# Appendix

Conceptual paper.

Topic: Rubber-growing areas identification from LANDSAT satellite imagery: a case study of RRI – MACRES project.

Author: Choong Wai Cheong.
Temporary Research Officer in Malaysian Centre for Remote Sensing (MACRES)

Date: 28 April – 03 May 2003

1. Introduction.
   RRI (Rubber Research Institute of Malaysia), MRRDB and MRELB were previously three main agencies under the name of MRB, which are now merged as one[1]. The primary objective of MRB (Malaysian Rubber Board) is to assist in the development and modernization of the Malaysian rubber industry in all aspects from cultivation of the rubber tree, the extraction and processing of its raw rubber, the manufacture of rubber products and the marketing of rubber and rubber products. In early 2003, MACRES was appointed by RRI to inventory the rubber-growing areas in the Peninsular Malaysia utilizing LANDSAT multi-spectral data (30 meter spatial resolution). In the first stage, MACRES is to deliver the 1:50,000 scale rubber-growing areas SIM (satellite image map) for the three northern states of Peninsular Malaysia ie. Kedah, Perlis and Perak and a 1:250,000 scale similar map for each of the states respectively. SIM, the state of the art of mapping technique in MACRES that integrates vector GIS data, raster satellite imageries and the cartographic component to reveal the best quality information of the earth surface.

2. Objective.
   The main objective of the study is to analyze the current land-cover identification technique in Remote Sensing, to design and to customize a suitable technique for the purpose of the RRI – MACRES project. This study focused mainly on the production of 1:50,000 rubber growing area SIM.

3. Current technique and the statement of the problem.
   Production of each 1:50,000 rubber-growing area map starts with the visual interpretation of the particular Landsat image. Rubber-growing areas were identified in this stage by an expert in imagery interpretation, later on being input into a computerized pattern recognition software namely E-cognition. As a quality control measure, the result is to be examined visually by the expert once again after the pattern recognition process. This rubber-growing polygon contributes as a data layer in the final output Satellite Image Map.
   The main problem of this technique is time consuming especially in interpreting visually and manually each and every image of the targeted states plus the E-cognition processing. In average, for each map sheet (30km x 30km), it requires 1-2 days in visual interpretation and 5-6 days in E-cognation process. Perlis comprises 3 map sheets, Kedah 11 map sheets and Perak >< 33 map sheets and these entire map sheets are given three months duration to finish. Due to the

[1] http://www.lgm.gov.my/general/aboutmrb.html

almost impossible tight schedule, there is a need to study and review the current technique and introduce an alternative solution, which is less time consuming but sustaining the accuracy of its result.

4. Methodology.
System Analysis and Design concept is being employed in this study. The rubber-growing area identification technique is being analyzed and an alternative technique is being designed in this study.
In order to reduce the time consumption in the identification process, one same sample set of rubber-growing areas should be applied as a training area to all map sheets instead of interpreting each and every images involved. The problem is, each image is "time and site-specific", which means training area derived from the first image does not work well in extracting the same feature in other image or even the same image at a different acquisition time. For example, DN (digital number) for rubber –growing area in band 4 of this image is 125-137 but in other image might appears as 135-148 due to the different topographic, atmospheric and radiometric effects.
Does it mean that there is no other alternative in solving the problem? No. Choong 2003 had done a study in land-cover information extraction from remotely sensed data and suggested that in order to reduce the topographic and atmospheric effect, training set could be prepared from a ratio-image instead of the original bands. This concept is the main model in this study.
The formulation of the alternative technique in this study also included decision rules such as the logical spectral reduction and spatial reduction process during the land-cover identification. Spatial reduction is the data slicing procedure to ignore irrelevant area and concentrated on the masked area. Spectral reduction does almost the same except that it reduced in the spectral range instead of a physical area.
Software used in the study is Idrisi by Clark University.

5. Image Processing.
5.1 Study area.
Gurun area in Kedah (Topo-map 3367) is chosen as the study area because It is part of the LGM-MACRES project and there are well-distributed classes of rubber-growing areas in this place, which made the comparison of the current and the alternative technique available. Image used in the study is Landsat TM dated 24 February 2001.

5.2 Spatial Reduction.
The main objective of this process is to reduce unnecessary area from the image. It is important, acting as a mask to ignore irrelevant areas in order to minimize confusion in the classification process. The target feature in the study is the rubber-growing areas; therefore, non-vegetation has no significant value in the study and should be removed. One simple solution to this is to generate a NDVI (Normalized Differentiate Vegetation Index) from the original Landsat bands.

## Appendix

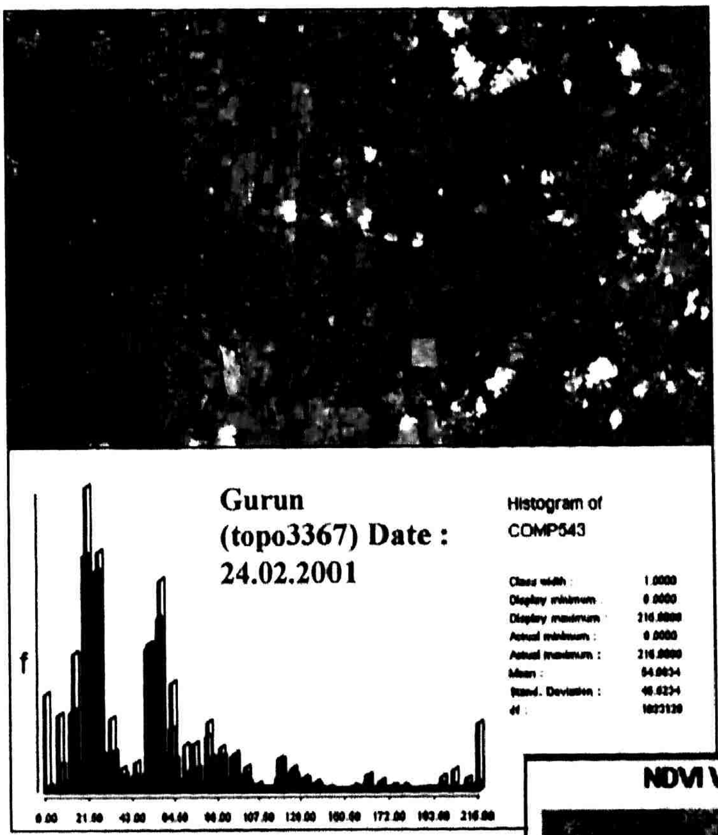



Figure 1:
The original “Gurun” image with the composite display palette

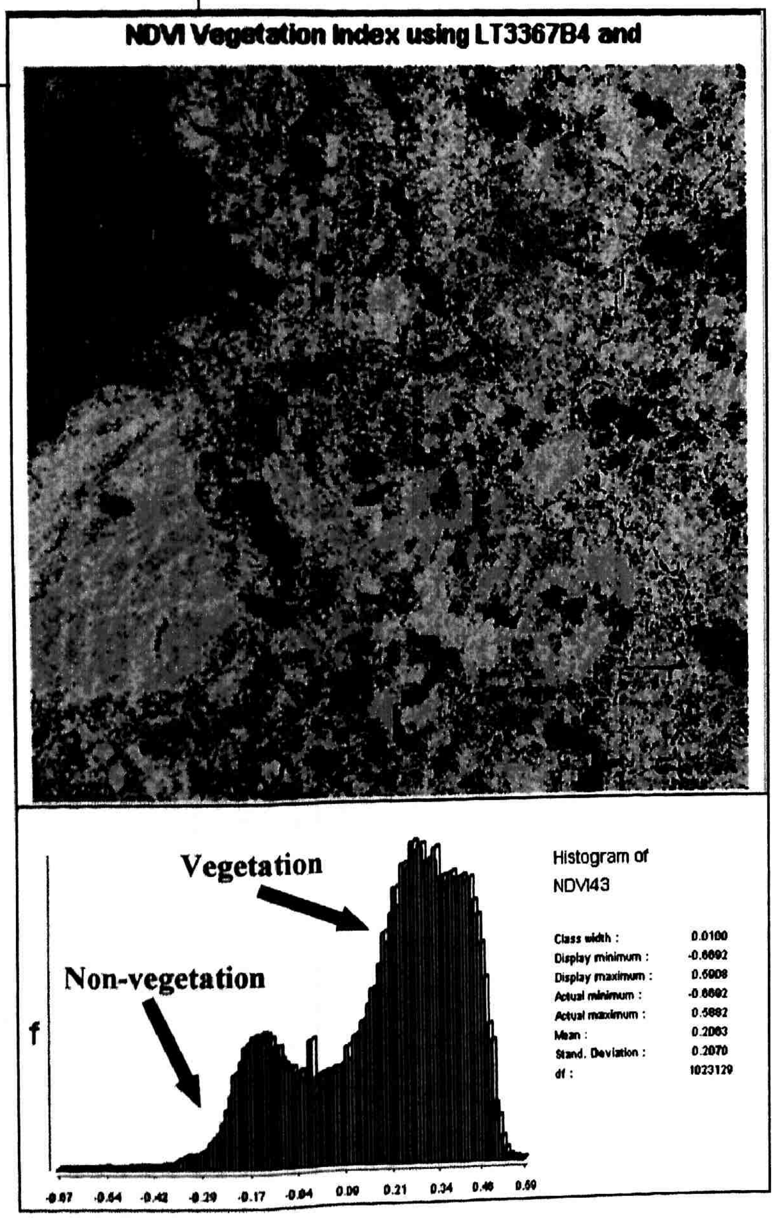
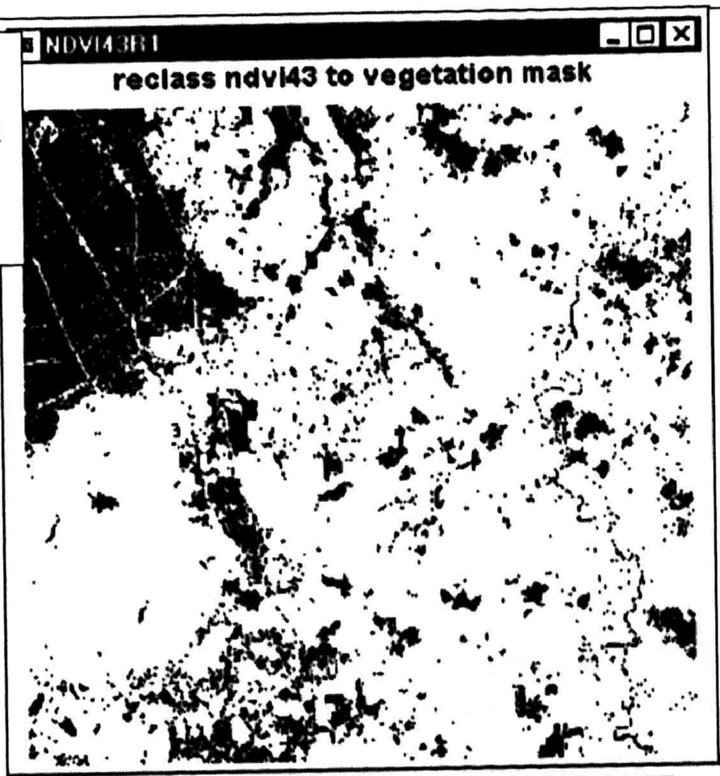
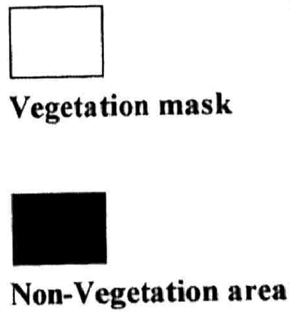



Figure 2:
NDVI result and its histogram demonstrating the distribution pattern of its data.

Figure 3:
A mask (white color) prepared for vegetation area.




**Appendix**



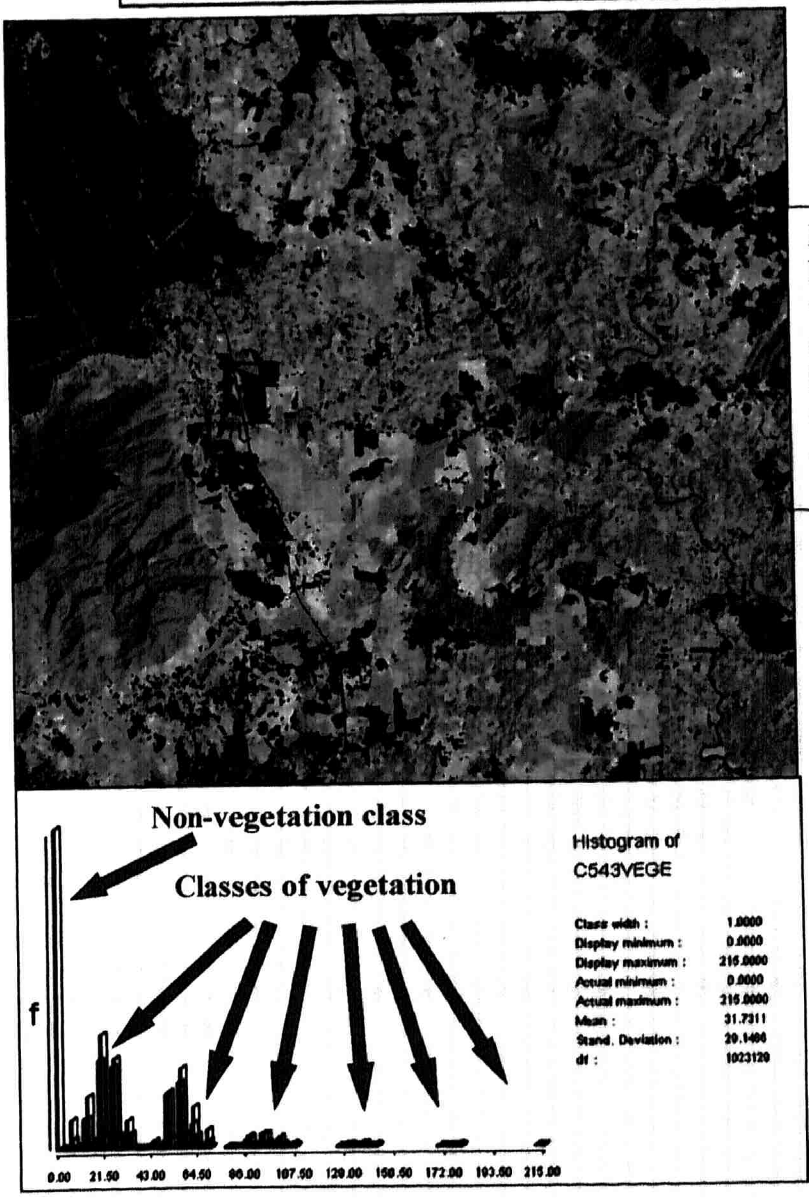




Figure 4:
Boolean overlay of the vegetation mask on the original Image and the histogram (below) demonstrating the data distribution.

Figure 1 shows the display of the original "Gurun" image with the composite display palette. Figure 2 shows the NDVI result and its histogram demonstrating the distribution pattern of its data. There is a clear distinction between vegetation and non-vegetation in the NDVI image. From this result, the non-vegetation areas, which are not relevant to the study, can be removed. Figure 3 displays a vegetation mask derived from the NDVI image. As illustrated in Figure 4, the data distribution shows the different cluster of vegetation in the masked area.

5.3 Ratio-image.

Two map overlays with some basic arithmetic and Boolean operators are powerful tools for examining the spatial patterns caused by interactions of one map with another (Bonham-Carter G. F. 1996). NDVI was designed to extract and enhance vegetation information from a remote sensing data. Its formula is shown in Figure 5 It was found to be reliable in overcoming the topographic and atmospheric effects of a Landsat data and It is used in this study as a base in training area preparation.

Figure 5: The formula of NDVI.

$$\boxed{NDVI = \frac{NIR - VIS}{NIR + VIS}\quad [2]}$$

For Landsat TM data, band 4 and 3 are determined to be the best option in computing the NDVI values. However, other bands are used as well in the study for the purpose other then the distinction of vegetation and non-vegetation group. It is discussed under item 5.4.

5.4 Spectral Reduction and Training area.

Spectral reduction is the concept of slicing the selected region of a data distribution and ignoring the others, which are not significant in the study. Training area prepared in this study is the result of the Spectral reduction process.

Experiments and Observations in Choong's study found that NDVI45 (Figure 6) –0.02 to 0.1 is suitable for extract rubber feature from remote sensing data. Therefore in the spectral reduction process, the particular range of NDVI45 –0.02 to 0.1 was extracted as the first step in training area preparation for rubber-growing area (Figure 7).

In order to minimize the salt and pepper effect, the training area is filtered using a median filter (size 5X5 panel) (Figure 8)

[2] NIR = Near Infrared band; VIS = Visible band. For the Landsat TM data, band 4 is the NIR and band 1, 2 and 3 are the Visible bands.

## Appendix

Figure 6:
The NDVI 45 displayed in the VI display palette.

NDVI 45: input for extraction of rubber information

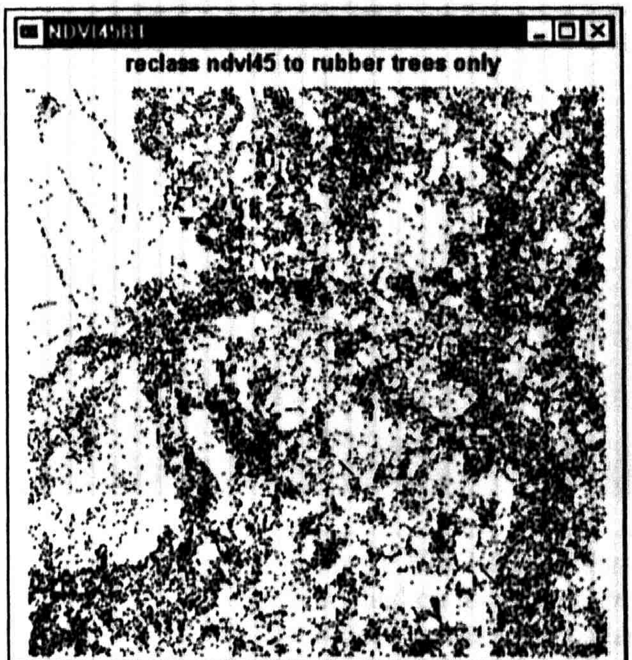



Figure 7: Reclass NDVI 45 ranging from –0.02 to 0.1 as integer 1, others to 0.

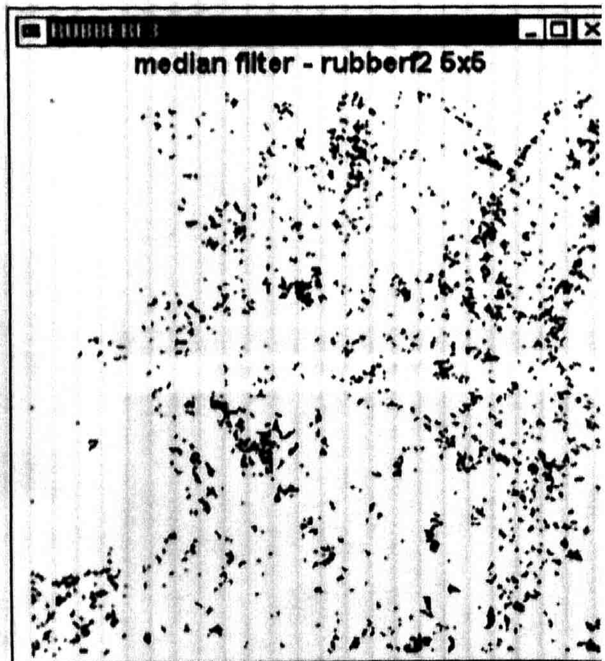



Figure 8: Median filter (size 5X5) applied to NDVI45R1.

5.5 Baysian Soft Classification.

Baysian is a soft classifier, instead of producing one output map with all the definite classes; it produces several images depending on the number of signatures (training area) input in it and each image shows the probability level in all pixels to be identified as one land-cover.

5.6 Result Inspection.

The Baysian result shows probability of each pixel belonging to the rubber class (Figure 9). In this study, probability greater then 0.9 (as 0 = most unlikely and 1 = certainly) was chosen as the output rubber-growing area (Figure 10). Accuracy assessment was carried out, comparing the Baysian result with the topographic map.

Figure 9: Baysian classification; probability of each pixel belonging to the rubber class.

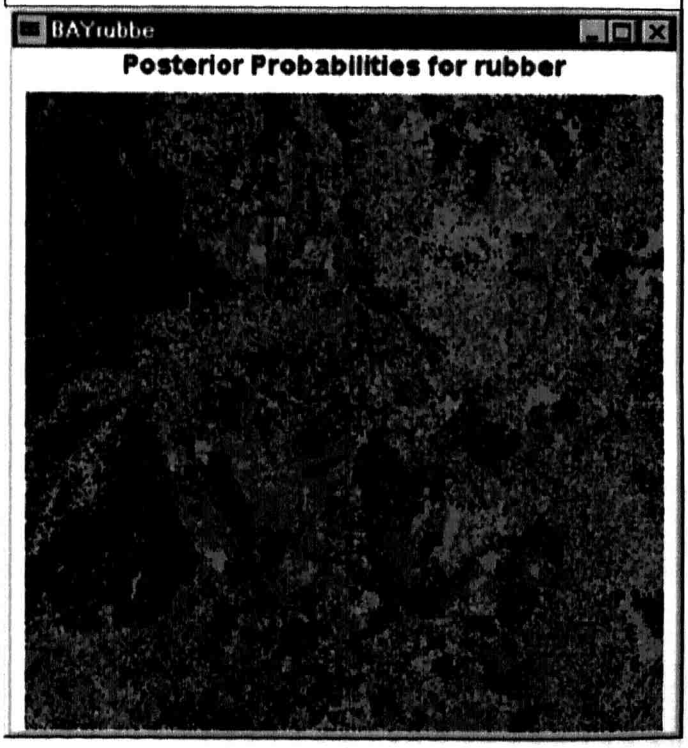
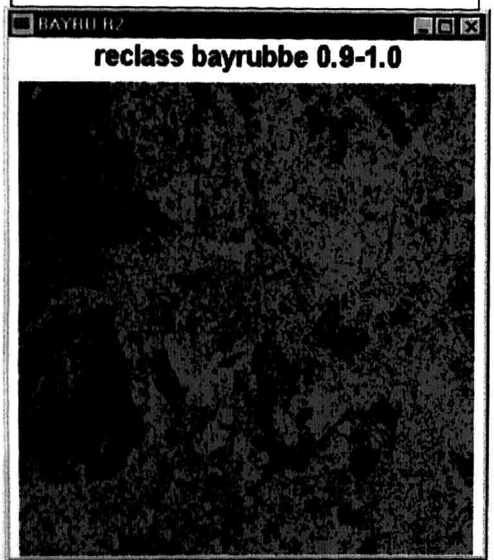



Figure 10a: Reclass selected pixel (with probability of 0.9 to 1.0) as rubber class.



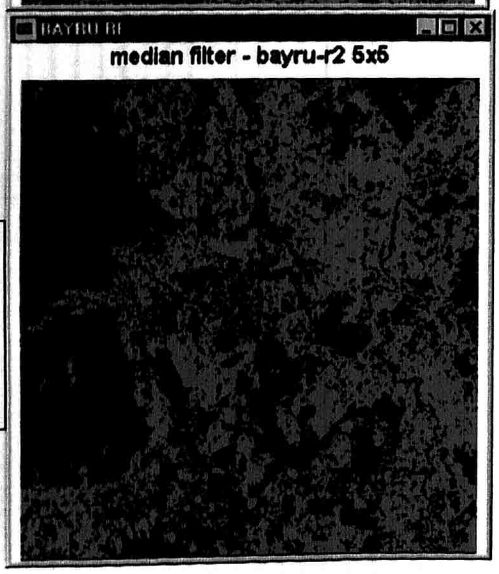




Figure 10b: Median filter applied on the rubber class in order to reduce the salt and pepper effect.

5.7 Classes of rubber-growing area.

After the general rubber growing area is identified, further processing classified different clusters within the rubber area. Unsupervised classification examined the distribution pattern of a data set and grouped pixels with similar spectral pattern as a cluster. For the study area, there are 2 major and 4 minor clusters as shown in Figure 11. The first (on the left) and also the largest cluster represent the non-rubber area; while the second peak in the histogram represents the major rubber class (yellow colored area in Figure 12) following by 4 other rubber growing clusters.

Figure 11: Histogram of rubber growing clusters

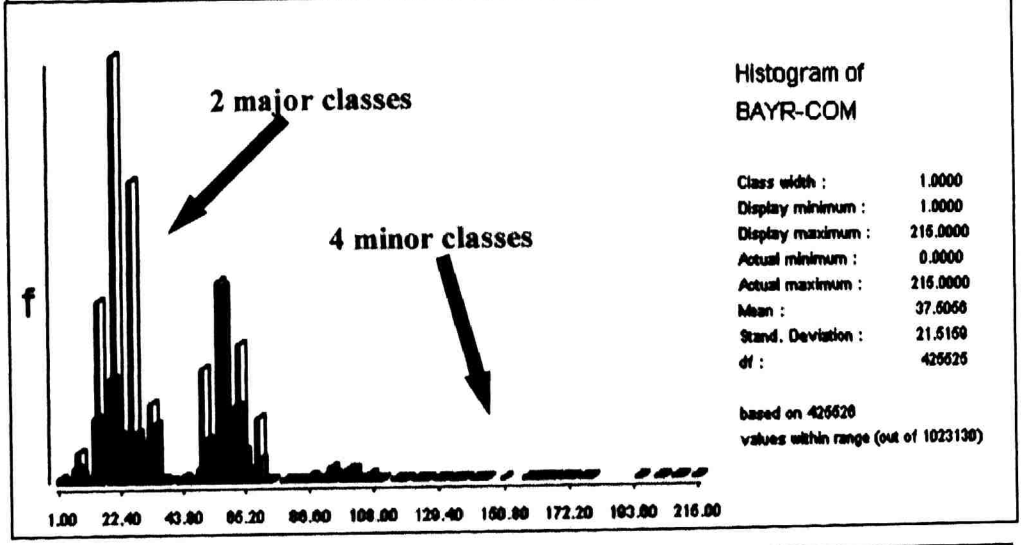
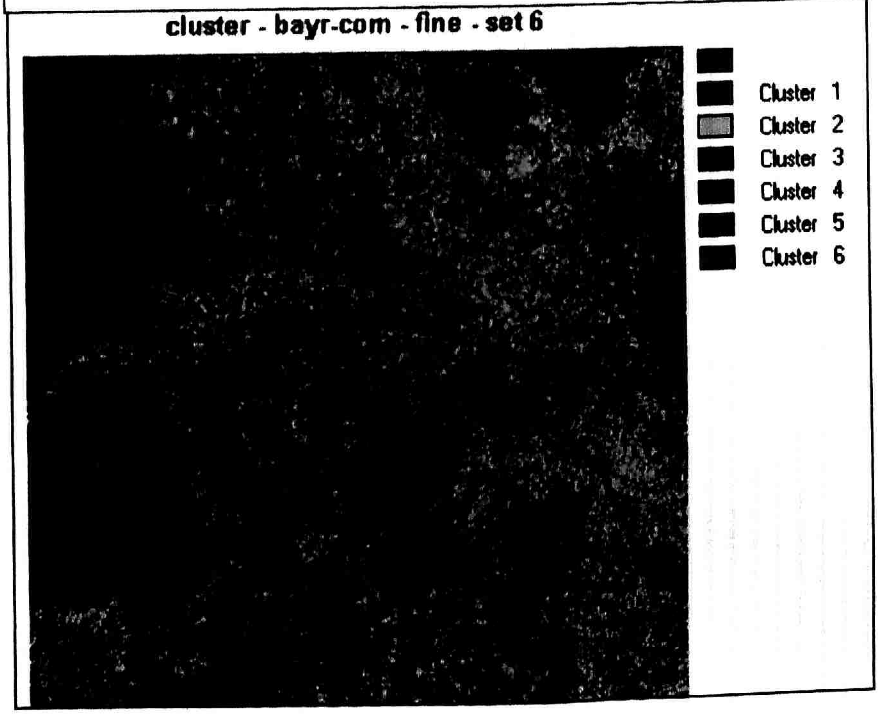



Figure 12: One non-rubber cluster (Cluster 1)and 5 rubber growing clusters



Result and discussion.

Accuracy assessment was carried out, comparing the Baysian result with the topographic map. Selection of 40 sample points, which are rubber area identified in the topography map, were overlaid and examined in the Baysian result. It archived an accuracy of 97.5% where only one rubber sample was classified as non-rubber. Due to the concept of building training area from the sliced image ratio, the rubber growing area resulted from this alternative classification procedure is expected to ignore many small holding rubber growing area, which is minority or a mixture of rubber with other vegetations.

With this alternative classification procedure, it is possible for RRI – MACRES project to produce rubber growing area maps (1:50,000) without consuming too much time in the feature recognition process.

Item 5.7 show that it is also feasible to differentiate various spectral clusters within a rubber growing area. It is important cause the project team may wish to study deeper on the distribution of the rubber growing area; answering questions such as where are the matured rubber growing areas? Where are the young rubber growing areas? What is the estimated production volume if the size of the matured rubber growing area can be identified?

Additional field verification is necessary in order to correlate these different clusters of rubber growing area with the real world.