Symposium. IGARSS '99 Proceedings. IEEE International*, **5**, 2483 – 2485.

Campbell, J. B., 2002, *Introduction to Remote Sensing,* Third edition, Taylor and Francis, London.

Campbell, J.B., 1981, Spatial correlation effects upon accuracy of supervised

classification of land cover, *Photogrammetric Engineering and Remote Sensing*, **47**, 335-357.

Cannon, R.L., Dave, J.V. and Bezdek, J.C., 1986, Efficient implementation of the fuzzy c-means clustering algorithm, *IEEE transactions on Pattern Recognition and Machine Intelligence*, **8**, 248-255.

Carlotto, M.J., 1995, Non-linear mixture model and application for enhanced resolution multispectral classification, *'Quantitative Remote Sensing for Science and Applications', International Geoscience and Remote Sensing Symposium*, **2**, 1168-1170.

Carpenter, G.A., Gopal, S., Macomber, S. and Woodcock, C.E., 1999, A neural network method for mixture estimation for vegetation mapping, *Remote Sensing of Environment*, **70**, 138-152.

Chang, C. and Plaza, A., 2006, A fast iterative algorithm for implementation of pixel purity index, *IEEE Geoscience and Remote Sensing Letters*, **3**, 63-67.

Chen, S., Samingan, A.K., and Hanzo, L., 2001, Adaptive multiuser receiver using a support vector machine technique, IEEE Transactions on Pattern Analysis and Machine Intelligence, 604-608

Cho S.B., and Kim, J.H., 1995, Combining multiple neural networks by fuzzy integral and robust classification, *IEEE Transactions on Systems, Man and Cybernetics*, **25**, 380-384.

Cihlar, J, Latifovic, R., Chen, J., Beaubien, J., Li, Z. and Magnussen, S., 2000, Selecting representative high resolution sample images for land cover studies. Part 2: application to estimating land cover composition, *Remote Sensing of Environment*, **72**, 127-138.

Congalton, R.G., 1991, A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, **37**, 35-46.

Côté, S. and Tatnall, A.R.L., 1997, The Hopfield neural network as a tool for feature tracking and recognition from satellite sensor images, *International Journal of Remote Sensing*, **18**, 871-885.

Cracknell, A.P., 1998, Synergy in remote sensing – what's in a pixel, *International Journal of Remote Sensing*, **19**, 2025-2047.

Cristianini, N., and Shawe-Taylor, J., 2000, *An Introduction to Support Vector Machines.* London, Cambridge University Press.

Curran, P.J., 1985, *Principles of Remote Sensing*, Longman, Harlow

Cutzu F., 2003, Polychotomous classification with pairwise classifiers: A new voting principle, *Lecture Notes Computer Science*, **2709**, 115-124.

Dai X. and Khorram S., 1999, Data fusion using artificial neural networks: a case study on multitemporal change analysis, *Computers, Environment and Urban Systems,* **23**, 19-31.

Denoeux, *T.*A, 1995, K-nearest neighbor classification rule based on Dempster-Shafer theory, *IEEE Transactions on Systems, Man and Cybernetics,* **25**, 804-813.

Diamond, I. and Jefferies, J., 2001, *Beginning Statistics: An Introduction for Social Scientists*, SAGE Publications, London.

Dimyati, M., Mizuno, K., Kobayashi, S. and Kitamura, T., 1996, An analysis of land use/cover change using the combination of MSS landsat and land use map – a case study in Yogyakarta, Indonesia, *International Journal of Remote Sensing,* **17**, 931-944.

Ebdon, D., 1985, *Statistics in Geography*, Second edition, Basil Blackwell, Oxford, UK.

Embashi, M.R.M., 1998, Mapping sub-pixel variation in land cover at the global scale using NOAA AVHRR imagery, *Ph.D. Thesis*, University of Southampton, Southampton.

Erol, H., 2000, A practical method for constructing the mixture model for a spectral class, *International Journal of Remote Sensing*, **21**, 823-830.

Erricker, B.C., 1971, *Advanced General Statistics*, The English Universities Press LTD.

Evensen G., 2003, The Ensemble Kalman Filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, **53**, 343-367.

Fan, J.L., Zhen W.Z., Xie, W.X., 2003, Suppressed fuzzy *c*-means clustering algorithm, *Pattern Recognition Letters*, **24**, 1607-1612.

Fisher, P.F. and Pathirana, S., 1990, The evaluation of fuzzy membership of land cover classes in the suburban zone, *Remote Sensing of Environment*, **34**, 121-132.

Fisher, P.F., 1997. The pixel: a snare and a delusion, *International Journal of Remote Sensing*, **18**: 679-685.

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**, 2-12

Foody G.M., 1996a, Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data, *International Journal of Remote Sensing*, **17**, 1317-1340.

Foody , G.M., 1996b, Relating the land cover composition of mixed pixels to artificial neural network classification output , *Photogrammetric Engineering and Remote Sensing*, **62**, 491-499.

Foody, G.M., 1997, Fully fuzzy supervised classification of land cover from remotely sensed imagery with an artificial neural network, *Neural Computing and Applications*, **5**, 238-247.

Foody, G.M., 2001a, The effect of a non-exhaustively defined set of classes on neural network classifications, *Geoscience and Remote Sensing Symposium, IGARSS '01. IEEE International*, **6**, 2731 – 2733.

Foody, G. M., 2001b. Monitoring the magnitude of land-cover change around the southern limits of the Sahara, Photogrammetric Engineering and Remote Sensing, 67: 841-847.

Foody, G.M., 2002a, Hard and soft classifications by a neural network with a non-exhaustively defined set of classes, *International Journal of Remote Sensing*, **23**, 3853-3864.

Foody, G.M., 2002b, The role of soft classification techniques in the refinement of estimates of ground control point location, *Photogrammetric Engineering and Remote Sensing*, **68**, 897-903.

Foody G.M. , 2002c, Soft thematic mapping, *Geospatial Pattern Recognition*, 169-188.

Foody G.M., 2002d, Status of land cover classification accuracy assessment, *Remote Sensing of Environment*, **80**, 185-201.

Foody, G.M. and Arora, M.K., 1996, Incorporating mixed pixels in the training, allocation and testing stages of supervised classifications, *Pattern Recognition Letters*, **17**, 1389-1398.

Foody, G.M. and Arora, M.K., 1997, An evaluation of some factors affecting the accuracy of classification by an artificial neural network, *International Journal of Remote Sensing*, **18**, 799-810.

Foody, G.M. and Cox, D.P., 1994, Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions, *International Journal of Remote Sensing*, **15**, 619-631.

Foody, G.M. and Mather, A., 2004, A relative evaluation of multiclass image classification by support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 1335-1342.

Foody, G.M., and Atkinson, P.M., 2002, *Uncertainty in Remote Sensing and GIS*, Wiley, UK.

Foody, G.M., Campbell, N.A., Trodd, N.M., and Wood, T.F., 1992, Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification, *Photogrammetric Engineering and Remote Sensing*, **58**, 1335-1341

Foody, G.M., Jackson, R.G. and Quine, C.P., 2003, Potential improvements in the characterisation of forest canopy gaps caused by windthrow using fine spatial resolution multispectral data: comparing hard and soft classification techniques, *Forest Science*, **49**, 444-454.

Foody, G.M., Lucas, R.M., Curran, P.J. and Honzak, M., 1997, Non-linear mixture modeling without end-members using an artificial neural network. *International Journal of Remote Sensing*, **18**, 937-953.

Foody, G.M., McCulloch, M.B. and Yates, W.B., 1995, Classification of remotely sensed data by an artificial neural network: issues related to training data characteristics, *Photogrammetric Engineering and Remote Sensing*, **61**, 391-401.

Foody, G.M., Muslim, A.M. and Atkinson, P.M., 2005, Super-resolution mapping of the waterline from remotely sensed data, *International Journal of Remote Sensing*, **56**, 5381-

5392.

Foschi, P.G. and Smith, D.K., 1997, Detecting sub-pixel woody vegetation in digital imagery using two artificial intelligence approaches, *Photogrammetric Engineering and Remote Sensing*, **63**, 493-500.

Friedl, M. and Broley, C., 1997, Decision tree classification of land cover from remotely sensed data, *Remote Sensing of Environment*, **61**, 399-409.

Friedl, M.A., Brodley, C.E., and Strahler, A.H., 1999, Maximizing land cover classification accuracies produced by decision tree at continental to global scales, *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 969-977.

Gader, P.D., Mohamed, M.A., and Keller, J.M., 1996, Fusion of handwritten word classifiers, *Pattern Recognition Letters*, **17**, 577-584.

Ghosh, A., Pal, N.R., Das, J., 2003, Fuzzy Rule Based Approaches for Cloud Cover Estimation Using METEOSAT 5 Images, *International Geoscience and Remote Sensing Symposium*, **6**, 3438-3440.

Giacinto, G. and Roli, F., 1997a, Ensembles of neural networks for soft classification of remote sensing images, *Proceeding of the European Symposium on Intelligent Techniques*, Bari, Italy, 166-170.

Giancinto, G., and Roli, F, 1997b, Adaptive selection of image classifiers, University of Cagliari, Italy.

Giancinto, G. and Roli, F., 2001a, An approach to the automatic design of multiple classifier systems, *Pattern Recognition Letters*, **22**, 25-33.

Giacinto, G. and Roli, F., 2001b, Design of effective neural network ensembles for image classification purposes, *Image and Vision Computing*, **19**, 699-707.

Giancinto, G. and Roli, F., 2001c, Dynamic classifier selection based on multiple classifier behaviour, *Pattern Recognition*, **34**, 1879-1881.

Giacinto, G., Roli, F., and Bruzzone, L., 2000, Combination of neural and statistical algorithms for supervised classification of remote-sensing images, *Pattern Recognition Letters*, **21**, 385-397

Gong, P. and Howarth, P.J., 1990, An assessment of some factors influencing multispectral land-cover classification, *Photogrammetric Engineering and Remote Sensing*, **56**, 597-603.

Gong, P., Pu, R. and Chen, J., 1996, Mapping ecological land systems and classification uncertainties from digital elevation and forest-cover data using neural networks, *Photogrammetric Engineering and Remote Sensing*, **62**, 1249-1260.

Gopal, S. and Woodcock, C., 1994, Theory and methods for accuracy assessment of thematic maps using fuzzy sets, *Photogrammetric Engineering and Remote Sensing*, **60**, 181-188.

Gopal, S., Woodcock, C. E., and Strahler, A. H., 1998, Fuzzy neural network classification of global land cover from a 1° AVHRR data set, *Remote Sensing of Environment,* **67**, 230-243.

Gu, T., Anderson, J.M. and Monk, J.G.C., 1999, An approach to the spectral and radiometric calibration of the VIFIS system, *International Journal of Remote Sensing*, **20**, 535-548.

Gualtieri, J.A., and Cromp, R.F., 1998, Support vector machines for hyperspectral remote sensing classification. *In Proceeding of the 27th AIPR Workshop: Advance in Computer Assisted Recognition, Washington, DC, Oct. 27, 1998*, 221-232.

Gunn, S.R., 1998, Support vector machines for classification and regression, Technical report, School of Electronics and Computer Science, University of Southampton.

Haertel, V., Shimabukuro, Y. E. and Almeida-Filho, R., 2004, Fraction images in multitemporal change detection, *International Journal of Remote Sensing*, **25**: 5473-5489.

Hansen L.K. and Salamon, P., 1990, Neural network ensembles, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 993-1001.

Hansen M.C and Reed, B, 2000, A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products, *International Journal of Remote Sensing*, **21**, 1365-1373.

Hansen, M., Dubayah, R., and Defries, R., 1996, Classification trees: an alternative to

traditional land cover classifiers, *International Journal of Remote Sensing*, **17**, 1075-1081.

Hastie, T., Tibshirani, R., and Friedman, J., 2001, *The Element of Statistical Learning: Data Mining*, Inference, and Prediction, Springer-Verlag New York.

Hermes, L., Frieauff, D., Puzicha,, J., and Buhmann, J..M., 1999, Support vector machines for land usage classification in Landsat TM imagery, *Proc. IGARSS'99*, **1**, 348-350.

Hopfield, J. and Tank, D.W., 1985, Neural computation of decision in optimization problems, Biological Cybernetics, **52**, 141-152.

Huang, C., Davis, L.S., and Townshend, J.R.G., 2002, An assessment of support vector machines for land cover classification, *International Journal of Remote Sensing*, **23**, 725-749.

Huang, Z. and Lees, G., 2004, Combining non-parametric models for multisource predictive forest mapping, *Photogrammetric Engineering and Remote Sensing*, **70**, 415-425.

Husak, G.J., Hadley, B.C., and McGwire, K.C., 1999, "Landsat thematic mapper registration accuracy and its effects on the IGBP validation", Photogrammetric Engineering and Remote Sensing, **65**, 1033-1039.

Ibrahim, M. A., Arora, M. K., and Ghosh, S. K., 2003, Improving the quality of remotely sensed derived land cover maps by incorporating mixed pixels in various stages of a supervised classification process, *Proceedings of the IGARSS 2003 Symposium (Toulouse, France: IEEE Gescience and Remote Sensing)*, 3447-3449.

Ibrahim, M. A., Arora, M. K., and Ghosh, S. K., 2005, Estimating and accommodating uncertainty through the soft classification of remote sensing data, *International Journal of Remote Sensing*, **26**, 2995-3007.

Ishibuchi, H., Nozaki, K. and Tanaka, H., 1992a, Distributed representation of fuzzy rules and its application to pattern classification, *Fuzzy Sets and Systems*, **3**, 260-270.

Ishibuchi, H., Nozaki, K., Tanaka, H., 1992b, Pattern classification by distributed representation of fuzzy rules, *92 IEEE International Conference Fuzzy Systems*, 643-650

Jager, G., and Benz, U., 2000, Measures of classification accuracy based on fuzzy

similarity, *IEEE Transactions on Geoscience and remote Sensing*, **38**, 1462-1467.

Ji, C., and Ma, S., 1997, Combinations of weak classifiers, *IEEE Transactions on Neural Networks*, **8**, 32-42.

John A.R. and Xiuping J., 1999, *Remote Sensing Digital Image Analysis: an Introduction*, Third edition, Springer – Verlag Berlin Heidelberg, Germany.

Ju, J., Kolaczyk, E.D., and Gopal, S., 2003, Gausian mixture discriminant analysis and subpixel land cover characterization in remote sensing, *Remote Sensing of Environment*, **84**, 550-560.

Kanellopoulos, I., and Wilkinson, G.G., 1997, Strategies and best practice for neural network image classification, *International Journal of Remote Sensing*, **18,** 711-725.

Kavzoglu T. and Mather P.M., 2003, The use of backpropagating artificial neural networks in land cover classification, *International Journal of Remote Sensing*, **24**, 4907-4938.

Klecka, W.R., 1980, *Discriminant Analysis*, Sage Publications, Beverly Hills, London.

Klir, G.J. and Folger, T.A., 1998, *Fuzzy Sets, Uncertainty and Information*, Prentice-Hall International, London.

Klir, G.J., and Yuan, B., 1995, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, USA.

Krishnapuram, R. and Keller, J.M., 1996, The possibilistic c-means algorithm: insights and recommendations, *IEEE Transactions and Fuzzy Systems*, **4**, 385-393.

Kuncheva, L.I., Kountchev R.K., 2002, Generating classifier outputs of fixed accuracy and diversity, *Pattern Recognition Letters*, **23**, 593-600.

Kuncheva, L.I., Skurichina, M., and Duin, R.P.W., 2002, An experimental study on diversity for bagging and boosting with linear classifiers, *Information Fusion*, **3**, 245-258

Lam, L., and Suen, S.Y., 1997, Application of majority voting to pattern recognition: an analysis of its behavior and performance, *IEEE Transactions on Systems, Man and Cybernetics,* **27**, 553-568.

Lark, R.M., 1994, Sample size and class variability in the choice of a method of discriminant analysis, *International Journal of Remote Sensing*, **15**, 1551-1555.

Lee, T., Richards, J. A., and Swain, P.H., 1987, Probabilistic and evidential approaches for multisource data analysis, *IEEE transactions on Geoscience and Remote Sensing*, **25**, 283-293.

Lein, J. K., 2003, Applying evidential reasoning methods to agricultural land cover classification, *International Journal of Remote Sensing*, **10**, 4161-4180.

Lewis, H.G. and Brown, M., 1998, *FLIERS Software Framework and Toolkit Version 2.0; User Guide*, University of Southampton.

Lewis, H.G., 1998, The use of shape, appearance and the dynamics of clouds for satellite image interpretation, *Ph.D. Thesis*, University of Southampton, Southampton.

Lewis, H.G., and Brown, M., 2001, A generalized confusion matrix for assessing area estimates from remotely sensed data, *International Journal of Remote Sensing*, **22**, 3223-3235.

Lewis, H.G., Nixon, M.S., Tatnall, A.R.L., Edwards, M. and Hughes, M., 1999, On the role of uncertainty in empirical mixture modelling, In *Proceedings of the 25th Remote Sensing Society Conference,* Taylor and Francis, London, 709-716.

Lillesand, T. M. and Kiefer, R. W. (2000) *Remote Sensing and Image Interpretation*, Fourth Edition, Wiley, New York.

Lin, C., Lee, Y. and Pu, H., 2000, Satellite sensor image classification using cascaded architecture of neural fuzzy network, *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 1033-1043.

Liu, X., Skidmore, A.K., and Oosten, H.V., 2002, Integration of classification methods for improvement of land-cover map accuracy, *ISPRS Journal of Photogrammetry and Remote Sensing*, **56**, 257-268.

Loveland, T.R. and Belward, A.S., 1997, "The IGBP-DIS global 1 km land cover data set, DISCover: first results," *International Journal of Remote Sensing*, **18**, 3289-3295.

Lu, D., Moran, E., and Batistella, M., 2003, Linear mixture model applied to Amazonian

vegetation classification, *Remote Sensing of Environment*, **87**, 456-469.

Ma, Z. and Redmond, R.L., 1995, Tau coefficients for accuracy assessment of classification of remote sensing, *Photogrammetric Engineering and Remote Sensing*, **61**, 435-439.

Manolakis, D., Siracusa, C., and Shaw, G., 2001, Hyperspetral subpixel target detection using the linear mixing model, *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 1392-1409.

Manslow, J., 2001, On the Extraction and Representation of Land Cover Information Derived From Remotely Sensed Imagery, *Ph.D Thesis*, University of Southampton, Southampton.

Manslow, J.F., and Nixon, M.S., On the ambiguity induced by a remote sensor's PSF, *Uncertainty in Remote Sensing and GIS*, edited by G.M. Foody and P.M. Atkinson 2002, 37-57.

Maselli, F., Conese, C., De Filippis, T., and Norcini, S., 1995, Estimation of forest parameters through fuzzy classification of TM data, *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 77-84.

Mather, P. M., 2004, *Computer Processing of Remotely Sensed Images*, Third edition, Willey, New York.

Mcaffrey, T.M., and Franklin, S.E., 1993, Automated training site selection for large-area remote-sensing image analysis, *Computers and Geosciences*, **19**, 1413-1428.

McIver, D.K., and Friedl, M.A., 2002, Using prior probabilities in decision-tree classification of remotely sensed data, *Remote Sensing of Environment*, **81**, 253-261.

Melgani, F.; Al Hashemy, B.A.R.; Taha, S.M.R., 2000, An explicit fuzzy supervised classification method for multispectral remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 287-295.

Mertens, K.C., De Baets, B. Verbeke, L.P.C., and De Wulf, R.R., 2006, A sub-pixel mapping algorithm based on sub-pixel/pixel spatial attraction models, *International Journal of Remote Sensing*, **27**, 3293-3310.

Mood, A.M., Graybill, F.A. and Boes, D.C., 1974, *Introduction to the Theory of Statistics*, Third edition, McGraw-Hill Book Company.

Muslim, A.M., Foody, G.M. and Atkinson, P.A., 2006, Localized soft classification for super-resolution mapping of the shoreline, *International Journal of Remote Sensing*, **27**, 22271-2285.

Muslim, A.M., Foody, G.M., and Atkinson, P.M., 2006, Localised soft classification for super-resolution mapping of the shoreline, *International Journal of Remote Sensing*, **27**, 2271-2285.

Ng, G.S., Singh, H., 1998, Democracy in pattern classifications: combinations of votes from various pattern classifiers, *Artificial Intelligence in Engineering*, **12**, 189-204.

Nguyen, M.Q., Atkinson, P.M. and Lewis, H.G., 2006, Superresolution mapping using a Hopfield neural network with fused images, *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 736-749.

Nishii, R. and Tanaka, S., 1999, Accuracy and inaccuracy assessments in land-cover classification, *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 491-498.

Nozaki, K., Ishibuchi, H., Tanaka, H., 1996, Adaptive fuzzy rule-based classification systems, *IEEE Transactions on Fuzzy Systems*, **4**, 238 – 250

Oki, K., Uenishi, T.M., Omasa, K., and Tamura, M., 2004, Accuracy of land cover area estimated from coarse spatial resolution images using an unmixing method, *International Journal of Remote Sensing*, **25,** 1673-1683.

Pal, M. and Mather, P.M., 2001, Decision tree based classification and remotely sensed data, *22nd Asian Conference on Remote Sensing, 5-9 November 2001, Singapore, Asian Association on Remote Sensing.*

Pal, M., Mather, P.M., 2003, Support vector classifiers for land cover classification, *Map India 2003 Image Processing and Interpretation.*

Pal, S.K. and Mitra, S., 1992, Multilayer perceptron, fuzzy sets, and classification, IEEE *Transactions on Neural Networks*, **3**, 683-697.

Paola, J.D. and Schowengerdt, R.A., 1995, A review and analysis of backpropagation

neural networks for classification of remotely-sensed multi-spectral imagery, *International Journal of Remote Sensing*, **16**, 3033-3058.

Pasika, H., Haykin, S., Clothiaux, E., and Stewart, R., *1999,* Neural networks for sensor fusion in remote sensing, *Neural Networks. IJCNN '99. International Joint Conference*, **4**, 2772 – 2776.

Patterson, D.W., 1996, *Artificial Neural Networks: Theory and Applications*, Prentice Hall, London.

Peddle, D. R., 1995a, Knowledge formulation for supervised evidential classification, *Photogrammetric Engineering and Remote Sensing*, **61**, 409-417.

Peddle, D.R., 1995b, Mercury: An evidential reasoning image classifier, *Computer and Geosciences*, **21**, 1163-1176.

Peddle, D. R., and Ferguson, D. T., 2002, Optimisation of multisource data analysis: an example using evidential reasoning for GIS data classification, *Computer and Geosciences*, **28**, 45-52.

Perkins, S., Harvey, N.R., Brumby, S.P., and Lacker, K., 2001, Support vector machines for broad area feature classification in remotely sensed images, *Proceedings of SPIE - The International Society for Optical Engineering VII*, **4381**, 286-295.

Peterson, S.H., and Stow, D.A., 2003, Using multiple endmember spectral mixture analysis to study chaparral regrowth in southern California, *International Journal of Remote Sensing*, **24**, 4481-4504.

Petrakos, M., Benediktsson, J. A. and Kanellopoulos, I., 2001, The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion, *IEEE Transactions on Geoscence and Remote Sensing*, **39**, 2539-2546.

Petrou, M. Foschi, P.G., 1999, Confidence in linear spectral unmixing of single pixels, *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 624-626.

Plaza, A., Martinez, P., Perez, R., and Plaza, J., 2004, A quantitative and comparative analysis of end-member extraction algorithms from hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 650-663.

Raeth, P.G., Gustafson, S.C., Little, G.R., 1994, A basic function approach to programming concurrent voting systems to perform selection tasks, *Mathematical and Computer Modelling*, **20**, 73-88.

Rees, W.G., 1990, *Physical Principles of Remote Sensing*, Cambridge University Press, Cambridge.

Richard M.D., and Lippmann, R.P., 1991, Neural network classifiers estimate Bayesian a posteriori probabilities, *Neural Computation*, **3**, 461-483.

Richards, J. A., Landgrebe, D.A., and Swain, P. H., 1982, A means for utilizing ancillary information in multispectral classification, *Remote Sensing of Environment*, **12**, 463-477.

Roberts, D. A., Gardner, M., Church, R., Ustin, S., Scheer, G. and Green, R. O., 1998, Mapping Chaparral in the Santa Monica Mountains Using Multiple Endmember Spectral Mixture Models, *Remote Sensing of Environment*, **65**, 267-279.

Rogova, G., 1994, Combining the results of several neural network classifiers, *Neural Networks*, 7, 777-781.

Roli, F., Giacinto, G., and Vernazza, G., 1997, Comparison and combination of statistical and neural network algorithms for remote-sensing image classification, *Neurocomputation in Remote Sensing Data Analysis*, Berlin: Spring-Verlag, 117-124.

Rosenfield, G.H., and Fitzpatrick-Lins, K., 1986, A coefficient of agreement as a measure of thematic classification accuracy, *Photogrammetric Engineering and Remote Sensing*, **52**, 223-227.

Rueda, I.E.A., Arciniegas, F.A., and Embrechts, M.J., 2004, SVM sensivity analysis: an application to currency crises aftermaths, *IEEE Transactions on Systems, Man, and Cybernetics – part A: Systems and Humans*, **34**, 387-398.

Sabins, F.F., 1996, *Remote Sensing Principles and Interpretation*, Third edition, Quebecor Semline, USA.

Schalkoff, R.J., 1992, Pattern recognition: statistical, structural and neural approaches, Wiley, New York.

Schowengerdt, R. A., 1996, On the estimation of spatial-spectral mixing with classifier

likelihood functions, *Pattern Recognition Letters*, **17**, 1379-1387.

Schowengerdt, R.A., 1997, *Remote Sensing: Models and Methods for Image Processing*, Academic Press, San Diego.

Schwenk, H., Bengio, Y., 2000, Boosting neural networks, *Neural Computation*, **12**, 1869-1887.

Serpico, S.B., Bruzzone, L., and Roli, F., 1996, An experimental comparison of neural and statistical non-parametric algorithms for supervised classification of remote-sensing images, Pattern Recognition Letter, **17**, 1331-1341.

Serpico, *S.B.; Roli, F., 1995,* Classification of multisensor remote-sensing images by structured neural networks, IEEE Transactions on Geoscience and remote sensing, **33,** 562-578

Settle, J.J. and Drake N.A., 1993, Linear mixing and the estimation of ground cover proportions, *International Journal of Remote Sensing*, **14**, 1159-1177.

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

Shalan, M. A., Arora, M. K., and Ghosh, S. K., 2003, An evaluation of fuzzy classifications from IRS 1C LISS III imagery: a case study, *International Journal of Remote Sensing,* **24**, 3179-3186.

Simard, M., Saatchi, S.S., and Grandi, G.D., 2000, The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical, *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 2310-2320.

Skole, D. and Tucker, C., 1993, Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988, Science, **260**, 1905-1910.

Small, C., 2001, Estimation of urban vegetation abundance by spectral mixture analysis, *International Journal of Remote Sensing*, **22**, 1305-1334.

Small, C., 2004, The Landsat ETM+ spectral space, *Remote Sensing of Environment*, **93**, 1-17.

Smits, P.C., 2002, Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection, *IEEE Transaction on Geoscience and*

Remote Sensing, **40**, 801-813.

Song, C., 2005, Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate end-member variability?, *Remote Sensing of Environment,* **95**, 248-263.

Souza J.R.C. And Barreto, P., 2000, An alternative approach for detecting and monitoring selectively logged forests in the Amazon, *International Journal of Remote Sensing*, **21**, 173-179.

Steele B. M., 2000, Combining multiple classifiers: an application using spatial and remotely sensed information for land cover type mapping, *Remote Sensing of Environment*, **74**, 545-556.

Stoodley, K.D.C., Lewis, T. and Stainton, C.L.S., 1980, *Applied Statistical Techniques*, Ellis Horwood Limited.

Suarez, A. and Lutsko, J.F., 1999, Globally optimal fuzzy decision trees for classification and regression, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**, 1297-1311.

Tadjudin, S. and Landgrebe, D.A., 2000, Robust parameter estimation for mixture model, *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 439 – 445.

Tatem, A.J., 2002, Super-resolution land cover mapping from remotely sensed imagery using a Hopfield neural network, *Ph.D. thesis*, University of Southampton, Southampton, UK.

Tatem, A. J., Lewis, H. G., Atkinson, P. M. And Nixon, M.S., 2001, Super-resolution target identification from remotely sensed images using a Hopfield neural network, *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 781-796.

Tatem, A. J., Lewis, H. G., Atkinson, P. M. And Nixon, M.S., 2002, Super-resolution land cover pattern prediction using a Hopfield neural network, *Remote Sensing of Environment,* **79**, 1-14.

Theseira, M. A., Thomas, G., Taylor, J. C., Gemmell, F., and Varjo, J., 2003, Sensitivity of mixture modelling to end-member selection, *International Journal of* Remote Sensing, **24**, 1559-1575.

Theseira, M. A., Thomas, G., Sannier, C. A. D., 2002, An evaluation of spectral mixture modelling applied to a semi-arid environment, *International Journal of Remote Sensing*, **23**, 687-700.

Tom, C.H., and Miller, L.D., 1984, An automated land-use mapping comparison of the Bayesian maximum likelihood and linear discriminant analysis algorithms, *Photogrametric Enginnering and Remote Sensing*, **50**, 193-207.

Townshend, J.R.G., 1992, Land cover, *International Journal of Remote Sensing*, **13**, 1319-1328.

Townshend, J.R.G., Huang, C., Kalluri, S. N. V., Defries, R. S., Liang, S., and Yang, K., 2000, Beware of per-pixel characterization of land cover, *International Journal of Remote Sensing*, **21**, 839-843.

Tso, B. and Mather, P. M., 2001, *Classification Methods for Remotely Sensed Data*, Taylor and Friancis, London.

Tzeng, Y.C., and Chen, K.S., 1998, A fuzzy neural network to SAR image classification, *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 301-307.

Van Der Meer, F. and De Jong, S.M., 2000, Improving the results of spectral un-mixing of Landsat Thematic Mapper imagery by enhancing the orthogonality of end-members, *International Journal of Remote Sensing*, **21**, 2781-2797.

Vapnik, V.N., 1995, *The Nature of Statistical Theory*. New York: Springer-Verlag.

Veitch, N., Treweek, J.R. and Fuller, R.M., 1995, The land cover map of Great Britain – a new data source for environmental planning and management, in Danson, F.M. and Plummer, S.E. (eds.), *Advances in Environmental Remote Sensing*, John Wiley, Chichester, 157-170.

Verhoeye, J. and De Wulf, R., 2002, Land cover mapping at sub-pixel scales using linear optimisation techniques, *Remote Sensing of Environment,* **79**, 96-104.

Wang, F., 1990a, Fuzzy supervised classification of remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 194-201.

Wang, F., 1990b, Improving remotely sensed image analysis through fuzzy information

representation, *Photogrammetric Engineering and Remote Sensing*, **56**, 1163-1169.

Wang, F., 1994, The use of artificial neural networks in a geographical information system for agricultural land-suitability assessment, *Environment and Planning A*, **26**, 265-284.

Warner, T.A. and Shank, M., 1997, An evaluation of the potential for fuzzy classification of multispectral data using artificial neural networks, *Photogrammetric Engineering and Remote Sensing*, **63**, 1285-1294.

Wilkinson, G. G. and Megier, J., 1990, Evidential reasoning in a pixel classification hierarchy – a potential method for integrating image classifiers and expert system rules based on geographic context, *International Journal of Remote Sensing*, **11**, 1963-1968.

Windeatt, T., 2003, Vote counting measures for ensemble classifiers, *Pattern Recognition Letters*, **36**, 2743-2756.

Woodcock, C.E. and Gopal, S., 2000, Fuzzy set theory and thematic maps: accuracy assessment and area estimation, *International Journal of Geographical Information Science*, **14**, 153-172.

Woods, K., Kegelmeyer, W.P., and Bowyer, K., 1997, Combination of multiple classifiers using local accuracy estimates, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, 405-410.

Wu, C., and Murray, A. T., 2003, Estimating impervious surface distribution by spectral mixture analysis, *Remote Sensing of Environment*, **84**, 493-505.

Wyatt, B.K., Greatorex-Davies, J.N., Hill, M.O., Parr, T.W., Bunce, R.G.H. and Fuller, R.M., 1993, Comparison of land cover definitions, *Institute of Terrestrial Ecology report to the Department of the Environment*, Institute of Terrestrial Ecology, Cambridge.

Xu, L., Krzyzak, A., and Suen, C.Y., 1992, Methods of combining multiple classifiers and their applications to handwriting recognition, *IEEE Transactions on Systems, Man and Cybernetics*, **22**, 418-435.

Yamakawa, T., 1993, A fuzzy inference engine in nonlinear analog mode and its application to a fuzzy logic control, *IEEE Transactions on Neural Networks*, **4**, 496-521.

Yool, S.R., 1998, Land cover classification in rugged areas using simulated

moderateresolution remote sensor data and an artificial neural network, *International Journal of Remote Sensing*, **19**, 85-96.

Zadeh, L., 1965, Fuzzy sets, *Information Control*, **8**, 338-353.

Zadeh, L., 1973, Outline of a new approach to the analysis of complex or imprecise concepts. *IEEE Transactions on Systems, Man and Cybernetics*, **3**, 28-44.

Zhang, J., Foody, G.M., 1998, A fuzzy classification of sub-urban land cover from remotely sensed imagery, *International Journal of Remote Sensing*, **19**, 2721-2738.

Zhang, J. and Foody, G. M., 2001, Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches, *International Journal of Remote Sensing*, **22**, 615-628.

Zhang, J. and Goodchild, M.F., 2002, *Uncertainty in Geographical Information*, Taylor and Francis, London.

Zhang, J., and Kirby, R. P., 1997, An evaluation of fuzzy approaches to mapping land cover from aerial photographs, *ISPRS Journal of Photogrammetry and Remote Sensing*, **52**, 193-201.

Zhou, A., Wu, J., Tang, W., 2002, Ensembling neural networks, *Artificial Intelligent*, **137**, 239-263.