THE USE OF CONTEXT IN THE CLASSIFICATION OF URBAN AERIAL IMAGERY

Alastair James Buchanan

Doctor of Philosophy

Aston University

November 2000

Acknowledgements

I would like to thank several of people for their help in completing this thesis, but in particular I would like to thank my supervisor John Elgy. I would also like to thank the following people for their assistance with certain aspects related to the project, and for their friendship throughout: Peter Hedges, Rob Poole, Tom Charnock, Rob Ellis, Simon Gregory, and Phil Purnell. The project was funded by the EPSRC, to whom I extend my gratitude. Lastly I would like to thank Rhiannon for doing small bits of proof reading and digging into her deep reserves of patience.

Aston University

The Use of Context in the Classification of Urban Aerial Imagery

Alastair Buchanan

Doctor of Philosophy, November 2000

Summary

Urban regions present some of the most challenging areas for the remote sensing community. Many different types of land cover have similar spectral responses, making them difficult to distinguish from one another. Traditional per-pixel classification techniques suffer particularly badly because they only use these spectral properties to determine a class, and no other properties of the image, such as context.

This project presents the results of the classification of a deeply urban area of Dudley, West Midlands, using 4 methods: Supervised Maximum Likelihood, SMAP, ECHO and Unsupervised Maximum Likelihood. An accuracy assessment method is then developed to allow a fair representation of each procedure and a direct comparison between them.

Subsequently, a classification procedure is developed that makes use of the context in the image, though a per-polygon classification. The imagery is broken up into a series of polygons extracted from the Marr-Hildreth zero-crossing edge detector. These polygons are then refined using a region-growing algorithm, and then classified according to the mean class of the fine polygons. The imagery produced by this technique is shown to be of better quality and of a higher accuracy than that of other conventional methods. Further refinements are suggested and examined to improve the aesthetic appearance of the imagery.

Finally a comparison with the results produced from a previous study of the James Bridge catchment, in Darleston, West Midlands, is made, showing that the Polygon classified ATM imagery performs significantly better than the Maximum Likelihood classified videography used in the initial study, despite the presence of geometric correction errors.

Keywords: Remote Sensing, Per-Polygon Classification, Accuracy Assessment, ATM.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Introduction

Precise land-use information about urban catchments is a primary input into urban drainage models. The complexity of the model determines the nature of this input - for less sophisticated models, such as BEMUS (Maksimovic *et al*, 1995) the percentage of impermeable area (PIMP) is required, whereas more sophisticated models, such as distributed physically-based models require the exact location of land use type, such as roads. Models of non-point sources of pollution also require this type of information.

Urban catchments are prone to considerable land use changes over short time-scales. The general trend is for an increase in the population density of urban areas and the urbanization of former sub-urban and rural catchments (Maksimovic, 1993). This is a global tendency, as illustrated in Figure 1.1. As land usage becomes more urban, demand for an efficient and comprehensive drainage system is increased, since the capacity of the system may be insufficient to cope with increased runoff.

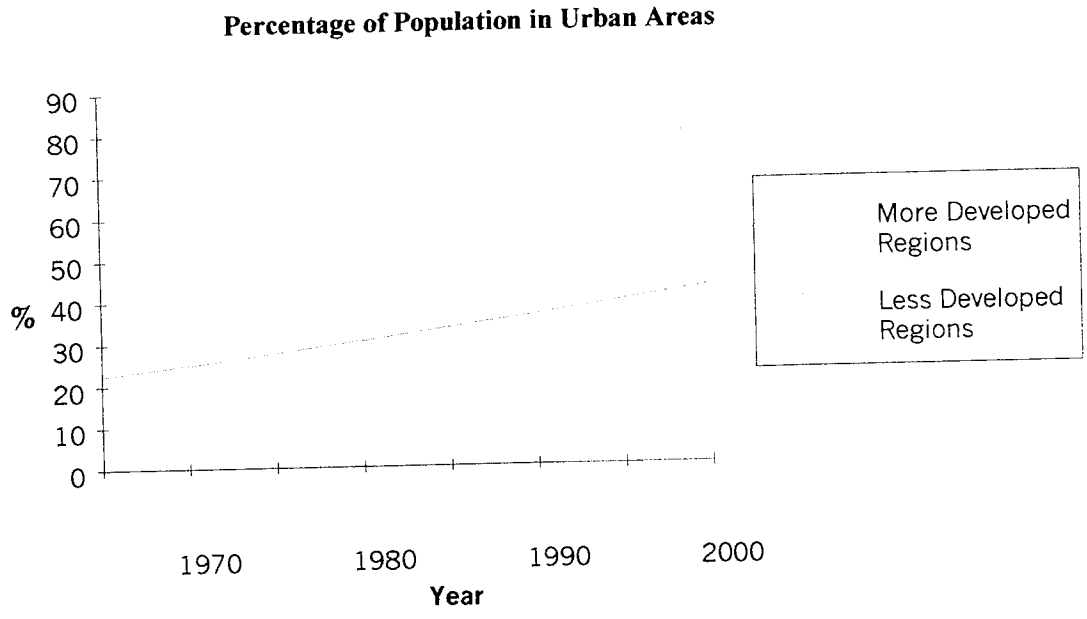


Figure 1.1 Projected figures demonstrating the trend towards urbanisation (Maksimovic 1993)

The development of urban drainage models over recent years has allowed the accurate prediction of runoff for these areas. However, acquiring spatial data for these models is an expensive and time-consuming process. Remote sensing can provide up to date information at a fraction of the cost of traditional methods, and ultimately can become an important tool in the design of urban drainage networks.

There are several commercially available computer packages that calculate surface runoff for a given rainfall intensity and duration (for instance HYSTREM-EXTRAN, Fuchs and Sheffer, 1993). While they are dependent on a number of catchment characteristics including soil type and catchment wetness, of most importance is the accurate determination of the amount of permeable and impermeable cover within the catchment. However, while the models offer good reliability, the calculation of the percentage of impervious surface (PIMP) is either expensive, in the case of photo-interpretation, inaccurate if pixel based computer classification procedures are used, or both if out of date maps are used (Finch *et al*, 1989).

Because remote sensing data sets are frequently very large, computationally efficient statistical techniques are required for pixel classification. For the past 20 years the major image processing techniques used have revolved around statistical pattern recognition and analysis, extending the usual methods of multivariate analysis to the remote-sensing problem. However these techniques do not exploit the spatial component of the data. They are primarily concerned with the multi-spectral features of the individual pixels, while neglecting the effect that size, shape and texture of the surrounding objects will have on the correct identification of these pixels. These properties are referred to as context, and are necessary in order to identify objects within the imagery, in the same manner as a photo-interpreter might.

The normal statistical approach to pixel classification assumes that data vectors in neighbouring pixels are independent. This is not so. Furthermore, spatial dependencies between pixels may be caused by scattering of reflected electro-magnetic radiation from the surface of the earth resulting in contamination from over-sampling and re-sampling, or spatial continuity of the ground classes (Curran, 1985). Data from neighbouring pixels will not be independent even if they are independent conditional on the ground classes. Unless training pixels are selected to be sufficiently far apart, these spatial dependencies can result in biased estimates of the covariance matrix, and hence in an increased classification error

rate. Additionally, statistical analysis is performed on pixels independently and so may result in classified images that are patchier than in reality.

Several *ad hoc* approaches have been used to incorporate context into the procedure of pixel classification. These include simple augmentation of the data vectors by the data vectors of surrounding pixels, smoothing algorithms and various contextual classification methods. Contextual classification exploits the tendency for certain ground cover classes to occur more frequently in certain contexts than in others. The classification of a given pixel is influenced by the probable classifications of surrounding pixels.

1.2 Background to the Project

There has been a considerable amount of research carried out by the Environmental Systems Research Group at Aston University into the effect of context on the classification of remotely sensed imagery (Booth, 1989; Oldfield, 1988). Further work has also been concerned with integrating the data into an urban drainage Geographical Information System (Elgy *et al*, 1993).

The data examined in this project were acquired during a previous study of an area of Dudley, West Midlands (Ellis, 1997). The imagery is line-scan data from the Airborne Thematic Mapper (ATM), and covers a region of deep urbanisation. The imagery is of a highly complex nature and contains many of the features associated with urban regions, such as many different types of land use with similar spectral patterns.

Recent papers on the subject (Finch *et al*, 1989; Scott, 1994; Fankhauser, 1998) have demonstrated that existing techniques are insufficiently sophisticated to classify adequately urban imagery due to its complex nature. Additionally, these papers are characterised by the weak assessment of the accuracy of these methods.

1.3 Aims and Objectives

There is clearly a requirement for a classification method that will provide the type of high accuracy information that today's urban drainage models require. As these models become more sophisticated they also require more detailed information about the catchment to which they are applied. The provision of this data can only currently be supplied through the use of manual classification procedures. However, these procedures may be prohibitive due to their cost.

Consequently the primary purpose of the thesis is to find a way of improving the quality of automatically classified data supplied to the drainage models

Thus the main aim of the project is to develop a classification procedure that automates as much of the photo-interpretation process as possible, and consequently improves upon existing methods for urban image classification. In order to do this contextual information must be incorporated into the classification procedure. A further aim of this project is to determine a method of accuracy assessment that allows an unbiased comparison between classification methods.

Having achieved these objectives, a further goal is to compare the entire procedure with an existing scheme for urban classification.

1.4 Thesis Structure

Chapter 2 examines the published literature on several subjects, all of which are closely related to classification procedures. A review of some recent papers that attempt to classify urban imagery to provide urban drainage inputs is followed by an overview of several of the classification procedures currently in use. Low level image processing is then examined, along with context and methods for incorporating this into a classifier. Techniques for assessing the success of the classification methods are followed by some applications of classified imagery.

Chapter 3 introduces four classification procedures, and examines how they operate, presenting the results with a discussion of their qualitative merits and drawbacks.

Chapter 4 is concerned with developing a method for assessing the accuracy of the classifications in order that they may be directly compared with one another. The classified imagery presented in Chapter 3 is comprehensively examined with several techniques, and a comparison of the accuracy of each is made.

Chapter 5 presents the development of a new polygon classifier, illustrating in a stage by stage fashion the reasoning behind its development. This is followed by an assessment of the accuracy of this classifier, using the methods developed in Chapter 4. A comparison between the new classifier and the existing methods examined in Chapter 3 is then made.

Chapter 6 investigates the areas that remain for improvement of the classifier.

Chapter 7 makes use of the classifier for the determination of the percentage of impermeable area (PIMP) of an urban catchment in Darleston, West Midlands, and compares the results achieved with some results from a previous study.

Chapter 8 presents the conclusion.

The literature concerning the subjects mentioned will now be examined and discussed.

Chapter 2: Review and Evaluation of Existing Classification Schemes

2.1 Introduction

" Existing urban drainage models are of sufficient detail and complexity to model accurately the effects of storms on a urban drainage basin. Their widespread use is curtailed mainly by problems of data acquisition, storage and manipulation.....remote sensing offers an excellent possibility to obtain land cover data that is up to date and easily ported into a GIS"

Elgy et al, 1993.

The paper by Elgy *et al* examined the use of a variety of Geographic Information Systems for creating the input files for two different urban drainage models. As the statement above indicates, the main weakness in the procedure lies in the acquisition of high quality land cover data, the fundamental problem being concerned with the trade-off between cost and accuracy. Photo-interpretation of aerial imagery, although highly accurate, is a time-consuming and costly process - it can take more than one day to classify a single photograph.

The Ordnance Survey (OS) can supply digital maps of high quality for the United Kingdom, although they may be out of date, costly and have certain copyright and user license restrictions. Ordnance Survey digital maps are not GIS products. An examination of Figure 2.1 demonstrates this. The map does not contain polygons, continuous line data, land cover information or the potential for links to other attribute databases. For instance, the road polygon remains open at (a). There are also spurious boundaries (b), such as the barriers indicated. There is no inclusion of land use information (c) - the path at this point contains paved blocks, whereas the path at point (d) is covered with gravel. Additionally, there is no attribute data, such as unique identifiers for individual buildings that can be linked to a database (e). Consequently, the maps provided by the OS are not suitable for hydrological GIS without considerable post-processing. An added problem is that the data

remains under the ownership of the OS – any subsequent use requires subsequent licensing and expense (Ordnance Survey, 1997).

In comparison, multi-spectral data provided by the Airborne Thematic Mapper can be acquired cheaply, and is completely owned by the purchaser. Videography can be produced even more cheaply, although it is in analogue form and requires extensive pre-processing to be of use (Scott, 1994; Blagojevic *et al*, 1994). The data, once acquired, can subsequently be used to produce a classified land usage map which can then be ported into a Geographical

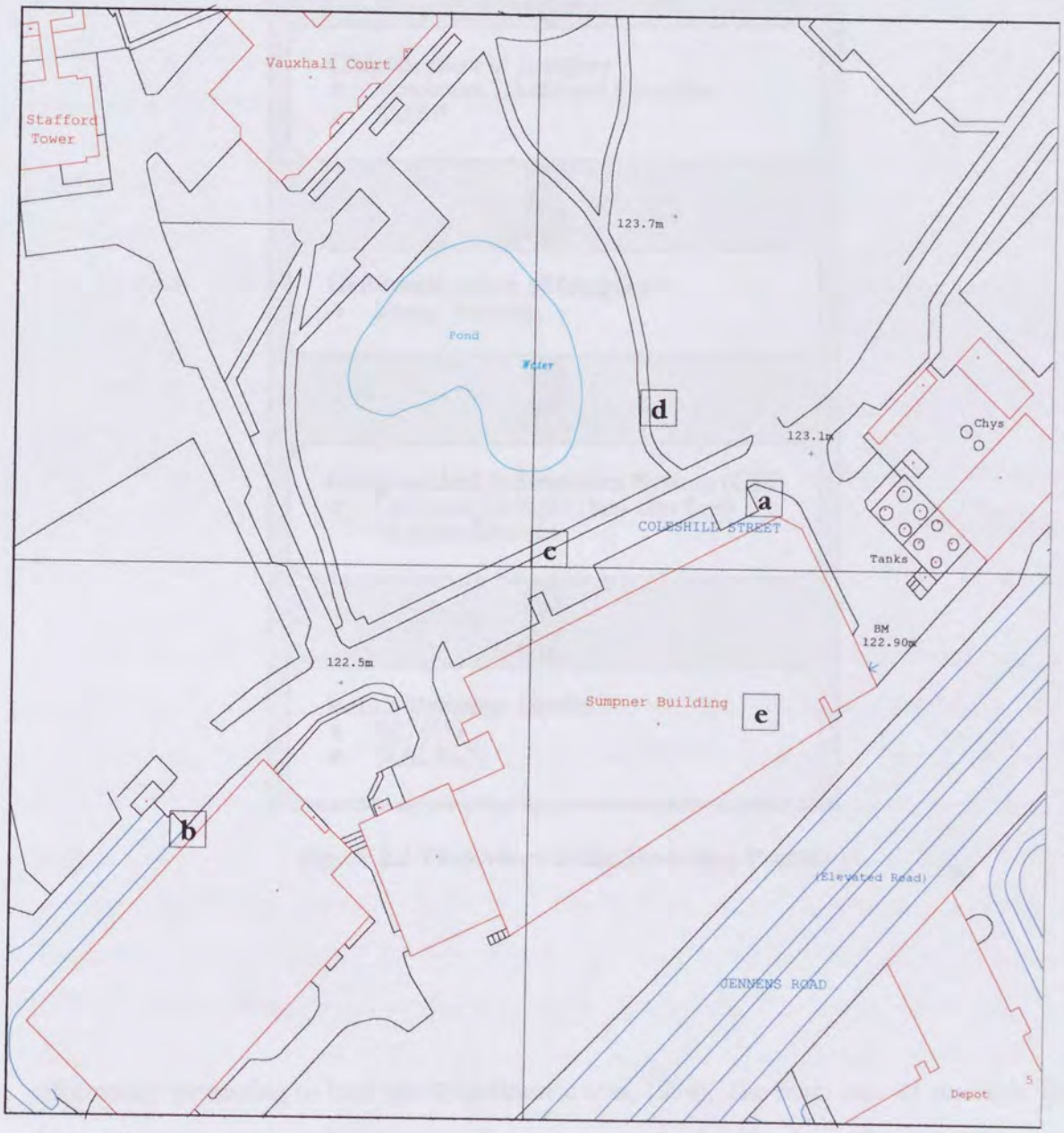


Figure 2.1 An Ordnance Survey map of Aston University campus (original scale 1:1250)

Information System (GIS).

Remotely Sensed imagery is an important source of digital data for the assessment of the quantity of water resources for planning purposes, for the updating of aged maps, and for

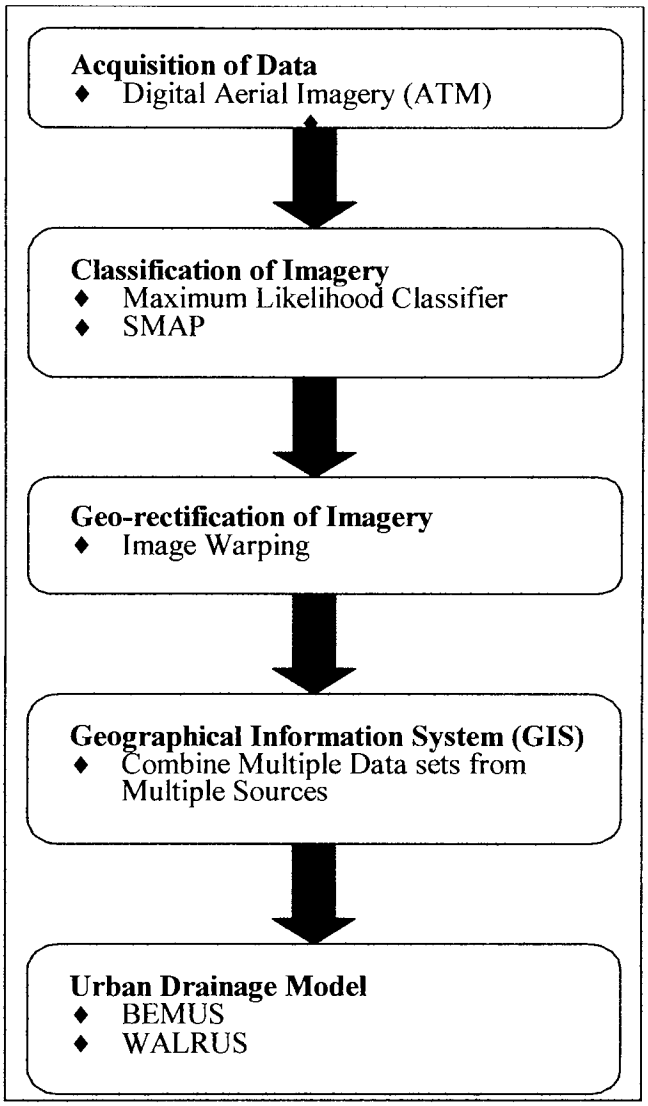


Figure 2.2 Flow chart of the Modelling Process

information pertaining to land use (Maksimovic *et al*, 1994). The main area of research, for this project, is concerned with providing an accurate classification of remotely sensed

airborne data, producing a high quality input into an urban drainage model, combined through the use of a Geographical Information System. The procedure for doing this is outlined in Figure 2.2

This chapter examines the literature covering the methods and tools used to generate and manipulate the data. The theory of per-pixel classification is briefly explained along with some of the most popular and recent developments in this area. The term 'context' is then introduced, followed by a look at some of the attempts to incorporate the information it provides into the classification procedure and discusses the relative success of these approaches. Methods for assessing the accuracy of the classification process are then introduced, followed by applications of the classified image.

2.2 Background

2.2.1 Introduction

The paper by Finch *et al* (1989) illustrates perfectly many of the problems associated with image classification. Finch used a supervised Maximum Likelihood method to classify the percentage of impervious surface of Daedalus ATM imagery, with a pixel resolution of 2m.

Five spectral bands were used (3, 5, 7, 9 and 11) for the analysis of nine catchment areas. Initially eleven training classes were defined, with these being combined to eventually give three classes: roofs, roads and paths, and pervious area.

A traditional survey using a 1:2500 scale Ordnance Survey map of the area was made for use as reference data, with additional information being provided by aerial photographs and, where necessary, ground surveys. Using this map, ground control points were matched with the ATM imagery and through the use of cubic convolution interpolation, the image was geometrically rectified, with a transformation root mean square error of 3m. To achieve a geometric rectification accuracy of this level, the flight must have taken place in excellent conditions. The data acquired for this project could not be rectified to anywhere near this level of accuracy due to greater platform instability - see section 2.2.6 for a discussion on the difficulties of geo-rectifying ATM Imagery, and section 3.1.3 for details specific to this project.

The mean differences between the results of a manual survey and the classification for the three classes were found to be 15.5% (roof), -7.4% (roads and paths) and 8.1% (pervious area) respectively. This was calculated for each individual sub-catchment by calculating the area of PIMP for the manual image, calculating the area of PIMP for the classified image and calculating the difference between the two. The difference for each sub-catchment was then averaged to provide this value. So if, for example, the manual classification was 50% PIMP, and the Maximum Likelihood classification was 40% PIMP, the difference would be 10%. However, the error, which is not mentioned, would be 20%. Furthermore, the averaging of this data does not take account of catchment size. Consequently, very small catchments have a disproportionate effect on the calculation of the mean difference value.

Unfortunately, while these results indicate some measure of success, closer inspection revealed some disturbingly large errors in some of the catchments. For instance, the error in the calculation of road and path (in catchment 2) underestimated the manually derived amount by over 50% (manual = 39.0%, classification=16.6%). Finch *et al* explain that in this catchment there is an unusually large number of trees overhanging the road, resulting in a large misclassification. Other problems encountered include large errors in regions with many different types of roof materials, cars and areas lacking homogeneity.

The situations that cause misclassification point to many of the shortcomings of per-pixel classification methods, and indicate that additional information is necessary. Further to this, however, it is clear that there were serious errors in the methodology used by Finch *et al.* The presence of overhanging trees, for instance, could have been avoided if the imagery was acquired during a more suitable season, a fact that he acknowledges. What is not acknowledged is that the presentation of the accuracies given is based on several dubious assumptions. The first of these is that the manual survey was definitive and the second is that geo-rectification of the digital imagery was of sufficient accuracy to allow quantitative comparison between the two.

The third, and most important, assumption, is that the percentage of pixels over the whole catchment can be used directly for accuracy assessment. As we will see in chapter 4, a far higher accuracy value will be obtained if the overall figure is quoted rather than using a measure of the total correct pixels combined with some indication of the errors of omission

and commission involved. The use of a 'mean difference' value is particularly misleading - we are given no real measure of accuracy in the first instance, so we are forced to make inferences through the difference between values which, while perhaps not meaningless, are deceptive.

Unfortunately, these results were actually intended to provide the input for the WASSP drainage model. The results, as presented by Finch, show that the digital classification overestimated the amount of pervious area by 8.0 %. This figure could, however, have been achieved through a completely inaccurate classification, since we are never shown if any individual pixels are actually correctly classified. This being the case, one must treat these results with great caution. The errors in the methodology mean that it is not possible to transfer these results or conclusions to another area.

Scott (1994) compares the use of multi-spectral airborne videography and manual methods of classification. The purpose was to provide the input for a WALLRUS drainage model (Wallingford Software, 1991.) Since this model needs land use information to distinguish between roads, roofs and permeable areas, Scott reasoned that multi-spectral data was overly sophisticated. He suggested that vidoegraphy could provide a low cost alternative, especially useful for small catchments.

The imagery was collected using a true colour camera and an infrared camera, with still images grabbed and mosaiced to provide a single image. He found that the multi-band classification (he does not mention which one) was poor, but that density-slicing techniques were of an acceptable accuracy.

The presentation of the data, however, is once more misleading, giving a permeable area measurement masquerading as overall an accuracy figure in a similar fashion to Finch *et al.*

Some of the problems encountered with multi-band classification are discussed, particularly with regard to the misclassification of certain areas, such as greenhouses, sheds and patios. He suggests that these areas should be excluded under the conditions given for use of the drainage model and can be rectified by applying a correction factor over the whole imagery.

It is perhaps an indication of the inherent lack of success of the technique used that he suggests a value for this correction factor of 25%.

Fankhauser (1998) used aerial photographs as the main data source, due to the unavailability of any other source of high-resolution imagery in the region of Switzerland under examination. The photographs were split into the three primary bands (red, green and blue) and analysed using a Maximum Likelihood classification. The number of classes were initially determined by inspection, and subsequently grouped into three classes: impervious, pervious and shadow. After the classification, a modal filter was employed to eliminate small clusters of pixels that were mis-classified.

He reported that the automatic classification of the aerial photography, when compared with a manual classification, overestimated the results by up to 15%, while the combination of aerial photography with colour-infrared photography produced similar estimations. There was, however, no measure of accuracy over and above this 'fit for use' reasoning. This, in effect, provides no proper analysis of the method employed and gives no indication as to where the misclassification is occurring. Indeed, it gives no indication at all of how many incorrectly classified pixels there actually are. In essence, the conclusion to be drawn is that the manual classification and the automatic classification could be completely unrelated and still produce 'acceptable' results. The images presented by Fankhauser at the conference and not reproduced in the paper show much permeable area mis-classified as permeable and vice versa. Fortunately, this was to about the same amount for both.

Traditionally, the major statistical image processing techniques used in remote sensing have revolved around statistical pattern recognition and analysis. However, these methods have been primarily concerned with the multi-spectral features of individual pixels, and have not examined what effect the size, shape and location of surrounding areas may infer. This results in the characterisation of spectral classes while neglecting the description and identification of objects - in other words the techniques used in visual photo-interpretation. Multi-spectral classification methods such as the Maximum Likelihood classifier will result in the misclassification of a significant percentage of pixels due to variations in shade and emissivity within the same class.

It has been argued that the misclassification does not matter as long as the percentages for each class of land cover obtained through automatic classification are close to those of the manual classification (Blagojevic *et al*, 1998). In a catchment containing equal amounts of permeable and impermeable areas, for instance, these errors will cancel each other out. As a result it can be asserted that the automatically classified map performs as well as the manual one. However, if there is a large amount of permeable area and a small amount of impermeable area, these errors will not cancel each other out.

Therefore it is important that land-use is correctly calculated. In addition, the data used by Blagojevic *et al* (1998) was closely linked to the urban drainage model. Different models require more precise land-use information.

Although errors may be large, this method of automatic interpretation is still used due to the high cost of human interpretation. It can take more than one day to interpret a single photograph by manual means, making these methods very time consuming. However, some of the misclassification experienced by pixel-based methods are due to the assumption that surrounding pixels and objects within an image have no relationship to individual pixels.

When compared to the results achieved by a skilled photo-interpreter, statistical pattern recognition and image processing procedures yield inadequate results. This is because image interpretation is more complicated than reading individual pixels from images. It involves seeing and understanding, which can be aided by identifying the effect that image pattern elements (i.e. size, shape, texture, situation, homogeneity) have. We group these elements together and refer to them as the context of the image.

2.2.2 Manual Preparation of Land Use Maps

Typically, a map of a region is subdivided into discrete parcels, each labelled as a single category, known as a cloropeth map (Campbell, 1983). For many users, land use maps prepared in this manner appear to be the simplest of all maps. Consequently, land use maps are conceived to be the most basic to compile and interpret. Yet, simplicity of form and content conceal complexities that may emerge only as critical attention is devoted to the meaning and usefulness of specific maps. Once encountered, these complexities are interwoven with the geographical patterns of land use, the classification system, and the

qualities of the aerial images used to make the map. Intricate errors in form and content may influence the usefulness of the map and are seldom subject to convenient correction, because they flow from decisions regarding the map purpose, detail scale and use.

The cartographic model used to portray land use patterns identifies each mapped parcel with a single category. In reality, most mapped parcels are composed of several categories of land use - the inevitable consequence of cartographic generalisation. This means that detail, perhaps important detail, is withheld from the map-reader.

Land use information is, of course, of great significance in scientific research. Geographers, economists and others have long regarded land use patterns as fundamental in their studies of economic systems. Land use patterns are recognised as influential elements in hydrological and meteorological processes

However, maps in their paper form quickly become redundant, due to the difficulty encountered when linking the geographical data with other attribute data. The advent of GIS has meant that paper maps must be digitised to be of any use. In addition, maps (particularly those of urban regions) quickly become outdated and consequently need constant revision. In particular, maps of urban regions, due to the rapid growth and decline of industrial buildings can become redundant very quickly indeed.

2.2.3 The Photo-interpretation Procedure

Most aerial imagery presents a map-like representation of the landscape that seems to form a natural and convenient base for delineating land use. Usually, however, the photo-interpreter is presented with much more information than can be accurately and legibly presented on a map. As a result, the interpreter defines a working model for relating detail on the image to specific land cover categories, then attempts to apply this model consistently throughout the image, using the model to assign areas of the image to categories in the classification system. This working model can be represented as a kind of filter that separates relevant from irrelevant detail or, alternatively, as a translator that can assign image detail to its correct informational category, much as a linguist assigns words and concepts in one language to corresponding words and concepts in another.

These filters/translators are devised and applied informally, almost intuitively, by photo-interpreters in a manner that is tailored to the scale and resolution of the imagery, to the detail in the classification system, and to the publication scale of the final land use map. It is essential that each interpreter apply the strategy in a disciplined, systematic and consistent manner throughout the image, so that the final map is uniform in accuracy and representation of detail.

Interpreters study remotely sensed data and attempt through logical processes to detect, identify, measure and evaluate the significance of environmental and cultural objects, patterns, and spatial relationships. Interpreters use elements of image interpretation to detect, recognise and identify objects and patterns. These elements traditionally include size, shape, tone texture, pattern, and association - that is those qualities that permit us to recognise features we see on aerial images.

2.2.4 Urban Imagery

Urban areas provide the remote sensing community with its most challenging problem due to the extremely complicated nature of the land-use variation. They constitute some of the most difficult and time-consuming tasks for photo-interpreters and provide a rich environment of varied structures and natural terrain features against which to test the robustness of new approaches. Furthermore, many types of cover have similar spectral properties, making discrimination by existing multi-spectral methods difficult.

The primary difficulty, when dealing with urban imagery, is that materials such as concrete, asphalt, roofing materials, plaster, paint etc., are used to produce a diverse array of land uses which, unfortunately, may produce similar spectral responses. It becomes clear that it is difficult to discriminate one specific urban class of information from another based solely on spectral characteristics (Colwell, 1983). The importance of size, shape, texture/pattern information is evident, with these elements of image interpretation being essential to accurate identification of urban features.

2.2.5 The Airborne Thematic Mapper

The Dædalus 1268 Airborne Thematic Mapper (ATM) provided the source data for this study. It is a multi-spectral line-scanner whose 11 channels cover the visible and near-

infrared (bands 1-8), short-wave infrared (SWIR - bands 9 and 10) and thermal infrared (TIR - band 11) and includes channels that closely match the seven spectral channels of the Landsat Thematic Mapper. Table 2.1 shows the wavelengths of each band. More specific details relating to the system itself can be found in Wilson (1995).

The data generated by the ATM may require pre-processing to account for atmospheric effects and other problems that arise from the instability of the platform (Richards, 1986). An important step when analysing geo-spatial data is in the integration of the data into a Geographical Information System which requires accurate geometric rectification of the imagery. Unfortunately, achieving this with ATM imagery is extremely difficult. Whereas satellite data can be effectively rectified using polynomial equations, aircraft motion causes disturbances of too great a severity to be compensated for by even high-order polynomials.

Channel Number	Wavelength (μm)	Applications
1	0.42 - 0.45	Bathymetry; soil/vegetation differentiation
2	0.45 - 0.52	Coniferous/deciduous differentiation
3	0.52 - 0.60	Green reflectance by healthy vegetation
4	0.605 - 0.625	Secondary green reflectance band
5	0.63 - 0.69	Plant species differentiation by chlorophyll absorption
6	0.695 - 0.75	Geo-botanical surveys; water body delineation
7	0.76 - 0.90	Bio-mass surveys; water body delineation
8	0.91 - 1.05	Secondary near infra-red band
9	1.55 - 1.85	Altered rock mapping; vegetation moisture measurement
10	2.08 - 2.35	Hydro-thermal mapping
11	8.5 - 13.0	Plant heat stress detection; thermal mapping
12	8.5 - 13.0	Plant heat stress detection; thermal mapping half the gain of band 11

Table 2.1 The Band Wavelengths of the Airborne Thematic Mapper (Wilson, 1995).

Methods to improve the rectification of existing ATM imagery have been the focus of much recent research (Gregory, 1996; Ehlers, 1994). These will be covered in the following section on Geo-rectification.

2.2.6 Geo-rectification

An important step in the integration of multi-spectral imagery into a GIS is the accurate geometric rectification of the image data. The main sources of geometric error commonly found in remotely sensed images can be categorised in the following way (Richards, 1986);

1. the rotation of the earth during image acquisition,

2. the curvature of the earth,

3. the wide field of view of some sensors,

4. panoramic effects relating to the imaging geometry,

5. sensor imperfections,

6. the finite scan rate of some sensors, and

7. variations in the platform altitude, attitude and velocity.

These errors can be classified into two groups; those which are deterministic in their nature (numbers 1-6), and those resulting from fluctuations in platform stability giving rise to varying degrees of distortion throughout the image. Deterministic errors can generally be successfully compensated for by constructing mathematical models that reflect the nature and magnitude of the image distortions. However, variations in platform stability cause potentially the most severe problems in airborne remotely sensed imagery due to the unpredictable pattern with which they distort the image geometry. These errors are most severe in images obtained from airborne line scanning sensors, since distortions continually vary from one scan line to the next as the sensor platform is buffeted by atmospheric currents during image acquisition. Mounting the sensor upon a stabilised platform may reduce attitude effects, although these stabilisation systems cannot guarantee compensation of very rapid attitude movements.

A number of efforts have been made to address the problems of geo-rectification of airborne line scanner imagery. These can be classified into two groups - parametric and non-parametric. Non-parametric methods attempt to correct imagery by developing general expressions to describe the image distortions, irrespective of the nature of error. In the absence of platform ephemeris data, image rectification is most commonly achieved by using polynomial warping algorithms. This procedure transforms the raw image data to a map co-ordinate system by developing mapping polynomials based on a series of ground control points (GCPs). Although polynomial correction offers a convenient and computationally efficient method for image correction, airborne line-scanner imagery contains local regions of distortion due to platform instability, and these errors are generally too severe to be represented by a first or second order polynomial. This fundamental drawback has meant that historic data sets of the same region can rarely be accurately co-registered, limiting the ability to quantitatively assess temporal change.

However, a number of efforts are currently being made to address the problems of geo-rectification of airborne line-scanner imagery. Zhang *et al.* (1994) have shown that an accuracy of 2-3 pixels can be achieved by utilising flight parameters obtained from on-board motion sensors. The Natural Environment Research Council (NERC) have recently introduced a remote sensing system using GPS satellite technology to provide platform ephemeris data (Wilson, 1994). However, a number of teething problems have emerged since the introduction of this facility and the problems of precise geo-rectification of airborne imagery remain largely unresolved.

Ehlers (1994) developed a method that relied on Multi-quadratic Interpolation techniques. Whereas satellite data can be effectively rectified using polynomial equations, the aircraft motion causes disturbances of too great a severity to be compensated for by even, high-order polynomials. Although computationally intensive, the Multi-quadratic Interpolation approach was shown to improve accuracy, quantitatively and qualitatively.

The improvement of this important aspect of ATM imagery is, in fact, the concern of ongoing research within the Environmental Systems Research Group at Aston University.

2.2.7 Geographical Information Systems

Geographical Information Systems (GIS) are software programs that provide one or more spatial data models, usually based on raster or vector structures. Associated with every spatial feature or entity is a set of descriptive attributes, which collectively describe geographical phenomena. Special functions allow users to query, analyse and display complex spatial information (Burrough, 1986). Remote sensing provides a major source of geographical data which, when interpreted, can be merged with other data layers in a GIS (such as urban water drainage models) (Maksimovic *et al*, 1994; Fuchs *et al*, 1994; Blagojevic *et al*, 1994).

There are a variety of computer packages available that combine image processing features with GIS capabilities, such as ERDAS (Erdas, 1998), IDRISI (Eastman, 1995), ERMapper (ERMapper, 1998), ARCInfo (ArcInfo, 1998), and GRASS (Shapiro *et al*, 1993). The choice of which one to use is difficult, since they all perform many of the same features. However, some are more suited to GIS work, where as others have a large range of image processing features and sacrifice some GIS capabilities.

At various stages other packages were also used. ERDAS was used to prepare the data for input into a package called Multispec (Landegrebe, 1980), which performed the ECHO classification (see section 2.3.4.4). Houghtool (Kälviäinen *et al*, 1996) was used to produce straight-line segments from various Hough transform algorithms.

The Geographical Resources Analysis Support System (GRASS) is a public domain image processing and raster based GIS originally developed by researchers in the Environmental Division of the US Army Construction Engineering Research Laboratories (USACERL) in Champaign, Illinois (Shapiro *et al*, 1993). Unlike many GIS software toolkits, GRASS is open; it has well-defined accessible data formats; the source code is supplied, allowing users to modify and extend code at will.

GRASS is written in the C language and is accompanied by an extensive list of program libraries and complete source code. As well as containing GIS facilities, the user system provides a broad range of functions, currently numbering over 400, to perform various geo-processing operations. Most analytical functions operate on raster data, and include such

capabilities as map algebra, statistical reporting, filtering, classification and so on. Special features include tools to build more complex applications using shell scripts - GRASS contains a comprehensive list of image processing modules, allowing easy development of any program for geo-processing. Additionally, it has a large user base and extensive support in the form of Internet mailing list and newsgroups. The open source system makes it an ideal base for image processing/GIS research.

2.3 Classification

2.3.1 Introduction

Preparation of a land use map from aerial imagery is essentially a process of segmenting the image into a mosaic of parcels, with each parcel assigned a land use class. The most useful categories, in practice, are those that match the informational needs of the map user. Typically categories such as 'suburban land' are completely unsatisfactory for the user who requires division of land into functional classes such as 'residential', 'commercial' and 'industrial' land.

The image interpreter must prepare a classification system that is simultaneously compatible with the needs of the map user and consistent with image detail and map scale.

2.3.2 Scale of Imagery and Classification Accuracy

When embarking on scientific research, it is usual for the scale at which observations are collected to be carefully selected by the investigator. In remote sensing, however, there are limitations regarding the type of platforms available, and the actual data available to the researcher. In this project, for instance, the data was collected prior to the project commencing. However, the complicated nature of urban elements under study, and the level of detail required of the output from the study are such that there can be little doubt as to the appropriateness of the scale of the data.

Woodcock and Strahler (1987) propose an approach based on the spatial structure of images, which they suggest is an indication of the relationship between environments and spatial resolution. The spatial structure of the images is expected to be primarily related to the relationship between the size of the objects in the scene and spatial resolution. Consequently graphs of local variance may be used to measure spatial structure in images.

Interestingly, they also indicate that at finer spatial resolutions more pixels fall on the boundaries between classes. This will result in lower classification accuracy. Additionally, the increased spatial variance of land-cover classes associated with finer spatial resolution will also influence classification accuracy since within-class variance decreases the spectral separability of classes. Thus we would expect lower accuracy from the classification of high-resolution urban imagery than we would from low-resolution agricultural imagery.

2.3.3 Classification of Ground Cover Classes

The classification of remotely sensed data provides the user with a map detailing the land cover over a region of interest. Theoretically, a given cover type gives a unique pixel value in each spectral band, so across a given number of spectral bands a cover type can be identified by the pattern of pixel values. We term this the 'spectral response pattern' for that specific land cover type. In reality a more complex situation exists, with there being a statistical variation in the pixel values for that land cover type, which produces a cluster of points, as opposed to a single point, on a scatter plot (see Figure 2.3). A classification algorithm separates the feature space into region enclosing these clusters and produces a map that discriminates between the land cover classes.

We term classifiers that operate on this principle (in either supervised or unsupervised mode) per-pixel classifiers. That is, they discriminate each pixel individually, ignoring the effect surrounding pixels may have on the choice of class.

2.3.4 Per-Pixel Classifiers

2.3.4.1 Introduction

The term 'per-pixel' is used to refer to a classification procedure that classifies each pixel individually, based solely on the multi-spectral properties of each particular pixel. The following section presents some examples of the more commonly used techniques, as well as some more recent innovations. A comprehensive review of per-pixel classification can be found in Colwell (1983) and Lillesand and Kiefer (1994).

2.3.4.2 Minimum Distance to Means

One of the simplest classification procedures is the Minimum Distance to Means Classifier. For each category the mean spectral value is calculated for each waveband under consideration. As the name suggests, a pixel of unknown identity may be classified

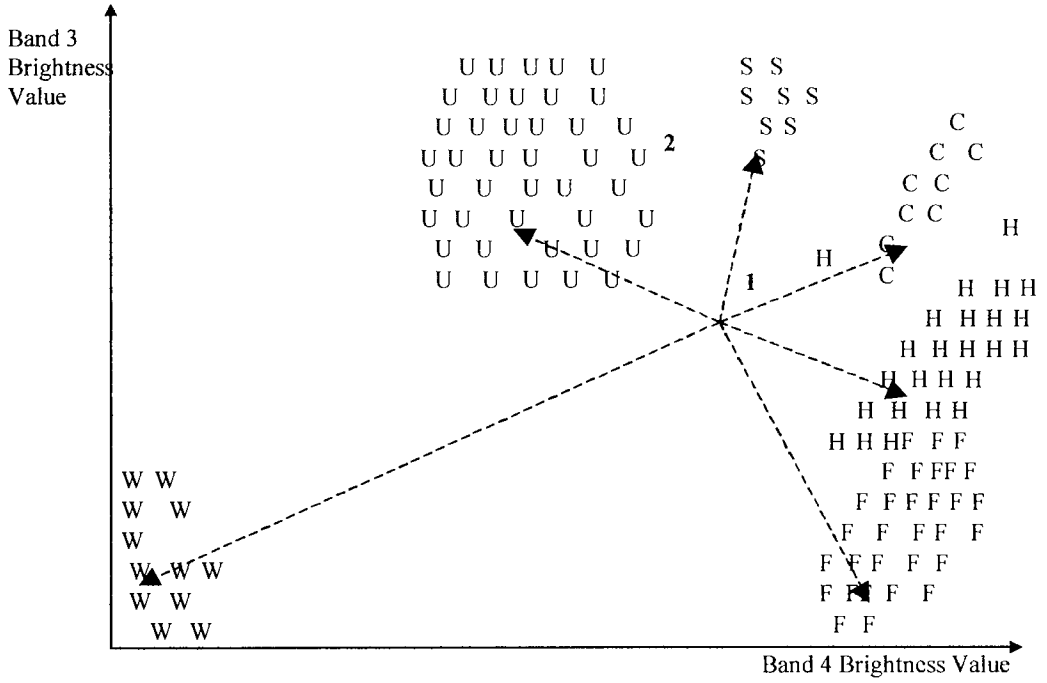


Figure 2.3. The Minimum Distance to Means Classifier (from Lillesand and Keifer, 1994)

by computing the distance between the value of the unknown pixel and the mean of each category. For instance, in Figure 2.3 the pixel plotted at 1 would be assigned to the Sand (S) category.

The mathematical simplicity of this strategy produces a fast computation time but it has a major drawback. In Figure 2.3, the pixel plotted at point 2 would be assigned to the Sands category (S), whereas the greater variability of the Urban category (U) suggests this would be a more appropriate assignment. As a consequence, this is not a widely used classifier.

2.3.4.3 Maximum Likelihood Method

When classifying an unknown pixel, this method evaluates both the variance and correlation of the category spectral response patterns. This is based on the assumption that the distribution of points in the category training data is Gaussian. This is a reasonable assumption for common training class spectral distributions. In a supervised image classification it makes use of the region means and covariance matrices from the training areas selected to determine which category each cell in the image has the highest possibility of belonging to. It is then assigned to that category.

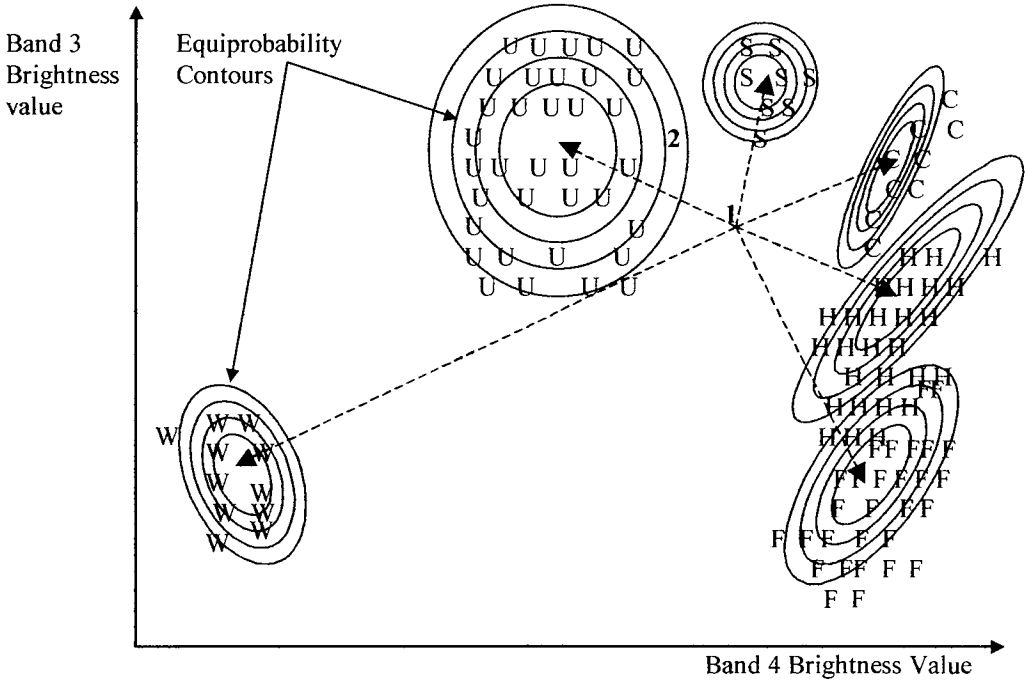


Figure 2.4 Equiprobability contourse of the Maximum Likelihood method (Lillesand and Keifer, 1994)

In Figure 2.4 we can see that the pixels at points 1 and 2 would be assigned to the Sand category (S) and Urban category (U) respectively.

2.3.4.4 ECHO

The ECHO (extraction and classification of homogeneous object) is a two stage conjunctive approach to image segmentation. It may be used in supervised or unsupervised mode.

The first stage, cell selection, consists of dividing the data into small groups called cells (consisting of four or more pixels), using a rectangular grid, then subjecting the pixels of each cell to a mild test of statistical homogeneity. Cells that fail this test are referred to as singular and are assumed to overlap a boundary in the data. They are classified on a pixel-by-pixel basis. In the second step, adjacent non-singular cells are tested for statistical similarity using a generalised likelihood ratio test. Cells found to be similar are merged or annexed. The results of extensive testing (Landegrebe, 1980) indicate that the ECHO classification provided greater classification accuracy at greater computational efficiency than the Maximum Likelihood classifier.

In essence, this classifier is an example of a 'split and merge' method, where statistically similar neighbouring pixels are merged if they exhibit similar spectral responses. Consequently, this classifier is not strictly contextual, although some spatial analysis is incorporated through this neighbourhood scheme. The relevance of this classifier is that it attempts to extract homogeneous regions from imagery through this scheme.

2.3.5 Bayesian Classifiers

An extension to the maximum likelihood classifier is the Bayesian classifier. This technique applies two weighting factors to the probability estimate. First the analyst determines the *a priori* probability for each class. Secondly a weight associated with the cost of misclassification is applied to each class. Together, these factors act to minimise the cost of misclassification, resulting in a theoretically optimum classification.

Besag (1986) describes a classification scheme that makes use of the fact that pixels close together tend to have the same or similar values. In order to do this two information sources are combined by Bayes' Theorem. The first of these sources, is that associated with each pixel there is a multivariate record, which provides data on the individual pixel brightness value, assuming the records follow a known statistical distribution. The second source makes the assumption that the local characteristics can be represented

probabilistically by a non-degenerative Markov random field, which crudely represents the underlying scene.

The combination of these sources by Bayes' Theorem is achieved by choosing (i) the brightness value that has the overall maximum probability, and (ii) that in which each brightness value at each individual pixel has maximum probability, given the records. In a Bayesian framework (i) corresponds to maximum *a posteriori* estimation, and (ii) maximises the posterior marginal probability at each pixel.

However, due to reservations concerning the computational burden and the undesirable large scale characteristics of the random field, a simple iterative method (Besag, 1986) is shown which, although still based on probabilistic considerations, avoids these problems.

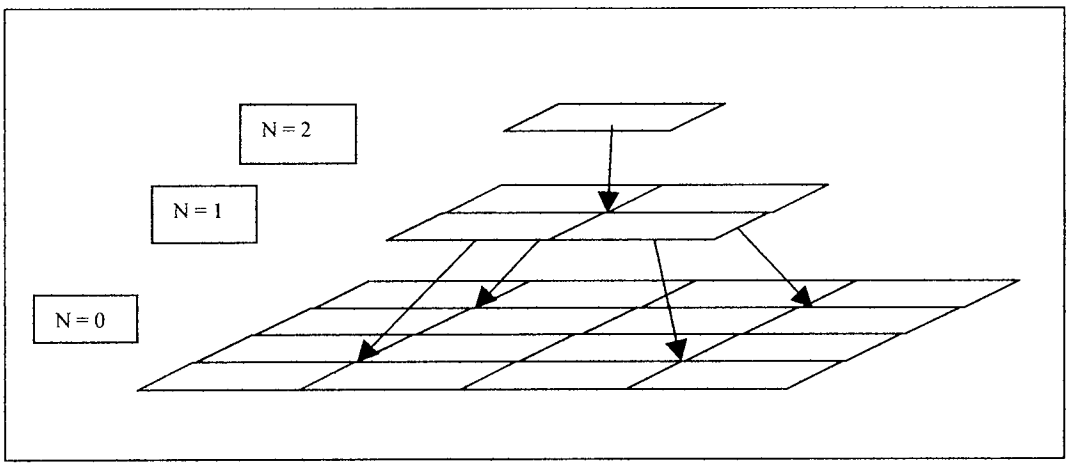


Figure 2.5 The pyramidal structure of the MSRF. The random field at each scale is causally dependent on the coarser scale field above it (Bouman and Shapiro, 1992)

2.3.5.1 SMAP Classifier

The Sequential Maximum *A Posteriori* estimator (Bouman and Shapiro, 1992; Bouman and Shapiro, 1994) is a recently developed algorithm for the Bayesian segmentation of images. The approach of Besag (1986) used a maximum a *posteriori* (MAP) estimation in conjunction with Markov random fields. There are, however, several problems, namely that exact MAP estimates cannot be computed, approximate MAP estimates are still computationally expensive, and unsupervised parameter estimation of the Markov random field is difficult.

Bouman and Shapiro propose a new method which replaces the MRF model with a new multi-scale random field (MSRF) and replace the MAP estimator with a sequential MAP (SMAP) estimator derived from a new estimation criteria. The MSRF is composed of a series of random fields, progressing from coarse to fine scale, with each finer field only dependent upon the previous coarser field (Figure 2.5). A Markov chain is thus formed by the series of fields. The result is a rich model with computationally tractable properties. The new SMAP estimator minimises the expected size of the largest region and is calculated recursively, hence the name.

The SMAP method may be used in either supervised or unsupervised modes, the details of which, along with comparisons to the Maximum Likelihood method and the method suggested by Besag (1986) may be found in Bouman and Shapiro (1992). A further comparison with the Maximum Likelihood method and the ECHO method indicates the success of this method with overall classification accuracies of 93.4, 92.1 and 89.8% for the SMAP, ECHO and Maximum Likelihood methods respectively (McCauley and Engel, 1995). However, these accuracies relate to aerial imagery of a region of farmland with large homogeneous tracts and not an urban region. As we have seen, urban imagery presents us with the most rigorous test of classifier performance.

2.3.6 Neural Network Classifiers

The motivation behind the use of artificial neural networks for remote sensing classification is based on the realisation that the human brain is very efficient at processing vast quantities of data from a variety of different sources. This section will not concentrate on the specifics of these networks, since these details can be found in the standard texts (Kohenen, 1988; Simpson, 1990; Bishop, 1995). The focus here is on the application of these networks, and the pros and cons of this approach.

Neural networks, being an artificial intelligence technique, come from the same family as expert systems and knowledge based approaches to learning. However, neural networks are only concerned with mathematical analysis of the data, rather than higher level heuristics. When applied to classification they are concerned with the transformation of data from feature space to class space. Hence, they really belong to the same class as other automated pattern recognition techniques (Atkinson and Tatnall, 1997).

According to Atkinson and Tatnall (1997) the capabilities of neural networks are affected by:

i) The number of nodes – the larger the number of nodes, the better represented the training data, but the poorer the ability of the neural network to generalise.

ii) The size of training set – the data must be representative of the entire distribution of values likely to be associated with a particular class. If the extent of the distribution of the data in feature space is not adequately covered the network may fail to classify new data accurately. As a result, large quantities of data are often required for training.

iii) Training time – the longer a network is trained the more accurately it can classify that data, but the poorer it can classify previously unseen data. It is possible to over-train a network so that it is able to memorise the training data but is unable to generalise when applied to different data.

Ersoy and Hong (1990) applied a hierarchical network to classify airborne multi-spectral scanner system imagery. Despite the fact that this was low resolution agricultural data (see section 2.3.2) the best accuracy achieved, after using several different types of neural network, was 59.6%. This also confirms the findings of Civco (1993) that in certain circumstances neural networks were actually less accurate than conventional statistical approaches for classifying land cover.

There are few papers that discuss the use of neural networks for urban aerial imagery classification. One of the reasons for this is that the complexity of the data is such that developing a classifier that can operate on multiple image sets is extremely difficult. It is likely that for each data-set, considerable retraining of the network would be required. Consequently there is little to suggest that this approach, for this study, would yield any improvement on existing classification procedures.

2.3.7 Fuzzy Classification

Fuzzy sets are appropriate where the boundaries between phenomena are not distinct. Generally there is a trend towards the investigation of fuzzy sets that include various degrees of set membership (Gopal and Woodcock, 1994). In neuro-fuzzy techniques, the power of neural networks are combined with fuzzy logic to enable fuzzy rules to be incorporated into

the classification and enable the intrinsic uncertainty in classification to be represented and minimalised. A common problem with classification in remote sensing is that many observed pixels represent a mixture of classes. Methods of fuzzy classification for dealing with sub-pixel mixing, including the use of neural networks, are described in Atkinson *et al* (1997).

A development of the technique is the fuzzy c-means classification. This can be considered as a modification of 'hard' classification where each pixel is assigned to a class based on some measure of the distance in feature space from the point representing the class means (for instance, the Maximum Likelihood method). However, given information on the distances to various class means the classification can be extended to assign certain pixels certain amounts to certain classes (Atkinson *et al*, 1997)

2.3.8 Spectral Mixture Modelling

There is a group of classification methods that make use of proportion estimation, or mixture models (Woodcock and Strahler, 1987). These are appropriate under conditions of high local variance where the contrast between measurements is maximised. In these models, the proportion of several elements is estimated for each pixel. Since they are formulated for high variance conditions they should be suited to urban imagery classification. However, the mixture models provide information about the elements in the lower level of the scene model, and as such are not a substitutes for spectral classifiers at these resolutions. In urban environments they would provide information about the elements comprising the scene, such as house, lawns, roads, etc., and would not classify the general land-use or land-cover. As this study is concerned with the correct determination of the land-use in the imagery, they are not appropriate for this project.

2.3.9 Incorporation of Spatial Information - Contextual Algorithms

The per-pixel spectral classifiers have found favour because of the wealth of information that can be obtained from the spectral domain at relatively low computational cost. However, their limitations become apparent when compared with the techniques used by a trained photographic observer, particularly as they fail to utilise other types of information such as texture and the interrelationships between pixels. There have been some attempts to

incorporate this type of information into the segmentation procedure, although the correct classification of these segments based on this information is still poor.

The term contextual can be misleading, as we will discover in section 2.5.1. A more appropriate term would be neighbourhood algorithms, since in reality these classifiers simply take into account the effect of directly surrounding pixels on each pixel.

The algorithms developed to represent context attempt to incorporate the following influences, with varying degrees of success (Oldfield, 1988).

1. Distance - The pixel is likely to be associated with pixels of its own class. Objects, which are separated by some distance from an object of their own class, are likely to be mis-classified.

2. Direction - Pixels are likely to be associated with pixels separated by some angular direction or are less likely to be associated with some pixels at some other angular direction.

These two influences can be applied not only to single pixels but also 7to objects and therefore be used either before or after classification. A description of contextual algorithms developed for image classification, which make use of these neighbouring functions, is provided in Appendix 4.

From the descriptions of the various algorithms, it becomes apparent that they only consider context in broad terms. They do not fully utilise information about the objects classified, do not attempt to describe regions, and do not take into consideration the shape of regions. Therefore, when classifying high-resolution urban data, they are of limited use. For instance, in daytime imagery, the multi-spectral signatures of shadow regions and water regions are very similar. However, the shape, size, position, and location of the shaded region is the type of information that may assist us when determining its correct class. These contextual algorithms are fundamentally unable to represent this expertise and cannot be truly described as contextual.

2.4 Texture

2.4.1 Introduction

Texture is an important low-level characteristic for the analysis of many types of images. It can be described qualitatively as having properties of roughness, coarseness, smoothness, granulation, randomness, lineation or being mottled, irregular or hummocky (Haralick, 1979). In other words, textural features contain information about the spatial distribution of tonal values.

Texture is the organisation of a surface as a set of repeated elements. It is of importance because distortions of a texture in an image are likely to correspond faithfully to orientation and depth changes in the original scene.

Several methods have been developed for extracting information about texture. These fall into two categories: Structural Analysis and Statistical Analysis. Structural Analysis depends on the prior recognition of the texture, that is, analysis of it as consisting as a set of elements that are basically the same. The idea behind Statistical Analysis is to measure the amount of texture, giving us a value that tells us how rough or smooth the analysed region is.

2.4.2 Structural analysis

This concerns the use of blocks of textural primitives (texels) to cover a region of an image. Much work has been concentrated in this area, mainly involving the matching and analysis of Brodatz textures (Brodatz, 1966), a large catalogue of natural texture samples ranging from straw to bricks. Image texels originate from the perspective projection of (usually 3-D) texture surfaces onto the image plane, and their appearance consequently depends on mainly geometric and photo-metric effects that distort the properties of the original 3-D texels.

Much of the work in this area has been concentrated on the discrimination and subsequent extraction of texels from simple artificial and natural scenes (Strand and Taxt, 1994; Taxt *et al*, 1989; Hild and Shirai, 1993). However, while good segmentation can be achieved in simple scene analysis, the ambiguous location of region borders and the absence of standard textures throughout make this type of analysis unsuited to urban aerial imagery. It is mentioned only for completeness.

2.4.3 Statistical Analysis

This approach regards texture as a sample from a probability distribution on the image space and is defined by a stochastic model or characterised by a set of statistical features. The most common features are based on the tonal properties and the pattern properties. These are measured from first and second order statistics and have been used for discrimination between textures.

The first order statistics, such as the grey level, are the simplest statistical example serving as a means of classifying a set of textures. The pixel being considered is directly replaced by some measure of the pixels around it. Measures include mean, variance, and skew (Oldfield, 1988).

The second order statistics, such as the grey level co-occurrence matrix and the grey level difference histogram are also powerful texture measures. Higher order statistics, such as the grey level run length, auto-regression model and the auto-correlation function that can be measured. Brief descriptions are given in Appendix 3, with a more detailed description of each being given in Oldfield (1988) and Haralick (1979).

2.5 Context

2.5.1 Introduction

The term context was introduced into pattern recognition to describe the explicit incorporation of dependencies between elements, for example, to reduce ambiguities in classifications or fill gaps in segmentations. The elements may be individual pixels, groups of pixels (in the form of a segment), or properties of the segment.

Our understanding of space can be affected by context. When we normally discuss spatial problems, we commonly refer to the arrangement of objects or their configuration. However, the situation or state of events that are perceived are as important as the mere physical structure. Information pertaining to the objects, such as the function, situation and environment of the objects, will have a profound effect when we attempt to classify them. These components taken together provide the context associated with a given region of interest.

In effect, these components (along with the more obvious ones such as size and shape) are what an image interpreter makes use of when interpreting an image. Knowledge with regard to the regional environment is referred to as background knowledge, but for our purposes would still be described as contextual. The subtlety of spatial context becomes apparent, and we can see that it is therefore difficult to automate. For instance, an expert system approach (see section 2.6) suffers from the difficulty in developing an operator set (or knowledge-based rules) general enough to be used for multiple data sets or scenes (CSS, 1989).

Another problem arises, due to the many different interpretations of the term context. As stated by Krinn (1992): " The most successful approach to the definition of operator sets would be based on shared conceptual definitions of space and analysis of objects in space." In other words, it is important for these definitions to be standardised.

2.5.2 Methods for Representing Context

Previously the attempts to represent context have been split into two categories, namely contextual algorithms, and expert systems. It is important to distinguish between the two approaches at this point. The contextual algorithms, as we have already seen, refer to an image consisting of pixels, and attempt to solve the problem of misclassification by using mathematical solutions. The expert system approach has a much larger scope, attempting to describe objects by modelling human expertise in the form of heuristics, or "rules - of - thumb" (Tailor *et al*, 1986; Argialas and Harlow, 1990). It allows us to combine a range of strategies for this purpose and choose the most appropriate one based on a pre-ordained decision strategy.

2.6 Expert Systems

2.6.1 Introduction

Artificial intelligence and knowledge-based expert systems have been used to represent expertise and are particularly suited to tasks that are hard to break down into an algorithm. The progress of research in the area of artificial intelligence has been aided by the development of techniques that allow the modelling of information at higher levels of abstraction (Nagao and Matsuyama, 1980).

These utilise whatever domain-specific knowledge is available about the class of scenes they are to "understand". This takes many forms, including descriptive definitions of entities, concepts and objects, and their relationship to each other and criteria for making decisions. The shift is now from procedural knowledge representation to declarative. A procedural representation of a fact is a set of instructions that, when carried out, arrive at a result consistent with the fact, while a declarative representation of a fact is an assertion that the fact is true.

Expert systems represent knowledge as a set of rules - rules being in the form of IF-THEN statements. Each rule has a condition part, which consists of one or more antecedent clauses, and an action part, the consequent, which may create or modify working memory elements. The order in which these rules are activated is dependent of the control mechanism of the system.

Rules are appropriate for image representation because a major part of the domain - specific knowledge results from empirical associations (heuristics) developed through years of experience in a particular area, which may be expressed as heuristic rules. However, representing knowledge as an unordered and unstructured set of rules has certain disadvantages. For example, one cannot easily express the structure of the domain in terms of taxonomic, or part - whole, relations that hold between objects and between classes of objects (CSS, 1989). Further details on the available architecture of general expert systems can be found in Giarratano and Riley (1994). Further information regarding the development of expert systems for image processing is given by Levine (1985), Tailor *et al* (1986) and Argialas and Harlow (1990).

2.6.2 Expert Systems use in Remote Sensing

A thorough review of some of the attempts to incorporate contextual information into an expert system follows.

Nagao and Matsuyama (1980) developed one of the first expert vision systems incorporating rule-based techniques into the analysis of complex aerial scenes. This performed structural analysis of aerial photographs in a 'segmentation by recognition' manner. Objects and regions were recognised by spectral and spatial features such as size, shape, location, colour

and texture. Whenever possible, the spatial characteristics of regions were used rather than spectral properties so that the system was capable of giving stable results despite changes in photographic conditions. Knowledge about location constraints and spatial arrangements was used so that context sensitive objects, such as cars on roads and regularly arranged houses, could be recognised.

A further aerial scene analysis system called SIGMA (Matsuyama, 1987; Matsuyama and Hwang, 1990) demonstrated mechanisms for the focus of attention, conflict resolution and the correction of early segmentation errors. SIGMA emphasised spatial reasoning and top-down goal-directed image segmentation, while an earlier system (Nagao and Matsuyama, 1980) incorporated sophisticated image analysis procedures to recognise objects based on their various spectral and spatial properties. Although the system of Nagao and Matsuyama could analyse fairly complex scenes, its capabilities were limited; the types of recognisable objects were limited and various recognition errors occurred due to noise and the imperfection of the segmentation algorithms. The SIGMA system analysed black and white aerial photographs of suburban housing developments and located cultural structures such as houses, roads and driveways.

A similar method using explicit knowledge about the geometry of objects and their spatial relationships in the form of rules was developed by Eklundh (1983). This structured approach iterates to a solution, being similar in nature to a relaxation approach. The main goal, the improvement in classification, was achieved, with an increase of correctly classified regions by a factor of ten (although only a small number of categories were used).

Nicolin and Gabler (1987) developed a knowledge-based system for the analysis of aerial imagery of suburban scenes. Their aim was to develop and implement a complete system, starting from the digitised picture and producing a complete, unambiguous and reliable labelling of essential image segments. They incorporated semantics against a generic model of the scene leading to explicit labelling of the image structures. The semantic network was structured using hierarchical relationships, allowing the structure of complex objects to be given from less complex objects.

In Nicolin and Gablers' system, processing is carried out by a series of modules. Low-level modules determine bright, dark and border areas. Medium-level modules incorporate segmentation algorithms and high-level modules perform object identification. In this system the invocation of any processing method is done by means of a bi-directional control mechanism, which is a mixture of bottom-up (data-driven) and top-down (model-driven) control. The data driven component is activated to generate the first results of the image (e.g. to establish interpretations of image segments). As soon as a sufficient number of image fragments have been identified to form a hypothesis, the top-down component of the control model is invoked.

Nazif and Levine (1984) developed a rule-based system for the low-level segmentation of imagery into uniform regions and connected lines. They differentiate between the low-level segmentation and a high level interpretation by the knowledge available to each. General-purpose models were used which contain knowledge about the imagery and rules for grouping.

A focusing method was employed to concentrate on significant parts of the image. An interesting feature of this technique is that regions and lines are represented by the same data structure. In a further paper, Levine and Nazif (1984) present an optimal set of image segmentation rules. This is a direct extension to image classification and interpretation adding further knowledge modules capturing domain-dependent information.

Kontoes *et al* (1991) used an expert system and supervised relaxation techniques to improve land use mapping in a post-classification refinement process. In order to overcome some of the difficulties encountered in classifying SPOT satellite imagery, statistically separable classes were matched forcibly, with respect to their radiometry, to natural classes perceived and mapped by the photo-interpreter. This is of interest because SPOT imagery suffers from many similar problems to aerial imagery, such as the adequacy of the training sample and the performance of the clustering algorithm. The aim of this was to achieve in a machine assisted photo-interpretation method all the characteristics of a "human remote sensing system". This applied logical reasoning, took account of contextual relationships, assessed similarities and disparities among the various objects, perceived and occasionally estimated bio-geophysical parameters using either absolute or subjective measurements.

However, the method is not yet in operational use and still has many components missing, such as the addition of spectral signatures. The imagery is also of poor quality, with roads, for instance, being removed in the spectral generalisation process. It is also worth noting that it uses only pixel context, rather than segment context.

Matsuyama (1989) describes several expert systems for image processing which he distinguishes from image understanding systems by precisely defining each. These operate at a higher level than image understanding systems, in that they decide which image understanding methods to apply based on user requirements. The main features of an expert system for image processing are summed up thus:

1. Objective. Effective image analysis processes are developed by combining basic image processing functions from program libraries. The expert system should make full use of available image processing techniques.

2. Knowledge Sources. The knowledge used should concern the choice of image processing techniques as well as the image domain knowledge.

3. Goal Specification. A typical goal might be 'find rectangles'. Since there are many possible methods to extract rectangles from an image, knowledge is required about primitive operators in order to select promising ones, and know-how about image processing techniques is needed so as to combine them effectively.

McKeown *et al* (1985) presents a rule-based system, SPAM, that uses map and domain specific knowledge to interpret airport scenes. The domain specific knowledge is based on spatial constraints concerning airport design and observations made about the aerial imagery. This domain specific knowledge cannot, however, be applied to urban imagery due to the lack of codified spatial organisation. An attempt to remedy this situation was the subject of further research by McKeown *et al* (1989). The procedure of knowledge acquisition is automated by analysing the spatial constraints and scene primitives of aerial imagery in general. The tool developed, RULEGEN, compiles spatial and structural knowledge from several sources, stores these as collections of rule schemata, and generates production rules that are executed by the SPAM system.

More recent work by McKeown *et al* (1994) has focused on improving the interpretation accuracy and computational performance of the SPAM system. In addition to this, the design of effective algorithms to extract man-made features has been investigated by incorporating general knowledge of typical scene geometry and specific knowledge about the specific scene of interest. The incorporation of photogrammetric techniques into a building extraction system is described.

2.6.3 Expert System Implementation

Based on the literature regarding expert systems, one may be drawn to two conclusions:

1) the problem of context, due to its imprecise nature, can be solved through the use of heuristics;

2) the most suitable method for capturing heuristic information is an expert system.

However, while section 2.6.2 has indicated that this approach may be used successfully for other types of imagery, such as aerial photography, (Nagao and Matsuyama, 1980; Matsuyama and Hwang, 1990; Nicolin and Gabler, 1987; McKeown *et al*, 1985; McKeown *et al*, 1994) in this section it will be demonstrated that the approach does not lend itself well to digital ATM imagery. In order to evaluate this promising approach a simple expert system was implemented.

There are many systems and languages available that allow the user to easily build expert systems (Gilmore and Roth, 1988). For this initial evaluation, the popular CLIPS tool was used (Giarantino and Riley, 1994). CLIPS is a development and delivery expert system tool which provides a complete environment for the construction of rule-based systems. It was chosen due to its widespread use and support, as well as its similarity to the C language, with which the author is familiar.

The rules in Figure 2.6 were generated by Matsuyama and Hwang (1990) for the classification of urban aerial photographs. As they are of a general nature, it would seem appropriate to apply them to the urban ATM imagery used for this project.

These rules were implemented and applied to the Dudley dataset (see section 3.1.1). However, the results were extremely poor – the following discussion will outline the reasons for this.

We can see that the rules (Figure 2.6) are extremely *ad hoc* in nature, as we would expect, but also are heavily dependent on the resolution of the imagery. However, rule 3 doesn't work for this imagery due to the lack of discernible texture on the roofs of buildings – a typical house in the ATM imagery is covered by 10-12 pixels (at the resolution used).

1. Road = (elongated region) and *not* (vegetation region) and *not* (water region)

2. River = (elongated region) and (water region)

3. House = (high contrast texture area) and *not* (large homogeneous region) and *not* (large vegetation area)

4. Building = *not* (vegetation region) and *not* (water region) and (shadow-making region) and size > threshold), *where the threshold is empirically determined.*

Figure 2.6 The rules used for the expert system (Matsuyama and Hwang, 1990)

Another characteristic of ATM imagery is that it is extremely difficult to geo-rectify (section 2.2.6). This leads to difficulties in making use of the shape characteristics of various structures within the image, such as roads. For instance, warping of the scanlines in the imagery removes the straight edges normally associated with man-made objects. This makes the development of further rules, which might make use of the contextual features of man-made objects, very difficult.

In fact the most serious shortcoming of the expert system approach is in the development of a suitable rule set. If the rules are to be applied to previously unclassified imagery, then a significant number of complex rules need to be developed in order to characterise the properties of certain classes. For instance, a rule to determine which regions are water

regions must take into account the low reflectivity of water in certain bands, the thresholds of which must be empirically determined. So in reality, an impracticably large number of *ad hoc* rules are required just to determine the land-cover, before any attempt to determine land-use is made.

2.7 Feature Extraction

2.7.1 Introduction

Having described some of the approaches used to combine the contextual information, it becomes apparent that low-level methods must be used to extract the bulk of this information. Low level analysis refers to the absence of *a priori* knowledge about scene content (Levine and Nazif, 1984), and is concerned with extracting data from the pixel values in the imagery.

2.7.2. Edge Detection

One of the most valuable sources of information for many Remote Sensing applications is that which is contained in the edges surrounding objects or features of interest.

Unlike the real world, images do not actually have edges. Images have abrupt changes in intensity, but since the aim is to locate edges in the real world via an image, these changes are commonly referred to as 'edges'. All edges are locally directional. Therefore the goal in edge detection is to find out what occurs perpendicular to the edge.

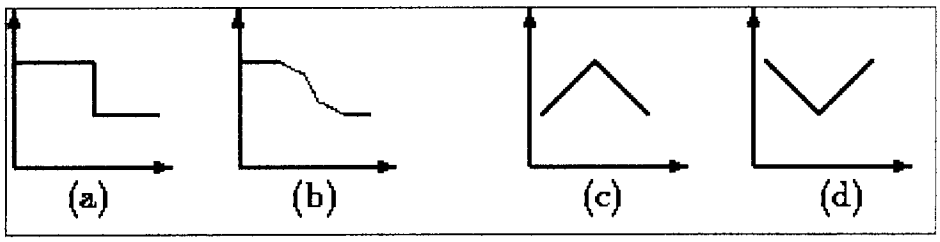


Figure 2.7 Commonly found edge profiles

There is a variety of different types of edge, with intensity profiles ranging from that of a sharp step (a), through a gradual step (b) and on to a 'roof' (c) or 'trough' (d) shaped profile (Figure 2.7). Edges can also be found to be any combination of these.

There is no single image filter that will detect all of the previously mentioned image edges and all their intermediates. This is because edges may have many different profile shapes and spatial scales. There follows a brief review of some of the simpler edge detection methods and a thorough review of some of the more sophisticated methods more pertinent to this study

2.7.3 Simple Edge Operators.

Edge detection at its simplest level involves the use of filters in the form of kernels to emphasise the edge information within the image. The Roberts Cross operator (Roberts, 1965) performs a simple, computationally efficient two-dimensional spatial gradient measurement on an image. The Sobel operator (Sobel, 1970; Danielsson and Seger, 1990) performs in a similar manner, although a larger kernel is used. This results in a slower performance but reduces the noise associated with the Roberts Cross. The Prewitt gradient edge detector (Prewitt, 1970) is again based on the same principle but uses a slightly different kernel. A full description of the these methods can be found in many standard texts, such as Jenson (1986).

2.7.4 Canny Edge Detector

There are several common criteria relevant to edge detector performance, three being described by Canny, 1986. These are:

i) Low-error rate. Good detection is that with a minimum number of false positives and false negatives. It is important that edges that occur in the image should not be missed and that there are no spurious responses.

ii) Good Localisation. The points marked by the detector should be as close as possible to the correct edge position.

iii) Unique response to a each edge.

Based on the above criteria Canny developed an optimal filter. This was shown to be a very close approximation to the first derivative of a Gaussian distribution function. Non-maximum suppression in a direction perpendicular to the edge was applied, to retain

maxima in the image gradient. Finally weak edges were removed using thresholding. Edge contours are processed as complete units, with two thresholds defined. If a contour being tracked has gradient magnitude above the higher threshold then it is still passed as an edge at those parts where the strength falls below this threshold, as long as it does not go below the lower value. This reduces streaking in the output edges.

2.7.5 Laplacian of Gaussian (LoG)

Edge enhancement without regard to the edge direction may be obtained by applying a Laplacian convolution mask to the imagery (Marr, 1982). The Laplacian is a two-dimensional isotropic measure of the second derivative of an image. The edges are found by the highlighting of regions of rapid intensity change. The Laplacian $L(x,y)$ of an image with pixel intensity values $I(x,y)$ is given by:

$$L(x,y) = \frac{\delta^2 I}{\delta^2 x} + \frac{\delta^2 I}{\delta y^2} \qquad \text{Equation 2.1}$$

The image is represented by a set of discrete pixels so a discrete convolution kernel is used to approximate the second derivatives in the definition of the Laplacian. By using kernels to approximate a second derivative measurement, we also introduce a great sensitivity to noise. This is countered by Gaussian smoothing of the image prior to the application of the Laplacian filter, to reduce the high frequency noise component.

Additionally, because the convolution operation is of an associative nature, the Gaussian smoothing filter can be convolved along with the Laplacian filter initially, with this hybrid filter being convolved with the image to achieve the required result. The two-dimensional LoG function centred on zero and with Gaussian standard deviation σ has the form:

$$LoG(x,y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2+y^2}{2\sigma^2}\right]e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad \text{Equation 2.2}$$

It should be noted that as the Gaussian standard deviation is made increasingly narrow, the LoG kernel becomes the same as the simple Laplacian kernels.

By itself, the effect of the filter is to highlight edges in an image. The fact that the output of the filter passes through zero at edges can be used to detect those edges. The use of this technique is discussed in the following section on zero-crossing edge detection.

2.7.6 Zero Crossings

Marr and Hildreth (1980) proposed the use of zero crossings of the Laplacian of a Gaussian (LoG). They developed several criteria that they thought were desirable properties for a feature extractor.

These criteria are:

1. Localisation in space - the location of the edge must be correct.

2. Localisation in frequency - the width of the edge must be spatially precise.

3. Independence of orientation - there is no preference for an edge in a particular direction

It should be noted that criteria 1 and 2 are mutually inconsistent, so there must be a fundamental trade-off when trying to satisfy them both.

The generic approach in edge detection is to apply a first derivative operator, look for extrema, and compare these with a threshold value. Another approach is to look for zero-crossings in the second derivative (Figure 2.8).

The only non-directional (i.e. rotationally independent) linear derivative operator is the Laplacian.

Having satisfied criteria 3, the localisation in space and frequency must be addressed. The Gaussian filter is the optimal filter for a particular trade-off between the localisation in space and frequency. When applied to an image, the Gaussian filter will spread out all the values in the image by the shape of the filter. Of note is that the variance is reciprocal between the

original Gaussian and the Fourier transform of the Gaussian. Therefore the wider the Gaussian gets in space, the thinner its transform gets in frequency.

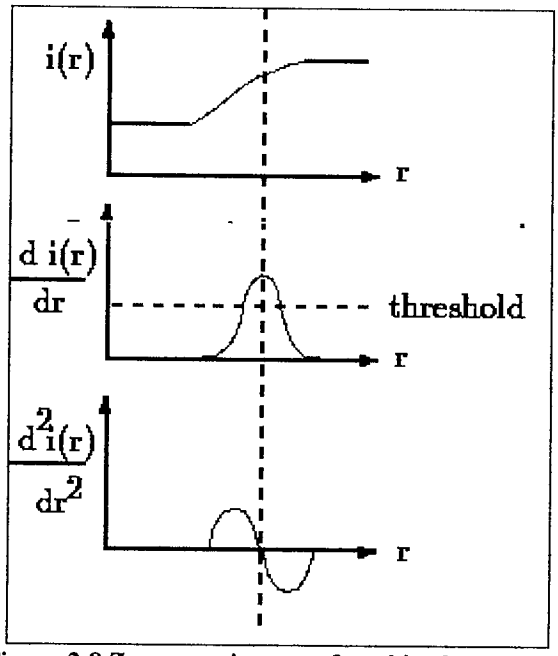


Figure 2.8 Zero crossings are found in the second derivative

Although points where the Laplacian changes sign occur at edges, they also occur at places not normally associated with edges. This is due to the use of non-directional derivatives which result in the measurement of edge responses parallel to each edge, reducing the signal to noise ratio. Connectivity at junctions is good, although corners have rounded edges. The main use of this type of detector is for feature detection, since its desirable property is that the zero crossings always lie on closed contours. The output from the zero crossing detector is usually a binary image with single pixel thickness lines showing the positions of the crossing points.

The GRASS4.1 module *i.zc* is an implementation of this edge detector (Shapiro *et al*, 1993). The x-y pixel extent of the Gaussian filter can be adjusted by the user, with edges representing more gradual changes in pixel values associated with a greater width, and vice versa. The output from the procedure is a thinned (i.e. single pixel width) contour map of the image.

The full procedure is outlined in the following steps:

The Fourier transform of the image is taken.

The Fourier transform of the Laplacian of a two-dimensional Gaussian function is used to filter the transformed image.

The result is run through an inverse Fourier transform.

The resulting image is traversed in search of places where the image changes from positive to negative (or negative to positive).

Each pixel in the image where the value crosses zero (with a value greater than a set threshold value) is marked as an edge and, optionally, an orientation is assigned to it.

2.7.7 SUSAN approach

A recent approach to low level image processing, in particular edge and corner detection and structure preserving noise reduction, based on the SUSAN principle, is described by Smith and Brady (1995). Non-linear filtering is used to define which parts of the image are closely related to each individual pixel, and each pixel has associated with it a local image region which is of similar brightness to that pixel. The local area or Uni-value Segment Assimilating Nucleus (USAN) contains much information about the structure of the image. From the size, centroid and second moments of the USAN two-dimensional features and edges can be detected. The acronym SUSAN arises from the determination of the smallest USAN.

In contrast to the other methods, no image derivatives are used and no noise reduction is needed. This approach therefore represents a significant departure from previously developed feature extraction and noise reduction methods.

2.7.8 Edge Thresholding

A drawback with using edges is that not only do edge detectors extract meaningful and useful edges, but also many other spurious ones which arise from noise and minor changes in intensity values. If all such edges are kept, then the resulting clutter is hard for subsequent processing stages to analyse, while the large number of edge points can seriously degrade computational performance. The alternative is to select a subset of edges for further

analysis, and disregard the remainder. A threshold in the gradient magnitude of the pixels generally achieves this.

Surprisingly little attention has been paid to the problem of automating the thresholding of edges. Most standard image intensity algorithms cannot be applied since they assume bi-modal (or multi modal) intensity histograms while gradient magnitude histograms are more likely to be uni-modal. Moreover, other inappropriate assumptions are often made such as modelling the two populations by Gaussian distributions, and not expecting the sizes of the populations to be too dissimilar. Unfortunately in practice edge thresholding is often done in an *ad-hoc* manner, frequently requiring user tuning of parameters. To enable the building of robust machine vision systems it would be preferable to automate the edge thresholding process so that the systems can adapt to different scenes and imaging conditions without manual intervention.

There are trade-offs in choosing operators to pursue the best overall performance. The Laplacian of Gaussian, according to Yuille and Poggio (1986), should be used because it promises the best zero crossing behaviour in scale space, which presumably will be helpful in later scale space manipulations. However, an isotropic operator like the LoG picks up more noise in the directions where no signal components are present than a directional operator that suppresses the outputs in those directions. As a result, the choice of the LoG is not optimal in terms of signal-to-noise ratio (SNR) and edge localisation accuracy (ELA). The Canny edge detector has better SNR and ELA performances than the LoG. However, the local extrema of the Canny detector's outputs may have unconstrained behaviours in the scale space

In addition, the image produced by most edge detectors is largely lacking in connectivity. This means that subsequent processing must be applied if useful information is to be extracted from the edges. This generally involves some form of shape detector, such as a Hough Transform, which can locate and extract shapes such as circles, ellipses and rectangles, as well as straight lines. Clearly there is a distinct advantage to be gained through the use of an edge detector that produces continuous edges, such as the Zero-Crossing edge detector.

2.8 Straight Line Extraction

2.8.1 Introduction

When attempting to model human vision on a computer, it can be useful to extract straight lines from the image in order to build up information about specific objects contained within the image. This can potentially be applied to remotely sensed imagery, in particular urban imagery, where we might expect to find lines and distinct corners associated with man made objects, for instance buildings, roads and canals. The most common methods for doing this involve the use of the Hough Transform or one of its many variants.

2.8.2 Hough Transform

The Hough Transform was developed by Paul Hough in 1962 and patented by IBM (Ballard and Brown, 1982; Levine, 1985; Boyle and Thomas, 1988). It became, in the last decade, a standard tool in the domain of artificial vision for the recognition of straight lines, circles and ellipses. The Hough Transform is particularly robust to missing and contaminated data. It can also be extended to non-linear characteristic relations and made resistant to noise by use of anti-aliasing techniques.

The Hough Transform is a technique that can be used to isolate features of a particular shape within an image. The main advantage of this method is that it is very robust to noise and is not affected when part of the shape is occluded. In other words, the evidence for the existence of a shape does not require the pixels to be closely associated in the image. An example of this would be a road occluded by many bridges, which would still produce enough evidence for the presence of a straight line in the Hough Transform. A basic explanation of the transform follows.

A general line can be expressed as $y = ax + b$, and also in terms of its parameters, $b = -ax + y$. So, in Cartesian space, any line can be represented by a single point (a,b) in parameter space. Instead of solving the many line equations for all the points in an image we translate them to parameter space using the Hough Transform, thus reducing the number of calculations.

The Hough Transform compares the lines between each pair of points, with a point is placed in the parameter space representing the line. If the same position in parameter space is produced more than once it represents three pixels in line with each other.

As an aid to the interpretation of the points, the parameter space is split into a grid, where each cell represents lines that are almost collinear. As the transform progresses, each accumulator cell counts a tally of points, a high tally representing a line with a lot of pixels along it. This can then be applied to the vectorisation of the image.

The main weakness of the Hough Transform is that noise, or pixels associated with other features in the image, can also contribute to the detection of a certain feature. In the case of lines, a set of pixels that seem to make a well-defined line can end up voting for a skewed line, due to interference from other structures elsewhere in the image.

Implementations of the Hough Transform suffer from two problems related to the use of global edge information in the image to find local edges.

1) Edge elements on collinear but unconnected line segments vote for the same line.

2) The transformation of an image with many noisy edge points or many irregular blobs may contain spurious lines linking distant edge elements. This is because the threshold on peak detection in the Hough space must be set low in order to obtain these lines.

2.8.3 The Muff Transform

Another approach is the Muff Transform (Wallace, 1985). The image is assumed to be bounded by a rectangle parallel to the x and y axes and extending from the origin to the furthest point (xmax, ymax). This bounding box provides the basis for the parameterisation. The two points where the line intersects the perimeter of the bounding box parameterise a line passing through the image. These points are given by their distance along the perimeter of the bounding box, with the distance measures counter-clockwise along the box, starting at the origin. The line has two parameters, S1, and S2, representing the two points where the line intersects the box. To preserve uniqueness of the representation, it is assumed that $S1<S2$. The range of possible values, therefore, is

$$0 \leq S1 \leq S2 \leq (x\max, y\max)$$

The approach used by the Muff transform is to divide the image into a number of smaller rectangles and to compute the transform for each. One immediate advantage of the Muff Transform is graphical, in that it maps back into points on the images bounding rectangle.

Chapter 6 will show the weaknesses with both types of transform when applied to urban images.

2.9 Accuracy Assessment

2.9.1 Introduction

The accuracy of traditional remote sensing techniques, such as photo-interpretation, has generally been accepted as high, without any confirmation. With the advent of more sophisticated digital classification methods, however, there has become a need to assess the accuracy and reliability of these methods, due to their complexity. Since it is often assumed that photo-interpretation provides us with a 'correct' classification, this often provides the reference to which the assessment of digital classification is compared. This assumption is rarely valid, however, and can lead to a poor assessment of the digital classification (Congalton, 1991).

The accuracy of information generated from remotely sensed data has been of interest for many years, but recent widespread evaluation of the results of digital image classification has probably been responsible for the major growth in accuracy assessment. In a way, the focus upon accuracy of digital classifications is unfair, because traditionally the usefulness of manual interpretations has often been accepted in the absence of substantive evidence of accuracy.

There have been few systematic investigations of the relative accuracies of manual and machine interpretations, accuracies of different individuals, accuracies of the same interpreter at different times, accuracies associated with separate pre-processing and classification algorithms, or accuracies associated with different images of the same area.

Most accuracy assessment methods compare the remote sensing derived map with some form of ground truth. Various relations between the two sets of data can be derived providing the user with an idea of the success of the classified map.

2.9.2 Accuracy Assessment Methods

The oldest and most frequently used method of evaluation is a qualitative assessment. The only method available is a simple visual comparison between two maps to determine whether similar features are present. This is, by its nature, inexact, but nevertheless very important. Chapter 3 will examine this further.

Another method of evaluation is simply to compare the two maps in respect to the areas that match when the two maps are superimposed. The result of such a comparison is to report the spatial proportions of the two patterns that match. These values report the extent of the agreement between the two maps in total area in each category, but do not take into account compensating errors in misclassification that can cause this kind of accuracy measure to be inaccurate itself. This method is sometimes referred to as 'non-site-specific accuracy' because it does not consider agreement between the two maps at specific locations, but only overall agreement.

Booth (1989) and Congalton (1991) present comprehensive reviews of literature concerned with assessing the accuracy of classified remotely sensed imagery. Some of the techniques more relevant to our study are presented and discussed below.

Hay (1979) identified five problems that arise when testing the accuracy of qualitative characteristics determined from remotely sensed data.

1. What proportion of all the sample predictions proved to be correct?

2. What proportion of the sample predictions of single categories proved to be correct?

3. What proportion of land truly (in the ground truth sense) in a category is correctly predicted?

4. Is the net effect of questions 2 and 3 for predictions to overestimate or underestimate a given category?

5. Are the errors randomly distributed?

Point 5 may arise in a multi-category case where some categories are acknowledged to be very similar; in such a case, mis-classification between similar categories may be high, although overall accuracy is quite high.

A very large manual survey of the imagery will answer all the problems above, but the purpose of the classification algorithm being tested is to avoid classifying the entire image manually. To this end Hay demonstrates that any class sample of less than 50 pixels will be an unsatisfactory guide to the true error rate, and recommends that a sample size of 50-100 pixels is used. This assumes that all categories are of equal interest.

Jenson (1986) identifies two procedures for measuring overall and site specific accuracy. In order to determine the overall accuracy it is necessary to ascertain whether the map meets or exceeds some predetermined classification criteria. For earth resource management the overall accuracy of land use maps should be 85%, with approximately equal accuracy for most categories (Jenson, 1986; Campbell, 1983). The agreement between the two maps in total area in each category is derived. The measurements do not usually evaluate compensating errors that occur in the various categories.

Site-specific error analysis compares the accuracy of the remote sensing derived map pixel by pixel with the assumed true land use map. If test locations in the study are identified and labelled prior to classification and are excluded from the training stage they can be used to evaluate the accuracy of the classification map. To do this additional ground truth is required for these test site areas. It is important to determine how many pixels are necessary in each test site class along with the method of identifying the location of the test sites. Many statistical tests require that locations be randomly selected prior to classification so that the analyst does not bias their selection in any way.

Two methods for determining the individual category accuracies can be determined from the error matrix (Story and Congalton, 1986). They are:

1) The number of correctly classified samples in a category divided by the number of category samples in the reference data.

2) The number of correctly classified samples in a category divided by the total number of samples classified as that category.

The first method measures the 'error of omission' and has been referred to as the producer's accuracy, because the producer of the classified image is interested in how well a specific area on the Earth can be mapped. The second method measures the 'error of commission', and has been referred to as the user's accuracy, because a map user is interested in the reliability of the map, or how well the map represents what is really on the ground. It is obvious that these measures of accuracy can vary massively.

Although these measures may seem simple, it is critical that they both be considered when assessing the accuracy of a classified image. It is often the case that only one measure is reported. This value can be extremely misleading, as has been seen in section 2.2.

There is a requirement for a single value (or coefficient) that adequately represents the accuracy of thematic classification. As well as requiring a single figure to measure the overall accuracy of the map, it is desirable to calculate the accuracies of the individual categories within the map. One way of doing this is to address the problem as a measure of agreement between classification and verification. This can be measured using the Kappa coefficient of agreement (Rosenfield and Fitzpatrick-Lins, 1986).

The estimate of Kappa (K) is the proportion of agreement after chance agreement is removed from consideration:

$$K = (p_o - p_c)/(1 - p_c) \qquad \text{Equation 2.3}$$

where

p_o = proportion of units which agree,

p_c = proportion of units for expected chance agreement.

When the obtained agreement is chance agreement, K=0. Positive values of Kappa occur from greater than chance agreement; negative values of Kappa are from less than chance agreement. The upper limit of Kappa (+1.00) occurs only when there is perfect agreement.

Kappa coefficient values give a significantly lower value for classification accuracy than the total percentage correct classification. It uses information in the classification error matrix resulting from errors by commission and of omission. For these reasons, Rosenfield and Fitzpatrick-Lins (1986) recommend that the coefficient of Kappa be adopted by the remote sensing community as a measure of accuracy for thematic classification as a whole, and for the individual categories.

Ma and Redmond (1995) developed the Tau coefficient, which measures the improvement of a classification over a random assignment of pixels to groups. The difference between Tau and Kappa is that Tau is based on *a priori* probabilities of group membership, whereas Kappa is based on *a posteriori* probabilities. As a result they claim that Tau, as well as being easier to calculate, better adjusts the percentage agreement than Kappa.

Another method is a 'fit for use' method, whereby the classification is tested by passing the land cover data through a urban drainage model and comparing the results with those from hand classified data, using the same model (Blagovic *et al*, 1998). Though this appears the most realistic test what we are in fact doing is testing the model/land cover data together as an inseparable pair.

2.9.3 Conclusion

There are a great many considerations to be made when attempting to provide an accuracy figure for a digital classification. It is commonly stated, for instance, that land use maps must be 85% accurate to be of any use to the earth resource community (Campbell 1983; Jenson, 1986). However, what this figure means and how it relates to those derived from accuracy assessment techniques is rarely discussed. This is of critical importance to remote sensing, since the quantitative analysis of our techniques is what provides confidence to the end users of our technology. As Congalton (1991) points out, the days of casual assessment of the "it looks good" nature are no longer acceptable. A classification is not complete until it

has been assessed, and it is this assessment which validates the information that remote sensing has provided.

As a consequence, it is suggested that as much time and effort is concentrated on this often neglected area as on the classification itself. An assessment of the accuracy should be as comprehensive as possible, allowing true comparison with other classifications as well as restoring confidence in the method itself.

2.10 Urban Drainage Modelling

2.10.1 Introduction

The aim of this project is to provide a high accuracy land-use classification for input into an urban drainage model. As a result, it is useful to assess what type of data input is required by the various drainage models available, in particular the land-use information. Therefore the purpose of this section is to provide an overview, rather than a rigorous critique, of these models from this perspective.

2.10.2 Urban Drainage

There is a general tendency worldwide for the increase in the percentage of the population living in urban areas. This causes not only an increase in the density of the population but also the urbanisation of former sub - urban and rural areas. Demand for a comprehensive drainage system in such areas is evident and the system becomes more and more complex as the urbanised area increases.

Rehabilitation of the sewer network may consequently be required. There are two fundamental reasons why this may be the case. The first of these is that the sewer may be structurally inadequate, due to ageing, and have poor drainage qualities as a result of poor design of the existing system. The second reason is that the system may have an inadequate capacity. This may be because:

1. In the design phase the water quantity was underestimated (bad modelling).

2. Newly urbanised areas, not accounted for during the design phase, have been connected to the existing system without its reconstruction.

3. Land use within the existing catchment has changed because an increase in urbanisation has caused an increase in the percentage of impermeable areas, resulting in increased runoff.

4. Environmental requirements, public health concerns and other statutory reasons.

2.10.3 Urban Drainage Models

There are several popular urban drainage models in use today. Different models are preferred in different countries, although they all broadly follow the same principles. They range from MOUSE (Mark, *et al*, 1997; DHI, 1995) and BEMUS (Maksimovic *et al*, 1995) to WALLRUS (Wallingford Software, 1991) and SWMM (Jewell and Adrian, 1978).

A explanation of the principles behind the design of these models and a brief description of each follows.

The BEMUS model is a physically based deterministic model that consists of several modules. Each module deals with one of the phases of flow processes i.e. surface detention, infiltration, surface runoff, flow along gutters and flow in the underground network. All surfaces belonging to a sub-catchment are divided into sub-surfaces such as roofs, other impervious surfaces (streets, pavements, parking lots etc.) and pervious surfaces (parks, gardens etc.). Contributions to runoff is the whole rainfall from roofs, and effective rainfall from other impervious surfaces (Maksimovic, 1993).

The most widely used package in the UK is the Wallingford Storm Sewer Package (WASSP), which was developed by HR Wallingford. The original WASSP (as used by Finch *et al*, 1989) consisted of interrelated computer programs to assist engineers in the hydraulic design of storm sewer networks in the UK. It included software to assess the performance of an existing or designed system. This has evolved into the WALLRUS package (Wallingford Software, 1991) which is applicable to a wide range of climatic types, and includes new facilities such as spatially varied rainfall, free surface backwater effects, and sediment depths.

The US Storm Water Management Module (SWMM) is a comprehensive model for continuous and single-event simulation of runoff quantity and quality. It applies to all hydrologic, hydraulic and water quality aspects of urban drainage, including storm sewers,

open channels, combined and sanitary sewers. It covers dynamic storms, snowmelt, pollutant build-up, wash off and transport, dynamic routing and surcharging in sewer networks and a variety of other systems.

MOUSE is a software package for the simulation of surface runoff, flows, water quality and sediment transport in urban catchments and sewer systems. It can be applied to any type of pipe network with alternating free surface and pressurised flows. It contains computational models for surface runoff in urban catchments and hydrodynamic analysis of flows in pipe networks. A variety of add-on modules are available, including MOUSE GIS, for network data modelling and results display within the ArcView environment.

In almost all of the models the percentage of paved and roof areas in each sub-catchment must be determined. If the development is uniform, the paved and roof area in a selection of the sub-catchments can be measured, with the same percentages of development applied throughout the catchment. The calculation of runoff and flows is particularly sensitive to the impervious area, making this a time consuming procedure requiring high accuracy.

Paved areas and pitched roofs must be treated separately. Pitched roofs give a much faster response to the sewer system, creating more likelihood of flooding. The difference this makes to the results is small, but may be significant if more than 30% of the impervious area is pitched roof connected directly to the sewer system (Wallingford Software, 1991). Flat roofs are treated as paved areas. Roof area which is not drained directly into the sewerage system but which drains across permeable areas should not be included, as part of the roof area but should be included in the permeable area.

The WALLRUS package (Wallingford Software, 1991) has an option that allows the user to specify the characteristics of permeable regions in detail. For example, the user can specify that the permeable area be categorised as high, medium or low-density housing, or open area. This demonstrates the move towards more sophisticated drainage models and emphasises the need for higher detail land-use classification of urban imagery.

2.11 Conclusion

This chapter has examined, in depth, the principles and techniques behind existing classification methods, and a critique of these when applied to the field of urban drainage has also been undertaken. Consequently it has outlined the motivating factors for the development of a new classification routine. Further to this, several common methods for assessing the accuracy of these existing classification techniques have been discussed. The application of these accuracy assessment techniques will allow a more detailed critique of the shortcomings of these classifiers.

The literature has shown that, for highly complex data, context is an important factor in determining the correct land use. It has also been shown that urban imagery is indeed highly complex. However, it is clear that there have been few attempts to make use of context in classification, due to the difficulties encountered when attempting to represent it. Where contextual features have been incorporated into the classification process it has been done so through the use of expert systems applied to scanned aerial photography. This is of considerably higher resolution than ATM imagery and does not suffer from the same geo-rectification problems. These problems mean that effective rules are difficult to generate – for instance, shape – based rules cannot be reliably defined (due to rectification errors), and texture-based rules are not applicable (due to image resolution). Consequently we have seen that expert systems are inappropriate for incorporating context when classifying imagery of the type used in this project.

This project will attempt to rectify the failings of traditional classification techniques, through the development of a new classifier particularly focused on the urban drainage application. Contextual features in the imagery will be used to enhance both the qualitative and quantitative aspects of the classification, with a discussion of both improvements being presented along with imagery and numerical data.

The remainder of this thesis will concentrate on describing the research behind the development of the classifier, along with a detailed analysis of its performance when compared to existing ones

Chapter 3: Data Management and the Classification of Urban Imagery.

3.1 Introduction

In this chapter the imagery used in this project is introduced and analysed, leading to an examination of some of the classification methods available. The purpose of this is to assess where the various classification schemes succeed, and where they fail. This will illustrate areas where improvements are necessary that can then be developed and incorporated into the new classification scheme proposed in Chapter 5. A comparison between these existing schemes and the technique developed in this project can then be made, from both qualitative and quantitative viewpoints.

3.1.1 The Study Area: Dudley, West Midlands.

The imagery selected for this project covers a region of the West Midlands near Dudley, a deeply industrial urban area. This was selected because a variety of land use classes are represented and the complex nature of the imagery was considered representative of urban imagery. The image contains heavily industrial zones as well as large residential ones. The imagery is typical of the type used for urban drainage modelling.

Urban areas are extremely difficult data sets to deal with. They are dynamic, in the sense that the land usage can change rapidly in a short time-scale. Buildings such as warehouses can be constructed in a short time, and become derelict and disused equally as quickly. This results in maps fast becoming outdated. Furthermore, these buildings tend to be large and densely packed, producing shaded regions between them that are difficult to classify spectrally (they have a similar spectral response to water), and contextually (uniform elongated regions similar to segments of canals and rivers).

As mentioned in Chapter 2, many similar materials are used to produce an array of diverse land uses, which have similar spectral responses. This causes a multitude of problems to classification methods that rely on per-pixel analysis alone, as demonstrated later in this chapter.

3.1.2 Daytime ATM Data

The Department of Civil Engineering at Aston University has several suitable data sets, archived and in current use, that cover urban areas. The data used in this study was acquired through the National Environmental Research Council (NERC) as part of the Airborne Remote Sensing Programme of 1992 (Ellis, 1997; Wilson, 1995). A flight containing several passes took place on 25/6/92, starting at 9.30 in the morning, from which the imagery used in this project was taken. The image was selected because it was clear that it was representative of many urban areas while containing a broad range of land usage. The region covered by this image is shown in Figure 3.1. The accompanying colour stereo photography was of approximately 1:5000 scale, with each photograph having ground coverage of approximately 1.5 km by 1.5 km.

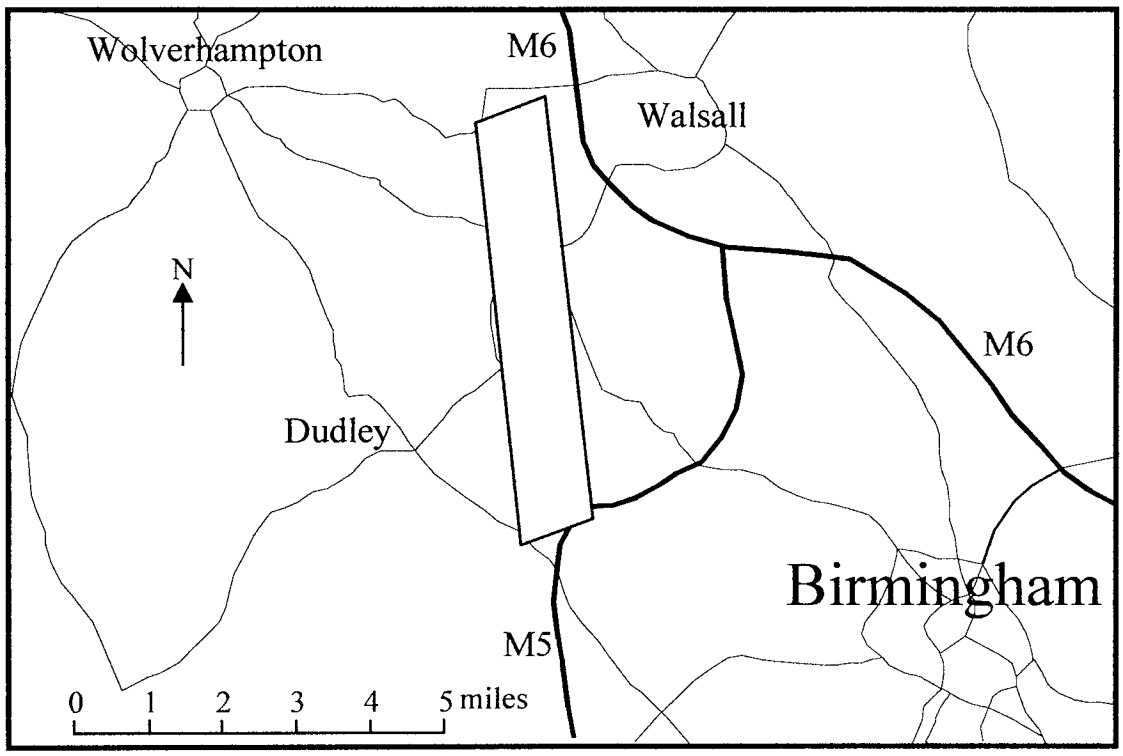


Figure 3.1 The area covered by the aerial imagery.

Prior to any manipulation of the data the imagery was displayed and examined so that any problems could be highlighted and resolved. The detail regarding the image bands can be found in Table 2.1. Band 1 was found to be of very poor visual quality, due to excessive noise attributable to Rayleigh scattering (Lillesand and Kiefer, 1994). This is a common problem in the 0.42μm - 0.45μm range (Finch *et al*, 1989). The remaining visible bands (2, 3, 4, and 5) were of high quality, as were the three near infra-red bands (6, 7, and 8). Band 9 had a similar spectral response to the near infra-red bands (see Figure 3.3), but contained a substantial amount of noise. Bands 10 and 11 were once more of high quality, with band 12 being identical to band 11, but with half the gain. It was therefore decided that bands 1, 9 and 12 were unusable and were therefore discarded.

3.1.3 Dawn Thermal ATM Imagery

Dawn thermal imagery was acquired over the same region on 8th April 1994. The intention was to use this imagery in conjunction with the daytime ATM imagery, since recent research (Elgy *et al*, 1995) has indicated that it is particularly good for discriminating between road and roof - this is demonstrated by Figure 3.2. The addition of this imagery to the daytime imagery should provide a particularly rich dataset for classification. However, difficulties encountered in the co-registration of the imagery meant that it could not be used with the daytime imagery.

Some work with this imagery alone, however, has proved successful (Elgy *et al*, 1995). The nature of dawn thermal imagery is such that successful discrimination between regions of road and roof can be made. Research into methods to resolve the co-registration of the imagery is currently part of an ongoing project at Aston University (Elgy *et al*, 1998).

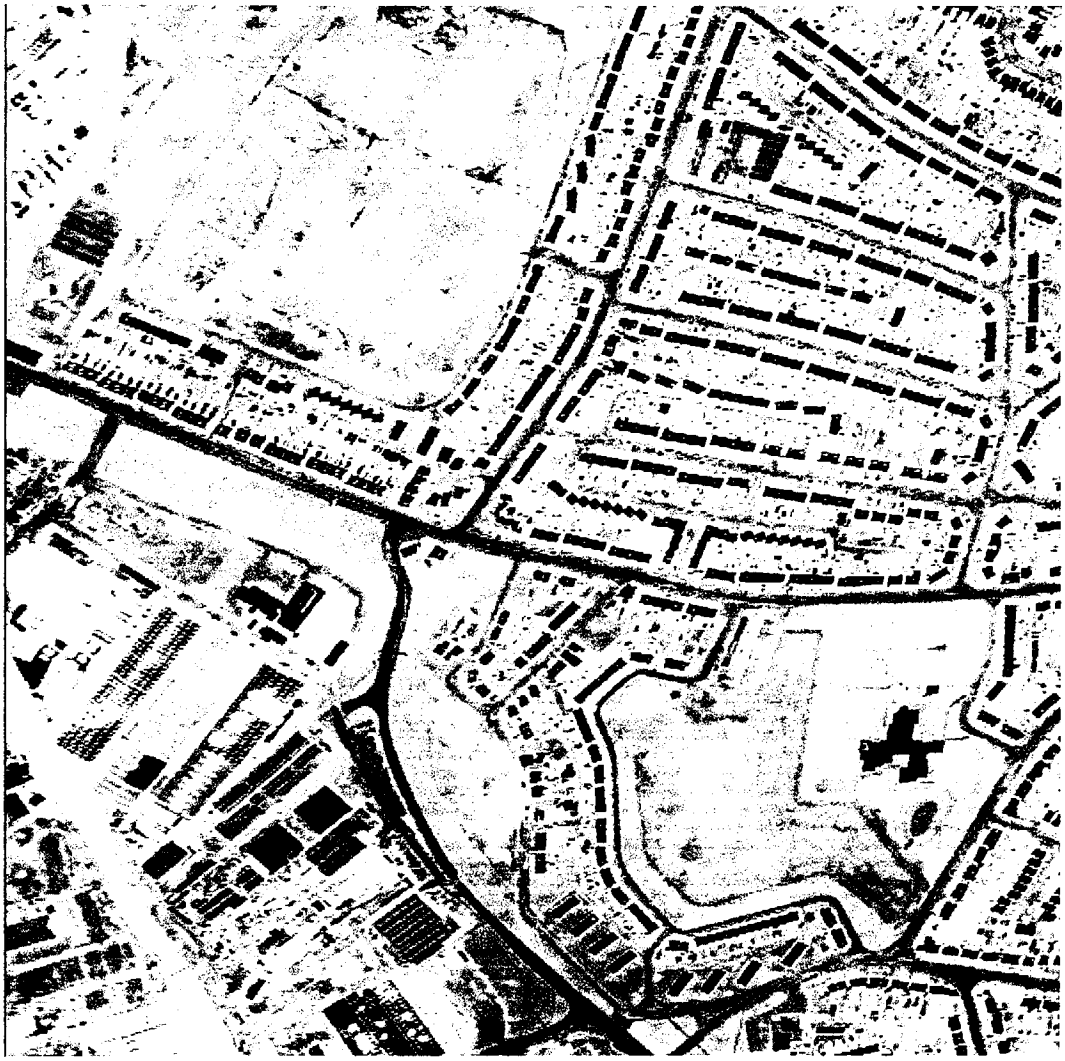


Figure 3.2 Dawn Thermal Imagery.

3.1.4 Pre-processing

The raw imagery used in this project was found to suffer from an inconsistent intensity variation across the flight path. The image brightness either increased towards one side, or exhibited an increase or decrease down the centre of the image. This variation in pixel value across the image in know as the 'scan angle effect', and is related to the wide scan angle of the ATM (Barnsley and Barr, 1993).

The imagery was consequently radiometrically corrected to compensate for this disparity. For details of the correction specific to the imagery used in this project the reader is referred to Ellis (1997). Further documentation of the scan angle effect can be found in Danson (1986) and Barnsley and Barr (1993).

Section 2.2.6 detailed the problems associated with the geo-rectification of line-scan imagery. As a consequence it was decided that classified maps of the raw imagery would be examined, since at this stage it is the performance of the classifier and not the geo-rectification procedure that we wish to assess. However, for reference to other data-sets, geo-rectification is essential. Therefore in Chapter 7, when such a comparison is made, a corrected classified image is presented. For that image, the geo-rectification took place after the classification using a first-order polynomial fit, using nearest neighbour resampling (categorical data). While higher order polynomials are generally required for line-scan imagery to account for the various distortions, the lack of a large enough number of reference points meant that the first order polynomial performed better. Ongoing work to remedy the geo-rectification problem is detailed in Elgy *et al*, 1998.

3.2 Band Selection

3.2.1 Introduction

While we have already been forced to discard bands 1, 9 and 12, it is useful to examine the histograms of each band and compare them to one another. This gives us a qualitative 'feel' for the imagery and gives an indication of the pixel distribution of the data (see Figures 3.3 – 3.5). It is important to note that using more bands for a classification does not necessarily yield better results (Swain and Davis, 1978). The information content of an important band can be contaminated by the insignificant data in another band, this problem being particularly acute when large numbers of similarly distributed bands are used.

3.3 Training Area Determination

3.3.1 Introduction

The supervised per-pixel classification provided by the SMAP estimator and the Maximum Likelihood method rely on the accurate determination of the spectral response pattern for each land cover category of interest. The generation of the training data is crucial to the success of the classification process, although the selection of representative training areas is very much an *ad hoc* procedure. The selection relies heavily on a thorough knowledge of the geographical region of interest.

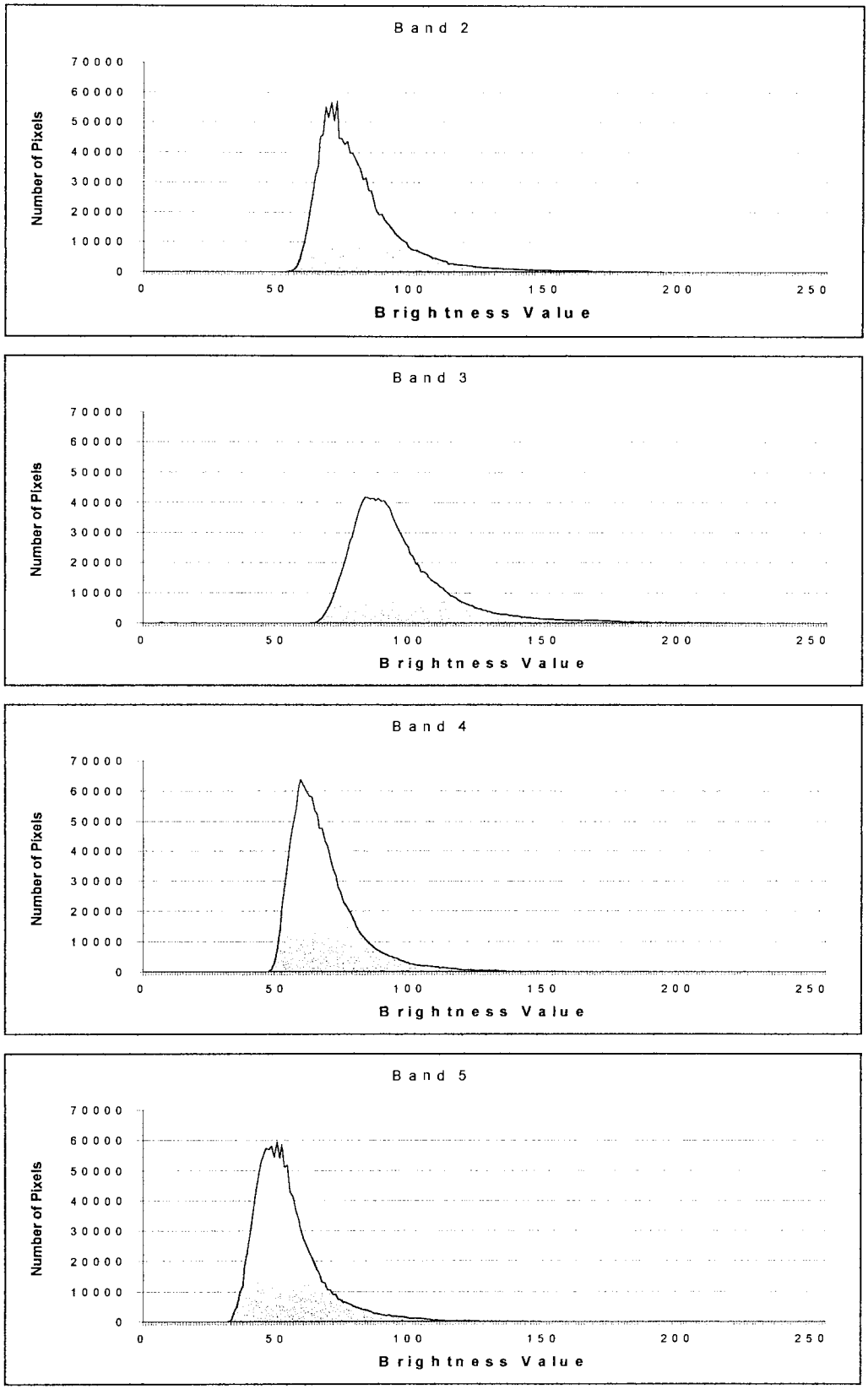


Figure 3.3 Pixel Distribution Histograms for Bands 2, 3, 4 and 5

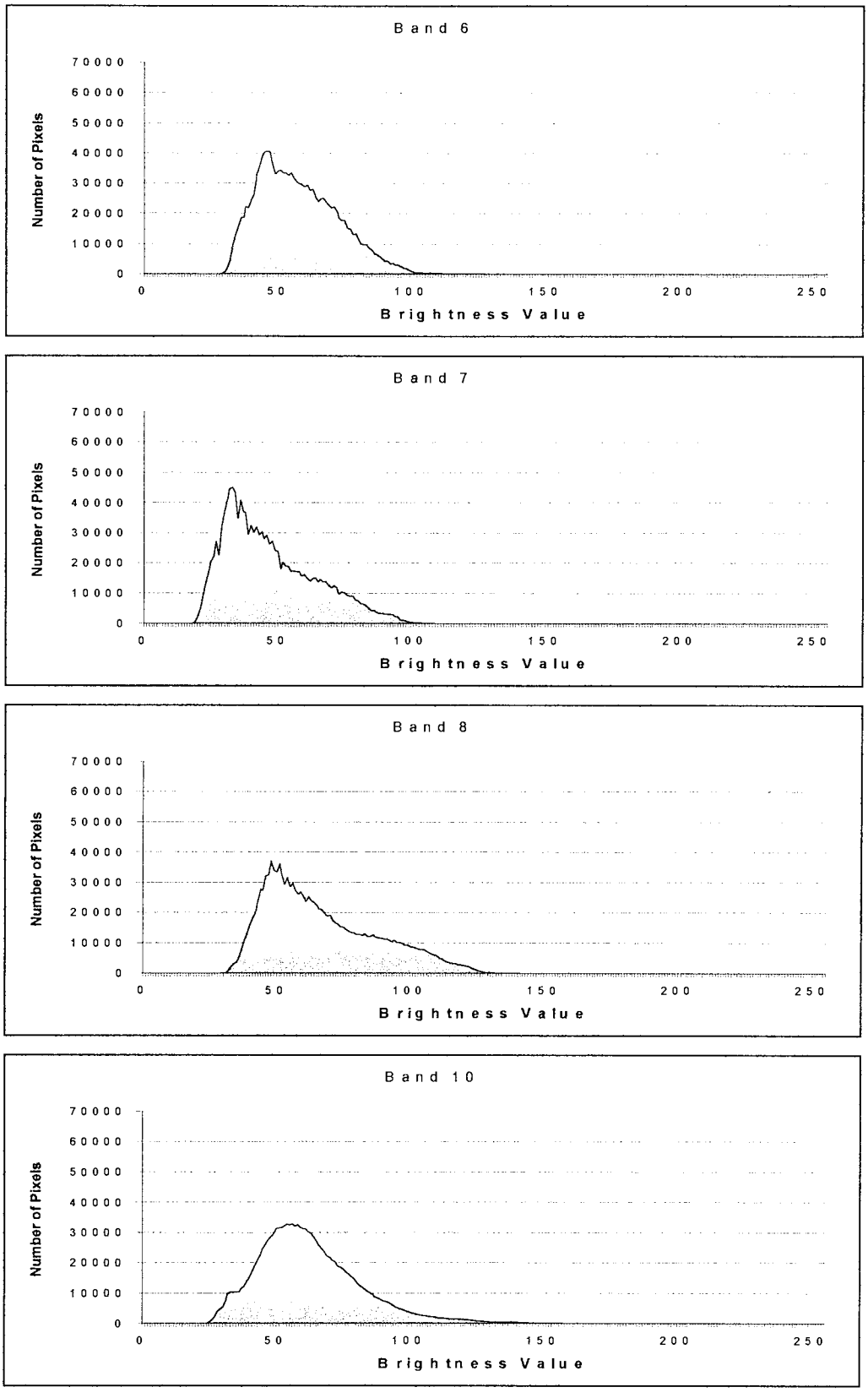


Figure 3.4 Pixel Distribution Histograms for Bands 6, 7, 8 and 10

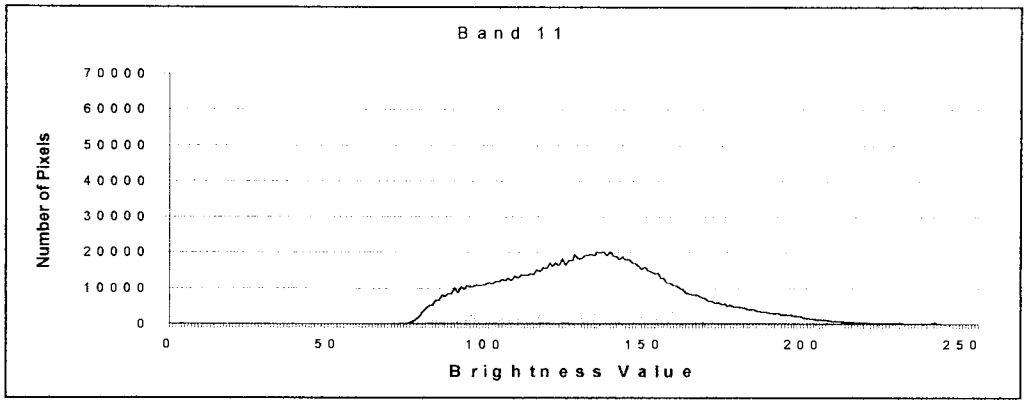


Figure 3.5 Pixel Distribution Histogram for Band 11

In addition, the spectral characteristics of the features under analysis must be examined and evaluated. The statistical data generated by this method is domain specific. There are several ways of establishing the spectral response patterns of the training areas. Direct field measurements are a good source, although the correlation between these and the image data can be difficult due to effects of illumination, atmospheric interference, platform stability, etc. Generally, it is preferable to establish the category response patterns using the image data itself. The training patterns will precisely correspond only to the individual image data-set from which it has been generated, however, compounding the lack of portability of the training data. Thus for each subsequent classification the training data must be recompiled.

The training process begins with the selection of training areas that are representative examples of each information category to be interpreted. These areas are normally selected by consulting the reference data sources, such as topographic maps, aerial photographs and ground observation. The corresponding features in the image are identified and assigned to a class.

It is important to analyse several training sites throughout the scene when selecting areas for inclusion in the training set. Dispersion of the sites increases the chance that the training data will be representative of all the variations in cover type throughout the scene. The digital values of the pixels in each selected feature define the spectral response patterns of the

various land use categories in either a parametric, or non-parametric manner. In other words, they are either based on simple descriptive parameters of the normal distribution, such as mean and variance, or attempt to retain the non-normality found in training data. Consequently, the determination of the training areas is a time-consuming affair, since each one must be analysed thoroughly and digitised manually. The training areas selected for the per-pixel classifications performed in this project are shown in Figure 3.6.

The separability of the spectral response patterns for each class must also be evaluated, whilst at the same time checking that the data are normally distributed if the classifier used depends on the assumption of normality (i.e. Maximum Likelihood method). In Figure 3.7 the distribution of a region of road selected from the training data is shown.

If a category appears to be of a bi-modal distribution, it indicates that the training site is composed of two subclasses with slightly different spectral characteristics. The classification accuracy can be improved if the subclasses are each treated as a separate category, or a suitable multi-modal classifier is used.

The use of a co-spectral plot can conveniently be used to assess the spectral separation between categories. This shows the mean radiometric response of each category and the spread of the distribution. This type of plot shows the potential overlap between category response patterns, and gives an indication of which combination of bands might be best suited for discrimination. Figures 3.8 – 3.10 show these plots for the training classes selected for each band used. These indicate that the major urban classes overlap each other significantly, particularly the 'factory roof', ' car park', and 'industrial forecourt' classes. The large standard deviation of these classes, and their relative spectral heterogeneity, means that the potential for an accurate automatic classification is likely to be limited. In order to improve the classification accuracy, certain spectrally similar classes should therefore be merged (see section 3.4).

With this many classes identified in this type of imagery, per-pixel classification techniques are likely to be inadequate for mapping urban land-use. An alternative to merging the classes is to incorporate contextual information, or to enhance the data with some measure of texture.

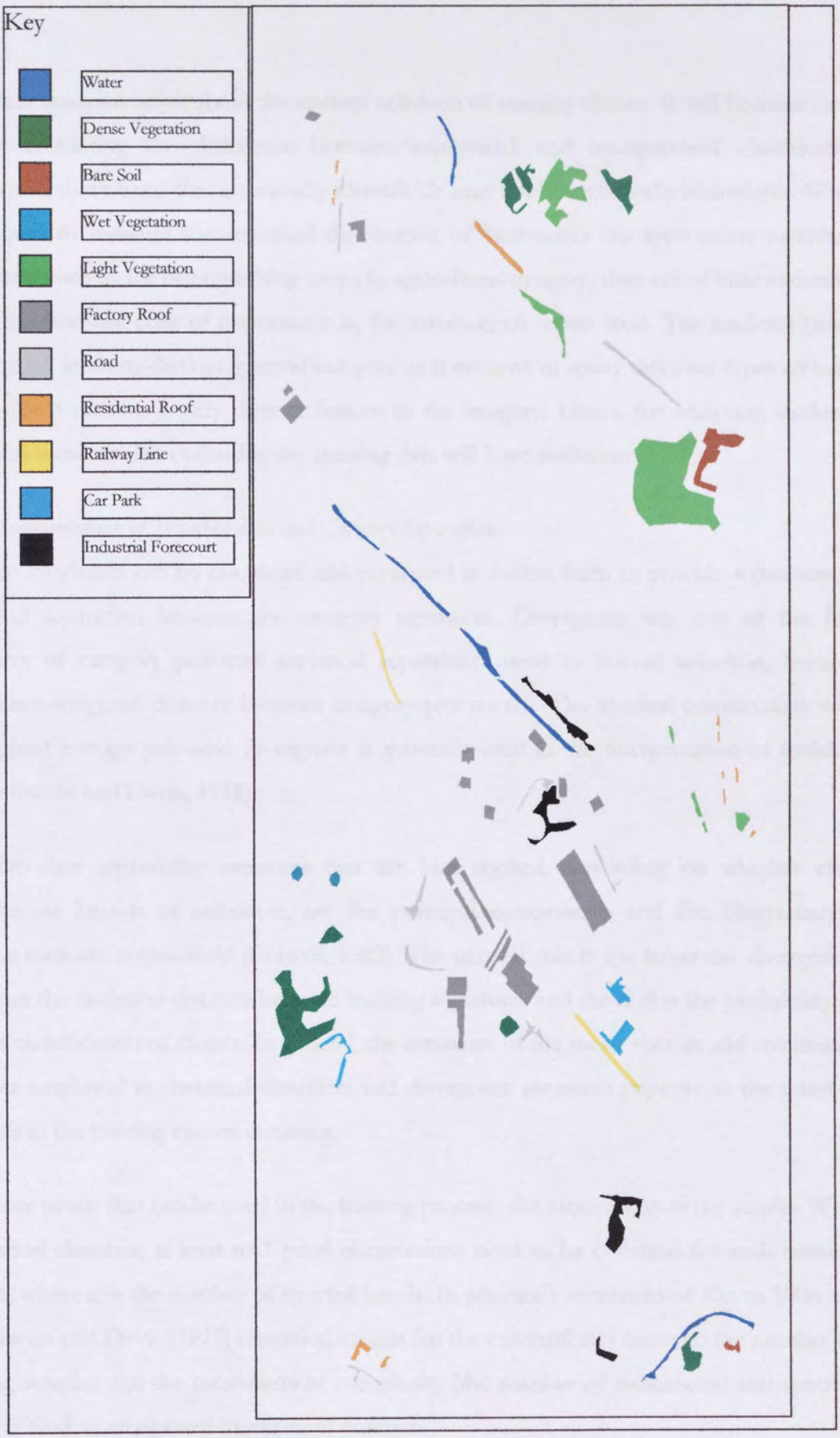


Figure 3.6 The Training Areas used for the supervised Classifications.

A further question arises about the correct selection of training classes. It will become clear, when considering the difference between supervised and unsupervised classification techniques, that classes that are visually identifiable may not be statistically identifiable. While techniques to measure the statistical distribution of land-cover are appropriate in certain situations, such as for distinguishing crops in agricultural imagery, they are of little assistance when the land-use class of importance is, for instance, an urban road. The land-use (road) will not fall into any distinct spectral category as it consists of many different types of land-cover, yet it remains a very distinct feature in the imagery. Hence for analysing land-use, statistical measures for evaluating the training data will have limitations.

3.3.2 Substantiation of Training data and Category Separation

All pairs of classes can be calculated and presented in matrix form to provide a measure of statistical separation between the category signatures. Divergence was one of the first measures of category pair-wise statistical separability used in feature selection, being a covariance-weighted distance between category-pair means. The spectral combination with the highest average pair-wise divergence is generally used in the interpretation of multiple classes (Swain and Davis, 1978).

The two-class separability measures that are best applied, depending on whether class statistics are known or unknown, are the principal components and the Bhattacharyya distance measure respectively (Colwell, 1983). The general rule is the larger the divergence, the larger the statistical distance between training signatures and the higher the probability of correct classification of classes. In general, the estimates of the mean vectors and covariance matrices employed in statistical classifiers and divergence measures improve as the number of pixels in the training classes increases.

The more pixels that can be used in the training process, the more accurate the results. With a statistical classifier, at least n+1 pixel observations need to be collected for each training pattern, where n is the number of spectral bands. In practice a minimum of 10n to 100n are used. Swain and Davis (1978) identified criteria for the compatibility between the number of training samples and the measurement complexity (the number of radiometric and spectral levels) as finding an optimal (maximum) number.

One can see, however, that the selection of training areas is not a trivial task and, although it is by no means as time-consuming as manual classification, it occupies most of the time taken in the supervised classification procedure. The use of unsupervised classification methods does not require the identification of training areas, but instead requires what is in effect a post-classification determination of what the classes actually represent. This, once more, is not a trivial task, since many areas that the interpreter may consider to belong to a single class in fact belong to several. As a consequence, this raises questions about the normality of the data itself.

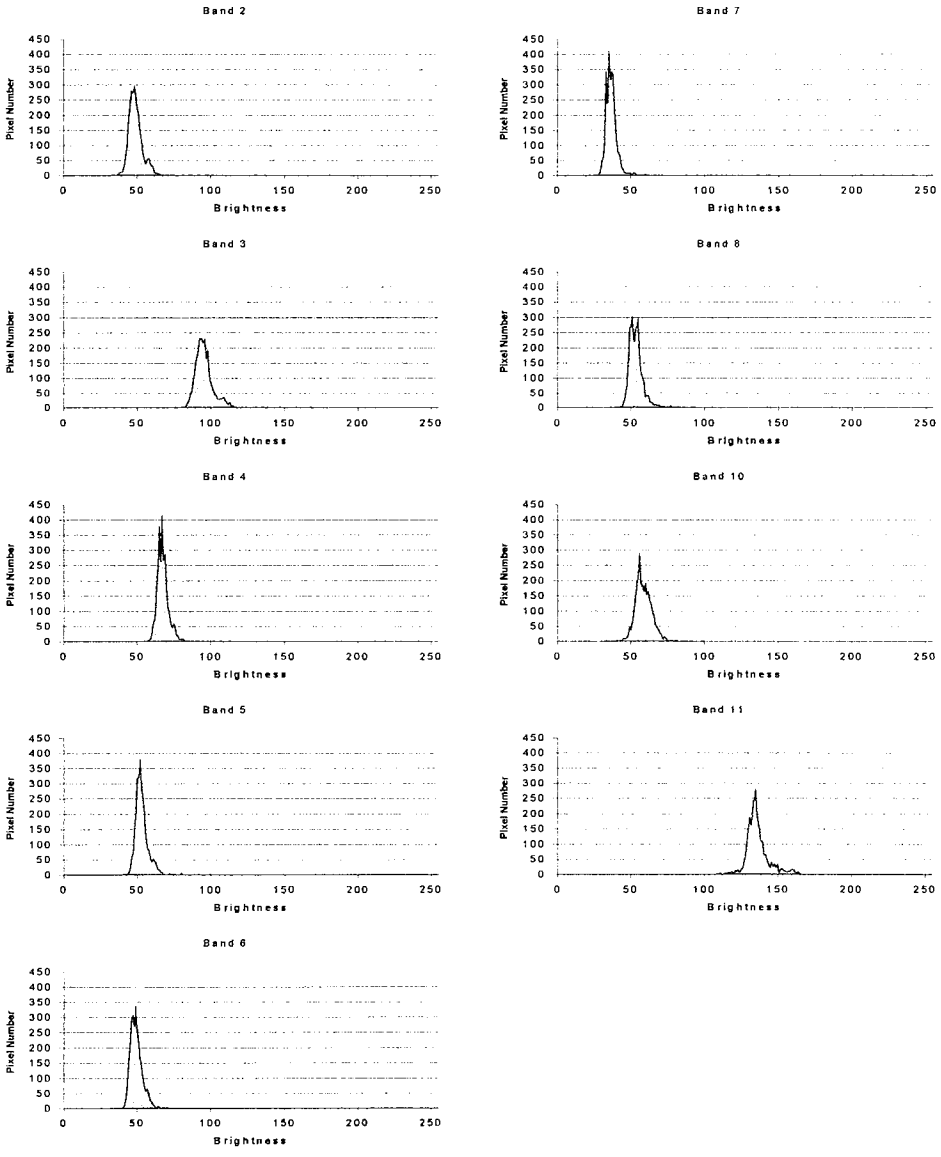
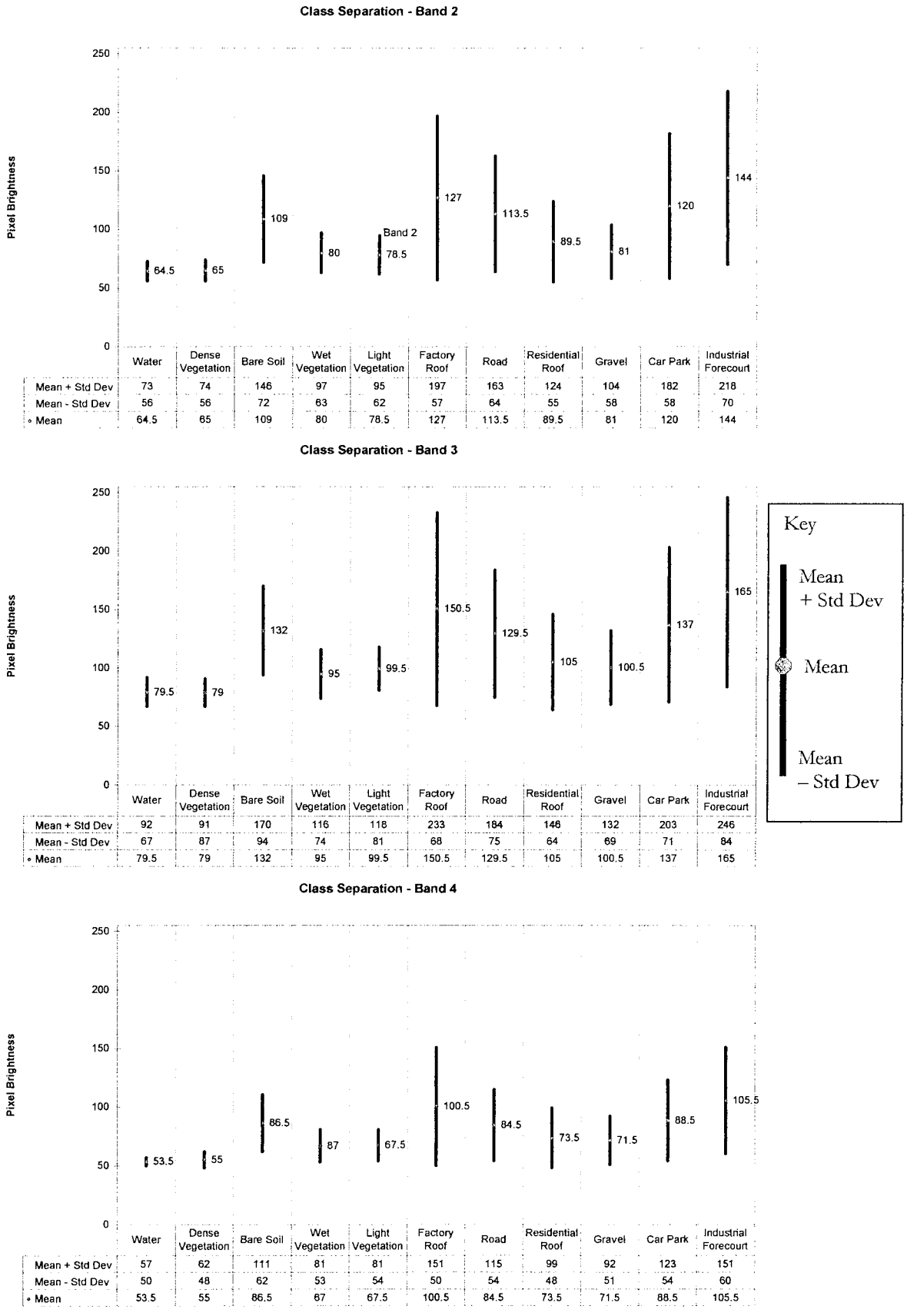


Figure 3.7 Pixel distribution histogram for bands 2, 3, 4, 5, 6, 7, 8, 10, and 11 for a training region of the 'road' class



	Water	Dense Vegetation	Bare Soil	Wet Vegetation	Light Vegetation	Factory Roof	Road	Residential Roof	Gravel	Car Park	Industrial Forecourt
Mean + Std Dev	73	74	146	97	95	197	163	124	104	182	218
Mean - Std Dev	56	56	72	63	62	57	64	55	58	58	70
◦ Mean	64.5	65	109	80	78.5	127	113.5	89.5	81	120	144

	Water	Dense Vegetation	Bare Soil	Wet Vegetation	Light Vegetation	Factory Roof	Road	Residential Roof	Gravel	Car Park	Industrial Forecourt
Mean + Std Dev	92	91	170	116	118	233	184	146	132	203	246
Mean - Std Dev	67	87	94	74	81	68	75	64	69	71	84
◦ Mean	79.5	79	132	95	99.5	150.5	129.5	105	100.5	137	165

	Water	Dense Vegetation	Bare Soil	Wet Vegetation	Light Vegetation	Factory Roof	Road	Residential Roof	Gravel	Car Park	Industrial Forecourt
Mean + Std Dev	57	62	111	81	81	151	115	99	92	123	151
Mean - Std Dev	50	48	62	53	54	50	54	48	51	54	60
◦ Mean	53.5	55	86.5	67	67.5	100.5	84.5	73.5	71.5	88.5	105.5

Figure 3.8 Class Separability for Bands 2, 3 and 4

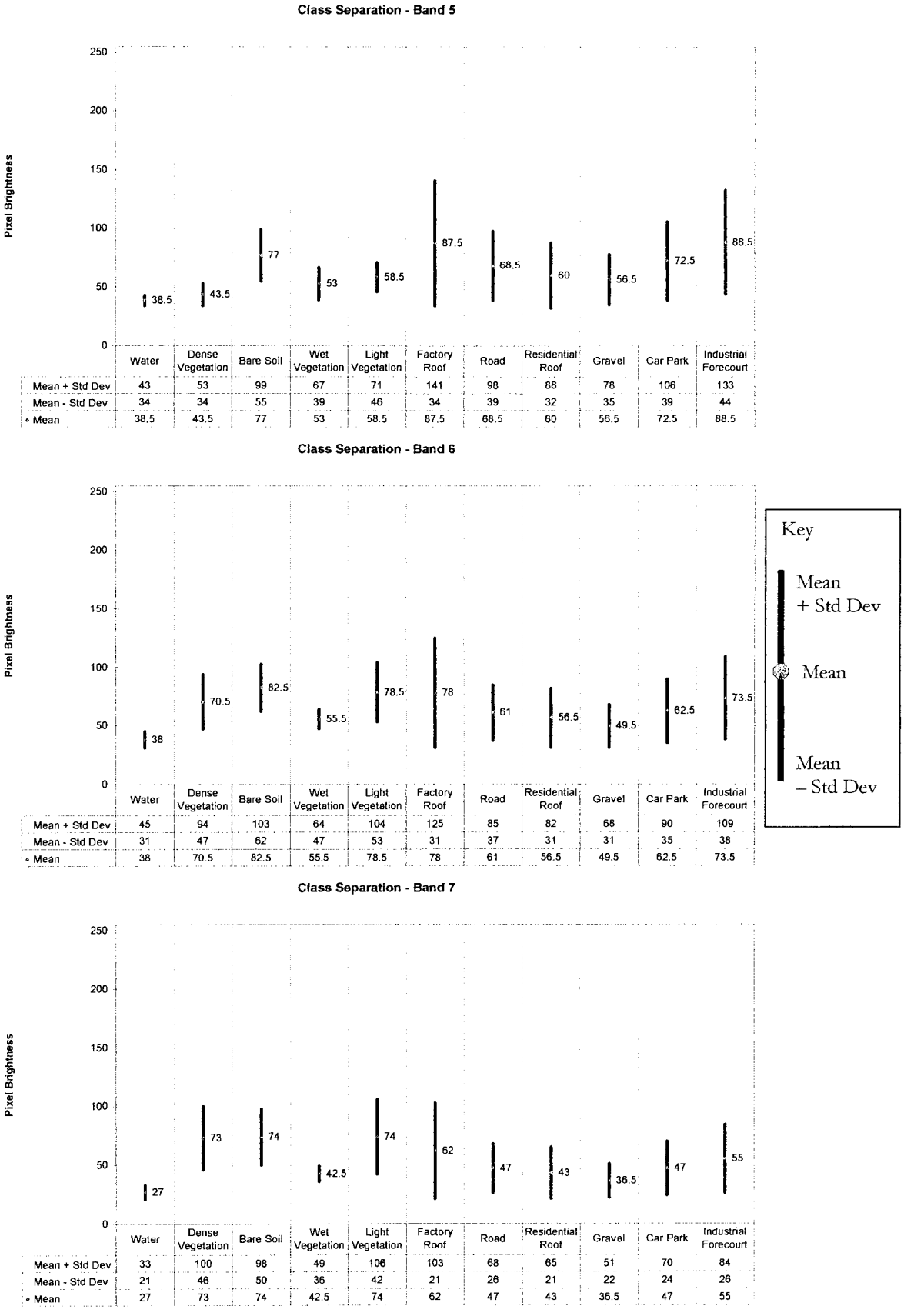
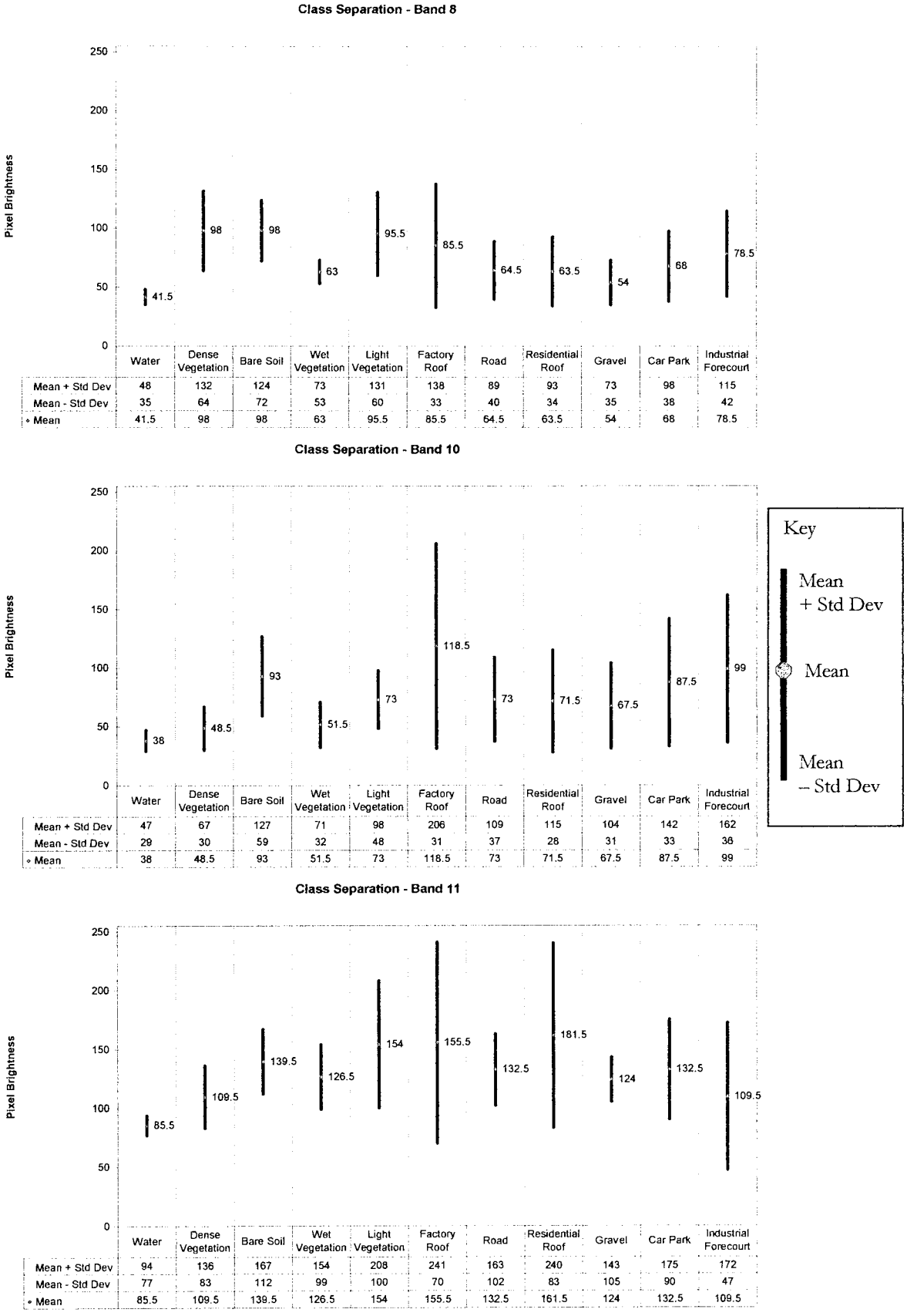


Figure 3.9 Class Separability for Bands 5, 6 and 7.



	Water	Dense Vegetation	Bare Soil	Wet Vegetation	Light Vegetation	Factory Roof	Road	Residential Roof	Gravel	Car Park	Industrial Forecourt
Mean + Std Dev	48	132	124	73	131	138	89	93	73	98	115
Mean - Std Dev	35	64	72	53	60	33	40	34	35	38	42
◦ Mean	41.5	98	98	63	95.5	85.5	64.5	63.5	54	68	78.5

	Water	Dense Vegetation	Bare Soil	Wet Vegetation	Light Vegetation	Factory Roof	Road	Residential Roof	Gravel	Car Park	Industrial Forecourt
Mean + Std Dev	47	67	127	71	98	206	109	115	104	142	162
Mean - Std Dev	29	30	59	32	48	31	37	28	31	33	36
◦ Mean	38	48.5	93	51.5	73	118.5	73	71.5	67.5	87.5	99

	Water	Dense Vegetation	Bare Soil	Wet Vegetation	Light Vegetation	Factory Roof	Road	Residential Roof	Gravel	Car Park	Industrial Forecourt
Mean + Std Dev	94	136	167	154	208	241	163	240	143	175	172
Mean - Std Dev	77	83	112	99	100	70	102	83	105	90	47
◦ Mean	85.5	109.5	139.5	126.5	154	155.5	132.5	161.5	124	132.5	109.5

Figure 3.10 Class Separability for Bands 8, 10 and 11.

There have been concerns over the validity of the statistical approach. Crane, Malila and Richardson (1972) evaluated the suitability of the normal density assumption in a series of tests. They found that many data sets (e.g. of training classes) were non-normal. They concluded that a decision rule based on the assumption of multivariate normal distributions of scanner signals performed sufficiently well, in comparison with a more accurate but more complicated rule, to warrant its continued use in recognition processing.

3.3.3 Training Stage confusion matrix

A confusion matrix, when applied to the training data, gives us a guide to the quality of the subsequent classification. It also provides a measure of the spectral stability of the data. This information allows us to study errors of omission and commission (inclusion). If there is no misclassification, then the non-diagonal elements of the matrix will be zero. It is important to note that this stage only informs us of how well the classifier can classify the training data. Since these should be homogeneous examples of each cover type, they we should expect them to be interpreted more accurately than less pure examples found elsewhere in the scene.

The matrix should not be confused with a measure of how accurate the classification is. The developments of several methods for assessing the accuracy are developed in Chapter 4.

The confusion matrix for the training stage for the Maximum Likelihood classifier is shown in Table 3.1. From this Table it becomes apparent that the Light and Dense Vegetation classes are slightly confused. This indicates a possible source of error, which may be eliminated if the two classes are combined. This is discussed in the following section.

Training Stage Confusion Matrix											
Class	**Water**	**Dense Vegetation**	**Bare Soil**	**Damp Vegetation**	**Light Vegetation**	**Factory Roof**	**Road**	**Residential Roof**	**Gravel**	**Car Park**	**Industrial Forecourt**
Water	3521	2	0	3	0	2	1	22	0	1	0
Dense Vegetation	0	8537	0	0	1	1	0	2	0	2	0
Bare Soil	0	24	1458	0	49	209	0	27	0	0	223
Wet Vegetation	0	0	0	413	0	0	6	3	0	0	1
Light Vegetation	0	3931	398	0	9621	0	0	71	0	0	0
Factory Roof	0	0	0	16	0	9431	166	815	19	980	156
Road	0	1	2	0	0	5	2176	11	45	89	159
Residential Roof	16	2	2	0	2	10	59	1122	45	40	15
Railway	0	0	0	0	0	0	14	15	1068	0	74
Car Park	0	8	0	1	0	46	37	35	20	1564	46
Industrial Forecourt	0	12	1	1	0	81	495	33	78	63	3256
Total	3537	12517	1861	434	9673	9785	2954	2156	1275	2739	3930

Table 3.1 The Training Stage Confusion Matrix for the Maximum Likelihood Classifier.

3.4 Existing Classification Schemes applied to the Dudley data-set

Table 3.2 shows the classes identified for the classification of the imagery. While they do not include every type of land cover in the imagery, as this would be impractical due to the sheer number of classes present, they are broad enough to account for almost all of the pixels in the imagery and fine enough to discriminate between different classes.

At this stage it is important to differentiate between land-cover and land-use. Land-cover describes the material which extends over the region, such as concrete or asphalt. Land-use, however, refers to the use which the land is put, for instance a road or roof. As this project is concerned with determining the amount of surface runoff that can occur (for urban drainage applications), the categories of concern will be such things as sloped roofs, irrespective of their material of construction.

The initial classification is divided into 11 categories. These represent the most distinctive features in the imagery. Subsequent reclassifications are easily achieved by grouping these together into macro classes. The purpose of this is to provide suitable data for input into urban drainage models. It is likely that this will lead to an improvement in the classification accuracy, as statistically similar classes (such as Light Vegetation and Dense Vegetation) are merged. This avoids the misclassification of one as the other and vice versa.

From the 11 classes, we can group several similar categories together to give four macro classes: Road, Water, Roof and Permeable area. Road and roof are included, rather than being grouped together as impermeable area because the amount of paved and roofed area in the catchment is an important input into some urban drainage models. The reclassification is achieved by combining these classes with, for instance, all the area previously classified as "Industrial Forecourt" being renamed "Road".

This reclassification is shown in Table 3.3.

Class	Land-Use Description
1	Water
2	Dense Vegetation
3	Bare Soil
4	Wet vegetation (i.e. Sewage Beds/ Canal path)
5	Light Vegetation
6	Factory Roof
7	Road
8	Residential Roof
9	Railway
10	Car Park
11	Industrial Forecourt

Table 3.2 Description of 11 Classes.

Reclassified Land-Use	Previous Class	Previous Land-Use
1 - Road	7	Road
	10	Car Park
	11	Industrial Forecourt
2 - Water	1	Water
3 - Roof	6	Factory Roof
	8	Residential Roof
4 - Permeable Area	3	Bare Soil
	4	Wet Vegetation
	2	Dense Vegetation
	5	Light Vegetation
	9	Gravel

Table 3.3 Redefinition of the 11 classes into 4 classes.

Four classification methods were selected for examination: Supervised Maximum Likelihood, SMAP, ECHO and Unsupervised Maximum Likelihood. Details on these classification methods can be found in Chapter 2.

The reason these four methods were selected was because it was felt that each of them was an example of the most widely used approaches used to tackle the classification problem. The Maximum Likelihood method (Lillesand and Keifer, 1994) is the most commonly used one, and as a result was tested in supervised and unsupervised modes. The SMAP Estimator (Bouman and Shapiro, 1992; Bouman and Shapiro, 1994) is a recent development and represents the 'state of the art', whereas the ECHO classification (Landegrebe, 1980) represents one of the more traditional region-growing methods.

These four methods, and the different approaches used, also give a platform from which the development of a new classification procedure can be started. Chapter 4 will assess the quantitative aspects of these classifiers from an accuracy assessment perspective.

The results are presented in three sections. The first part presents the classifications using 11 classes. The second part presents 4 class classifications, and the final section presents the PIMP image. The look-up tables applied to the imagery were selected so as to be a similar colour to the class they represent. As such, they give a good 'feel' for the quality of the classification, with certain classes being instantly identifiable.

3.5 Results

As discussed in Chapter 2, a visual inspection of the imagery can reveal a great deal of information about a classified map, from a qualitative point of view. Many features of the classification, such as how well it distinguishes road and roof, and the homogeneity of these regions, can be quickly ascertained in this manner. Consequently, the discussion of these results in this chapter will focus on this type of assessment. A detailed quantitative assessment is presented in Chapter 4.

3.5.1 Supervised Maximum Likelihood Classification

The Classification of the imagery was carried out using bands 2, 3, 4, 5, 6, 7, 8, 10 and 11, as detailed in the previous section. The training regions from the map shown in Figure 3.6 were used. The result of this classification for 11 classes is shown in Figure 3.11 and the result for 4 classes is shown in Figure 3.12.

3.5.2 Sequential Maximum a posteriori *Estimator*

This was a supervised classification. The same training map and ATM bands that were used for the Maximum Likelihood Method were used as the input for this classification. The result of this classification for 11 classes is shown in Figure 3.13 and the result for 4 classes is shown in Figure 3.14.

3.5.3. ECHO Classification

This was a supervised classification. The same training map and ATM bands that were used for the Maximum Likelihood Method were used as the input for this classification. The result of

this classification for 11 classes is shown in Figure 3.15 and the result for 4 classes is shown in Figure 3.16.

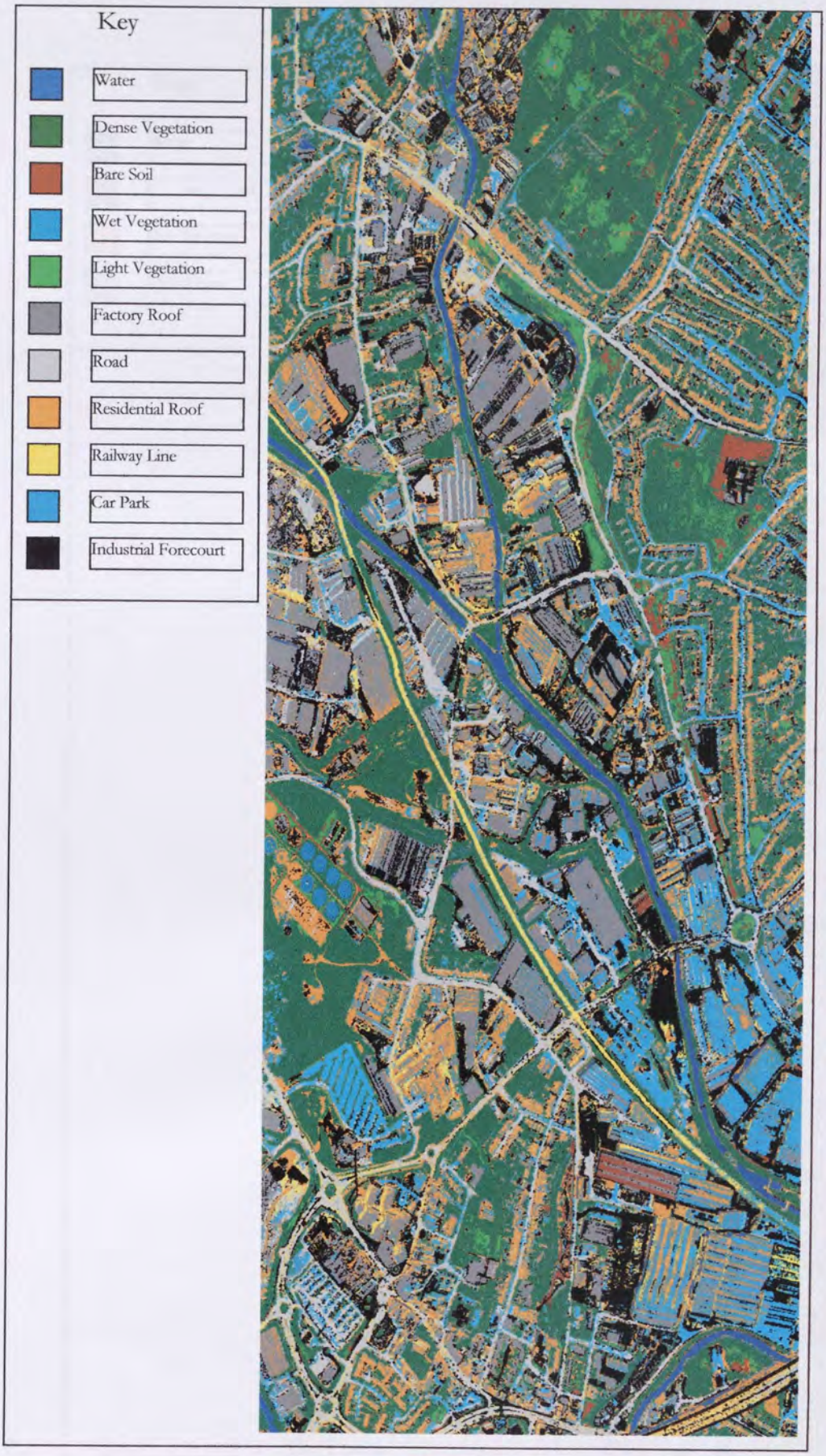


Figure 3.11 Supervised Maximum Likelihood Classification, 11 Classes.

Figure 3.12 Supervised Maximum Likelihood Classification, 4 Classes.

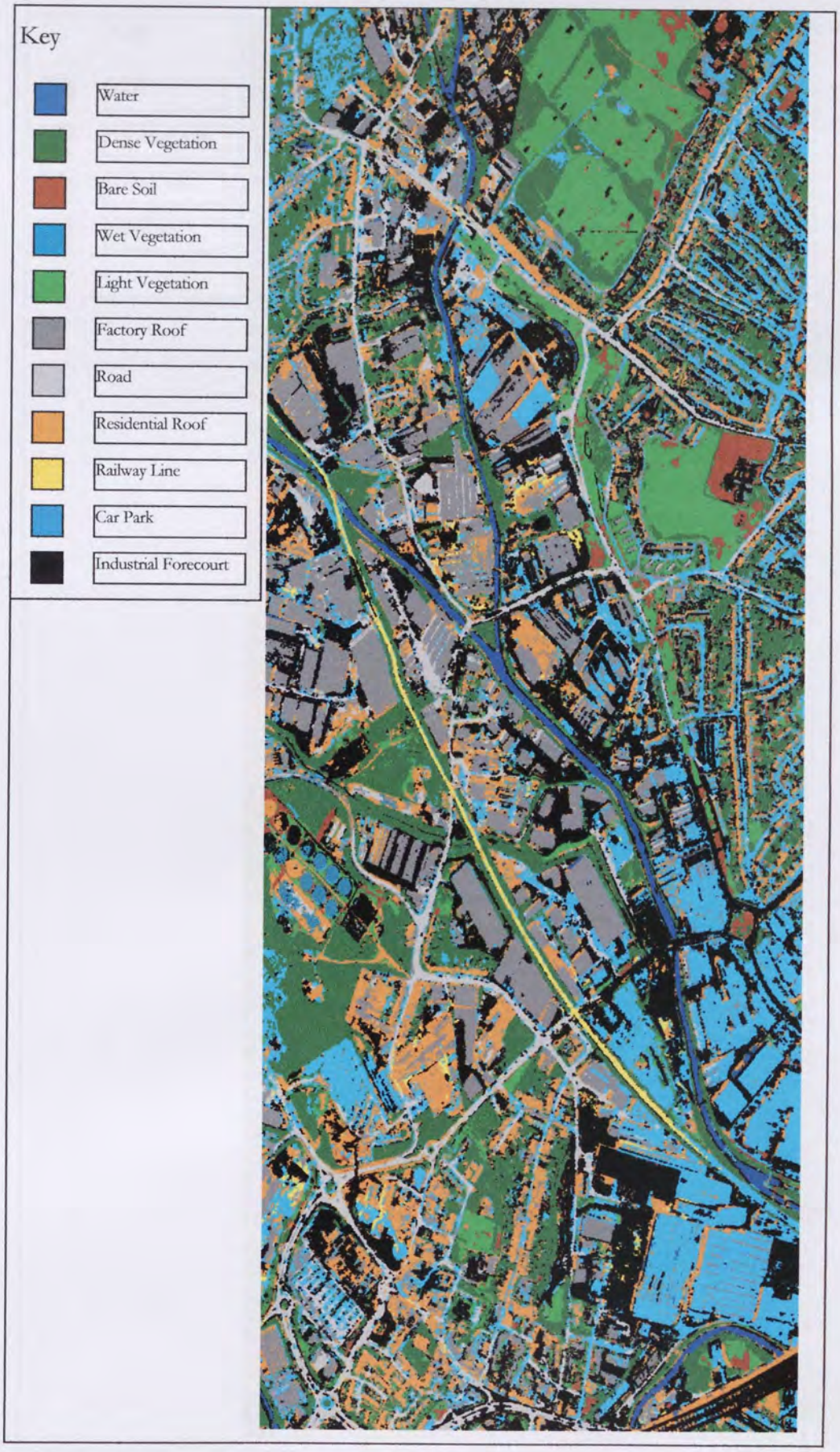


Figure 3.13 SMAP classification, 11 classes.

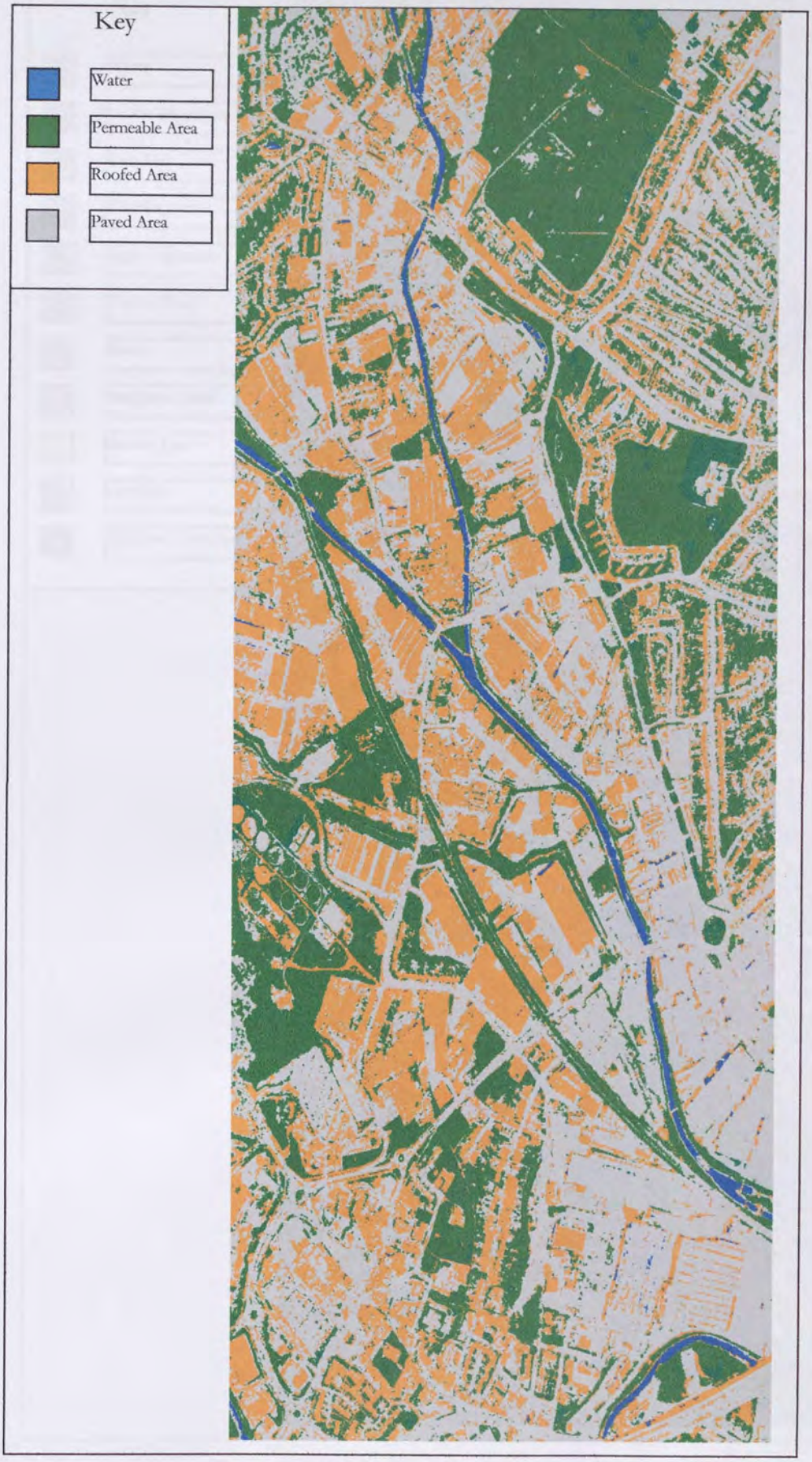


Figure 3.14. SMAP classification, 4 Classes.

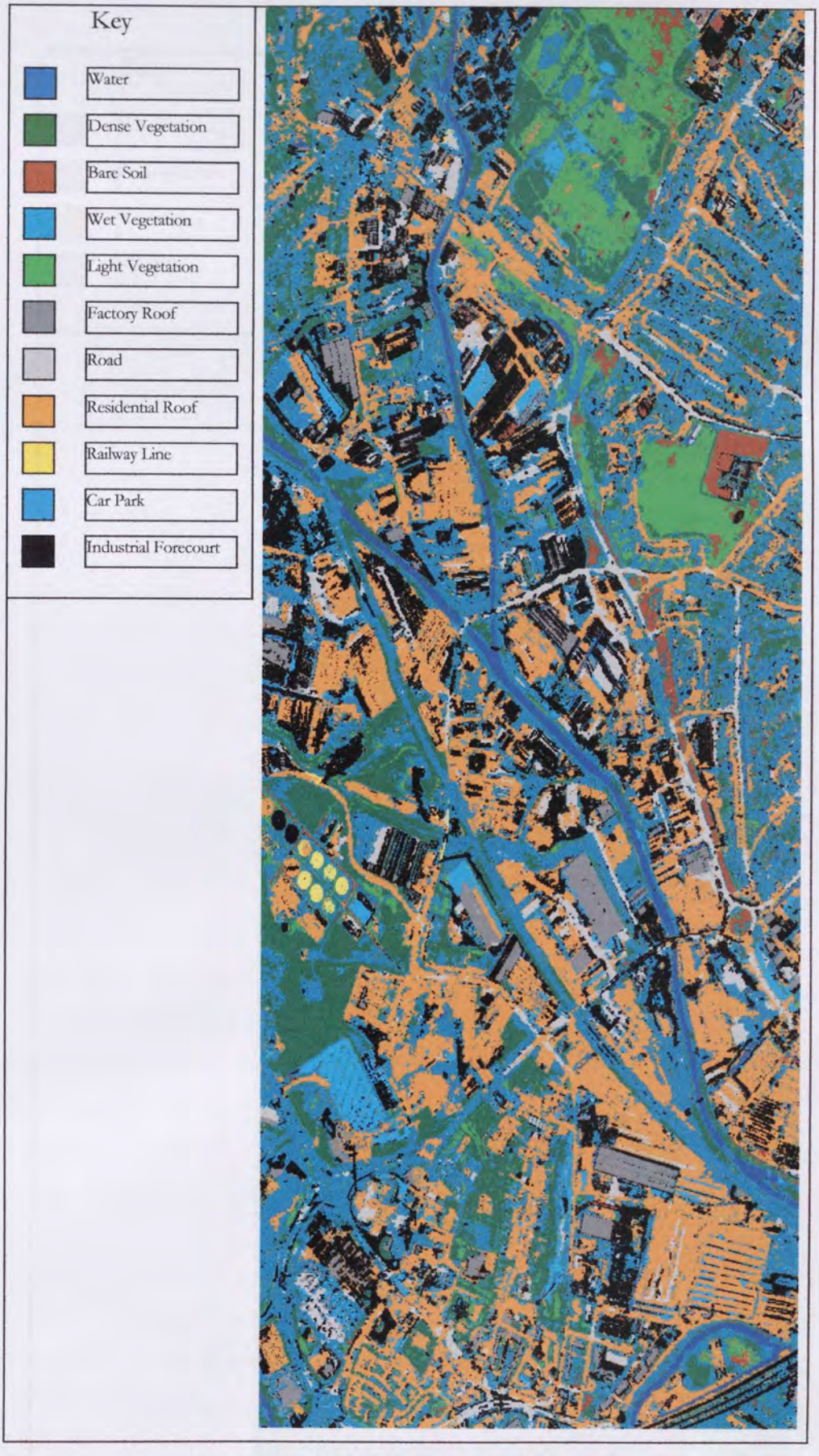


Figure 3.15 ECHO Classification , 11 Classes.

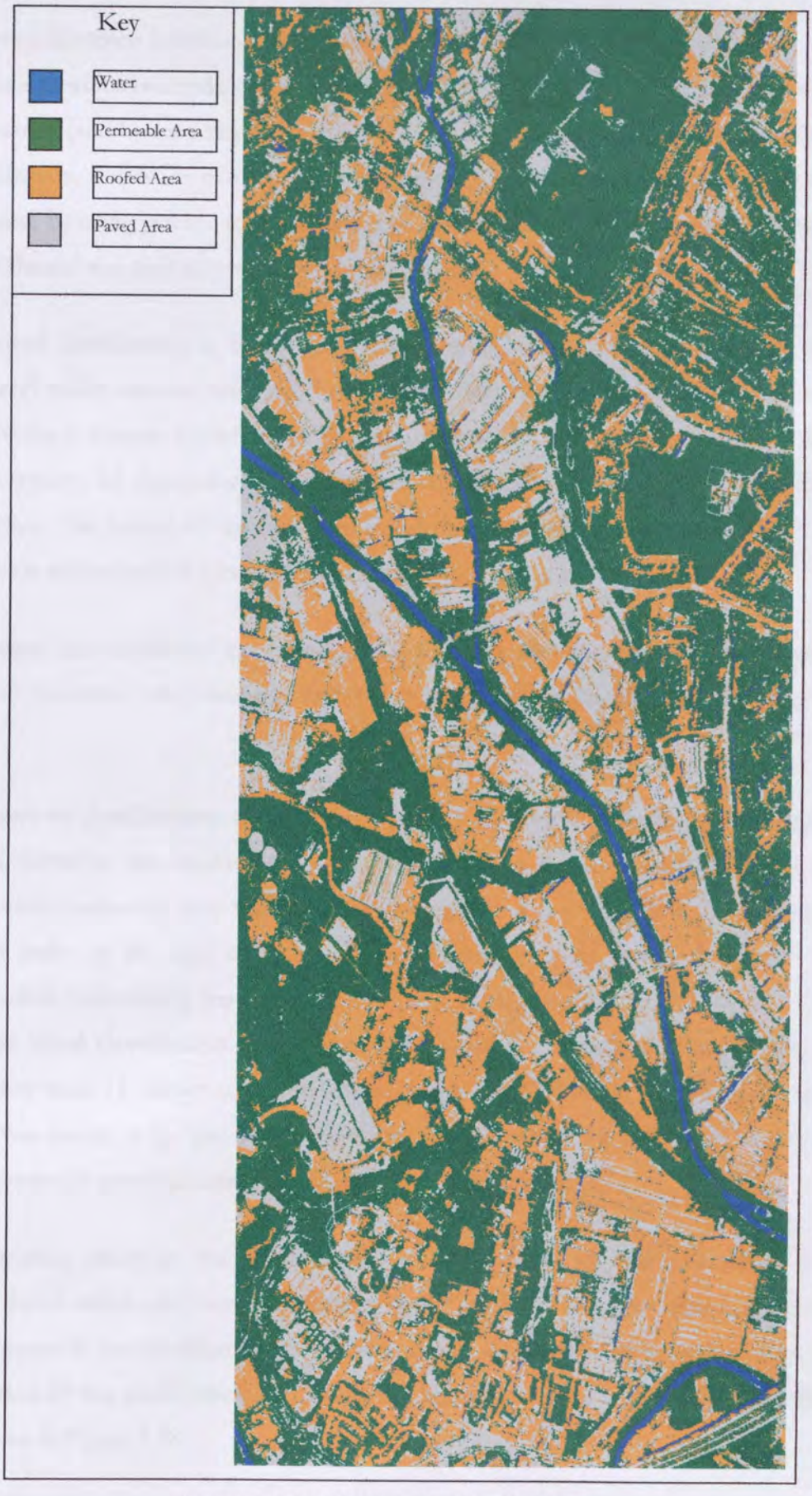


Figure 3.16 ECHO Classification, 4 Classes.

3.5.4 Unsupervised Maximum Likelihood Classification

Unlike the three previous methods, this was an unsupervised classification. The aim here was to achieve a similar (and hence comparable) classification to the other methods. By examining statistical parameters, a number of clusters are identified. These are then classified after the map is produced, by using ground truth or, as in this instance, aerial photography. The Grass 4.1 function 'i.cluster' was used to perform this classification.

The unsupervised classification is based on a clustering algorithm that reads through the imagery data and builds clusters based on the spectral reflectance of the pixels. The spectral distributions of these clusters are influenced by six parameters. These parameters include the number of clusters to be discriminated, the percent convergence, and the row and column sampling intervals. The default values were used for all the parameters – for specific details of these the reader is referred to the Grass 4.1 users manual (Shapiro *et al*, 1993.)

The cluster means and covariance matrices from the signature files generated by this process are then used to determine which category each cell in the image has the highest probability of belonging to.

As with the previous classifications, the aim was to extract 11 classes. With unsupervised per-pixel methods, however, the resulting map will generally have a fewer number of distinct classes than initially requested, since some are clearly members of the same class. For instance, the pixels that make up the road class may be spread across several groups that must be combined if a class representing road is to be produced. As a consequence, 20 groups were selected for the initial classification. These were subsequently combined to give 10 distinct classes. 10 rather than 11 classes were used because the class identified as 'railway' in the training areas was found to be spread across several of the other groups. Whereas the road regions were distinct in several groups, the railway was distinct in none.

It becomes apparent, therefore, that unsupervised classification is a hit and miss affair. The operator must decide which class certain indistinct groups of pixels belong to. This is certainly an area where errors in the classification are introduced, and involves a certain degree of good fortune. The result of this classification for 10 classes is shown in Figure 3.17 and the result for 4 classes is shown in Figure 3.18.

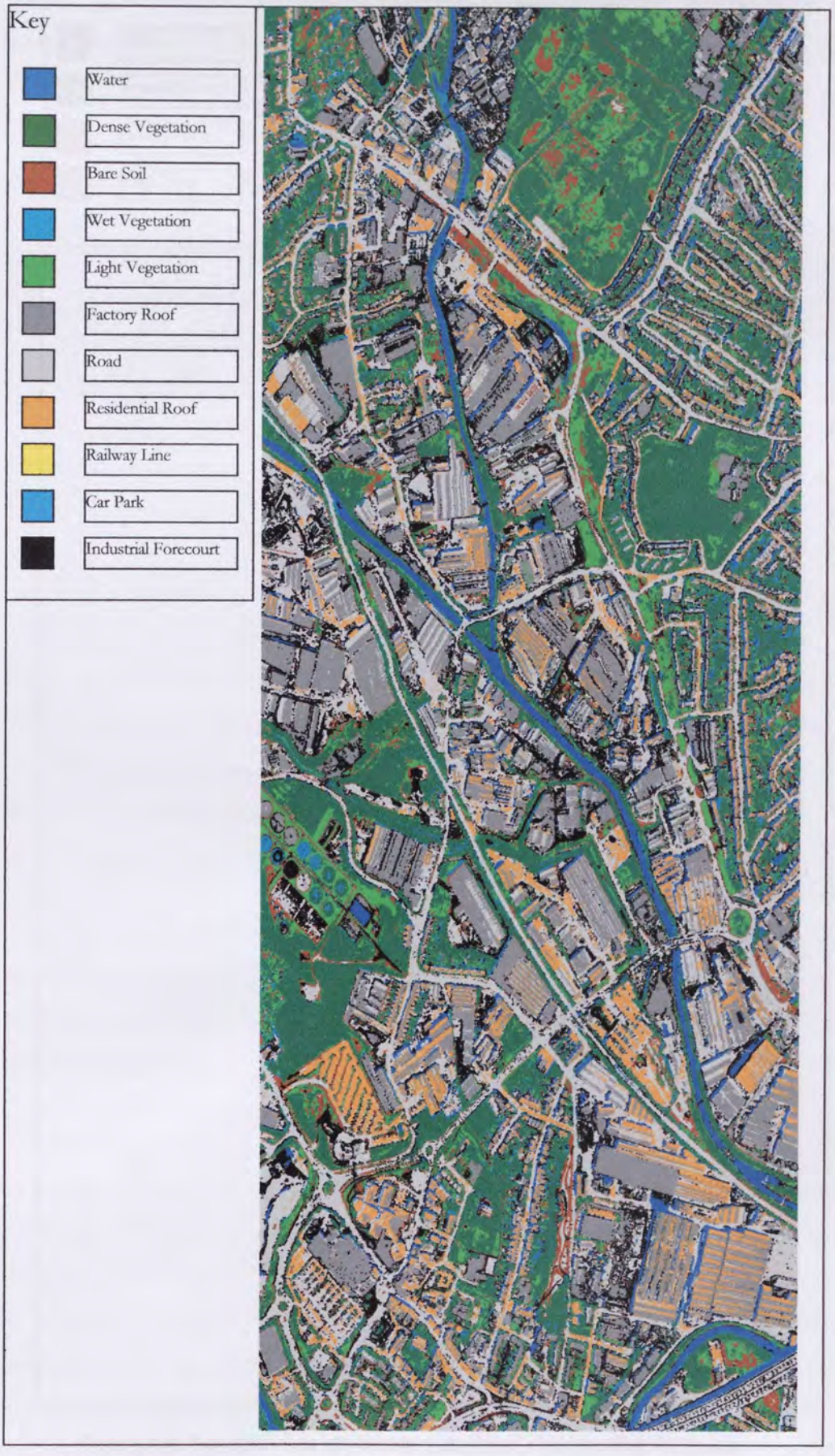


Figure 3.17 Unsupervised Maximum Likelihood Classification, 9 Classes.

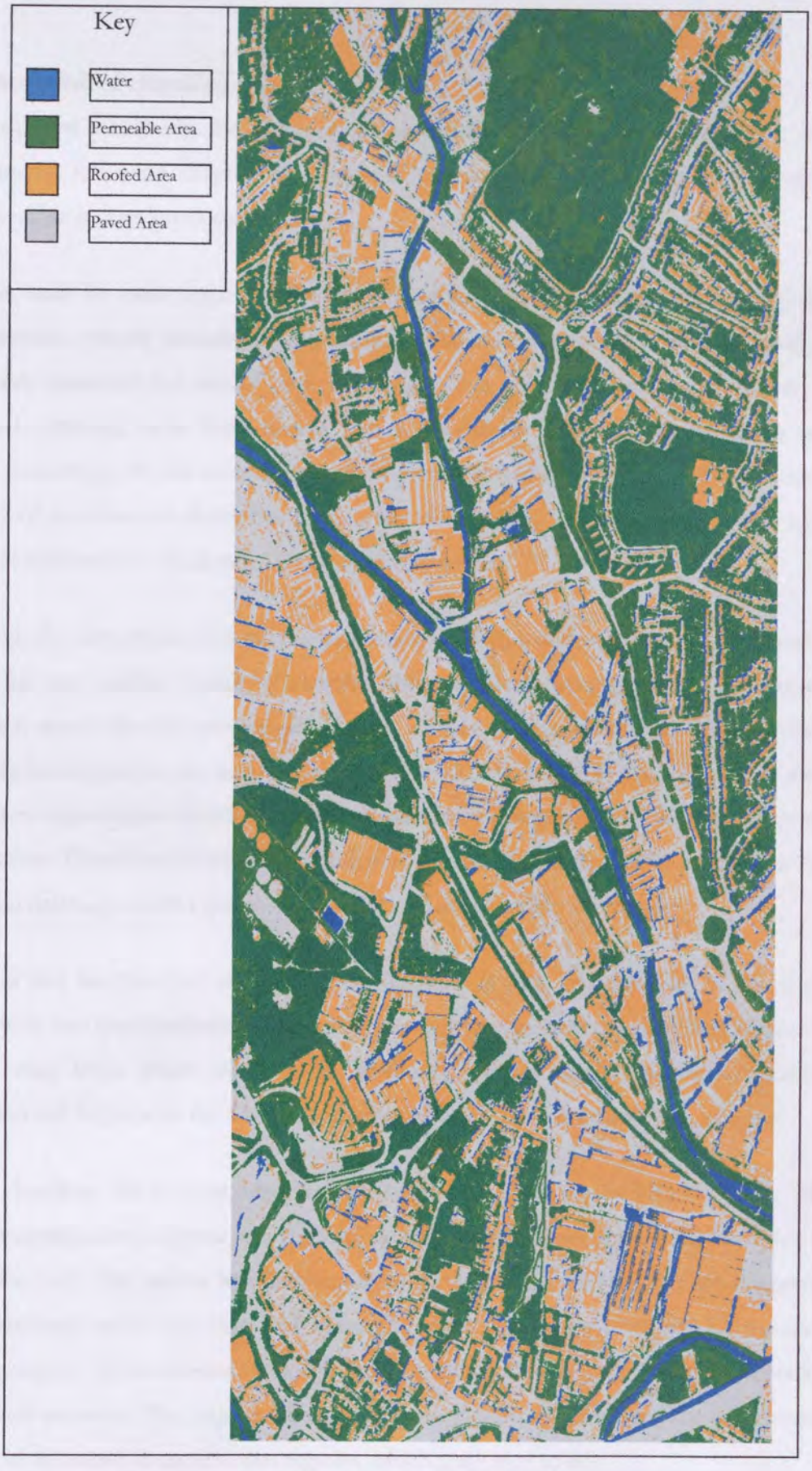


Figure 3.18. Unsupervised Maximum Likelihood classification, 4 Classes.

3.6 Discussion of classified imagery

As mentioned previously, the accuracy of the classifications will be assessed in a variety of ways, in the following chapter. The purpose of this section is to point out the qualitative aspects of the images produced.

A point must be made regarding the amount of speckle observed in the imagery. This is a characteristic typically associated with per-pixel classification routines, where a pixel may be incorrectly classified for one of several reasons. One reason is that it may simply be mis-classified. Another, more likely reason, is that the pixel is 'mixed' – that is it may contain a certain percentage of one class and a certain percentage of another. As a result it may not fit into any of the classes defined, but some other 'mixed' class. This is not necessarily a problem for most applications - section 2.3.8 discusses this further.

However, the elimination of speckle is important for the classified imagery in this project, since urban drainage models require information about regional coverage within the imagery. The important inputs are the percentages of road, roof and impervious area – mixed pixels must therefore be assigned to the most appropriate class. This will lead to homogeneity in the image, as the most appropriate class for stray pixels should be determined by the context surrounding those pixels. Therefore an important qualitative measure of the suitability of the classified data for urban drainage model input is the amount of speckle within the image.

It is clear that the classified images vary considerable in their interpretation of the data - both qualitatively and quantitatively. Yet it is also apparent that each of the classifiers has produced a distinct map from which most of the features of this imagery can be recognised. This discussion will begin with the 11-class imagery and conclude with the 4-class imagery.

Of the classifiers, the poorest, from a qualitative viewpoint is the ECHO classifier. The road network appears non-existent as a result of confusion with several other classes, most notably residential roof. The railway line (running from top-left to bottom-right in the imagery) in the image has been entirely mis-classified, as has the distinctive sewage plant area (to the centre-left of the imagery). These classes are small, though, and as a consequence may not greatly affect the overall accuracy. The map produced is not speckled in appearance since this classifier was developed to extract homogeneous regions, which it appears to do.

The SMAP classifier produces a map of better quality. The railway line is almost fully extracted, and there is evidence of a road network in the image. However, there is still confusion between road and several other classes. The image has a more speckled appearance than the ECHO method, although the quality is generally high.

The supervised Maximum Likelihood Classifier successfully extracts almost all of the road network, railway and distinct sections of residential housing. However it is much more speckled in appearance than the SMAP and the ECHO classifiers.

The Unsupervised Maximum Likelihood Classifier is, like the supervised one, speckly in nature. The roads and roofs in the imagery can be clearly distinguished throughout the imagery although, for reasons already mentioned, the railway is also classed as road. A result of the procedure used is also that the edges of the road regions are classified as Industrial Forecourt. This is because certain clusters in the 20-class image contain two classes, yet they must still be assigned to a single class upon reclassification to 10 classes.

Reducing the number of classes (from 11 to 4 for supervised classifications, 10 to 4 for unsupervised one) highlights the misclassification of houses and road in the imagery while, to some degree, masking the speckly nature of some of the classified images.

The supervised Maximum Likelihood classifier distinguishes between roads and roofs very well, although there is some confusion between road and permeable area on the motorway segment in the lower right side of the imagery. The image is, as would be expected, less speckly than with 11 classes, suggesting that much of the previous mis-classification be due to classes of similar properties being confused.

The SMAP classifier mis-classifies large regions of road and roof, particularly the factory roof areas. The road is well represented, but appears to be suffering from a large error of commission.

The ECHO classifier performs the poorest classification, with large areas of road mis-classified as permeable along with many small areas of roof. There is also some confusion between permeable areas and water, particularly alongside some roofed areas. This is indicative of the presence of shadow regions.

The shadow regions are much more noticeable on the unsupervised classification. This is a result, once more, of two classes being combined into one. Otherwise, however, the road and roof classes are clearly distinguishable from one another. The speckling of the 10-class image has also been reduced.

3.7 Conclusion

The classified imagery is of varying quality. Clearly the Maximum Likelihood method is of the poorest quality in terms of homogeneity, due to its speckly nature. This is even more so with the unsupervised classification.

The homogeneity of the Maximum Likelihood image is inferior to that of the SMAP and ECHO classifications. This is because the SMAP and ECHO classifications have an element of contextual information incorporated, albeit a small one. This is effectively of the form of a neighbourhood-smoothing algorithm, which is implemented differently in the two classifiers. In the SMAP classifier it is achieved through the use of scaling, whereas in the ECHO classifier it is achieved through its inherent region-growing algorithm.

The homogeneity of the images aside, the Maximum Likelihood classifications, both supervised and unsupervised, appear to be those of the highest quality. The supervised method does not suffer from the errors at the edges of roads and houses in the same manner as the unsupervised one. These are particularly difficult areas for a classification method. Consequently the classification method most suited to this type of imagery, from those presented, is the supervised Maximum Likelihood.

There is a need, however, to measure the accuracy of these classifications objectively, in such a way that the qualitative impression gained from this chapter can be allied with a quantitative measure.

Chapter 4: Testing Methods and Evaluation

4.1 Introduction

Perhaps the most important stage of analysis is concerned with finding out how successful the classification has been. The evaluation of the accuracy of the classification stage is a difficult procedure, and one in which the avoidance of bias is difficult. A detailed review of methods developed to assess the accuracy of classified imagery, not only in terms of direct measurement but also in terms of the end user suitability, can be found in Booth (1989).

While the main concern here is in quantifying the absolute accuracy of a classified image, a further aim is to show that a single value can be misleading. It is important to know what the value represents. For instance, a commonly quoted statement is "for an automatically classified land use map to be acceptable, its accuracy must exceed 85%" (Jenson, 1986; Campbell, 1983; Colwell, 1983). Does this value represent a straight comparison between the amount of area determined by the classifier compared with that obtained by a manual classification? Or is it a more robust test that shows where misclassification is occurring and provides the user with a qualitative view as well as a quantitative one? When deciding upon a method, consideration should be made as to what information the test is actually providing.

It would be useful to know what has happened to those pixels that have been mis-classified, for instance whether they have been assigned to a statistically similar class or not. Since the nature of urban imagery is such that many classes have a similar spectral response this may not actually detract from the performance of the classifier, but it will provide the user with pertinent information about the absolute quality of the classification procedure.

As a consequence, the most important information is provided by the error of commission and omission for each class. This gives information about whether the classifier is performing well in one class because it is a good classifier, or whether this is due to the classifier having an abnormally large number of pixels assigned to that class. It is clear that the greater the number of pixels assigned to one class, the greater the chance of a higher number of pixels being correctly identified. Equally, if the number of pixels in the class is underestimated we may expect a lower number of correctly assigned pixels than otherwise.

This additional information is just as important as the overall accuracy in assessing the performance of the classifier.

In this section, we will examine the methods for generating a sample for assessment, and the two most popular techniques for measuring the accuracy, applied to the imagery presented in Chapter 3. While there are many more methods for assessing the accuracy of data (Booth, 1989; Congalton, 1991; Gopal and Woodcock, 1994; Ma and Redmond, 1995) this study is only concerned with determining a suitable measure that allows an objective comparison between image classification techniques. Consequently, rather than perform a rigorous analysis of all the available methods, a decision, based on the literature, was made to assess only the most widely used techniques – namely the absolute accuracy (with its confidence level) and the Kappa statistic. Similarly, for the same reason, a further decision was made to use only the most common sampling methods, these being the single random pixel and the region correlation methods.

This examination is followed by a discussion of their merits and shortcomings. The classification schemes presented in Chapter 3 will be assessed comprehensively, detailing which is the most successful at providing a classified map of an urban region.

4.2 Sampling Methods.

4.2.1 Region Correlation

The first approach involves the comparison of a selection of regions from the classification procedure with those from a manual aerial photograph interpretation. The presentation of this data in a confusion matrix allows a detailed examination of the accuracy of the computer-guided classification.

There are several sources of bias that can affect our determination of accuracy using this method. These can be either optimistic or conservative.

a) Sources of optimistic bias.

i) Selection of reference data that is related to the training data - this can be reduced if we select areas that are different from, and considerably more extensive than, the training areas.

ii) The restriction of the testing of pixel samples to those that are relatively easy to classify, for instance those in the centre of objects.

b) Sources of conservative bias (assuming many classes and relatively high classification accuracy).

i) Any class-assignment error in the reference data

ii) Any class differences due to temporal changes between the reference data and the classified image

iii) Positional errors that are inherent in both the reference data and the classified image

iv) Differences in the pixel size of the classified image and the minimum unit of reference polygons derived from aerial photography

Care must be taken to reduce the effect these errors have on the assessment. It is, however, impossible to completely eliminate them, although the presence of one source will, to some degree, counter the effect of another. Furthermore, in comparing the automatic classification with a manual one, we are assuming that the manual one is 100% correct.

Where we are unable to determine the correct land-use of a pixel, we should assign the pixel to an 'unclassified' class. The purpose of the manual classification is to determine a reference against which to compare the automatically classified results. We have to make the assumption that the reference is 100% correct for the comparison to work. Therefore, we must remove difficult pixels from the manual image – this is one of the shortcomings of this procedure, as mentioned above.

4.2.2 Single Random Pixel

The second method is a single pixel (point) approach to error assessment, as demonstrated in Jenson (1986). A set of randomly distributed points is generated and overlaid on the classified imagery. This subset is then classified manually, using field measurements, photo-interpretation and direct examination of the digital imagery. The accuracy measure is then provided by the percentage of correctly matching pixels between the classified imagery and

the manually classified random pixels. Note that small areas consequently have a small number of pixels in both the classified image and ground truth.

The number of pixels generated will affect the accuracy measure. It is important to have enough pixels to provide a representative sample, but it should be noted that it is a time consuming and tedious procedure. For the accuracy measurements presented here the manually classified pixels (direct from the ATM imagery) amount to 500, 1000 and 1500 pixels. This represents, from an image of 1360400 pixels (712 x 1900), approximately 0.0035%, 0.07% and 0.11% of the coverage. Obviously, to classify as much as even 1% of the pixels (13,604 pixels) is unrealistic.

The purpose of varying the number of test pixels is to observe any changes in the accuracy measure that may result from the increase in sample sizes. This should provide an idea of both the effectiveness of this method and an idea as to what the optimum number of pixels required is. A rough guide as to whether or not we have enough pixels is to see what percentage of the manually classified image is of one class (say, water), and what percentage of an automatically classified image is the same class. The results indicate that with 500 manually classified pixels the distribution is such that approximately 2% of the image is classified as water. This is a figure that matched by the result from the automatic classifications of the imagery. While the automatic classification itself may be incorrect it provides a guide as to how representative the sample size is.

4.3 Measuring the Accuracy

4.3.1 Confidence Levels

The allocation of a confidence level to a classification product is considered essential in many applications of remote sensing. Each class can have associated with it a minimum threshold above which the pixels may confidently be expected to be members of that class. The assignment of a specific minimum threshold to each class is a decision that must be made by the analyst, although a commonly used measure is 95%.

The success of a classification can be influenced by a variety of factors – sensor, software, and human. In addition, the real world land-use may be difficult to determine. The

assignment of a confidence level to a classification must recognise that it represents such a combination of influences.

The accuracy of the classified map depends on the ability to extrapolate successfully from the training areas to the whole mapped area. Unless there is some statistical measure of the efficiency of the extrapolation process, the producer of the imagery cannot provide a level of confidence in the classification. Once a confidence level is so quantified, then a user of the classification data can relate it, via the probability of correct classification, to actuality over the entire classified area.

When checking the efficiency of a classification, we are concerned with the summation of the probabilities for all stipulated pixels between *n* and a lower level *(i)*. That is, we wish to know the probability that at least *i* pixels are correctly classified, when a random sample of pixels is selected.

This probability is called the confidence level (CL) for that classification, and is usually expressed as a percentage. Thus, if the CL is the integrated probability expressed as a percentage, we can say that we are CL% confident that the pixels are classified correctly at least *i* times out of *n*.

4.3.2 Calculation of the Confidence Level

Thomas *et al* (1987) describe a method for determining the confidence level of an accuracy value. A simplified explanation is given here - a more in depth discussion can be found in the paper by Thomas *et al* (1987).

As the sample size n becomes larger, the discrete Binomial Distribution approaches the continuous Normal Distribution, as the limiting case for n tends towards infinity. Provided large sample sizes are used (i.e. over 50), we can use the Normal Distribution to simplify things. This is useful, as when the total area under the curve is normalised to 1.0, the probability we want is the integrated area between the limits n and i. The Equation 4.1 shows the unit-area Normal Distribution.

$$probabilitydensity = \frac{1}{s(2\pi)^{1/2}} \exp\left(\frac{-(i-m)^2}{2s^2}\right) \qquad \text{Equation 4.1}$$

Under such a curve the integrated area from 1.65 standard deviations below the mean to infinity is 0.95. Therefore for a 95% confidence level, the lower bound to the number of pixels that must be correctly classified in the check sample is equal to the mean minus 1.65 standard deviations. So the lower acceptable limit to give a 95% confidence level is:

$$(m - 1.65e_m) - 1.65(s + 1.65e_s) \qquad \text{Equation 4.2}$$

where:

m = the estimated mean of the standard distribution,

s = the estimated standard deviation of the mean,

e_m = the standard error of the estimate of the mean and,

e_s = the standard error of the estimate of the standard deviation.

4.3.3 Cohen's Kappa statistic

In this study the use of correct percentage as an indication of map accuracy was rejected, because under certain circumstances even a random classifier can produce ostensibly reasonable values. Cohen (1960) suggested an index that reports the agreement between two images in relation to results expected from a random assignment of pixels to categories. Cohen's Kappa is a statistic that varies from +1 (perfect agreement) to –1 (complete disagreement). A value of zero indicates that the results are not distinguishable from those obtained by random classification. Further measures of this kind are reported in Rosenfield and Fitzpatrick-Lins (1986).

4.4 Results

4.4.1 Region Correlation

The first stage of this assessment method is to classify manually a region of the imagery. The selection of a region must take care to avoid the pitfalls discussed in section 4.2.1. The areas selected are shown in Figure 4.1, with the training regions highlighted. The manually classified image is shown in Figure 4.2. This region was chosen because it avoided these pitfalls and also represented most of the classes identified in the imagery; as such it provides us with an accuracy value for the majority of classes in the image.

The initial assessment was made using all 11 classes (Figure 4.2). However, the region selected contained only 9 of these classes (class 4, wet vegetation, and class 9, railway, were not present in this region). Selecting a single region of the image meant that it was difficult to find an area of the image containing all the classes. The reason that a single area was preferred over a selection of regions (such as with training areas) was that the bias of selecting only easily identifiable regions was eliminated. An area of the image was simply selected in an area that was not covered by the training areas, and manually classified.

The tables and graphs detailing the assessment can be found in Appendix 2. Only a brief description of the trends and features in this assessment are presented here, since the classes selected must subsequently be reduced for input into drainage models. However, the indication is that drainage models are becoming more sophisticated and will eventually require input of this level of detail. Furthermore, it allows us to see what effect the reduction of the classes has on a classification, and whether the errors are caused by confusion between spectrally similar classes or not.

4.4.1.1 11-Class Imagery.

The overall accuracies of the classifiers are low, although this is directly attributable to the confusion between similar classes. This being the case, the accuracies should improve when similar classes are merged and re-assessed.

Of the classifiers, the one that performs best is the Supervised Maximum Likelihood method (Table A2-1). While differentiating between areas of factory roof, residential roof and car park causes problems, the accuracy in the individual classes is high. In fact the Kappa values,

for the most part, indicate that there are low errors of commission and omission for most of the classes (Figure A2-1). The difference between the accuracies suggested by the confidence level and the Kappa value support this.

The SMAP estimator (Table A2-2, Figure A2-2), however, performs markedly less well. The main reason, once more, is the level of confusion between the car park and industrial forecourt classes. There is also confusion between other classes, notably road and roof.

The ECHO Classifier (Table A2-3, Figure A2-3) performs very poorly (the worst of all the classifiers) in all the classes except water. There is considerable confusion in all of the remaining classes, producing low measures of accuracy in each.

The Unsupervised Maximum Likelihood Classifier (Table A2-4, Figure A2-4), in contrast, performs very well in most of the classes. There are a large amount of commission errors in several of the classes, however.

4.4.1.2 4-Class Imagery

The confusion matrices (or coincidence tables) (Table 4.1 - 4.4) were produced from the GRASS program *r.coin* (Shapiro *et al*, 1993) and show the amount of pixels correctly identified. They are a reclassification of the original 11 classes selected.

From these tables we are able to produce the errors of commission and omission for each class for each classifier. Unlike the single random pixel method not every pixel is assigned to a class. This is because of the difficulty in classifying ATM imagery manually, and as we can see in Figure 4.2 leads to us effectively having an 'unclassified' class.

In order to assist in the description of the results for the region correlation sampling method, the tables have their most salient features lettered.

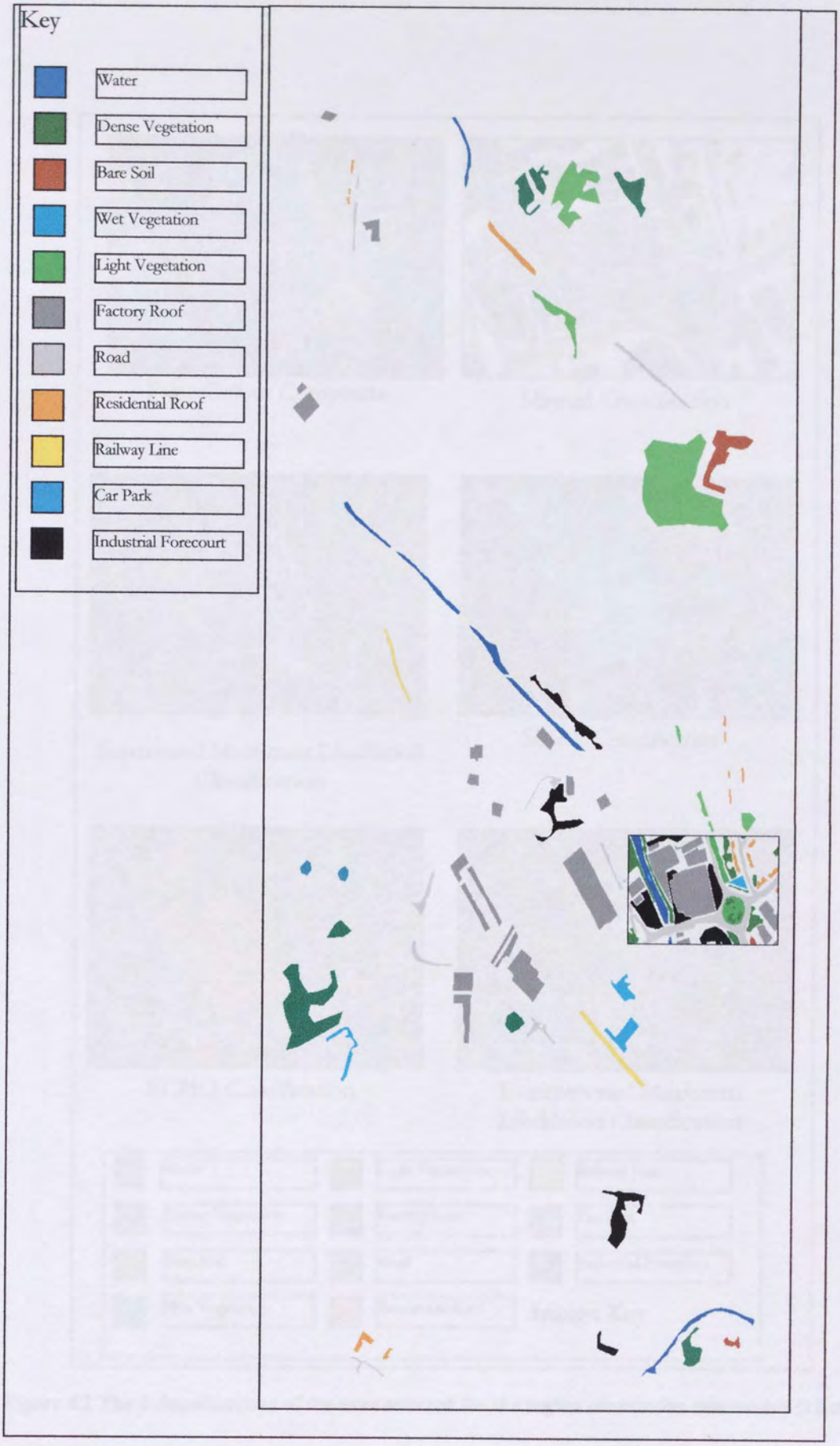


Figure 4.1 The training area map with the manually classified region superimposed.

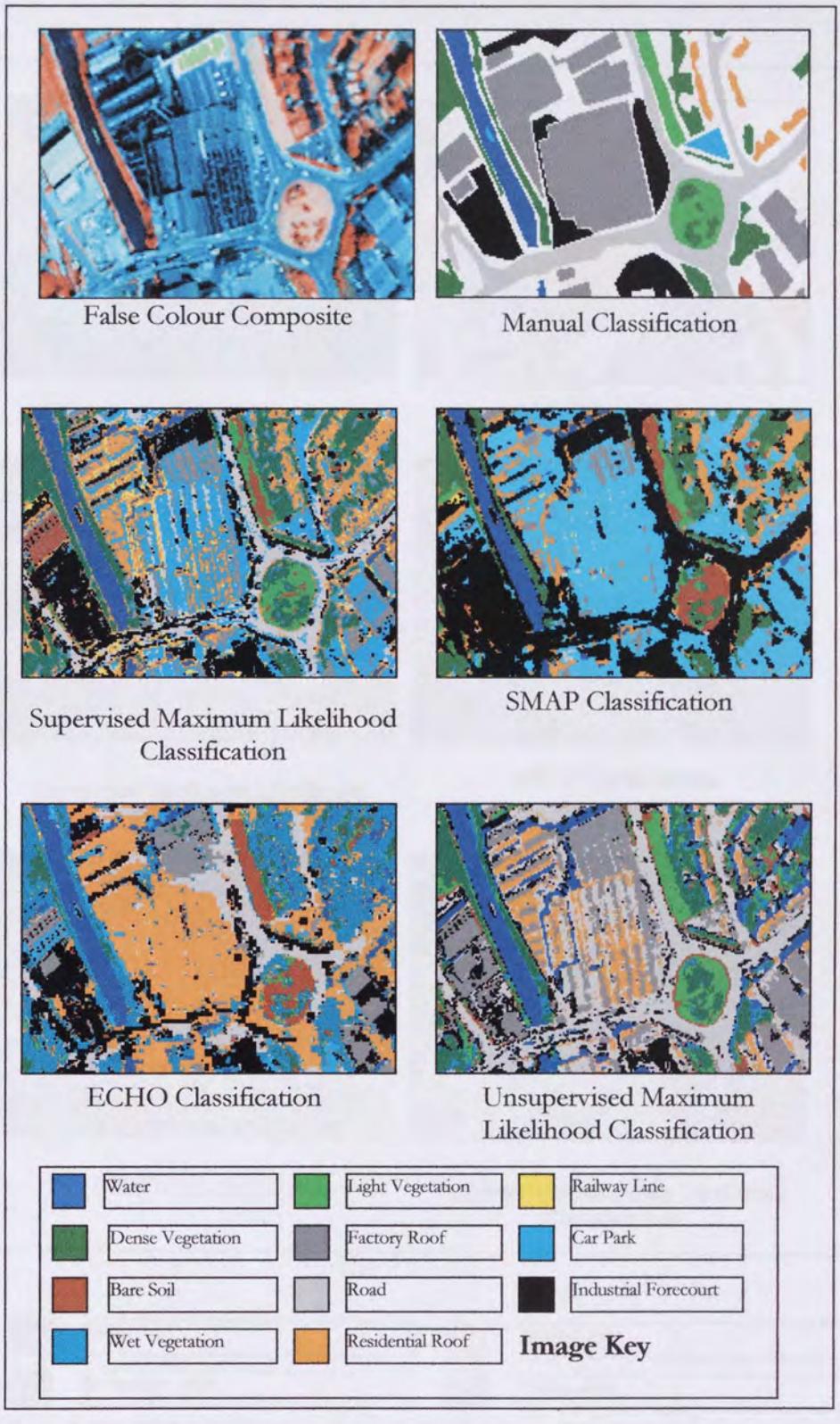


Figure 4.2 The 4 classifications of the area selected for the region correlation assessment (11 classes)

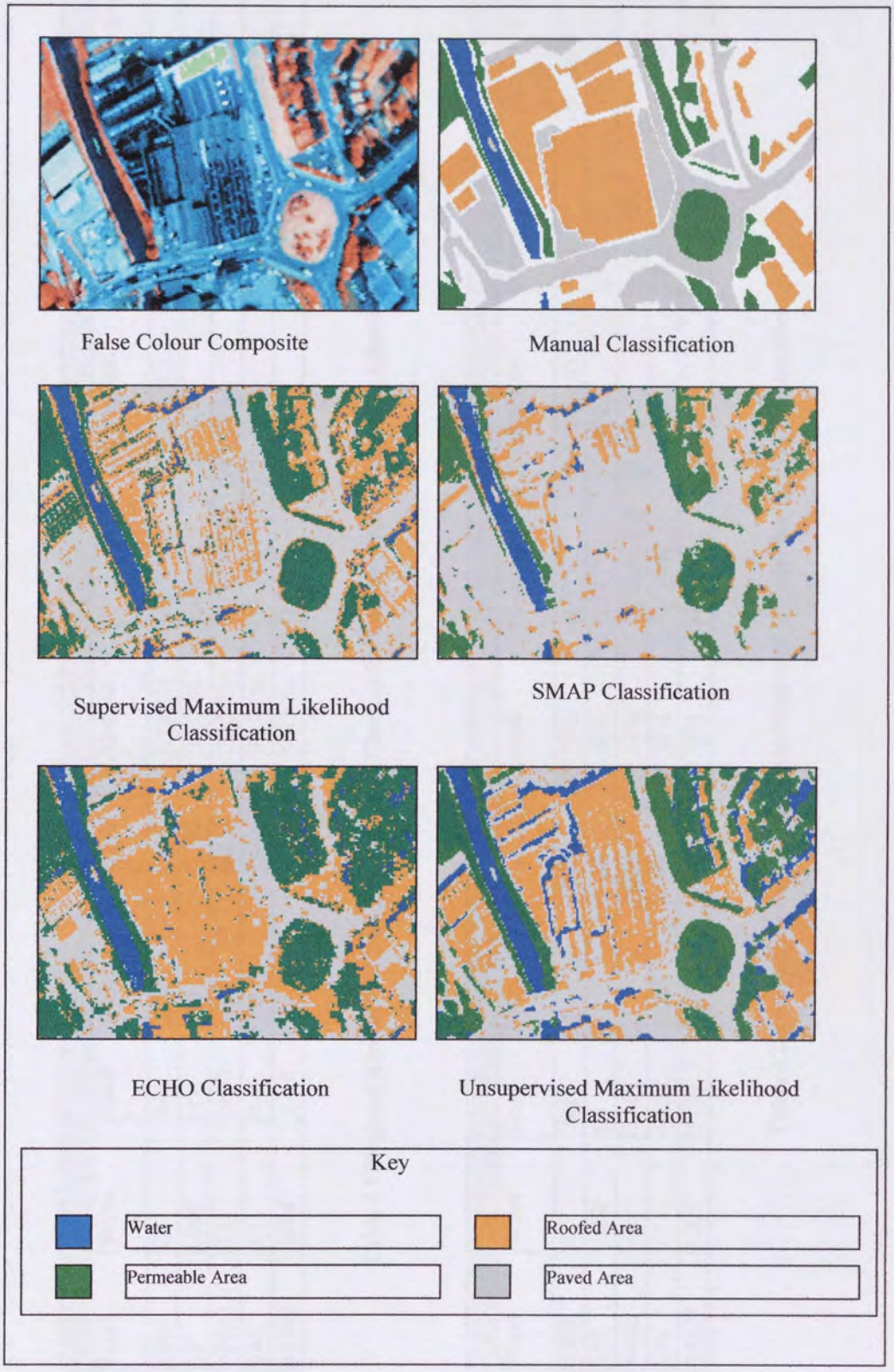


Figure 4.3 The 4 classifications of the area selected for the region correlation assessment (4 classes)

Maximum Likelihood Classifier										
	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95% CL
Road	6699	26	643	272	7640	87.68%	12.32%	62.62% **(C)**	70.81%	87.05%
Water	19	1048	161	20	1248	83.97%	16.03%	9.29%	82.98%	82.16%
Roof	4562 **(A)**	71	3004	587	8224	36.53%	63.47% **(A)**	15.19%	19.24% **(B)**	35.63%
Permeable	203	19	445	3431	4098	83.72%	16.28%	21.45%	79.21%	82.74%
Total	11483	1164	4253	4310	19866	71.39%	27.02%	27.14%	63.06%	70.85%

Table 4.1 Region Correlation Confusion Matrix (4 Classes) for the Supervised Maximum Likelihood Classifier

SMAP Classifier										
	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95% CL
Road	7352	12	221	55	7640	96.23%	3.77%	99.18% **(A)**	84.83%	95.86%
Water	152	1024	67	5	1248	82.05%	17.95%	5.69%	81.00%	80.15%
Roof	6802	45	1342	35	8224	16.32%	83.68% **(A)**	5.57%	7.98% **(B)**	15.63%
Permeable	623	14	170	3291	4098	80.31%	19.69%	2.32%	76.26%	79.25%
Total	14929	1095	1800	3386	19866	65.48%	31.27%	28.19%	62.52%	64.92%

Table 4.2 Region Correlation Confusion Matrix (4 Classes) for the SMAP Classifier.

ECHO Classification										
	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	4788	13	2279	560	7640	62.67%	37.33% **(B)**	23.00%	44.33%	61.73% **(A)**
Water	53	1115	24	56	1248	89.34%	10.66%	13.54%	88.61%	87.81%
Roof	1433	126	5965	700	8224	72.53%	27.47%	32.05%	51.56%	71.70% **(A)**
Permeable	271	30	333	3464	4098	84.53%	15.47%	32.11%	79.63%	83.57%
Total	6545	1284	8601	4780	19866	77.18%	22.73%	25.18%	66.03%	76.68%

Table 4.3 Region Correlation Confusion Matrix(4 Classes) for the ECHO Classifier.

Unsupervised Max Likelihood Classifier										
	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	5705	103	1658	174	7640	74.67%	25.33%	36.51%	55.76%	73.83%
Water	65	1118	0	65	1248	89.58%	10.42%	61.62%	88.49%	88.07%
Roof	2546	577	5073	28	8224	61.69%	38.31% **(A)**	20.37%	41.98%	60.78%
Permeable	178	89	17	3814	4098	93.07%	6.93%	6.52%	91.28% **(B)**	92.39%
Total	8494	1887	6748	4081	19866	79.08%	20.25%	31.25%	69.37%	78.60%

Table 4.4 Region Correlation Confusion Matrix (4 Classes) for the Unsupervised Maximum Likelihood Classification.

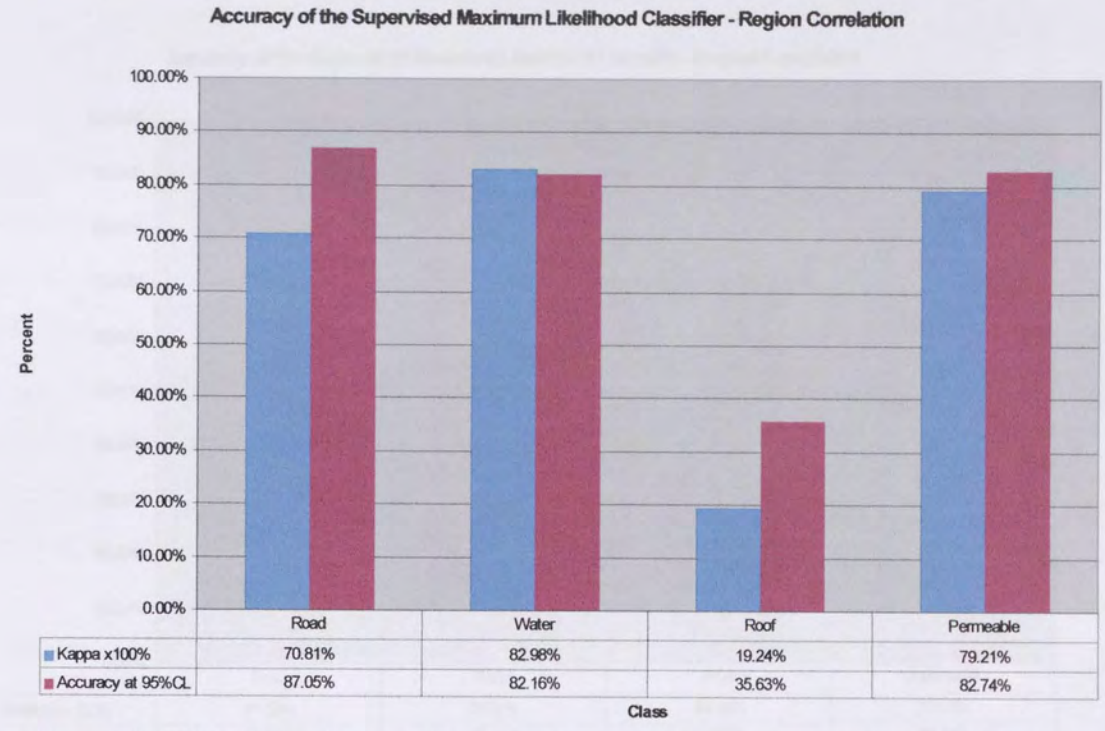


Figure 4.4 The Class Accuracies for the Supervised Maximum Likelihood Classifier (Region Correlation Method).

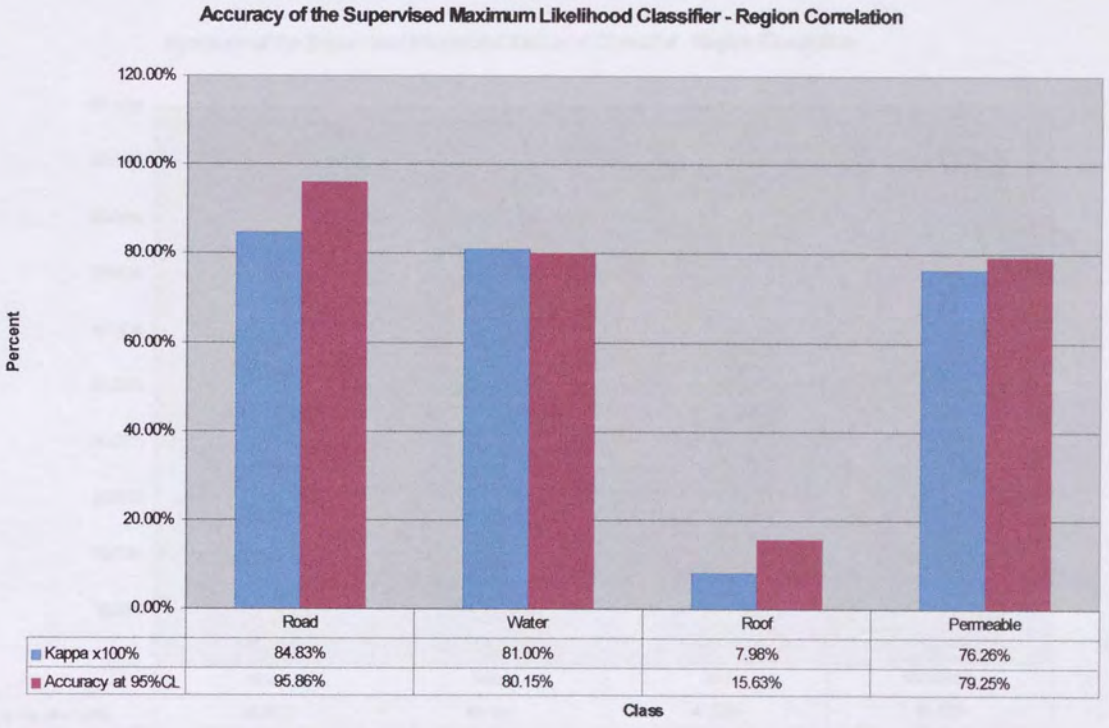


Figure 4.5 The Class Accuracies for the SMAP estimator (Region Correlation Method).

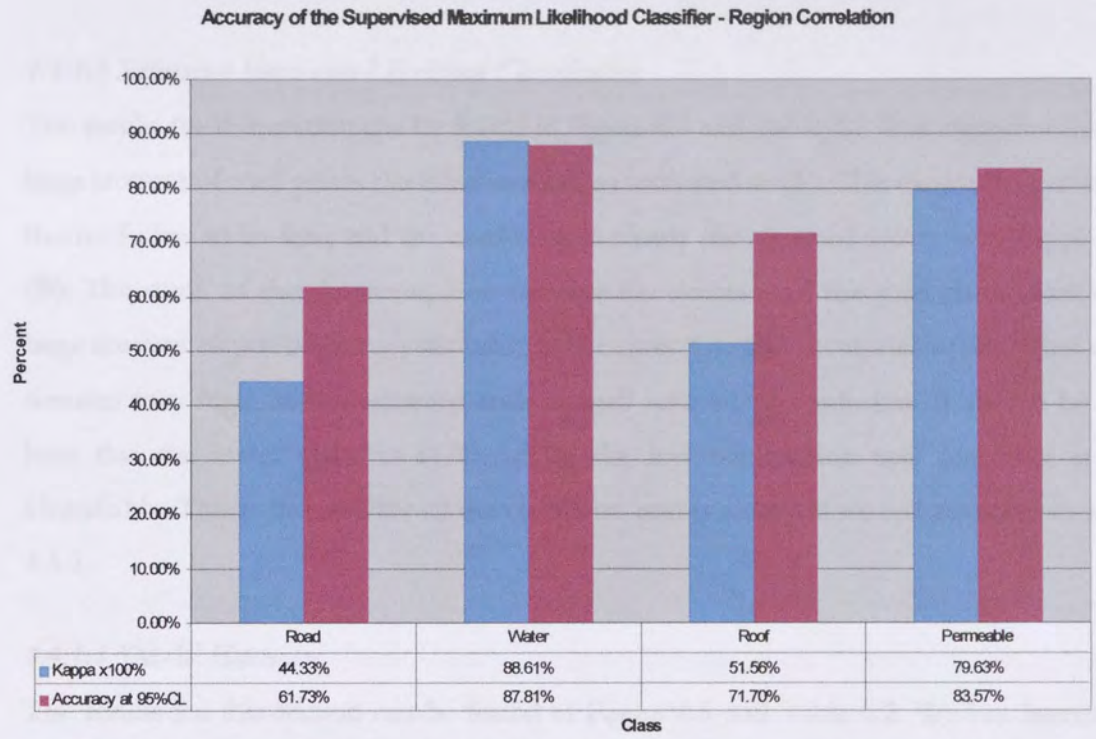


Figure 4.6. The Class Accuracies for the ECHO (Region Correlation method).

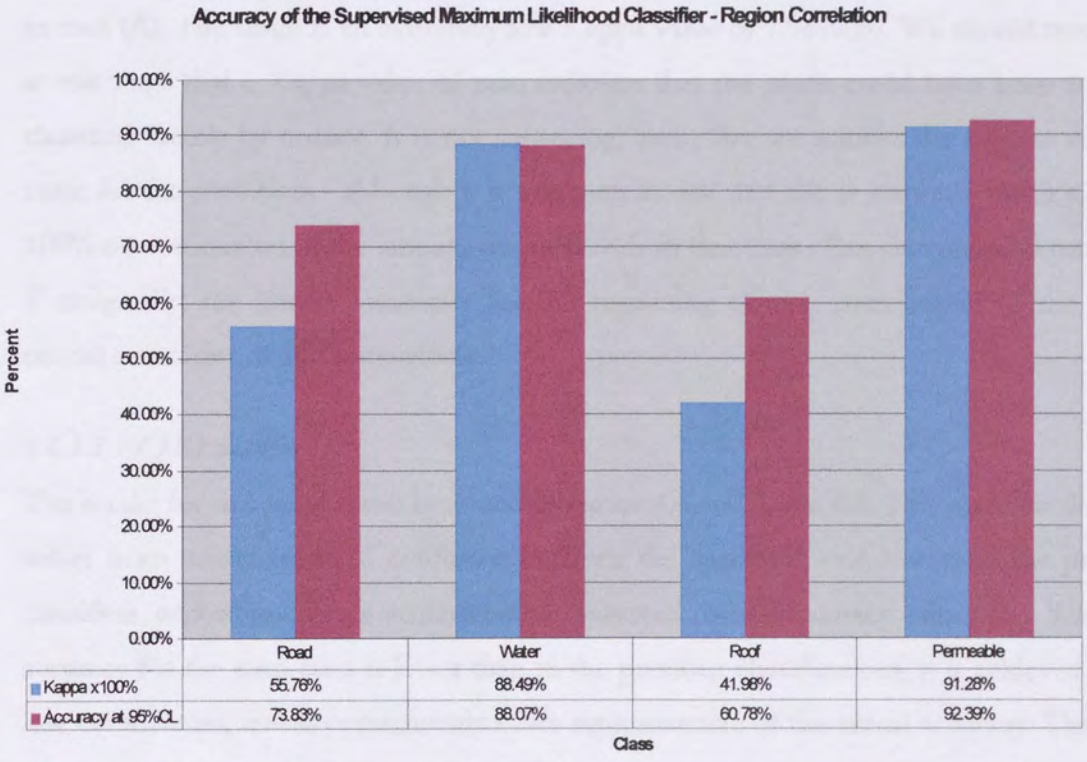


Figure 4.7. The Class Accuracies for the Unsupervised Maximum Likelihood Classifier (Region Correlation method).

4.4.1.3 Supervised Maximum Likelihood Classification

The results for this section can be found in Figure 4.4 and Table 4.1. It is clear that there is a large amount of roof pixels classified as road, as indicated in (**A**). This causes the accuracy of the roof class to be low, and the confusion is clearly the cause of a very low Kappa Value (**B**). The result of this, however, is to increase the accuracy of the road class, albeit with a large amount of pixels wrongly included in the class (**C**). The accuracies of the other classes remains very high, and is achieved with a small amount of confusion. It should be noted here that the water class, as indicated by the low commission and omission, is easily identifiable. This is the case for all the classifiers, and is a subject we will return to in section 4.4.2.

4.4.1.4 SMAP Estimator

The results for this section can be found in Figure 4.5 and Table 4.2. We can immediately see the confusion between the road and roof classes experienced with the previous classifier is increased in this classification, with almost 5 times more roof pixels classified as road than as roof (**A**). The result is an extremely low Kappa value of 7.98%(**B**). We should remember at this stage that a Kappa value of zero indicates that the pixels could have been correctly classified merely by chance. It is not surprising, then, that we achieve the highest accuracy value for the road class - although it is not coincidental that this is associated with almost a 100% overestimation of the amount of pixels truly in that class. This classifier also produces, if marginally, the lowest accuracies for the remaining classes, contributing to the lowest overall accuracies of all the classifiers.

4.4.1.5 ECHO classifier

The results for this section can be found in Figure 4.6 and Table 4.3. This classifier does not suffer from nearly as much confusion between the road and roof classes as the previous classifiers, and consequently achieves more balanced overall accuracy values (**A**). While the accuracy for the road class is lower than in the previous classifications, it is achieved with a low commission, and is consequently more representative of the actual accuracy. The cause of the low Kappa value is the amount of omission for this class (**B**). This will be discussed further in this chapter. The values for the water and permeable classes are once more very high, although the permeable class has the highest commission of any of the classifiers (**C**).

4.4.1.6 Unsupervised Maximum Likelihood Classifier

The results for this section can be found in Figure 4.7 and Table 4.4. This Classifier achieves the highest performance of all of the classifiers presented. Low values of omission, except in the roof class (A), contribute to this. As with the other classifiers, the confusion of the road class with the roof class (and vice versa) has the largest effect on the overall classification accuracy. With this classifier, however, the effect is less pronounced. The accuracies of the other classes, however, are exceptional, particularly in the permeable class (B).

4.4.2 Random single pixel

Classifying a map of disparate random pixels is by no means a trivial exercise. Locating each pixel and subsequently identifying it correctly, with the assistance of aerial photographs, is a time-consuming affair. Furthermore, the repetitive nature of the task means that errors will inevitably occur. However, if a truly unbiased assessment of accuracy is to take place, in the absence of a manually classified image covering the area of interest (from which correctly classified pixels can easily be extracted), this must be done.

Consequently, it is unrealistic to classify manually a vast amount of pixels. Conversely, Hay (1979) and Thomas *et al* (1987) state that it is desirable for each class to contain more than 50 pixels. So the question we must ask is: " How many random pixels do we require to fulfil this criteria?" Considering the classification of the imagery into 4 classes we can see, by inspection of the imagery, that class 2 (water) is likely to be a limiting region in this sense, since it has the smallest spatial coverage.

In order to answer the question posed above, we need to classify a number of pixels that is neither prohibitively excessive nor too few to give reasonable class sample sizes. The Grass program *r.random* allows us to generate a pixel set randomly distributed across the entire image. A stratified approach was taken, by manually identifying and classifying 500 and 1000 random pixels, giving us a comparison of the effect that 500, 1000 and 1500 pixels will have on the assessment of the accuracy of the classification procedures carried out in Chapter 3.

The results of the manual classification of the random pixels are given in Table 4.5, and the distributions are shown in Figures 4.8, 4.9 and 4.10.

Class 1 Road	Class 2 Water	Class 3 Roof	Class 4 Permeable	Total Pixels
180 (36.0%)	13 - (2.6%)	173 (34.6%)	134 (26.8%)	500
299 (29.9%)	15 (1.5%)	354 (35.4%)	332 (33.2%)	1000
479 (31.43%)	28 (1.87%)	527 (35.1%)	466 (31.1%)	1500

Table 4.5 The number of 'correct' pixels contained in each class for the manual classification of 500, 1000 and 1500 random pixels

We can see that for 500 classified pixels the number of pixels classified as water is 13. One would intuitively expect there to be approximately twice that amount for 1000 random pixels. Table 4.5, however, indicates that this is not the case. So while, from the figure given for 1500 pixels (28) we might expect to find at least 50 pixels out of 3000 classified as water, this may not be necessarily so. From the supervised Maximum Likelihood classification, the results indicate that water covers 2% of the image. This indicates that to get 50 single random pixels classified as water, 2500 pixels should be classified. In any event, it becomes obvious that to represent adequately such a small class, the overall amount of pixels that must be manually classified becomes unmanageable.

This raises questions about the significance of such a small class. It is perhaps fortunate that, when dealing with urban drainage, the amount of area classified as water is of no great significance. Ultimately the value that we require is the percentage of impermeable area, which we are able to extract. However, for other applications (for instance rural hydrology) the spatial extent of the pixels classified as water is an important value.

In the next section we will examine what confidence we can place in the accuracy of an under-represented sample size. Clearly, since the method for determining the accuracy at 95% assumes a large sample size (over 50), we expect poor results for this class. It is perhaps important to point out, however, that water, of all the types of land cover, is the most simple

to distinguish due to its particular spectral properties. As a result, we should not be overly concerned about the points raised above.

4.4.2.1 Effect of Sample Size on Distribution

Using the grass program r.random, an initial 500 pixels, randomly distributed across the entire image, were manually classified. The same procedure, again across the entire image, was repeated for 1000 different pixels. Combining these two maps provided us with a map of 1500 random pixels spread throughout the image.

We can observe that there is an effect. Rising from 500 to 1000 pixels we see a 5% decrease in the amount classified as road, and a 6% increase in the amount classified as permeable. Clearly the larger the sample the better represented the image is. To this end we would ultimately like to classify a greater number of random pixels. However, since this is unrealistic, it may prove the case that some combination of the random pixel method and the region correlation method may be made, causing a reduction in the bias of the first method while improving the overall representation of the second method. Clearly the results from either method, if taken out of context, are flawed.

While the initial manual classification classified each pixel into one of 11 classes, it became apparent that some classes were underrepresented. With 500 pixels divided into 11 classes, one could expect, with equal distribution, less than 50 pixels for each class. Since the distribution was far from equal, several classes were not represented at all. As a consequence, the 11 classes were combined into 4, as with the region correlation method. Even so, the water class was still poorly represented.

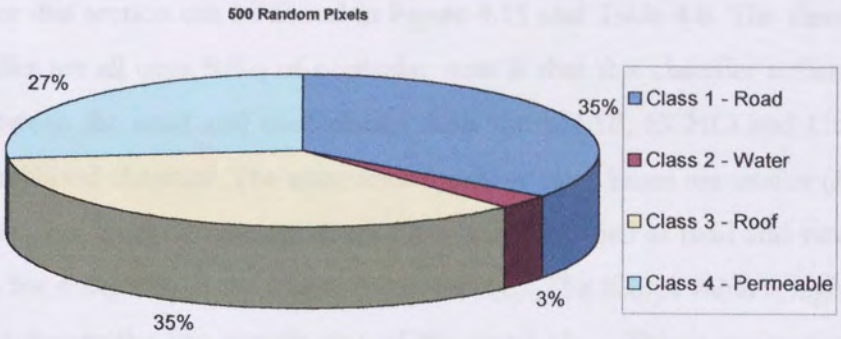


Figure 4.8 The Distribution of 500 manually classified pixels

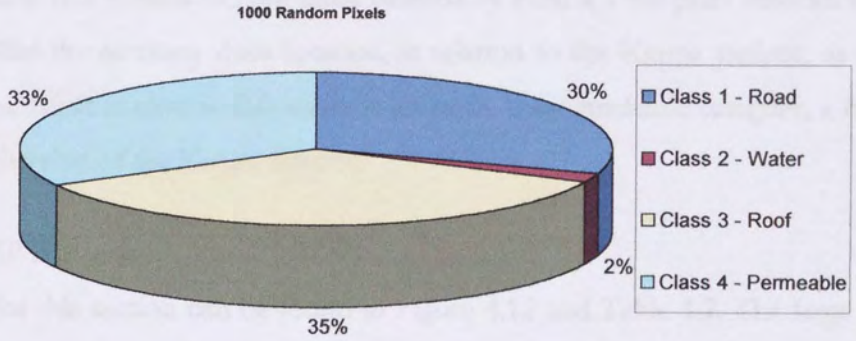


Figure 4.9 The Distribution of 1000 manually classified pixels

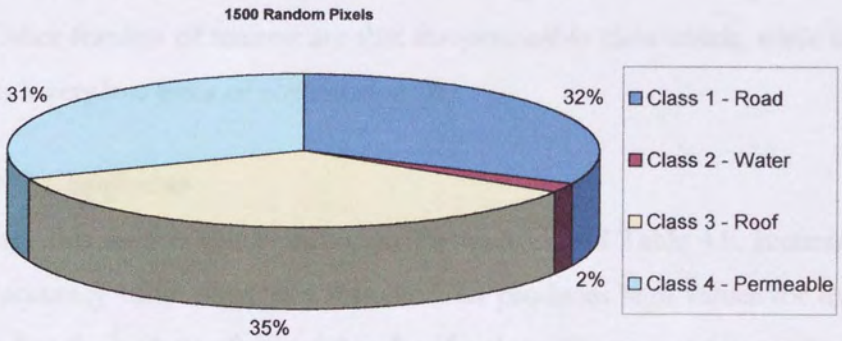


Figure 4.10 The Distribution of 1500 manually classified pixels

4.4.2.2 Supervised Maximum Likelihood Classification

This results for this section can be found in Figure 4.11 and Table 4.6. The class accuracies for this classifier are all over 50%; of particular note is that this classifier suffers less from confusion between the road and roof classes than the SMAP, ECHO and Unsupervised Maximum Likelihood classifier. The accuracies for these two classes are similar (**A**). There is some confusion, but an equal amount of roof has been classified as road and vice versa (**B**). The measures for water raise some interesting points (**C**). The Kappa value is higher than the accuracy value due to the low sample size of the water class. This is as we expected, and proves to be true for all the classifiers. This is because the assumptions and simplifications used to calculate the accuracies with associated confidence levels require a sample size of over 50. Clearly this criteria has not been fulfilled by even a 1500 pixel random sample, but we can see that the accuracy does increase, in relation to the Kappa statistic, as the sample size increases. What is clear is that water is an easily distinguishable category, a fact pointed to by the high value of the Kappa statistic.

4.4.2.3 SMAP Estimator

The results for this section can be found in Figure 4.12 and Table 4.7. The large amount of roof pixels classified as road has a large effect on the accuracy of this classifier. They result in high errors of commission for the road class, high errors of omission for the roof class and consequently a low accuracy and Kappa value for the roof class (**A**). It is interesting that the inclusion of road pixels into the roof class is, while significant, not as high as with the other classifiers. Other features of interest are that the permeable class which, while having a high accuracy, has a very low error of commission (**B**).

4.4.2.4 ECHO Classification

The results for this section can be found in Figure 4.13 and Table 4.8. Examination of the Kappa and accuracy value show that this classifier produces high values for the Permeable Class (**A**), in fact the highest of any of the classifications. Of greatest interest here is the high value of commission for this class. This would appear to be associated with the high values of omission for the remaining classes. The confusion matrix shows that a large amount of pixels from the road and roof classes (**B**) have been included in the permeable class, while

large amounts of road has been classified as roof and vice versa (**C**). The result of this is that the accuracies of the road and roof Classes are low, and the Kappa values are the lowest of all the classifications. This results in a poor overall accuracy and a very low overall value for Kappa.

4.4.2.5 Unsupervised Maximum Likelihood Classifier

The results for this section can be found in Figure 4.14 and Table 4.9. This classifier proves, with this method of accuracy assessment, to be the most accurate (**A**). This figure is a result of high individual class accuracy in three of the categories. The error of commission for the water class is, however, particularly striking (**B**). This is a consequence of the classification procedure, where 20 classes were extracted automatically and, by inspection, assigned to larger class. It is fortunate that the class that has an excessive error of commission is the smallest, and thus has the least effect on the remaining classes. Once more, with this classifier, we see that a large number of roof pixels have wrongly been labelled as road (**C**). As a result the Kappa value for this class is low and the omission high.

Supervised Maximum Likelihood Classification										
500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95% CL
Road	112	3	44	21	180	62.22%	37.78%	33.33%	42.41%	55.30%
Water	0	10	2	1	13	76.92%	23.08%	30.77%	76.26%	46.06%
Roof	57	0	99	17	173	57.23%	42.77%	29.48%	38.89%	50.00%
Permeable	3	1	5	125	134	93.28%	6.72%	29.10%	90.01%	89.05%
Total	172	14	150	164	500	72.41%	27.59%	30.67%	61.89%	68.80%
1000 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95% CL
Road	200	5	68	26	299	66.89%	33.11%	44.82%	50.28%	61.84%
Water	0	13	2	0	15	86.67%	13.33%	60.00%	86.37%	64.08%
Roof	112	3	217	22	354	61.30%	38.70%	32.20%	42.15%	56.54%
Permeable	22	1	44	265	332	79.82%	20.18%	14.46%	70.62%	75.75%
Total	334	22	331	313	1000	73.67%	26.33%	37.87%	62.36%	71.21%
1500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95% CL
Road	312	8	112**(B)**	47	479	65.14%	34.86%	40.50%	47.39%	61.19%**(A)**
Water	0	23	4	1	28	82.14%	17.86%	46.43%	81.70%**(C)**	65.31%**(C)**
Roof	169**(B)**	3	316	39	527	59.96%	40.04%	31.31%	41.06%	56.11%**(A)**
Permeable	25	2	49	390	466	83.69%	16.31%	18.67%	76.09%	80.58%
Total	506	36	481	477	1500	72.73%	27.27%	34.23%	61.56%	70.73%

Table 4.6 500, 1000 and 1500 random pixel confusion matrix for the Supervised Maximum Likelihood Classification.

SMAP Classifier										
500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95% CL
Road	143	1	31	5	180	79.44%	20.56%	61.67%	58.22%	73.67%
Water	4	9	0	0	13	69.23%	30.77%	7.69%	68.60%	35.42%
Roof	90	0	78	5	173	45.09%	54.91%	25.43%	27.36%	37.82%
Permeable	17	0	13	104	134	77.61%	22.39%	7.46%	71.00%	70.56%
Total	254	10	122	114	500	67.84%	32.16%	25.57%	56.30%	64.06%
1000 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95% CL
Road	231	2	55	11	299	77.26%	22.74%	78.93%	57.33%	72.76%
Water	0	14	1	0	15	93.33%	6.67%	26.67%	93.21%	76.76%
Roof	169	1	175	9	354	49.44%	50.56%	22.03%	32.31%	44.55%
Permeable	67	1	22	242	332	72.89%	27.11%	6.02%	63.27%	68.39%
Total	467	18	253	262	1000	73.23%	26.77%	33.41%	61.53%	70.76%
1500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95% CL
Road	374	3	86	16	479	78.08%	21.92%	72.44%**(A)**	57.79%	74.65%
Water	4	23	1	0	28	82.14%	17.86%	17.86%	81.80%	65.31%
Roof	259**(A)**	1	253	14	527	48.01%	51.99%	23.15%	30.68%**(A)**	44.08% **(A)**
Permeable	84	1	35	346	466	74.25%	25.75%	6.44%**(B)**	65.63%	70.57%**(B)**
Total	721	28	375	376	1500	70.62%	29.38%	29.97%	58.98%	68.57%

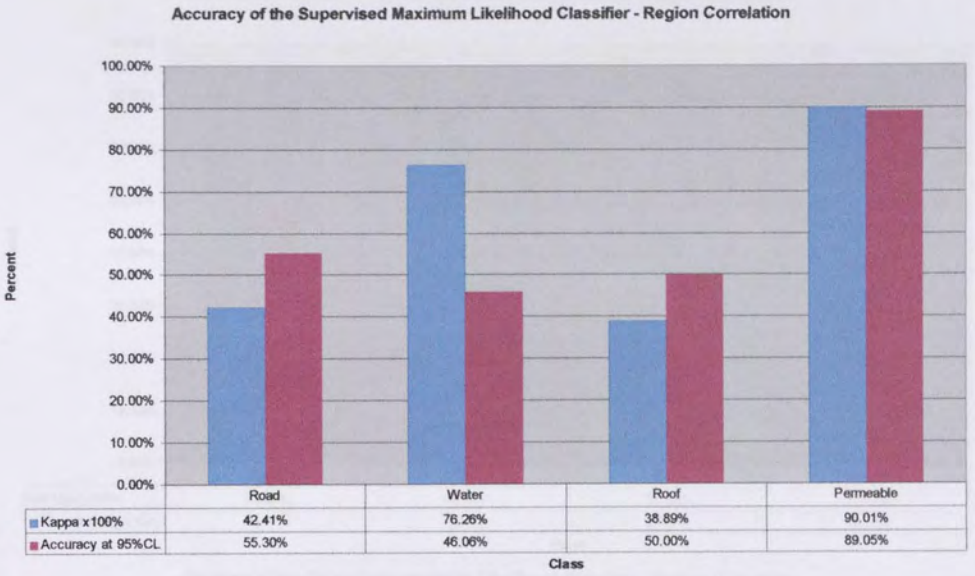
Table 4.7 500, 1000 and 1500 random pixel confusion matrix for the SMAP Classifier.

ECHO Classification										
500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	65	3	65	47	180	36.11%	63.89%	34.44%	14.36%	29.25%
Water	1	10	2	0	13	76.92%	23.08%	61.54%	76.06%	46.06%
Roof	55	4	72	42	173	41.62%	58.38%	41.62%	18.00%	34.42%
Permeable	6	1	5	122	134	91.04%	8.96%	66.42%	84.51%	86.21%
Total	127	18	144	211	500	61.42%	38.58%	51.00%	48.23%	57.48%
1000 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	118	9	99	73	299	39.46%	60.54%	40.13%	20.56%	34.22%
Water	0	15	0	0	15	100.00%	0.00%	93.33%	100.00%	100.00%
Roof	112	5	167	70	354	47.18%	52.82%	33.05%	26.22%	42.29%
Permeable	8	0	18	306	332	92.17%	7.83%	43.07%	85.79%	89.45%
Total	238	29	284	449	1000	69.70%	30.30%	53.40%	58.14%	67.14%
1500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	183	12	164**(C)**	120**(B)**	479	38.20%	61.80%	38.00%	18.33%	34.18%
Water	1	25	2	0	28	89.29%	10.71%	78.57%	88.94%	75.69%
Roof	167**(C)**	9	239	112**(B)**	527	45.35%	54.65%	35.86%	23.53%	41.44%
Permeable	14	1	23	428	466	91.85%	8.15%	49.79%	85.44%**(A)**	89.54%**(A)**
Total	365	47	428	660	1500	66.17%	33.83%	50.56%	54.06%	64.04%

Table 4.8 500, 1000 and 1500 random pixel confusion matrix for the ECHO Classification.

Unsupervised Maximum Likelihood Classifier										
500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	140	6	26	8	180	77.78%	22.22%	45.00%	60.18%	71.84%
Water	1	10	0	2	13	76.92%	23.08%	161.54%	75.40%	46.06%
Roof	72	14	71	16	173	41.04%	58.96%	15.03%	26.85%	33.85%
Permeable	8	1	0	125	134	93.28%	6.72%	19.40%	90.38%	89.05%
Total	221	31	97	151	500	72.26%	27.74%	60.24%	63.20%	68.63%
1000 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	209	16	64	10	299	69.90%	30.10%	51.17%	52.82%	64.97%
Water	0	14	0	1	15	93.33%	6.67%	233.33%	92.99%	76.76%
Roof	128	16	188	22	354	53.11%	46.89%	19.49%	36.89%	48.23%
Permeable	25	3	5	299	332	90.06%	9.94%	9.94%	85.12%	87.03%
Total	362	49	257	332	1000	76.60%	23.40%	78.49%	66.95%	74.24%
1500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	349	22	90	18	479	72.86%	27.14%	48.85%	55.61%	69.18%
Water	1	24	0	3	28	85.71%	14.29%	200.00%**(B)**	84.91%	70.33%
Roof	200**(C)**	30	259	38	527	49.15%	50.85%	18.03%	33.44%	45.21%
Permeable	33	4	5	424	466	90.99%	9.01%	12.66%	86.71%	88.58%
Total	583	80	354	483	1500	74.68%	25.32%	69.89%	65.16%	72.72%**(A)**

Table 4.9 500, 1000 and 1500 random pixel confusion matrix for the Unsupervised Maximum Likelihood Classifier.



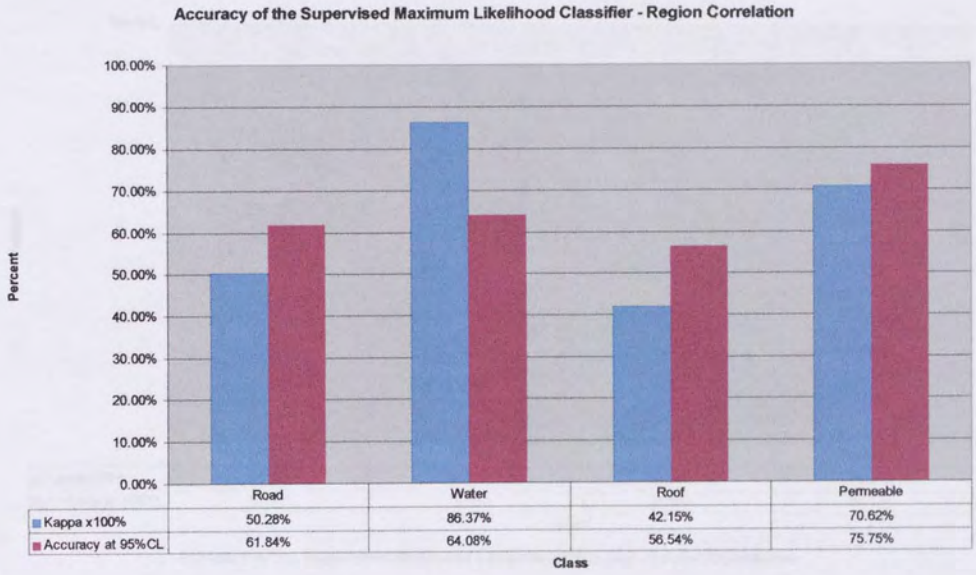
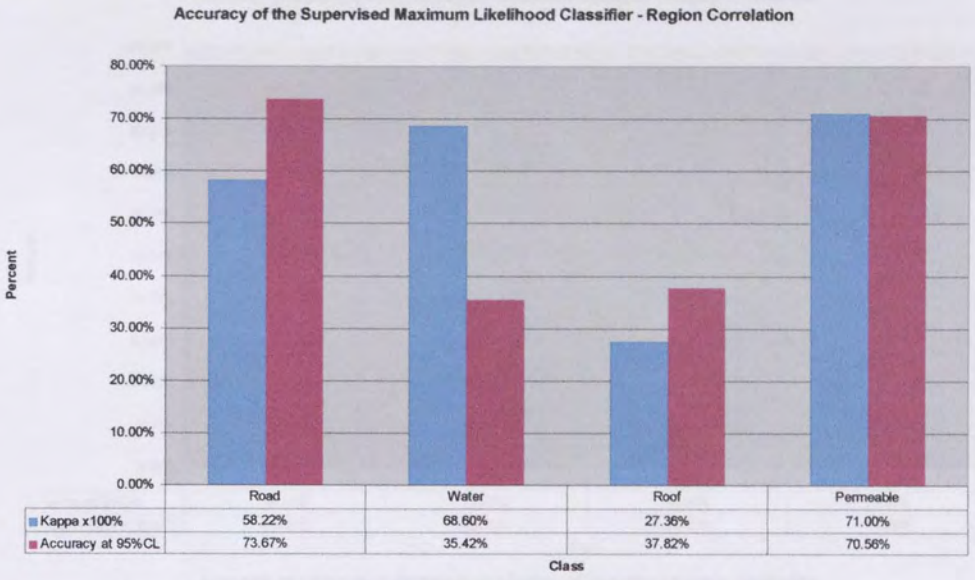
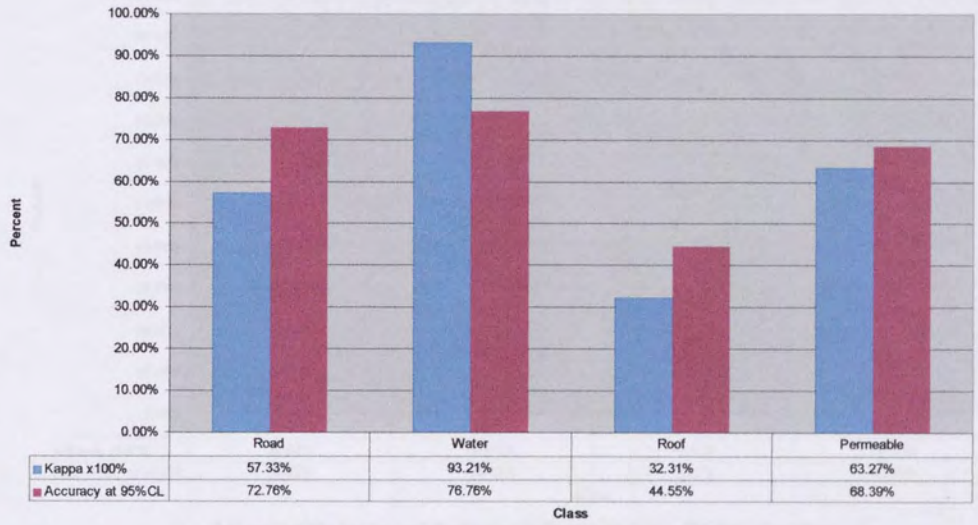


Figure 4.11 Accuracy of the Supervised Maximum Likelihood Classifier (single random pixel method).





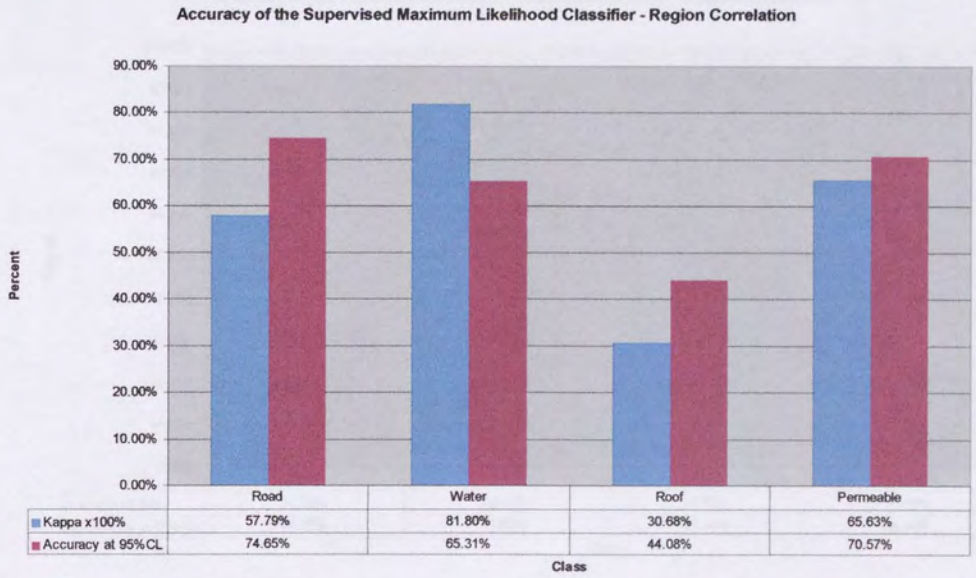


Figure 4.12. Accuracy of the SMAP Classifier (single random pixel method).

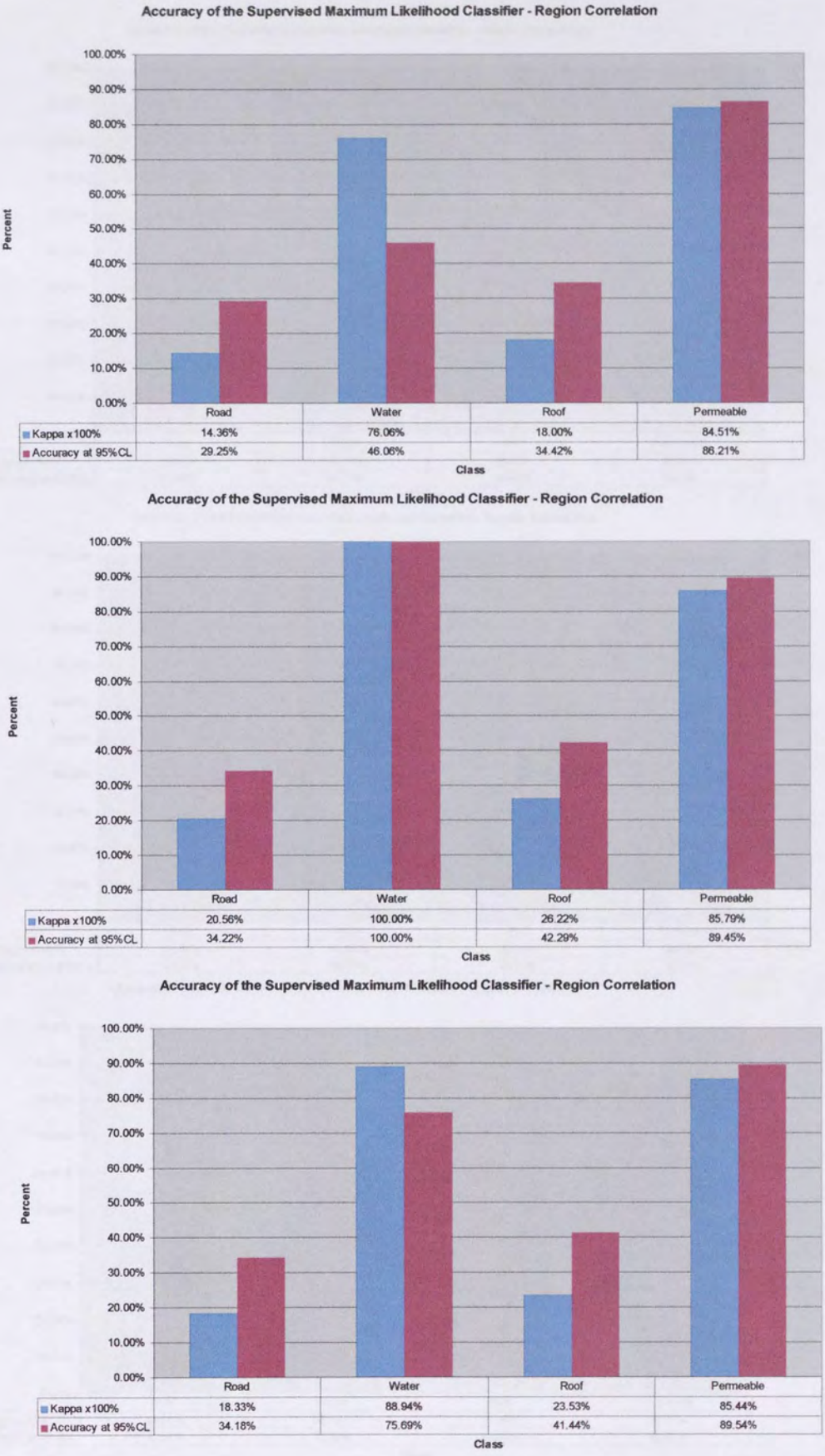
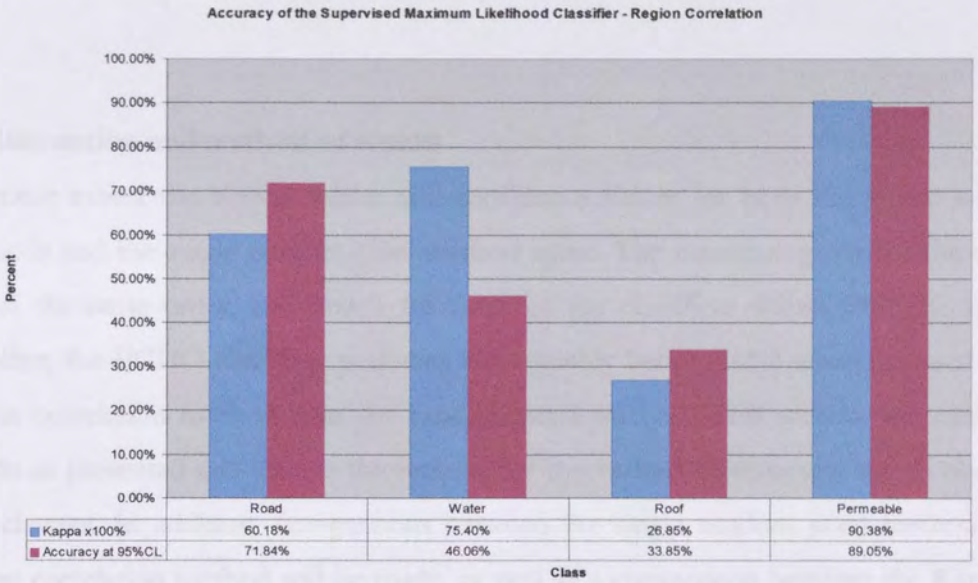
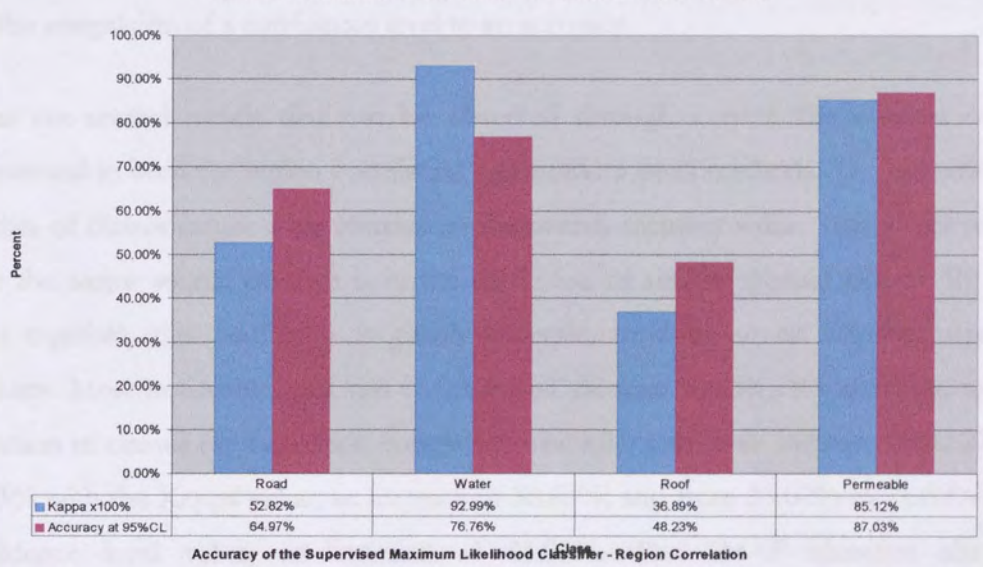


Figure 4.13. Accuracy of the ECHO Classifier (single random pixel method)





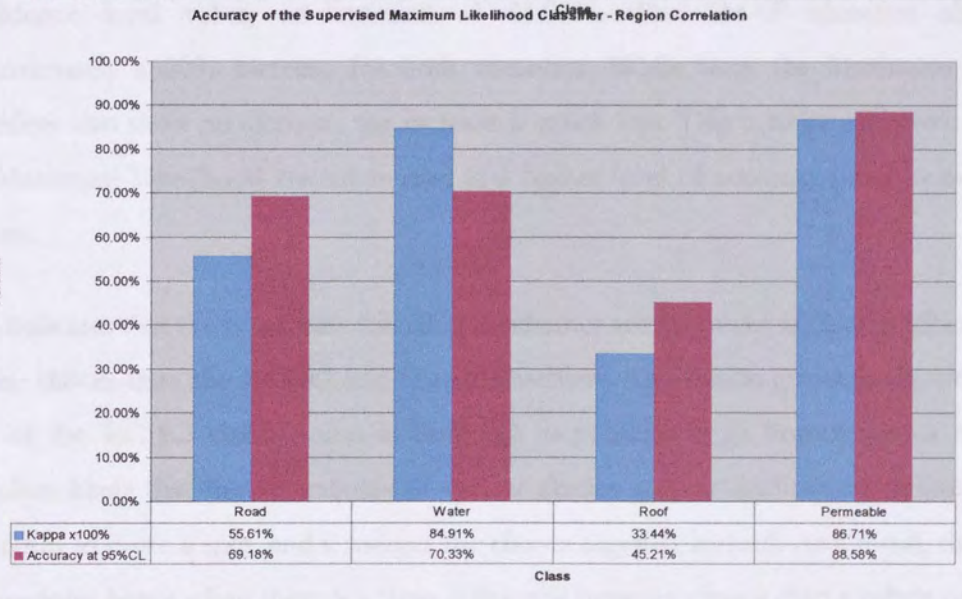


Figure 4.14 Accuracy of the Unsupervised Maximum Likelihood Classifier (Single Random Pixel).

4.5 Discussion and analysis of results

To some extent the Kappa values and confidence values for both the region correlation methods and the single random pixel method agree. The measures given for the classifiers are of the same order, and match for three of the classifiers within 5%. The remaining classifier, the ECHO classifier, performs considerably better (16%) when analysed with the region correlation method than the random pixel method. This section will examine the results as presented and discuss the reasons for the various features and trends observed in this chapter. In addition, comparisons between the single random pixel method and the region correlation method will be made, as well as a comparison between the Kappa value and the assignation of a confidence level to an accuracy.

There are several trends that can be observed through varying the number of classes represented in both the region correlation and random pixel methods. The reduction in the number of classes causes a big increase in the overall accuracy value. This is not surprising, since the major source of error is in the confusion of similar spectral classes. By merging these together, this confusion is greatly reduced, resulting in an improvement in the accuracy. Most noticeably, this can be observed through the ECHO classifier, where the reduction in classes (in the region correlation example) causes an increase from 27.38% to 66.03% with the Kappa value, an increase of 38.65%, and from 33.62% to 76.68% with the confidence level value, an increase of 43.06%. The SMAP classifier also shows approximately a 20% increase for both measures. While both the Maximum likelihood classifiers also show an increase, the increase is much less. This is to be expected, given that the Maximum Likelihood classifiers start at a higher level of accuracy prior to merging the classes.

This indicates that the Maximum Likelihood schemes are far better at distinguishing between similar classes than the ECHO and SMAP classifiers. The region growing algorithm that is part of the ECHO classification is designed to produce large homogeneous areas. It is therefore likely that the boundaries of similar classes are not sufficiently different for the algorithm to cause a split, and it merges the classes together instead. As a result, the classifier will perform better when there is a large difference between classes than a subtle one.

The SMAP classifier is also designed to produce large homogeneous regions. This is achieved by re-sampling the image from coarse resolution to fine. This means that when spectrally similar classes are of a small spatial extent, they are likely to be classed as belonging to the same class. In agricultural satellite imagery, where the purpose is to extract large homogeneous regions such as fields, this is apparently successful (McCauley and Engel, 1995). However, when there are many spatially small classes, such as in urban imagery, this approach is less successful and hence a less appropriate classification method.

The Maximum Likelihood classifiers, then, through their absence of any neighbourhood smoothing, are better equipped to deal with this. There is still an improvement to be gained from reducing the number of classes, however. The better performance achieved with the unsupervised method over the supervised is also misleading. The initial number of clusters (20) identified by the unsupervised method were combined to give 10 classes, the selection being made by the operator as seen in Chapter 3. The reason for this was that some classes (particularly the railway class) did not appear as a single cluster, but were made up of many different clusters. In effect this meant that the railway class had to be discarded. As a result, although this was a small class, if the imagery in Chapter 3 is examined, the railway line is classed as road. In effect, if the random pixels or region selected had included any members of this class (it did not), the unsupervised method would have performed less well. Consequently, for this to be reflected in the single region pixel method, a greater number of pixels must be classified.

Having established the reasons for fewer classes resulting in greater accuracy, the effect of sample size on the accuracy should be examined. The agreement between the two methods suggests that a sample size of 1500 pixels can provide similar results to that of nearly 20,000 pixels, as used in the region correlation method. The smaller samples tested, however, indicate that this is the very smallest amount that can be used. The results appear valid, but an increase in the sample size clearly affects the accuracy assessment.

Where the single random pixel method excels, however, is in the absence of bias. This impartiality may mean that a much smaller sample is required. While attempts were made to reduce the bias of the region correlation method, through selecting an entire area and classifying the pixels in it of known land use, the distribution may not have been

representative of the entire image itself. For instance, a much higher proportion of factory roof and car park was contained in the area than in the rest of the image. This results in better representations of some classes than others, with this bias being reflected in the accuracy of those classes.

If Figure 4.15 is examined, it becomes apparent that the region correlation method produced a sample that was particularly favourable towards the ECHO classifier. The other classifiers perform very similarly with either sampling method, which suggests that the single random pixel method is the most appropriate method for assessing accuracy. The absence of bias means that, if the sample size is sufficient, it provides a truly objective account of the performance of each classifier, making direct comparison between the classifiers possible.

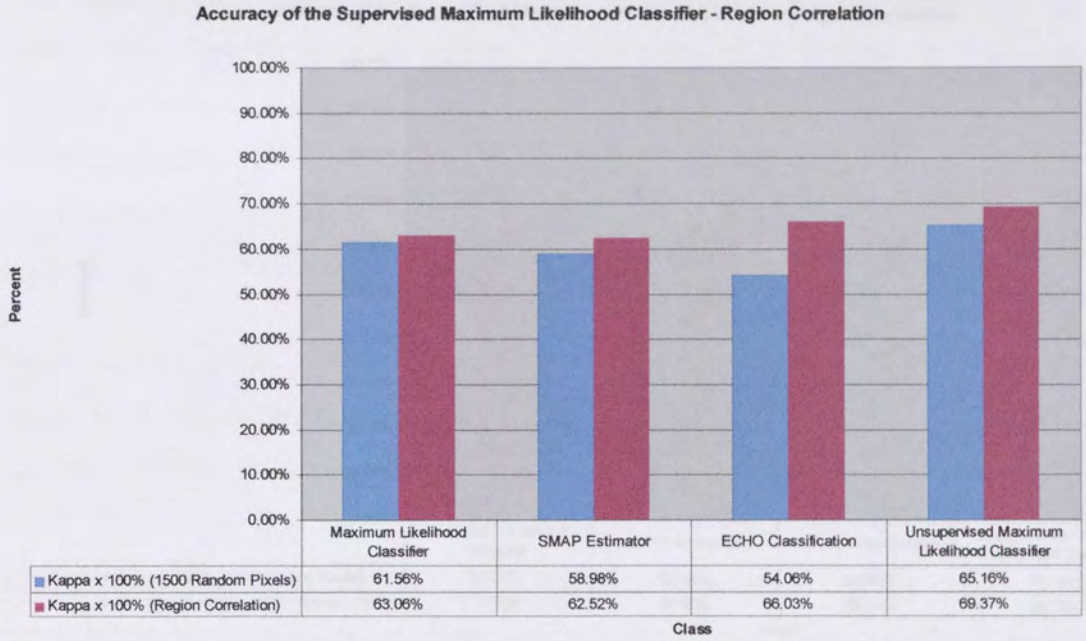
The Kappa value and accuracy at 95% confidence also appear to support this. In Figure 4.16, both measures follow an almost identical trend for the Single Random Pixel method. For the Region Correlation method, however, both measures follow a similar trend except with the SMAP classifier. The high Kappa value and low accuracy at 95% confidence level indicate, once more, that for this classifier the errors of omission and commission were relatively low, while the standard deviation in the sample was also small. This backs up the assertion of the previous paragraph.

Accepting the Single Random Pixel method as the most suitable sampling method, there appears to be no real difference between the Kappa value and the confidence level when it comes to direct comparison between classification methods. Both measurement techniques indicate the same trend almost exactly on an unbiased sample. It is only in poorly represented samples, such as the individual accuracy of water, that there is a great difference between the two measurement techniques. The Kappa value is not greatly affected by a small sample size whereas the calculation of the confidence level is.

As a result it is difficult to choose between the two accuracy assessment methods. They both have their merits, namely that the Kappa value is very responsive to large errors in omission and commission, whereas the confidence level is very sensitive to small sample sizes. The information that both techniques can provide in an adequately sampled class is very similar, however, and hence there is little to choose between the two. However, the fact that both

are sensitive to different scenarios means that quoting both provides a more complete picture. If there is a small sample size, this will be reflected in the confidence level of the accuracy, providing essential information when using the single random pixel method. Additionally, the Kappa value is easier to calculate than the confidence level, making mistakes less likely.

The provision of two values may, however, be impossible to achieve in a real world usage, as it may be misleading and difficult to interpret. Consequently, assuming the sample size is adequate, the single random pixel method with accuracy measured by the Kappa statistic is recommended for urban drainage applications.



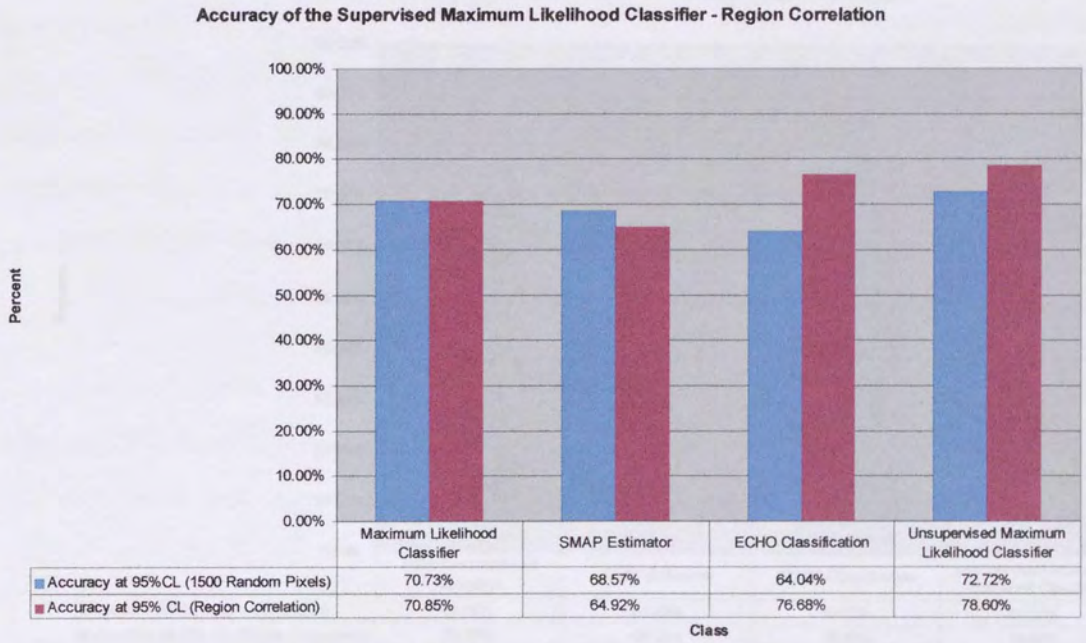
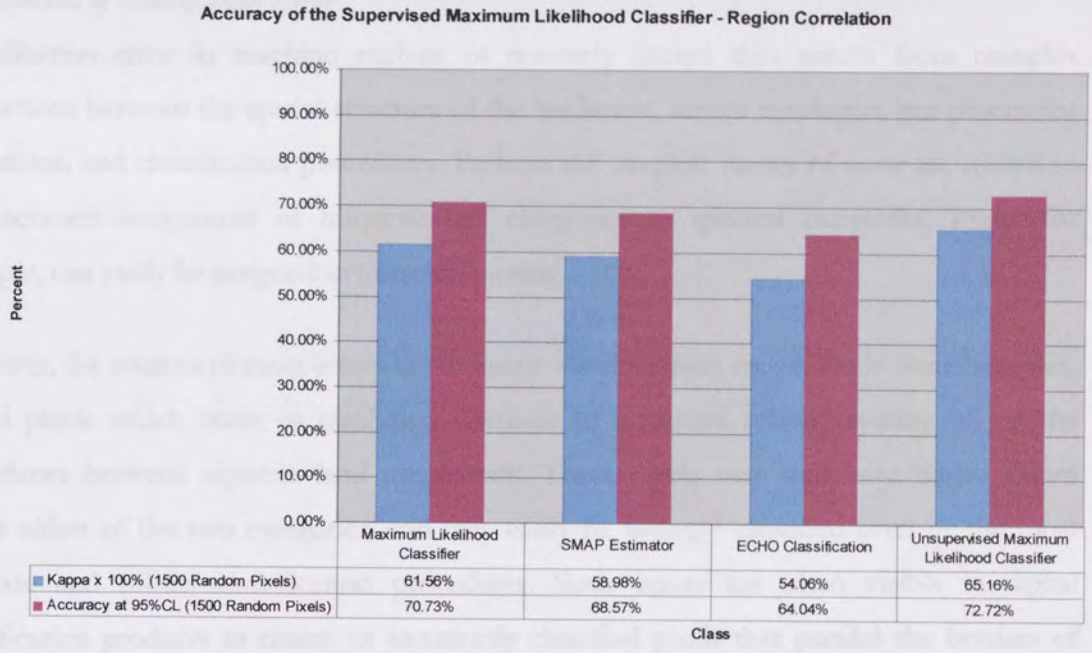


Figure 4.15 A Comparison between the Single Random Pixel method and the Region Correlation method for both measurement techniques.



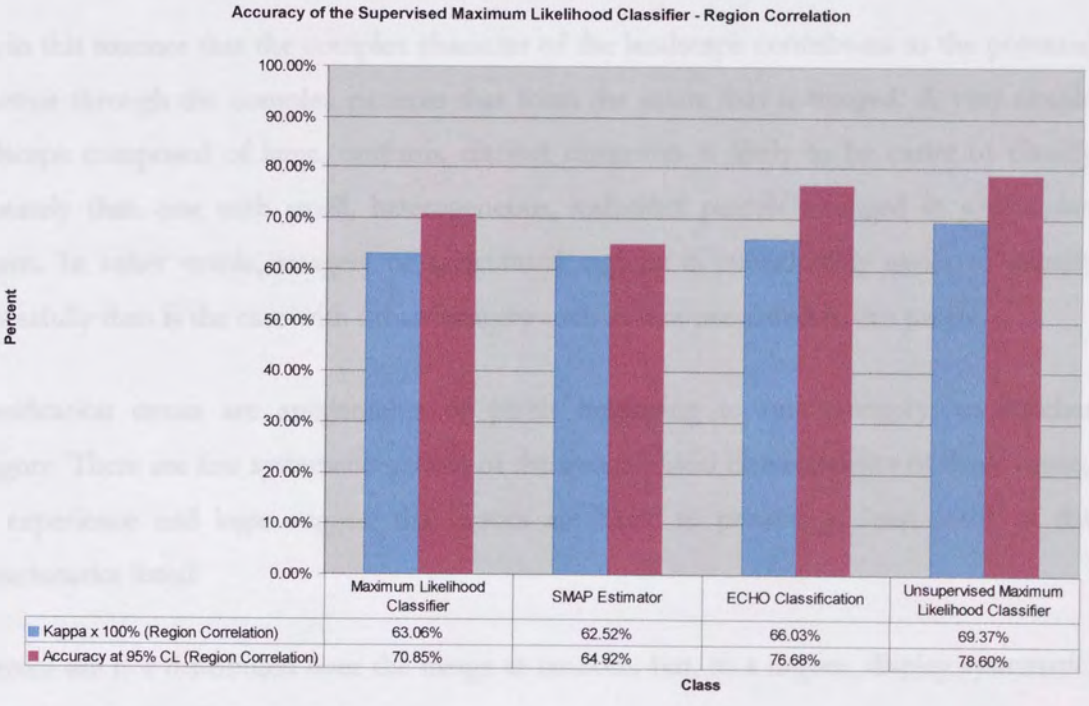


Figure 4.16. A comparison between the Kappa Value and the Confidence level at 95% for the two sampling methods.

4.5.1 Sources of Classification Error

Classification error in machine analysis of remotely sensed data results from complex interactions between the spatial structure of the landscape, sensor resolution, pre-processing algorithms, and classification procedures. Perhaps the simplest causes of error are related to the incorrect assignment of informational categories to spectral categories. Road, for example, can easily be assigned to terraced housing.

However, the sources of most errors in automatic classifications are probably more complex. Mixed pixels which occur as resolution elements of a remote sensing system fall on the boundaries between separate land use parcels. These pixels may well have digital values unlike either of the two categories, and may easily be wrongly classified even by the most accurate and robust classification procedures. Such errors are often visible in digital classification products as chains of incorrectly classified pixels that parallel the borders of rather large, homogeneous, parcels.

It is in this manner that the complex character of the landscape contributes to the potential for error through the complex patterns that form the scene that is imaged. A very simple landscape composed of large, uniform, distinct categories is likely to be easier to classify accurately than one with small, heterogeneous, indistinct parcels arranged in a complex pattern. In other words, imagery of agricultural regions is considerably easier to classify successfully than is the case with urban imagery such as that presented in this project.

Classification errors are assignments of pixels belonging to one category, to another category. There are few systematic studies of the geographical characteristics of these errors, but experience and logic suggest that errors are likely to possess at least some of the characteristics listed:

1. errors are not distributed over the image at random, but, to a degree, display systematic, ordered occurrence in space;

2. errors are not distributed at random among the various classes on the image, but may be preferentially associated with certain classes;

3. erroneously assigned pixels are often not spatially isolated, but occur grouped in areas of varied size and shape;

4. errors are not distributed randomly among the various parcels on the image, but may occur in parcels with certain sizes, shapes, locations, and arrangement in respect to other parcels;

5. errors may have specific spatial relationships within the parcels to which they pertain; for example, they may tend to occur at the edges or in interiors of parcels.

Todd (1980) experienced three categories of classification error, in the mapping of a wildlife project. He summarised the quality assurance checks and error analyses, which preferably accompany each computer interpretation summary map as:

1) geometric and radiometric problems in the image accounted for about 5 - 15 percent of the errors;

2) excessive category detail (i.e. attempts to extract classes whose spectral characteristics approached the noise level of the data) accounted for another 30-40 percent;

3) analyst decisions in identifying spatial clusters, which required considering a complex interaction between vegetative and terrain characteristics, accounted for the rest of the errors.

It should be noted that in some cases there might be sources of error other than the classification system. The matching of sites on the imagery with exact locations in the field is subject to error. The data used in this project, due the problems described in Chapter 2, is not geo-rectified, however, and is not affected by this problem. Of greater significance is the determination of the exact pixel class from the ATM imagery, where the only source of ground truth comes from aerial photographs. The accuracy of this method is not 100 % due to the presence of mixed pixels and the occasional pixels whose correct class cannot be identified. These errors have not been included in the assessment of accuracy, however, because they are very small.

4.6 Conclusion

This chapter has examined several methods for assessing the accuracy of a automatically classified image. The use of random and non-random reference data has been compared, along with, in the case of the former, varying sample size.

This section has demonstrated that the use of enough single random pixels as a basis for the accuracy assessment can lead to an unbiased measurement. This must surely be the aim if any faith is to be put in the procedure.

However, a single measure of accuracy is both misleading and unsuitable. Without suggesting that the accuracy figure should be as complicated as the imagery itself, the provision of several key values, such as omission, commission and an overall accuracy value would seem the very least that one should accept when making use of automatically classified imagery.

As a result it is suggested that the use of the single random pixel method be implemented when gathering a sample upon which to perform an accuracy assessment. Since the only problem with this method is that certain categories may be underrepresented, it would appear prudent to measure the confidence level at 95% for each sample. Additionally the Kappa value should be measured, since it provides information about the errors of commission and omission of the data. When both measures are used the suitability of the sample size can be determined by comparing the two values. If the confidence level is lower that the Kappa value, the sample size should be increased. This results in an adequate sample that can then be assessed in an unbiased manner.

This analysis of the data provides information about the classification accuracy that is thorough and demonstrably unbiased. Measurement of data in this fashion will place true confidence in the accuracy of a classification due to its unbiased nature and unambiguous measurement.

Chapter 5. Data Processing : Contextual classification using homogeneous polygons

5.1 Introduction

In this chapter the principles behind the development of the polygon classifier are presented and discussed. The classifier is a multi-spectral contextual classifier; that is, it utilises the statistical properties of the dataset *and* augments this data with the introduction of context, based on visual characteristics inherent in the imagery.

The development of the classifier is described as a series of stages. Figure 5.1 gives a simplified outline of the procedure. A Marr Hildreth zero-crossing edge detector (Marr and Hildreth, 1980) is applied to a single image of the data-set. This image should contain as much information about the data as possible in order to produce an edge map rich in detail. This edge map is then used to produce a series of coarse polygons, each of which can be extracted individually from the image.

These polygons are tested for homogeneity through the application of a region growing algorithm. This algorithm splits regions that are not homogeneous into finer polygons until they become homogeneous. These homogeneous areas are then classified on a per-polygon basis, assigning them the class which is predominant in each polygon. The assignment of a class to each pixel in the original imagery is made through use of either supervised or unsupervised Maximum Likelihood statistics. As a result of this procedure a classified map containing a series of homogeneous regions is produced.

The following chapter explains this procedure in greater detail, and examines the reasoning behind the approach to each stage.

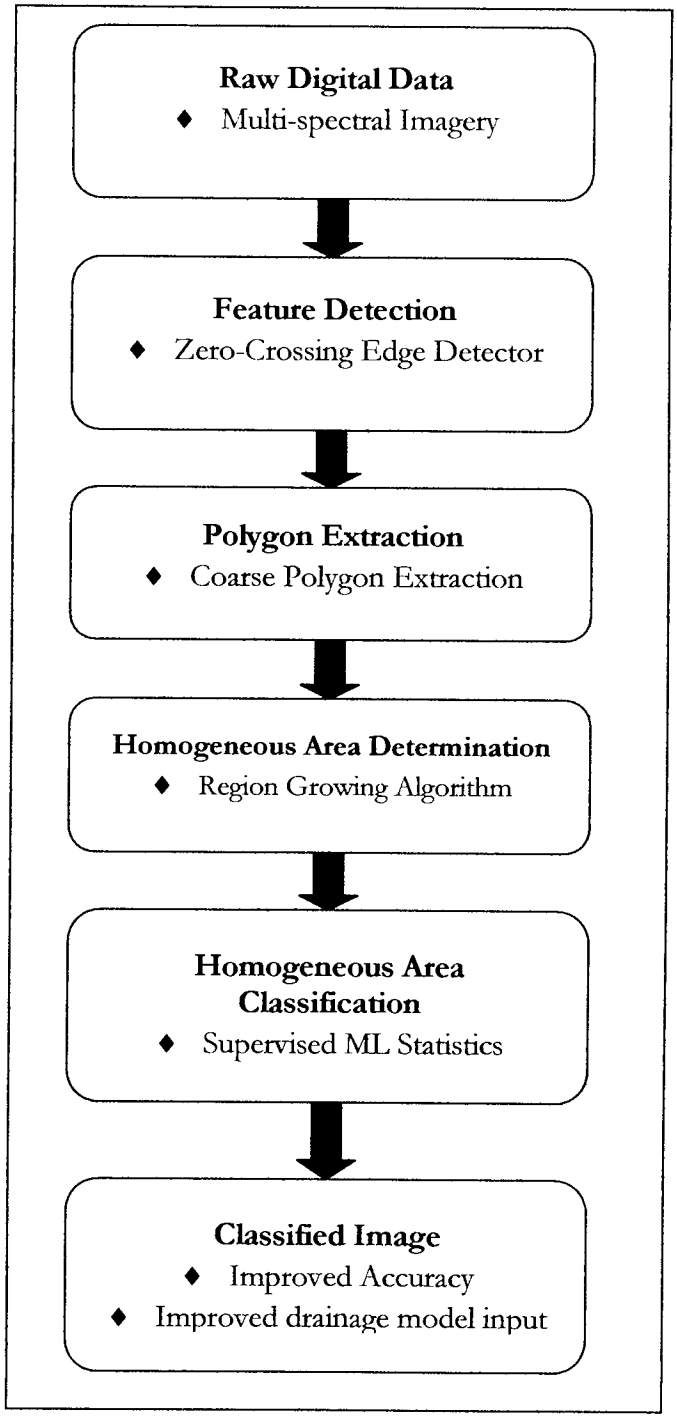


Figure 5.1. The Flowchart describes a simplified overview of the Classification technique developed.

5.2 Edge Detection

5.2.1 Introduction

The first step in the classification process is to detect lines representing the boundaries between individual homogenous polygons. The Marr-Hildreth edge detector (Marr and

Hildreth, 1980), is based on the zero-crossings of a Laplacian convolution mask (Chapter 2). To recapitulate: the Laplacian is a two dimensional measure of the second derivative of an image. This introduces an increased sensitivity to noise, which is countered by incorporating Gaussian smoothing into the edge detector. By applying non-maximal suppression to the zero-crossings of the image, we are left with an isocline map of the edges found within the image. This means that the image produced has edges that always lie on closed polygons in a binary image. Further detail can be found in Chapter 2.

While other edge detectors (Canny, 1986; Smith and Brady, 1995) are capable of providing considerably more accurate edge location, they are of little use when spatial information about objects (for instance, roads and houses) is required. The reason is that the criterion for the development of these edge detectors is concerned with the mathematical properties of the imagery, namely the exact location of the image brightness gradients that constitute edges. As we have seen in Chapter 2, the Canny edge detector was designed to be an optimal edge detector in terms of quality and computational efficiency.

The Marr-Hildreth edge detector, however, was developed using visual perception theory that maintains that the edge isoclines contain the most important structural properties of the imagery. According to this theory, the purpose of early visual processing is to construct a primitive but rich description of the image. This is to be used to determine the reflectance and illumination of the visible surfaces, and their orientation and distance relative to the viewer. This first primitive description is known as the raw primal sketch. These zero-crossing segments are found to be extremely rich in information (Marr and Hildreth, 1980).

An edge detected image produced by the Canny edge detector is presented in Figure 5.2. Upon first inspection, the image appears to be rich in detail. However, closer inspection reveals that many edges are unconnected at both ends. Furthermore, inspection of the road network and railway line show that they are not continuous throughout the image. This means that the each polygon, that we would like to contain an entire class, is broken down into several smaller polygons and must be treated separately. In effect, the features in the image have not been located.

If the image is to be used to produce a map of enclosed polygons, the lines that are not connected at both ends must be discarded. As a result, much of the detail in the image is

lost and a very coarse polygon map is left. This map contains little of the structure visible in the original image and is therefore unsuitable for polygon extraction.

Figure 5.2 The Canny Edge Detector applied to a section of the Dudley Imagery (Band 11)

5.2.2 Zero – crossings

This section presents a series of zero-crossings taken from the same image, Band 11 (Figure 5.3). A minor point is that there are a few areas of burnout in the image. This does not cause a loss of data as Band 12 can be used instead, although with lower resolution. The purpose is purely to illustrate which scaling of the feature detector is best suited to our classification. The selection of Band 11 is arbitrary at this stage - we will examine the use of different bands/combinations of bands in section 5.2.3. At this stage

we are merely concerned with finding a suitable feature detector for this type of imagery at this scaling *per se.*

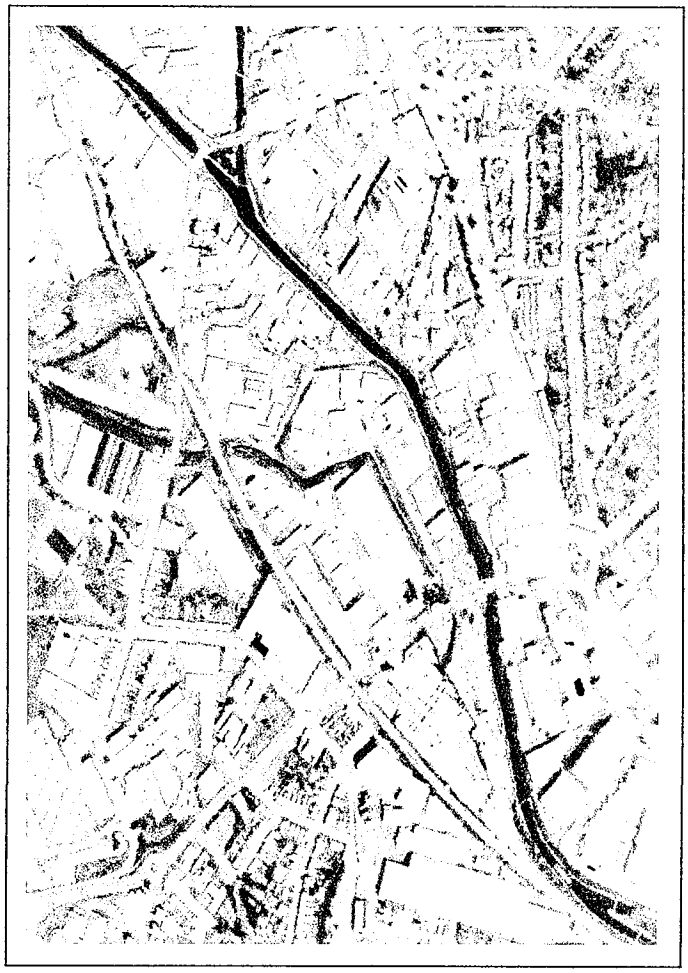


Figure 5.3. A section of the Dudley Imagery (Band 11).

It is possible to alter several parameters of the Marr-Hildreth edge detector, so that it can detect features of varying scale. One of these parameters is the width of the Gaussian component, making it more or less sensitive to the edge gradient. An examination of these adjustments follows.

We see that the edge detection at coarse scales (Gaussian = 15) is unsuited to imagery of this complexity. We lose most of the detail in the imagery and are given only a slight indication of the relationship between crudely delineated objects (Figure 5.4).

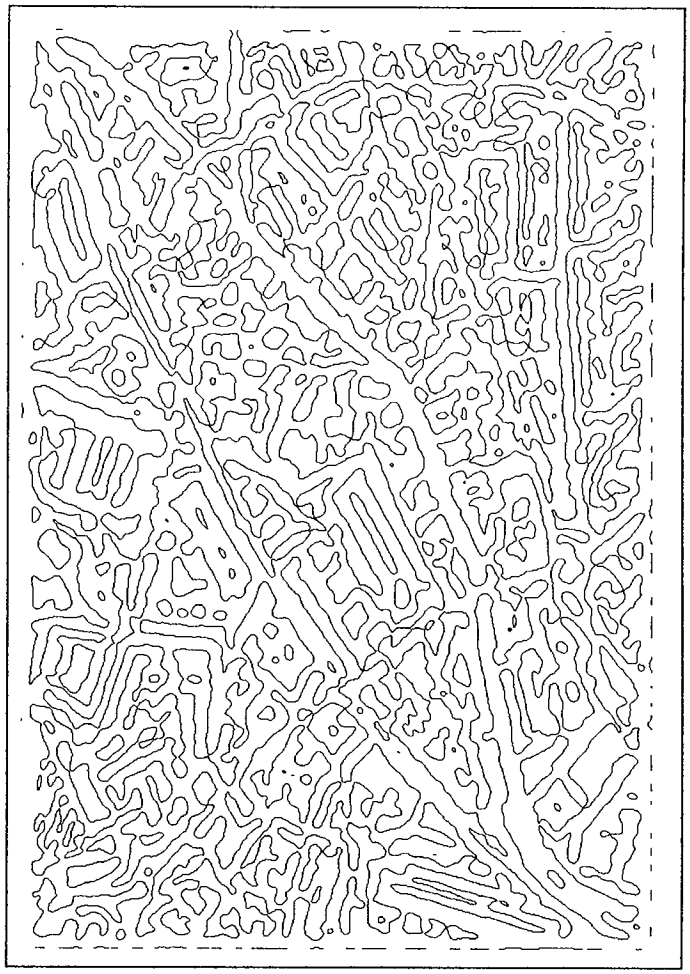


Figure 5.4. Zero-Crossings of image 5.3, Gaussian = 15.

Making the edge detector finer (Gaussian = 9) results in more of the image structure being present, although we still lack fine detail of the kind necessary for accurate feature detection (Figure 5.5).

At the finest level (Gaussian = 5, Figure 5.6), we are presented with a rich image, in which the structure of the image in clearly visible. We are able to identify various features of the imagery. The canal and railway tracks are, for instance, well represented.

This is in keeping with Marr and Hildreth's findings. The reason that the coarse scale (Gaussian = 15, Figure 5.4) zero-crossings present little useful information is that the image is highly detailed and structurally complicated in nature.

A further problem can also be observed if the zero-crossing edge images in this section (Figures 5.4 – 5.6) are closely examined. The effect of forcing the image to contain closed intensity contours is such that edges on the border of the image are either ignored or forced to fit in such a way that one very large contour is made to exist. This provides continuity throughout the image, but at a cost of discarding useful edge information at

the border of the image. In order to overcome this loss, we can artificially crop the image after the edge detection and thus remove these spurious edges. It should be remembered that the images covering a catchment will all overlap, so there is no loss of data as a result of this procedure.

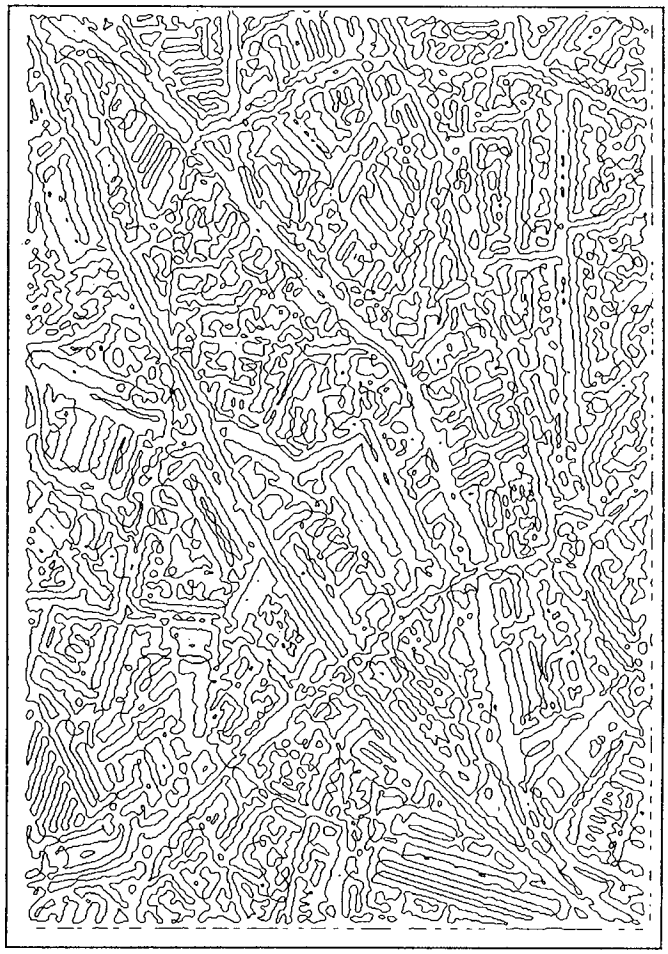


Figure 5.5 Zero-Crossings of Image 5.3, Gaussian = 9

5.2.3. Band selection

Having determined which scales and parameters are best suited to imagery of this type, we need to make sure that the most salient imagery is utilised. Initially, the approach favoured involved taking the zero-crossing of each band and combining them. Those edges that were supported by the greatest number of bands would then be chosen as the most representative features. However, it was found that each image band indicated vast quantities of different features, and very few related features. This was clearly not surprising, as the wavelength ranges of each band are carefully selected for this reason.

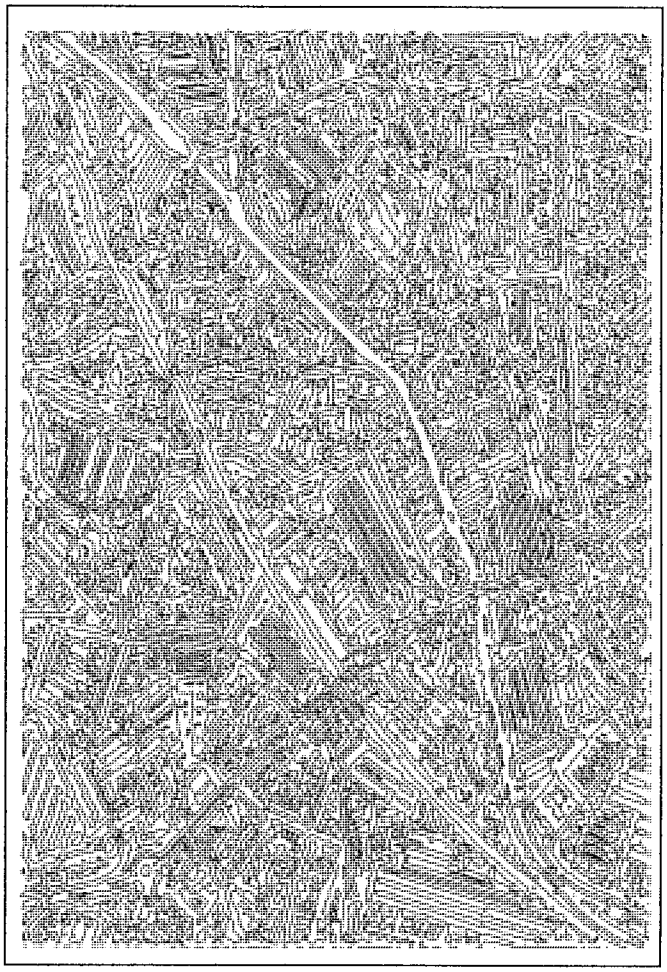


Figure 5.6. Zero-Crossings of Image 5.3, Gaussian = 5.

What was surprising, however, was the lack of correlation between the edge images. This was because of the inexact nature of the feature detector. The likelihood of the same feature extracted from a different band giving an edge at the same point is low, since we must remember that the Marr-Hildreth edge detector provides us with contours of image variation, and not exact edges.

The method detailed above, while being a perfectly reasonable approach, does not utilise the main property of the Marr-Hildreth edge detector. Being a feature detector, based on visual perception theory, its purpose is to extract the features that humans see in a *single* image, not a collection of images.

Another method is to combine the data prior to the edge detection through the use of principal component analysis. This is a standard remote sensing technique (Lillesand and Keifer, 1994) for reducing the dimensionality of the data by creating a single image that contains most of the variance in the original multi-band dataset. Unfortunately, it can also create an image with excessive noise, which in effect confuses the edge detection process. While strong edges are emphasised, weak ones are too, resulting in too many

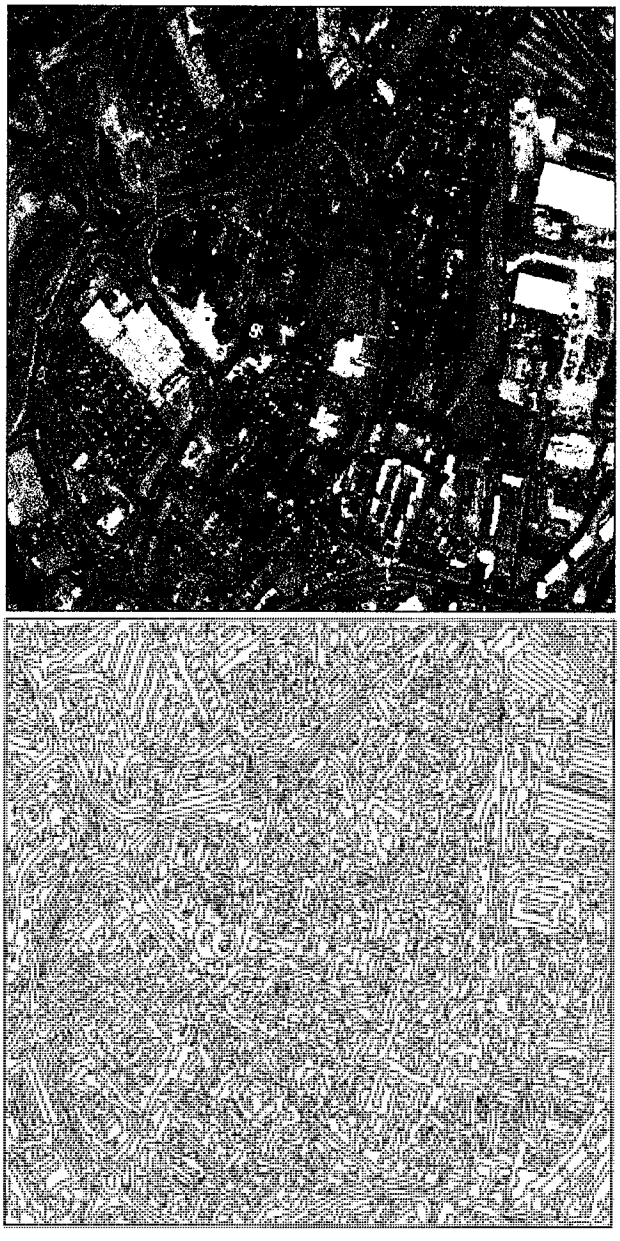


Figure 5.7. The first principal component and its zero-crossings (Guassian=5).

irrelevant edges to make the image useful. An example of this is shown in Figure 5.7.

The image contains many small polygons which do not reflect the structure of the imagery. An examination of the road network for instance, shows that instead of being one large polygon continuous throughout the imagery, there are many small regions. This means that the initial 'crude' feature extraction contains little contextual information about the road network.

The higher order principal components suffer from the converse problem, namely that they contain very little useful information for the edge detection process to perform adequately, producing worse results than the coarse edge extraction seen in Figure 5.4.

Other combinations of bands were tried, with a focus on making use of the imagery in which road, roof and permeable area are best discriminated. These are the required outputs from the classifier for urban drainage models.

The combinations used included individual bands such as Band 11 (thermal – see Figure 5.3) and Band 7 (near infrared), as well as all the other bands. Bands 11 and 7 performed better than the other bands, so the imagery was combined in order to take advantage of this. However, this did not yield any improvement in the extraction of structural features in the imagery, since the data once more contained too much variance and hence produced too many edges (similar to the first principal component).

Other standard methods for combining information from multiple bands include the true colour imagery (red (Band 5), green (Band 3) and blue (Band 2)) and the Normalised Difference Vegetation Index (NDVI), which allows us to combine the red (Band 5) and near-infrared bands (Band 7) (Lillesand and Kiefer, 1994). The true colour imagery was, as expected, poor. However, we might have expected some useful results from the NDVI imagery, since the purpose of this band combination is to highlight vegetation within the imagery. However, the results were once more unsatisfactory, with insufficient edges being extracted. Section 5.2.4 will examine a further method of combining bands.

5.2.4 False Colour Composite Image

False colour near infrared imagery is widely used in a variety of applications which include classification of urban areas (Curran, 1985), monitoring of soil moisture, mapping of soil and environmental disaster assessment.

The False Colour Composite (FCC) imagery is a combination of Bands 3, 5 and 7 and is favoured by photo-interpreters because it provides good contrast between vegetation and man made structures. The green band (Band 3) is assigned shades of the primary colour blue, the red band (Band 5) is assigned shades of the primary colour green and the near infrared band (Band 7) is assigned shades of the primary colour red (Figure 5.8). To avoid confusion with the ATM bands, the following section will refer to the red, green and blue colour assignments as channels rather than bands.

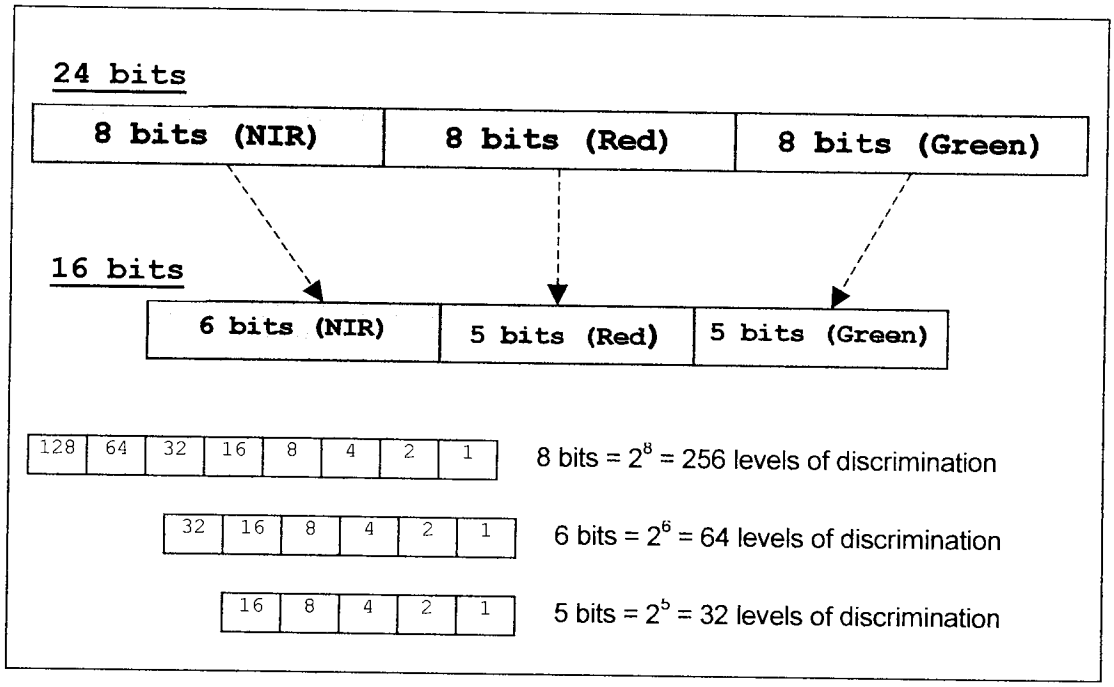


Figure 5.9 The resampling of the false colour composite imagery from 24 bits to 16 bits

The false colour composite image used is created by combining the three channels, each of which are 8 bit imagery. The image created can then be reduced from 24 bits to 16 bit imagery by re-sampling, by taking 6 bits for the near infra-red, 5 bits for the red and 5 bits for the green. This makes the image considerably more easy to handle, since the size of the image is reduced, along with its complexity, which means any subsequent processing takes much less time to complete. An additional benefit of this is that much of the low variance content of the scene is removed – Figure 5.9 illustrates this. Essentially the discriminatory resolution of the bands is reduced, leaving only the information that we are concerned with.

As a consequence, the most significant part of the image is the near infrared component, which is also the most discriminating band for roads and houses, followed by the red and then the green. Only if the near infra-red and the red have identical values is the green used for discrimination. Consequently, by using the type of imagery favoured by photo-interpreters we expect to produce an edge map rich in the features that they themselves look for, such as a continuous road network.

The imagery is most applicable to studies in which the determination of the extent of vegetation within the image is an important factor, such as with urban environments. In itself it is not a feature detector, but an artificial way of enhancing the features in an image so as to make them easier to locate. Further more, it makes some use of the multi-spectral nature of the dataset, incorporating three bands. We would therefore expect the edge detector to extract the structural features favoured by photo-interpreters. Our results bear this out, with an edge image that emphasises the structural features in the image with which we are concerned (Figure 5.10). Of all the combination methods attempted, this proved to be the most suitable for this purpose.

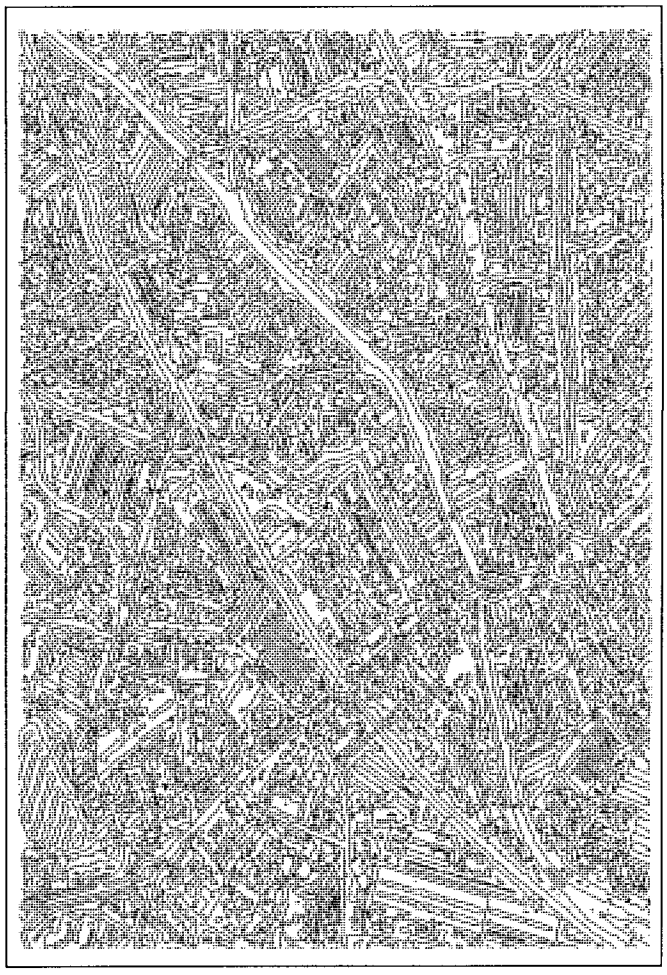


Figure 5.10. Zero-Crossings of the False Colour Composite Image (Gaussian = 5)

5.2.5 Conclusion

We have shown that the Marr-Hildreth edge detector, when applied to remotely sensed imagery, can produce an isocline map of the complex features in the imagery, particularly when the parameters are selected for the determination of fine scale elements. We have

also managed to optimise the detection of these features by utilising the same type of imagery used by a photo-interpreter, and consequently extracted similar features to those used in manual classification.

5.3 Polygon extraction

5.3.1 Introduction

One of the many advantages of using a feature detector as part of our classifier is that we are provided with a 'continuous' contour map – in other words, every contour within in the image is closed. This leaves us with an image containing large polygonal regions, the area within being largely homogeneous. It is possible to classify this map, assigning classes to each polygonal region.

5.3.2 Simple extraction

It is a relatively straightforward procedure to convert the isocline map produced by the feature detector into a classified land use map. We can make use of the Maximum Likelihood classification method to classify each region individually. In order to do this we classify the original multi-spectral imagery and then calculate the most common class for each polygon, assigning it to that class. If we assume that each region is

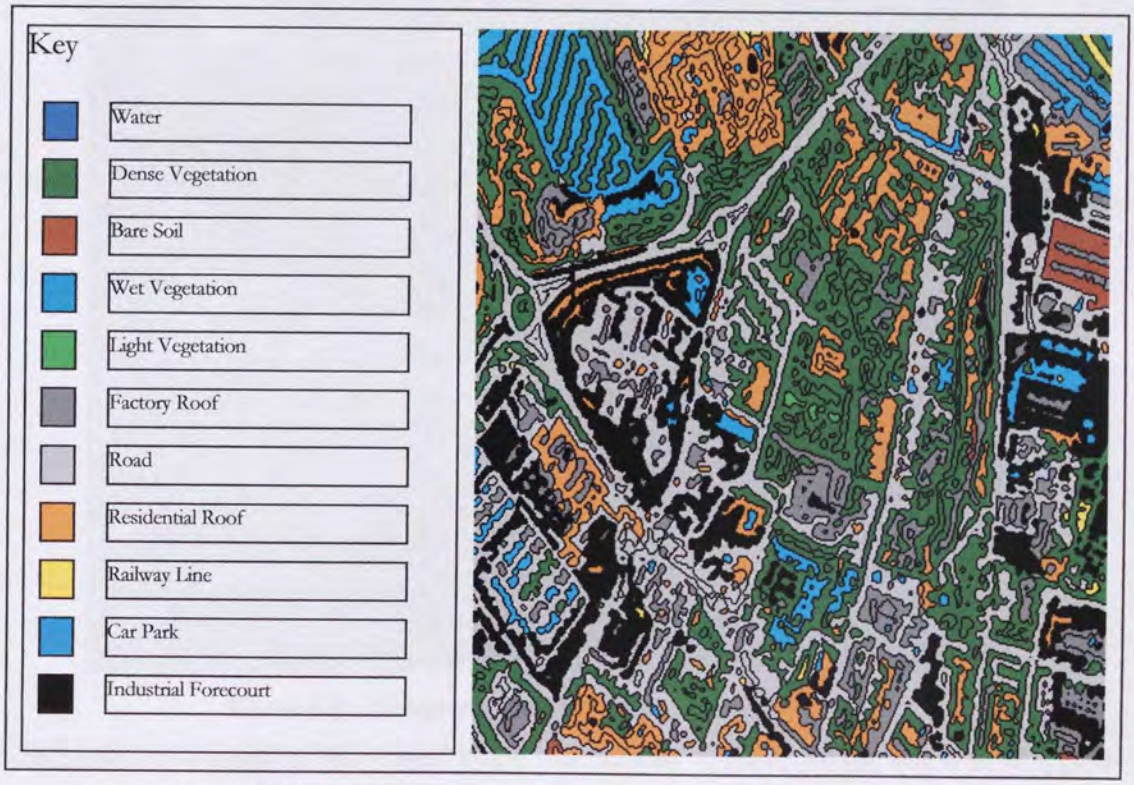


Figure 5.11. Simple Polygon Classification, using 11 classes (see key).

homogeneous and then classify each region into its most likely class we end up with a map as shown in Figure 5.11. This procedure is described in more depth in section 5.4.2

Upon first inspection, the map appears to be of reasonable quality. The image has none of the speckling associated with per-pixel methods, and is clearly similar in appearance to the original imagery in as much as the structural features are maintained.

Unfortunately, there are several fundamental problems with this approach. Firstly, from a qualitative point of view, it is clear that the edges of the polygons in this image are not sufficiently adequate at representing what we see in the original image; that is, they are not located along the correct edges. The edges are located close to the true edges, but are crude and in need of refinement. This is a problem we will address in the next stage of the classifier's development.

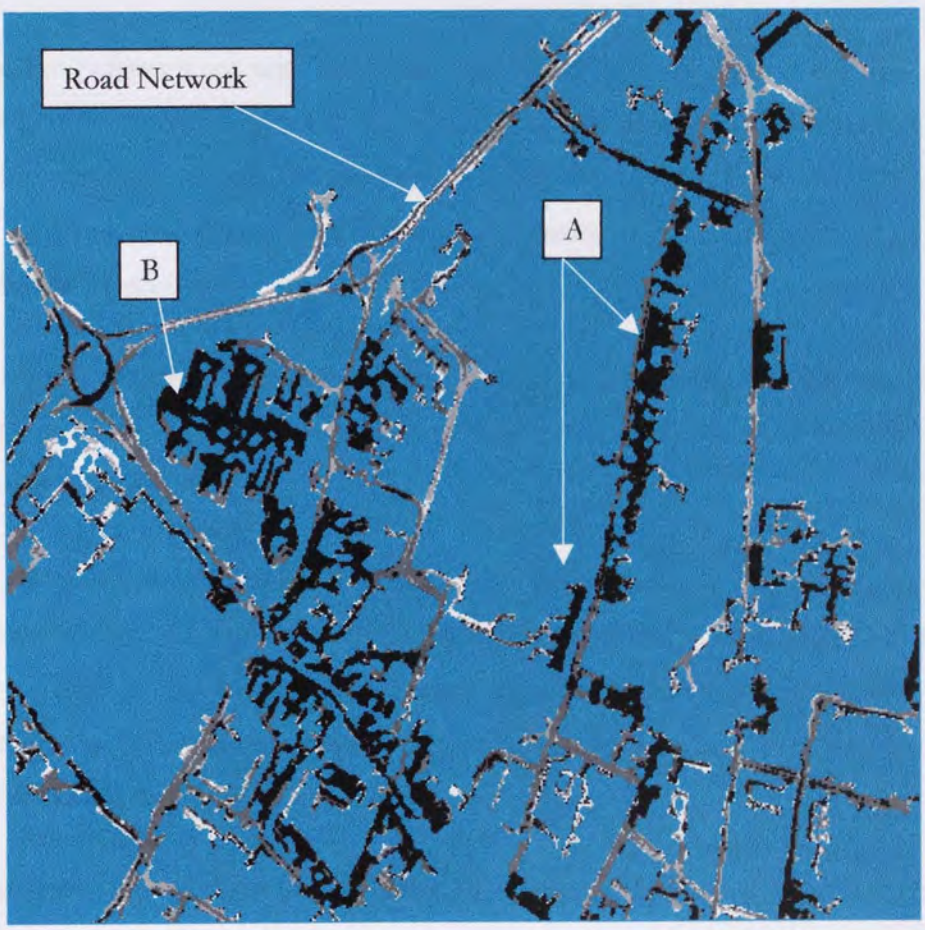


Figure 5.12. Composite Image of Road Network polygon.

Secondly, as a result of the 'edges' being inaccurately located, we are left with polygons that are not entirely homogeneous in nature. A closer examination of one of the polygons in the classified map indicates this (Figure 5.12). In this image, the largest

polygon has been used as a mask overlaid on with the composite image. This polygon has good continuity throughout the image and appears to represent the road network in the image. However, it can be seen that the polygon bounds not only road segments but segments of residential housing (A) and some areas of larger buildings (B). It becomes clear that, although we have located the features in the image with a reasonable degree of accuracy, we must find a way to improve this accuracy if we are to produce a classified map of any use.

5.3.3 Incorporation of a Seed growing algorithm

If we examine the crude polygons, we can see that they are not all entirely homogeneous in nature. However, we have an excellent platform from which to progress, since these crude polygons are representative of the structure of the imagery and, as small individual entities, of a much less complicated spectral makeup than the entire imagery which, when reconstructed, they represent.

Since the polygons are not entirely homogeneous, it is necessary to divide them into homogeneous regions. This will then provide several polygons that can be individually classified as single classes. When reconstructed, this should provide us with an image map in which the edges of the buildings and roads in the image are well defined and accurate in location.

This can be achieved if a region-growing algorithm is implemented. The algorithm used in this classifier is of a relatively simple nature; more sophisticated algorithms can be found in Jenson (1986).

The algorithm progresses sequentially across the polygon, joining neighbouring pixels together if their brightness values are within a pre-determined threshold. Clearly, too high a threshold will result in many small polygons, whereas too low a threshold will result in very few. It is important at this stage to strike a balance between the two. It is possible to use one or several bands for this purpose.

Initially, a variety of band combinations (the same as those used in section 5.2.3) were used, but this was found to give far too many polygons. It became clear that homogeneous regions may be contained in a single image alone, but do not tend to correlate well over multiple images. Consequently, single images were used for the region growing algorithm. Of these, Band 11 produced the most useful results, especially when

polygon has been used as a mask overlaid on with the composite image. This polygon has good continuity throughout the image and appears to represent the road network in the image. However, it can be seen that the polygon bounds not only road segments but segments of residential housing (A) and some areas of larger buildings (B). It becomes clear that, although we have located the features in the image with a reasonable degree of accuracy, we must find a way to improve this accuracy if we are to produce a classified map of any use.

5.3.3 Incorporation of a Seed growing algorithm

If we examine the crude polygons, we can see that they are not all entirely homogeneous in nature. However, we have an excellent platform from which to progress, since these crude polygons are representative of the structure of the imagery and, as small individual entities, of a much less complicated spectral makeup than the entire imagery which, when reconstructed, they represent.

Since the polygons are not entirely homogeneous, it is necessary to divide them into homogeneous regions. This will then provide several polygons that can be individually classified as single classes. When reconstructed, this should provide us with an image map in which the edges of the buildings and roads in the image are well defined and accurate in location.

This can be achieved if a region-growing algorithm is implemented. The algorithm used in this classifier is of a relatively simple nature; more sophisticated algorithms can be found in Jenson (1986).

The algorithm progresses sequentially across the polygon, joining neighbouring pixels together if their brightness values are within a pre-determined threshold. Clearly, too high a threshold will result in many small polygons, whereas too low a threshold will result in very few. It is important at this stage to strike a balance between the two. It is possible to use one or several bands for this purpose.

Initially, a variety of band combinations (the same as those used in section 5.2.3) were used, but this was found to give far too many polygons. It became clear that homogeneous regions may be contained in a single image alone, but do not tend to correlate well over multiple images. Consequently, single images were used for the region growing algorithm. Of these, Band 11 produced the most useful results, especially when

road segments were examined. The road network stretches over the whole image, and should prove homogeneous in nature. Band 11, the thermal band, is particularly useful for discriminating between roads, roofs and vegetation (as we have seen from the dawn imagery in Figure 3.2). As a consequence, Band 11 imagery was used.

The results of this addition to the classifiers are shown in Figure 5.13. In this Figure, the extracted 'coarse' polygon has been refined using the region growing algorithm, and classified into a series of smaller homogeneous polygons. The large road network has now been broken down into smaller homogeneous regions, not all of which are actually

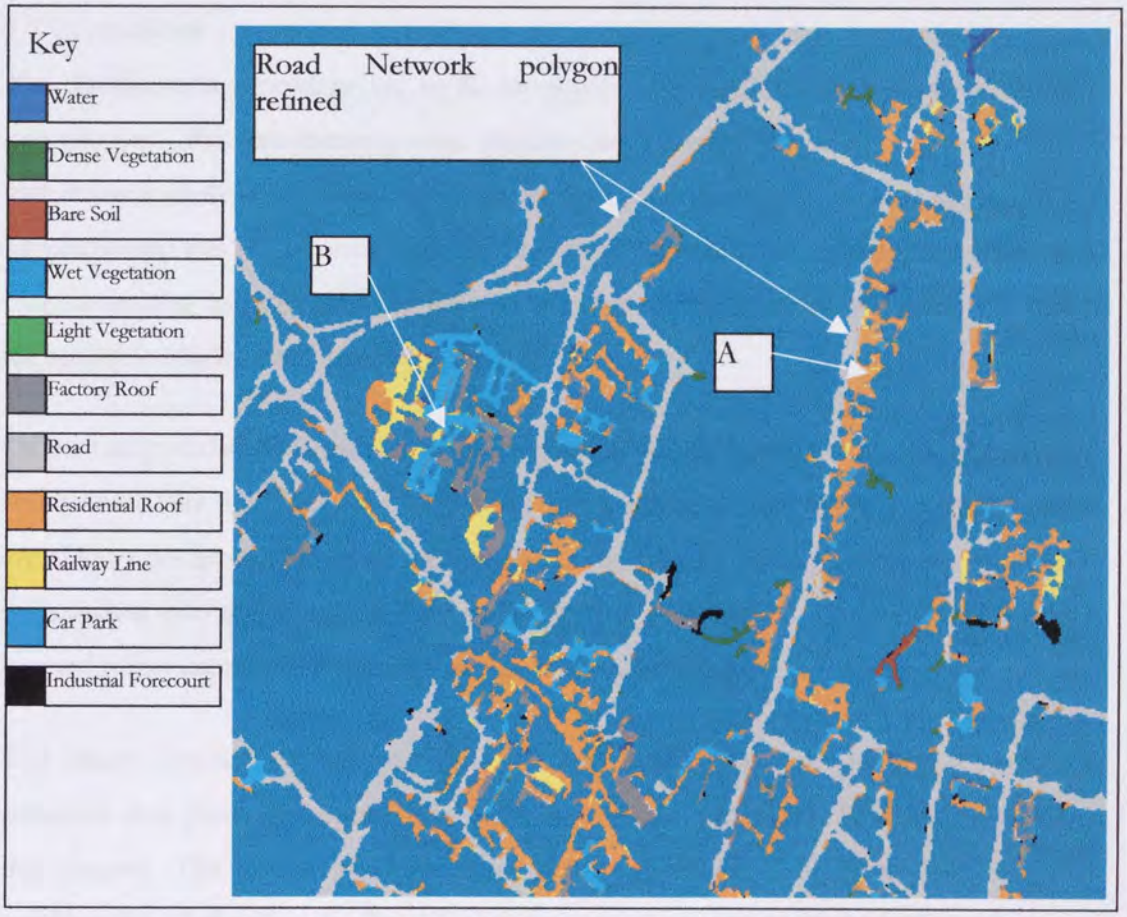


Figure 5.13. The Results of the Region Growing Algorithm

found to be road. The majority of the network remains, but in area A, for instance, a new polygon has been extracted and classified as residential roof. An examination of the composite imagery from Figure 5.8 will confirm that this is indeed the case. Furthermore, the large region situated at B has been broken up into several different regions. While this has been less successful, it can still be seen that areas of factory roof and car park have been extracted, as opposed of areas of roofing.

5.3.4 Conclusion

It has been demonstrated that the Marr-Hildreth edge detection routine can be used to produce a coarse polygon map of an entire image. This map can, after subsequent refinement of the edge detection procedure, through the introduction of a region growing algorithm, be used as a basis for the extraction of homogeneous regions. The homogeneous nature of these polygons makes it possible to assign each to a single class.

The problem of correctly classifying each of these polygons must next be solved.

5.4 Classification of extracted polygons

5.4.1 Introduction

The classification procedure has so far incorporated spatial information provided from two sources - the zero-crossing edge detector has provided an initial map containing a vast amount of structural detail about the imagery, and the region growing algorithm has supplemented this information with neighbourhood information, refining the map into a map containing homogeneous polygons. In order to classify these polygons the multi-spectral information must next be utilised.

Of the many statistical classification schemes available, the earlier results from Chapters 3 and 4 show that the Maximum Likelihood will produce a map that is as accurate as we are able to get from a per-pixel technique; that is, without incorporating any contextual information into the classifier. This makes it ideal as a way of representing the multi-spectral information in the imagery on a purely per-pixel basis.

The other classification schemes (ECHO and SMAP) are unsuitable for providing statistical data precisely because they make use of some of the techniques described in this chapter. The SMAP classifier has a scaling approach that, in the manner of a neighbourhood classifier, produces large homogeneous regions, although these are suited to imagery with many rectangular regions (such as agricultural satellite imagery). The ECHO classifier is effectively a region growing method, and once more makes use of neighbouring pixels to produce homogeneous regions. As a consequence, the statistics produced by these methods are not of a purely per-pixel nature. Further to this, however, they are not classifiers that perform particularly well on urban imagery, as can be seen from the results in Chapter 3.

5.4.2 Maximum Likelihood

The statistics produced by the Maximum Likelihood algorithm have already provided the spectral characteristics of the multi-spectral image set. We are now in a position to make use of this per-pixel information on a per-polygon basis; that is, we can classify each polygon as a unique class. Since the polygons are now homogeneous in nature, the Maximum Likelihood scheme should be able to classify each pixel in each polygon as belonging to the same class. However, this will not in fact be the case, since the Maximum Likelihood classifier contains no contextual information and is prone to producing 'speckly' classification maps. A resolution to this problem is discussed in the following section.

The Maximum Likelihood classifier can be used in a supervised or unsupervised fashion. This provides two sources of spectral statistics. The classifier developed in this chapter can make use of either. The results of both classification schemes are presented in section 5.5, and the relative merits of both are discussed in section 5.6.

5.4.3 Determining the modal value of the polygon

In order to overcome the problems of the Maximum Likelihood classifier, it is possible to examine all the pixels in each polygon and to assign the most frequently occurring class to that polygon. The homogeneity of the polygons ensures that one class will be predominant in each pixel. We can therefore assign the overall class of the polygon to that single predominant class.

5.5 Implementation of the Classifier

The classifier was implemented in the C language, and the code is included in Appendix 1. The code makes use of the Grass libraries (Shapiro *et al*, 1993) for displaying the image and selecting the geographical region of the raw imagery. As a result, it is capable of classifying imagery that has been geo-rectified as well as that which has not, within the Grass framework. In addition, several Grass modules are used to create the required inputs into the classifier. These inputs are a map delineating the polygonal regions in the imagery, a map containing the Maximum Likelihood statistics of the imagery (either supervised or unsupervised), and a raw image (Band11) for the discrimination of homogeneous regions in the image.

The procedure is outlined below and in Figure 5.14.

1. A False colour composite image is created using the Grass Module *i.composite*. The zero-crossings of this image are then calculated through the use of the Grass module *i.zc*.

2. A Maximum Likelihood classification is performed using the Grass Module *i.maxlik*. This can be performed in either supervised or unsupervised mode. If it is to be performed in supervised mode, a training map must also be created prior to the determination of the statistics.

3. The zero-crossing image is then converted into a polygon map, using a script routine which numbers the polygons in ascending order to make subsequent extraction and analysis of these regions simpler. Prior to this, the edge of the image is cropped so that the intensity contours at the edge of the image are not distorted by the zero-crossing edge detector.

4. Along with a thermal image, these two maps are used as the input for the classifier. The classifier extracts each polygon in the image in a sequential order. A test of homogeneity is then applied to the image through the use of a region growing algorithm applied to the thermal image. If the polygon is found not to be homogeneous in nature, it is broken up into smaller polygons that are homogeneous.

5. The homogeneous polygons are then assigned to the modal class (from the Maximum Likelihood map) determined from the pixels contained within the polygon.

6. The boundary image (the zero-crossing edge map) is then classified in the same manner, effectively being treated as one large coarse 'polygon' and then refined through the test of homogeneity. The result is then combined with the polygon image. This results in the entire image being classified. This makes the assumption that edge points can be classified as a single class. As we have seen from the discussion of Spectral Mixture modelling (section 2.3.8) there are different method for classifying (and assessing) these boundary points. This will be expanded on in Chapter 6.

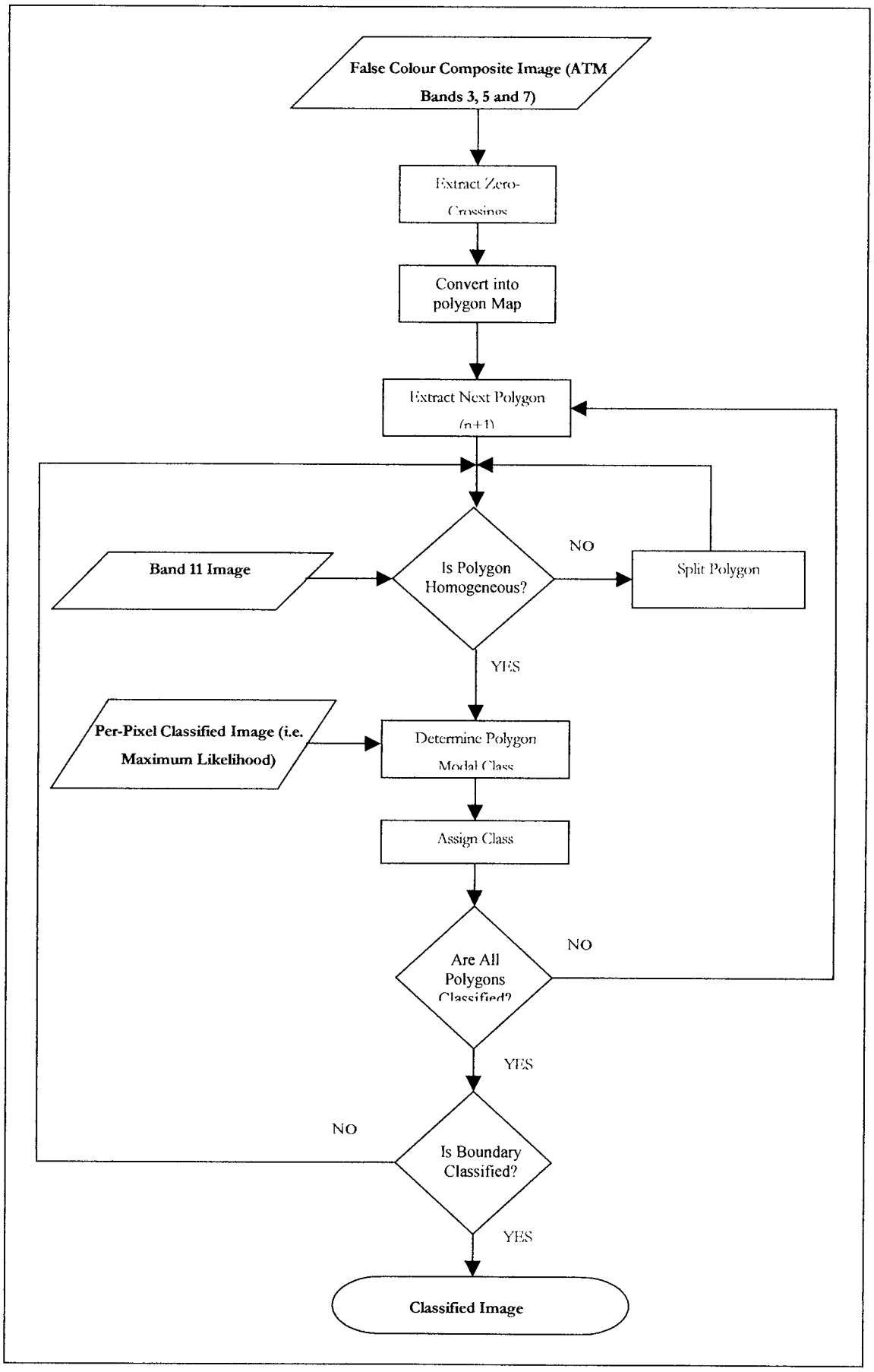


Figure 5.14 The flowchart of the Polygon Classification Process

Consequently a classified map is produced. The production of this map currently takes approximately three days from start to finish, running on a SUN Sparc 3. This is not particularly fast, although it should be noted that the Maximum Likelihood classification has only become a computationally viable procedure in recent years due to the improvement in computer processing. However, the code for the polygon classifier was originally written using a Grass script file, from which it was converted into C. As a result, the code was written for ease of implementation rather than efficiency. Optimisation of the code should result in a considerable increase in performance. The reasons for this are detailed in Appendix 1.

5.6 Results

Four images are presented in this section. The supervised and unsupervised images are presented, initially with 11 classes (Figures 5.15/5.16), and finally 4 classes (Figures 5.17/5.18). The image key indicates what each colour in the look-up table represents.

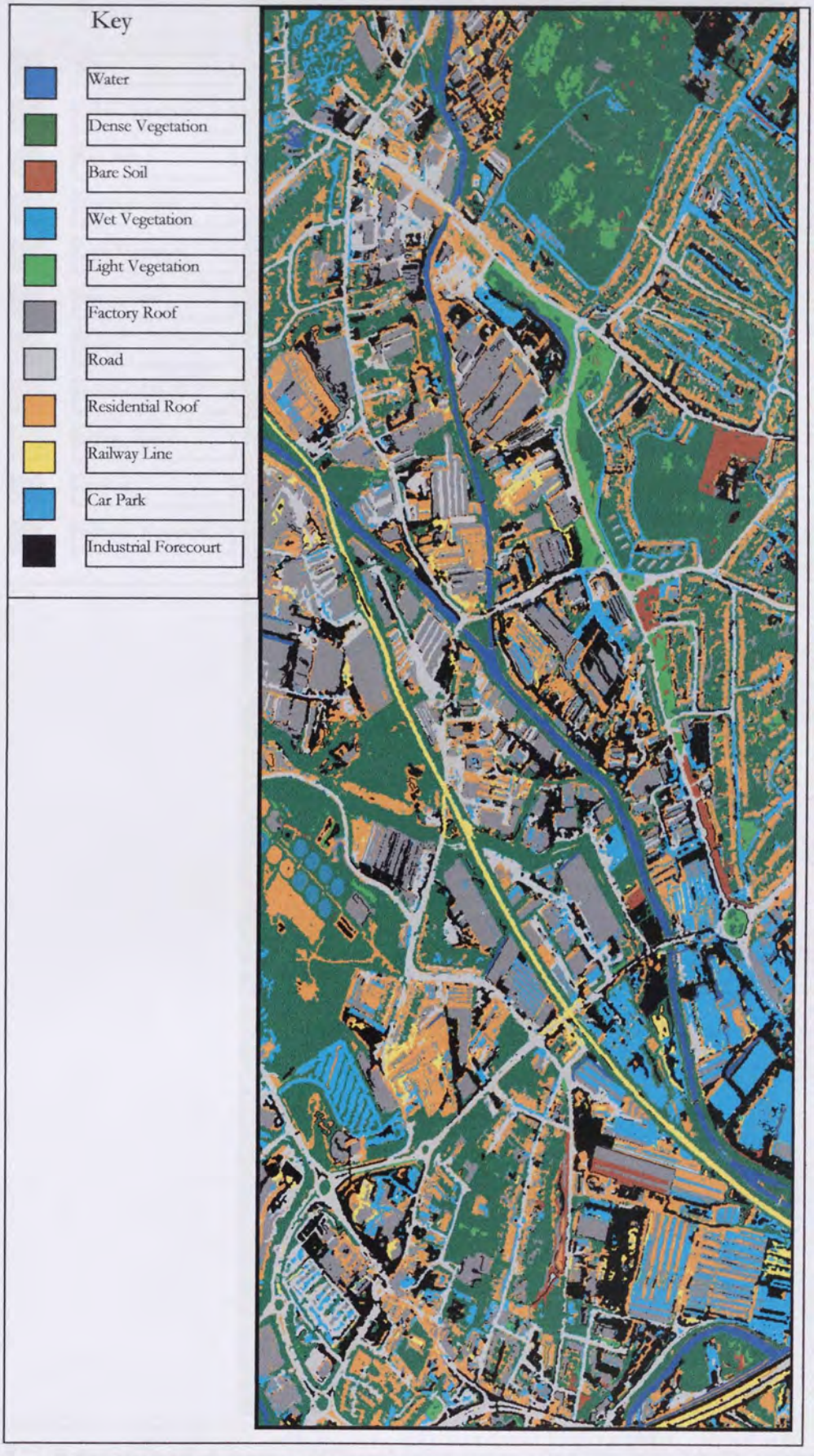


Figure 5.15. Polygon Classified Image (11 Classes) using supervised Maximum Likelihood statistics.

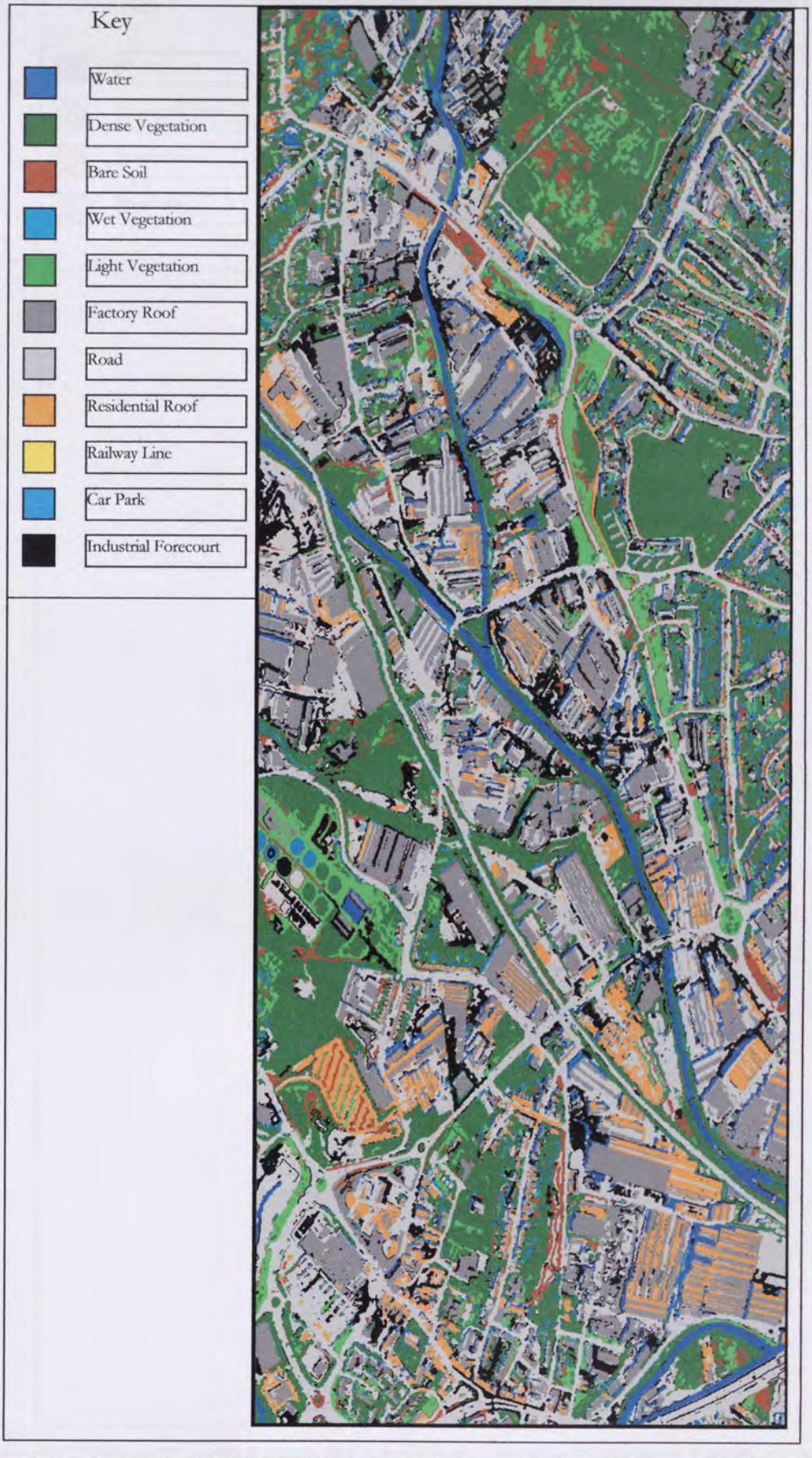


Figure 5.16. Polygon Classified Image (10 Classes) using unsupervised Maximum Likelihood Statistics

Figure 5.17. Polygon Classified Image (4 Classes) Using Supervised Maximum Likelihood Statistics

Figure 5.18 Polygon Classified Image (4 Classes) using unsupervised Maximum Likelihood Statistics.

5.7. Analysis of results

5.7.1 Introduction

In Chapter 4 an accuracy assessment scheme was developed that produced an objective measure of accuracy. This allows a comparison of various classification methods to be made. This section presents an analysis of the accuracy of the classifications and discusses the accuracy figure obtained.

5.7.2 Accuracy Assessment

The tables and charts detailing the accuracy of the Polygon classification of 11 classes (in both supervised and unsupervised more) can be found in Appendix 2.

Using the region correlation method, the supervised classification performs particularly well, in fact producing the highest accuracy of all the classifiers. There is high accuracy attained for most of the classes, except the bare soil class and the car park class. These are, as with many of the classes analysed, confused with spectrally similar classes. The unsupervised classification suffers mostly from bad confusion between road and factory roof while also confusing large amounts of residential roof with factory roof. Consequently this classification performs less well. For the 4 class imagery, the region correlation shows the supervised polygon classifier has an accuracy of 65.23% (Kappa, Table 5.1) and the unsupervised polygon classifier has an accuracy of 70.99% (Kappa, Table 5.2).

The reduction of the number of classes to four, as with the other classifiers, improves the accuracy, showing (for the single random pixel method) the supervised polygon classifier having a Kappa value of 63% and the unsupervised having a Kappa of 62% (Table 5.3). The region correlation method indicates that the classification is generally of a high quality, in fact indicating that the unsupervised Polygon classifier (Table 5.3) is the most accurate. An examination of the confusion matrices reveals there is less confusion between roof and road for the unsupervised method than with any of the other methods (Table 5.3). The results from both sampling methods are shown visually in Figures 5.20 and 5.21.

A comparison of the accuracies of the 6 classifications as determined by the region correlation method indicates, however, that the difference between the Kappa values and

confidence levels at 95% is not consistent (Figure 5.22). The different pattern in each, especially considering the sample size, indicates the presence of a bias in the sample. This reinforces the view that the better assessment method is the single random pixel one.

The single random pixel method shows the same difference between the two measurement methods, indicating that the sample is free from bias, and that the sample is well balanced. As an indication, if the road and roof pixels calculated from the manual classification are added together, they come to 15864 (7640 + 8224). The same sum for the supervised classifier comes to 15693 (11355+4338). The confusion indicated by the matrix suggests, however, that the classification error is a result of mis-classification between the road and roof classes.

5.7.3 Qualitative Assessment

The images produced by the Polygon classifier exhibit a considerable improvement in appearance over any of the other classifiers examined in this project. There is a clear degree of homogeneity present in the imagery, resulting in an image with few speckled areas. The detailed images shown in Figure 5.19 confirm this.

An close examination of the entire imagery (Figures 5.15 - 5.18) shows that the detection of certain features improves dramatically when compared to the per-pixel classifiers. The railway line in the supervised image is entirely assigned to one class. In the unsupervised image, the motorway section in the lower part of the image is also assigned to a single class, whereas the other classifiers assigned it to several different classes (Figures 3.11 – 3.18). There are other examples of this throughout the imagery which, from a qualitative viewpoint, indicate a large improvement in the classification procedure.

It is important that single objects, such as a section of road, are assigned entirely to one class. The number of pixels that are reassigned through the addition of context may not necessarily be large. As a result, this may not be reflected greatly in the accuracy assessment method, particularly one using a small sample size. While it is clear that the region correlation method is biased, it does reflect the improvement in accuracy achieved through the incorporation of context.

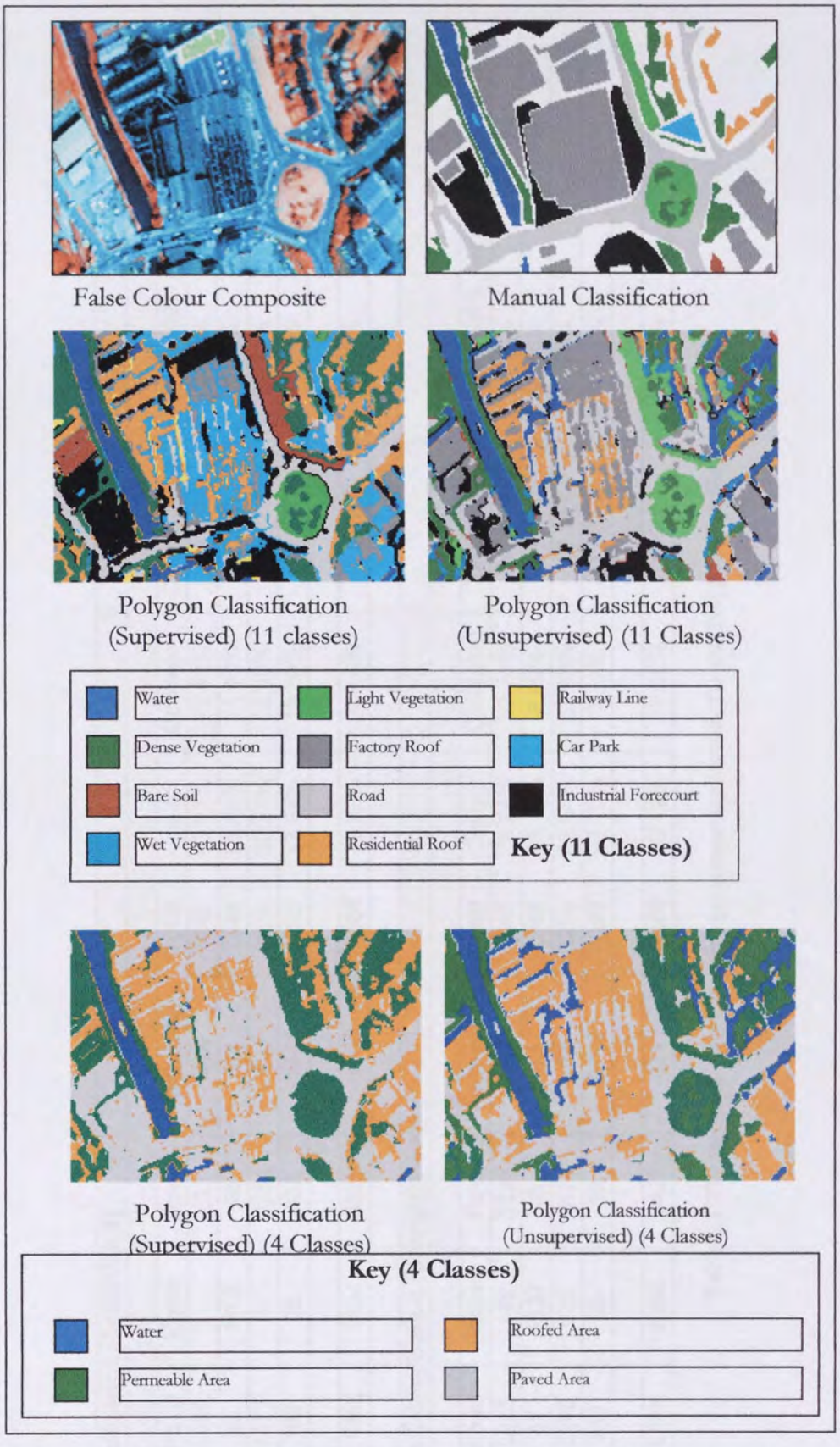


Figure 5.19. Region Correlation Accuracy Assessment for the Polygon Classifier.

Polygon Classification (Supervised)										
	Road	**Water**	**Roof**	**Permeable**	**Total**	Correct	**Omission**	**Commission**	**Kappa x100%**	**Accuracy at 95%CL**
Road	6690	1	521	428	7640	87.57%	12.43%	61.06%	**70.98%**	**86.93%**
Water	23	1105	96	24	1248	88.54%	11.46%	2.32%	**87.85%**	**86.96%**
Roof	4477	2	3254	491	8224	39.57%	60.43%	13.18%	**22.68%**	**38.66%**
Permeable	165	26	467	3440	4098	83.94%	16.06%	23.01%	**79.40%**	**82.97%**
Total	11355	1134	4338	4383	19866	72.93%	25.10%	24.89%	**65.23%**	**72.41%**
Polygon Classification (Unsupervised)										
	Road	**Water**	**Roof**	**Permeable**	**Total**	Correct	**Omission**	**Commission**	**Kappa x100%**	**Accuracy at 95%CL**
Road	5825	146	1226	443	7640	76.24%	23.76%	34.54%	**58.61%**	**75.42%**
Water	47	1137	6	58	1248	91.11%	8.89%	70.27%	**90.10%**	**89.69%**
Roof	2380	636	5176	32	8224	62.94%	37.06%	15.14%	**45.24%**	**62.04%**
Permeable	212	95	13	3778	4098	92.19%	7.81%	13.01%	**90.03%**	**91.48%**
Total	8464	2014	6421	4311	19866	80.12%	19.38%	33.24%	**70.99%**	**79.64%**

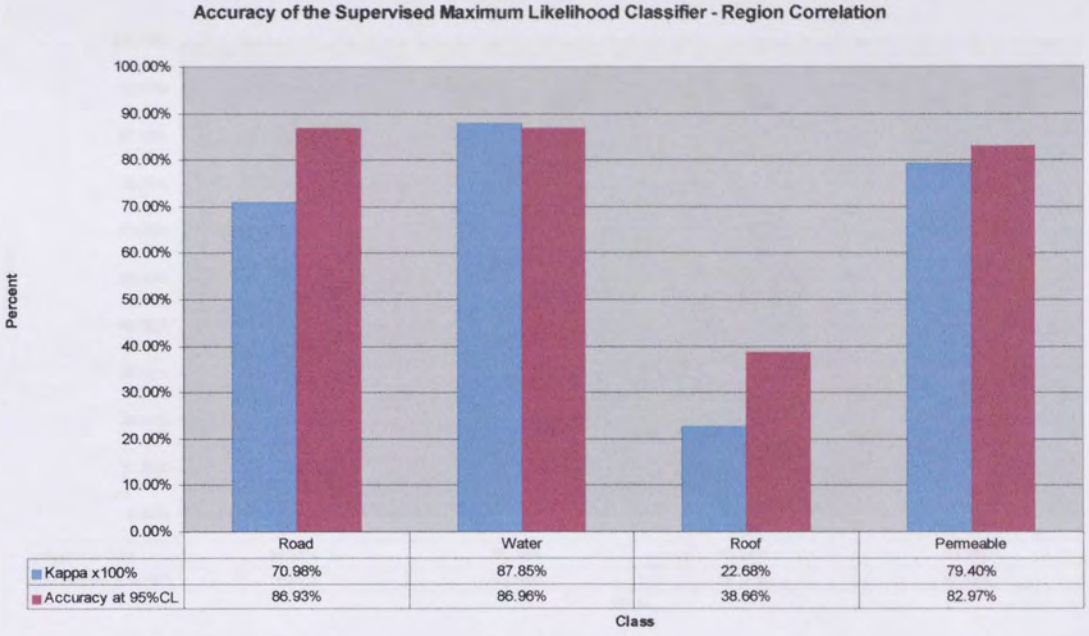
Table 5.1 The region correlation accuracy assessment results for the Polygon Classifier.

Polygon Classification (Supervised)										
500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	114	1	37	28	180	63.33%	36.67%	30.00%	**44.78%**	**56.45%**
Water	0	11	2	0	13	84.62%	15.38%	15.38%	**84.20%**	**58.18%**
Roof	49	0	102	22	173	58.96%	41.04%	25.43%	**42.03%**	**51.77%**
Permeable	5	1	5	123	134	91.79%	8.21%	37.31%	**87.45%**	**87.15%**
Total	168	13	146	173	500	74.67%	25.33%	27.03%	**64.62%**	**71.15%**
1000 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	195	3	69	32	299	65.22%	34.78%	46.49%	**47.77%**	**60.10%**
Water	0	14	0	1	15	93.33%	6.67%	26.67%	**93.21%**	**76.76%**
Roof	115	1	215	23	354	60.73%	39.27%	30.79%	**41.91%**	**55.96%**
Permeable	24	0	40	268	332	80.72%	19.28%	16.87%	**71.48%**	**76.73%**
Total	334	18	324	324	1000	75.00%	25.00%	30.20%	**63.60%**	**72.59%**
1500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	309	4	106	60	479	64.51%	35.49%	40.29%	**46.66%**	**60.54%**
Water	0	25	2	1	28	89.29%	10.71%	21.43%	**89.06%**	**75.69%**
Roof	164	1	317	45	527	60.15%	39.85%	29.03%	**41.97%**	**56.30%**
Permeable	29	1	45	391	466	83.91%	16.09%	22.75%	**75.93%**	**80.81%**
Total	502	31	470	497	1500	74.46%	25.54%	28.38%	**63.40%**	**72.50%**

Table 5.2 500, 1000 and 1500 random pixel confusion matrix for the Polygon Classification (Supervised).

Polygon Classification (Unsupervised)										
500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	129	8	18	27	182	70.88%	29.12%	47.25%	48.91%	64.43%
Water	1	13	0	2	16	81.25%	18.75%	118.75%	79.97%	56.43%
Roof	72	10	68	19	169	40.24%	59.76%	11.83%	27.47%	32.98%
Permeable	13	1	2	117	133	87.97%	12.03%	36.09%	82.04%	82.44%
Total	215	32	88	165	500	70.08%	29.92%	54.48%	59.60%	66.38%
1000 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct			Kappa x100%	Accuracy at 95%CL
Road	202	15	55	27	299	67.56%	32.44%	48.83%	50.24%	62.53%
Water	0	13	0	2	15	86.67%	13.33%	273.33%	85.91%	64.08%
Roof	115	21	191	27	354	53.95%	46.05%	17.23%	38.44%	49.08%
Permeable	31	5	6	290	332	87.35%	12.65%	16.87%	80.66%	83.98%
Total	348	54	252	346	1000	73.88%	26.12%	89.07%	63.81%	71.43%
1500 Random Pixels										
Class	Road	Water	Roof	Permeable	Total	Correct	Omission	Commission	Kappa x100%	Accuracy at 95%CL
Road	331	23	73	54	481	68.81%	31.19%	48.23%	50.08%	64.99%
Water	1	26	0	4	31	83.87%	16.13%	193.55%	82.89%	68.73%
Roof	187	31	259	46	523	49.52%	50.48%	15.49%	34.73%	45.57%
Permeable	44	6	8	407	465	87.53%	12.47%	22.37%	81.08%	84.74%
Total	563	86	340	511	1500	72.43%	27.57%	69.91%	62.19%	70.42%

Table 5.3 500, 1000 and 1500 random pixel confusion matrix for the Polygon Classification (Unsupervised)



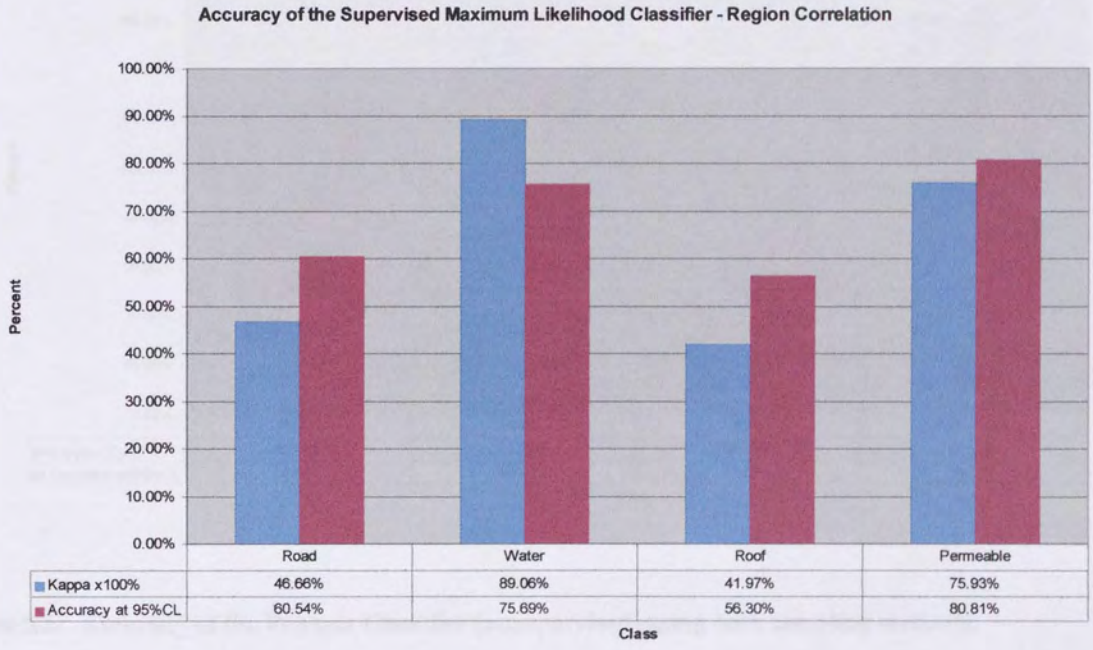
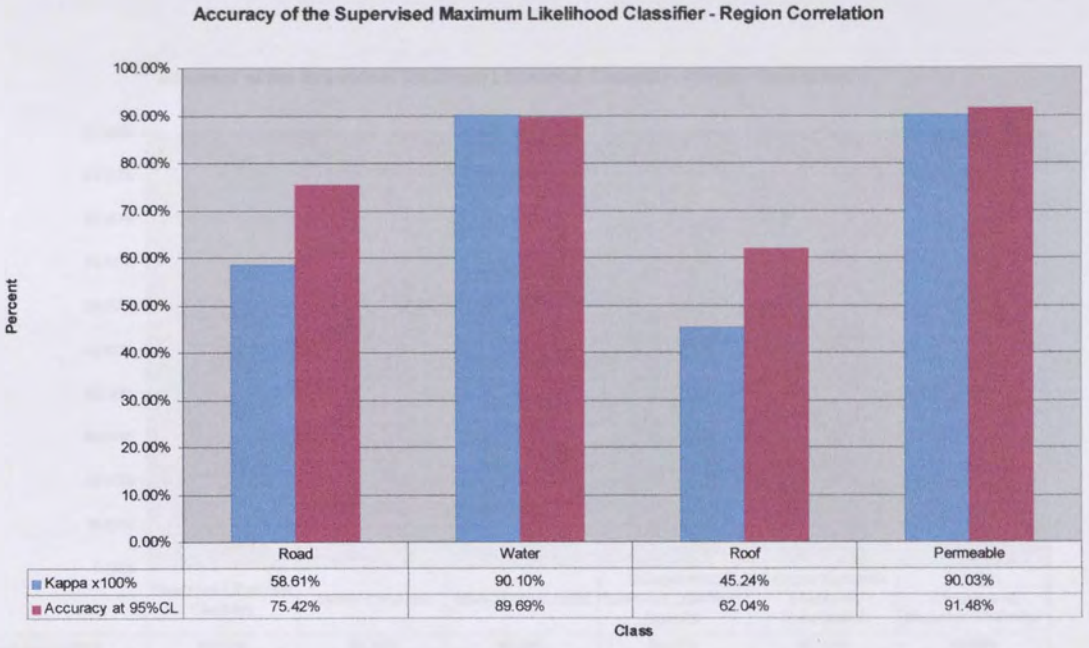


Figure 5.20 Accuracy of the Polygon Classifier (supervised) assessed with both sampling methods.



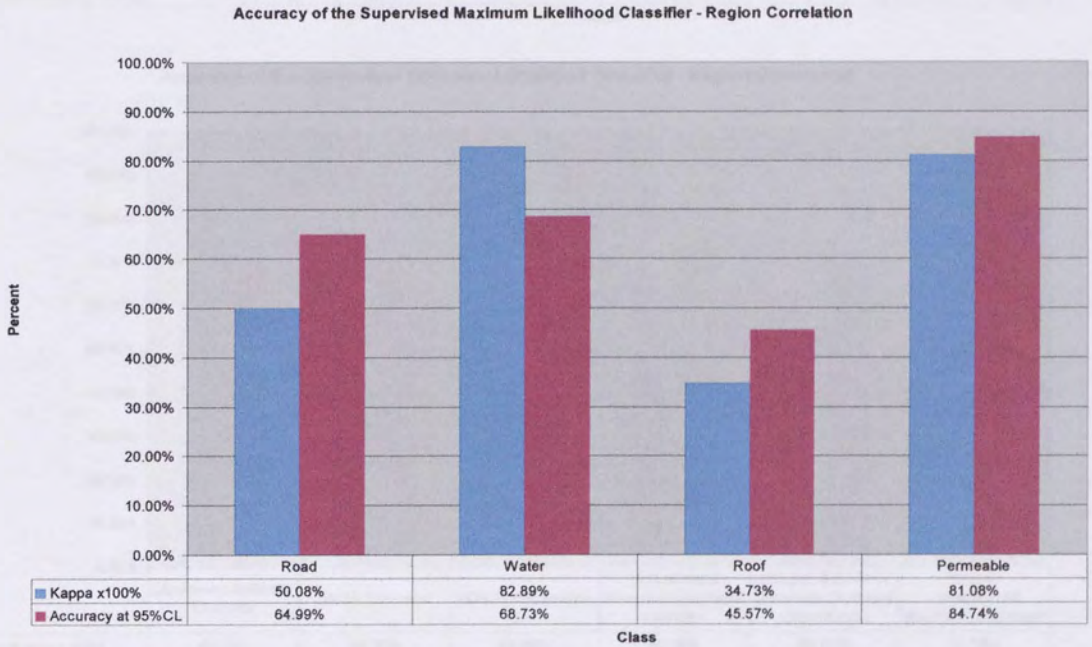
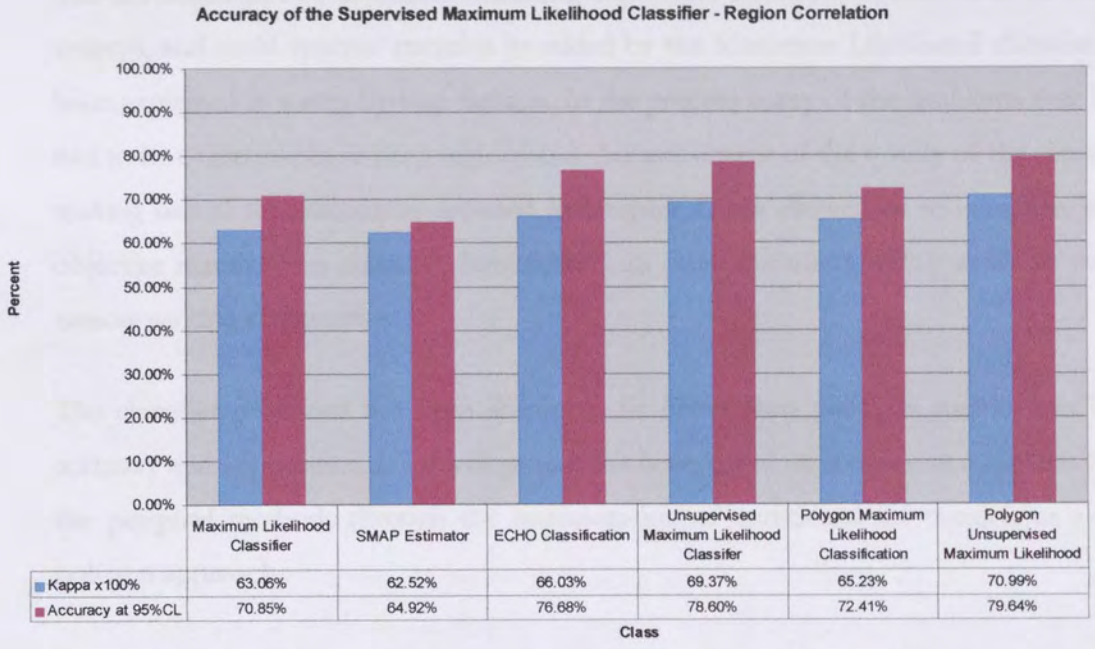


Figure 5.21. Accuracy of the Polygon Classifier (unsupervised) using both sampling methods.



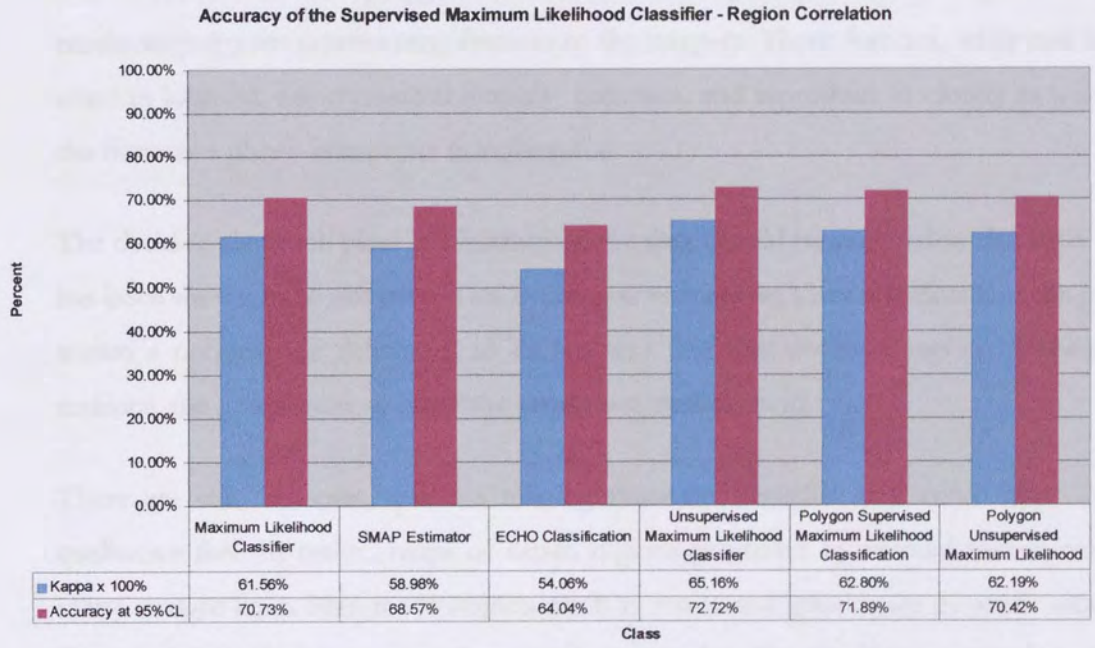


Figure 5.22 Comparison of the Accuracies for all six classifiers presented in this project.

5.8 Conclusion

The development of a classifier containing contextual information, derived from several sources, and multi-spectral statistics provided by the Maximum Likelihood classifier has been explained in a step-by-step fashion. In the process many of the problems that have had to be overcome have been highlighted. An assessment of the quality of the classifier, making use of the techniques assessed in Chapter 4, has allowed us to compare, in an objective manner, our classifier favourably with other classifiers widely available to the remote sensing community.

The classifier produced has been shown to be better than per-pixel methods in both accuracy and appearance. It's development has been based on a different approach from the per-pixel methods through the incorporation of contextual information in a per-polygon approach.

The extraction of these polygons is based on visual perception methods, and as such produces polygons representing features of the imagery. These features, while not being exact in location, are connected intensity contours, and reproduce as closely as possible the features a photo-interpreter is looking for.

The decision that each pixel in a homogeneous area should be assigned to the same class has been shown to be effective. This decision is reasonable, since it is clear that the pixels within a polygon are delineated in such a way that if it covers a region of road, for instance, the pixels must all be of the same class, namely road.

There are still, however, qualities missing from the classifier that could improve its qualitative feel. In reality, maps of urban regions produced by manual means contain many straight lines. Man made objects, such as roads and houses, are generally straight. Large buildings, for instance, are in general, rectangular. The classifier presented does not account for this, and as such does not come close to that provided by a photo-interpreter. An attempt to incorporate this is discussed in the next chapter.

Chapter 6: Further improvements

6.1 Introduction

The classifier produced has demonstrated some techniques that can be used to improve both the accuracy and the quality of the final classification map. However, a map classified manually (such as an Ordnance Survey Map) is of distinctly different appearance. Man made objects such as roads and houses (in urban developments) generally have straight edges. If the classified imagery is examined, it can be seen this is not the case.

This is partially due to the absence of geometric rectification of the imagery. The buildings and roads in the imagery do not have straight sides due to the instability of the aeroplane during the acquisition of the ATM data. There are, however, other causes. Roads, for instance, are occluded at various points of the imagery by overhanging trees. The classifier currently has no method for coping with this. Shadows from large buildings also cause some confusion to the classifier, once again distorting the man-made features in the imagery. This chapter will examine some methods for improving the appearance of the imagery.

6.2 Straight Line Extraction

The basic theory of the Hough transform and the Muff transform are examined in Chapter 2 - see section 2.8.

6.2.1 The Hough Transform

Hough transforms (Ballard and Brown, 1982; Levine, 1985) are particularly useful for extracting straight lines in images. They are used in a variety of computer vision applications to parameterise several geometric shapes such as straight lines, circles, ellipses, and rectangles. They are generally used in conjunction with an edge detector, such as the Canny edge detector (Canny, 1986), to provide a description of the features in the imagery. The lack of continuity in the edge detector does not affect the results of the Hough Transform.

The Hough transform is a robust parameter estimator of multi-dimensional features in images. It provides robustness against discontinuous or missing features. This would seem ideal when attempting to deal with partially occluded features, such as roads with overhanging trees.

A noteworthy characteristic of the Hough Transform is that it does not require the connection of the co-linear points. Segmented lines will generate a peak in the parameter space and the lacking segments simply do not contribute to the transform. On the other side, artefact peaks might be generated in the presence of noise and high density of features by coincidental intersections in the parameter space.

It would appear that the incorporation of this line detector into the classifier would improve the appearance of the imagery. If it were applied prior to the polygon extraction, the coarse polygon map produced by the edge detection procedure should resemble a manually produced map much more closely, with man made objects given straight edges. Subsequent extraction of homogeneous areas, in theory, would produce a map of high accuracy and of a highly pleasing appearance. In essence, such an approach is very similar to that of manual cartography. The cartographer locates objects, identifies them and then assigns straight lines to the borders as appropriate.

6.2.2. Results of the Hough transform

The first approach was to take an area of the imagery and apply a Hough Transform to it. In this case a region of Dudley was selected that was high in man made urban content, such as canals, roads, houses and factories. All of these features contain distinct straight edges. The results of the application of the transform on this region are presented in Figure 6.1. This figure indicates, however, that the high density of features has resulted in a great many false lines being detected. The tuning of several of the parameters in the Hough Transform (such as minimum line length / maximum distance between lines) does little to improve this. Inspection of the imagery does indicate that many correct lines have also been extracted too (in particular the large factories). The lines around them are also skewed due to the effect of noise in the surrounding areas.

This approach clearly has a great many drawbacks. The image produced is unacceptable since there are a vast number of incorrectly identified lines, rendering this technique as unusable.

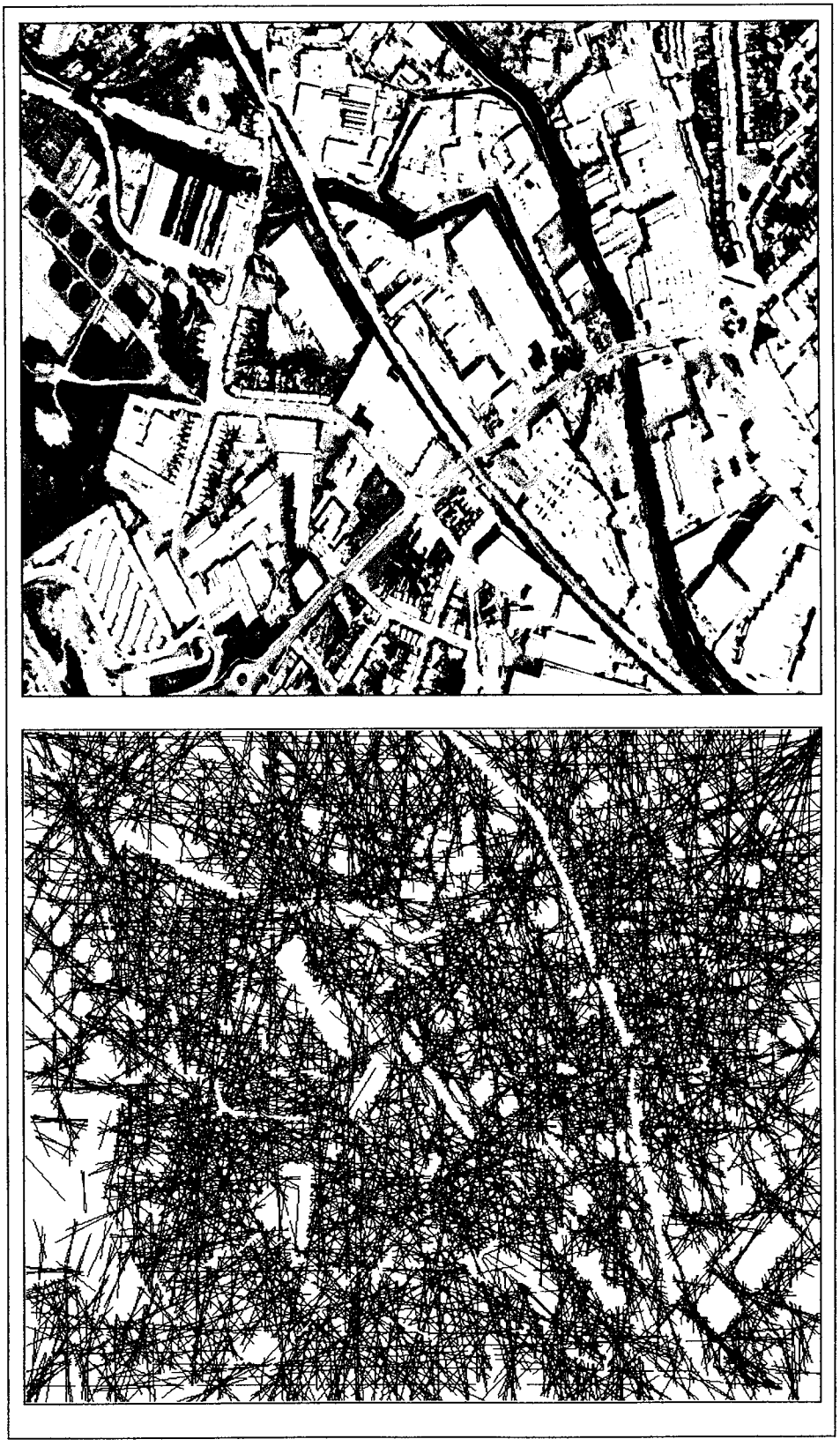


Figure 6.1 The lines produced from a Hough Transform of a region of Dudley (Band11)

6.2.3 The Muff Transform

One of the problems seen in Figure 6.1 is that there are many lines crossing the entire image. This is a result of the use of global edges to determine the location of local lines. A better approach to tackle this problem is to break the image into smaller regions and to extract the straight lines in those regions. It will then subsequently be possible to combine co-linear lines if they meet an appropriate criterion. The Muff transform (Wallace, 1985) was developed with this in mind.

It is small linear features rather than large linear features that are of interest here. The purpose of this approach is the same as the Hough Transform, namely to 'smooth out' the classified map to improve its appearance. The improved appearance, with buildings having straight edges, for instance, would lead to other improvements being possible, making use of the enhanced description of objects within the scene.

6.2.4. Results of the Muff Transform

The results of the Muff Transform are shown in Figures 6.2 a-c. A road side curb is highlighted in Figure 6.2(a). The same region is also highlighted (Figure 6.2 (b)) in the zero-crossing edge detection of the original image. The extraction of this straight line is shown in Figure 6.2 (c).

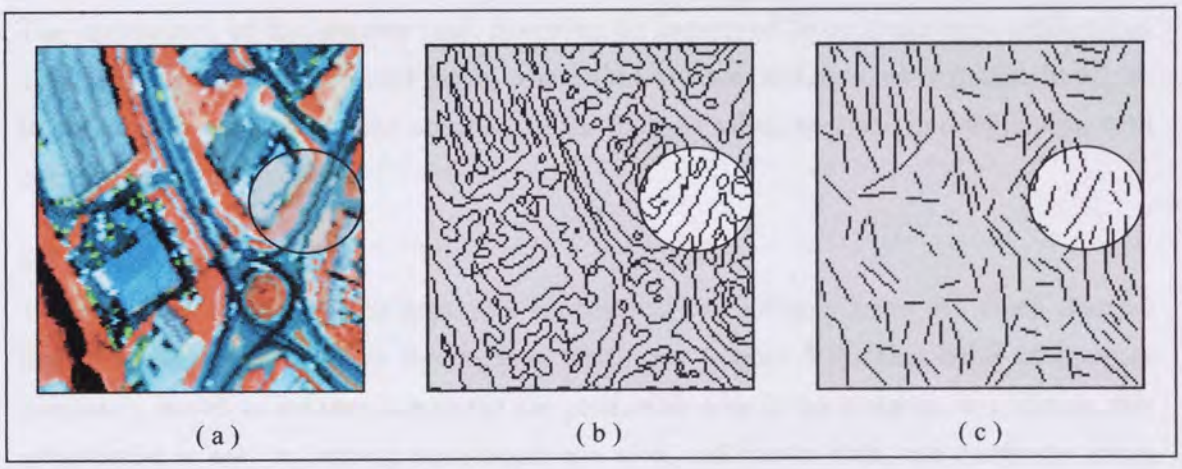


Figure 6.2. The Extraction of straight lines from aerial imagery using the MUFF transform.

Unfortunately, while the extraction of some of the lines in the imagery is possible, the image is complicated by the many edges detected. This results in the extraction of a

considerable number of spurious lines, as well as some genuine ones. As a result, the image is of little use. Any snapping of the lines to produce polygons would result in many polygons of little use in the classification procedure, and would certainly not improve the appearance of the image.

6.2.5 Analysis of results

Part of the problem is that the imagery suffers from distortion due to the motion characteristics of the plane. However, it is doubtful whether the line extraction procedure would be of any benefit if the imagery was accurately rectified and consequently had a plethora of straight lines for examination. The reason, as we have demonstrated in the previous sections, it that urban imagery is of a particularly complicated nature.

6.2.6 Conclusion

The Hough and Muff transforms are both better suited to simple scene analysis, where the imagery is of a much less complicated nature. The imagery used for this project is of a deeply urban region with many overhanging trees and areas of wasteland, which resemble noise, and increase its complexity still further. This is unfortunately representative of the problems faced when dealing with this type of imagery. In addition, the difficulty in geo-rectifying the line-scan imagery (Chapter 3) reduces the amount of straight lines present in it, since many roads and large buildings are warped in appearance.

The appearance of the imagery may, however, be improved by accurate geo-rectification. This will result in buildings and roads, where they are not occluded, having straight edges. In addition, it will improve the accuracy of the imagery when analysed in conjunction with correct ground truth data.

6.3 Dawn Thermal ATM

The extension of the data to include a further band of imagery from the dawn thermal flights would vastly enhance the daytime image set (Figure 3.2). This band of imagery contains a wealth of information about the permeable area in the imagery. In addition, this information is easy to extract, since roads are cold, and hence dark, and roofs are either cold too, or show up as very hot objects, in the case of factories operating overnight.

The problem with this imagery is that while it has a considerable amount of information about land use, it cannot currently be rectified to an accuracy of greater 10m (Gregory, 2000). Consequently the subject of geo-rectification is a pertinent one, as without it the

two temporally different image sets cannot be co-registered. The problem of co-registration, therefore, is a non-trivial task, as seen in section 2.2.6.

The result of this is that it is difficult to measure the success of the classification procedures in real applications, since the accuracy is always lowered as a result of geo-rectification errors. This is an area that has a real need to be addressed, since there are countless ATM data sets available, all of which suffer from this problem. Furthermore, the co-registration of imagery, such as might be required for overlapping or multi-temporal imagery, is impossible with any acceptable accuracy.

An example of the type of error caused by instability of the aeroplane when acquiring the data is shown in Figure 6.3. The factory roof highlights the error, which is reproduced across the affected scan lines and causes warping in the objects that are present.

Figure 6.3. Platform instability results in warping of the line-scan imagery.

Several methods have been presented in chapter 2, which claim to improve the situation. A further method has been developed by Gregory (2000) and has been applied to the same imagery as that used in this project. The method is a two-stage procedure:

1. Identify the platform attitude errors in line-scanner imagery.

2. Correct the imagery by compensating for the platform errors.

Because aerial photographs are instantaneous 'snapshots' of the surface under observation, it is assumed that geometric errors are deterministic in nature. As a result, they can be compensated for by constructing appropriate mathematical models. Once aerial

photographs have been corrected, they in effect become accurate digital maps to which additional data sets can be referenced. This is a standard photogrammetric procedure

Platform attitude errors can now be identified by directly comparing ground control points in the line-scanned ATM imagery with those in the corresponding aerial photography. Ground control points can be identified for the ends of every scan line in the image and this procedure can be automated using correlation techniques. Having identified the image disparities, the raw image data can be re-sampled to the corrected grid, producing ATM imagery that is co-registered with the corrected aerial photography.

The advantage of this method is that, providing there is aerial photography accompanying the digital imagery, existing data-sets can be registered. This method is the basis of ongoing research at Aston University. Recent reports (Elgy *et al*, 1998; Gregory 2000) suggest that this method may meet with considerable success.

6.4 Conclusion

There are still several areas in need of improvement. The appearance of the imagery needs improving, which may be achieved by geo-rectification of the data. This is a problem that is relevant to the other classification methods examined also, since it is an inherent problem with ATM imagery. The most promising improvement is in the incorporation of the dawn thermal imagery into the data-set, which is unfortunately linked to the geo-rectification problem. The location of the straight edges may, however, still be difficult to achieve due to the complicated nature of urban imagery.

This chapter has highlighted some of the problems that still remain in the classified imagery. It has reviewed several methods, each with varying degrees of success, that may improve both the accuracy and the aesthetics of the imagery. The methods discussed provide areas for further research, some of which is ongoing at Aston University as part of a separate project (Elgy *et al*, 1998; Gregory, 2000).

Chapter 7: Applications of the Polygon Classifier

7.1 Introduction

Classified imagery has a variety of geographical applications. The purpose of this research is to develop a classifier for urban imagery so that an accurate Percentage of Impermeable Area (PIMP) measure can be extracted and subsequently inputted into an urban drainage model. The majority of models in use require several inputs, of which the PIMP is one. Table 7.1 shows the PIMP measures extracted from imagery classified using the supervised and unsupervised Polygon Classifier.

7.2 Calculation of PIMP

The reclassification of the imagery from 4 classes into 2 classes (permeable and impermeable) provides an image from which the PIMP can be extracted. The calculation of the PIMP is made by counting the number of pixels classed as impermeable within the area.

Classifier	Polygon Classifier (Supervised) (Figure 7.2)	Polygon Classifier (Unsupervised)(Figure 7.1)
Percentage of Impermeable Area (PIMP)	66.64%	57.82%

Table 7.1 The PIMP for the Polygon Classification

The supervised polygon classifier produces a PIMP value 10% higher than that produced by the unsupervised classifier. This can make a big difference to the urban drainage capacity that the model will calculate as being required. The reasons for the difference are clear from the images presented in Figure 7.1 and 7.2. The railway running diagonally across the image is classified correctly as permeable by the supervised polygon classifier, and impermeable by the unsupervised one. This is because the railway was not included as a class in the unsupervised maximum likelihood classifier that was used to provide the statistics.

This error, however, is countered by the larger amount of roof area indicated by the supervised classifier. There are large areas of residential and factory roof that have been extracted by the supervised method. Consequently this method produces a higher value for the PIMP.

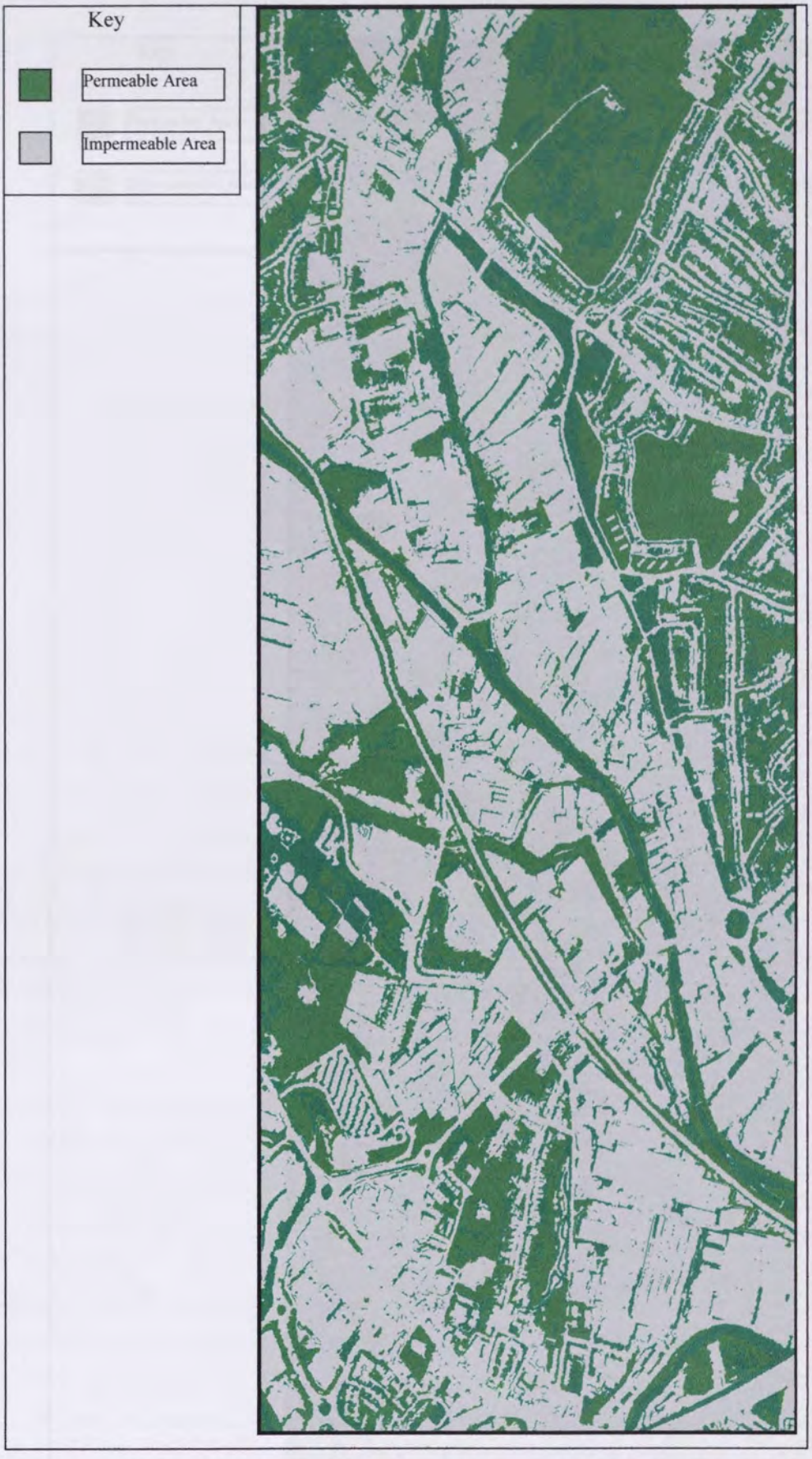


Figure 7.1. The Polygon Classified PIMP Image (2 Classes) using Unsupervised Statistics.

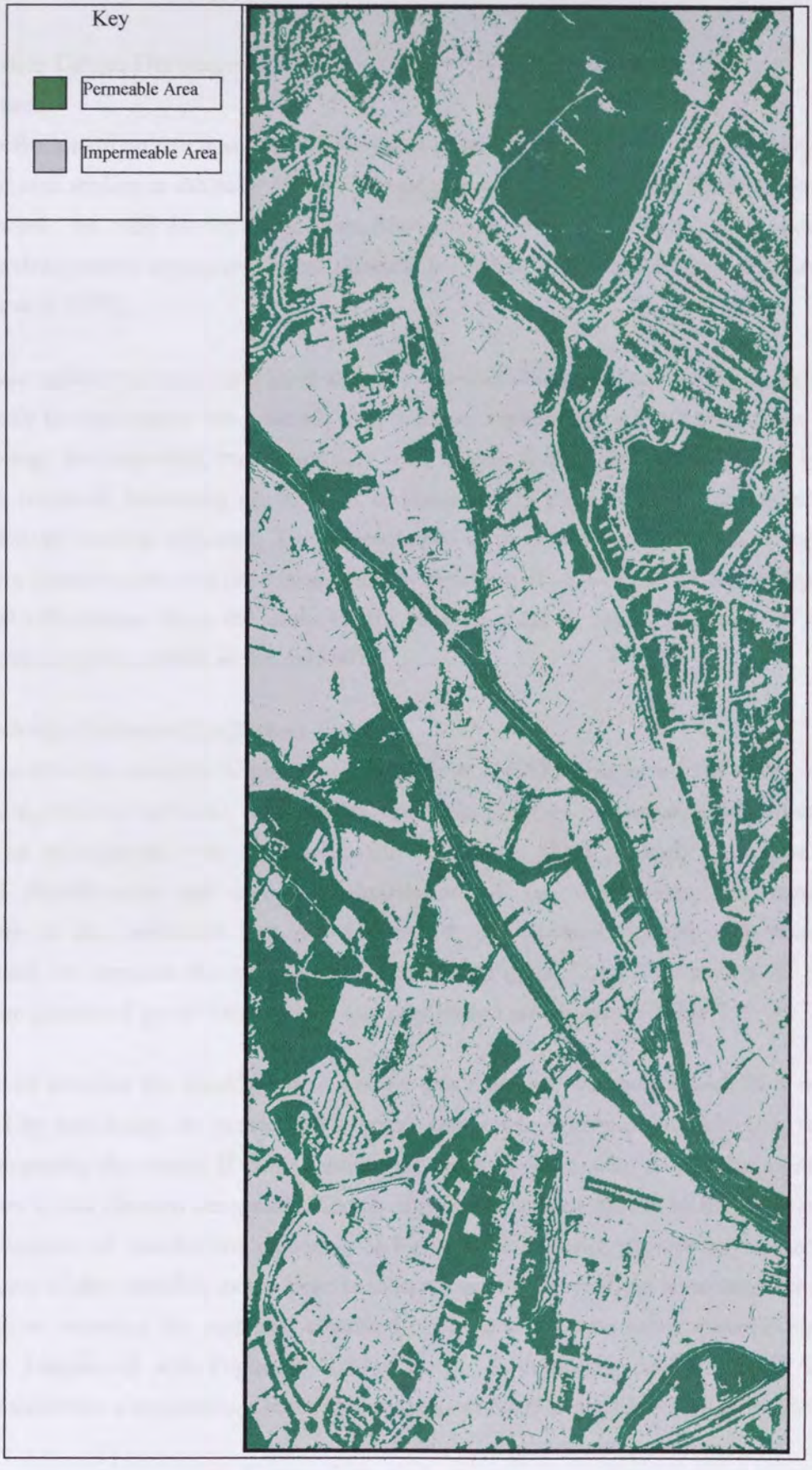


Figure 7.2. The Polygon Classified PIMP Image (2 Classes) using Supervised Statistics.

7.3 Input into Urban Drainage GIS

7.3.1 Introduction

The James Bridge urban catchment in Darlaston, Walsall, West Midlands, was the subject of drainage area studies in the early 80's, as part of plans to reconstruct the Black Country sewer network. As well as traditional overflow structures and retention tanks, novel systems, hydrodynamic separators, were planned for urban areas in the West Midlands (Blagojevic *et al*, 1998).

The land use within the catchment used as a test case has changed considerably in recent years. Initially the catchment was a heavily industrialised region, with a small percentage of dense housing. Subsequently, many factories were closed during the recession and had their roofs removed, increasing the amount of wasteland in the catchment. In addition, new low-density housing appeared. The consequence of these temporal land-use changes was that the impermeable area decreased, thereby reducing the runoff. Thus hydrological studies and calculations from the early 1980's have resulted in the over-design of the hydrodynamic separators built in the mid-80's.

7.3.2 James Bridge Catchment Classification

The data used in the study by Blagojevic (Blagojevic *et al* 1998; Blagojevic *et al* 1994) was acquired using low cost airborne videography, due to its accurate geo-registration and ease of use. The videography was classified with 11 classes, through both a Maximum Likelihood classification and on-screen digitisation of the videography. A manual classification of the catchment area was performed and digitised, for use as a source against which to compare the classification accuracy (Figure 7.3). The results of the classification produced by the videography and digitisation are shown in Table 7.2.

A measure of whether the classified videography was adequate was performed. This was determined by calculating the percentage of roof, permeable, and impermeable area, and directly comparing the values. If these values were close to those achieved by the manual classification it was deemed acceptable. Clearly this is not an assessment of the accuracy, but an assessment of whether the data was 'fit for use'. This cannot provide any measure of the success of the classifier, as we have seen in Chapter 4, since there is no measure of the correlation between the manually classified image and the two other classifications (Maximum Likelihood and Digitised Videography). A complete assessment of the accuracy is therefore a requirement if any comparison with other methods is to be made.

The availability of a manually classified map of the entire region means that the assessment of the accuracy of the image is considerably easier than would otherwise be the case. It should be noted that this is not normally the case. The measurement can be made using a confusion matrix, without the need for a sampling strategy such as the single random pixel method suggested in chapter 4. As a result it is possible to determine an absolute measure of accuracy for this imagery, given the assumption that the accuracy of the manual classification is 100%.

Class	Digitised Map(%coverage)	Digitised Videography (%coverage)	Maximum Likelihood. (%coverage)
Roofs	16.14	15.50	34.62
Impermeable	26.09	24.80	13.83
Permeable	34.47	36.40	28.24
Total	76.7	76.7	76.69

Table 7.2: Classification results for the James Bridge Catchment (Blagojevic *et al*, 1998).

Figure 7.3 Manual Classification of the James Bridge Catchment (Blaojevic *et al*, 1998). The white highlighted region indicates the area for which ATM imagery was available.

7.3.3 Polygon Classification (Supervised)

Aston University has an extensive archive of ATM imagery, particularly over the West Midlands region. Of this imagery, a dataset was located that covered approximately 1/3 of the Darleston Catchment examined by Blagojevic *et al* (1998). The area for which imagery was available is highlighted in Figure 7.3. While the coverage of the region was not as extensive as initially thought, it allows an examination of the techniques used by Blagojevic and those developed in this project. Furthermore, the higher resolution of the ATM imagery leads to improvements in the classification in that region of the imagery.

The raw ATM imagery was co-registered prior to the classification procedure using a second order polynomial fit. This proved to be the best geo-rectification method available, in terms of RMS error, despite the normal assertion that higher order polynomial fits are better suited to aerial imagery. The reason for this was due to the absence of recognisable features in certain sections of the imagery, which caused under-representation in those areas and consequential distortion of the imagery. The causes of the line-scan errors and the accurate geo-rectification of these have been discussed in Chapters 2 and 6.

For the purposes of assessing how the developed classification technique performs in a 'real' situation, however, it is essential that the imagery is geo-rectified. Clearly the greater the inaccuracy of the rectification, the less well the classifier will be seen to perform if assessed in comparison to geo-referenced manual imagery. However, the geo-rectification was of sufficient accuracy when taking into account the difference in the two image resolutions.

7.3.4 Results

The results of the polygonal Maximum Likelihood classification are shown in Figure 7.4. Figure 7.5 shows the results of the maximum likelihood classification of the original videography (Blagojevic *et al*, 1998) for the same region. Table 7.2 shows a comparison between the three methods.

The Polygon classification was performed in supervised mode, as the overall performance, in terms of accuracy and quality, was deemed to be the most effective. The bands used for the classification were the same as those used in Chapter 3. The classifier was applied to the raw imagery and subsequently geo-rectified as outlined above.

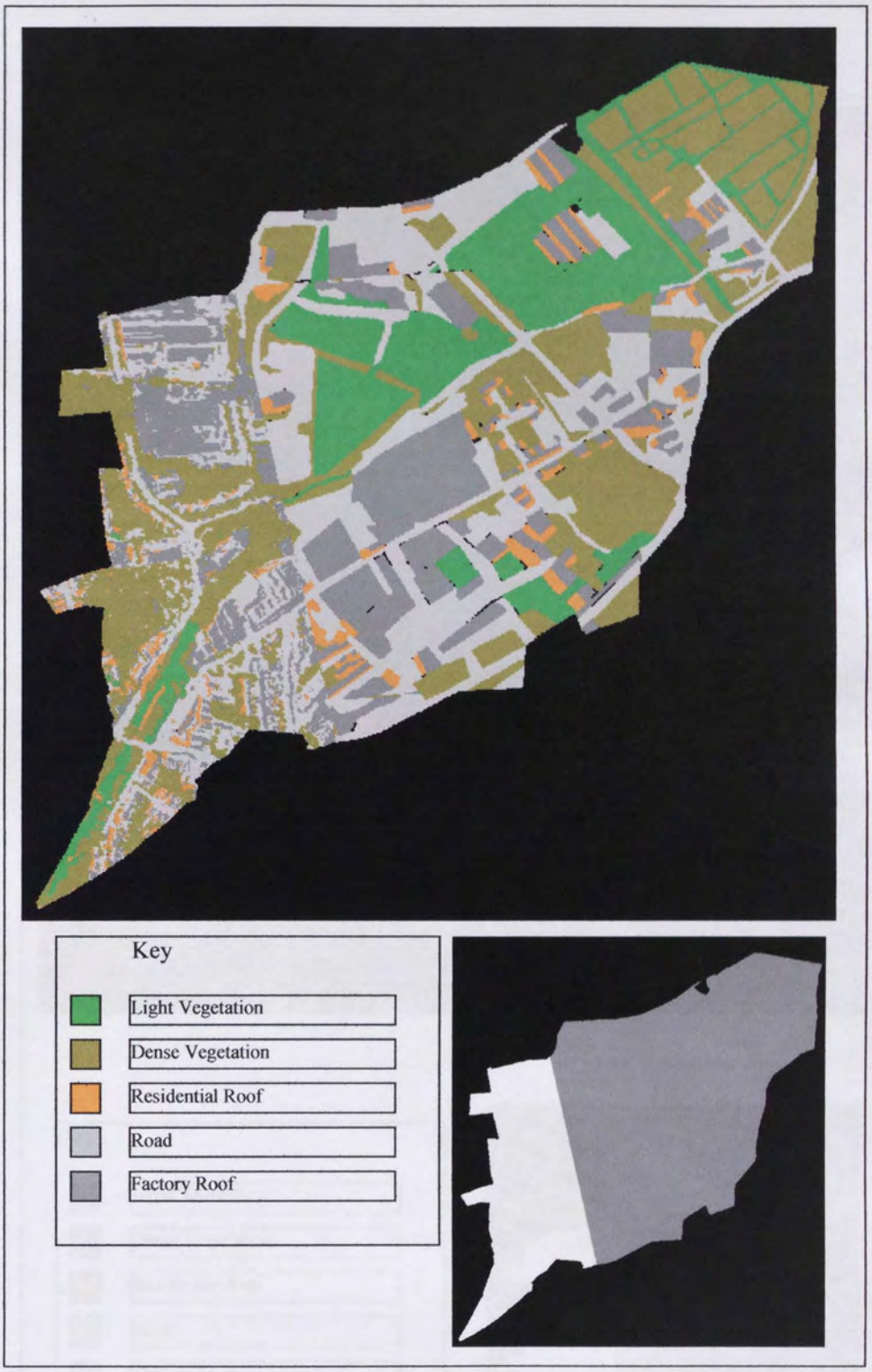


Figure 7.4 The Polygon Classified ATM Imagery overlaid on the manually classified Map of the James Bridge Catchment.

Figure 7.5 The Maximum Likelihood classification of the Vidoegraphy (Blaojevic *et al*, 1998) overlaid on the manually classified map.

7.3.5 Analysis of Results

The comparison of the classifiers shown in Table 7.3 indicates that the Polygon classification of the ATM imagery produces a classification of similar quality to the manual classification, which vastly out-perform the Maximum Likelihood classification of the videography. If the PIMP value calculated by the polygonal classifier is used as an input into an urban drainage model, it would produce good results - the value obtained is only 5 percent lower than that obtained by the manual classification.

Class	Manual	Polygon Classifier (ATM data)	Maximum Likelihood (Videography)
Light Vegetation	2.68%	3.87%	5.81%
Dense Vegetation	34.51%	36.49%	26.26%
Factory Roof	21.28%	25.94%	37.25%
Road	35.80%	29.20%	12.73%
Residential Roof	5.73%	4.50%	17.95%
PIMP	62.81%	59.64%	67.93%

Table 7.3 The percentage cover of the sub-area is shown for each method .

However, it would be incorrect to suggest that the accuracy of this classifier was 95 percent. As we have seen in Chapter 4, this value could be attainable without any of the pixels classified by the polygonal classifier matching any of those produced by the manual method. There is clearly a danger associated with accepting the results as presented in Table 7.3, since there is no guarantee that the same 'accuracy' can be repeated with different imagery. A more rigorous assessment of the accuracy must be sought if any claims regarding the quality of the classification are to be made.

An assessment of the accuracy can be undertaken by examining both the Kappa Values and the confidence level at 95%. The presence of the manually classified map also means

that there is a large sample size unaffected by bias since it covers the entire region. This also allows us to examine further which of the two measures is the most appropriate to use.

The confusion matrices for the different classifications are shown in Tables 7.4 and 7.5, and those for the PIMP calculation are shown in tables 7.6 and 7.7. Figure 7.6 illustrates the original classification and Figure 7.7 the PIMP calculation. It is evident that the Polygon classification greatly outperforms the Maximum Likelihood classifier for both these classifications. The overall accuracy for each class is certainly low, however. This merely confirms that the 'fit for use' method for determining the accuracy is extremely misleading if used as a measure of classification accuracy.

There is considerable confusion between all the classes for both of the classified images. This can be explained, in part, for the Polygon Classifier through errors in geo-rectification - a misalignment of just one pixel can affect the accuracy significantly. Given this, it appears that the reason for the improved classification is due to the quality of the ATM data over the Videography, and an improvement in the classification procedure.

The poor performance of the Maximum Likelihood classifier cannot be solely attributed to geo-rectification errors. Videography has a great advantage over line-scan imagery in this respect, since the achievement of accurate geo-rectification is not compromised by platform instability. The image acquired is a 'snapshot', much the same as an aerial photograph. Consequently, standard photogrammetric techniques can be used to rectify the image.

Class	Light Vegetation	Dense Vegetation	Factory Roof	Road	Residential Roof	Total	correct	omission	commission	Kappa	Accuracy at 95% CL
Light Vegetation	7	599	267	336	111	1320	0.53%	99.47%	151.89%	-3.72%	0.18%
Dense Vegetation	1169	8774	2127	3956	933	16959	51.74%	48.26%	53.75%	24.12%	51.09%
Factory Roof	302	2086	4152	3482	436	10458	39.70%	60.30%	81.93%	18.65%	38.90%
Road	534	5733	5269	5437	619	17592	30.91%	69.09%	50.48%	2.50%	30.32%
Residential Roof	0	697	905	1106	110	2818	3.90%	96.10%	74.49%	-0.62%	3.28%
Total	**2012**	**17889**	**12720**	**14317**	**2209**	**49147**	**37.60%**	**74.64%**	**82.51%**	**8.19%**	**24.75%**

Table 7.4 Confusion Matrix for the Polygon Classification of a section of the James Bridge Catchment (5 Classes).

Class	Light Vegetation	Dense Vegetation	Factory Roof	Road	Residential Roof	Total	correct	omission	commission	Kappa	Accuracy at 95% CL
Light Vegetation	50	258	727	212	71	1318	3.79%	96.21%	212.97%	-2.14%	2.87%
Dense Vegetation	1822	3688	7427	1864	2158	16959	21.75%	78.25%	54.35%	-6.12%	21.22%
Factory Roof	453	2918	3742	1257	2088	10458	35.78%	64.22%	139.25%	-2.33%	34.99%
Road	498	5451	5654	2560	3431	17594	14.55%	85.45%	21.00%	2.09%	14.10%
Residential Roof	34	591	755	362	1076	2818	38.18%	61.82%	274.95%	24.66%	36.61%
Total	**2857**	**12906**	**18305**	**6255**	**8824**	**49147**	**22.62%**	**77.19%**	**140.51%**	**3.23%**	**21.96%**

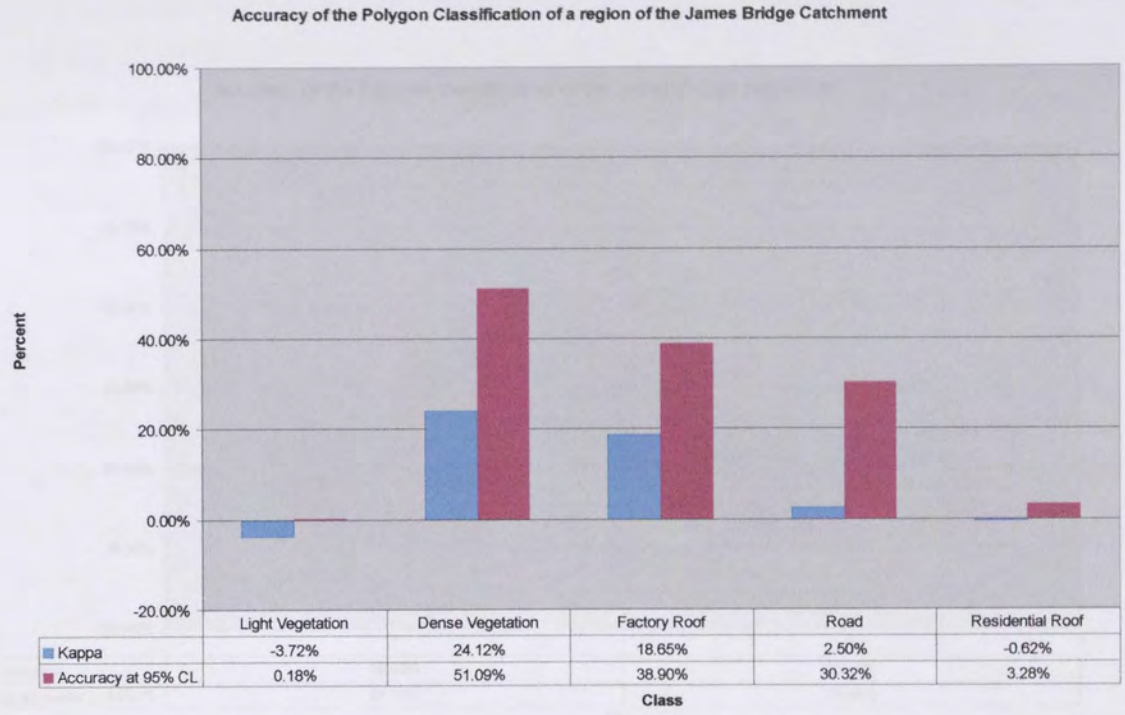
Table 7.5 Confusion Matrix for the Maximum Likelihood Classification of a section of the James Bridge Catchment (5 Classes).

	Permeable	Impermeable	Total	Correct	Omission	Commission	Kappa	**Accuracy at 95% CL**
Permeable	10549	7730	18279	57.71%	42.29%	51.16%	28.93%	57.10%
Impermeable	9352	21516	30868	30.30%	69.70%	25.04%	25.18%	29.86%
Total	**19901**	**29246**	**49147**	**65.24%**	**56.00%**	**38.10%**	27.06%	43.48%

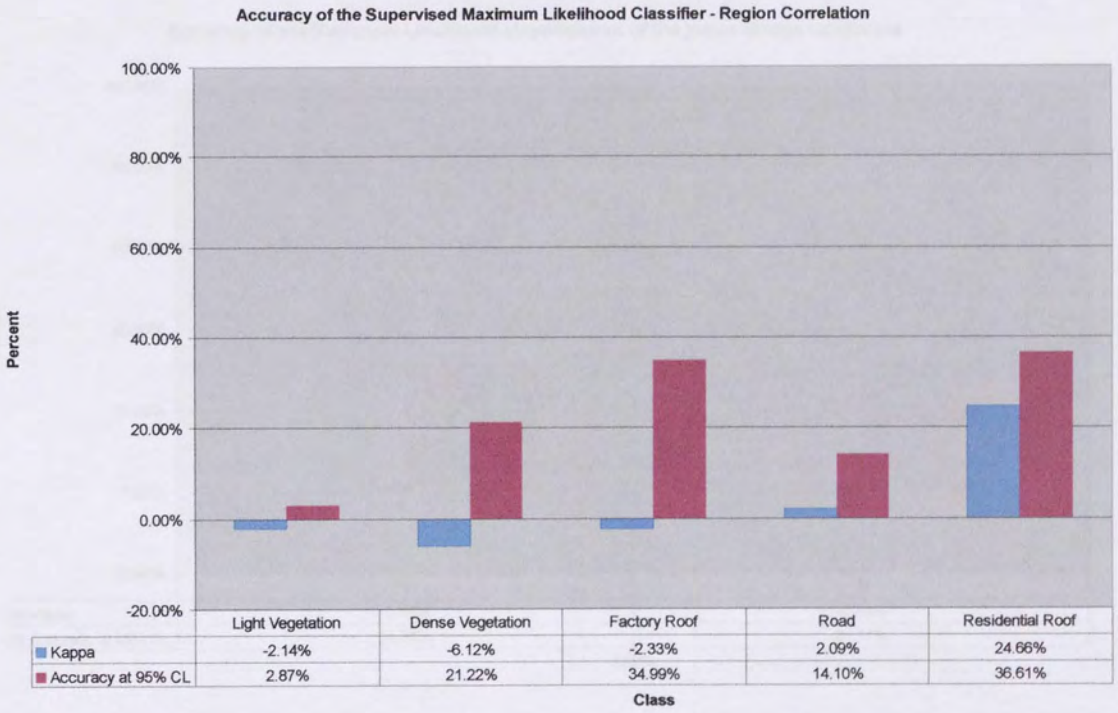
Table 7.6 Confusion Matrix for Polygon Classifier (PIMP).

	Permeable	Impermeable	Total	Correct	Omission	Commission	Kappa	**Accuracy at 95% CL**
Permeable	5818	12459	18277	31.83%	68.17%	54.41%	-0.35%	31.25%
Impermeable	9945	20925	30870	32.22%	67.78%	40.36%	-0.44%	31.77%
Total	**15763**	**33384**	**49147**	**54.41%**	**67.98%**	**47.39%**	-0.40%	**31.51%**

Table 7.7 Confusion Matrix for Maximum Likelihood Classifier (PIMP).

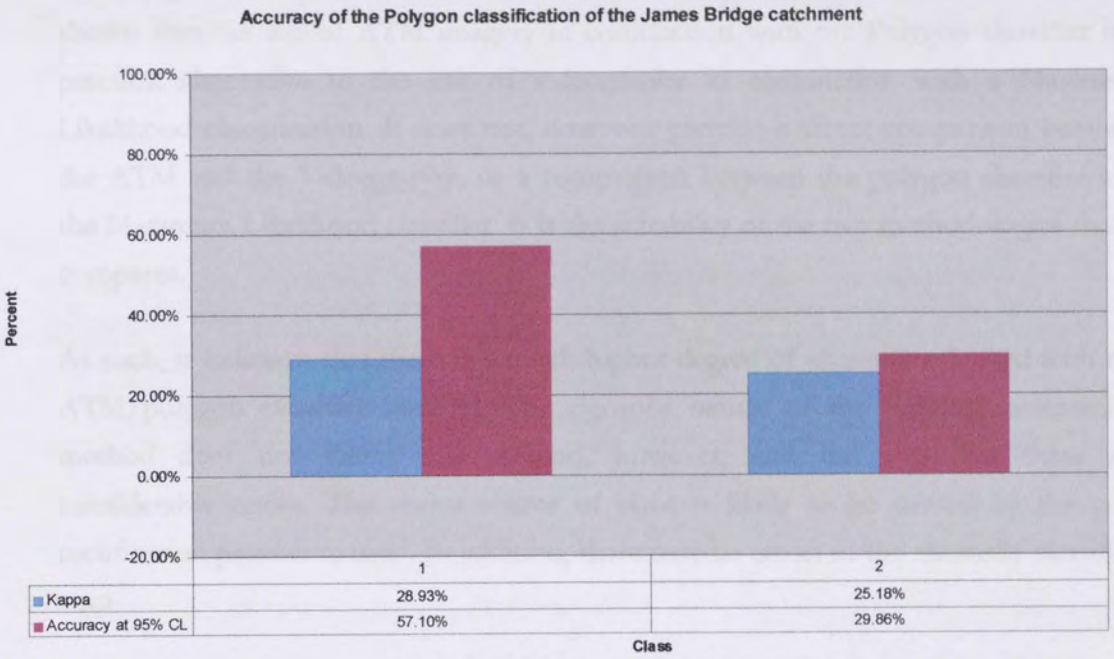


	Light Vegetation	Dense Vegetation	Factory Roof	Road	Residential Roof
Kappa	-3.72%	24.12%	18.65%	2.50%	-0.62%
Accuracy at 95% CL	0.18%	51.09%	38.90%	30.32%	3.28%



	Light Vegetation	Dense Vegetation	Factory Roof	Road	Residential Roof
Kappa	-2.14%	-6.12%	-2.33%	2.09%	24.66%
Accuracy at 95% CL	2.87%	21.22%	34.99%	14.10%	36.61%

Figure 7.6 A comparison between the Polygon classifier and Maximum Likelihood classifier for a region of the James Bridge catchment (5 Classes)



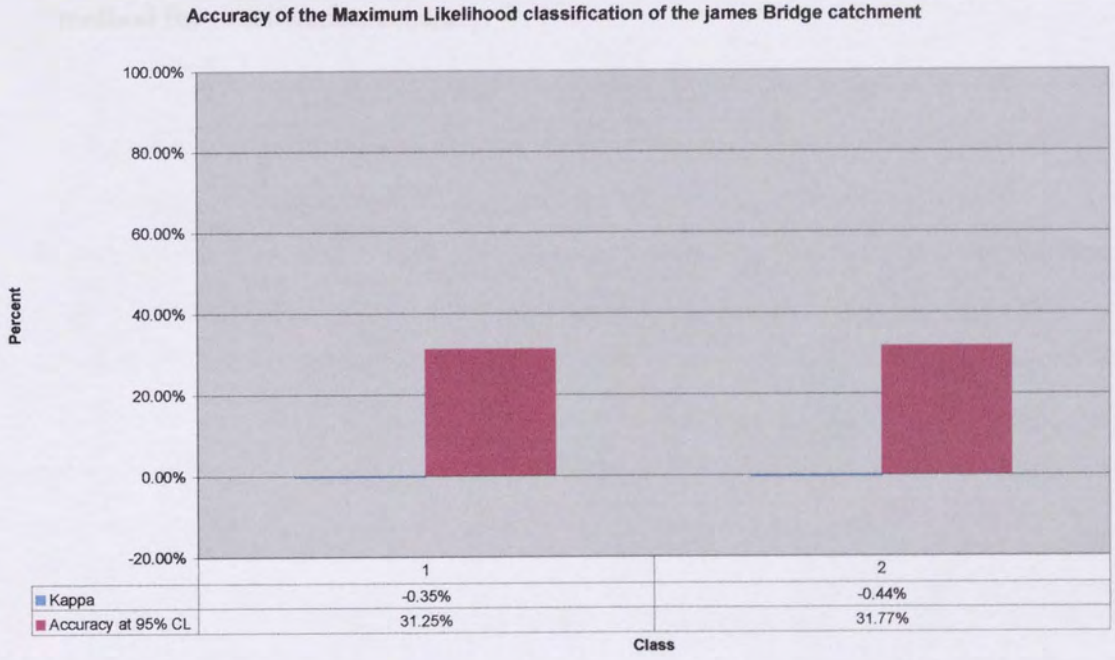


Figure 7.7: Comparison of the accuracy of the Polygon classifier and the Maximum Likelihood classifier for a region of the James Bridge catchment (2 classes)

7.4 Conclusion

This chapter has illustrated the differences between the two methods used. It has shown that the use of ATM imagery in conjunction with the Polygon classifier is a practical alternative to the use of videography in conjunction with a Maximum Likelihood classification. It does not, however, provide a direct comparison between the ATM and the Videography, or a comparison between the polygon classifier and the Maximum Likelihood classifier. It is the suitability of the two methodologies that it compares.

As such, it indicates that there is a much higher degree of accuracy achieved with the ATM/polygon classifier method. The rigorous nature of the accuracy assessment method does not flatter this method, however, and indicates that there are considerable errors. The major source of error is likely to be caused by the geo-rectification procedure used. In addition, there may be errors in the manually classified map

This comparison has also highlighted the effectiveness of the Kappa statistic as a method for classification accuracy.

Chapter 8: Conclusion

The purpose of this project has been to investigate what improvements the introduction of context to the classification procedure can provide. In particular, the focus has been on the classification of urban ATM imagery for providing input into an urban drainage model.

Initially four conventional classification procedures (supervised Maximum Likelihood, SMAP, ECHO and unsupervised Maximum Likelihood) were performed, in order to find out the suitability of each for this task. This highlighted the inadequacies of per-pixel classification. From a qualitative viewpoint the Maximum Likelihood classifiers produced the best classification in terms of discrimination of the classes, although the appearance of the classified map was lacking in homogeneity. In order for the quality of the classification to be qualified, there arose a need to determine an accuracy assessment method that would provide comprehensive detail about the classification procedure, while remaining unbiased so that a fair comparison between classifiers could be made.

This aspect was developed and consequently produced the recommendation that the use of the Kappa value combined with the single random pixel method, providing the sample size was adequate, would provide this comparison. The confidence level, while being an important measure of the sample size, does not encapsulate information about the errors of omission and commission of a classification. The region correlation sampling method was found to inherently introduce bias from a variety of sources, and as such is of little scientific use. The development of the assessment method allows an unbiased and comprehensive comparison between the existing classification methods and the new one developed in this project.

From this research a platform was provided from which a new classifier, based on a per-polygon approach, could be developed. As the review in Chapter 2 indicated, the edges of objects in imagery contain much important information about the objects themselves. The polygons contained within the edges have mostly homogeneous properties that can be used by the classifier. These polygons provide continuity throughout the image and allow the features in the image, such as roads and roofs, to be easily extracted.

The new classifier is based on a per-polygon classification, which incorporates contextual information through the extraction of polygons produced by the Marr Hildreth zero-crossing edge detector. This edge detector, while not optimal in the mathematical sense, locates intensity contours in the imagery, which represent the features that a photo-interpreter looks for when classifying an image. The best results for this edge detection were achieved when using a false colour composite image, which combined band 3 (green), band 5 (red) and band 7 (near infrared). This imagery highlights vegetation and man made features in the imagery and enables the production of an edge map that richly describes these features.

This edge map is used to incorporate contextual information into the classification procedure. The classification of each individual pixel is dependent on the class to which the polygon it is contained in is assigned. Each polygon is subjected to a test of homogeneity - if it is found to fail this test it is further refined until it becomes homogeneous. This test of homogeneity is tuned towards the urban drainage model requirements, which are that roads and roofs contain the most important information. Consequently the thermal band, which is the most discriminating band for these features, is used for this task.

The new per-pixel classification procedure was then subjected to the rigorous accuracy assessment approach developed in Chapter 4, and found to produce a classified image of improved accuracy and appearance. Ease of implementation indicates that this is a practical approach to land-use classification

Subsequently some potential improvements to the classifier were examined. Attempts to further improve the appearance of the classified image through the use of a Hough transform and the Muff transform were found to be unsuccessful, due to the complex nature of urban imagery. The presence of many closely packed edges in the imagery resulted in many spurious edges being extracted by these techniques.

Dawn thermal imagery could potentially be used to enhance the dataset. In this imagery, roads and roofs have a high contrast relative to other features in the imagery. As such it is likely that these features could be extracted with relative ease (perhaps through thresholding the imagery.) However, problems encountered with geometric correction of the ATM data

mean that it is currently impossible to combine both data sets. The problem of ATM image geo-rectification is part of ongoing research at Aston University.

A comparison of the method developed and that used in a previous study was then made. This found that the developed method was considerably more effective than the use of per-pixel classification combined with videography.

From this research it is recommended that the supervised Polygon classifier is used with ATM imagery to classify land-use information for urban drainage modelling.

References

ArcInfo, (1998). ArcInfo User Manual, *Available from* ESRI, 380 New York Street, Redlands, CA 92373-8100, USA.

Argialas, D.P., and Harlow, C.A., (1990). 'Computational Image Interpretation Models: An Overview and a Perspective', *Photogrammetric Engineering and Remote Sensing*, **56**, No. 6, June 1990, 871-886.

Atkinson, P.M., and Tatnall, A.R.L, (1997). 'Neural Networks in Remote Sensing: Introduction', *International Journal of Remote Sensing*, **18**, No. 4, 699-709.

Atkinson, P.M., Cutler, M., and Lewis., (1997). 'Mapping sub-pixel variation in land cover in the U.K. from AVHRR imagery', *International Journal of Remote Sensing*, **18**, 917 – 935.

Ballard, D.H., and Brown, C.M., (1982). Computer Vision, Englewood Cliffs, N.J., Prentice-Hall.

Barnsley M.J. and Barr S.L., (1993). 'Mapping urban areas using high spatial resolution remotely sensed images. 1. Data Processing' *Proceedings of the NERC Symposium on Airborne Remote Sensing*, 20-21 Dec, 1993. Dundee. 1-13.

Besag, J. (1986). 'On the Analysis of Dirty Pictures', *Journal of the Royal Statistical Society*, **48**, No.3, 259-302.

Bishop, C.M., (1995), Neural Networks for Pattern Recognition, Oxford University Press, Oxford.

Blagojevic, B., Elgy, J., Chen, Z., and Maksimovic, C. (1994), 'Airborne Videography as Data Source for an Urban Hydrological Model', in Remote Sensing and GIS in Urban Waters (UDT '94) Ed. Maksimoic, C., Elgy, J., and Dragalov, V, p121 – 129.

Blagojevic, B., Maksimovic, C., and Elgy, J., (1998). 'The Use of Airborne Videography for Simulating Urban Drainage Runoff', *Hydrology in a changing environment – Proceedings of the First International Conference of the British Hydrological Society, **III**, p375 – 382.*

Booth, D.J., (1989). Classification and Contextual Enhancement of Remotely Sensed Data, Ph.D. Thesis, University of Aston in Birmingham, Birmingham, UK.

Bouman, C., and Shapiro, M. (1992), 'Multispectral Image Segmentation Using a Multiscale Image Model', *Proceedings of the IEEE International Conference o. Acoustics, Speech and Signal Processing*, San Fransico, California, March 23-26, 1992, **3**, 565-568.

Bouman, C.A., and Shapiro, M. (1994), 'A Multiscale Random Field Model for Bayesian Image Segmentation', *IEEE Transactions on Image Processing*, **3**, No. 2, March 1994, 162-177.

Boyle, R.D., and Thomas, R.C. (1988),Computer Vision: A First Course, Blackwell Scientific Publications, 210p.

Brodatz, P. (1966). Textures - a Photgraphic Album for Artists and Designers. Dover, New York.

Burrough, P.A. (1986). Principles of Geographical Information Systems for Land Resources Assessment, Monographs on Soil and Resources Survey No. 12, Clarendon Press, Oxford.

Campbell, J.B., (1983). Mapping the Land - Aerial Imagery for Land Use Information, Resource Publications in Geography.

Canny, J.F., (1986). 'A Computational Approach to Edge Detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **PAMI-8**, No. 6, 679-698, November 1986.

Civco, D.L., (1993). 'Artificial neural networks for land cover classiication and mapping', *International Journal of Geographical Information Systems*, **7**, 173-186.

Cohen, J., (1960). 'A Coefficient of Agreement of Nominal Scales', *Educational and Psychological Measurement*, **20**, no.1, 37-46.

Colwell, R.N. (1983). Manual of Remote Sensing Volume II, American Society of Photogrammetry.

Congalton, R.G., (1991). 'A review of Assessing the Accuracy of Classifications of Remotely Sensed Data', *Remote Sensing of the Environment*, **37**, 35-46.

Crane, R.B., Malila, W.A., and Richardson, W., (1972). 'Suitability of the normal density assumption for processing multi-spectral scanner data', *IEEE Transactions on Geoscience Electronics*, **GE-10**, 158-165.

CSS (1989). Benefits and Risks of Knowledge-Based Systems, Report of a Working Party Council for Science in Society, Oxford University Press.

Curran, P.J., (1985). Principles of Remote Sensing, Longman, London.

Danielsson, P-E and Seger, O. (1990). 'Generalized and Separable Sobel Operators', in Machine Vision for Three-Dimensional Scenes, Ed Freeman, H., Academic Press.

Danson, F.M. (1986), 'Remotely sensed red and near infra-red response to forest canopy cover', *Proceedings of the NERC 1985 Airborne Campaign Workshop*, 16th Nov., Ripton, Cambs, UK. A1-A10.

Davis, W.A. and Peet, F.G. (1977). 'A method of smoothing digital thematic maps', *Remote Sensing of the Environment*, **6**, 45-49.

DHI (1995). Technical Reference Manual of MOUSE, *Danish Hydraulic Institute*, Hoersholm, DK-2970, Denmark.

Eastman, J.R., (1995). IDRISI for Windows User's Guide (Version 1.0), Clark Labs for Cartographic Technology and Geographic Analysis, Clark University, 950 Main St, Worcester, MA, 01610-1477 USA.

Ehlers, M., (1994). 'Geometric Registration of Airborne Scanner Data using Multi-Quadratic Interpolation Techniques', *Proceedings of the 1st International Airborne Remote Sensing Conference and Exhibition*, **II**, 12-15 September, Strasbourg, France, 494-502.

Eklundh, J-O, (1983). 'A Structured Approach to Segmentation of Aerial Photographs', *Proceedings of SPIE International Society of Optical Engineering*, Part 397, 1983, 21-27.

Elgy, J., Buchanan, A.J., and Gregory, S. (1998). 'Airborne Remote Sensing for Urban Waters', UDM '98, p461 – 474.

Elgy, J., Charnock, T.W., and Hedges, P.D. (1995), Chapter VIII, New Technologies for Large Water Supply Systems, NATO Advanced Studies Institute, 2 Environment, Vol 15, Ed Maksimovic, C., Calomino, F., and Snoxell, J., p 465 – 555.

Elgy, J., Maksimovic, C., and Prodanovic, D. (1993). 'Matching Standard GIS Packages with Urban Storm Drainage Simulation Sortware', in HydroGIS 93: Application of Geographic Information Systems in Hydrology and Water Resources Management (Proceedings of the Vienna Conference, April 1993). IAHE publ. no. 211.

Ellis, R.J., (1997). Evaluation of Remote Sensing for the Detection of Landfill Gas and Leachate in an Urban Environment, Ph.D. Thesis, Aston University, Aston Triangle, Birmingham, UK.

ERDAS, (1998). ERDAS User manual, *Available from* ERDAS (UK) Limited, Telford House, Fulbourn, Cambridge, CB1 5HB.

ERMapper, (1998). ERMapper Reference Manual, *Available from* ERMapper, Blenheim House, Crabtree Office Village, Eversely Way, Egham, Surrey, TW20 9RY, UK.

Ersoy, O.K., and Hong, D., (1990). 'Parallel self-organising, hierarchical neural networks', *IEEE Transactions on Neural Networks*, **1**, 167-178.

Fankhauser, R., (1998). 'Automatic Determination of Imperviousness in Urban Areas from Digital OrthoPhotos', UDM '98, p321 – 326.

Finch, J., Reid, A., and Roberts, G. (1989), 'The Application of Remote Sensing to Estimate Land Cover for Urban Drainage Catchment Modelling', *Journal of the Institute of Water Environmental Management.*, **3**, No. 6, 558-564.

Fuchs, L. and Sheffer, C. (1993), 'HYSTEM-EXTRAN, improvements to EPA-EXTRAN', *Proc. 6th Int. Conf. Urban Storm Drainage*, **I**, 1993, 237-242.

Fuchs, L., Sheffer, C., and Spöneman, P. (1994). ' User Needs in Interfacing GIS and Satellite Data fo Urban Storm Drainage', in Remote Sensing and GIS in Urban Waters (UDT '94), p113 – 120.

Galloway, M.M. (1975). 'Texture Analysis using Gray Level Run Lengths', *Computer Graphics and Image Processing*, **4**, 172-179.

Giarratano, J.C. and Riley, G. (1994). Expert Systems - Principles and Programming, PWS Boston.

Gilmore, J.F., and Roth, S.P., (1988). GEST (Generic Expert System Tool) Reference Manual, Version 3.0, Georgia Technical Research Institute, Atlanta, Georgia, USA.

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.

Gregory, S., (1996). Geo-Rectification of Airborne Thematic Mapper Imagery using Aerial Photography, Final Year Research Project, Aston University, Birmingham, UK.

Gregory, S., (2000). The Geometric Correction and Registration of Airborne Line-Scanned Imagery for Temporal Thermal Studies, Ph.D. Thesis, Aston University, Birmingham, UK.

Haralick, R.M. (1979), 'Statistical and Structural Approaches to Texture', *Proc. IEEE*, **67**, No 5, May 1979, 786-803.

Haralick, R.M, and Joo, H. (1985), 'Context Classifier', *IGARSS '85*, 247-254.

Hay, A.M., (1979). 'Sampling Designs to Test Land-Use Map Accuracy', *Photogrammetric Engineering and Remote Sensing*, **45**, 529-533.

Hild, M., and Shirai, Y. (1993), 'Extraction of Textured Elements from Images of Shaded Scenes', *Pattern Recognition*, **26**, No. 8, 1177-1191.

Jensen, J.R. (1986). Introductory Digital Image Processing - A Remote Sensing Perspective, Prentice Hall.

Jewell, T.K., and Adrian, D.D., (1978). 'SWMM Stormwater pollutant washoff functions', *Journal of the Environmental Engineering Division*, **104(5),** 489-499.

Kälviäinen, H., Hirvonen, P., and Oja, E., (1996). 'Houghtool - a Software Package for the Use of the Hough Transform', *Pattern Recognition Letters*, **17**, No. 8, pp. 889-897.

Krinn, G.J. (1992). 'GIS Operators Requiring Spatial Context and their Implications for Remote Sensing', *SPIE*, **1819**, 1992, 126-132.

Kohonen, T., (1988), 'An introduction to neural computing', *Neural Networks*, **1**, 3-16.

Kontoes, C.C., Rokos, D., Wilkinson, G.G., and Mégier, J. (1991). 'The Use of Expert System and Supervised Relaxation Techniques to Improve Spot Image Classification Using Spatial Context', IGARSS '91, **3**, 1855-1858.

Landgrebe, D.A., (1980). 'The Development of a Spectral-Spatial Classifier for Earth Observational Data', *Pattern Recognition*, **12**, 165-175.

Levine, M.D. (1985). Vision, Man and Machine, McGraw Hill.

Levine, M.D., and Nazif, A.M., (1984). 'An Optimal set of Image Segmentation Rules', *Pattern Recognition Letters*, 1984, 243-248.

Lillesand, T.M., and Kiefer, R.W., (1994). Remote Sensing and Image Interpretation, John Wiley and Sons.

Ma, Z., and Redmond, R., (1995). 'Tau coefficients for accuracy assessment of classification of remote sensing data', *Photogrammetric Engineering and Remote Sensing*, **61**, 435-439.

Maksimovic, C., (1993). 'Postgraduate Training Course No.1: New Technologies in Urban Drainage - Design, Simulation and Sanitary Aspects Management', in Hydrology and Hydraulics of Urban Drainage Systems, 1-12 March, Sofia, Bulgaria.

Maksimovic,C., Elgy, J., Fuchs, L., Prodanovic, D. and Djordjevic, S., (1994). 'Gluing Routines for Matching Standard GIS Packages with Simulation and Design Models for Water Projects', in Remote Sensing and GIS in Urban Waters (UDT '94), p91 – 104.

Maksimovic, C., Rajcevic, A., Djordjevic, S., Prodanovic, D., and Draskovic, M. (1995). 'Results of Simulation with updated data and modified BEMUS Model', in F.Calomino, C.Maksomovic and B, Molino (Eds), Urban Drainage Experimental Catchments in Italy, Editoriale Bios, Cosenza, 263-276.

Mark, O., Van Kalken, T., Rabbi, K., and Kjelds, J., (1997). 'A MOUSE GIS Study of the drainage in Dhaka City', *Technical Report,* Danish Hydraulic Institute, Denmark.

Marr, D., and Hildreth, E.,(1980). 'Theory of Edge Detection', *Proceedings Royal Society of London B.,* **207**, 187-217.

Marr, D. (1982). Vision: a computational investigation into the human representation and processing of visual information, W.H. Freeman, San Francisco, CA, USA.

Matsuyama, T., (1987). 'Knowledge-Based Aerial Image Understanding Systems and Expert Systems for Image Processing', *IEEE Transactions on Geoscience and Remote Sensing,* **GE-25**, No. 3, May 1987, 305-316.

Matsuyama, T., (1989). 'Expert Systems for Image Processing: Knowledge-Based Composition of Image Analysis Processes', *Comp. Vision, Graphics and Image Processing,* **48**, 1989, 22-49.

Matsuyama,T., and Hwang, V.S.S.,(1990). SIGMA - A Knowledge-Based Aerial Image Understanding System, Plenum Press, New York and London, 277p.

McCauley, J.D., and Engel, B.A., (1995), 'Comparison of Scene Segmentations: SMAP, ECHO, and Maximum Likelihood', *IEEE Transactions on Geoscience and Remote Sensing,* **33**, No. 6, 1313-1316.

McKeown, D.M., Cochran, S.D., and Ford, S.J., (1994). 'Research in the Automated Analysis of Remotely Sensed Imagery: 1993-1994', *Proceedings of the ARPA Image Understanding Workshop*, November 14-18, Monterey CA.

McKeown, D.M., Harvey, W.A., and McDermott, J., (1985). 'Rule-Based Interpretation of Aerial Imagery', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-7**, No. 5, 570-585, September 1985.

McKeown, D.M., Harvey, W.A., and Wilson, L.E., (1989). 'Automating Knowledge Acquisition for Aerial Image Interpretation', *Computer Vision, Graphics, and Image Processing*, **46**, 37-81, 1989.

Nagao, M., and Matsuyama, T., (1980). A Structural Analysis of Complex Aerial Photographs, Plenum Press, New York, 199p.

Nazif, A.M. and Levine, M.D., (1984). 'Low Level Image Segmentation: An Expert System', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**, No. 5, September 1984, 555-577.

Nicolin, B., and Gabler, R. (1987). 'A Knowledge-Based System for the Analysis of Aerial Images', *IEEE Transactions on Geoscience and Remote Sensing*, **GE-25**, No. 3, May 1987, 317-329.

Oldfield, R.B., (1988). Lithological Mapping of Northwest Argentina with Remote Sensing Data using Tonal, Textural and Contextual Features, Ph.D. Thesis, University of Aston in Birmingham, Birmingham, UK.

Ordnance Survey, (1997). 'Products and Services Catalogue, 1997, for the Business and Professional User' *Available from* Ordinance Survey, Romsey Road, Southampton, UK, SO16 4GU Tel: 0345 33 00 11

Prewitt, J.M.S. (1970). 'Object Enhancement and Extraction', in Picture Processing and Psychopictorics, Ed Lipkin, B.S., Academic Press.

Richards, J.A., (1986). Remote Sensing Digital Image Analysis, Springer-Verlag.

Roberts, L.G., (1965). 'Machine Perception of Three Dimensional Solids', in J.T. Tippet (Ed), Optical and Electro-Optical Information Processing, 159-197, MIT Press.

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.

Scott, A., (1994). 'Low-Cost Remote Sensing Techniques Applied to Drainage Area Studies', *Journal of the Institute of Water and Environmental Management*, **8**, 497-501.

Shapiro, M., Westervelt, J., Gerdes, D., Larson, M., and Brownfield, K.R., (1993). GRASS 4.1 Programmers Manual, US Army Research Construction Engineering Research Laboratory.

Simpson, P.K., (1990). Artificial Neural Systems, Pergamon Press.

Smith, S.M. and Brady, J.M., (1995). 'SUSAN - A New Approach to Low Level Image Processing', DRA Technical Report TR95SMS1c.

Sobel, I, (1970). 'Camera Models and Machine Perception', Technical Report AIM-21, Stanford University Artificial Intelligence Lab, Stanford, California.

Story, M., and Congalton, R.G., (1986). 'Accuracy Assessment: A User's Perspective', *Photogrammetric Engineering and Remote Sensing*, **52**, No.3, March, 397-399

Strand, J., and Taxt, T., (1994). 'Local Frequency Features for Texture Classification', *Pattern Recognition*, **27**, No. 10, 1994, 1397-1406.

Swain, P.H. and Davis, S.M., (1978). Remote Sensing: The Quantitative Approach, McGraw Hill, New York.

Tailor, A., Cross, A., Hogg, D.C., and Mason, D.C., (1986). 'Knowledge-Based Interpretation of Remotely Sensed Images', *Image and Vision Computing*, **4**, No. 2, May 1986, 67-83.

Taxt, T., Flynn, P.J., and Jain, A.K., (1989). 'Segmentation of Document Images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-11**, No. 12, Dec 1989, 1322-1329.

Thomas, I.L., Benning, V.M., and Ching, N.P., (1987). Classification of Remotely Sensed Images, Adam Hilger, Bristol.

Todd, W.J, (1980). 'Landsat Wildland Mapping Accuracy', *Photogrammetric Engineering and Remote Sensing*, **46**, 509 - 520.

Wallace, R.S., (1985). 'A Modified Hough Transform for Lines', *Computer Vision and Pattern Recognition*, 665 - 667.

Wallingford Software, (1991). WALLRUS User Manual 4th Edition, Wallingford Software, Hydraulics research Ltd, Wallingford, Oxon, UK.

Wilson, A.K., (1995). NERC Scientific Series Airborne Remote Sensing Facility User Guide Handbook, CSS ARF Facility Handbook Version 1.0 July 1995, National Environmental Research Council.

Wilson, A.K., (1994). 'The NERC Integrated ATM/CAS/GPS System', *Proceedings of the 1st International Airborne Remote Sensing Conference and Exhibition,* **II**, 12-15 September, Strasbourg, France, 249-259.

Woodcock, C., and Strahler, A., (1987). 'The factor of scale in remote sensing', *Remote Sensing of the Environment*, **21**, 311-332.

Yuille, A.L., and Poggio, T.A., (1986). 'Scaling Theorems for Zero Crossings', *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **PAMI-8,** No. 1, 15-25, January 1986.

Zhang,W., Albertz, J., and Li, Z., (1994). 'Rectification of Airborne Line-Scanner Imagery Utilising Flight Parameters', *Proc. 1st Int. Airborne Remote Sensing Conference and Exhibition,* **II**, 12-15 September, Strasbourg, France, 447-456.

Appendix 1 – C Listing for the Polygon Classifier

This appendix contains the code for the various computer programs used in this project. The code was written in C and compiled in the Grass4.1 environment. The code will require minor modifications to run outside of this environment.

The code presented in this appendix is discussed in section 5.5. It was written for ease of implementation, rather than efficiency, and as a result has not been optimised. However, the reason for this is due to the memory allocation for each polygon that is extracted. Most of the polygons in the polygon map image are spatially small, although a few (such as a road network) have a large spatial extent. The current program uses a static array for each extracted polygon, rather than a dynamically allocated one, the size of which is enough to hold the entire image. Consequently, each polygon takes the same amount of time to process. Implementing dynamic memory allocation will vastly improve the performance of the classifier.

List of Files :

`Polygon.c` - Main program.
`Polygon.h` - Header file.
`Gmakefile` - Makefile for compilation in a Grass4.1 environment.

Gmakefile

```
OBJ= seed2.o

EXTRA_CFLAGS=$(VECT_INCLUDE)


r.seed2:$(OBJ) $(GISLIB)
$(CC) $(LDFLAGS) -o $@ $(OBJ) $(GISLIB) $(VECTLIB) -I$(INCLUDE_DIR) -
I/usr/include $(EXTRA_CFLAGS) $(VECTLIBFLAGS) $(XDRLIB) $(MATHLIB)


$(GISLIB):#in case library changes
seed2.h:
$(VECTLIB):
```

Polygon.h

```
/* Header file for r.polygon */


#include <stdio.h>
#include <gis.h>


#define MATRIX_SIZE 512

#define A image[row-1][col-1]
#define B image[row-1][col]
#define C image[row-1][col+1]

#define D image[row][col-1]
#define E image[row][col]
#define F image[row][col+1]

#define G image[row+1][col-1]
#define H image[row+1][col]
#define I image[row+1][col+1]

#define Ab1  buf1[row-1][col-1]
#define Bb1  buf1[row-1][col]
#define Cb1  buf1[row-1][col+1]

#define Db1 buf1[row][col-1]
#define Eb1 buf1[row][col]
#define Fb1 buf1[row][col+1]

#define Gb1 buf1[row+1][col-1]
#define Hb1 buf1[row+1][col]
#define Ib1 buf1[row+1][col+1]


int nrows;
int ncols;
int file1_d;
int file2_d;
int file3_d;

int filed_new;

CELL *cell1;
CELL *cell2;
CELL *cell3;

CELL image[MATRIX_SIZE][MATRIX_SIZE];

CELL buf1[MATRIX_SIZE][MATRIX_SIZE];
CELL buf2[MATRIX_SIZE][MATRIX_SIZE];
CELL buf3[MATRIX_SIZE][MATRIX_SIZE];
CELL buf4[MATRIX_SIZE][MATRIX_SIZE];
CELL maxlik[MATRIX_SIZE][MATRIX_SIZE];
```

```
int row;
int col;
```

Polygon.c

```
/* r.polygon  15.1.97 By Alastair Buchanan

This program classifies an image. A polygon map of the areas generated
by the zero-crossing edge detection program is used as a mask. An area
is selected and extracted. A seed is then planted and grown from the
edge of the area, the criteria for growth being that the neighbouring
pixels fall within a threshold of homogeneity. Once a sub-region has
been grown, the underlying classification statistics determined by a
Maximum Likelihood classifier is examined. The mode value for the sub-
region is taken as the true class for that region. The next seed in the
extracted area is then grown and further sub areas are extracted and
classified. */


#include "polygon.h"
#include <time.h>



#define CLASS 500
#define ITERATIONS 400
#define THRESHOLD 10
#define number_of_classes 20
#define HOWMANY 100000

struct Option* input1;
struct Option* input2;
struct Option* input3;
struct Option* output;
struct Flag* quiet;
int x;
int SEEDROW;
int SEEDCOL;
int seed1;
long int n;
int find_mode();
int findseed();
int grow();
int condition();
int flag=0;
int grow_region();
int i;
int mode=0;
long int time1,time2;
int endflag=0;
int area_count=0;
int a;

main(argc,argv)
```

```
int argc;
char *argv[];

{
char file1[50];
char file2[50];
char file3[50];
char filter_new[50];
char *mapset;

G_gisinit(argv[0]);
time1=clock();


/***************Call Parser and set it up ***********************/


  input1=G_define_option();
  input2=G_define_option();
  input3=G_define_option();


  output=G_define_option();
  quiet=G_define_flag();

  input1->key="input1";
  input1->description="input thermal image";
  input1->required=YES;
  input1->type=TYPE_STRING;
  input1->gisprompt=("old,cell,raster");

  input2->key="input2";
  input2->description="input statistics map";
  input2->required=YES;
  input2->type=TYPE_STRING;
  input2->gisprompt=("old,cell,raster");

  input3->key="input3";
  input3->description="input polyon map";
  input3->required=YES;
  input3->type=TYPE_STRING;
  input3->gisprompt=("old,cell,raster");

  output->key="output";
  output->description="output filefilter";
  output->required=YES;
  output->type=TYPE_STRING;

  quiet->key='q';
  quiet->description="suppress verbosity?";


if(G_parser(argc,argv))
    {
    G_fatal_error("parser choked");
    }
```

```
/*****************The rest of the program*********************/


if (!quiet->answer){fprintf(stderr,"r.seed - december 1996 \n");}

  file1[0]='\0';
  G_strcat(file1,input1->answer);

  if((file1_d=G_open_cell_old(input1->answer,G_mapset()))<0)
    {
    G_fatal_error("cannot read file1");
    }

  file2[0]='\0';
  G_strcat(file2,input2->answer);

  if((file2_d=G_open_cell_old(input2->answer,G_mapset()))<0)
    {
    G_fatal_error("cannot read file2");
    }
  file3[0]='\0';
  G_strcat(file3,input3->answer);

  if((file3_d=G_open_cell_old(input3->answer,G_mapset()))<0)
    {
    G_fatal_error("cannot read file3");
    }


filter_new[0]='\0';
G_strcat(filter_new,output->answer);



nrows=G_window_rows();
ncols=G_window_cols();

printf("rows=%d cols=%d\n",nrows,ncols);


filed_new=G_open_cell_new (filter_new);



/* Open matrix for buffer */

cell1=G_allocate_cell_buf();


for (row=0; row<(nrows); row++)
{
G_get_map_row(file1_d, cell1,row);
```

```
for(col=0; col<(ncols); col++)
{
  image[row][col]=cell1[col];
}
cell2=G_allocate_cell_buf();
for (row=0; row<(nrows); row++)
{
G_get_map_row(file2_d, cell2,row);


for(col=0; col<(ncols); col++)
{

  maxlik[row][col]=cell2[col];



}
}
cell3=G_allocate_cell_buf();
for (row=0; row<(nrows); row++)
{
G_get_map_row(file3_d, cell3,row);


for(col=0; col<(ncols); col++)
{

buf4[row][col]=cell3[col];

}
}


/* find areas to work on */
/* How many areas are there? */

for (row=0; row<nrows; row++)
  {
  for (col=0; col<ncols; col++)
    {
    if (buf4[row][col]>area_count)
       {
       area_count=buf4[row][col];
       }
    }
  }

printf("\nNumber of Areas=%d", area_count);
/* create buffer containing band11 values */


for (a=1; a<=area_count;a++)
{
for (row=0; row<nrows; row++)
  {
  for (col=0; col<ncols; col++)
```

```
    {
       {
       buf1[row][col]=0;
       }
    }
  }

for (row=0; row<nrows; row++)
  {
  for (col=0; col<ncols; col++)
    {
    if (buf4[row][col]==a)
       {
       buf1[row][col]=image[row][col];
       }
    }
  }
printf("\nProcessing area %d ",a);
endflag=0;




/* find a starting point */

for (n=1;n<HOWMANY;n++)
{


if (findseed()==1)
{

row=SEEDROW;
col=SEEDCOL;

/*printf("at %d %d file1: Seed value = %d\n",row,col,E);*/
seed1 = E;


Eb1 = CLASS;


grow_region();

/*perform region growing on the polygon.*/

for (i=1; i<ITERATIONS; i++)
{
flag=0;
/*A  Dynamically  allocated  array  can  be  used  here  to  improve
efficiency*/
  for (row=0;row<=nrows;row++)
    {
    for(col=0; col<(ncols); col++)
      {
```

```
      if (Eb1==CLASS)
        {
        seed1=E;

        grow_region();



        }
      }
    }

if (flag==0)
  {
  i=(ITERATIONS-1);
  /*printf("Bailout\n");*/
  }
}


/* assign maximum likelihood values to extracted region */


for (row=0; row<nrows; row++)
{
for(col=0; col<ncols; col++)
{
         if (buf1[row][col]==CLASS)
           {
            buf2[row][col]=maxlik[row][col];
           }
        }
}
/* find mode value */
mode=find_mode();


/* assign mode value to extracted region */


for (row=0; row<nrows; row++)
{
for(col=0; col<ncols; col++)
{
         if (buf1[row][col]==CLASS)
           {
            buf3[row][col]=mode;
           }
        }
}



/* produce leftover image for next loop and clean buffers*/
```

```
for (row=0; row<nrows; row++)
{
for(col=0; col<ncols; col++)
{
          if (buf1[row][col]==CLASS)
            {
             buf1[row][col]=0;
            }

          buf2[row][col]=0;
         }
}




}
else
 {
  time2=clock();
  printf("\n%i seconds",(time2-time1)/CLOCKS_PER_SEC);
  endflag=1;
 }
if (endflag==1)
 {
  break;
 }
}

}



/* Write new image to new matrix for output */

for (row=0; row<nrows; row++)
{


for(col=0; col<ncols; col++)
{

cell1[col]=buf3[row][col];


}
G_put_map_row(filed_new,cell1);
}

/* Close files .... finished*/


if (!quiet->answer){fprintf(stderr,".....done. ");}

G_close_cell(file1_d);
G_close_cell(file2_d);
G_close_cell(file3_d);
```

```
G_close_cell(filed_new);
time2=clock();
printf("\n%d    Areas    processed    in    %i    seconds",a,(time2-
time1)/CLOCKS_PER_SEC);
exit (1);



}


grow_region()
{


 if ((Ab1>=(seed1-THRESHOLD))&&(Ab1<=(seed1+THRESHOLD)))
                                        {if (A!=CLASS){A=CLASS;flag=1;}}
    if ((Bb1>=(seed1-THRESHOLD))&&(Bb1<=(seed1+THRESHOLD)))
                                        {if (Bb1!=CLASS){Bb1=CLASS;flag=1;}}
    if ((Cb1>=(seed1-THRESHOLD))&&(Cb1<=(seed1+THRESHOLD)))
                                        {if (Cb1!=CLASS){Cb1=CLASS;flag=1;}}
    if ((Db1>=(seed1-THRESHOLD))&&(Db1<=(seed1+THRESHOLD)))
                                        {if (Db1!=CLASS){Db1=CLASS;flag=1;}}
    if ((Fb1>=(seed1-THRESHOLD))&&(Fb1<=(seed1+THRESHOLD)))
                                        {if (Fb1!=CLASS){Fb1=CLASS;flag=1;}}
    if ((Gb1>=(seed1-THRESHOLD))&&(Gb1<=(seed1+THRESHOLD)))
                                        {if (Gb1!=CLASS){Gb1=CLASS;flag=1;}}
    if ((Hb1>=(seed1-THRESHOLD))&&(Hb1<=(seed1+THRESHOLD)))
                                        {if (Hb1!=CLASS){Hb1=CLASS;flag=1;}}
    if ((Ib1>=(seed1-THRESHOLD))&&(Ib1<=(seed1+THRESHOLD)))
                                        {if (Ib1!=CLASS){Ib1=CLASS;flag=1;}}


if (flag==1) {return 1;}
else {return 0;}

}

findseed()
{

for (row=0;row<=nrows;row++)
  {
  for (col=0; col<=ncols; col++)
    {
    if (Eb1!=0)
      {
       SEEDROW=row;
       SEEDCOL=col;
       return 1;
      }
    }

  }
}
```

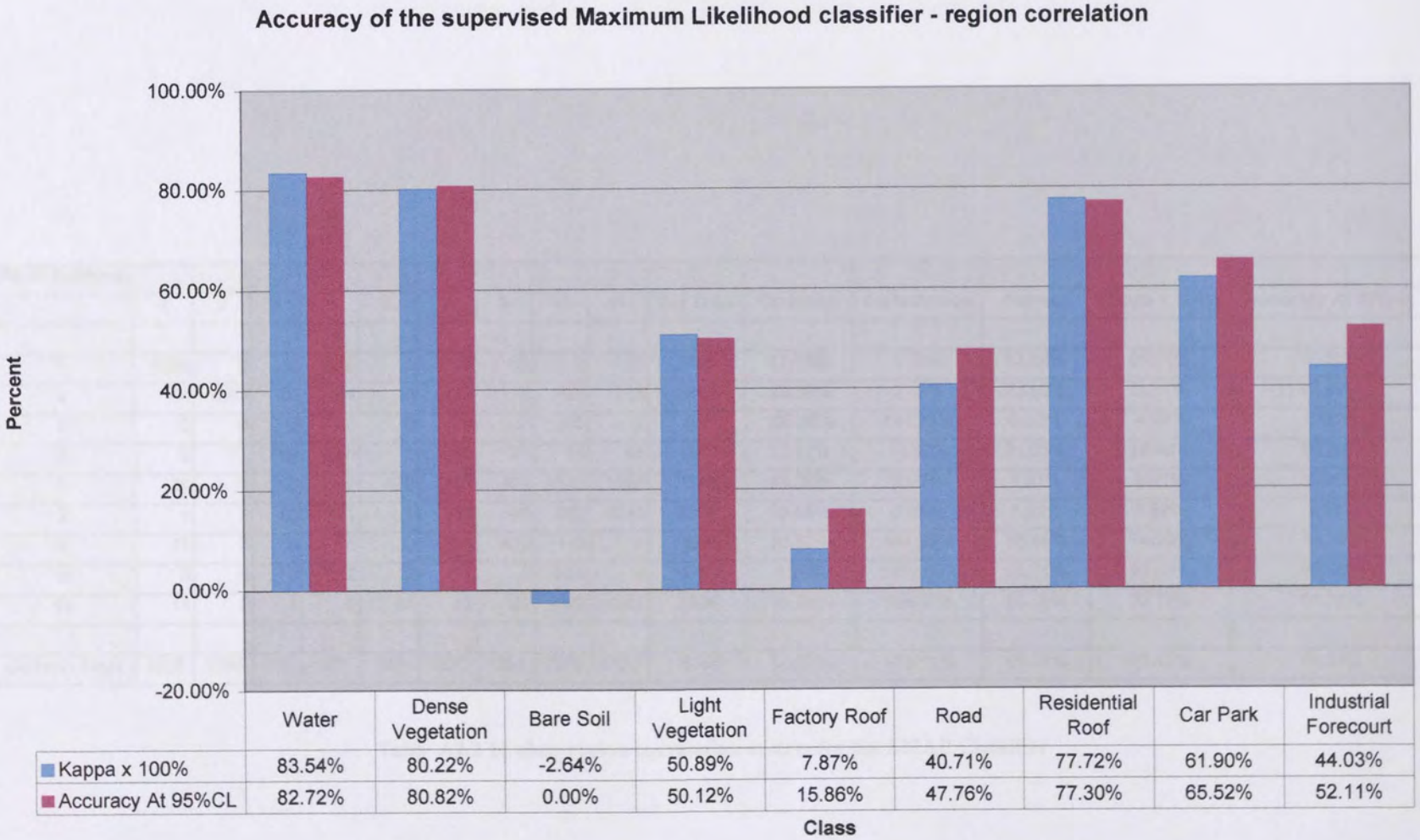
```
find_mode()
{
int area[number_of_classes];
int index=1,count=0;
int class=0;
int biggest=0;

 for (index=1; index<number_of_classes; index++)
   {
   for (row=0; row<nrows;row++)
     {
     for (col=0;col<ncols;col++)
       {
       if (buf2[row][col]==index)
         {
          count++;
         }
       }
     }

area[index]=count;

/*printf("area %d = %d\n",index,area[index]);*/
count=0;
}

biggest=0;

for (index=1; index<number_of_classes; index++)
{

if (area[index]>biggest)
  {
  biggest=area[index];
  class=index;
  }
}

/*printf("biggest= class %d with %d pixels\n",class,biggest);*/
return (class);

}
```

Appendix 2 – Class accuracies for the 11 class imagery

This appendix contains the confusion matrices for the assessment of the 11 class classified imagery region correlation method. The results for each of the 4 classifiers examined in Chapters 3 and 4 are presented along with the results for the newly developed Polygon classifier from Chapter 5.

Supervised Maximum Likelihood Classification															
	1	2	3	5	6	7	8	10	11	Col Total	Omission	Commission	Correct	Kappa x 100%	Accuracy At 95%CL
1	1048	11	0	1	52	0	109	12	7	1240	15.48%	10.16%	84.52%	83.54%	82.72%
2	19	1805	8	37	129	2	126	69	0	2195	17.77%	9.70%	82.23%	80.22%	80.82%
3	0	1	0	0	127	0	3	61	0	192	100.00%	266.15%	0.00%	-2.64%	0.00%
5	0	184	244	622	3	35	57	5	31	1181	47.33%	8.13%	52.67%	50.89%	50.12%
6	25	0	251	9	1204	452	1339	3185	788	7253	83.40%	9.36%	16.60%	7.87%	15.86%
7	2	13	3	44	56	2015	158	808	1006	4105	50.91%	19.27%	49.09%	40.71%	47.76%
8	56	1	0	0	13	1	438	36	0	545	19.63%	352.66%	80.37%	77.72%	77.30%
10	0	0	0	3	21	11	11	150	14	210	28.57%	2294.76%	71.43%	61.90%	65.52%
11	24	3	5	2	278	290	119	643	1581	2945	46.32%	62.68%	53.68%	44.03%	52.11%
Column Total	1174	2018	511	718	1883	2806	2360	4969	3427	19866	45.49%	336.99%	54.51%	**49.36%**	**52.47%**

Table A2.1 11 class region correlation matrix for the Supervised Maximum Likelihood Classifier.

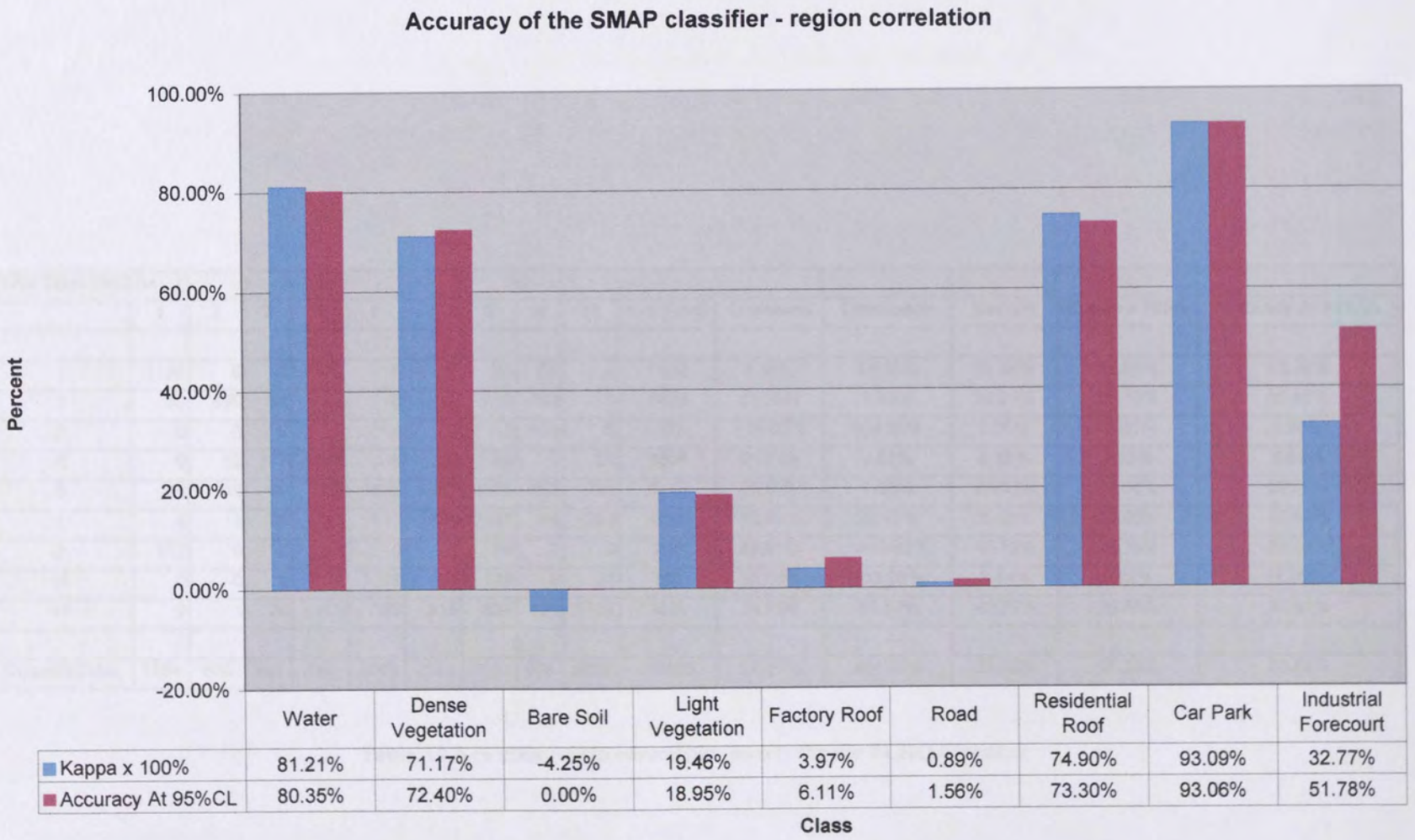


	Water	Dense Vegetation	Bare Soil	Light Vegetation	Factory Roof	Road	Residential Roof	Car Park	Industrial Forecourt
Kappa x 100%	83.54%	80.22%	-2.64%	50.89%	7.87%	40.71%	77.72%	61.90%	44.03%
Accuracy At 95%CL	82.72%	80.82%	0.00%	50.12%	15.86%	47.76%	77.30%	65.52%	52.11%

Figure A2.1 Accuracy of the Supervised Maximum Likelihood Classification (11 Classes) Region Correlation method

SMAP Estimator															
	1	2	3	5	6	7	8	10	11	Col Total	Omission	Commission	Correct	Kappa x 100%	Accuracy At 95%CL
1	1024	2	0	0	1	0	66	2	150	1245	17.75%	5.70%	82.25%	81.21%	80.35%
2	14	1820	42	134	0	1	115	123	213	2462	26.08%	3.17%	73.92%	71.17%	72.40%
3	0	0	0	0	0	0	23	168	1	192	100.00%	421.88%	0.00%	-4.25%	0.00%
5	0	71	719	250	1	9	31	14	94	1189	78.97%	11.52%	21.03%	19.46%	18.95%
6	34	0	8	0	479	103	424	5142	1059	7249	93.39%	0.92%	6.61%	3.97%	6.11%
7	1	2	40	1	1	81	87	762	3244	4219	98.08%	2.99%	1.92%	0.89%	1.56%
8	11	3	0	0	0	0	439	117	4	574	23.52%	141.99%	76.48%	74.90%	73.30%
10	0	0	0	2	0	1	2	201	4	210	4.29%	3499.05%	95.71%	93.09%	93.06%
11	11	0	1	0	64	12	67	1020	1351	2526	46.52%	188.80%	53.48%	32.77%	51.78%
Column Total	1095	1898	810	387	546	207	1254	7549	6120	19866	54.29%	475.11%	45.71%	**41.47%**	**44.17%**

Table A2.2 11 class region correlation matrix for the SMAP Classifier

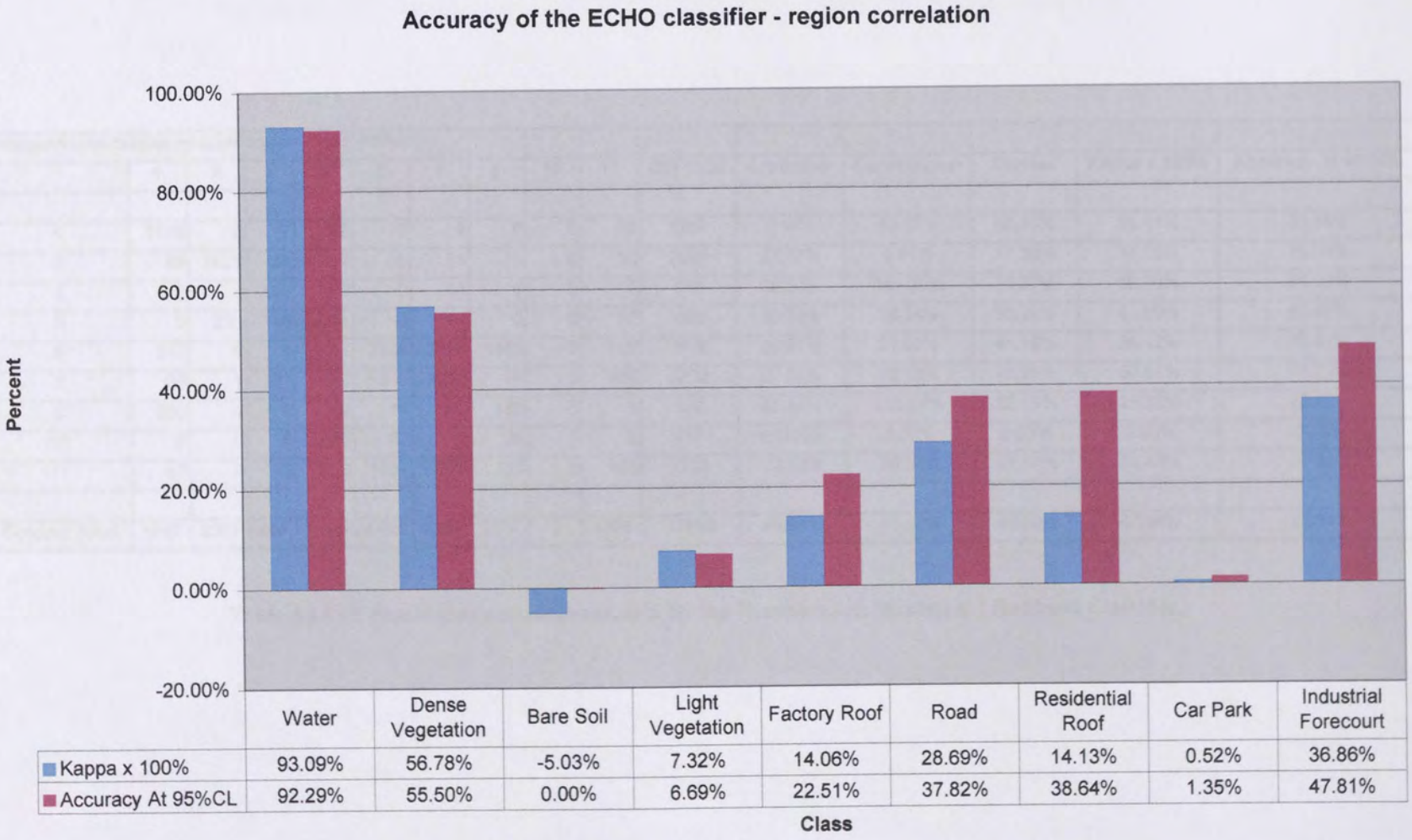


	Water	Dense Vegetation	Bare Soil	Light Vegetation	Factory Roof	Road	Residential Roof	Car Park	Industrial Forecourt
Kappa x 100%	81.21%	71.17%	-4.25%	19.46%	3.97%	0.89%	74.90%	93.09%	32.77%
Accuracy At 95%CL	80.35%	72.40%	0.00%	18.95%	6.11%	1.56%	73.30%	93.06%	51.78%

Figure A2.2 Accuracy of the SMAP Classifier (11 Classes). Region Correlation method

Echo Classification															
	1	2	3	5	6	7	8	10	11	Col Total	Omission	Comission	Correct	Kappa x 100%	Accuracy At 95%CL
1	1115	0	0	0	3	0	21	52	1	1192	6.46%	14.18%	93.54%	93.09%	92.29%
2	30	599	51	99	0	4	213	32	1	1029	41.79%	5.83%	58.21%	56.78%	55.50%
3	0	0	0	0	4	8	12	166	0	190	100.00%	500.53%	0.00%	-5.03%	0.00%
5	0	12	777	85	5	38	99	1	21	1038	91.81%	9.83%	8.19%	7.32%	6.69%
6	14	46	62	0	1806	388	4383	253	790	7742	76.67%	4.35%	23.33%	14.06%	22.51%
7	4	1	35	2	117	1801	1357	44	1252	4613	60.96%	23.46%	39.04%	28.69%	37.82%
8	112	0	0	0	0	0	94	2	0	208	54.81%	3409.62%	45.19%	14.13%	38.64%
10	0	0	0	1	19	28	120	8	27	203	96.06%	333.00%	3.94%	0.52%	1.35%
11	9	1	26	0	189	616	887	126	1797	3651	50.78%	57.30%	49.22%	36.86%	47.81%
Column Total	1284	659	951	187	2143	2883	7186	684	3889	19866	64.37%	484.23%	35.63%	**27.38%**	**33.62%**

Table A2.3 11 class region correlation matrix for the ECHO Classifier

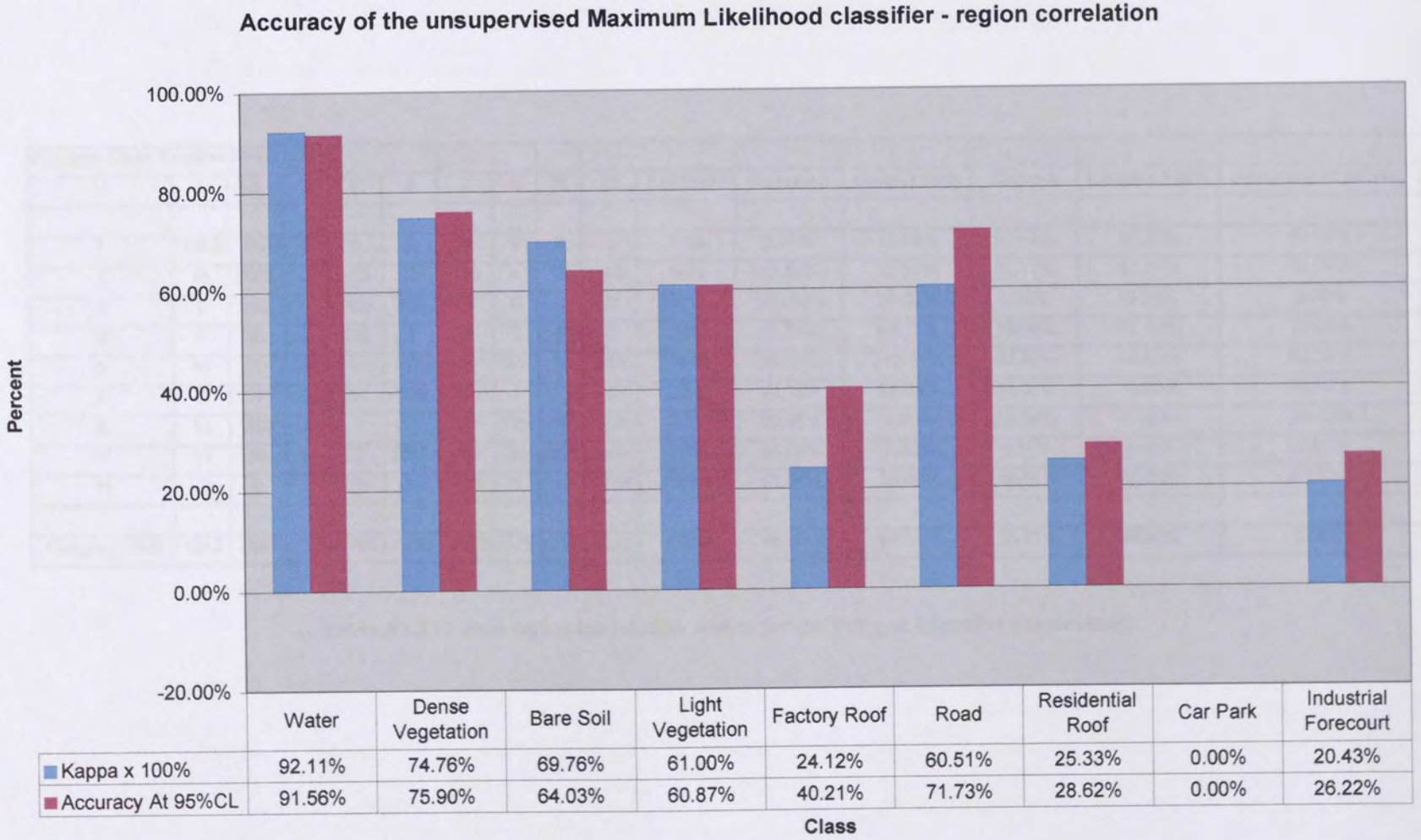


	Water	Dense Vegetation	Bare Soil	Light Vegetation	Factory Roof	Road	Residential Roof	Car Park	Industrial Forecourt
Kappa x 100%	93.09%	56.78%	-5.03%	7.32%	14.06%	28.69%	14.13%	0.52%	36.86%
Accuracy At 95%CL	92.29%	55.50%	0.00%	6.69%	22.51%	37.82%	38.64%	1.35%	47.81%

Figure A2.3 Accuracy of the ECHO Classifier (11 Classes) Region Correlation method

Unsupervised Maximum Likelihood Classification															
	1	2	3	5	6	7	8	10	11	Col Total	Omission	Commission	Correct	Kappa x 100%	Accuracy At 95%CL
1	1118	2	7	12	0	9	0	0	56	1204	7.14%	63.87%	92.86%	92.11%	91.56%
2	89	1827	38	354	0	31	10	0	12	2361	22.62%	9.91%	77.38%	74.76%	75.90%
3	0	0	133	2	0	46	1	0	7	189	29.63%	143.39%	70.37%	69.76%	64.03%
5	0	231	116	751	6	63	0	0	19	1186	36.68%	36.34%	63.32%	61.00%	60.87%
6	317	0	1	7	2956	2126	1459	0	310	7176	58.81%	21.08%	41.19%	24.12%	40.21%
7	36	1	91	50	341	2800	14	0	505	3838	27.05%	90.18%	72.95%	60.51%	71.73%
8	260	0	0	0	14	102	182	0	8	566	67.84%	288.87%	32.16%	25.33%	28.62%
10	0	0	6	3	98	79	21	0	3	210	100.00%	0.00%	0.00%	0.00%	0.00%
11	67	0	12	3	1054	1005	130	0	865	3136	72.42%	29.34%	27.58%	20.43%	26.22%
Column Total	1887	2061	404	1182	4469	6261	1817	0	1785	19866	46.91%	75.89%	53.09%	**47.56%**	**51.01%**

Table A2.4 11 class region correlation matrix for the Unsupervised Maximum Likelihood Classifier.

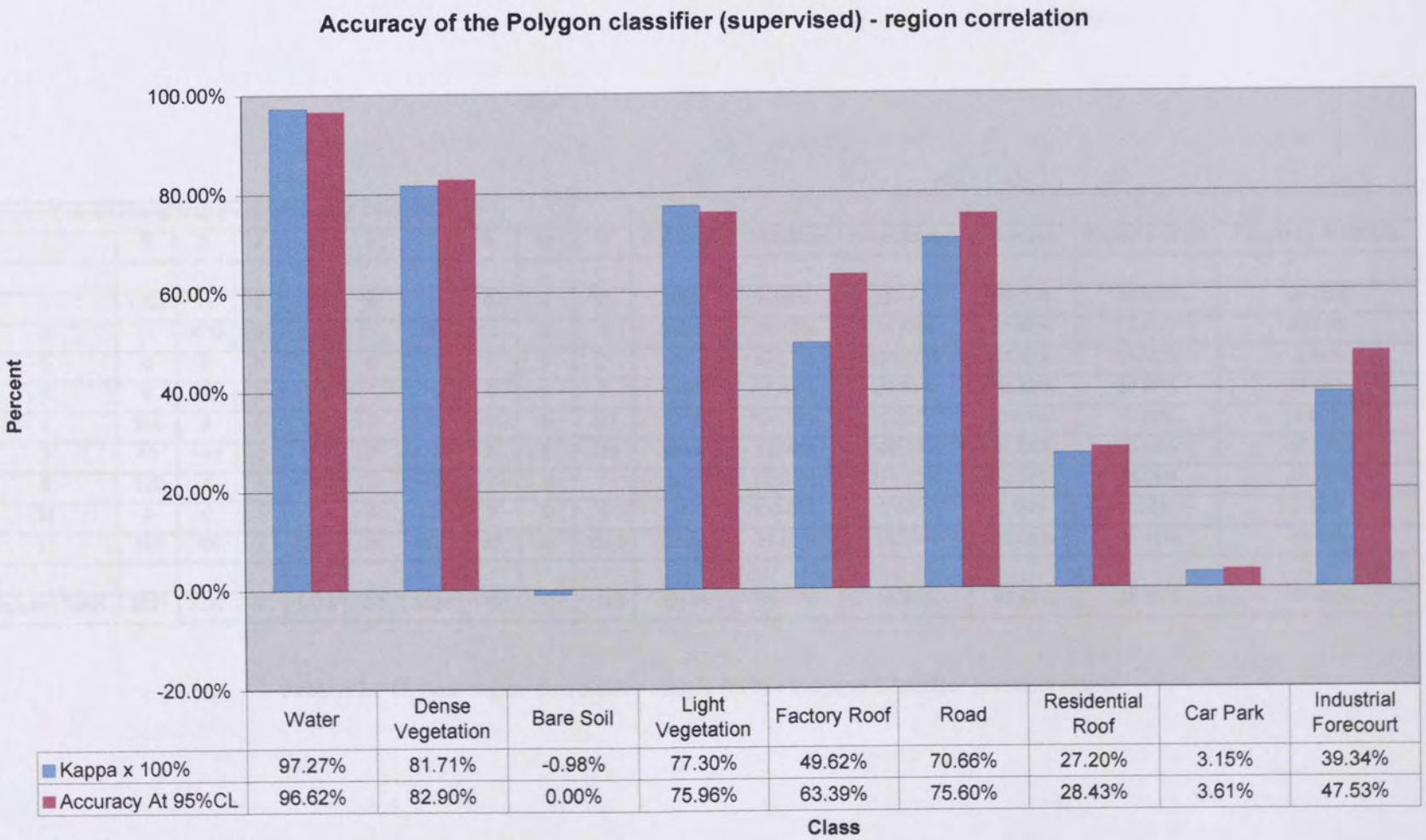


	Water	Dense Vegetation	Bare Soil	Light Vegetation	Factory Roof	Road	Residential Roof	Car Park	Industrial Forecourt
Kappa x 100%	92.11%	74.76%	69.76%	61.00%	24.12%	60.51%	25.33%	0.00%	20.43%
Accuracy At 95%CL	91.56%	75.90%	64.03%	60.87%	40.21%	71.73%	28.62%	0.00%	26.22%

Figure A2.4 Accuracy of the Unsupervised Maximum Likelihood Classifier (11 Classes) Region correlation method

Polygon Classification (Supervised)															
	1	2	3	5	6	7	8	10	11	Col Total	Omission	Commission	Correct	Kappa x 100%	Accuracy At 95%CL
1	1105	26	0	0	2	0	0	0	1	1134	2.56%	12.08%	97.44%	97.27%	96.62%
2	18	2089	0	103	3	156	4	0	109	2482	15.83%	23.37%	84.17%	81.71%	82.90%
3	0	109	0	431	319	17	0	0	0	876	100.00%	21.92%	0.00%	-0.98%	0.00%
5	0	86	0	575	1	66	0	3	0	731	21.34%	84.13%	78.66%	77.30%	75.96%
6	44	50	156	0	1077	76	11	13	219	1646	34.57%	313.37%	65.43%	49.62%	63.39%
7	0	19	0	54	426	2489	1	13	236	3238	23.13%	53.00%	76.87%	70.66%	75.60%
8	52	260	0	1	1172	97	729	8	108	2427	69.96%	1.90%	30.04%	27.20%	28.43%
10	14	30	36	0	2307	450	30	156	716	3739	95.83%	1.44%	4.17%	3.15%	3.61%
11	9	0	0	26	928	854	0	17	1759	3593	51.04%	38.66%	48.96%	39.34%	47.53%
Column Total	1242	2669	192	1190	6235	4205	775	210	3148	19866	46.03%	61.10%	53.97%	**49.47%**	**52.67%**

Table A2.5 11 class region correlation matrix for the Polygon Classifier (supervised).

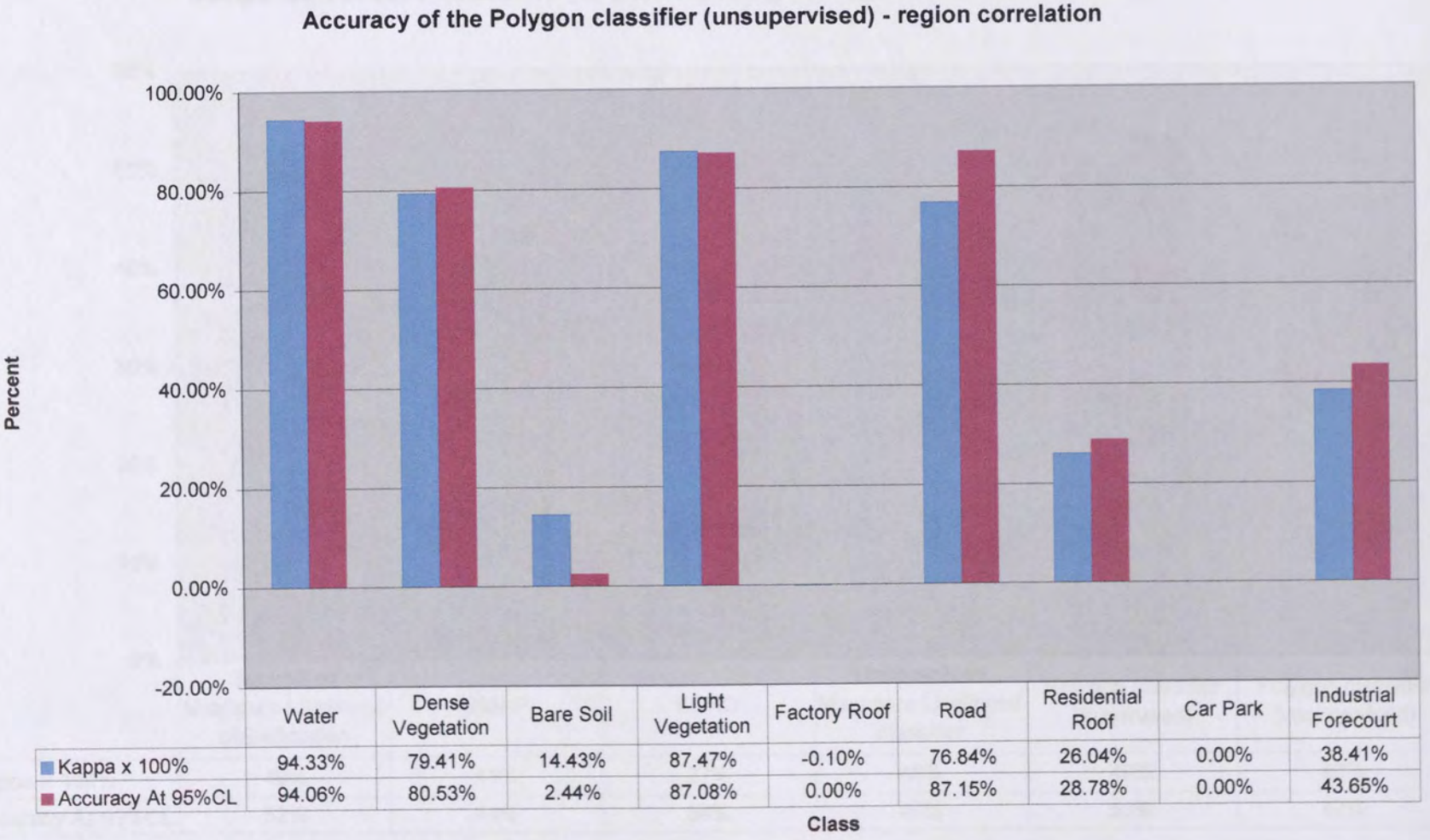


	Water	Dense Vegetation	Bare Soil	Light Vegetation	Factory Roof	Road	Residential Roof	Car Park	Industrial Forecourt
Kappa x 100%	97.27%	81.71%	-0.98%	77.30%	49.62%	70.66%	27.20%	3.15%	39.34%
Accuracy At 95%CL	96.62%	82.90%	0.00%	75.96%	63.39%	75.60%	28.43%	3.61%	47.53%

Figure A2.5 Accuracy of the Polygon Classifier (supervised) (11 Classes) Region Correlation method

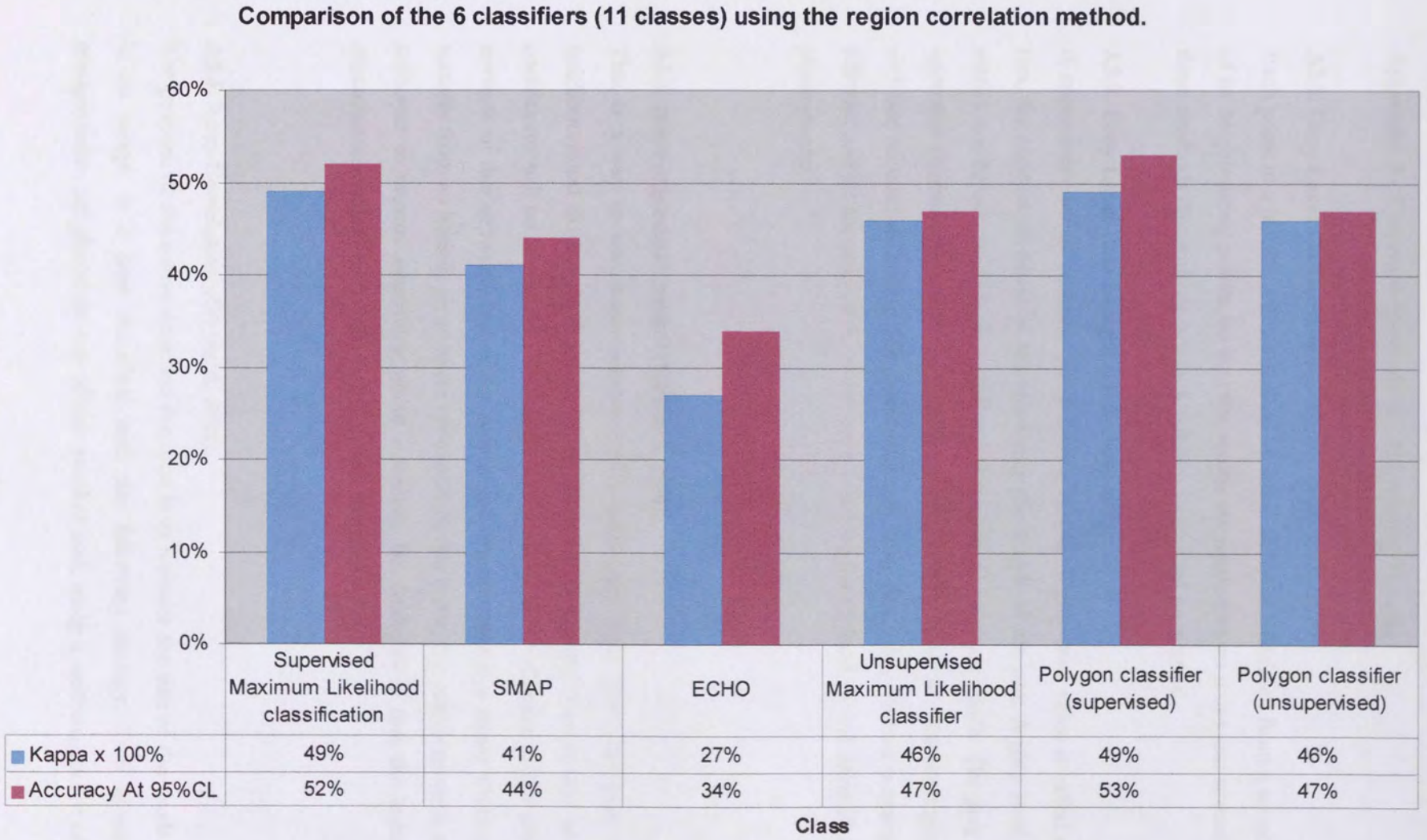
Polygon Unsupervised Maximum Likelihood Classification															
	1	2	3	5	6	7	8	10	11	Col Total	Omission	Commission	Correct	Kappa x 100%	Accuracy At 95%CL
1	1504	12	7	7	6	7	0	0	40	1583	4.99%	55.40%	95.01%	94.33%	94.06%
2	95	2018	9	233	1	90	11	0	8	2465	18.13%	14.28%	81.87%	79.41%	80.53%
3	0	0	6	0	0	31	1	0	3	41	85.37%	100.00%	14.63%	14.43%	2.44%
5	0	86	5	1319	0	75	0	0	5	1490	11.48%	23.83%	88.52%	87.47%	87.08%
6	314	3	0	0	0	3217	1472	0	211	5217	100.00%	0.38%	0.00%	-0.10%	0.00%
7	41	142	8	101	0	4259	3	0	289	4843	12.06%	108.71%	87.94%	76.84%	87.15%
8	322	4	1	0	5	42	179	0	0	553	67.63%	275.23%	32.37%	26.04%	28.78%
10	0	0	4	3	3	130	0	0	0	140	100.00%	0.00%	0.00%	0.00%	0.00%
11	105	105	7	11	5	1673	35	0	1593	3534	54.92%	15.73%	45.08%	38.41%	43.65%
Column Total	2381	2370	47	1674	20	9524	1701	0	2149	19866	50.51%	65.95%	49.49%	**46.31%**	**47.08%**

Table A2.6 11 class region correlation matrix for the Polygon Classifier (unsupervised).



	Water	Dense Vegetation	Bare Soil	Light Vegetation	Factory Roof	Road	Residential Roof	Car Park	Industrial Forecourt
Kappa x 100%	94.33%	79.41%	14.43%	87.47%	-0.10%	76.84%	26.04%	0.00%	38.41%
Accuracy At 95%CL	94.06%	80.53%	2.44%	87.08%	0.00%	87.15%	28.78%	0.00%	43.65%

Figure A2.6 Accuracy of the Polygon Classifier (unsupervised) (11 Classes) Region Correlation.



	Supervised Maximum Likelihood classification	SMAP	ECHO	Unsupervised Maximum Likelihood classifier	Polygon classifier (supervised)	Polygon classifier (unsupervised)
Kappa x 100%	49%	41%	27%	46%	49%	46%
Accuracy At 95%CL	52%	44%	34%	47%	53%	47%

Figure A2.7 A comparison of the performance of the classifiers with 11 categories (Region Correlation).

Appendix 3 - Common Methods for Measuring Texture

A3.1. Grey Level Co-Occurrence Matrix (Haralick, 1979).

Each point in a picture is characterised by some statistics calculated from a set of grey levels of its neighbouring points, so that the results are insensitive to noise and texture or, rather, these methods discriminate points according to textural properties.

A3.2. Grey Level Run Length (Galloway, 1975)

A consecutive set of co-linear pixels having the same grey-level value is called a grey level run, the number of pixels in the run being the length of the run. A grey level run length matrix can be calculated in four different directions for a given image. The grey run length approach characterises coarse textures as having as many pixels in a constant grey tone run, and fine textures as having few pixels in a grey tone run. This method is computationally efficient and an accuracy of 83% has been achieved for six land classes using digitised aerial photography.

A3.3. Auto-regression model (Oldfield, 1988).

This is a way to use linear estimates of a pixel grey tone given the grey tones in a neighbourhood containing it in order to characterise texture. For coarse textures, the coefficients will be similar. For fine textures, the coefficients will have wide variations. The strength of this approach is that it is easy to use the estimates in a mode which synthesises textures from an initially given linear estimator. In this sense, the auto-regressive approach is sufficient to capture everything about a texture. Its weakness is that the textures it can characterise are likely to consist mostly of micro textures.

A3.4. Auto-Correlation (Oldfield, 1988).

The purpose of the auto-correlation function is to measure the size of the texels contained in an image. It is best described with the following analogy. Two identical image transparencies are placed on top of one another and, using a uniform source of light, the

average light transmitted through the double transparency is measured. One transparency is then rotated relative to the other and transmitted light, as a function of the rotation, is measured. The two dimensional auto-correlation of the transparencies is the graph of these measurements. The auto-correlation and the power spectral density are Fourier transformations of one another. The auto-correlation will drop off slowly with distance if the tonal primitives are relatively large. If the tonal primitive are small the auto-correlation will drop off quickly.

Appendix 4 - An Overview of Contextual Algorithms

One method of representing context involves other pixels or groups of pixels in the classification of the image by using the neighbourhood dependency of class labels. The contextual algorithms described below use the spatial information provided by the surrounding pixels. Basic methods include mode filters and median filters, which can be used as contextual post classifiers, as simple smoothing filters removing isolated stray pixels.

A4.1. Minimal Area Replacement Filter (Davis and Peet, 1977).

A smoothing technique based on the size of the classified objects in the image. All connected regions are identified - any region smaller than a certain size is declassified.

A4.2. Unclassified Boundaries (Oldfield, 1988).

Related to the above idea is the theory that the largest errors of the classification occur at the edges or boundaries of the objects. A simple contextual rule is used to label all boundary areas as unclassified.

A4.3. Modified Minimum Area Replacement Filter (Oldfield, 1988).

A variation on the Minimal Area Replacement Filter is to identify noise and boundary pixels on original imagery. This is done by :

i) First passing a suitable edge detector over the image (i.e. Roberts/Sobel/Laplacian high pass).

ii) Thresholding this to separate the image into edges/non-edges.

iii) Declassifying the pixels identified as edges.

iv) Applying a reclassification algorithm to the declassified image (such as one of those already mentioned).

A4.4. Relaxation Labelling (Booth, 1989).

The image is reclassified according to the probabilities of a pixel and its neighbours belonging to each class. Probabilistic relaxation - an extension of the maximum likelihood classification algorithm to cover a neighbourhood rather than single pixels. However this is slow over large neighbourhoods - so iterative implementation must be done over small neighbourhoods.

A4.5. Non-probabilistic methods (Booth, 1989).

Modifying the minimum distance classifier in a similar manner to the above, but classifying according to the distance to means rather than probabilities.

Another alternative is to use ranked classes for each pixel, since rank order of classes can be calculated faster than the probabilities using an equation derived from the maximum likelihood probability estimate.

A more sophisticated two pass algorithm for the contextual classification of images has been developed (Haralick & Joo, 1985). The algorithm takes the form of a recursive neighbourhood operator first applied in a top down scan of the image and then in a bottom up scan of the image. This assigns a class to each pixel, based on the following properties. Associated with each pixel is a best path which passes through the pixel and generally flows from top to bottom (or left to right). The one dimensional context parameters for such paths are those estimated from the ground truth averages along the four types of paths, the horizontal, vertical, and the two diagonal directions. When applied to a Landsat image, the context classifier shows better classification accuracy compared to a non-contextual Bayes classifier with correct *a priori* values (80.0% compared to 68.8%).