SCIENCE LETTERS

Volume 2 Issue 1 2005

ISSN 1675 - 7785

SCIENCE LETTERS
Volume 2, Issue 1, June 2005

CONTENTS

Deadhesion Mechanism Of Commercial Unpigmented Epoxy And Alkyd Coating Under Cathodic Polarisation — 1
M.K. Harun and M.Z.A. Yahya

Thermo–Optical Properties Of Polymer Planar Waveguides — 13
Mohd Kamil Abd Rahman, Aiman Kassir, Abang Annuar Ehsan, Noraspalelawati Razali and Sahbudin Shaari

Conductivity Study On Plasticized Chitosan Acetate Based Electrolytes Doped With Silver Triflate Salt System — 23
A.M.M. Ali, M.Z.A. Yahya, M.K. Harun and A.A. Mohamad

Enhanced Levitation Force Of Partial Melted $(Y,Er)Ba_2Cu_3O_{7-\delta}$ Superconductors With Formation Of Minor 211 Phase — 31
Ahmad Kamal Yahya, Suzi Ahmad and Imad Hamadneh

Bi-Based Mono-Core Superconductor Tapes: Fabrication Process And The Effect Of The Thermomechanical Treatment On The Transport Properties — 41
K.T. Lau and R-Abd Shukor

Physical And Mechanical Properties Of Fibreboards From Oil Palm Fibres And Polyethylene — 49
Mohd Ariff Jamaludin, Yul Haidzir Razali Bujang, Ridzuan Ramli, Kamarulzaman Nordin and Mansur Ahmad

Developing A Solid Phase Microextraction Method For The Identification Of Accelerants In Fire Debris — 55
Abdul Aziz Haron and Nor'ashikin Saim

Chrysophanol From Aloe Sap: Isolation, Characterization And Biological Activity Study — 63
Pat M. Lee, Hannis Fadzillah Mohsin and Kong Hung Lee

Antioxidative Activity Of Ginger And Coriander In Cooked Patties Of Mackerel (*Scomber Scombrus)* During Storage — 71
Normah I, Diana W, Norisuliana I and Nur Azura A

Logistic Regression Modelling Of Thematic Mapper Data For Rubber (*Hevea Brasiliensis*) Area Mapping — 79
Mohd Nazip Suratman, Valerie M. LeMay, Gary Q. Bull, Donald G. Leckie, Nick Walsworth and Peter L. Marshall

LOGISTIC REGRESSION MODELLING OF THEMATIC MAPPER DATA FOR RUBBER (*HEVEA BRASILIENSIS*) AREA MAPPING

Mohd Nazip Suratman[a]*, Valerie M. LeMay[b], Gary Q. Bull[b], Donald G. Leckie[c], Nick Walsworth[c] and Peter L. Marshall[b]

[a]*Faculty of Applied Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.*
[b]*Faculty of Forestry, University of British Columbia, Vancouver, B.C. V6T 1Z4 Canada*
[c]*Pacific Forestry Centre, Canadian Forest Service, Victoria, B.C. V8Z 1M5 Canada*

**Corresponding author: Tel : +603-5544 4538, Fax : +603-5544 4562*
email : nazip@salam.uitm.edu.my

ABSTRACT

Logistic regression modelling of Landsat Thematic Mapper (TM) was applied for mapping the area of rubber plantations in the study area of Selangor, Malaysia. TM bands 2 – 5 and 7 were included in the final logistic regression model, and all were highly significant at the 0.0001 level. The χ^2 *value (23247.9) for the model was highly statistically significant (*$P<0.0001$*), which implies the estimated model fitted the model building data. TM bands 4 and 5 were the two most influential variables affecting the odds of rubber area occurrence on the imagery. Using probabilities of* ≥ 0.5*, the model correctly classified 94.5% of the observations in both the training and validation data sets. This high accuracy suggests that the model is appropriate for predicting the presence of rubber trees in the pixels based on selected spectral bands measured by Landsat TM.*

Keywords: Landsat TM, logistic regression, rubber plantation, area modelling

1. INTRODUCTION

Rubber tree crops (*Hevea brasiliensis* [Wild. ex Adr. de Juss.] Muell Arg.) play pivotal roles in providing natural rubber, sources of wood products, and other benefits that support basic human needs and for economic development[1-4]. Despite their importance, little attention has been given to mapping and monitoring the crops. Land cover mapping using remote sensing techniques are now becoming essential tools in forest management activities[5]. These techniques rely on the assumption that land cover produces reflectance[6], which allows the

ISSN 1675-7785

identification and characterization of the cover types[7]. A widely used method in modelling probability is logistic regression analysis[8,9]. This approach proved useful in examining the relationship between a set of predictor variables and a response variable which takes only two dichotomous values[10,11]. In the case of rubber plantation area mapping, logistic model can be structured to estimate the probability based on whether or not a pixel belongs to rubber area, and consequently be classified as rubber or non-rubber. The objective of this study is to develop a modelling method for mapping the area of rubber plantations using Landsat TM data.

2. EXPERIMENTAL METHODS

2.1 Study Area, Image Acquisition and Pre-Processing

The study area chosen was the state of Selangor (796,084 km^2), Malaysia, located between latitudes of 2° 35' to 3° 55' N and longitudes of 100° 45' to 102° 00' E. The image used was a full scene of Landsat-5 Thematic Mapper (TM), path 157 and row 58, recorded on February 11, 1999. Spatial resolution was 30 m × 30 m for TM bands 1 – 5, and 7. The Landsat scene was geometrically corrected using a total of 36 ground control points (GCP) gathered throughout the image using PCI software (Version 7.0)[12]. A root mean squared error (RMSE) of less than 0.5 pixels was attained for these control points. Available reference data to support this work consisted of ground truth data, topographic and land use/cover maps produced by the Department of Survey and Mapping and Department of Agriculture, Malaysia, respectively.

2.2 Logistic Regression Model for Rubber Area Mapping

Logistic regression is a method that was devised for dichotomous outcomes[8]. In this study, the response variable for rubber pixels is coded to the value 1, while for the non-rubber pixels the response variable is coded to the value 0. The set of the predictor variables consists of the spectral channels of rubber and non-rubber and consequently the values of each predictor variable are the digital numbers (DN) of the pixel of each spectral band (TM bands 1 – 5, and 7). The probability of an event (i.e., rubber or non-rubber) occurring is given by the following general logistic regression model:

$$\hat{p} = \frac{\exp(b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k)}{1 + \exp(b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k)}$$

where $\hat{p}$ is the predicted probability of the outcome, b_1, …, b_k are the coefficients associated with each predictor variable x_1, …, x_k.

2.3 Establishment of the Data Set

For logistic model development, the sample areas of rubber and non-rubber were collected and delineated throughout the study area and then randomly split into training and validation areas. The

ISSN 1675-7785

boundaries of the selected areas were delineated on the imagery and corresponding spectral measurements for each area were extracted from the TM bands for subsequent analyses.

The training areas contained 6,589 pixels of rubber and 26,884 pixels of non-rubber collected from 28 separate rubber and 53 non-rubber sites from four vegetation classes: forest, grassland, mixed-crops, and oil palm. Model validation areas contained 3,258 pixels of rubber and 13,536 pixels of non-rubber from 53 independent sites. Verification of these selected areas was made with the aid of reference data. The maximum-likelihood method was used to estimate the coefficients using the logistic regression procedure (PROC LOGISTIC) in the SAS statistical package, version 8.1[13]. The significance of each parameter was tested using the Wald χ^2 statistic (α = 0.05). The goodness of fit of the logistic model was evaluated by means of the -2 log L (L=likelihood) statistic. To obtain the final model, stepwise selection method was used to eliminate the least correlated variables. The predictive accuracy of the model was interpreted by comparing the number of correct and incorrect classifications in a 2 $\times$ 2 classification table. The final model was then applied to a demonstration area to display output bitmaps of predicted rubber pixels on the imagery.

3. RESULTS AND DISCUSSION

The χ^2 value (23247.90) for the logistic model was highly statistically significant ($P<0.0001$), implying that the estimated model containing five predictor variables explained some of the differences between rubber versus non-rubber on the imagery (Table 1).

Based on stepwise selection method, TM bands 2 – 5 and 7 were included in the logistic model, and all were highly significant at the 0.0001 level (Table 1). TM bands 4 and 5 were the two most influential variables affecting the odds of rubber area occurrence on the imagery. These results are consistent with a number of other studies that have found that near- and mid-infrared reflectance provide the best correlation with vegetation parameters[14,15].

Examining the TM band statistics extracted from the sample areas indicated that TM bands 4 and 5 had the largest dynamic ranges for both model building and the validation data sets. The corresponding model equation for this data is:

$$\hat{p} = \frac{\exp(14.9 - 0.58 * B2 - 0.32 * B3 - 0.33 * B4 + 0.47 * B5 - 0.06 * B7)}{1 + \exp(14.9 - 0.58 * B2 - 0.32 * B3 - 0.33 * B4 + 0.47 * B5 - 0.06 * B7)}$$

where: $\hat{p}$ is predicted probability of pixels being rubber.

The estimated logistic regression model was applied to the validation data set, reserved for testing. The predictive accuracy of the logistic regression as applied to each of the data sets is shown in Table 2. Correctly classified

ISSN 1675-7785

observations appear on the main diagonal. Using probabilities $\geq$ 0.5 as rubber, the model appeared to have good predictive power, correctly classifying 94.5% of the observations in both the model-building and validation data sets. The model did a better job in predicting non-rubber (96%) than that of rubber (87%) for both data sets (Table 2). This high accuracy suggests that the model is appropriate for predicting the presence of rubber plantation areas based on selected spectral bands measured by Landsat TM.

The model developed here was applied to an independent demonstration area consists of three districts in Selangor namely Petaling, Sepang, and Kuala Langat. This produced the predicted output bitmaps which display the spatial distribution representation of rubber plantations for the demonstration area (Figure 1).

Table 1. Prediction of rubber areas using a logistic model (n=33,473 pixels).

Predictor variables	Estimates of β	Standard error	Wald χ^2	P-value	Odds ratio (Exp(β))
Intercept	14.901	0.367	1634.292	<0.0001	
Band 2	-0.575	0.028	421.779	<0.0001	0.56
Band 3	-0.323	0.020	261.890	<0.0001	0.73
Band 4	-0.326	0.005	3691.774	<0.0001	0.72
Band 5	0.472	0.009	2820.408	<0.0001	1.60
Band 7	-0.062	0.016	15.182	<0.0001	0.94

Model χ^2=23247.90

Table 2. Classification accuracy of logistic regression model for predicting rubber areas.

		Training (n=33,473 pixels)			Validation (n=16,794 pixels)		
		Predicted		Percent	Predicted		Percent
		0	1	correct	0	1	correct
Observed	0	25,905	979	96.4	13,027	509	96.2
	1	846	5,743	87.2	409	2,849	87.4
Overall				94.5			94.5

Notes: 0 denotes "non-rubber", 1 denotes "rubber"

ISSN 1675-7785

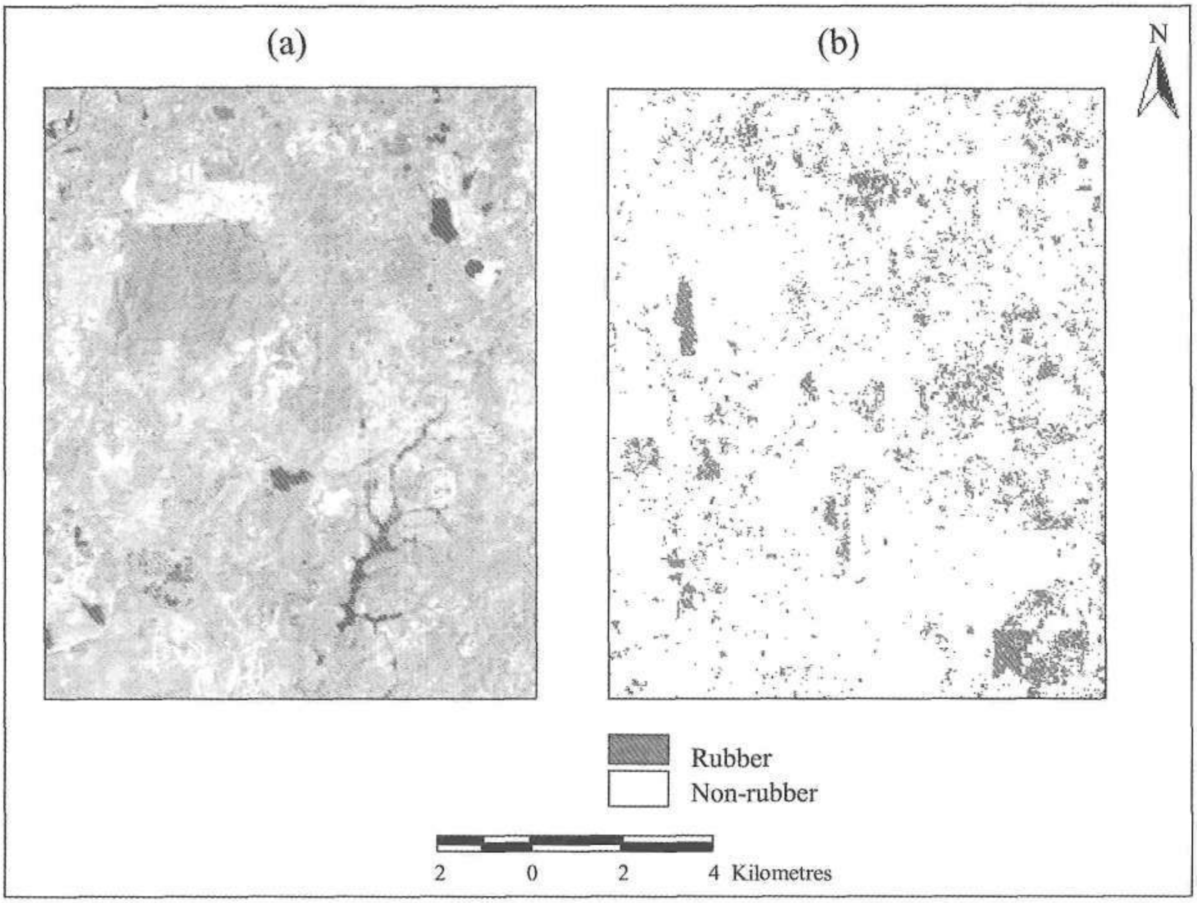

Figure 1: (a) Landsat TM pseudocolour composite (RGB = 5, 4, 3) image of the demonstration area (districts of Petaling, Sepang and Kuala Langat), and (b) output bitmaps of rubber area from the application of logistic regression model for the same area.

4. CONCLUSION

This study examined whether logistic modelling of Landsat TM data acquired could be applied in rubber area mapping. The statistical indications, which have arisen from the structure and application of logistic regression model and the estimated accuracies, show that the proposed methodology can be applied in predicting the presence of rubber trees in the pixels with promising results. In conclusion, logistic regression modelling applied to satellite data proved to be suitable statistic tool in predicting a response variable that takes only two values and consequently can be used in a classification process in a dichotomous way.

ACKNOWLEDGEMENTS

We acknowledged the International Council for Canadian Studies (ICCS),

ISSN 1675-7785

and the Canadian Forest Service for the financial support they offered to us. We also acknowledged the Canadian Forest Service and University of British Columbia for providing valuable resources for remote sensing data processing. We thank the Malaysian Rubber Board (MRB) for permitting us to use the rubber experimental station, the Rubber Industry Smallholding Development Authority (RISDA) for providing logistic assistance.

REFERENCES

1. Forestry Department Peninsular Malaysia. (1997). Country report – Malaysia. Forestry Department Headquarters, Peninsular Malaysia. Asia Pacific For. Sector Outlook Study Working Pap. Series. No. APFSOS/WP/07. FAO, Rome. 24 pp.
2. Killmann, W. and Hong, L.T. (2000). "Rubberwood – the success of an agricultural by-product". *Unasylva*. 201(51): 66–72.
3. MRB. (2002). Malaysian rubber boards rubber statistic (Web page). Available at: http://www.lgm.gov.my. 16 Sep. 2003.
4. RISDA. (2002). Rubber industry small landholders' development authority. (Web page). Available at: http://www.risda.gov.my. 16 Sep. 2003.
5. Rogan, J., Franklin, J. and Roberts, D.A. (2002). "A comparison of method of monitoring multitemporal vegetation change using Thematic Mapper Imagery". *Remote Sens. Environ.* 80: 143–156.
6. Yuan, D. and Elvidge, C. (1998). "NALC land cover change detection pilot study: Washington D.C. area experiments". *Remote Sens. Environ.* 66(2): 166–178.
7. Mertens, B. and Lambin, E.F. (1999). "Modelling land-cover dynamics: Integration of fine scale land cover data with landscape attributes". *Intl. J. of App. Earth Obs. and Geoinformation*. 1(1): 48–52.
8. Hosmer, D.W. and Lemeshow, S. (2000). Applied logistic regression. 2nd ed., John Wiley and Sons, NY. 128 pp.
9. Collet, D. (1991). Modelling binary data. Chapman and Hall, London. 369
10. Pereira, J.M.C. and. Itami, R.M. (1992). "GIS-based habitat modelling suing logistic multiple regression: a study of the Mt. Graham Red Squirrel". *Photogram. Engineering and Remote Sens.* 57:1475-1486.
11. Bian, L. and West, E. (1997). "GIS modelling of elk calving habitat in a prairie environment with statistics". *Photogram. Engineering and Remote Sens.* 63:161-167.
12. PCI. (1997). ImageWorks. Version 7.0. Richmond Hill, Ontario, Canada. 203 pp.
13. SAS Institute Inc. 2000. SAS/STAT®User's guide, version 8, vol. 1, Cary, NC, USA. 1031 pp.
14. Ripple, W.J., Wang, S., Isaacson, D.L, and Paine, D.P. (1991). "A preliminary comparison of Landsat TM and SPOT-1 HRV multispectral data for estimating coniferous forest

ISSN 1675-7785

volume." Intl. J. Remote Sens. 12(9): 1971–1977.
15. Brockhaus, J.A. and Khorram, S. (1992). "Comparison of SPOT and Landsat-TM data for forest inventories". *Intl. J. Remote Sens.* 13(16): 3035–3043.

ISSN 1675-7785