



THE UNIVERSITY OF  
WESTERN AUSTRALIA  
*Achieving International Excellence*



## LAND COVER CLASSIFICATION OF HIGH RESOLUTION IMAGERY FOR ESTABLISHING INPUTS TO URBAN WATER BALANCE

RONALDI SOETANTO (10886864)



### SUPERVISORS:

Winthrop Professor Carolyn Oldham (UWA)

Assistant Professor Bryan Boruff (UWA)

Dr. Don McFarlane (CSIRO)

Dr. Peter Caccetta (CSIRO)

Dr. Olga Barron (CSIRO)

## EXECUTIVE SUMMARY

Understanding of water balance in the urban area has been limited mainly due to its complex features which consist of both natural and built environment. Study has shown that urban development has affected the hydrological processes and the water cycle in general. One method by which urban water balance can be calculated is by identifying the land cover of the area. Each land cover feature can be associated with hydrological processes which act as an input and output to water balance.

Understanding the urban water balance would help planners in improving urban design and water management in the long run.

Remote sensing is recognized as an efficient way of obtaining land cover information over a wide geographical area. A project known as Urban Monitor (UM) uses digital aerial photography to obtain high resolution imagery over the Perth region. The high resolution imagery offers an opportunity and challenge of identifying and classifying the heterogeneous land cover features of the urban area in detail.

This project used remote sensing analysis to classify urban land cover features of high resolution imagery. A decision rule tree was developed for the classification process using three variables namely spectral band values, Normalized Digital Surface Model (nDSM) and Normalized Difference Vegetation Index (NDVI). The classification was done for the study area which includes the Southern River catchment. Southern River catchment is a rapidly developing area which also retains its undisturbed vegetated area. Geology and vegetation map were used to refine the classification of the urban land cover. Total area and percentage of the land cover was obtained as the final result. The accuracy of the urban land cover classification was found to be 50%.

This project has shown that land cover classification of high resolution imagery is achievable particularly in the urban area. Although improvements could still be made to refine the classification accuracy, the project proved the potential value of obtaining land cover information from high resolution imagery. The study ultimately aims to support the developing of a decision support tool of accounting urban water balance.

## ACKNOWLEDGEMENTS

This report is submitted as partial fulfilment of the requirements of the degree Bachelor of Engineering (Environmental) at the University of Western Australia.

The author would like to acknowledge the Commonwealth Scientific and Industrial Research Organisation (CSIRO) who provided the data required for this project. The author would also like to thank the following people:

Winthrop Professor Carolyn Oldham, Supervisor, University of Western Australia

For overseeing the aspects of the project and supervising the execution and progress of the project

Assistant Professor Bryan Boruff, Supervisor, University of Western Australia

For giving supervision and guidance in land cover classification process and providing access to the image processing software ERDAS Imagine

Dr Don McFarlane, Co-supervisor, CSIRO

For providing guidance and background information regarding integrated urban water balances and water management

Dr Peter Caccetta, Co-supervisor, CSIRO

For giving guidance and knowledge in technical processing of the data

Dr Olga Barron, Co-supervisor, CSIRO

For providing knowledge and expertise in the hydrology of urban catchment

## LIST OF ABBREVIATIONS

CSIRO	Commonwealth Scientific and Industrial Research Organisation
DEM	Digital Elevation Model
DSM	Digital Surface Model
EC	Expert Classifier
ER	Electromagnetic Radiation
GEM	Ground Elevation Model
NDSM	Normalized Digital Surface Model
NDVI	Normalized Difference Vegetation Index
NVIS	National Vegetation Information System
UM	Urban Monitor

## LIST OF FIGURES

Figure 1 Conceptual Model of urban water balance (McFarlane 1984) .....	15
Figure 2 Comparison of spatial resolution .....	19
Figure 3 Concept of extracting nDSM from DSM and DEM(GEM) (Geospatial Information Authority of Japan 2010).....	22
Figure 4 Decision process of Expert Classification .....	24
Figure 5 Extent of the digital aerial imagery and the boundaries of the Southern River Catchment....	26
Figure 6 A subset of the nDSM image with the original image above it.....	28
Figure 7 Determining the NDVI value from an image .....	30
Figure 8 Determining the nDSM value from an image.....	31
Figure 9 Determining the characteristic spectral band values of trees from its spectral profile .....	32
Figure 10 Variation of spectral band 3 values for each land cover.....	34
Figure 11 Soil map of the study area .....	37
Figure 12 Vegetaton map of Western Australia.....	39
Figure 13 Results of the dicing process .....	40
Figure 14 Image showing the produced NDVI.....	41
Figure 15 Initial decision rule tree .....	42
Figure 16 The produced classified image from the initial decision rule tree .....	43
Figure 17 Final decision rule tree .....	44
Figure 18 the Produced classified image from the final decision rule tree .....	45
Figure 19 Classified image of the study area .....	47
Figure 20 Refined map of the classified image using the soil map as ancillary data.....	49
Figure 21 Refined map of the classified image using the vegetation map as ancillary data.....	49
Figure 22 Distortion error of trees as compared to the normal image .....	52

Figure 23 Inconsistency in image quality showing patches of images of different spectral signature .	53
Figure 24 Missing data within an image .....	54
Figure 25 Error with the nDSM image .....	54
Figure 26 NDVI image showing some of the tree canopies were not captured as vegetation .....	55
Figure 27 NDVI image showing some red colour rooftops were identified as vegetation .....	55
Figure 28 Image showing variability of the colour of the rooftops .....	56
Figure 29 Deciding orders of classification and the resulting rules and conditions .....	57
Figure 30 Graph showing the spectral band values of the distorted trees .....	58
Figure 31 Classified image of dense residential area.....	61
Figure 32 Classified image of undisturbed vegetated area .....	62
Figure 33 Mismatch of the soil map and the classified image .....	64
Figure 34 A subSection of the study area showing dense urban area and some vegetated area.....	71
Figure 35 A subSection of the study area showing preserved vegetation area .....	72
Figure 36 A subSection of the study area showing some sort of agricultural area .....	73
Figure 37 A subSection of the study area showing dense urban areas .....	74
Figure 38 A subset image of the study area as a comparison for the classified image .....	75
Figure 39 A classified image of a subSection of the study area .....	76

## LIST OF TABLES

Table 1 Reference of common land cover classes used for water balance study.....	16
Table 2 Definitions of land cover based on the 3 variables.....	35
Table 3 Error of omission and commission.....	48
Table 4 Percentage and total area of land cover.....	63

## CONTENTS

Executive Summary .....	2
Acknowledgements .....	3
List of Abbreviations .....	4
List of Figures .....	5
List of Tables .....	7
1. Introduction.....	10
2. Literature Review.....	12
2.1. Urban Water Balance.....	12
2.2. Remote Sensing .....	18
2.3. Urban Monitor .....	21
2.4. Expert Classifier.....	23
2.5. Southern River Catchment.....	25
3. Methodology .....	27
3.1. Input Data.....	27
3.2. Data Processing.....	29
3.2.1. Dicing.....	29
3.2.2. Expert Classifier.....	29
3.2.3. Mosaic .....	35
3.2.4. Accuracy Assessment .....	35
3.3 Post-Classification Analysis .....	37
4. Results.....	40
5. Discussion.....	51
5.1. Improvement of Classification Accuracy.....	51

5.1.1.	Data Quality .....	51
5.1.2.	Heterogeneity of Land Cover.....	56
5.2.	Urban Water Balance .....	63
6.	Conclusions and Recommendations .....	66
	References.....	67
	Appendix I .....	71
	Appendix II .....	75

## 1. INTRODUCTION

More than half of the world's population will be living in the urban areas by 2030. United Nations (UN) has projected that the number will be 4.9 billion people which accounts for 60% of the world population (Mcgee 2001). Such rapid urban population growth puts the pressure on the governments to cope with the large increase of population in a relatively short period of time. The main challenge will be the provision of infrastructure services including water supply.

Perth's population has seen a significant growth in the last decades and it was predicted that it could reach 2.3 million by year 2030 (Syme & Nancarrow 2011). As identified by Water Corp, an additional 120 giga litres of water need to be supplied to Perth Metropolitan area by 2030 (Syme & Nancarrow 2011). Therefore, water planners and regulators are faced by challenges to provide sufficient water for Perth's population.

In recent times it was reported that the water table in the Perth urban area is declining (Syme & Nancarrow 2011). While the water balance has not been fully understood, it may suggest that there is not enough recharge to the water table or aquifer. Therefore, efforts to increase groundwater recharge may be required in view of providing water supply for the increasing population. Water balance represents a method of understanding water movement within a defined boundary (Kenway et al. 2011). Urban water balance provides an insight into the impact of urbanization on the hydrological processes within the catchment such as groundwater storage change, runoff and evapotranspiration (Mitchell et al. 2003). For example, increasing groundwater recharge could be achieved through directing storm runoff into drains and discharge basins within the catchment boundaries (McFarlane 1984).

While the importance of understanding urban water balance is clear, calculation of urban water balance is not that straightforward. A method of calculating urban water balance that will be explored in this report is through classification of land cover of the urban area. In the past, water balance of Perth urban area has been studied through land cover classification of analogue aerial photography. It was found that changes in land cover resulted in changes to water fluxes and storages (McFarlane 1984). However, the heterogeneous nature of urban land covers poses a challenge to the classification process as it requires information of the urban areas in as great detail as possible.

In 2007, digital image of the Perth region was started to be collected using aerial photography in a project called Urban Monitor (UM). The produced high resolution image has been identified for several applications such as monitoring change in vegetation conditions, irrigated areas, surface elevations and also land cover. Therefore, land cover classification will be applied to this high resolution imagery in establishing inputs to water balance.

Southern River catchment represents suitable study area because it consists of rapidly urbanized area and also areas of preserved vegetation. The land cover classification of high resolution imagery will be tested on both areas.

Therefore, this project aimed to:

1. Develop capability of land cover classification of high resolution imagery.
2. Improve the accuracy of the land cover classification.
3. Proof the feasibility of using high resolution imagery for urban water balance study.

These aims were achieved through the following work:

1. Identifying the key relevant land cover associated with hydrology processes which contribute to the urban water balance.
2. Developing decision rule tree to classify the land cover.
3. Accuracy assessments of the land cover classification.
4. Refinements of land cover classification using ancillary data.

In bigger picture this project can be used as a baseline for:

1. Developing a decision support tool for exploring alternative ways to manage urban water resources.
2. Promoting the use of high resolution imagery for other study including land, agriculture and vegetation management.

## 2. LITERATURE REVIEW

### 2.1. URBAN WATER BALANCE

Water balance can be defined as an account of water movement in the land phase pertaining to the hydrological cycle of a given area and time using the principle of mass conservation (Mitchell et al. 2003). There are several approaches for calculating water balance. It varies from a simple equation of inputs and outputs to complex models accounting for all hydrological processes. In general, water balance for a given time interval can be described by the formula:

$$\Delta S = (P + I) - (E + R) \quad (\text{Mitchell et al. 2003; Kenway et al. 2011})$$

Where,

$\Delta S$  = *Change in catchment storage*

$P$  = *Precipitation*

$I$  = *Imported Water*

$E$  = *Actual evapotranspiration*

$R$  = *Stormwater runoff*

The equation represents the change in catchment water storage which equals to the sum of inputs minus the sum of outputs. Change in catchment storage includes the storage within the soil profile as well as the groundwater aquifer (Mitchell et al. 2003).

Several studies have used the equation above as a baseline for simple groundwater balance model. Water accounting done in Singkarak-Ombilin river basin of Indonesia had used the modified version of the equation to account for groundwater and surface water separately (Peranginangin et al. 2004). A hydrological model called GROWA was used to study the water balance of the River Rur catchment basin near Aachen, Germany. Similar to the equation above, the principal water balance components of the model were actual evapotranspiration, total discharge, surface runoff, interflow and groundwater recharge (Montzka et al. 2008). Another study used groundwater modelling to investigate the surface and sub-surface water accounting in River Zenne, Belgium (Dujardin et al. 2011).

However, in the urban context quantitatively representing the water balance is not so straightforward. Urban area is a complex and dynamic system consisting of both natural and built environment. Urban development has replaced the pervious soil and vegetation with impervious surfaces and routed

stormwater runoff directly into streams and the ocean (Rose & Peters 2001). Studies have shown that urban development has changed stormwater runoff characteristics. The effects include decreased groundwater recharge, increased surface runoff, increased magnitude of peak runoff and decreased lag time between rainfall event and runoff response (Burns et al. 2005). Also, in a relatively flat urban areas which are affected by inundation and shallow groundwater tables, large volumes of water need to be drained (Barron, Pollock, et al. 2011). Before, the water was naturally stored in landscape and was lost to evaporation. Therefore, urbanization has affected the hydrological cycle in general.

In addition, defining the boundaries of an urban catchment is not simple. Aside from increasing the surface runoff, it was found that urbanisation also leads to expansion of catchment area which contributes to the river flow (Barron, Pollock, et al. 2011). This relates to the concept of hydrological connectivity where urbanisation has changed soil type, slope, and land use which affects the water transport (Barron, Pollock, et al. 2011). Boundaries are needed to determine the relevant inputs and outputs in applying the conservation of mass principle as represented in the equation above.

One of the methods of measuring water balance is by identifying the land cover of the catchment area. Land cover is generally physical objects which cover the surface of the earth. Water balances of Perth have been carried out where land cover was estimated using analogue (film) aerial photography (McFarlane 1984). Measurements of water fluxes from each land cover were used to approximate the water balance (McFarlane 1984). However, due to the complexities of urban hydrology, the water balance was only approximate and the fluxes were considered as a lumped process which is more suitable than spatially distributed process (McFarlane 1984). This is due to the scattered distribution of the pervious and impervious surfaces.

Similar studies have also used land cover information to define fluxes to water balance although not of urban areas. Land cover classification was used to quantify total fluxes in a modelling procedure done for a brown field case site in Vilvoorde, Belgium (Dujardin et al. 2011). A large scale water balance model of an agricultural site was done by using the land cover classification of agricultural vegetation and impervious surfaces as the inputs (Montzka et al. 2008).

The relevant land cover features used in the Perth water balance study include roofs and paths, roads and car parks, trees and large bushes, small bushes and grass, water bodies (McFarlane 1984). The flows of water through these land cover were estimated based on knowledge of flows from several other studies as well as the environmental conditions of the area. Each land cover feature affects the hydrological processes in one way or another resulting in water storage changes. Finally, it was discovered that changes in land cover resulted in changes to water fluxes and storages (McFarlane 1984).

The hydrological processes affected by the land cover features include evapotranspiration (E), Runoff (R), and also infiltration which is not represented in the water balance equation but may contribute as direct recharge to the storage. Precipitation which falls on rooftops and paths could lead to the following outcomes:

1. Loss to the atmosphere through evaporation. The amount of evaporative loss from rooftops was found to be as high as 20% of intercepted precipitation (McFarlane 1984).
2. Indirect aquifer recharge through roof runoff directed to soak wells.
3. Direct recharge to the soil from roof runoff directed onto the ground adjacent to the house through spoon drains.

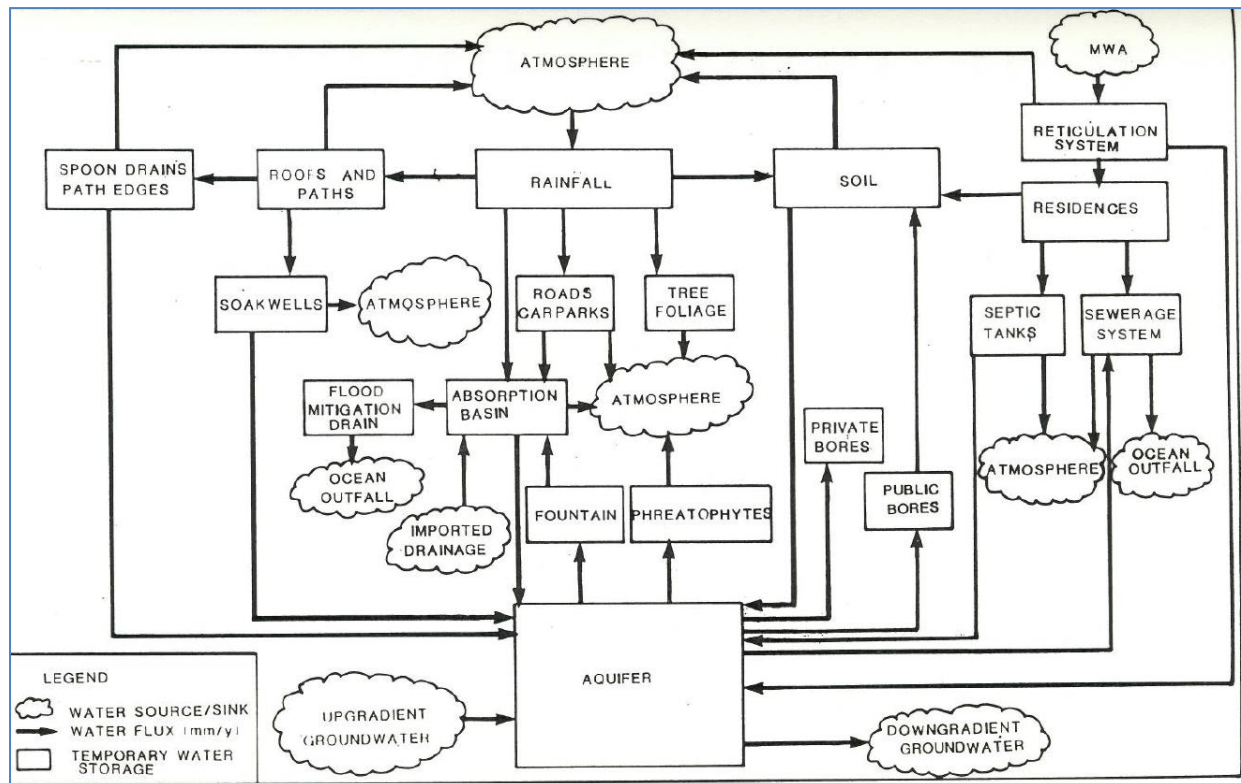
Precipitation which falls on roads and car parks also has several possible outcomes:

1. Direct recharge to the soil from runoff directed into drains and discharge basins within the catchment boundaries.
2. Evaporative loss from the road surface and any depressions.
3. Splattered onto adjacent soil which itself has two possible outcomes. The outcomes are either lost to the atmosphere through evaporation or infiltration to the soil. The study has found that 20% of the rainfall falling on roads flows onto the soil (McFarlane 1984).

Precipitation which falls on trees and large bushes has following outcomes:

1. Interception loss on the canopies which leads to less infiltration to the soil.
2. Through fall to the soil which then has the possible outcomes as explained above
3. Evaporative loss from the canopies

Finally, precipitation which falls on any water bodies is eventually lost to evaporation.



**Figure 1 Conceptual Model of urban water balance (McFarlane 1984)**

Figure 1 above summarizes all the possible fluxes of urban water balance. The figure shows the complexities of the processes involved in defining urban water balance. However, this model was built based on the study of the Subiaco-Shenton Park catchment. Different catchment may involve different fluxes depending on the land cover feature available.

**Table 1 Reference of common land cover classes used for water balance study**

Land Cover Classes	Reference of studies
Rooftops	(McFarlane 1984) (Barron, Pollock, et al. 2011) (Dujardin et al. 2011) (Montzka et al. 2008)
Roads	(McFarlane 1984) (Barron, Pollock, et al. 2011) (Montzka et al. 2008)
Grass	(McFarlane 1984) (Barron, Pollock, et al. 2011) (Contreras et al. 2011) (Dujardin et al. 2011) (Montzka et al. 2008)
Trees and vegetation	(McFarlane 1984) (Barron, Pollock, et al. 2011) (Contreras et al. 2011) (Dujardin et al. 2011) (Montzka et al. 2008)
Cleared areas	(McFarlane 1984) (Barron, Pollock, et al. 2011) (Dujardin et al. 2011)
Water bodies	(McFarlane 1984) (Barron, Pollock, et al. 2011) (Dujardin et al. 2011) (Montzka et al. 2008)

Table 1 above shows the typical land cover classes used for water balance study. These land covers classes are deemed necessary for the context of their study. Based on all considerations above, the relevant land cover features that will be classified in this project are:

### 1. Rooftops

The rooftops will be explored further based on the soil property. Clay soil is known to have poor infiltration rate which results in ponding problem (Moore 2001). The water usually needs to be drained somewhere which therefore has different processes compared to other type of soil. Therefore the houses will be classified whether they were built:

- a. On clay soil: in which water is likely to be disposed off to road drains.
- b. On non-clay soil: in which water is likely to be infiltrated on site.

### 2. Roads

Roads include driveway, car parks and basically other impervious surfaces.

### **3. Grass**

Grass includes lawns and parks.

### **4. Trees and vegetation**

Trees and vegetations have different interception losses, through fall and evaporative losses depending on the species and age. Banksia is a typical plant of a shallow sandy aquifer in Mediterranean climate. It is known to have a shallow root depth and interact with groundwater differently based on the season (Zencich et al. 2002). Therefore, the trees and vegetation will be classified whether they are:

- a. Banksia woodlands on the coastal plain.
- b. Other types of trees.

### **5. Cleared areas**

Cleared areas include mainly dry land agriculture with annual pasture which is dead in summer. Therefore similar to the rooftop, cleared areas will be classified whether they are:

- a. Clay soil: in which water ponding will be the most likely case.
- b. Non-clay soil: in which water is likely to be infiltrated on site.

### **6. Water bodies**

Ponds, lakes and swimming pools are considered here as water bodies.

The heterogeneous nature of urban land covers which include different kinds of roofs, roads, trees, grass, impervious areas, irrigated and non-irrigated areas and green space made the water balance more difficult to understand (Caccetta, McFarlane, et al. 2011). However, if the complex urban land cover features can be classified, the water fluxes and storages can be approximated as explained above. Therefore, identifying the water fluxes and storages will lead to better understanding of urban water balance. Nonetheless, there is a need for a method of obtaining information of the urban area in as great detail as possible.

## 2.2. REMOTE SENSING

Remote sensing can be defined as the measurement of information about a physical object using non-contact sensor systems (Jensen 2005). Generally, remote sensing instruments use sensor to record electromagnetic radiation (ER) reflected, emitted or back-scattered by an object (Jensen 2005).

Consequently, remote sensing has several advantages over in situ data collection methods. Firstly, as a passive process it is unobtrusive and does not disturb the object of interest. Secondly, it can cover a wide geographical area and is therefore more time efficient compared to single-point observations(Jensen 2005).

There are many parameters of the electromagnetic radiance that remote sensing process is able to record and each of them has different resolution. Spectral information and resolution represents the wavelength intervals which the instrument is able to detect. Most multispectral remote sensing systems record more than 4 bands of electromagnetic spectrum including red, green, blue and near infrared (Jensen 2005).

Spatial information and resolution is defined as the smallest linear separation of two objects that can be defined by the remote sensing process (Jensen 2005). In layman terms, it can be simply understood as the size of the pixels or the cells of the resulting image. Depending on the purpose of the remote sensing process, the spatial resolution can be as large as  $1\text{km} \times 1\text{km}$  and as small as  $10\text{cm} \times 10\text{cm}$ . Therefore, the smaller the spatial resolution, the greater is the spatial resolving power or in other words, the more detail represented by the resulting image. Figure 2 below explain the difference of spatial resolution of an image. The first figure has spatial resolution of  $0.2\text{m} \times 0.2\text{m}$  while the below one has spatial resolution of  $2\text{m} \times 2\text{m}$ . Both figures are at scale of 1:2500, but the first figure shows greater detail. Since fine scale resolution image covers more details, it is suitable for the study of urban area.



Figure 2 Comparison of spatial resolution

Temporal information and resolution identify how often the remote sensing process capture information about the area of interest (Jensen 2005). Again depending on the purpose of the remote sensing process, temporal resolution drives the ability for temporal comparisons of a location over time.

Radiometric information and resolution defines the sensitivity and the number of distinct brightness value of the reflected electromagnetic radiation spectrum that the sensor can capture (Jensen 2005). The latest sensor systems have been able to record 12-bit radiometric resolution with brightness values ranging from 0 to 4095 (Jensen 2005). Therefore, high radiometric resolution raises the probability of detecting the physical object more accurately.

Remote sensing generally uses either satellite or airborne sensor technology to acquire information of objects on the earth. One of its common uses is to obtain land use and land cover information. For example, remote sensing was used to study land use and land cover dynamic changes in Brazil and Bangladesh (Uddin & Gurung 2010; de Espindola et al. 2011). It has also been commonly used to study vegetation types and canopy cover (Borel 2010; Q. Yu et al. 2006). Observations of land cover information will define the land use information of a certain area which is of significant interest to planners, resource managers and decision makers (Rozenstein & Karnieli 2011).

However, the use of remote sensing for fine scale resolution (both spatial and radiometric) of urban areas has been limited. This was mainly due to the limitation of the analogue image processing method of the produced aerial image. Manual visual interpretation of the image is extremely tedious and subjective even at low resolution (Rozenstein & Karnieli 2011). Also, the spatial resolution provided by the analogue aerial photography is quite large (generally 30m × 30m) which makes it suitable for mapping of large area only (Rozenstein & Karnieli 2011). Some aerial photography has higher resolution but it only records 3 bands values. This also explains why it has only been used for monitoring of natural resources applications over large geographic areas but not urban areas (X. Wu et al. 2010).

### 2.3. URBAN MONITOR

The acquisition of multi-band aerial imagery serves the opportunity for quantitative mapping and monitoring applications at very fine scales which previously was not possible (X. Wu et al. 2010). The Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) and its partners are collecting aerial imagery over the Perth region for Urban Monitor (UM) project annually (Collings et al. 2011). The area covers new and existing urban areas, peri-urban agriculture, native forests and woodlands, wetlands, agricultural and industrial areas over 9,600 km<sup>2</sup>.

Started in 2007, the data was collected using airborne sensor at an altitude of 1300m over 19 cloud free days during dry summer weather (Caccetta, Collings, et al. 2011). To minimise the effect of solar angle and shadowing, the data was collected between 10 AM and 2 PM (Caccetta, Collings, et al. 2011). The resulting images represent a spatial resolution of 10 to 30 centimetres, making it suitable for mapping of complex urban areas (X. Wu et al. 2010). With radiometric resolution of up to 14 bit, it provided the opportunity to accurately classify most objects present on the surface of the earth.

The images have 4 spectral bands interval namely green, blue, red and near-infra red. Monitoring of changes in vegetation is possible since the near-infra red band is now detectable. The reflectance of vegetation is particularly low in the band 3 (red) but high in the near-infra red band (Barton 2011). This unique characteristic of spectral reflectance makes it easier for vegetation to be identified by capitalizing on the difference between the two bands

Normalized Digital Surface Model (nDSM) is another output obtained from the acquisition of aerial imagery. nDSM is derived from the Digital Surface Model (DSM) and Ground Elevation Model (GEM) of the same area. DSM basically contains the elevation information of the land cover features while GEM contains information of the elevation of the ground. By subtracting the GEM from the DSM, nDSM is obtained (H.-yang Yu et al. 2011). Essentially, nDSM is showing absolute height values of the land cover features since the elevation of the ground has been technically set to zero (H.-yang Yu et al. 2011). Figure 3 explains how nDSM is derived from DSM and DEM. The red line is DSM while blue line is DEM. By subtracting the blue line from the red line, technically the height of the objects is obtained. The terms DEM and GEM are essentially the same. GEM term is used for the rest of the study except in this figure.

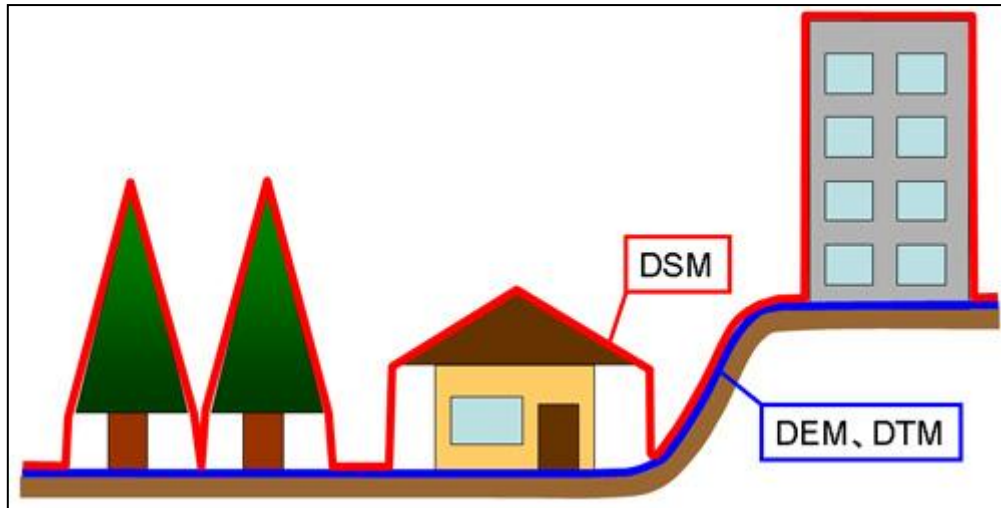


Figure 3 Concept of extracting nDSM from DSM and DEM(GEM) (Geospatial Information Authority of Japan 2010)

As identified by the consortium, the potential uses include monitoring of wetland condition, river foreshore condition, vegetation condition, weed invasion, tree density and growth, digital terrain model, irrigation efficiency, base map for planning and adoption of solar roof technology. In urban context, this includes monitoring of irrigation efficiency on public open spaces, land cover changes of impervious areas and groundwater and storm flow estimation (Caccetta, McFarlane, et al. 2011). One of its applications which will be explored in this project is classification of land cover information which will contribute to understanding the urban water balance. Essentially, 4 bands imagery provides the opportunity to classify the land cover more accurately.

The high resolution imagery provides advantages as well as challenges. It provides the opportunity for automated processing of the image (Caccetta, Collings, et al. 2011). Digital image processing provides more accuracy than visual interpretation, making use of threshold beyond which human can detect (Jensen 2005). Moreover, computer techniques can be used to perform repetitive task without being subjective and also to process large amount of data. However, since the Urban Monitor is a relatively new project, the data has not been used regularly. The method for land cover classification of high resolution imagery is relatively untested.

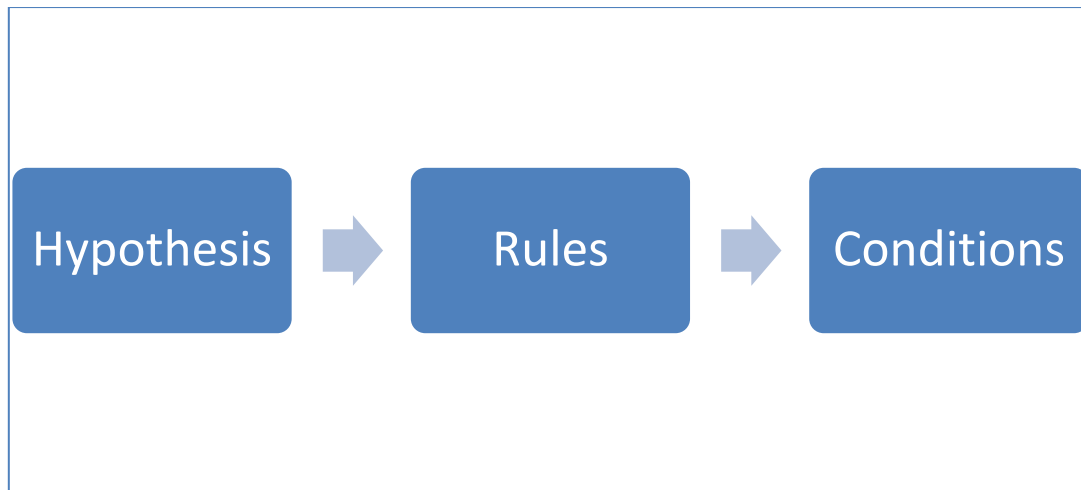
At broader scale, Urban Monitor represents an opportunity to develop a decision support tool for urban water accounting. A case in point would be to study the impact of urbanisation on water balance components.

## 2.4. EXPERT CLASSIFIER

A remotely sensed image can be examined in relation to the features on the earth it represents. These features include physical objects such as trees, buildings, roads and water bodies which can then be grouped into broader categories of land uses and land covers. Therefore, the main objective of a classification is to accurately identify these features and group them based on similar characteristics where possible. Classification methods involve a broad range of techniques and processes, ranging from simple interval measurements of reflected ER in discrete portions of the spectrum to processing of the image in order to identify a certain feature such as tree canopy (ERDAS 2009).

Multispectral classification is defined as classification which uses multiple bands of regions of the electromagnetic spectrum. In this project, multispectral classification is used to classify the different features of the high resolution image. The process involves sorting of pixels into a number of classes based on their data values. If a pixel's brightness value falls between a range of values, the pixel is then assigned to the corresponding class. The classes usually have a specific characteristic or feature which distinguish one from another. The result of the classification is commonly represented as an image such as land cover or land use map.

Expert classification is a classification system that uses human knowledge in developing knowledge base in form of data and rules to solve a problem (Jensen 2005). An expert system consists of several components namely human knowledge, user interface, knowledge base and data base. Human's knowledge of a specific area is applied in knowledge base by analysis of the data base using the user's interface. One way of conceptualizing an expert system is to use a decision-tree structure where rules and conditions are evaluated to determine the outcome of a hypothesis as seen in Figure 4. Therefore, the hypotheses, rules and conditions are determined by the expert's knowledge.



**Figure 4 Decision process of Expert Classification**

The expert classification has several advantages over other classification methods. Firstly is the flexibility in terms of deciding the hypothesis, rules and conditions (Jensen 2005). User can evaluate the output image and work backward to identify how a conclusion was reached. If the output is not satisfying, user is able to modify the rules and conditions. Secondly, the nature of the decision process is very clear as shown in Figure 4. The decision process of other method such as neural network is not as obvious (Jensen 2005).

ERDAS Imagine is a geospatial data analysis and remote sensing software. It is capable of performing advanced remote sensing analysis and spatial modelling to create new information and visualize them in various map representations. Expert Classifier (EC) is one of its classification modules which perform the core function of this project. The module takes a rules-based approach of multispectral expert classification where rule tree is developed to describe the conditions for a certain class of features. The Expert Classifier accepts a wide range of data base input such as raster imagery, vector coverage and spatial models (ERDAS 2001).

The rule may consist of one or more conditional statements about user-defined variables that subsequently determine the hypothesis. Numerous rules and hypotheses can be linked together to define a class which is more specific. EC is composed to two parts namely the Knowledge Engineer and the Knowledge Classifier. The Knowledge Engineer is the user interface component where user applies his knowledge to define the variables, develop the rules hierarchy and finally hypothesise the final output classes. The Knowledge Classifier is the module which processes the knowledge base developed in Knowledge Engineer and creates the output in form of classified image. However, the Knowledge Engineer has its own Test Mode where the accuracy of the classification can be analysed upon building the knowledge base (ERDAS 1999).

## 2.5. SOUTHERN RIVER CATCHMENT

Perth, the capital of Western Australia, is experiencing rapid urbanization with residential and commercial areas replacing previously undisturbed vegetated area. This is especially evident in the expanding regions of Southern River catchment. However, urban development in this area is continuously challenged by environmental problems such as seasonal inundation, preservation of wetlands and rising groundwater tables (Barron, Pollock, et al. 2011). The Southern River catchment also contributes nutrient loads significantly to the Swan-Canning estuary. This leads to environmental concerns that future urban development may increase the nutrient loads to the estuary. It is predicted that land use change will change the hydrological cycles substantially as inundation and shallow water tables are unsuitable with urban infrastructure in general (Barron, Barr, et al. 2011). These environmental issues highlight the need for extra effort and attention when altering the land use and land cover of the catchment.

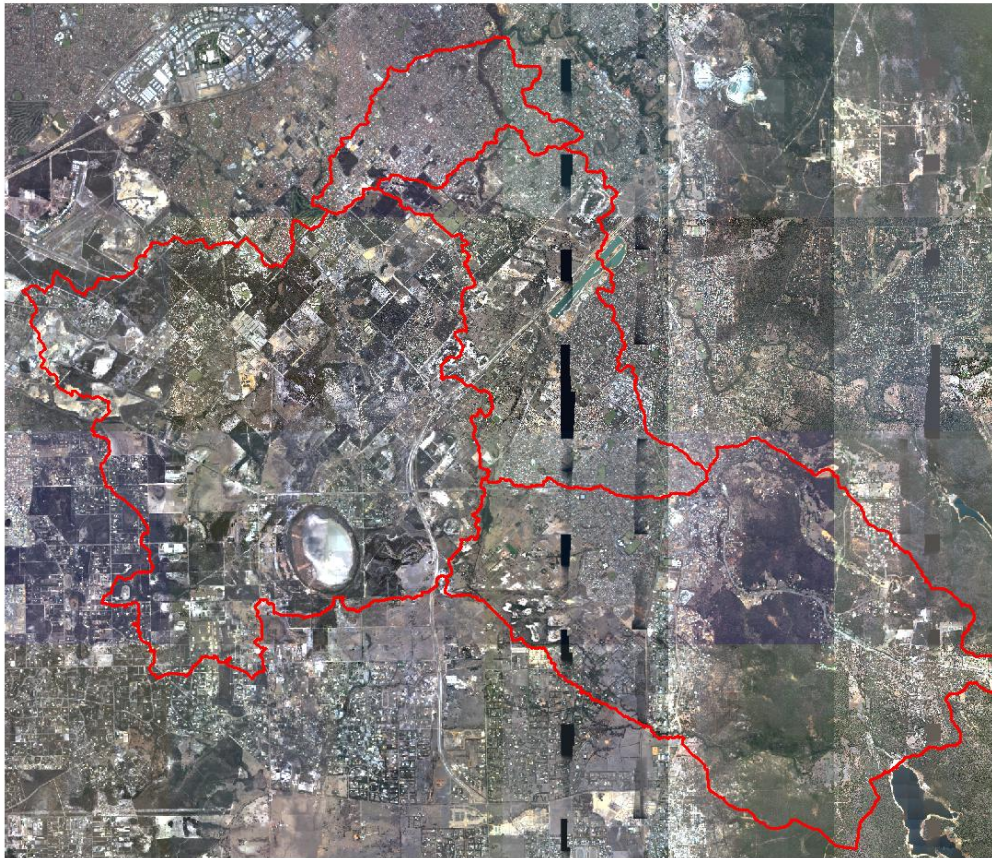
The south eastern part of the catchment includes an elevated upland area while the rest of the catchment is generally a flat low-lying area which is underlined by Quaternary fluvial and Aeolian deposits. The soil consists of Bassendean Sand in the west and the Guildford Formation to the east which is mainly clay. The catchment experiences the problem of seasonal inundation and formation of wetlands largely due to low surface gradient and poor drainage capacity (Barron, Pollock, et al. 2011).

The major land uses of the Southern River catchment are urban and rural. The urbanised area covers residential areas and infrastructure on the northern and eastern side of the catchment. The rest of the catchment is the rural area which encompasses state forest area, remnant vegetation area and some cleared areas used for livestock farming and horticulture (Barron, Pollock, et al. 2011).

The average annual precipitation in the catchment is 786 mm of which 87 mm becomes runoff (Barron, Barr, et al. 2011). More than 80% of the annual precipitation falls during the winter months while the summer months are warm and dry. During summer, the evaporation exceeds the rainfall which leads to decrease in water storage. During autumn and early winter, runoff is limited due to significant storage capacity of the sandy soils and the wetlands. On winter months, potential evaporation is usually lower than monthly rainfall. However, on annual basis, potential evaporation is greater than precipitation (Barron, Barr, et al. 2011).

The Southern River catchment makes a suitable area of study as it consists of an area which is rapidly urbanized but also contains areas of coastal and hills vegetation which are preserved. The urban areas has more heterogeneous and complex features which now can be identified and distinguished due to the fine resolution while the rural areas has more uniform features. This makes it ideal for the project

because land use classification of high resolution imagery can then be tested on both urban and rural areas.



**Figure 5** Extent of the digital aerial imagery and the boundaries of the Southern River Catchment

Figure 5 above shows the extent of the study area. The red line is the boundaries of the Southern River catchment along with the boundaries of the sub-catchments. Southern River catchment is shown to be within the coverage of the digital aerial imagery.

### 3. METHODOLOGY

#### 3.1. INPUT DATA

Digital aerial photography was flown in 2007 over the Perth metropolitan region and CSIRO provided the base for the land cover classification developed for this project. The study area encompasses the Southern River catchment including its surrounding area as outlined in Section 0 and can be seen in Figure 5.

As has been outlined in Section 0, the data has 4 spectral band values with radiometric resolution of signed 16-bit. The total area covered by the image is 378 km<sup>2</sup> with spatial resolution 20cm × 20cm. The image is represented in raster format which consists of 105,000 by 90,000 pixels (x, y).

CSIRO also supplied the DSM, GEM and nDSM derived from the digital aerial photos. The data has been orthorectified, radiometrically corrected and color balanced providing a ready to use product. Orthorectification is process of aligning the pixels in proper planimetric (x, y) map location by removing any geometric distortion (Jensen 2005). This allows for spatial comparison of the image with any maps with the same projection. Orthorectification is done by matching the image with its high resolution DSM which is automatically extracted (Caccetta, McFarlane, et al. 2011). Radiometric correction is standardization of the spectral signature of the images in order to make temporal comparison (Caccetta, Collings, et al. 2011). It is done by calibrating the sensor to ground reflectance consisting of in-situ fiberboard targets which have been measured using a spectrometer (Caccetta, McFarlane, et al. 2011). Figure 6 shows an example of an nDSM image. White indicates objects with height while black are objects on the ground. The image is just a Section taken from the original imagery.



Figure 6 A subset of the nDSM image with the original image above it

To further refine the land cover classification used in this study, ancillary data sources were also used during the expert classification process. Geology maps were obtained from the Department of Mines and Petroleum of the Government of Western Australia (Australian Government Department of Mines and Petroleum 2010). The map provides soil information for the study area. Vegetation information was acquired from the National Vegetation Information System (NVIS) of the Department of Sustainability, Environment, Water, Population and Communities of the Australian Government (Australian Government Department of the Environment and Water Resources 2002). NVIS provides spatial information on the extent and distribution of vegetation types in Australia.

### 3.2. DATA PROCESSING

#### 3.2.1. DICING

The original image was diced for convenience and efficiency purposes. The file size of the image was too large to be analysed quickly and efficiently in the viewer of ERDAS Imagine. This may be due to the limitations of computer and technical capability. The original imagery and its nDSM were cut into the size of 17,500 × 22,500 pixel. As a result, the image was cut into total of 24 diced images which will be classified individually.

#### 3.2.2. EXPERT CLASSIFIER

The land cover features were differentiated using 3 variables. They are the spectral band values, nDSM and Normalized Difference Vegetation Index (NDVI).

NDVI was used to distinguish vegetation and non-vegetation land cover features. It has been commonly used to classify vegetation type and coverage (Geerken et al. 2005; Whiteside & Ahmad 2008; J. Wu et al. 2007). NDVI is measureable since the near-infra red band can now be recorded. NDVI value is calculated from the formula:

$$NDVI = \frac{Mean\ NIR\ band - Mean\ red\ band}{Mean\ NIR\ band + Mean\ red\ band} \text{ (Whiteside \& Ahmad 2008; Jensen 2005)}$$

The NDVI value was obtained using band 3 and 4 values of the original image. The NDVI value was automatically calculated by selecting the appropriate function and sensor according to the band combinations needed. The NDVI value of the identified land cover was obtained by looking at the spatial profile of the NDVI image.

Figure 7 shows how the NDVI value was obtained from the spatial profile. The graph shows the variation of the NDVI value across a line transect, which is the white line seen in the figures above it. The top left figure is showing the line transect in the original image while the top right figure is showing the line transect in its NDVI image. The value is 1 for vegetation and 0 for non-vegetation objects.

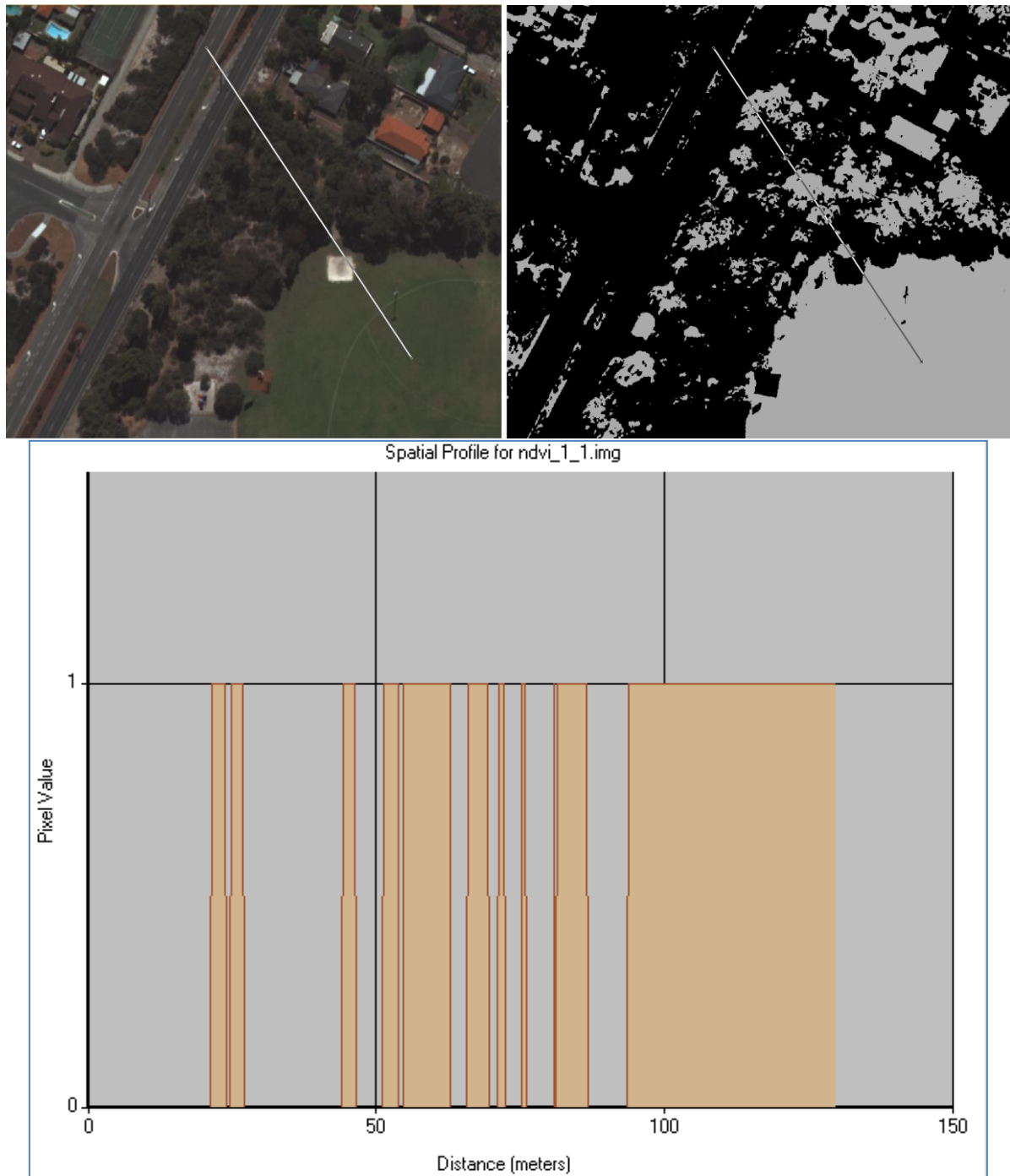
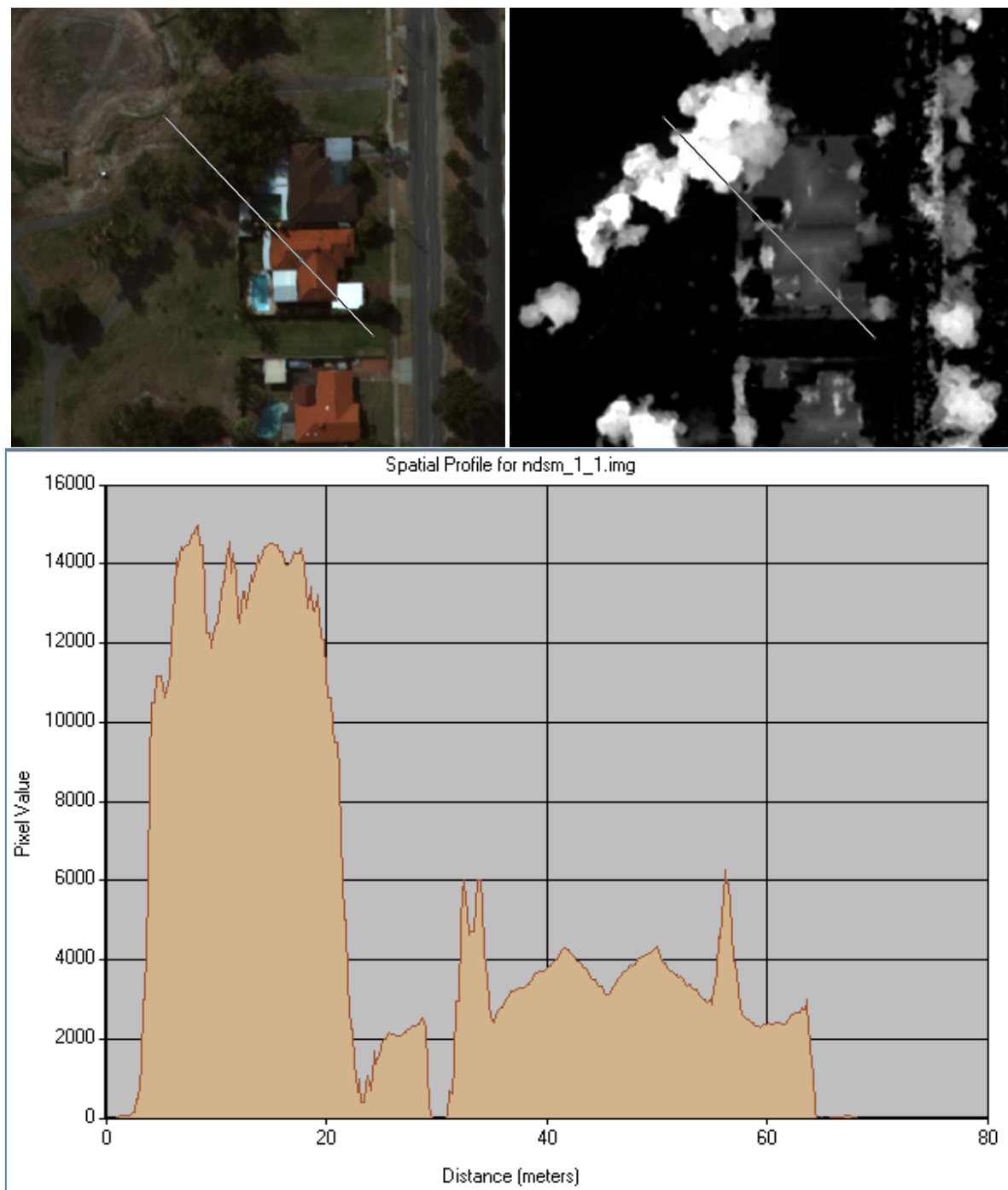


Figure 7 Determining the NDVI value from an image

Information of the height of the object was obtained from the nDSM image only. nDSM was used to tell apart object on the ground from those that are above the ground. Such height parameter has been commonly used in various classification studies (Taweessuk 2005). Similar to NDVI, the nDSM values were obtained from the spatial profile of the nDSM image as seen in Figure 8 below. The graph shows the variation of the object height across the line transect. The line transect is visible in the top left figure of original image and also the top right figure which is the nDSM image.



**Figure 8** Determining the nDSM value from an image

As mentioned before, the original image has 4 band values which can be used to differentiate the land cover features from each other. Some of the land cover features hold unique values for some of the band which can be used to characterize those specific land cover. Numerous classification studies have used band combination to classify specific land cover (Her 2007; Taweessuk 2005). The spectral band values were extracted by looking at the spectral profile of the original image as seen in Figure 9. To cover the variation of the trees, user may place the profile points as deemed necessary. The graph shows that trees have band 3 value of below 500.

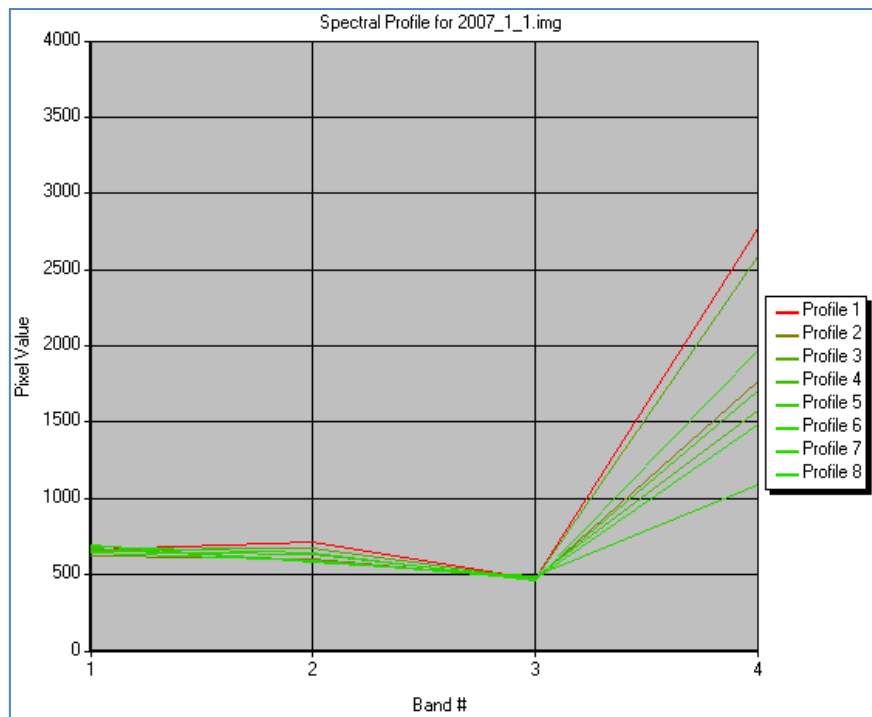


Figure 9 Determining the characteristic spectral band values of trees from its spectral profile

Essentially, the setting up of the rules in the Knowledge Engineer of the Expert Classifier was done by simple interval measurements of the values of these 3 variables namely spectral band values, nDSM and NDVI. It is important to ensure that the value interval measurement for a land cover feature represents the diversity of that specific feature. For example, the nDSM value interval for rooftop feature needs to consider the different height of the buildings and houses. This can be better represented by looking at Figure 10 below. The figure shows that a simple interval measurement of band 3 values was done for each land cover.

The order of the classification in the Knowledge Engineer also matters. A class that is specified first will not be reclassified by the other class that follows even if the rules are overlapping.

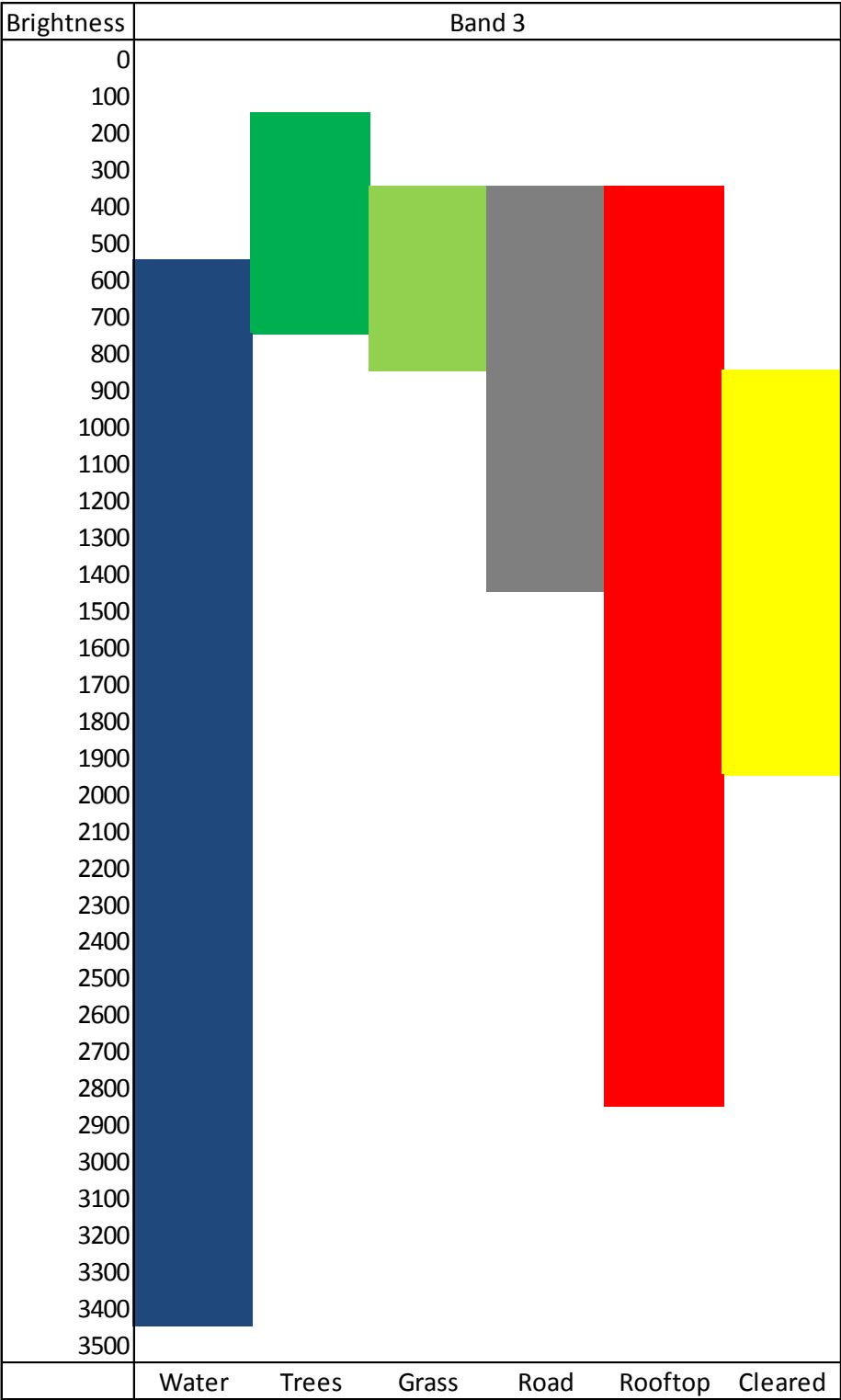


Figure 10 Variation of spectral band 3 values for each land cover

Therefore the rule tree of the Knowledge Engineer was built based on definitions summarized in the table below:

**Table 2 Definitions of land cover based on the 3 variables**

<b>Land Cover</b>	<b>Color (Spectral Band Values)</b>	<b>Height information (nDSM)</b>	<b>Vegetation (NDVI)</b>
<b>Rooftop</b>	Various (mixed)	Above ground	Non-vegetation
<b>Road</b>	Uniform, Grey to black	On the ground	Non-vegetation
<b>Grass</b>	Uniform, Mainly green	On the ground	Vegetation
<b>Trees</b>	Uniform, Mainly green	Above ground	Vegetation
<b>Cleared areas</b>	Various, White to yellow	On the ground	Non-vegetation
<b>Water bodies</b>	Uniform, Mainly blue	On the ground	Non-vegetation

The Knowledge Classifier was then executed based on the developed knowledge base. All 24 diced images were classified using the same rule tree.

### 3.2.3. MOSAIC

The image subsets were mosaicked together to form one classified image of the study area. All the classified diced images were used as the input. For this project, no seam line was required since there is no overlapping of the input images after the dicing process of the original imagery. The output mosaic area type contained the union of all the inputs in order to match the size of the original image of the study area. The projection and the cell size had to be the same as the original image. Finally, it is important to ensure that the background value is not of the same value as one of the land cover classes. If a land cover class happens to be set as background value, it will not be considered as a class in any further process.

### 3.2.4. ACCURACY ASSESSMENT

Next, checking of the classification accuracy was done. Random reference points need to be generated based on the required sampling scheme. The suitable sampling scheme for the reference points was the stratified random sample. Stratified random sampling was proven to be the most appropriate for remote sensing study because important minor feature class can be adequately represented using this

method (Maingi et al. 2002, Jensen 2005). The number of reference points needed was calculated using the formula:

$$N = \frac{Z^2 pq}{E^2} \quad (\text{Maingi et al. 2002; Jensen 2005})$$

Where,

N = Number of reference points

Z = Standard normal deviation for the 95% two-tail confidence level

p = Expected percent of accuracy of the classified image

q = (100-p)

E = Allowable error

The conventional value for allowable error (E) used in various studies was 5% (Maingi et al. 2002; Senseman et al. 1995). On the other hand, the value of expected accuracy (p) differs based on the confidence of the classification. The p value of several studies involving remotely sensed land cover classification ranged from 65%-90% (Senseman et al. 1995; Maingi et al. 2002). Since land cover classification of high resolution imagery has not been frequently studied, this justified greater allowable error in the project. Therefore, for this project the allowable error was 10% while the expected accuracy is 60%.

$$N = \frac{2^2 \times 60 \times 40}{10^2} = 96$$

After the 96 random points were generated on the viewer, manual visual inspection of each point was completed through visual comparison with the aerial imagery. This step compared the classification with what actually appears on the surface of the earth. Errors of omission and commission were then examined for each land cover class.

### 3.3 POST-CLASSIFICATION ANALYSIS

Post-Classification Analysis was done using ArcGIS. As mentioned in Section 3.1, geology map and vegetation map were used along with the classified image. By overlaying the map over the classified image, the land cover features can be classified further based on the geology and vegetation map.

The Armadale and Fremantle geology maps obtained from the Department of Mines and Petroleum were first stitched (Australian Government Department of Mines and Petroleum 2010). Next, the soil was generalised into either clay or non-clay soil as has been specified in Section 2.1 . New soil map was then generated based on these two general soil types. In the process, it is important to ensure that the resulting soil map has the same coordinate system and same cell size as the classified image. The raster soil map was then added together with the classified image to obtain land cover features which lay on either clay or non-clay soil (I. a Dar et al. 2011).

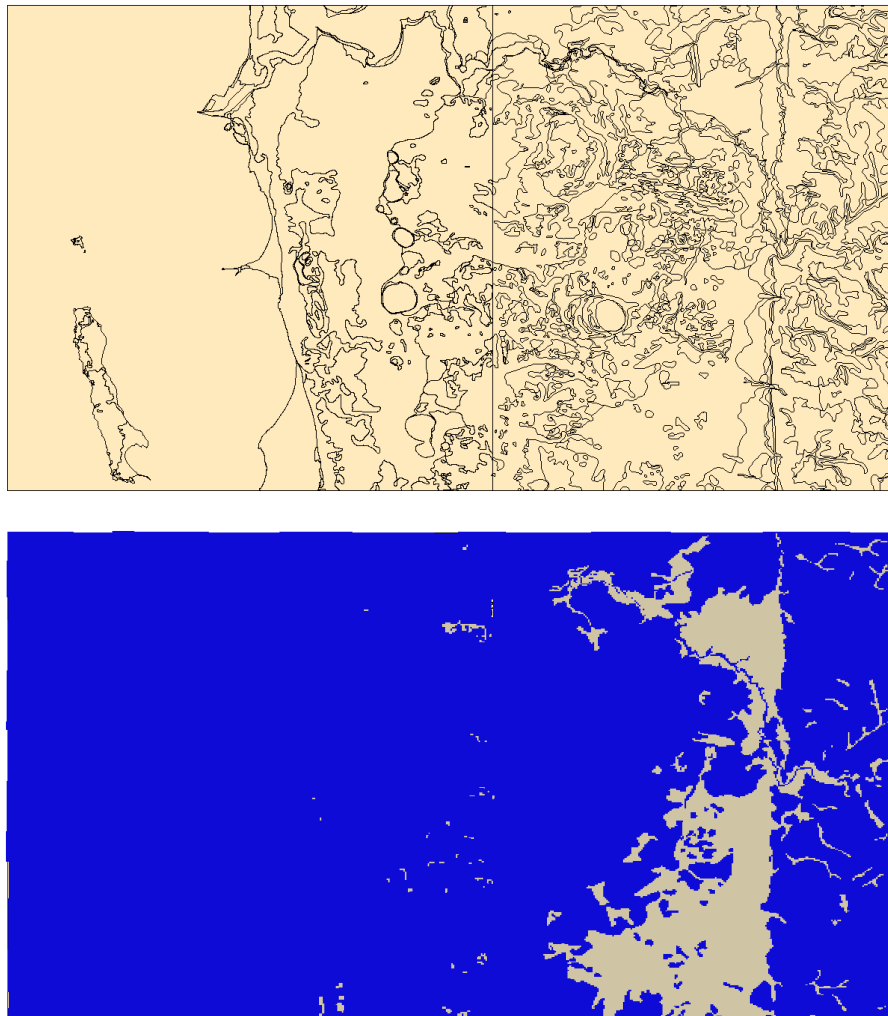


Figure 11 Soil map of the study area

Figure 11 above shows the soil map used to refine the classification of the land cover. The top figure is the original soil map which consists of various types of soils although represented as one colour. The bottom figure shows the soil map after it was generalized into either clay or non-clay.

Similar to the soil map, the trees and vegetation were classified further using the Western Australia vegetation map obtained from NVIS. The trees and vegetation was generalised into either Banksia woodlands or other trees as specified in Section 2.1. The coordinate system and the output cell size of the vegetation map were also matched to those of the classified image. The raster vegetation map was then added together with the classified image to classify whether the trees and vegetation are Banksia woodlands or other kinds of trees (Zhao et al. 2011). In Figure 12, the top figure shows the vegetation map of Western Australia which consists of all sorts of vegetation. The bottom figure shows the vegetation map after it was generalized into Banksia woodlands, non-vegetation or unknown and other native trees.

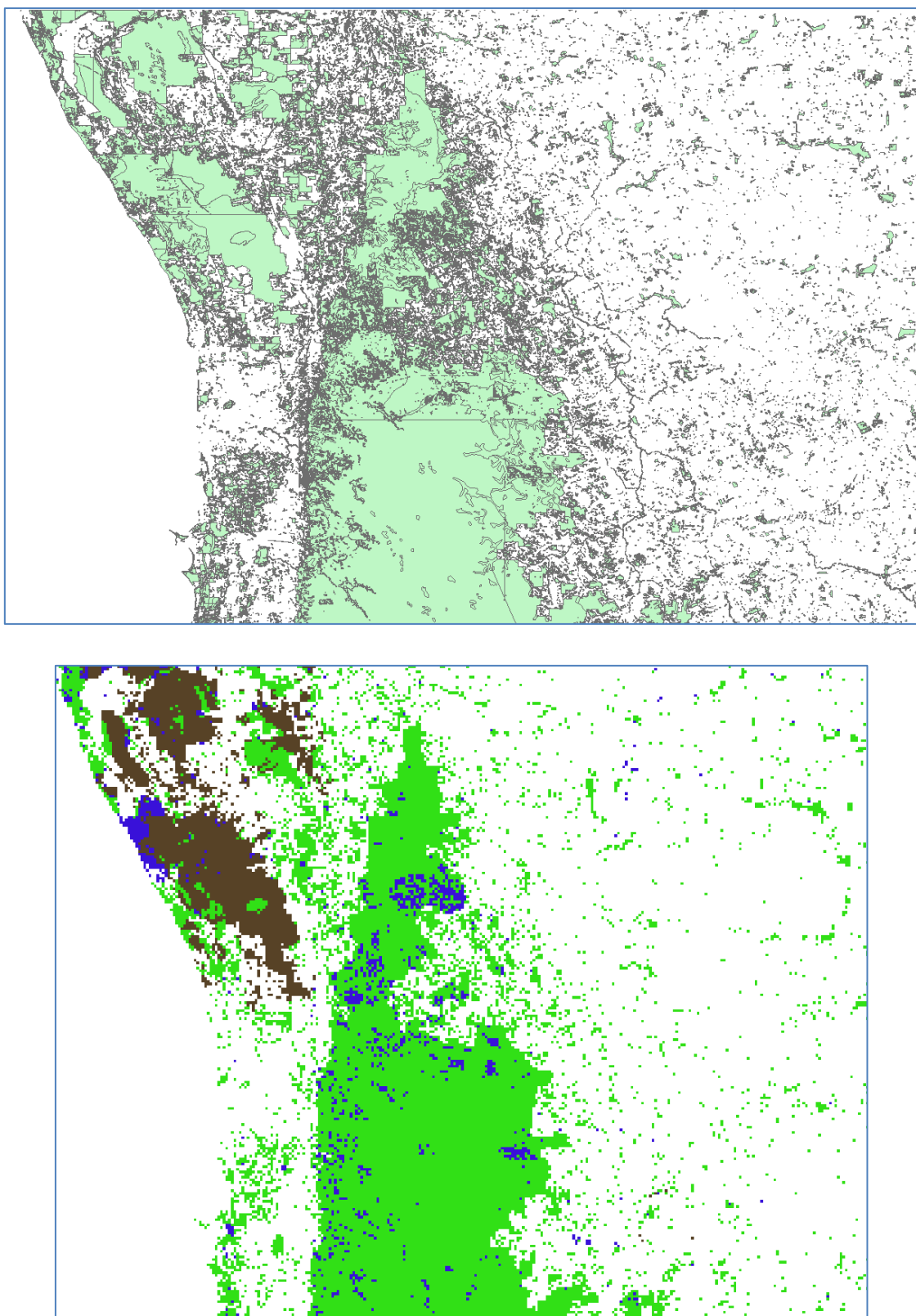


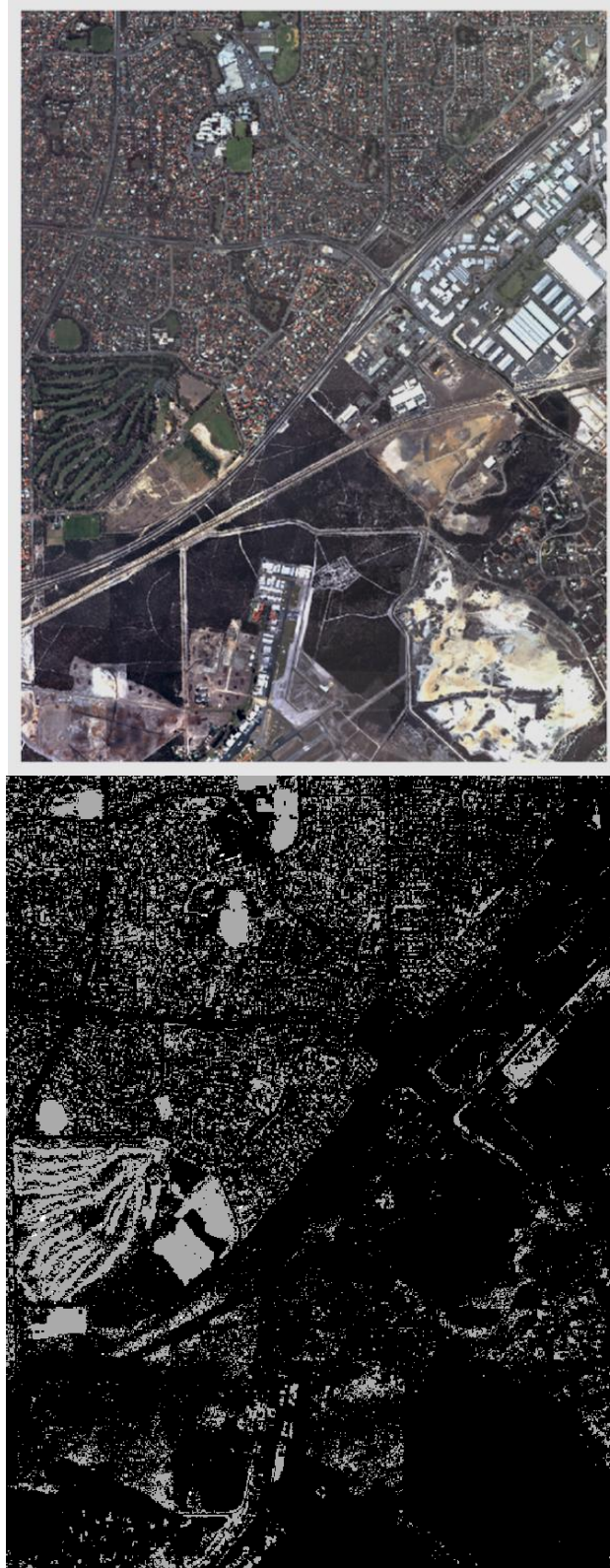
Figure 12 Vegetation map of Western Australia

## 4. RESULTS



Figure 13 Results of the dicing process

Figure 13 shows some of the diced images of the study area. The figure above shows the heterogeneity of the land cover within the Southern River catchment. The top left and right figures show the dense residential areas. The bottom left figure shows vegetated areas while the bottom right figure shows some kind of agricultural areas. Larger version of the image can be seen in Appendix I.



**Figure 14** Image showing the produced NDVI

Figure 14 shows the produced NDVI from its original image which is seen above. The white colour objects are the detected vegetation while the rest are non vegetation objects.

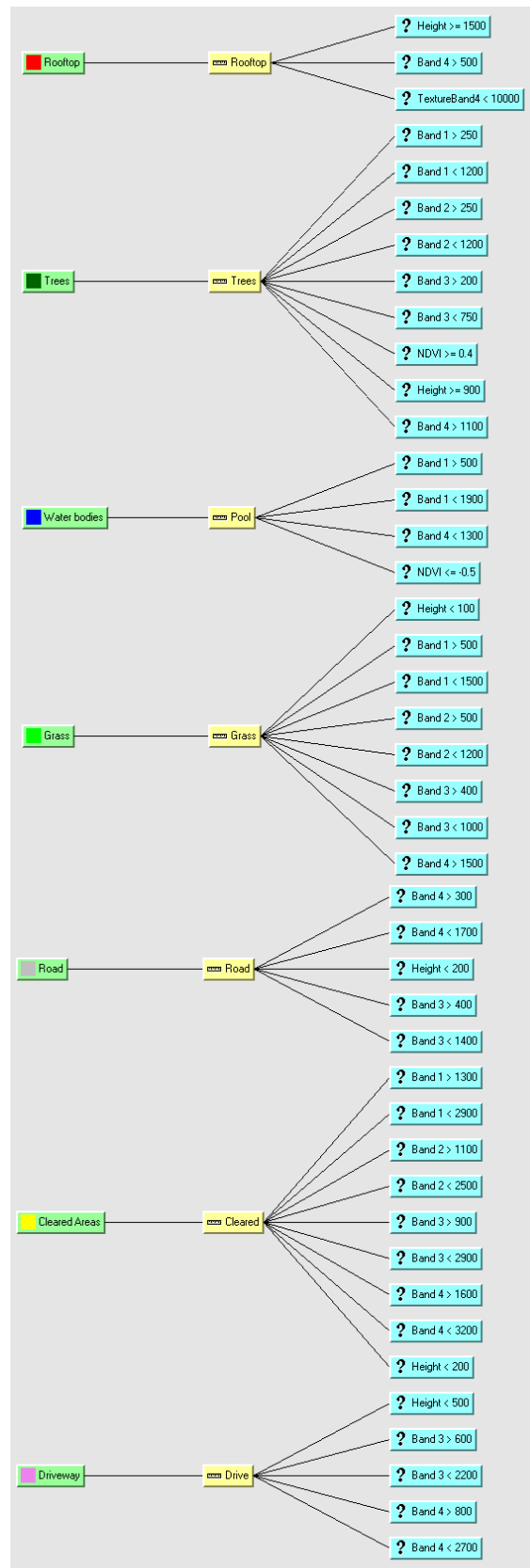


Figure 15 Initial decision rule tree

Figure 15 displays the initial decision rule tree developed for the classification. Driveway was added as an extra land cover class.



**Figure 16** The produced classified image from the initial decision rule tree

Figure 16 shows the classified image produced from the initial decision rule tree. Some problems visible in the image include misclassification of driveway and tree canopies classified as rooftops. The problem with the initial decision rule tree will be covered further in Section 5.1.

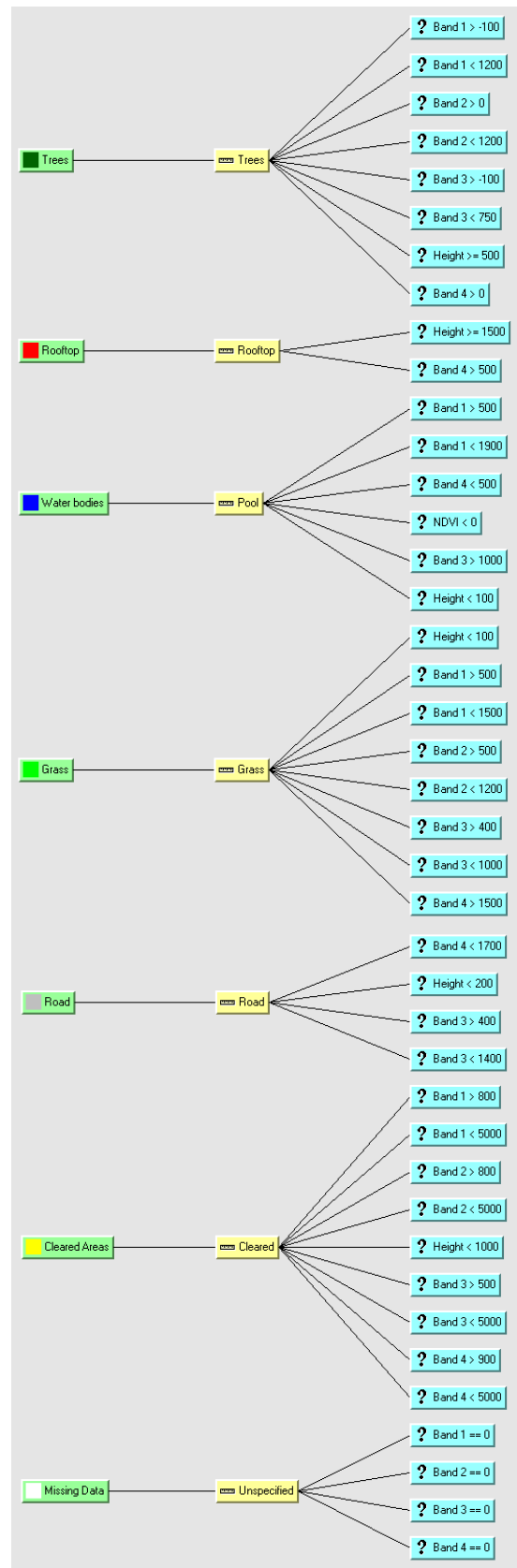


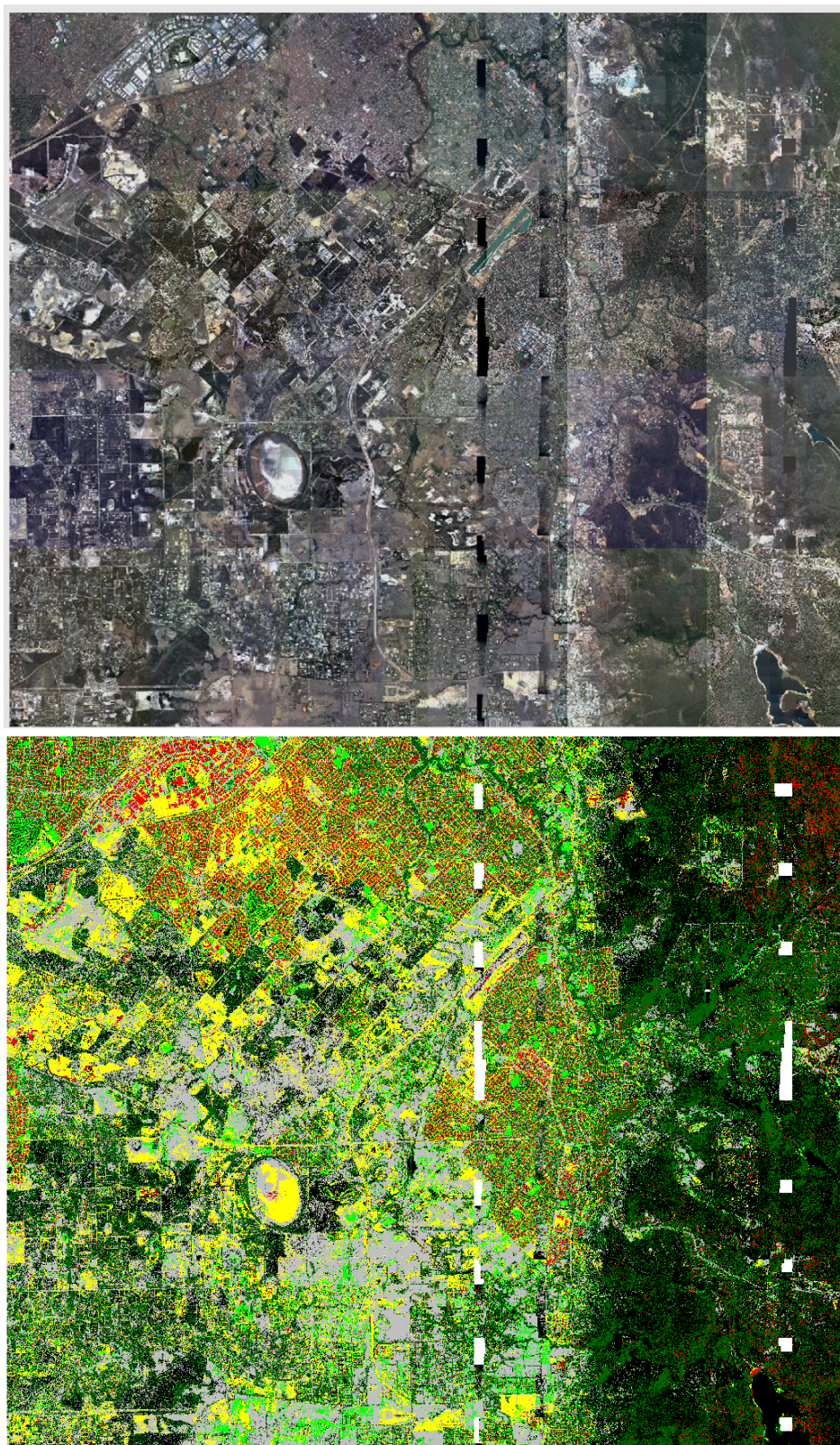
Figure 17 Final decision rule tree

As seen in Figure 17, some changes were made to the decision rule tree. Driveway land cover class was removed and the order of the classification was changed. The conditions of each land cover were also altered. The decisions taken to arrive at the final decision rule tree are discussed in Section 5.1.



Figure 18 the Produced classified image from the final decision rule tree

Figure 18 shows the produced classified image from the final decision rule tree. The improvement of the accuracy may be difficult to confirm but the final decision rule tree covers the diversity of the study area as much as possible. Section 5.1 will discuss the problem faced upon developing the final decision rule tree. The larger version of the classified image can be seen in Appendix II.



**Figure 19** Classified image of the study area

Figure 19 displays the result of mosaicking all the diced images which have been classified. This image is practically the classified image of the study area.

Table 3 Error of omission and commission

Classified Data	Reference Data								Row total
	Undefined	Water	Grass	Road	Cld Area	Trees	Rooftop	Miss.Data	
Undefined	0	0	3	1	8	9	0	0	21
Water bodies	0	0	0	0	0	0	0	0	0
Grass	0	0	6	0	3	0	0	0	9
Road	0	1	2	4	8	2	0	0	17
Cleared Areas	0	0	2	2	13	0	0	0	17
Trees	0	0	2	0	1	19	4	0	26
Rooftop	0	0	0	0	0	0	3	0	3
Missing Data	0	0	0	0	0	0	0	3	3
Column Total	0	1	15	7	33	30	7	3	96

Land Cover	Reference	Classified	Number Correct	Producers Accuracy	Omission Error	User Accuracy	Commission Error
Undefined	0	21	0	---	---	---	---
Water bodies	1	0	0	---	---	---	---
Grass	15	9	6	40.00%	60.00%	66.67%	33.33%
Road	7	17	4	57.14%	42.86%	23.53%	76.47%
Cleared Areas	33	17	13	39.39%	60.61%	76.47%	23.53%
Trees	30	26	19	63.33%	36.67%	73.08%	26.92%
Rooftop	7	3	3	42.86%	57.14%	100.0%	0.00%
Missing Data	3	3	3	100.00%	0.00%	100.0%	0.00%
Totals	96	96	48				
Overall Classification Accuracy		50.00%					
Overall Kappa Statistics		0.393					

It was observed that the omission error is above 50% for grass, cleared areas and rooftops. In other words, the number of correctly classified points are too small than there should be for grass, cleared areas and rooftops. This is proven by the low producer's accuracy. For example looking down the grass column of Table 3, there are only 6 points correctly classified as grass while in total there are 15 points which led to the low producer's accuracy of 40%.

The commission error was particularly high for roads, which is more than 70%. This suggests that too many points are classified as roads where in fact they are not. Looking across the road row, out of 17 points classified as roads only 4 of them are actually roads. This results in low user's accuracy of 23.53%.

Overall, the classification accuracy is 50% and the Kappa value is 0.393.

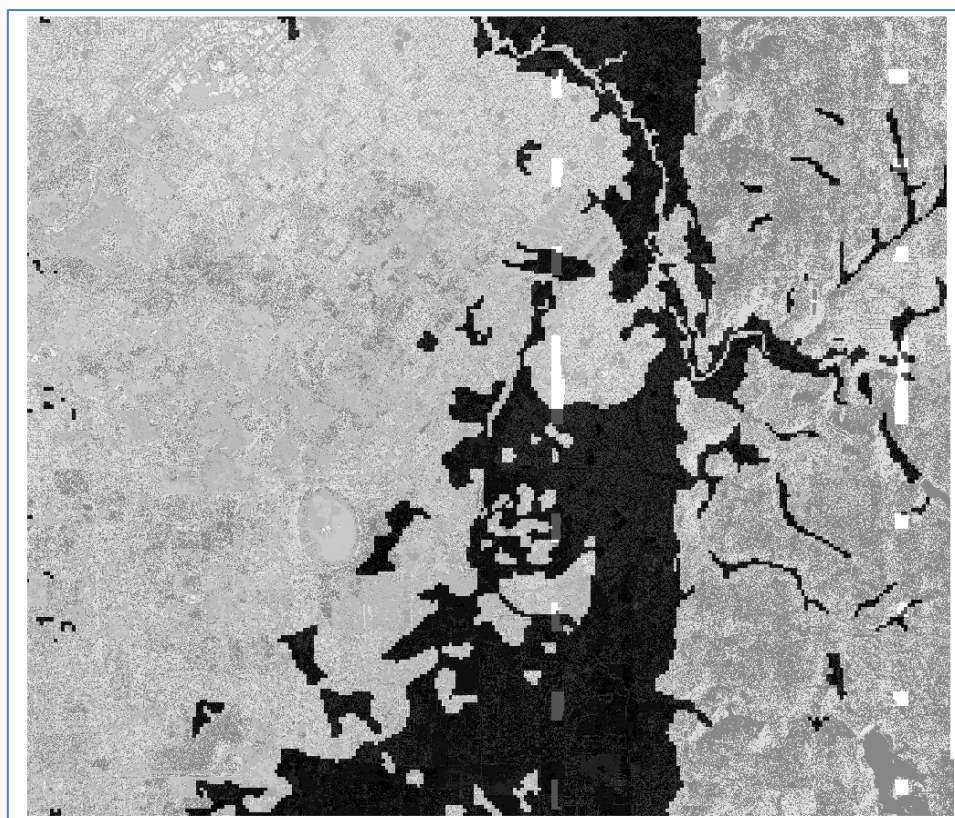


Figure 20 Refined map of the classified image using the soil map as ancillary data

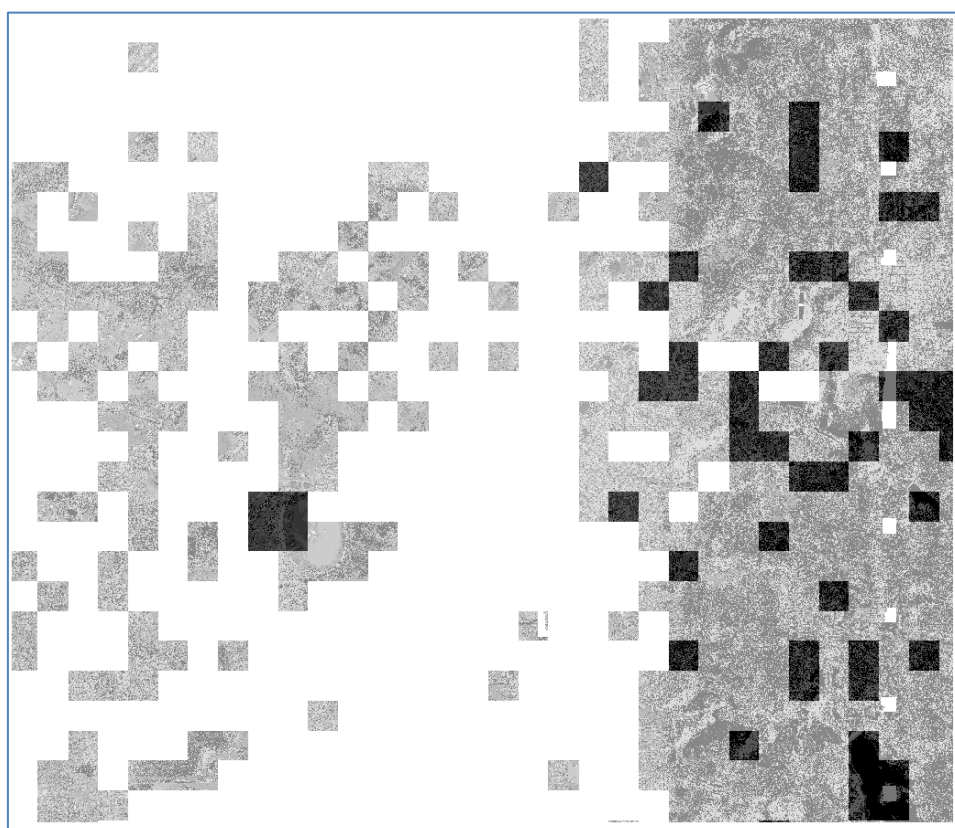


Figure 21 Refined map of the classified image using the vegetation map as ancillary data

Figure 20 and Figure 21 shows the refinement of the classification using the soil map and the vegetation map. The dark region in Figure 20 indicates the clay soil while the remaining is on non-clay soil. Therefore, all land cover within the dark region is built on clay soil. The dark region in Figure 21 represents the non-vegetation objects while the grey colour is other native trees. The white area represents the unspecified area of the image which is lost during the overlaying of the map. This will be discussed further in Section 0.

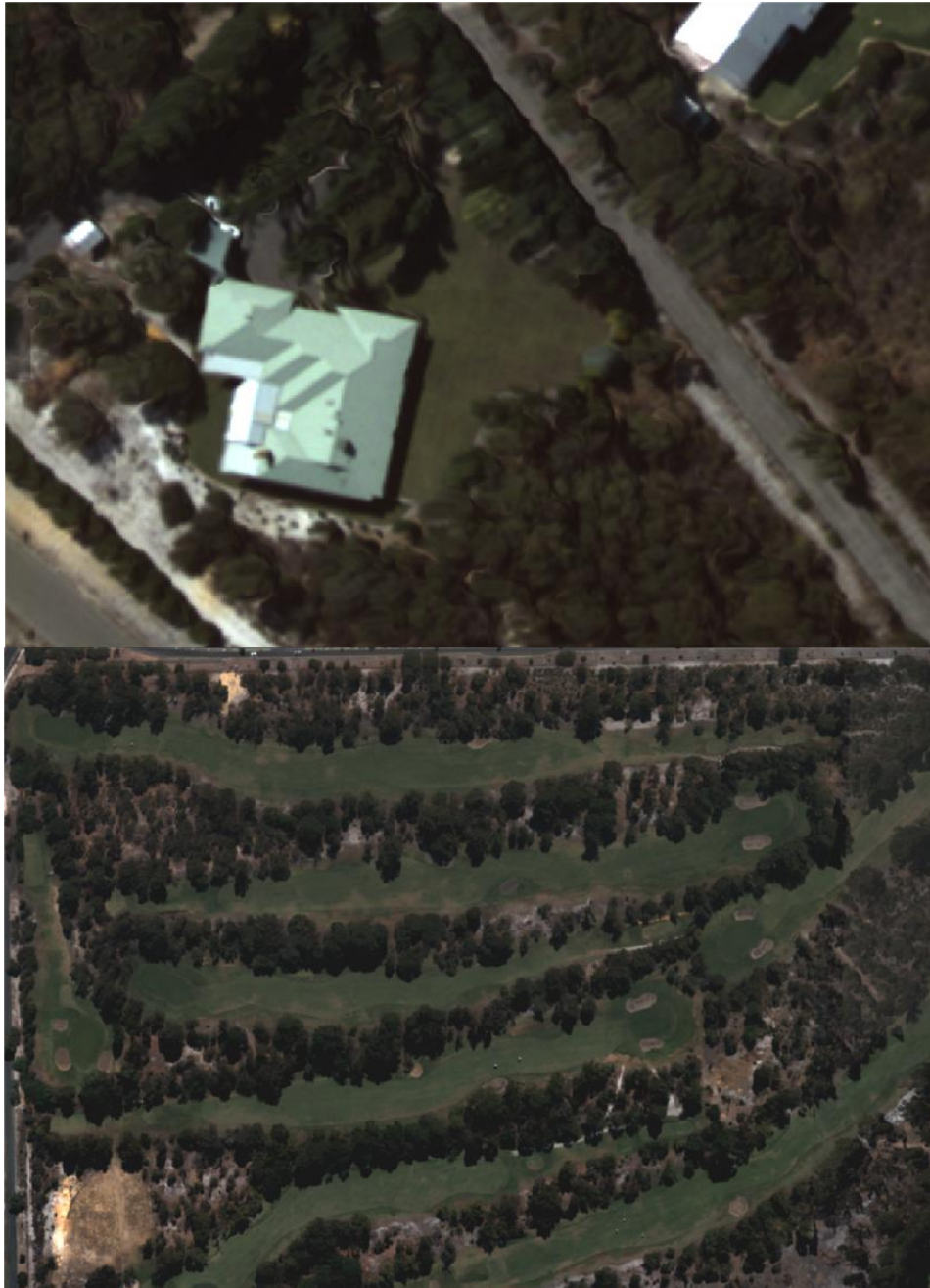
## 5. DISCUSSION

### 5.1. IMPROVEMENT OF CLASSIFICATION ACCURACY

Land cover classification of high resolution imagery poses a major challenge in terms of the accuracy of the classification. In this project, the challenges were in dealing with the diversity of the land cover and data quality which eventually compromised the accuracy of the classification.

#### 5.1.1. DATA QUALITY

It was observed that there is inconsistency with the quality of the image. Figure 22 shows a distortion of the tree which complicates the classification process. The imagery is collected as a series snapshot which were then mosaicked together. Differences in ER reflectance and airplane configuration results in the full scene of the Perth region becomes a patchwork of varying spectral signatures for the same feature. This is in spite of orthorectification, radiometric correction and colour balancing that has been done (Collings et al. 2011). It can be seen clearly in Figure 23 that the patches problem present upon stitching the snapshots together. Such problem can be seen throughout the whole image of the study area as seen in Figure 19.



**Figure 22 Distortion error of trees as compared to the normal image**

This was a problem because it has inconsistent spectral band values. On the other hand, it can still be visually confirmed as trees. Decision has to be made whether to still classify these land cover features in spite of their altered values or to just regard them as an error. In the end, the rules was generalised to cover these discrepancy. However, a certain threshold was still needed so that the conditions of the land cover are not overlapping with each other.

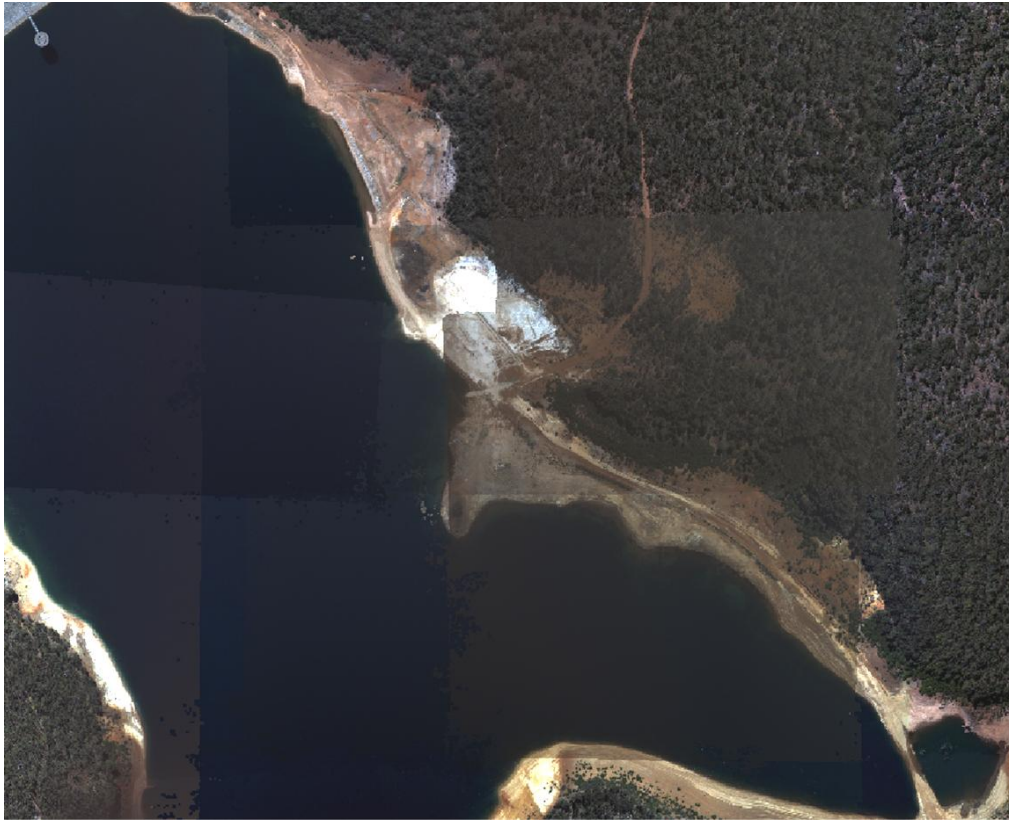


Figure 23 Inconsistency in image quality showing patches of images of different spectral signature

Moreover, there is also a problem of missing data in the image as seen in Figure 24. The black rectangle in the middle of the image is actually a missing data. The imagery was made by stitching together thousands of aerial photographs. Equipment noise and error in forms of variation and drift is quite common (Collings et al. 2011). Due to unforeseen circumstances the photos do not overlap and therefore a few patches of ground have been missed. In the end, these missing data was classified purely as a separate feature class as seen in Figure 17 Final decision rule tree.



Figure 24 Missing data within an image

There is also an error with the nDSM as a result of flaws from the original image. Figure 25 displays an error with the nDSM image which caused the rooftop to be undetected. Some parts of the rooftop are black which indicated that it is on the ground. However, the original image on the left shows that the rooftop is normal. As a result, it was classified as a cleared area (yellow colour) as seen in the right image.

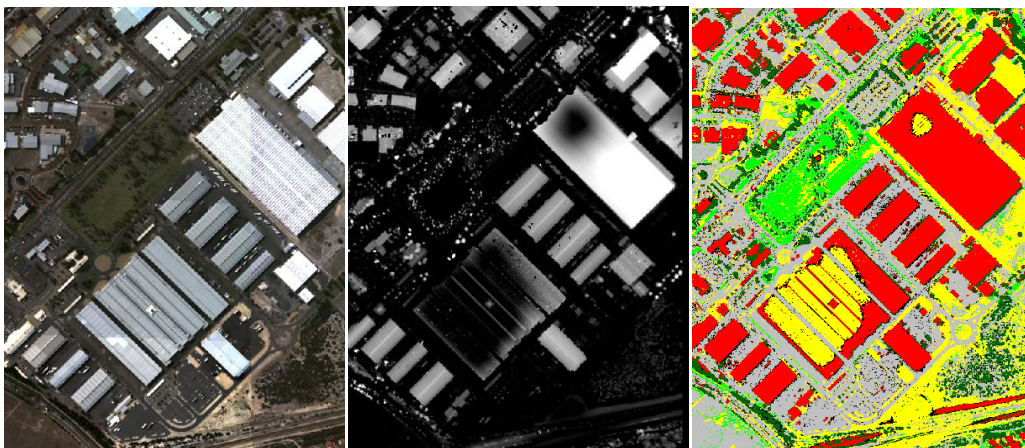


Figure 25 Error with the nDSM image

There is also a problem with the NDVI image. The NDVI is not able to pick up all the vegetations completely which they supposedly have to. Also, some red colour rooftops are identified by the NDVI as vegetation. Figure 26 is showing the overlaying of NDVI image on top of its original image. As

seen in the figure, some of the tree canopies were not identified as vegetation (white colour). Figure 27 is showing that red coloured rooftops were identified as vegetation instead. This might be because ratio-based index can be influenced by additive noise effects such as atmospheric path radiance (Huete et al. 2002). Also, NDVI is very sensitive to canopy background variations such as soil visible through the canopy (Huete et al. 2002). Therefore, the usefulness of the NDVI is quite limited.



Figure 26 NDVI image showing some of the tree canopies were not captured as vegetation

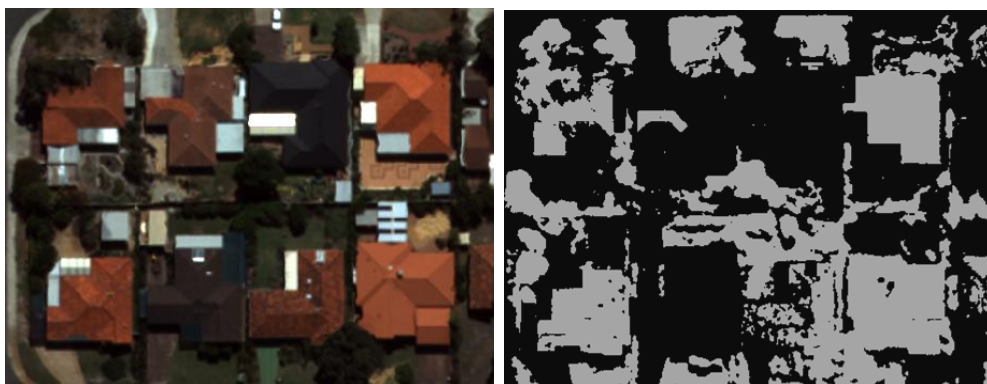


Figure 27 NDVI image showing some red colour rooftops were identified as vegetation

### 5.1.2. HETEROGENEITY OF LAND COVER

As mentioned in Section 3.2.2, the setting up of the rule tree in the Knowledge Engineer was done by simple interval measurements of the values of the 3 variables, spectral band values, NDVI and nDSM. It is also mentioned that the value interval measurement for a land cover feature has to represent the diversity of that specific feature. However, the initial rule tree was developed using only a Section of the original image of the study area. As a result, the developed knowledge base did not produce satisfying result when applied to the whole area. Upon observing all the diced images, it was observed that each diced image has different variability of land cover which has not been covered by the initial rule tree. As can be seen in Figure 28, the colour of the rooftops varied greatly. This image is only a subsection within a diced image. It is also important to note that there is a challenge of classifying both the complex urban areas and the rather uniform vegetated areas.



Figure 28 Image showing variability of the colour of the rooftops

There was a further problem with the order of the classification. As mentioned before, the order of the classification in the Knowledge Engineer matters. A class that is specified first will not be reclassified by the other class that follows even if the rules are overlapping. Fundamentally, this means that the conditions for a land cover feature that comes first need to be more specific and restricting than those class that follows. Therefore going down the rule tree, the rules can be more lenient and therefore

have larger value interval. Conclusively, land cover that is less heterogeneous should be classified first so that the land cover with more variability can be more easily classified. This is summarized in Figure 29 below.

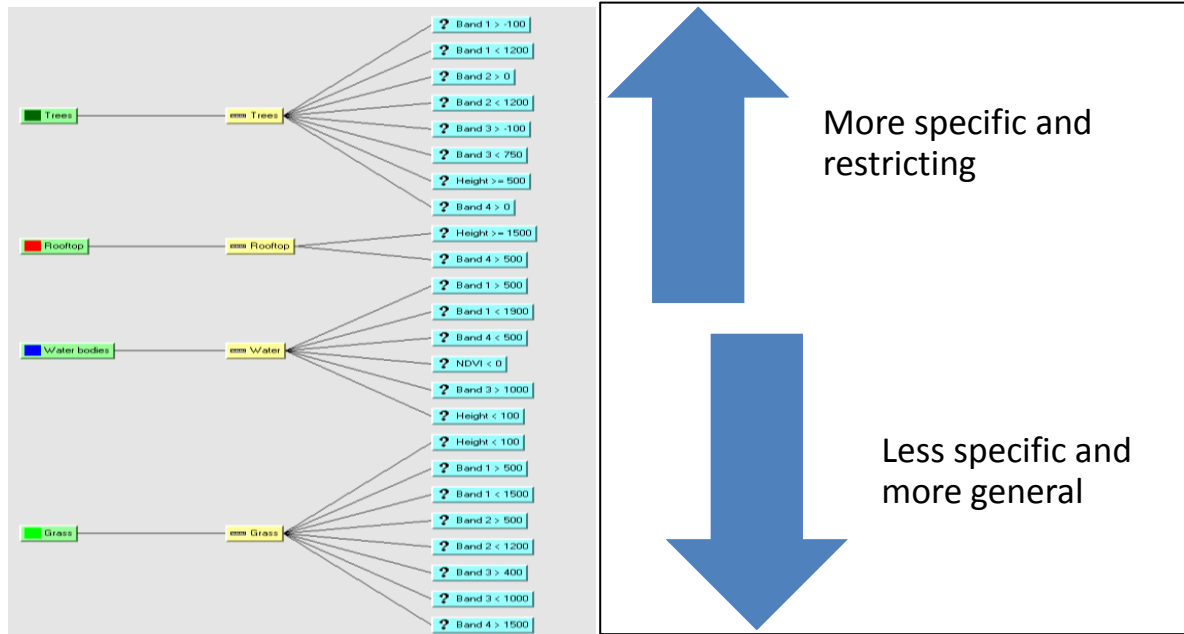


Figure 29 Deciding orders of classification and the resulting rules and conditions

There were two clashes of feature classes which are difficult to ascertain which one should be classified first. The first problem was between roads and cleared areas. Both are non-vegetation features on the ground which means NDVI and nDSM cannot be used to differentiate them. Spectral band values naturally were the only option. Although their colours are quite distinct, the road class in this case includes all impervious surfaces such as driveways and road pavements which can be quite similar in colour. Creating a separate feature class will not improve the classification since these features still need to be classified as either roads or cleared areas. This is why driveway was taken out from the initial decision rule tree. Moreover, there was also a problem of inconsistency of the image quality as discussed before. This was the main reason that roads are classified before cleared areas. The rules for cleared areas need to be more general in order to encompass the imperfect part of the cleared areas.

The second problem was between rooftops and trees. Principally this should be straight-forward since one is vegetation while the other is not although both are features above the ground. The idea of using NDVI was to distinguish between vegetation and non-vegetation features. However in this project, the NDVI images had an issue as discussed before. If NDVI was used, the features that are not picked up will be classified as land cover which is specified second. For example, some of the red rooftops that were identified as vegetation in the NDVI image will be classified as tree. Again in this case, spectral

band values were the only option. The problem of inconsistent image quality persists for the trees as well. Some of the trees are wrongly represented as having negative spectral band values as seen in Figure 30. Similarly, the rules developed for trees have to be generalised to encompass these flaws. However, in this case the trees are classified first before the rooftops simply because there is too large of variability in the colour of the rooftops.

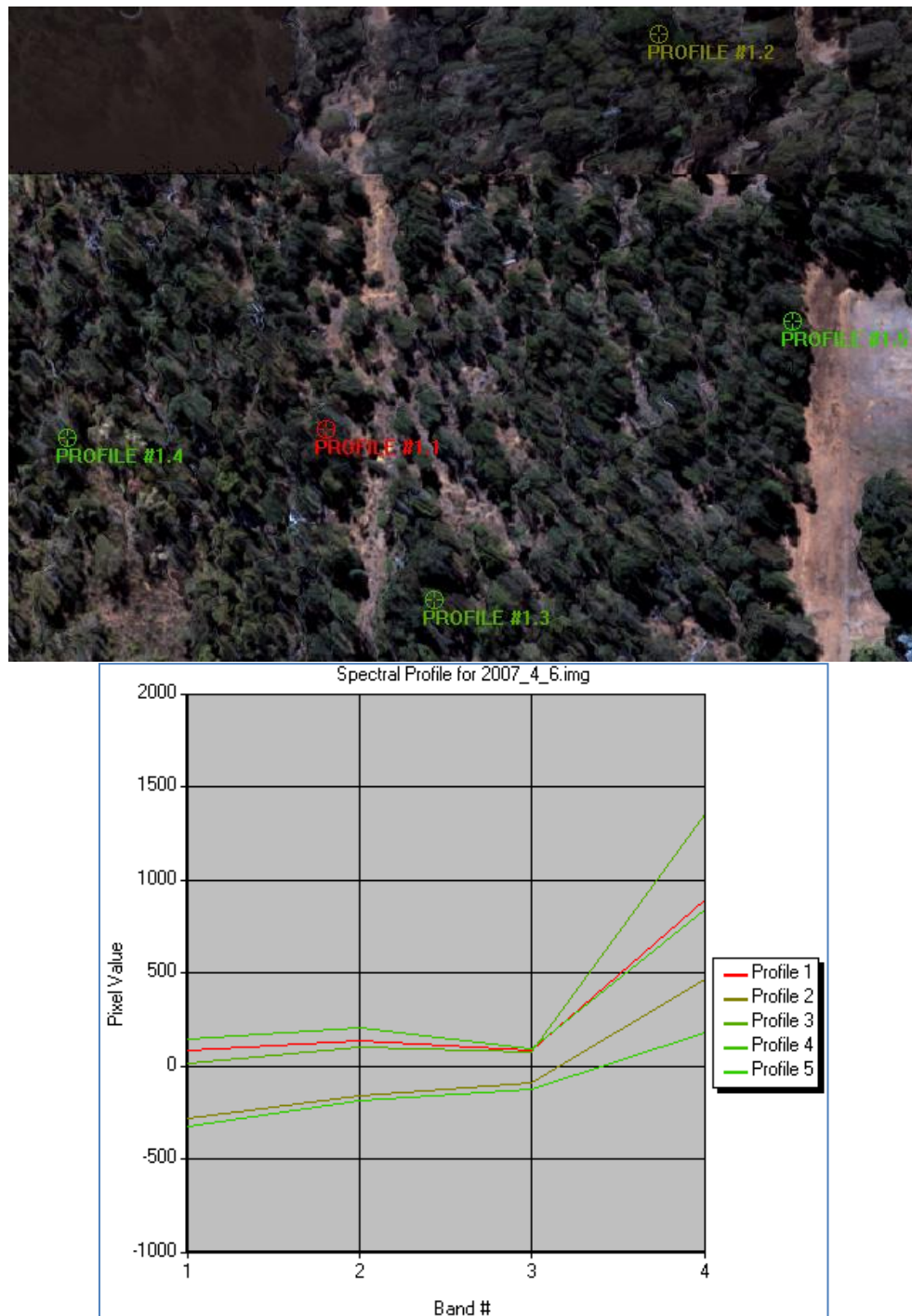


Figure 30 Graph showing the spectral band values of the distorted trees

Therefore, the final rule tree was developed by taking into account the diversity of each land cover features in terms of their variable values. This was done by closely observing each of the land cover feature in all the dived images through their spectral and spatial profile. The rules and conditions of the land covers were then adjusted accordingly. The output images were compared to see if the changes produced better classification.

The considerations above explain the final decision tree rule developed as seen in Figure 17. There is no significant difference observed in classifying the above ground or on the ground feature first. Therefore, the features above the ground were classified first. For the on the ground features, water bodies were classified first simply because the unique spectral band values and the utilisation of NDVI variable. Grass was next also because of its characteristically unique spectral band 3 and 4 values.

However, after all the efforts of trying to improve the accuracy of the classification, the overall accuracy was 50% as shown in Table 3. Compared to other land classification study, the accuracy result of this project is rather insubstantial. A land use/land cover classification for a watershed in Ohio has an accuracy of 67% (Ahn et al. 2008). In that study, urban lawn was misclassified as agricultural land and oppositely, pasture and bright soils were classified as urban land. A land cover change study of Southern Guam using supervised classification from same remote sensing software ERDAS Imagine is able to produce accuracy of 90.42% (Wen et al. 2011).

Analysing the error matrix and table of error of omission and commission in Table 3, it can be highlighted where the misclassification lies between the land covers. For roads which have low user's accuracy of 23.53%, it can be seen that the 8 points are misclassified as roads when it should be cleared areas. The producer's accuracy for cleared areas is also rather low, which is 39.39%. There are only 13 points correctly classified as cleared areas while there are 33 points. 8 points were undefined while they should be classified as cleared areas. This suggests that despite the effort made to solve the clash between roads and cleared areas, the problem remains.

Effort of solving the clash between trees and rooftops is showing quite respectable result. The producer's and user's accuracy for trees were 63.33% and 73.08% respectively. However, the table shows that there are still 4 points of rooftops misclassified as trees.

The high number of undefined points shows that the rules and conditions did not cover all features of the study area although again it can be attributed to the image quality problem. 21 out of 96 points (around 22%) were undefined when they should have been classified as any of the 6 land covers classes. 9 points should be classified as trees while 8 points should be classified as cleared areas. This

proves that the accuracy of the classification still needs to be improved although efforts have been made to improve them.

However, the accuracy result of this project is rather justifiable considering these following reasons.

1. All the discussions mentioned above explain that certain decisions have to be made while compromising the accuracy of the classification. Classifying the roads first leads to some of cleared areas classified as roads. Similarly, classifying the trees first leads to some of rooftops classified as trees. In both cases, the accuracy was compromised.
2. The issue with the quality of the image means that switching the order of the classification is unlikely to produce commensurately better result. There are issues with all three variables used which all contribute to the decrease of accuracy. It is quite reasonable considering that 2007 was the first time that the digital aerial imagery was collected over the Perth region. Therefore defects and imperfections were expected.
3. Visual inspection of the accuracy assessment. Again due to quality of the image, confirming what actually appears on the surface of the earth is quite instinctive. Even at points that were clear, it was quite difficult to tell what the land cover was in spite of the high spatial resolution. This obviously leads to more decrease in accuracy.

Looking at the final classified image of the whole study area, the classification worked particularly well at dense urban residential areas. It can be seen in Figure 31 that the land covers were identified quite accurately.



**Figure 31** Classified image of dense residential area

However, the classification did not work well for the undisturbed vegetated area which is rather uniform. It might be justified by the inconsistent image quality as discussed above. As shown in Figure 32 some of tree canopies were somehow classified as rooftops. Also, the soil underneath the vegetation was classified as undefined which is the black colour.

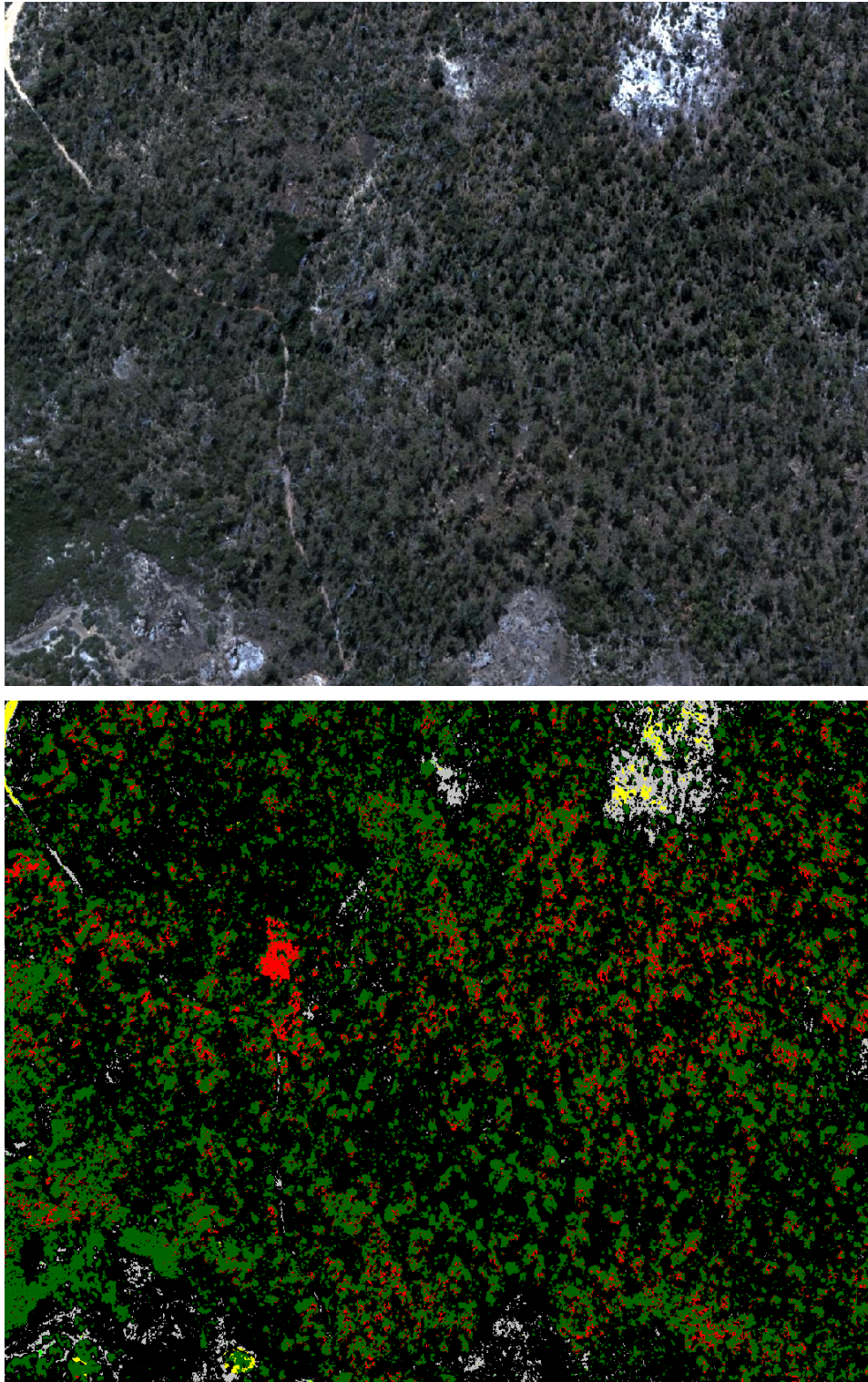


Figure 32 Classified image of undisturbed vegetated area

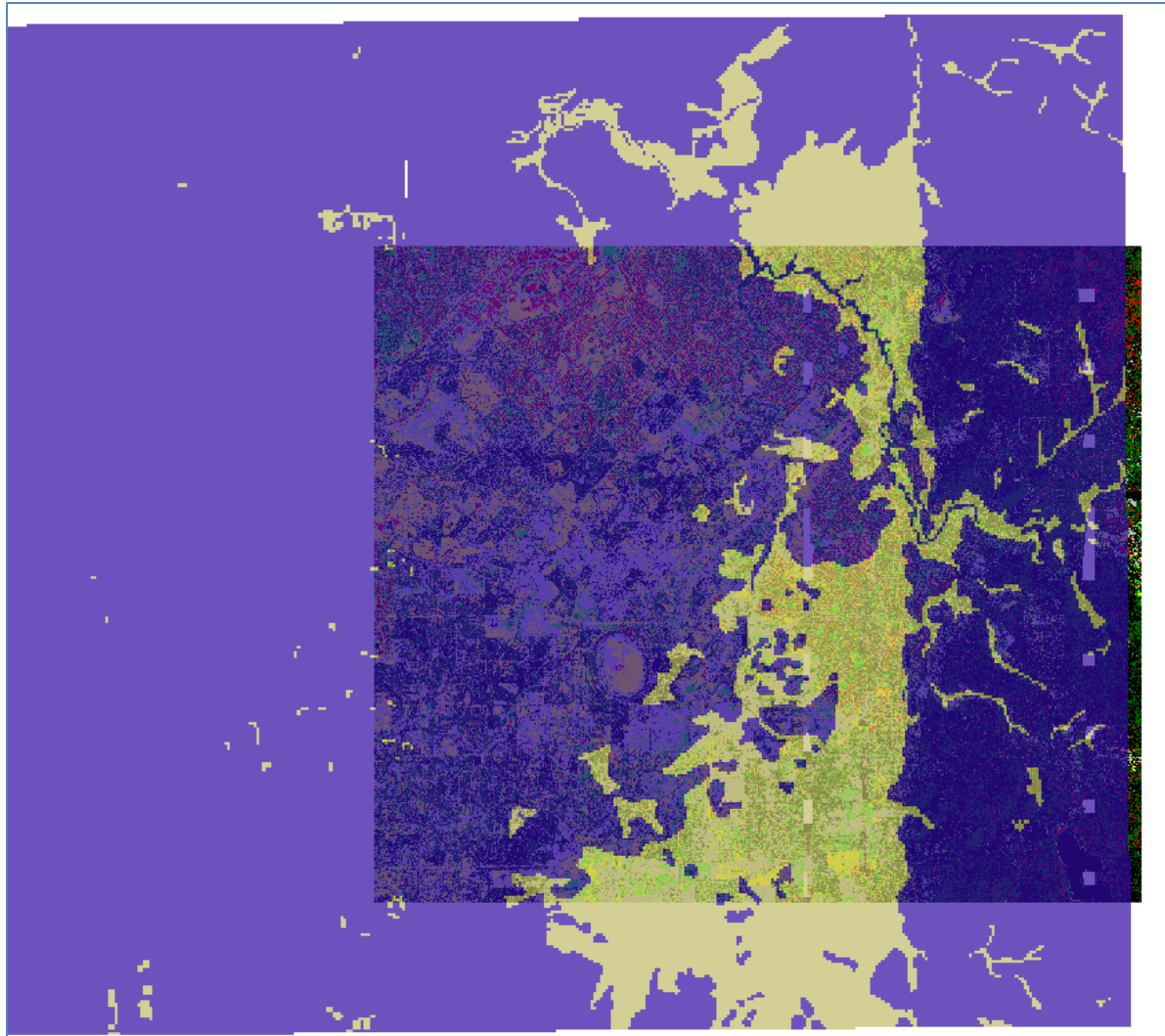
## 5.2. URBAN WATER BALANCE

Table 4 Percentage and total area of land cover

Land cover	Total area (km <sup>2</sup> )	%
undefined	90.61	23.97
water bodies	0.06	0.01
grass	26.82	7.10
roads	88.27	23.35
cleared areas	51.67	13.67
trees	96.99	25.66
rooftop	20.80	5.50
missing data	2.78	0.73
TOTAL	378.00	100.00
unknown trees / non-vegetation	7.53	7.76
other native trees (non-banksia)	47.68	49.16
TOTAL	55.21	56.92
rooftop on clay soil	4.16	20.00
rooftop on non-clay soil	16.18	77.81
TOTAL	20.34	97.81
cleared areas on clay soil	10.04	19.43
cleared areas on non-clay soil	41.56	80.44
TOTAL	51.60	99.86

Table 4 above is showing the percentage and total area of each land cover. The top half is the result of the classification using expert classification while the bottom half is the result of refining the classification using ArcGIS.

Putting the urban water balance in perspective, percentage of land cover is important in defining how much of the precipitation will fall on which land cover (McFarlane 1984). By looking at the table, it can be defined that 0.01% of the precipitation will fall on water bodies, 7.1% on grass, 23.35% on roads, 13.67% on cleared areas, 25.66% on trees and 5.5% on rooftops. This is assuming that precipitation falls vertically and evenly across all surfaces. From the 5.5% of rooftop land cover, 20% of the precipitation will fall on clay soil while the rest is on non-clay soil. From the 13.67% of cleared areas, 19.43% of the precipitation will fall on clay soil while the rest is non-clay soil. However, from the 25.66% of trees, it was found that none of them is actually Banksia woodlands while 7.76% is unknown trees or non-vegetation while the rest is other native trees.



**Figure 33 Mismatch of the soil map and the classified image**

There are some limitations of the result as seen at the Table 4 above. In refining the classification further, the total area of the rooftops and the trees did not account to 100% (97.8% and 56.9% respectively). There is an unaccounted loss of area in the map which is due to missing Sections of the geology and vegetation map. The geology map does not cover the study area completely as shown in Figure 33. The soil map is missing a section on the right hand part of the figure. Since there is no soil information for the classified image on this missing Section, it was cut during the overlaying process in ArcGIS. Similarly, the vegetation map does not define the non-vegetation and non-tree areas comprehensively as seen in Figure 12. The white areas in the figure are unspecified data which actually may be non-vegetation objects and therefore were lost during the process. This results in patches of missing data as shown in Figure 21. Also, the vegetation map was produced in 2002. In the span of 5 years, there may have been removal of the vegetation in the area which might explain why none of them is Banksia woodland. Note that the spatial resolution of the vegetation map is quite large as seen in Figure 21. This would mean a certain degree of generalization of the tree species

within a cell. Therefore, the usefulness of the result from the refinement of classification is quite restricted.

However, this project only evaluate thus far. For instance, defining recharge from a unit area of a roof or evapotranspiration from a unit area of trees and grass require a further study which is not covered in this project. The supplementary study may include variation in tree species and age for defining interception loss, through fall and evaporative loss from canopies and also study of the rooftops in defining whether the water goes to drains, soak wells or soil. This project limitedly ascertains lumped fluxes of inputs to the urban water balance and not the complete water balance itself. There are other fluxes which are not estimated such as groundwater and bore water use which also contributed to the storage change.

Nevertheless, recognizing the percentage and total area of the land cover alone can be useful.

Assuming that the classification accuracy is improved and the classification process can be done quickly, it could be used without considering the complete water balance. For example, it could benefit engineers designing stormwater systems or hydrogeologists doing groundwater modelling.

## 6. CONCLUSIONS AND RECOMMENDATIONS

In conclusion, land cover classification of high resolution imagery has been proven to be feasible. Percentage and total area of each land cover was the main result of the project. To a limited extent, the improvement of accuracy was proven to be successful although a great deal of work is still required to improve the classification accuracy.

This project is particularly significant in its application towards establishing inputs to urban water balance. Percentage of precipitation fall on each land cover was obtained as discussed in Section 5.2. While urban water balance has been determined through land cover before, land cover classification has never been done using high resolution imagery. This project provided the baseline for using land cover classification of high resolution imagery for future urban water balance studies. This project also provided basis for other studies to use high resolution imagery.

There are several recommendations for future studies of land cover classification of high resolution imagery. As has been discussed, classification accuracy requires more improvement. Given the current input data, iterative method of using the result of the accuracy assessment to improve the accuracy of the classification can be done although it will be quite time consuming. Colour balancing is also suggested to remove the variation of image quality. Logically, using better input data which has minimal quality problem would lead to higher accuracy of the classification. Also, methods of classifying cleared areas and roads more accurately has not been determined by this project.

Separate study could test the capability of classifying high resolution imagery on different area. Decision tree rules could be developed for urban areas and vegetation areas separately.

Further study which may use this project as the baseline could compare the land cover changes through time and their effects to the water fluxes and the urban water balance. Certainly, this requires improvement to the classification accuracy for the result to be considered significant.

## REFERENCES

- Ahn, G.-C., Merry, C.J. & Gordon, S.I., 2008. THE EFFECT OF URBANIZATION ON THE HYDROLOGIC REGIME OF THE BIG DARBY CREEK WATERSHED , OHIO. In *ASPRS 2008 Annual*. Portland, Oregon.
- Australian Government Department of Mines and Petroleum, 2010. GSWA data sets. *Western Australia*. Available at: [http://geodownloads.dmp.wa.gov.au/seriesmapping/digitalgeology\\_enh2.asp](http://geodownloads.dmp.wa.gov.au/seriesmapping/digitalgeology_enh2.asp) [Accessed April 27, 2012].
- Australian Government Department of the Environment and Water Resources, 2002. National Vegetation Information System (NVIS) Downloadable Data - Pre-European and Present Native Vegetation (Published Data - Stage 1, Version 2) National Land & Water Res Audit. *Metadata Publisher*. Available at: <http://www.environment.gov.au/erin/nvis/> [Accessed April 27, 2012].
- Barron, O., Barr, A., et al., 2011. Combined consideration for decentralised non-potable water supply from local groundwater and nutrient load reduction in urban drainage. *Water science and technology*, 63(6), pp.1289-97. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21436569> [Accessed May 3, 2012].
- Barron, O., Pollock, D. & Dawes, W., 2011. Evaluation of catchment contributing areas and storm runoff in flat terrain subject to urbanisation. *Hydrology and Earth System Sciences*, 15(2), pp.547-559. Available at: <http://www.hydrol-earth-syst-sci.net/15/547/2011/> [Accessed May 1, 2012].
- Barton, C.V.M., 2011. Advances in remote sensing of plant stress. *Plant and Soil*, pp.41-44. Available at: <http://www.springerlink.com/index/10.1007/s11104-011-1051-0> [Accessed March 29, 2012].
- Borel, C.C., 2010. Vegetative canopy parameter retrieval using 8-band data. In *DigitalGlobe 8-Band Research Challenge*. pp. 1-21.
- Burns, D. et al., 2005. Effects of suburban development on runoff generation in the Croton River basin, New York, USA. *Journal of Hydrology*, 311(1-4), pp.266-281. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0022169405000703> [Accessed May 1, 2012].
- Caccetta, P., Collings, S. & Hingee, K., 2011. Fine-Scale Monitoring of Complex Environments Using Remotely Sensed Aerial, Satellite, and Other Spatial Data. Available at: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6024217](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6024217) [Accessed March 15, 2012].
- Caccetta, P., McFarlane, D., et al., 2011. URBAN MONITOR- FINE-SCALE MONITORING OF COMPLEX ENVIRONMENTS. In *7th International Symposium on Digital Earth*. Perth.
- Collings, S. et al., 2011. Empirical Models for Radiometric Calibration of Digital Aerial Frame Mosaics. *IEEE Transactions on Geoscience and Remote Sensing*, 49(7), pp.2573-2588. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5729323>.

- Contreras, S. et al., 2011. Remote sensing estimates of supplementary water consumption by arid ecosystems of central Argentina. *Journal of Hydrology*, 397(1-2), pp.10-22. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0022169410007109> [Accessed May 29, 2012].
- Dar, I. a, Sankar, K. & Dar, M. a, 2011. Deciphering groundwater potential zones in hard rock terrain using geospatial technology. *Environmental monitoring and assessment*, 173(1-4), pp.597-610. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20364317> [Accessed May 29, 2012].
- Dujardin, J. et al., 2011. Improving surface-subsurface water budgeting using high resolution satellite imagery applied on a brownfield. *The Science of the total environment*, 409(4), pp.800-9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21112074> [Accessed May 29, 2012].
- ERDAS, 2001. ERDAS IMAGINE Expert Classifier Overview. *Imagine*. Available at: <http://www.gis.usu.edu/unix/imagine/ExpertClassifier.pdf>.
- ERDAS, 1999. Expert On-line Manual. *Imagine*. Available at: <http://std.kku.ac.th/5130205932/Remote/Lab01/ERDAS8.4/Documentation/Intro.pdf>.
- ERDAS, 2009. Feature Extraction & Classification. *Knowledge Creation Diffusion Utilization*. Available at: [http://www.erdas.com/Libraries/White\\_Papers/Feature\\_Extraction\\_and\\_Classification.sflb.ashx](http://www.erdas.com/Libraries/White_Papers/Feature_Extraction_and_Classification.sflb.ashx).
- de Espindola, G.M. et al., 2011. Agricultural land use dynamics in the Brazilian Amazon based on remote sensing and census data. *Applied Geography*, 32(2), pp.240-252. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0143622811000634> [Accessed March 5, 2012].
- Geerken, R., Zaitchik, B. & Evans, J.P., 2005. Classifying rangeland vegetation type and coverage from NDVI time series using Fourier Filtered Cycle Similarity. *International Journal of Remote Sensing*, 26(24), pp.5535-5554. Available at: <http://www.tandfonline.com/doi/abs/10.1080/01431160500300297> [Accessed March 11, 2012].
- Geospatial Information Authority of Japan, 2010. DSM figure. Available at: <http://www.gsi.go.jp/WNEW/TEC-NEWS/2007-tec172.html> [Accessed May 30, 2012].
- Her, Y., 2007. Land use classification in Zambia using Quickbird and Landsat imagery. In *2007 ASABE Annual International Meeting*. Minneapolis.
- Huete, A. et al., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, pp.195-213.
- Jensen, J.R., 2005. *Introductory Digital Image Processing: A remote sensing perspective* 3rd ed. K. C. Clarke, ed., New Jersey: Prentice-Hall Series.
- Kenway, S., Gregory, A. & McMahon, J., 2011. Urban Water Mass Balance Analysis. *Journal of Industrial Ecology*, 15(5), pp.693-706. Available at: <http://doi.wiley.com/10.1111/j.1530-9290.2011.00357.x> [Accessed March 15, 2012].

- Maingi, J.K. et al., 2002. *An Accuracy Assessment of 1992 Landsat-MSS Derived Land Cover for the Upper San Pedro Watershed ( U . S ./ Mexico ) An Accuracy Assessment of 1992 Landsat-MSS Derived Land Cover for the Upper San Pedro Watershed ( U . S ./ Mexico )*, Washington.
- McFarlane, D., 1984. *The effect of urbanisation on groundwater quantity and quality in Perth, Western Australia*. University of Western Australia.
- Mcgee, T., 2001. Urbanization Takes on New Dimensions in Asia's Population Giants. *Population Reference Bureau*, 29(7), pp.1-2.
- Mitchell, G.V., McMahon, T.A. & Mein, R.G., 2003. Components of the total water balance of an urban catchment. *Environmental Management*, 32(6), pp.735-746.
- Montzka, C. et al., 2008. Modelling the water balance of a mesoscale catchment basin using remotely sensed land cover data. *Journal of Hydrology*, 353(3-4), pp.322-334. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0022169408001042> [Accessed May 15, 2012].
- Moore, G., 2001. *SOILGUIDE A HANDBOOK FOR UNDERSTANDING AND MANAGING AGRICULTURAL SOIL* 4343rd ed., Agriculture Western Australia. Available at: [http://www.agric.wa.gov.au/objtwr/imported\\_assets/content/lwe/rpm/landcap/soilguide\\_introduction.pdf](http://www.agric.wa.gov.au/objtwr/imported_assets/content/lwe/rpm/landcap/soilguide_introduction.pdf).
- Peranginangin, N. et al., 2004. Water accounting for conjunctive groundwater/surface water management: case of the Singkarak–Ombilin River basin, Indonesia. *Journal of Hydrology*, 292(1-4), pp.1-22. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0022169403005237> [Accessed May 29, 2012].
- Rose, S. & Peters, N.E., 2001. Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): a comparative hydrological approach. *Hydrological Processes*, 15(8), pp.1441-1457. Available at: <http://doi.wiley.com/10.1002/hyp.218> [Accessed April 7, 2012].
- Rozenstein, O. & Karnieli, A., 2011. Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. *Applied Geography*, 31(2), pp.533-544. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0143622810001517> [Accessed March 15, 2012].
- Senseman, G.M., Bagley, C.F. & Tweddale, S.A., 1995. *Accuracy Assessment of the Discrete Classification of Remotely-Sensed Digital Data for Landcover Mapping*,
- Syme, G. & Nancarrow, B., 2011. *Overcoming water scarcity in Perth, Western Australia in Water Resources Planning and Management* R. Grafton & K. Hussey, eds., Cambridge: Cambridge University Press.
- Taweek, S., 2005. EXPERT CLASSIFICATION TECHNIQUE FOR MAPPING TEAK PLANTATION AREAS IN THAILAND. In *Pecora 16 "Global Priorities in Land Remote Sensing."* Sioux Falls.
- Uddin, K. & Gurung, D.R., 2010. Land cover change in Bangladesh- a knowledge based classification approach. In *10th International Symposium on High Mountain Remote Sensing Cartography*. pp. 41-46.

- Wen, Y., Khosrowpanah, S. & Heitz, L., 2011. Land cover change of watersheds in Southern Guam from 1973 to 2001. *Environmental monitoring and assessment*, 179(1-4), pp.521-9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21072586> [Accessed May 30, 2012].
- Whiteside, T. & Ahmad, W., 2008. ESTIMATING CANOPY COVER FROM EUCALYPT DOMINANT TROPICAL SAVANNA USING THE EXTRACTION OF TREE CROWNS FROM VERY HIGH RESOLUTION IMAGERY. In *Proceedings of GEOBIA 2008 - Pixels, Objects, Intelligence: GEOgraphic Object-Based Image Analysis for the 21st Century*. Calgary.
- Wu, J., Wang, D. & Bauer, M.E., 2007. Assessing broadband vegetation indices and QuickBird data in estimating leaf area index of corn and potato canopies. *Field Crops Research*, 102(1), pp.33-42. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0378429007000160> [Accessed March 3, 2012].
- Wu, X., Collings, S. & Caccetta, P., 2010. BRDF AND ILLUMINATION CALIBRATION FOR VERY HIGH RESOLUTION IMAGING SENSORS. In *Geoscience and Remote Sensing Symposium*.
- Yu, H.-yang et al., 2011. Object oriented land cover classification using ALS and GeoEye imagery over mining area. *Transactions of Nonferrous Metals Society of China*, 21, p.s733-s737. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1003632612616719> [Accessed May 28, 2012].
- Yu, Q. et al., 2006. Object-based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery. *Photogrammetric Engineering & Remote Sensing*, 72(7), pp.799-811.
- Zencich, S. et al., 2002. Influence of groundwater depth on the seasonal sources of water accessed by Banksia tree species on a shallow, sandy coastal aquifer. *Oecologia*, 131(1), pp.8-19. Available at: <http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1007/s00442-001-0855-7> [Accessed May 24, 2012].
- Zhao, N., Currit, N. & Samson, E., 2011. Net primary production and gross domestic product in China derived from satellite imagery. *Ecological Economics*, 70(5), pp.921-928. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0921800911000097> [Accessed May 29, 2012].

## APPENDIX I



Figure 34 A subSection of the study area showing dense urban area and some vegetated area

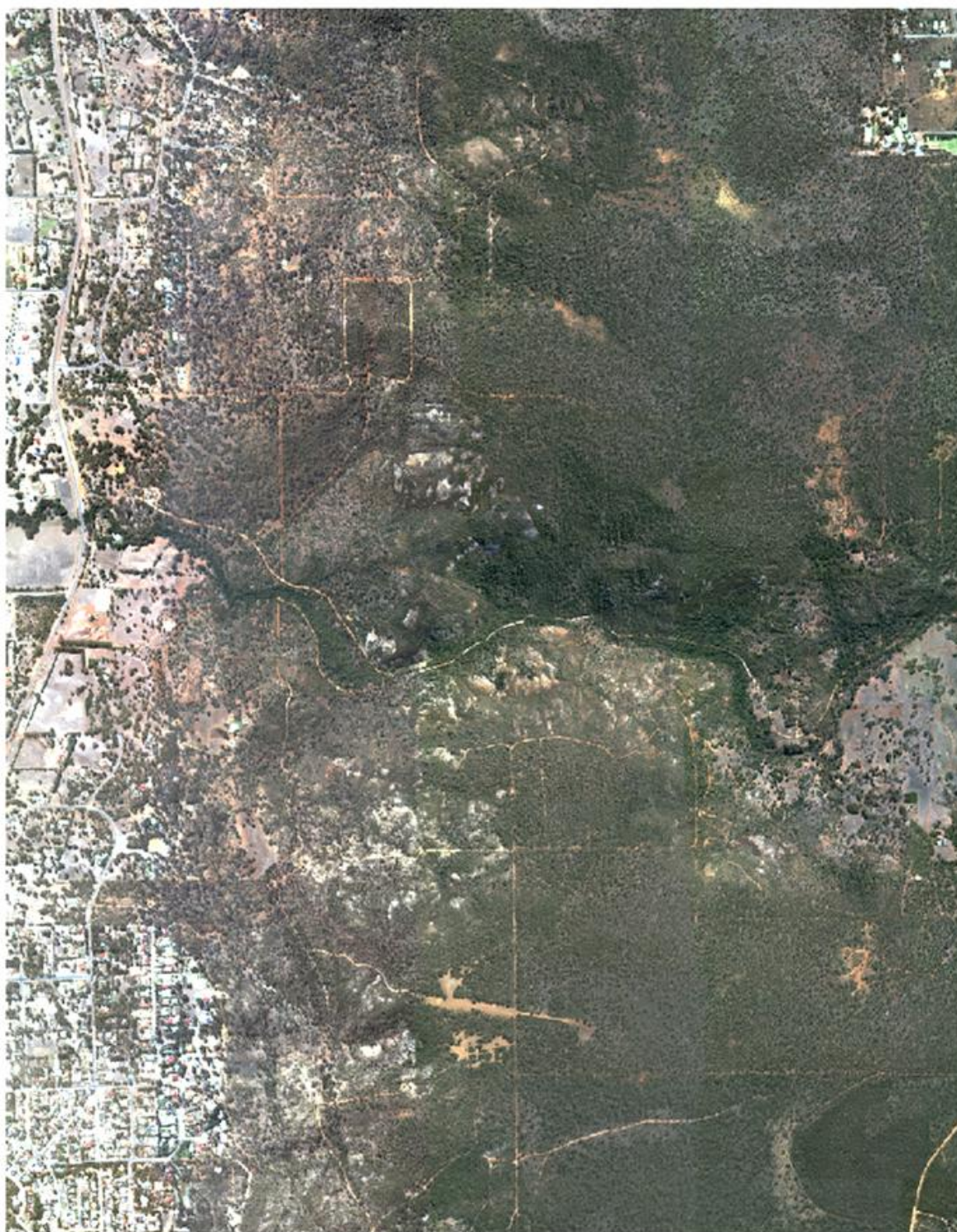


Figure 35 A subSection of the study area showing preserved vegetation area

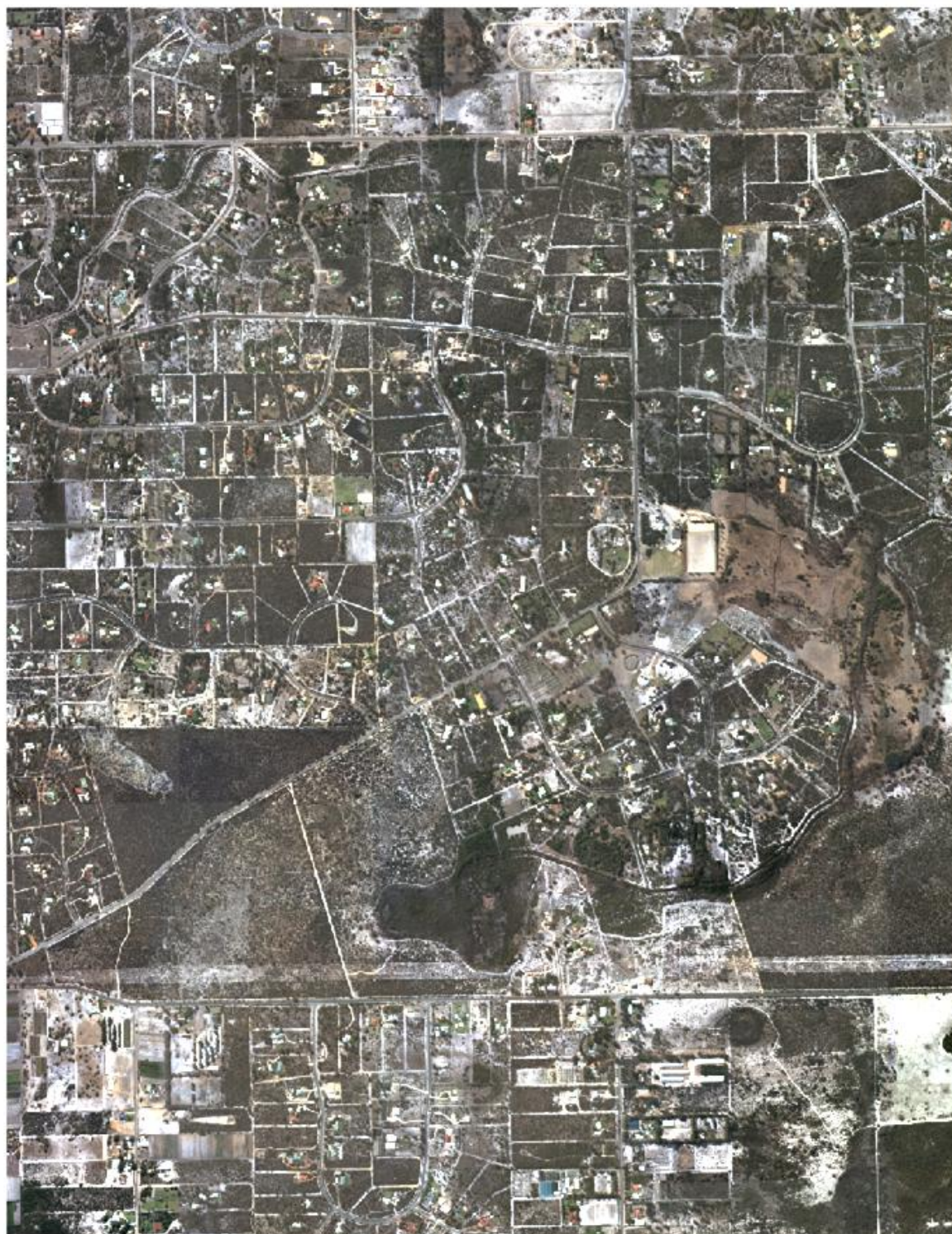


Figure 36 A subSection of the study area showing some sort of agricultural area



Figure 37 A subSection of the study area showing dense urban areas

## APPENDIX II



Figure 38 A subset image of the study area as a comparison for the classified image

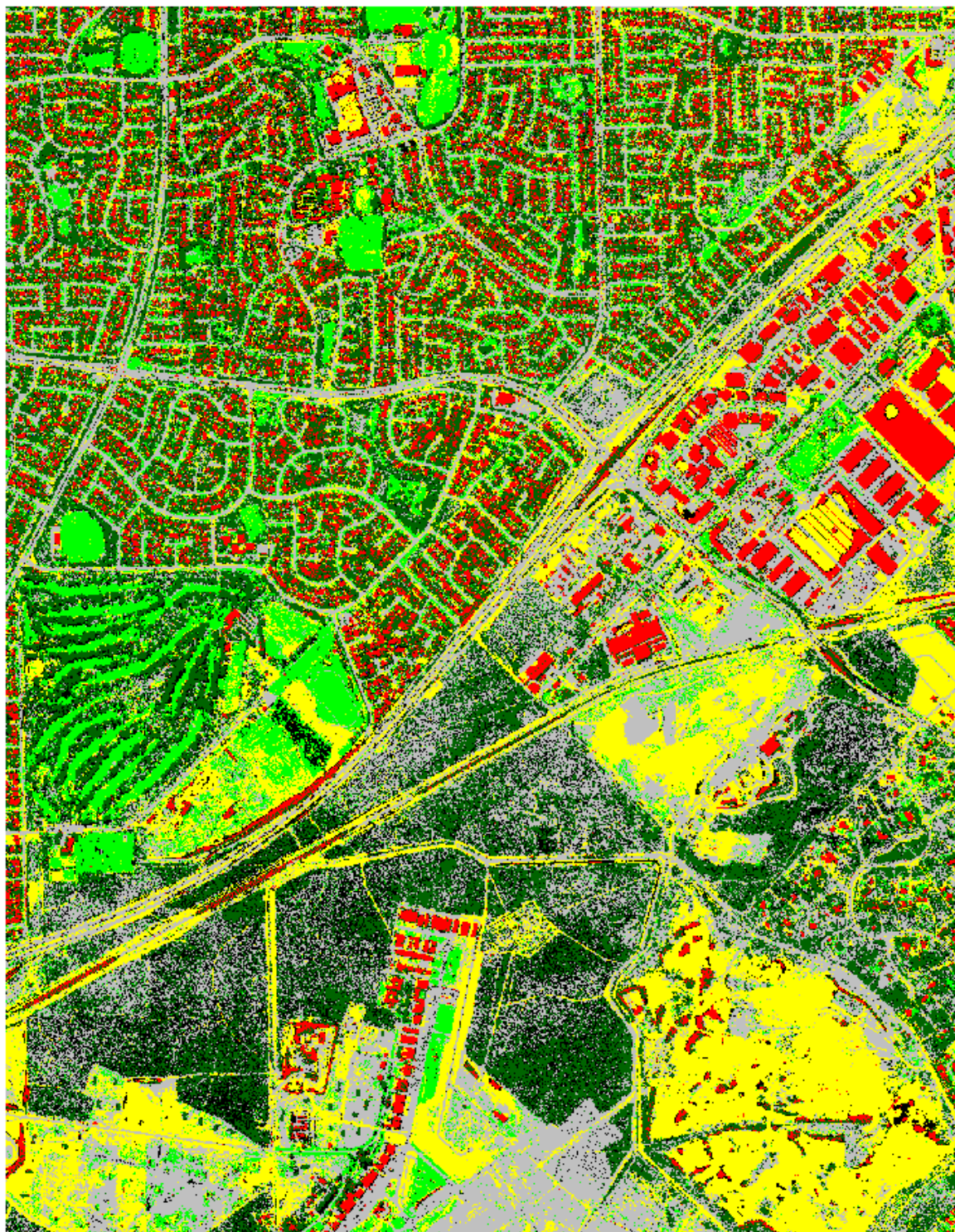


Figure 39 A classified image of a subsection of the study area