



## Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data

Mohd Hasmadi I<sup>1</sup>, Pakhriazad HZ<sup>1</sup>, Shahrin MF<sup>1</sup>

<sup>1</sup>Centre for Tropical Forest Airborne Observatory (TropAIR), Faculty of Forestry Universiti Putra Malaysia, Serdang, Selangor, Malaysia

Correspondence: Mohd Hasmadi I (email: [mhasmadi@putra.upm.edu.my](mailto:mhasmadi@putra.upm.edu.my))

### Abstract

Several methods exist for remote sensing image classification. They include supervised and unsupervised approaches. Accuracy assessment of a remote sensing output is a most important step in classification of remotely sensed data. Without accuracy assessment the quality of map or output produced would be of lesser value to the end user. However, supervised and unsupervised techniques show different levels of accuracy after accuracy assessment was conducted. This paper describes a study that was carried out to perform supervised and unsupervised techniques on remote sensing data for land cover classification and to evaluate the accuracy result of both classification techniques. The study used SPOT 5 satellite image taken on January 2007 for 270 / 343 (path / row) as a primary data and topographical map and land cover maps as supporting data. The land cover classes for the study area were classified into 5 themes namely vegetation, urban area, water body, grassland and barren land. Ground verification was carried out to verify and assess the accuracy of classification. A total of 72 sample points were collected using Systematic Random Sampling. The sample point represented 25% of the total study area. The results showed that the overall accuracy for the supervised classification was 90.28% where Kappa statistics was 0.86, while the unsupervised classification result was 80.56% accurate with 0.73 Kappa statistics. In conclusion, this study found that the supervised classification technique appears more accurate than the unsupervised classification.

**Keywords:** accuracy assessment, land cover mapping, remote sensing, SPOT 5 satellite image, supervised classification, unsupervised classification

### Introduction

Every parcel of land on the Earth's surface is unique in the cover it possesses. Land use and land cover are distinct yet closely linked characteristics of the Earth's surface. Land use is the manner in which human beings employ the land and its resources. Examples of land use include agriculture, urban development, grazing, logging, and mining. In contrast, land cover describes the physical state of the land surface. Land cover categories include cropland, forests, wetlands, pasture, roads, and urban areas. The term land cover originally referred to the kind and state of vegetation, such as forest or grass cover, but it has broadened in subsequent usage to include human structures such as buildings or pavement and other aspects of the natural environment, such as soil type, biodiversity, and surface and groundwater (Meyer, 1995).

A land cover map is constructed by partitioning a geographic area of interest into a finite set of map units and assigning a land cover class label to each unit. Land cover maps covering wide areas consisting of hundreds of units are often derived from satellite remotely sensed data. In

general, land cover is what covers the surface of the earth (Nerd, 2004). According to Comber *et al.*, 2005, there are two primary methods for capturing information on land cover, namely field survey and through analysis of remotely sensed imagery.

Remote sensing data has been use widely for land cover identification and classification of various features of the land surface from satellite or airborne sensor. Application of remotely sensed data for land cover and land use mapping and its changes is a key to many diverse applications such as environment, forestry, hydrology, agriculture, geology, etc. Acquired information from images remotely can be a valuable tool for a variety of resource-based applications. For example in forestry the images were used to assist in inventory assessment of their allotted forest stands. Ecologists may use image classification to categorize plant zones, wetland classification, etc. Data obtained from classification of images can then be utilized to assess changes in various ecosystems through time due to anthropogenic interference or global climate changes and natural disasters.

Classification in remote sensing involves clustering the pixels of an image to a (relatively small) set of classes, such that pixels in the same class are having similar properties. The majority of image classification is based on the detection of the spectral response patterns of land cover classes. Classification depends on distinctive signatures for the land cover classes in the band set being used, and the ability to reliably distinguish these signatures from other spectral response patterns that may be present (Eastman, 2003). There are many different approaches to classifying remotely sensed data. However, in common they all fall under two main topics: unsupervised and supervised classification technique.

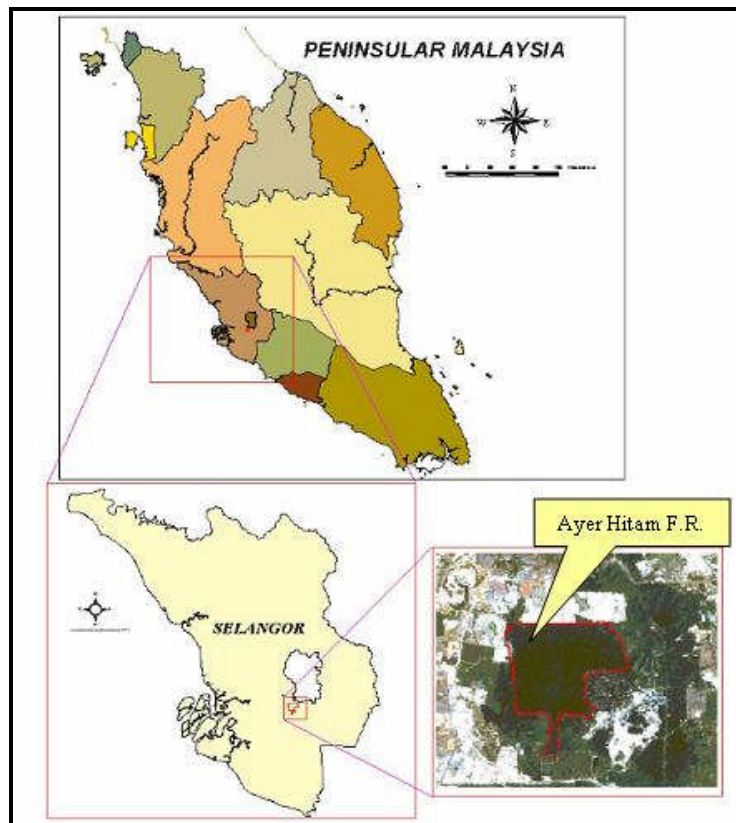
In unsupervised classification, an algorithm is chosen that will take a remotely sensed data set and find a pre-specified number of statistical clusters in multispectral or hyperspectral space. Although these clusters are not always equivalent to actual classes of land cover, this method can be used without having prior knowledge of the ground cover in the study site (Nie *et al.*, 2001). Supervised classification, however, does require prior knowledge of the ground cover in the study site. The multispectral or hyperspectral data from the pixels in the sample area or spectral signatures from spectral library are used to train a classification algorithm (Kamaruzaman *et al.*, 2009). Once trained, the algorithm can then be applied to the entire image and a final classification image is obtained.

In remote sensing-land cover mapping study, accuracy assessment is important to evaluate remote sensing final product. The purpose of assessment is important to gain a warranty of classification quality and user confidence on the product (Foody, 2001). Normally, accuracy result are derived from supervised or unsupervised or both techniques. However supervised and unsupervised technique relatively shows different level of accuracy after accuracy assessment was performed. Thus, the general objective of the study is to perform supervised and unsupervised technique on remote sensing data for land cover classification. The specific objectives of the study are two folds: (i) to evaluate land cover map of Ayer Hitam Forest Reserve and its surrounding areas using supervised and unsupervised classifications and (ii) to compare the accuracy assessment results from supervised and unsupervised classification techniques.

## **Study area**

The study was conducted in Ayer Hitam Forest Reserve (AHFR) and its surrounding area. The Ayer Hitam Forest Reserve is located in Puchong, Selangor covers a total area about 1248 hectare. The Selangor State Government leased this forest in 1996 for research and learning purpose especially to Faculty of Forestry, Universiti Putra Malaysia (UPM). Ayer Hitam Forest Reserve is about 20 minute drive from the UPM main campus in Serdang, Selangor. It is surrounded by development, making it an isolated patch of forest in the middle of modern

infrastructure and society of Puchong, Kinrara, Seri Kembangan, Serdang, Taman Desaminium, and Bandar Puteri (Paiman & Amat Ramsar, 2006). The location of the AHFR and its surrounding area is shown in Figure 1.

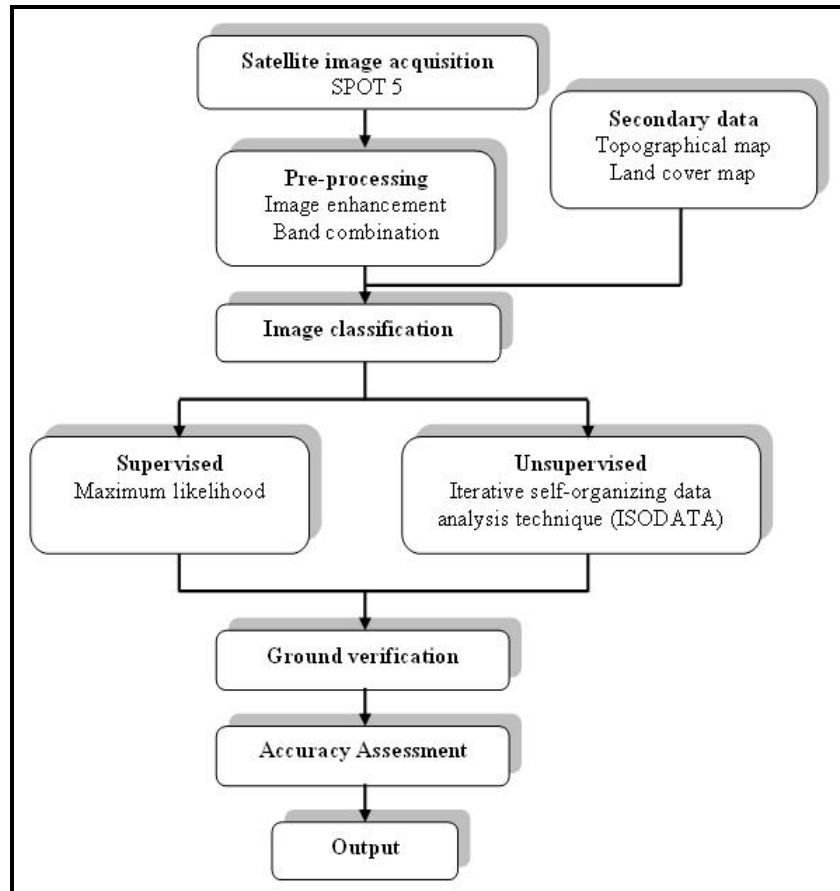


**Figure 1.** Location of the study area, Ayer Hitam Forest Reserve in Puchong, Selangor

## Materials and methodology

The primary data used in this study was SPOT 5 data with 10 m spatial resolution taken on 3<sup>rd</sup> January 2007. The path and row of the scene is K270 and J343. The satellite image was obtained from the Malaysian Remote Sensing Agency in Kuala Lumpur. In supporting the study secondary data were acquired which includes topographical map (scale 1:50000) obtained from the Department Survey and Mapping Malaysia (JUPEM) and land cover maps (scale 1:25000) issued by Petaling Jaya City Council. The Garmin GPS was used to navigate the location of selected samples during ground verification. Digital camera was used to photograph the land cover segments at visited sites for each class as reference data in the accuracy assessment. The digital image processing was carried out using a personal computer equipped with ERDAS IMAGINE 9.1 software for image processing, classification and analysis of the imagery.

Satellite image acquisition, pre-processing, secondary data collection, image classification, ground verification, accuracy assessment and output derivation are the main components involved in this study. Pre-processing was performed to improve qualitative judgements concerning the data. Image rectification and radiometric correction were done by the Malaysian Remote Sensing Agency. The summary of the methodology is shown in Figure 2.



**Figure 2.** The flow chart of the study

Image enhancement technique was performed to improve the quality of the image as perceived by a human. The process aims to edit the original image data by increasing the amount of information for visually interpretation from the data to create “new” image. These techniques are most useful because many satellite images when examined on a colour display give inadequate information for image interpretation. Contrast enhancement and band combination are two methods used in this study. A contrast enhancement technique is adequate to expand the range of brightness values to image so that the image can be efficiently displayed in a manner desired. The density values in a scene are literally pulled farther apart, that is, expanded over a greater range. The effect of this enhancement was increase the visual contrast between two areas of different uniform densities. However, there is no ideal or best image enhancement because the result are ultimately evaluated by humans, who make subjective judgments as to whether a given image enhancement is useful (Jensen, 1996).

Different combination bands of SPOT 5 image were tested and displayed to create different composite effects and increased information on land cover. In this study a false colour composite of 4-3-2 (Red-Green-Blue) band was selected for further analysis. Supervised and unsupervised were used in the image classification process. The algorithm used in supervised classification was the Maximum Likelihood Classification (MLC), while the unsupervised classification was the ISODATA technique (Iterative Self-Organizing Data Analysis).

#### *Development of land cover class*

Before attempting a classification it is important to define the land cover class theme based on the purpose of the study. The classification system was referred to the land cover map issued by the

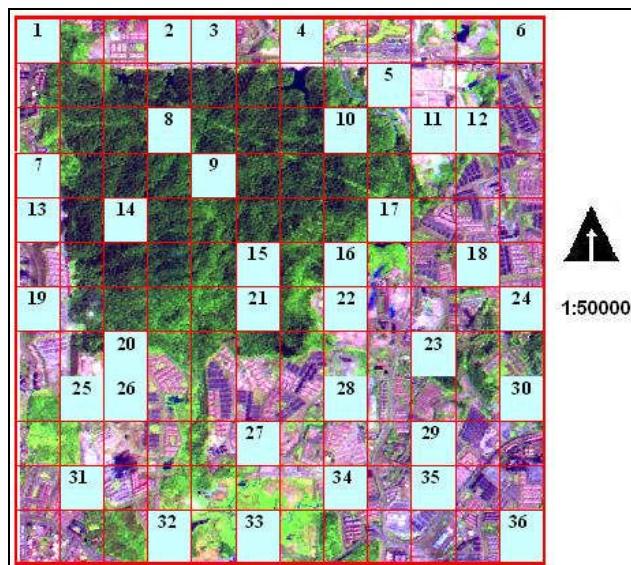
Petaling Jaya City Council. Land cover class for the study area was divided into five theme classes as shown in Table 1.

**Table 1. Land covers classification system used for this study**

| No. | Theme       | Associate Class   |
|-----|-------------|---|
| 1.  | Urban       | -Residential<br>-Industrial<br>-Transportation, communications, and utilities<br>-Industrial and commercial complexes<br>-Mixed urban or built-up land<br>-Road |
| 2.  | Grassland   | -Agricultural land<br>-Shrub and brush<br>-Turf & grass   |
| 3.  | Vegetation  | -Evergreen forest land<br>-Agricultural land<br>-Grassland  |
| 4.  | Water Body  | -Streams and canals<br>-Lakes<br>-Reservoirs  |
| 5.  | Barren Land | -Exposed soil<br>-Bare exposed rock<br>-Transitional areas  |

*Ground verification*

The ground verification was conducted based on area frame sampling where the basic unit is called a segment. It was conducted from 21<sup>st</sup> -23<sup>rd</sup> September 2007. A systematic random sampling method was adopted for collecting sample points of samples segment. Systematic random sampling enabled the samples be distributed evenly to all parts of the study. The size of



**Figure 3.** Area frame sampling with systematic random sampling (fillet blue) for 36 samples over Ayer Hitam Forest Reserve and its surrounding area

sample segment is 1 km by 1 km and sample size is 0.5 km by 0.5km. A total of 36 segments were selected and 72 sampling points were collected, where each sample has two samplings

(which represent 25% of the study area). Figure 3 shows the area frame sampling over the study area.

*Accuracy assessment: Supervised and unsupervised technique*

The accuracy assessments of both techniques were made through a confusion or error matrix. A confusion matrix contains information about actual and predicted classifications done by a classification system. The pixel that has been categorised from the image was compared to the same site in the field. The result of an accuracy assessment typically provides the users with an overall accuracy of the map and the accuracy for each class in the map. The percentage of overall accuracy was calculated using following formula:

$$\text{Overall accuracy} = \frac{\text{Total number of correct samples}}{\% \text{Total number of samples}} \times 100$$

Besides the overall accuracy, classification accuracy of individual classes was calculated in a similar manner. The two approaches are user's accuracy and producer's accuracy. The producer's accuracy is derived by dividing the number of correct pixels in one class divided by the total number of pixels as derived from reference data. In this study, the producer's accuracy measures how well a certain area has been classified. It includes the error of omission which refers to the proportion of observed features on the ground that is not classified in the map. Meanwhile, user's accuracy is computed by dividing the number of correctly classified pixels in each category by the total number of pixels that were classified in that category. The user's accuracy measures the commission error and indicates the probability that a pixel classified into a given category actually represents that category on ground. Producer's and user's accuracy are derived from following formula:

$$\text{Producer's accuracy (\%)} = 100\% - \text{error of omission (\%)}$$

$$\text{User's accuracy (\%)} = 100\% - \text{error of commission (\%)}$$

Kappa coefficient (K) is another measurement used in this study. It is calculated by multiplying the total number of pixels in all the ground verification classes (N) by the sum of the confusion matrix diagonals (  $X_{kk}$  ), subtracting the sum of the ground verification pixels in a class time the sum of the classified pixels in that class summed over all classes (  $\sum X_{k\Sigma} Y_{k\Sigma}$  ), where  $X_{k\Sigma}$  is row total and  $Y_{k\Sigma}$  is column total, and dividing by the total number of pixels squared minus the sum of the ground verification pixels in that class times the sum of the classified pixels in that class summed over all classes. The value of Kappa lies between 0 and 1, where 0 represents agreement due to chance only. 1 represents complete agreement between the two data sets. Negative values can occur but they are spurious. It is usually expressed as a percentage (%). Kappa statistic can be a more sophisticated measurement to classifier agreement and thus gives better interclass discrimination than overall accuracy. The calculation of Kappa statistic  $k$  is as follows:

$$k = \frac{\theta_1 - \theta_2}{1 - \theta_2}$$

$$\theta_1 = \frac{\sum_{i=1}^n x_{ii}}{N} \quad \theta_2 = \frac{\sum_{i=1}^n x_{i+} x_{+i}}{N^2}$$

where,

$x_{ij}$  = number of counts in the  $ij^{\text{th}}$  cell of the confusion matrix

N = total number of counts in the confusion matrix

$x_{i+}$  = marginal total of row i

$x_{+i}$  = marginal total of column i

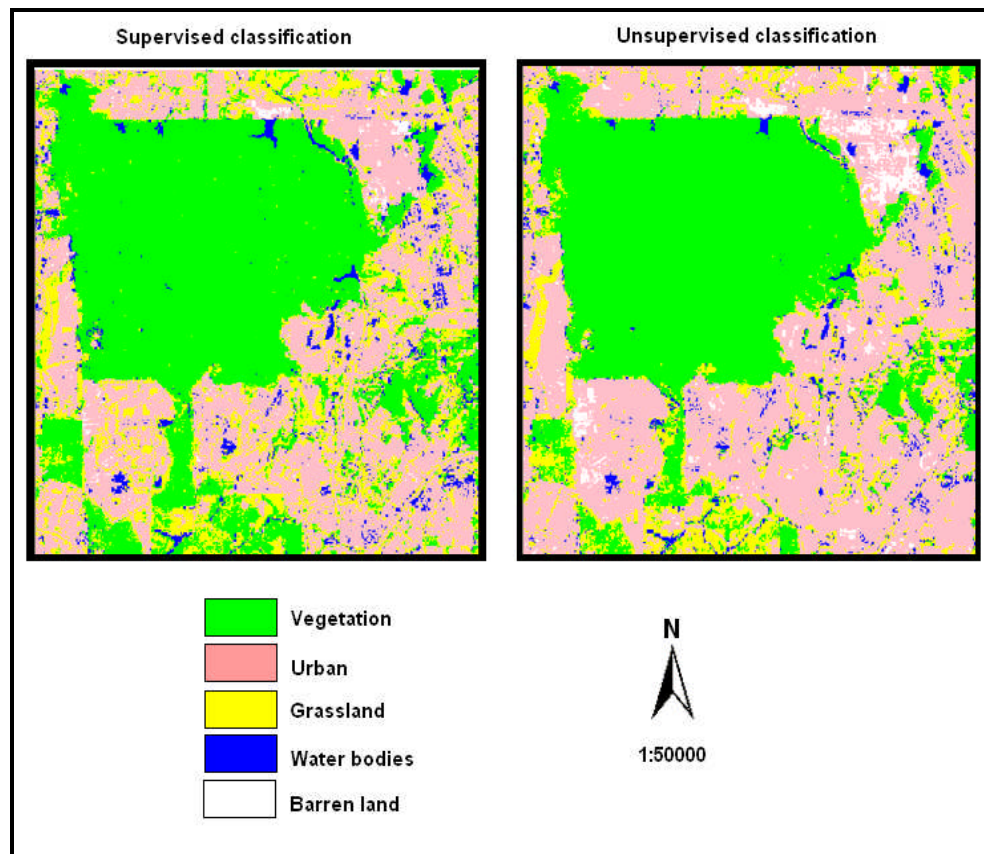


## Results and discussion

The selection of suitable band combination is essential for SPOT 5 image for visual interpretation. This enhancement technique was carried out in order to produce better an image classification for digital classification. The images of Ayer Hitam Forest Reserve and its surrounding area appear much better after performed enhancement technique. Brightness and contrast level panels were used to adjust the image view to an appropriate level. The image after enhancement clearly showed the separation of the land cover features. The view of the image after enhancement is better than before enhancement. In this study, the band combination of 4-3-2 (Red-Green-Blue) was selected as the best combination and later used for digital classification.

### *Image classification: Supervised and unsupervised*

The classification of the data was examined and divided into classes based on natural grouping and trained sample of the enhanced image of SPOT 5 data. Majority filter type was used to refine the image views and remove an unwanted isolated pixel. The results of the supervised and unsupervised classification techniques in five land cover classes are shown in Figure 4. The classes were vegetation, water body, urban, grassland, and barren land. The result showed that vegetation cover type was classified more in terms of area size by using supervised technique, but the barren land class type was classified more in area size using unsupervised technique. Meanwhile water bodies appear more expanded compared to the unsupervised technique. The difference in area classification of the themes are not just due to different approaches but the way the sample area was trained using the Maximum Likelihood Classifier (MLC) and the original mechanism of ISODATA classifier in the unsupervised technique.



**Figure 4.** *Supervised and unsupervised classification of SPOT 5 image of the study area*

*Accuracy assessment*

The accuracy of the land cover classification from supervised and unsupervised techniques were evaluated and presented as an error or confusion in the form of matrix table. The results of confusion matrix for supervised and unsupervised are presented in Tables 2 and 3. Meanwhile, Table 4 presents the comparison of overall accuracy and Kappa statistic between the supervised and unsupervised classification.

The accuracy assessment generated from the supervised classification technique showed an overall classification accuracy was 90.28% with Kappa Statistic of 0.87, which indicates a good agreement between thematic maps generated from image and the reference data. This amount of agreement is generally considered a good statistical return. For instance, in unsupervised classification technique, the accuracy decreased to 80.56% with Kappa statistic of 0.74. The supervised classification technique indicates a good agreement between the thematic map generated from SPOT 5 image and reference data. However, the accuracy of unsupervised classification was less than 85 %, which is below the acceptable level and standard of digital image classification for optical remote sensing data recommended by Paul (1991) and Jansen *et al.*, (2008). The vegetation mapping accuracy in both techniques referred to producer’s and user’s accuracies were ranging from 84.00% to 96.00% due to better spectral discrimination from other classes. This followed by urban class which accuracy ranging from 85.71% to 95.00%. However, from the result we found that all land covers classes were classified much better by supervised compared to unsupervised approaches. However, the accuracy value for both classification techniques on water bodies’ class were equal with 77.78% (producer’s accuracy) and 87.5% (user’s accuracy), respectively. Meanwhile, barren land class was well classified in supervised classification with Producer’s and User’s accuracy of 100% but decreased to 66.67% in unsupervised approach.

In Table 4 Kappa statistic illustrated that kappa accuracy and coefficient was better by using supervised compare to unsupervised, where kappa coefficient is slightly higher for supervised with 0.87 and unsupervised with 0.74.

**Table 2. Confusion matrix of the supervised classification image**

| Classified data         | Reference data |       |       |       |        | Row total | User’s accuracy (%) |
|-------------------------|----------------|-------|-------|-------|--------|-----------|---------------------|
|                         | V              | U     | WB    | GL    | BL     |           |                     |
| Vegetation (V)          | 24             | 0     | 0     | 1     | 0      | 25        | 96.00               |
| Urban (U)               | 0              | 19    | 2     | 0     | 0      | 21        | 90.48               |
| Water body (WB)         | 0              | 1     | 7     | 0     | 0      | 8         | 87.50               |
| Grass land (GL)         | 3              | 0     | 0     | 12    | 0      | 15        | 80.00               |
| Barren land (BL)        | 0              | 0     | 0     | 0     | 3      | 3         | 100.00              |
| Column total            | 27             | 20    | 9     | 13    | 3      | 72        | Overall accuracy    |
| Producer’s accuracy (%) | 88.89          | 95.00 | 77.78 | 92.31 | 100.00 |           | 90.28%              |



**Table 3. Confusion matrix of the unsupervised classification image**

| Classified data         |       | Reference data |       |       |        |           |                         |
|-------------------------|-------|----------------|-------|-------|--------|-----------|-------------------------|
| Land cover classes      | V     | U              | WB    | GL    | BL     | Row total | User's accuracy (%)     |
| Vegetation (V)          | 21    | 1              | 0     | 3     | 0      | 25        | 84.00                   |
| Urban (U)               | 0     | 18             | 2     | 0     | 1      | 21        | 85.71                   |
| Water body (WB)         | 0     | 1              | 7     | 0     | 0      | 8         | 87.50                   |
| Grass land (GL)         | 3     | 2              | 0     | 10    | 0      | 15        | 66.67                   |
| Barren land (BL)        | 0     | 1              | 0     | 0     | 2      | 3         | 66.67                   |
| Column total            | 24    | 23             | 9     | 13    | 3      | 72        |                         |
| Producer's accuracy (%) | 88.89 | 95.00          | 77.78 | 92.31 | 100.00 |           | Overall accuracy 80.56% |

**Table 4. Comparison of the overall accuracy and Kappa statistic between the supervised and unsupervised classification techniques**

| Classification | Overall accuracy (%) | Kappa accuracy (%) | Kappa coefficient |
|----------------|----------------------|--------------------|-------------------|
| Supervised     | 90.28                | 86.79              | 0.86              |
| Unsupervised   | 80.56                | 73.65              | 0.73              |

## Conclusion

The study on evaluation of accuracy from the supervised and unsupervised techniques had produced a baseline for the Ayer Hitam Forest Reserve area. The study concluded that the unsupervised classification was more “noisy” compared to the supervised classification. However the problem can be overcome by conducting a majority filter where isolated pixels were grouped into the closest value of spectral or digital number. SPOT 5 data is capable to map and classify the Ayer Hitam Forest Reserve and its surrounding area with five classes namely vegetation, urban, water body, grassland, and barren land for both classification techniques. The overall accuracy for supervised classification was higher than the unsupervised classification where the accuracy for supervised was 90.28% while the unsupervised produced 80.56%. Thus, the supervised classification appears more accurate than the unsupervised classification. In future it recommended that fusion technique of supervised and unsupervised is carried out in order to evaluate the area changes between two techniques of the land cover mapping. Later on the fusion images can be used to modify or determine the exact land cover class and investigate the advantage and disadvantage of the supervised and unsupervised approaches.

## References

- Comber AJ, Fisher PF, Wadsworth RA (2005) What is land cover? *Environment and planning. Planning and Design* **32**, 199-209.
- Eastman JR (2003) *Guide to GIS and image processing* **14**, 239-247. Clark University manual, USA.
- Foody GM (2001) Status of Land Cover Classification Accuracy Assessment. *Remote Sensing of Environment* **80**, 185-201.

- Jansen LJM, Bagnoli M, Focacci M (2008) Analysis of land-cover/use change dynamic in Manica Province in Mozambique in a period of transition (1990-2004). *Forest Ecology and Management* **254**, 308-326.
- Jensen JR (1996) *Introductory digital image processing: A remote sensing perspective*. Prentice-Hall Inc, Eaglewood Cliff, UK.
- Kamaruzaman J, Mohd Hasmadi I, Nurul Hidayah MA (2009) Spectral separability of tropical forest tree species using airborne hyperspectral imager. *J. of Environmental Science & Engineering* **3**(1), 37-41.
- Meyer WB (1995) Past and present land-use and land-cover in the USA consequences. *The Nature and Implications of Environmental Change* **1**(1), 24-33.
- Nerd H (2004) *Remote sensing resources: Remote sensing & Geographic Information System facility*. Center for Biodiversity and Conservation. American Museum of Natural History.
- Nie Y, Kafatos M, Wood K (2001) Estimating soil-type pattern from supervised and unsupervised classification: Case study in Cuprite, Nevada. Project Report, pp. 28. George Mason University.
- Paiman B, Amat Ramsar Y (2006) *Hutan simpan Ayer Hitam, warisan komuniti Koridor Raya Multimedia (in Malay)*, pp. 34. Universiti Putra Malaysia Press, Serdang.
- Paul M (1991) *Computer processing of remotely sensed images: An introduction*, pp. 352. Biddley Limited Publication, UK.