

Classification Simulation of RazakSAT Satellite

Asmala Ahmad

Center of Advanced Computing Technologies (C-ACT)
Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka (UTeM)
Hang Tuah Jaya, Melaka, Malaysia
asmala@utem.edu.my

Abstract — This study presents simulation of land cover classification for RazakSAT satellite. The simulation makes use of the spectral capability of Landsat 5 TM satellite that has overlapping bands with RazakSAT. The classification is performed using Maximum Likelihood (ML), a supervised classification method that is based on the Bayes theorem. ML makes use of a discriminant function to assign pixel to the class with the highest likelihood. Class mean vector and covariance matrix are the key inputs to the function and are estimated from the training pixels of a particular class. The accuracy of the classification for the simulated RazakSAT data is accessed by means of a confusion matrix. The results show that RazakSAT tends to have lower overall and individual class accuracies than Landsat mainly due to the unavailability of mid-infrared bands that hinders separation between different plant types.

Keywords - RazakSAT, Maximum Likelihood, Accuracy

I. INTRODUCTION

The Malaysian second remote sensing satellite, RazakSAT (Figure 1), is named after the Malaysian second prime minister - the late Abdul Razak, also known as the father of Malaysian Development. It was successfully launched into orbit by a Falcon 1 rocket on July 14, 2009, aiming to serve Malaysia with remote sensing data for applications mainly in agriculture, landscape mapping, forestry and urban planning.

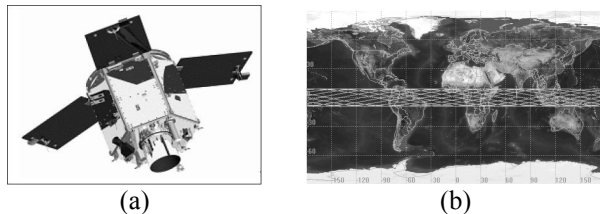


Figure 1. (a) RazakSAT satellite (Malaysian National Space Agency 2007) and (b) RazakSAT Near Equatorial Orbit (NEqO) [12].

As it is meant to serve the interest of Malaysia and other countries (Table 1) in the near-Equatorial belt, the satellite was launched into a Near Equatorial Orbit (NEqO) with inclination 9°, to take advantage of the corresponding opportunities to gather images at 14 times

daily and overcome the major obstacle for passive remote sensing, cloud cover [11].

TABLE 1. LIST OF COUNTRIES COVERED BY NEQO AT 9° INCLINATION [11]

Continent	Countries
Asia	India, Indonesia, Malaysia, Maldives, Philippines, Sri Lanka, Thailand
Africa	Angola, Benin, Burundi, Cameroon, Central African Republic, Chad, Cote d'Ivoire, Ethiopia, Gabon, Guinea, Kenya, Liberia, Nigeria, Rwanda, Sierra Leone, Somalia, Sudan, Tanzania, Uganda, Zaire
Latin America	Brazil, Colombia, Ecuador, French Guiana, Guyana, Panama, Peru, Surinam, Venezuela

The RazakSAT satellite carries a medium-sized aperture camera (MAC) to be used as an earth observation payload. It is a push-broom type high resolution camera with one panchromatic and four multispectral bands. The panchromatic band with 2.5 m ground sampling resolution operates at 0.510 to 0.730 μm wavelength. The other four bands operate at 0.450 to 0.520 μm (blue), 0.520 to 0.600 μm (green), 0.630 to 0.690 μm (red) and 0.760 to 0.890 μm (near infrared) wavelength respectively (Table 2). Unfortunately, due to unknown circumstances, the RazakSAT data is still available for users, although nearly after three and a half year since its launch.

TABLE 2. RAZAKSAT SPECIFICATIONS [13]

Parameter	Description
Altitude	685 km (nominal)
Inclination	9°
Spectral channels	5 visible and 1 panchromatic bands: Band 1: 0.450 – 0.520 μm (Blue) Band 2: 0.520 – 0.600 μm (Green) Band 3: 0.630 – 0.690 μm (Red) Band 4: 0.760 – 0.890 μm (Near infrared) Band 5: 0.510 – 0.730 μm (Panchromatic)
Ground Sampling	5.0 m (visible) 2.5 (panchromatic)
Swath Width (FOV)	> 20 km (1.675°)
IIFOV	7.30 μrad (visible) 3.65 μrad (panchromatic)
SNR	> 70
Modulation Transfer Frequency	15% (visible) 8% (panchromatic)
Data Storage	32 Gbit
Signal Quantization	8 bits

Instrument Mass	42.1 kg
Designed Mission Lifetime	> 3 years

One of the most common remote sensing satellites is Landsat, initiated by NASA (National Aeronautical and Space Administration) in 1972 [3]. The Landsat satellites have been providing optical data for almost 40 years. Landsat 1 – 3 launched in the 1970s and used Multispectral Scanner (MSS), while Landsat 4 – 5, launched in the 1980s, use Thematic Mapper (TM) as their main sensor. The latest Landsat 7, launched in 1999, uses the Enhanced Thematic Mapper (ETM+). Comparison between the specifications of these satellites is given in Table 3. Landsat 5 was launched on March 1, 1984 with an expected lifetime of 5 years, and, after more than 27 years of operation, has provided the global science community with over 900,000 individual scenes and is the longest running satellite of the series (Figure 2).



Figure 2. Landsat 5 satellite [16]

TABLE 3. LANDSAT 5 TM SATELLITE SPECIFICATIONS [3]

Parameter	Description
Spectral Bands	4 VNIR, 2 SWIR, 1 thermal
Spatial Resolution (IFOV)	30 m – VNIR, SWIR 120 m – thermal
Sampling	1 samples/IFOV along scan
Cross Track Coverage	185 km
Radiometric Resolution	8 bits
Radiometric Calibration	Internal lamps, shutter and black body
Scanning Mechanism	Bidirectional Scanning with Scan Line Corrector
Period of operation	Landsat 5: 1984 – present
Main sensor	MSS TM
Altitude	705 km
Repeat Cycle	16 days
Equatorial Crossing	9:45 AM +/- 15 minutes
Type	Sun synchronous, near polar
Inclination	98.2°

Landsat 5 TM level 1 data come in Product Generation System (LPGS) format and need to be converted into a physically meaningful common radiometric unit, representing the at-sensor spectral radiance. The Level 1 Landsat 5 TM data received by users are in scaled 8-bit numbers, Q_{cal} , or also known as

digital number (DN). Conversion from Q_{cal} to spectral radiance, L_{λ} , can be done by using [5]:

$$L_{\lambda} = \frac{(L_{max\lambda} - L_{min\lambda})}{(Q_{calmax} - Q_{calmin})} (Q_{cal} - Q_{calmin}) + L_{min\lambda} \quad (1)$$

where, L_{λ} is the spectral radiance at the sensor's aperture ($W/m^2 sr \mu m$), Q_{cal} is the quantized calibrated pixel value (DN), Q_{calmin} is the minimum quantised calibrated pixel value corresponding to $L_{min\lambda}$ (DN), Q_{calmax} is the maximum quantised calibrated pixel value corresponding to $L_{max\lambda}$ (DN), $L_{min\lambda}$ is the spectral at-sensor radiance that is scaled to Q_{calmin} ($W/m^2 sr \mu m$), $L_{max\lambda}$ is the spectral at-sensor radiance that is scaled to Q_{calmax} ($W/m^2 sr \mu m$) and Q_{calmin} and Q_{calmax} are 1 and 255 respectively.

Table 4 shows $L_{min\lambda}$, $L_{max\lambda}$ and the mean exoatmospheric solar irradiance (E_{λ}).

TABLE 4. LANDSAT TM SPECTRAL RANGE, POST-CALIBRATION DYNAMIC RANGES AND THE MEAN EXOATMOSPHERIC SOLAR IRRADIANCE [5]

Band	Spectral range	Centre wavelength	$L_{min\lambda}$	$L_{max\lambda}$	E_{λ}
	(μm)		(W/ m ² sr μm)		
1	0.452 – 0.518	0.485	-1.52	169	1983
2	0.528 – 0.609	0.569	-2.84	333	1796
3	0.626 – 0.693	0.660	-1.17	264	1536
4	0.776 – 0.904	0.840	-1.51	221	1031
5	1.567 – 1.784	1.676	-0.37	30.2	22.0
6	10.45 – 12.42	11.435	1.2378	15.3032	N/A
7	2.097 – 2.223	2.223	-0.15	16.5	83.44

Scene-to-scene variability can be reduced by converting the at-sensor spectral radiance to top-of-atmosphere (TOA) reflectance, also known as in-band planetary albedo. By performing this conversion, the cosine effect of different solar zenith angles due to the time difference between data acquisitions is removed, different values of the exoatmospheric solar irradiance arising from spectral band differences are compensated and the variation in the Earth–Sun distance between different data acquisition dates is corrected. The TOA reflectance can be computed by using [5]:

$$\rho_{\lambda} = \frac{\pi L_{\lambda} d^2}{E_{\lambda} \cos(\theta_s)} \quad (2)$$

where, ρ_{λ} is the planetary TOA reflectance (unitless), π is the mathematical constant equal to ~3.14159 (unitless),

L_λ is the spectral radiance at the sensor's aperture ($\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$), d is the Earth–Sun distance (astronomical units), E_λ is the mean exoatmospheric solar irradiance ($\text{W m}^{-2} \mu\text{m}^{-1}$) and θ_s is the solar zenith angle (degrees). d can be generated from the Jet Propulsion Laboratory (JPL) Ephemeris at <http://ssd.jpl.nasa.gov/?horizons> or can be obtained from the literature (e.g.[5]). Conversion to at-sensor spectral radiance and TOA reflectance can be performed by using built-in tools in high-end image processing software, such as ENVI.

This study attempts to simulate the RazakSAT land cover classification by making use of the Landsat data. The classification is carried out by means of Maximum Likelihood method.

II. MAXIMUM LIKELIHOOD (ML) CLASSIFICATION

ML is a supervised classification method derived from the Bayes theorem, which states that the a posteriori distribution $P(i|\omega)$, i.e., the probability that a pixel with feature vector ω belongs to class i , is given by:

$$P(i|\omega) = \frac{P(\omega|i)P(i)}{P(\omega)} \quad (3)$$

where, $P(\omega|i)$ is the likelihood function, $P(i)$ is the a priori information, i.e., the probability that class i occurs in the study area and $P(\omega)$ is the probability that ω is observed, which can be written as:

$$P(\omega) = \sum_{i=1}^M P(\omega|i)P(i) \quad (4)$$

where, M is the number of classes. $P(\omega)$ is often treated as a normalisation constant to ensure $\sum_{i=1}^M P(i|\omega)$ sums to 1. Pixel x is assigned to class i by the rule:

$$x \in i \quad \text{if } P(i|\omega) > P(j|\omega) \text{ for all } j \neq i \quad (5)$$

ML often assumes that the distribution of the data within a given class i obeys a multivariate Gaussian distribution. It is then convenient to define the log likelihood (or discriminant function):

$$g_i(\omega) = \ln P(\omega|i) = -\frac{1}{2}(\omega - \mu_i)^t C_i^{-1}(\omega - \mu_i) - \frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln(|C_i|) \quad (6)$$

Since \log is a monotonic function, Equation (3) is equivalent to:

$$x \in i \quad \text{if } g_i(\omega) > g_j(\omega) \text{ for all } j \neq i \quad (7)$$

Each pixel is assigned to the class with the highest likelihood or labelled as unclassified if the probability values are all below a threshold set by the user [7]. The general procedures in ML are as follows:

1. The number of land cover types within the study area is determined.
2. The training pixels for each of the desired classes are chosen using land cover information for the study area. For this purpose, the Jeffries-Matusita (JM) distance can be used to measure class separability of the chosen training pixels. For normally distributed classes, the JM separability measure for two classes, J_{ij} , is defined as follows [7]:

$$J_{ij} = \sqrt{2(1 - e^{-\alpha})} \quad (8)$$

where, α is the Bhattacharyya distance and is given by [7]:

$$\alpha = \frac{1}{8}(\mu_i - \mu_j)^t \left[\frac{(C_i + C_j)}{2} \right]^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left[\frac{|(C_i + C_j)|}{2\sqrt{|C_i||C_j|}} \right] \quad (9)$$

J_{ij} ranges from 0 to 2.0, where $J_{ij} > 1.9$ indicates good separability of classes, moderate separability for $1.0 \leq J_{ij} \leq 1.9$ and poor separability for $J_{ij} < 1.0$.

3. The training pixels are then used to estimate the mean vector and covariance matrix of each class.
4. Finally, every pixel in the image is classified into one of the desired land cover types or labeled as unknown.

In ML classification, each class is enclosed in a region in multispectral space where its discriminant function is larger than that of all other classes. These class regions are separated by decision boundaries, where, the decision boundary between class i and j occurs when:

$$g_i(\omega) = g_j(\omega) \quad (10)$$

For multivariate normal distributions, this becomes:

$$\begin{aligned}
 & -\frac{1}{2}(\boldsymbol{\omega} - \boldsymbol{\mu}_i)^t C_i^{-1} (\boldsymbol{\omega} - \boldsymbol{\mu}_i) - \frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln(|C_i|) - \\
 & \left(-\frac{1}{2}(\boldsymbol{\omega} - \boldsymbol{\mu}_j)^t C_j^{-1} (\boldsymbol{\omega} - \boldsymbol{\mu}_j) - \right. \\
 & \left. \frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln(|C_j|) \right) = 0
 \end{aligned} \tag{11}$$

which can be written as:

$$\begin{aligned}
 & -(\boldsymbol{\omega} - \boldsymbol{\mu}_i)^t C_i^{-1} (\boldsymbol{\omega} - \boldsymbol{\mu}_i) - \ln(|C_i|) + \\
 & (\boldsymbol{\omega} - \boldsymbol{\mu}_j)^t C_j^{-1} (\boldsymbol{\omega} - \boldsymbol{\mu}_j) + \ln(|C_j|) = 0
 \end{aligned} \tag{12}$$

This is a quadratic function in N dimensions. Hence, if we consider only two classes, the decision boundaries are conic sections (i.e. parabolas, circles, ellipses or hyperbolas).

III. ML USING LANDSAT DATA

The study area was located in Selangor, Malaysia, covering approximately 840 km² within longitude 101° 10' E to 101°30' E and latitude 2°99' N to 3°15' N. The satellite data come from bands 1, 2, 3, 4, 5 and 7 of Landsat-5 TM dated 11th February 1999 (Figure 3(b)). The data was chosen because have minimal cloud and free from haze, which normally occurs at the end of the year [2]. Prior to any data processing, the cloud and its shadow were masked out based on threshold approach [1]. Visual interpretation of the Landsat data was then performed, aided by a reference map (Figure 3(a)), produced in October 1991 using ground surveying and SPOT satellite data by the Malaysian Surveying Department and Malaysian Remote Sensing Agency. 11 main classes identified were water, coastal swamp forest, dryland forest, oil palm, rubber, industry, cleared land, urban, sediment plumes, coconut and bare land.

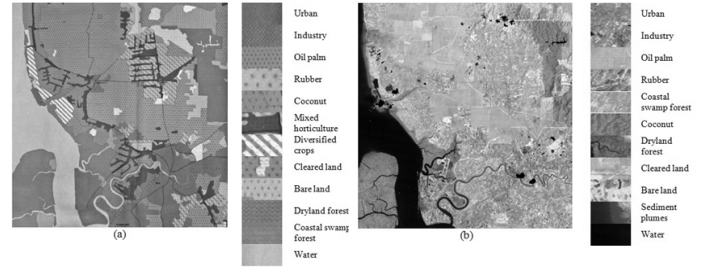


Figure 3. The study area from (a) the land cover map and (b) the Landsat-5 TM with bands 5 4 and 3 assigned to the red, green and blue channels. Cloud and its shadow are masked in black.

Training areas were established by choosing one or more polygons for each class. In order to select a good training area for a class, the important properties taken into consideration are its uniformity and how well they represent the same class throughout the whole image [8]. Class separability of the chosen training pixels (extracted from the training area) was determined by means of the JM distance. Fifty pairs have JM distance between 1.9 and 2.0 indicating good separability, four from 1.0 to 1.9 indicating moderate separability and one less than 1.0 indicating poor separability. The worst separability, possessed by the urban – industry pair (0.947), was expected since both have very similar spectral characteristics. For each class, these training pixels provide values from which to estimate the mean and covariances of the spectral bands used.

The outcome of ML classification after assigning the classes with suitable colours, is shown in Figure 4: coastal swamp forest (green), dryland forest (blue), oil palm (yellow), rubber (cyan), cleared land (purple), coconut (maroon), bare land (orange), urban (red), industry (grey), sediment plumes (sea green) and water (white). Clouds and their shadows are masked black. The areas in terms of percentage and square kilometres were also computed; the classes with the largest area are oil palm, cleared land and industry. Although being similar, coastal swamp forest and dryland forest can be clearly seen in the south-west and north-east of the classified image, as indicated by the reference map. Coastal swamp forest covers most of the Island and coastal regions in the south-west of the scene. Most of the dryland forest can be recognised as a large straight-edged region in the north-east. Oil palm and urban dominate the northern and southern parts respectively. Rubber appears as scattered patches that mostly are surrounded by oil palms. Industry can be recognised as patches near the urban areas, especially in the south-west and north-east. Coconut can be seen in the coastal area in the north-west of the image. A quite large area of bare land can be seen in the east, while cleared land can be seen mostly in the north, south and south-east of the image.

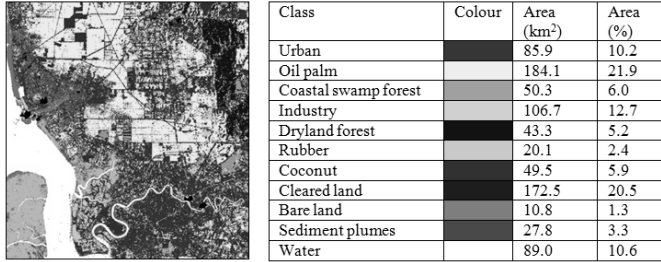


Figure 4. ML classification using band 1, 2, 3, 4, 5 and 7 of Landsat TM and the class areas in terms of square kilometre and percentage

Accuracy assessment of the ML classification was determined by means of a confusion matrix or sometimes called error matrix (Table 5) [10], which compares, on a class-by-class basis, the relationship between reference data (ground truth) and the corresponding results of a classification [15]. Such matrices are square, with the number of rows and columns equal to the number of classes, i.e. 11. For all classes, the numbers of reference pixels are: rubber (103), water (9129), coastal swamp forest (14840), dryland forest (6162), oil palm (10492), industry (350), cleared land (1250), urban (2309), coconut (159), bare land (313) and sediment plumes (1881). The diagonal elements in Table 5 represent the pixels of correctly assigned pixels and are also known as the producer accuracy. Producer accuracy is a measure of the accuracy of a particular classification scheme and shows the percentage of a particular ground class that is correctly classified. It is calculated by dividing each of the diagonal elements in Table 5 by the total of each column respectively:

$$\text{Producer accuracy} = \frac{c_{aa}}{c_{\bullet a}} \times 100\% \quad (13)$$

where,

c_{aa} = element at position a^{th} row and a^{th} column

$c_{\bullet a}$ = column sums

The minimum acceptable accuracy for a class is 90% [10]. It is obvious that all classes possess producer accuracy higher than 90%: bare land gives the highest (100%) and oil palm the lowest (92.4%). The relatively low accuracy of oil palm is mainly because 6% and 1% of its pixels were classified as coconut and cleared land. The misclassification of oil palm pixels to the coconut class is due to the fact that oil palm and coconut have a similar physical structure, so tend to have similar spectral behaviour and therefore can easily be misclassified as each other. User Accuracy is a measure of how well the classification is performed. It indicates the percentage of probability that the class which a pixel is classified to on an image actually represents that class on the ground [10]. It is calculated by dividing each of the diagonal

elements in a confusion matrix by the total of the row in which it occurs:

$$\text{User accuracy} = \frac{c_{ii}}{c_{i\bullet}} \times 100\% \quad (14)$$

where, $c_{i\bullet}$ = row sum . Coastal swamp forest, dryland forest, oil palm, sediment plumes, water, bare land and urban show a user accuracy of more than 90%. Rubber, cleared land and industry possess accuracy between 70% and 90%, while the worst accuracy is possessed by coconut (16%). The low accuracy of coconut is because the oil palm pixels tend to be classified as coconut because they having similar spectral properties to oil palm. A measure of overall behaviour of the ML classification can be determined by the overall accuracy, which is the total percentage of pixels correctly classified:

$$\text{Overall accuracy} = \frac{\sum_{a=1}^U c_{aa}}{Q} \times 100\% \quad (15)$$

where, Q and U is the total number of pixels and classes respectively. The minimum acceptable overall accuracy is 85% [6]. The Kappa coefficient, κ is a second measure of classification accuracy which incorporates the off-diagonal elements as well as the diagonal terms to give a more robust assessment of accuracy than overall accuracy. It is computed as [9]:

$$\kappa = \frac{\sum_{a=1}^U \frac{c_{aa}}{Q} - \sum_{a=1}^U \frac{c_{a\bullet} \cdot c_{\bullet a}}{Q^2}}{1 - \sum_{a=1}^U \frac{c_{a\bullet} \cdot c_{\bullet a}}{Q^2}} \quad (16)$$

where, $c_{a\bullet}$ = row sums . The ML classification yielded an overall accuracy of 97.4% and kappa coefficient 0.97, indicating very high agreement with the ground truth.

TABLE 5. CONFUSION MATRIX FOR ML CLASSIFICATION USING LANDSAT DATA

Overall Accuracy = 97.4%
Kappa Coefficient = 0.97

ML Classification (pixels)	Class	Ground Truth (Pixels)											Total classified pixels
		Coastal swamp forest	Dryland forest	Oil palm	Rubber	Cleared land	Sediment plumes	Water	Coconut	Bare land	Urban	Industry	
Coastal swamp forest	14801	0	0	0	0	0	1	0	0	0	0	0	14802
Dryland forest	0	6116	3	1	0	0	0	0	0	0	0	0	6120
Oil palm	0	9	9690	0	18	0	0	8	0	0	0	0	9725
Rubber	0	24	0	102	4	0	0	0	0	0	0	0	130
Cleared land	0	4	111	0	1173	3	0	0	0	123	1	0	1415
Sediment plumes	33	2	12	0	9	1804	0	4	0	0	0	0	1864
Water	0	0	0	0	0	0	9119	0	0	0	0	0	9119
Coconut	0	5	672	0	8	74	0	147	0	0	0	0	906
Bare land	0	0	0	0	4	0	0	0	313	0	0	0	317
Urban	0	0	4	0	11	0	0	0	0	2154	0	0	2169
Industry	6	2	0	0	23	0	9	0	0	32	349	421	421
Total ground truth pixels	14840	6162	10492	103	1250	1881	9129	159	313	2309	350	46988	

IV. ML USING SIMULATED RAZAKSAT DATA

Table 6 shows relationship between Landsat and RazakSAT satellite. For comparison, some of MODIS bands are also shown. Compared to Landsat, RazakSAT only has 4 multispectral bands and does not measure in mid-infrared wavelengths. Here, ML classification using bands 1, 2, 3 and 4 of Landsat dataset (i.e. correspond to bands 1, 2, 3 and 4 of RazakSAT) was performed in order to simulate RazakSAT classification. The labelling and colour scheme in Figure 4 were used. Figure 5 is the ML classification using (a) Landsat bands 1, 2, 3 and 4 and (b) same as shown in Figure 4, but is shown again for convenient. The bottom images are the enlarged versions of the region in the blue box in the top images. The circles indicate areas where apparent misclassification occurred; i.e. cleared land pixels (bottom circle) and industry pixels (top circle) being misclassified as urban.

TABLE 6. RELATIONSHIP BETWEEN MODIS, LANDSAT AND RAZAKSAT SATELLITE

Applications	MODIS			Landsat			RazakSAT		
	Spatial Resolution (m)	Band No.	Band Range (µm)	Spatial Resolution (m)	Band No.	Band Range (µm)	Spatial Resolution (m)	Band No.	Band Range (µm)
Land/Cloud/Aerosols Boundaries	250	1	0.620 – 0.670 (Red)	30	3	0.626 – 0.693 (Red)	5	3	0.630 – 0.690 (Red)
		2	0.841 – 0.876 (Near infrared)	30	4	0.776 – 0.904 (Near infrared)	5	4	0.760 – 0.890 (Near infrared)
Land/Cloud/Aerosols Properties	500	3	0.459 – 0.479 (Blue)	30	1	0.452 – 0.518 (blue)	5	1	0.450 – 0.520 (Blue)
		4	0.545 – 0.565 (Green)	30	2	0.528 – 0.609 (Green)	5	2	0.520 – 0.600 (Green)
		6	1.628 – 1.652 (Mid infrared)	30	5	1.55 – 1.75 (Mid infrared)	-	-	-
		7	2.105 – 2.155 (Mid infrared)	30	7	2.08 – 2.35 (Mid infrared)	-	-	-
Ocean Colour/Phytoplankton/Biogeochemistry	1000	10	0.483 – 0.493 (Blue)	30	1	0.452 – 0.518 (blue)	5	1	0.450 – 0.520 (Blue)
		11	0.526 – 0.536 (Green)	30	2	0.528 – 0.609 (Green)	5	2	0.520 – 0.600 (Green)
		12	0.546 – 0.556 (Green)	30					
		13	0.662 – 0.672 (Red)	30	3	0.626 – 0.693 (Red)	5	3	0.630 – 0.690 (Red)
		14	0.673 – 0.683 (Red)	30					
		15	0.743 – 0.753 (Near infrared)	30					
16	0.862 – 0.877 (Near infrared)	30	4	0.776 – 0.904 (Near infrared)	5	4	0.760 – 0.890 (Near infrared)		
Surface/Cloud Temperature	1000	31	10.780 – 11.280	120	6	10.40 – 12.50 (Thermal infrared)	-	-	-
		32	11.770 – 12.270	30					

The corresponding confusion matrix is given in Table 7; the overall classification accuracy declines from 97% (Figure 5(b)) to 88% (Figure 5(a)).

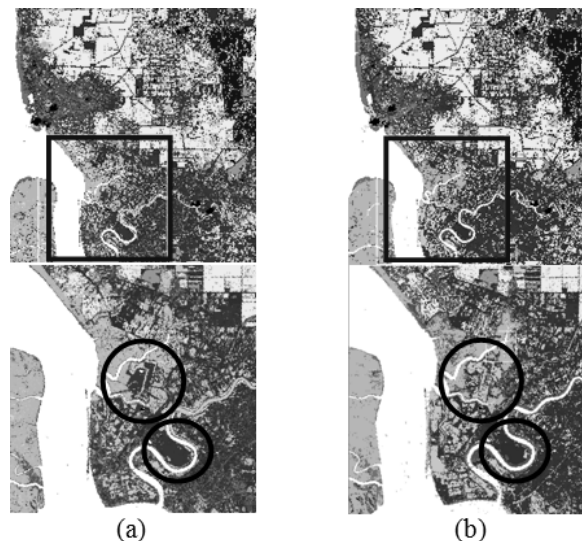


Figure 5. ML classification of the study area in Selangor, Malaysia using (a) bands 1, 2, 3 and 4 (top), (b) bands 1, 2, 3, 4, 5 and 7 (top) and enlarged versions of the area in the blue box (bottom). The Circles indicate areas where apparent misclassification occurred.

Table 8 shows the producer and user accuracy of simulated RazakSAT data for all classes; those of Landsat is also given for comparison. To better visualise the difference, plots of producer and user accuracy for the simulated RazakSAT and Landsat data are given in Figure 6. The highest producer accuracy is possessed by bare land, industry and water, while urban, cleared land and sediment plumes have the worst accuracy. For user accuracy, water, bare land and coastal swamp forest have the highest accuracy while rubber, coconut and cleared land have the lowest accuracy.

The lower user accuracies of rubber, coconut and cleared land indicates high commission error for these classes, meaning that there is a probability (proportionate to the error) that pixel classified as rubber, coconut and cleared land may not actually exist on the ground. On the other hand, urban, cleared land and sediment plumes have the worst producer accuracy due to the high omission error, meaning that there is a probability (proportionate to the errors) that the pixels for these classes were classified incorrectly. The maximum wavelength that RazakSAT can sense is about 0.9 µm, while that of Landsat is 2.35 µm. Wavelengths beyond 1.3 µm, e.g. 1.4 µm and 1.9 µm are known as water absorption bands are useful to differentiate between land and water boundary as well as between different plant types. The water in the leaf absorbs strongly at these wavelength. Clearly, these wavelengths are useful to differentiate plants due to the different amount of water stored in different plant leaves. This leads to the smaller producer and user accuracy of the simulated RazakSAT data compared to the Landsat

data in most plant classes. It can be seen that non plant classes, i.e. urban and cleared land, are also affected due to normally having minimal amount vegetation and surfaces with different moisture content. As expected the least difference is shown by industry and bare land due to having almost no vegetation and covered with uniform reflective surfaces.

This shows that when only considering the spectral capability of the bands, RazakSAT tends to produce classification with lower accuracy than Landsat. However, RazakSAT bands have a higher spatial resolution (5 m) than Landsat (30 m) and is equipped with a panchromatic band (2.5 m spatial resolution), which may be able to increase the classification accuracy.

TABLE 7. CONFUSION MATRIX FOR ML CLASSIFICATION USING SIMULATED RAZAKSAT DATA

Overall Accuracy = 88.5%
Kappa Coefficient = 0.86

Class	Ground Truth (Pixels)											Total classified pixels
	Coastal swamp forest	Dryland forest	Oil palm	Rubber	Cleared land	Sediment plumes	Water	Coconut	Bare land	Urban	Industry	
Coastal swamp forest	12675	132	267	7	0	0	1	8	0	0	0	13090
Dryland forest	238	5662	2	7	0	0	0	0	0	0	0	5909
Oil palm	317	7	9172	0	22	2	0	6	0	0	0	9526
Rubber	1163	355	130	89	11	0	0	1	0	0	0	1749
Cleared land	4	3	232	0	997	11	0	0	1	335	3	1586
Sediment plumes	75	0	41	0	6	1502	198	3	0	154	2	1981
Water	1	0	0	0	0	0	8919	0	0	0	0	8920
Coconut	261	0	645	0	16	305	0	141	0	2	0	1370
Bare land	0	0	0	0	6	0	0	0	312	4	0	322
Urban	106	2	3	0	161	56	5	0	0	1770	1	2104
Industry	0	1	0	0	31	5	6	0	0	44	344	431
Total ground truth pixels	14840	6162	10492	103	1250	1881	9129	159	313	2309	350	46988

TABLE 8. PRODUCER AND USER ACCURACY OF CLASSIFICATION USING SIMULATED RAZAKSAT AND LANDSAT DATA

Class	Producer Accuracy (%)		User Accuracy (%)	
	RazakSAT	Landsat	RazakSAT	Landsat
Coastal swamp forest	85.41	99.74	96.83	99.99
Dryland forest	91.89	99.25	95.82	99.93
Oil palm	87.42	92.36	96.28	99.64
Rubber	86.41	99.03	5.09	78.46
Cleared land	79.76	93.84	62.86	82.90
Sediment plumes	79.85	95.91	75.82	96.78
Water	97.7	99.89	99.99	100.00
Coconut	88.68	92.45	10.29	16.23
Bare land	99.68	100.00	96.89	98.74
Urban	76.66	93.29	84.13	99.31
Industry	98.29	99.71	79.81	82.90

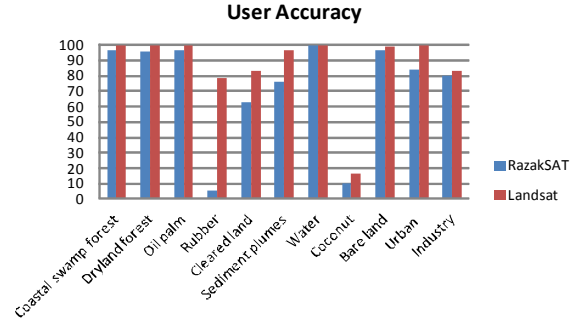
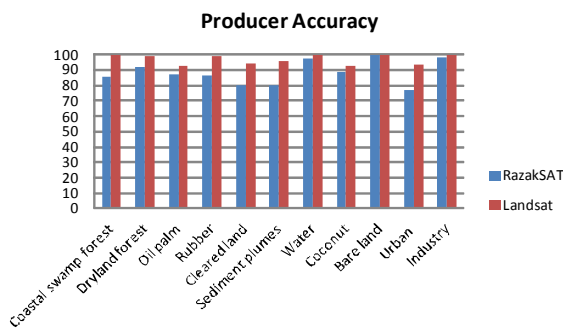


Figure 6. Plots of producer (top) and user (bottom) accuracy for simulated RazakSAT and Landsat data

V. CONCLUSION

The study presented here makes use of the moderate resolution Landsat 5 TM data to simulate the performance of RazakSAT classification. The unavailability of RazakSAT data allows limited analysis to anticipate its actual performance. The simulation of RazakSAT classification analysis using Landsat data shows that although without the mid-infrared bands, RazakSAT is still able to classify the study area with accuracy only slightly lower than Landsat. It is undeniable that the main shortcoming of RazakSAT is the unavailability of mid-infrared bands that hampers its capability to distinguish between vegetations and surfaces. Nonetheless, the availability of a panchromatic band and the higher spatial resolution of the multispectral bands confer major advantages of RazakSAT compared to the Landsat.

ACKNOWLEDGMENT

The author would like to thank the Malaysian Ministry of Higher Education (MOHE) and Universiti Teknikal Malaysia Melaka (UTeM) for funding this study and the Malaysian Remote Sensing Agency (ARSM), Ministry of Science, Technology and Innovation (MOSTI) for providing the data.

REFERENCES

- [1] A. Ahmad, and S. Quegan (2012), Cloud masking for remotely sensed data using spectral and principal components analysis, *Engineering, Technology & Applied Science Research (ETASR)*, 2, 221 – 225.
- [2] A. Ahmad, M.N. Ayof, S.B. Agus, H. Sakidin, and S.S.S. Ahmad (2006), Smoke detection using remote sensing technique. *Prosiding Seminar Pencapaian Penyelidikan KUTKM 2006 (REACH 06)*. 284-293.
- [3] B.L. Markham, J.C. Storey, D.L. Williams. and J.R. Irons (2004), Landsat sensor performance: history and current status, *IEEE Transaction on Geoscience and Remote Sensing*, 42, 2691 – 2694.
- [4] ENVI (2006), *User's guide*, Research Systems Inc., USA.
- [5] G. Chander, B.L. Markham and D.L. Helder (2009), Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+ and EO-1 ALI sensors, *Remote Sensing of Environment*, 133, 893 – 903.

- [6] J. Scepan (1999), Thematic validation of high-resolution global land-cover data sets, *Photogrammetric Engineering and Remote Sensing*, 65, 1051 – 1060.
- [7] J.A. Richards 1999, *Remote sensing digital image analysis: An introduction*. Springer-Verlag, Berlin, Germany.
- [8] J.B. Campbell 2002, *Introduction to remote sensing*, Taylor & Francis, London.
- [9] J.R. Jensen (1996), *Introductory Digital Image Processing: A Remote Sensing Perspective*, Pearson Prentice Hall, New Jersey, USA.
- [10] M. Story and R. Congalton (1986), Accuracy assessment: a user's perspective, *Photogrammetric Engineering and Remote Sensing*, 52, 397 – 399.
- [11] N.M. Yusoff (2005), RazakSAT - Technology advent in high resolution imaging system for small satellite, *Proceedings of the 26th ACRS (Asian Conference on Remote Sensing) Hanoi, Vietnam, Nov. 7-11, 2005*, AARS, Hanoi.
- [12] N.M. Yusoff, S.A. Ahmad and M. Othman (2002), Utilization of TiungSAT-1 data for environmental assessment and monitoring, *Canadian International Development Agency (CIDA)-Canadian Space Agency (CSA) Conference, Space Applications for Sustainable Development, Hull, Canada, May 21-22, 2002*, CIDA-CSA, Canada.
- [13] N. S. Wai, A. A.Tan, J. W. S. Mee, Ismail, M. & M.D. Subari, (2009). Preflight Radiometric Calibration of RazakSAT™. *Proceedings of International Conference on Recent Advances in Space Technologies, 2009 (RAST '09)*, 277 – 282.
- [14] S.A. Ackerman, K.I. Strabala, W.P. Menzel, R.A. Frey, C.C. Moeller and L.E. Gumley (1998), Discriminating clear-sky from clouds with MODIS, *Journal of Geophysical Research*, 103, 32141 – 32157.
- [15] T.M. Lillesand, R.W. Kiefer and J.W. Chipman (2004), *Remote Sensing and Image Interpretation*, John Wiley & Sons, Hoboken, NJ, USA.
- [16] USGS (2011), USGS Website, available at: http://landsat.usgs.gov/about_landsat5.php.