III

COPY

No 3

77-10245
CR-155035

AN ANALYSIS ON VEGETATION COVER BY USING LANDSAT MSS DATA

by

Shunji MURAI
Dr. Eng. Associate Professor
Institute of Industrial Science
University of Tokyo
Roppongi 7-22, Minatoku
Tokyo JAPAN

(E77-10245) AN ANALYSIS ON VEGETATION COVER BY USING LANDSAT MSS DATA (Tokyo Univ.) 11 p HC A02/MF A01 CSCL 08F
N77-33576
Unclas
G3/43 00245

ABSTRACT

The study deals with a mathematical model for estimating vegetation cover in Tokyo Districts by using LANDSAT MSS data. Two types of model, multi-regression analysis and parametric model, were tested for the geographically corrected test area with 45 kilometers by 18 kilometers in Tokyo Districts, where the ground survey for vegetation cover and classification was carried out by the Ministry of Construction, the Government of Japan in 1973.

The results obtained from the study conclude that vegetation cover can be well estimated from LANDSAT MSS data with the accuracy of not less than 80 per cent.

INTRODUCTION

Vegetation cover can be defined as the ratio of the area with vegetation cover to the unit area. However, recognition of vegetation varies depending on the definition of vegetation. In 1973, the Ministry of Construction carried out the ground survey associated with multi-band camera aerial photographs for mapping the vegetation and landuse in Tokyo Districts within 50 kilometers radius on the base of 500 meters mesh data.

Data for the unit area were digitized with respect to sixteen classes of vegetation and land use, which are utilized as the ground truth data in the study.

SELECTION OF TEST AREA

Test area was selected where Tokyo down town, suburb and hill are properly included as shown in Fig. 1. The test area corresponds to two frames of national base map, 1:50,000, namely ' HACHIOJI ' and ' TOKYO SEINANBU ' of which area makes about 800 square kilometers with 45 km in east-west direction by 18 km in north-south direction. Since so-called 500 meters mesh is made by deviding a frame of the base map of 1:50,000 into 40 grids in both direction of longitude and latitude, the test area involves 3200 mesh data.

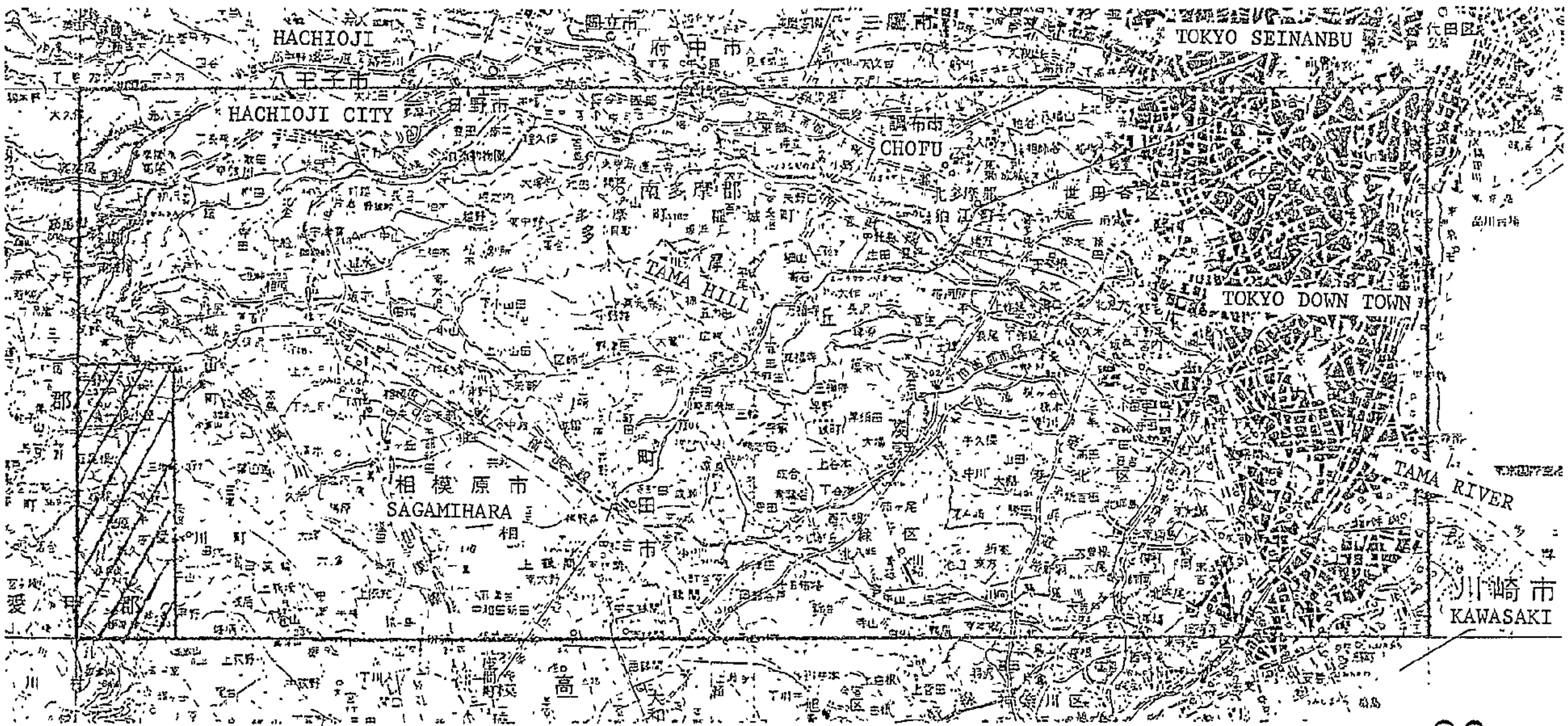

Fig. 1 Test Area

However, 3080 mesh data were utilized because 120 data were not valid for the ground survey.

LANDSAT MSS digital data were geometrically and geographically corrected so as to be registed to the above mensioned base map, of which unit pixel was resampled with the size of 100 meters by 100 meters. Therefore, approximetely 25 pixels of LANDSAT data corresponds to a unit 500 meters mesh data.

GROUND TRUTH DATA

Vegetation and land use were classified into sixteen classes some of which were selected as components of vegetation cover depending on the following three cases of definition, as shown in Table 1.

(1) Case 1: Vegetation cover, G_1.
G_1 includes ten classes except built-up area, disturbed area, street tree, and other.

(2) Case 2: Vegetation cover , G_2.
30 per cent of built-up area with vegetation is added to G_2.

(3) Case 3: Vegetation cover , G_3.
G_3 is corrected by subtracting 50 per cent of barren land from G_2, because actual barren land does not always involve full of vegetation.

Table 1 Definition of Vegetation Cover

Class \ Vegetation Cover	G_1	G_2	G_3
(1) Evergreen Forest (Replanted)	1	1	1
(2) Evergreen Forest (Natural)	1	1	1
(3) Deciduous Forest	1	1	1
(4) Mixed Forest	1	1	1
(5) Bamboo Forest	1	1	1
(6) Park	1	1	1
(7) Rice field	1	1	1
(8) Crop land	1	1	1
(9) Orchard	1	1	1
(10) Built-up area with green	0	0.3	0.3
(11) Built-up area without green	0	0	0
(12) Street tree	0	0	0
(13) Water	0	0	0
(14) Barren land	1	1	0.5
(15) Disturbed Area	0	0	0
(16) Other	0	0	0

MATHEMATICAL MODEL

The following two models were tested for estimating vegetation cover from LANDSAT MSS data.

(1) Multi-regression model

Average LANDSAT data with four bands, X (x_1, x_2, x_3, x_4) in a unit 500 meters mesh data and the corresponding vegetation cover, y are correlated by the multi-regression model as indicated as follows.

$$y = f(X) = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_0$$

Unknown parameters, a_0 to a_4 are estimated by the multi-regression analysis.

(2) Parametric Model

When variable X, which is derived from LANDSAT MSS data amounts m in the unit 500 meters mesh area, let n of m be subject to the condition that variable X exceeds parameter k. Supposed that vegetation cover can be estimated as n/m, parameter k should be determined so as to minimize the standard deviation of residuals between observed vegetation cover G and estimated vegetation cover n/m.

Eight variables were examined in the parametric model.

$$X_1 = x_3 / x_1$$

$$X_2 = x_4 / x_1$$

$$X_3 = x_3 / x_2$$

$$X_4 = x_4 / x_2$$

$$X_5 = (x_3 + x_4)/(x_1 + x_2)$$

$$X_6 = (x_1 + x_2 - x_3 - x_4)/(x_1 + x_2 + x_3 + x_4)$$

$$X_7 = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_0$$
(determined by multi-regression model)

$$X_8 = b_1x_1^* + b_2x_2^* + b_3x_3^* + b_4x_4^*$$
(second principals determined by principal component analysis)

Where

$$x^* = (x - \bar{x})/ \sigma$$

ESTIMATION OF VEGETATION COVER BY MULTI REGRESSION ANALYSIS

Multi-regression analysis was applied to 3080 500 meters mesh data with three cases of ground truth data. Table 2 summarizes the result of the analysis including coefficients a_0 to a_4 and accuracy of estimation represented by multi correlation coefficient and standard deviation. Coefficients in the lower column represent those for normalized data.

Best result was obtained for the case 3, in which multi-correlation coefficient γ is 0.871 and the standard deviation σ is $\pm$15.53 per cent.

Table 2 Result of Multi-regression Analysis

	Coefficient					Accuracy	
	a_1	a_2	a_3	a_4	a_0	γ	$\sigma(\pm\%)$
Case 1 : G_1	-3.844	-4.839	3.697	6.002	81.368	0.845	17.82
	-0.242	-0.378	0.331	0.377			
Case 2 : G_2	-3.722	-5.140	3.874	5.781	85.638	0.856	16.92
	-0.237	-0.406	0.351	0.367			
Case 3 : G_3	-3.340	-5.229	3.241	6.338	83.417	0.871	15.53
	-0.220	-0.428	0.304	0.417			

Fig. 2 shows comparison between observed vegetation cover and estimated vegetation cover for the cross section lines no. 10 , and no. 25 . Fig. 3 shows the computer generated map for observed vegetation cover and estimated vegetation cover.

ESTIMATION OF VEGETATION COVER BY PARAMETRIC MODEL

Most probable parameter k, which would be considered as an criterion to decide whether vegetation or not, was determined for each of eight cases of variables as mentioned before, so as to minimize the standard deviation between observation and estimation.

Table 3 summarizes the result of estimation by parametric model. Best estimation was obtained from the variable $X_3 = x_3 / x_2$ that is the ratio of MSS 6 to MSS 5. The standard deviation for this case was $\pm$15.36% for the most probable parameter k=1.14. This result is of great importance from the view point of radiance, that is, since radiances for full scale of CCT data in MSS 5 and MSS 6 band correspond to 1.76 and 2.00 $mW/cm^2.sr$ respectively, the criterion $X_3 = x_3/x_2 \geqq k = 1.14$ coincides with the criterion that the ratio of radiances between the two bands, $1.76\, x_3 / 2.00 x_2 \geqq 1$, or $x_3/x_2 \geqq 2.00/1.76 =$ 1.14. Therefore, if a slope for the ratio of MSS 6 to MSS 5 with respect to radiance is positive, it could be concluded that the corresponding data represents vegetation cover.

Table 3 Result of Estimation of Vegetation Cover by Parametric Model

Variable	Case 1: G_1		Case 2: G_2		Case 3: G_3	
	k	σ	k	σ	k	σ
$X_1 = x_3/x_1$	0.93	20.61	0.90	21.43	0.94	20.05
$X_2 = x_4/x_1$	0.46	18.97	0.44	18.99	0.45	18.13
$X_3 = x_3/x_2$	1.16	17.27	1.14	16.74	1.14	15.36
$X_4 = x_4/x_2$	0.57	17.80	0.55	17.23	0.58	16.12
$X_5 = (x_3+x_4)/(x_1+x_2)$	0.76	17.90	0.75	17.88	0.76	16.66
$X_6 = (x_1+x_2-x_3-x_4)/(x_1+x_2+x_3+x_4)$	0.14	17.93	0.15	17.86	0.14	16.75
$X_7 = a_1x_1+a_2x_2+a_3x_3+a_4x_4+a_0$	49	17.58	47	17.36	48	16.08
$X_8 = b_1\overset{*}{x}_1+b_2\overset{*}{x}_2+b_3\overset{*}{x}_3+b_4\overset{*}{x}_4$	-0.1	17.85	-0.2	17.69	-0.1	16.58

Where X_7 : multi-regression model
X_8 : second principal component

COMPARISON BETWEEN MULTI-REGRESSION MODEL AND PARAMETRIC MODEL

The following comparison between multi-regression model and parametric model can be pointed out.

(1) Multi-regression model utilizes the average data from the stand point of statistic, while parametric model accumulates point by point data.

(2) Parametric model results in a little better estimation, although the both model have less remarkable difference. Tendencies for error distribution in the two models look alike.

(3) Computing time is much more needed for parametric model, while flexibility could be expected more in the parametric model.

CONCLUSION

(1) Vegetation cover in Tokyo Districts could be well estimated by both of multi-regression model and parametric model. Multi-correlation coefficient between observed value and estimated value was 0.87 and standard deviation was ±15%.

(2) Vegetation cover in Tokyo Districts was mapped into five levels with equal intervals of 20 per cent.

(3) Definition of vegetation should be further studied, although the parametric model concluded that the decision rule to define vegetation should be whether the slope of ratio MSS 6/MSS 5 in the radiance unit is positive or not.

(4) With detail check for the result of estimation obtained from LANDSAT data, actual vegetation cover could be much more represented in the estimated map than the ground-surveyed map.

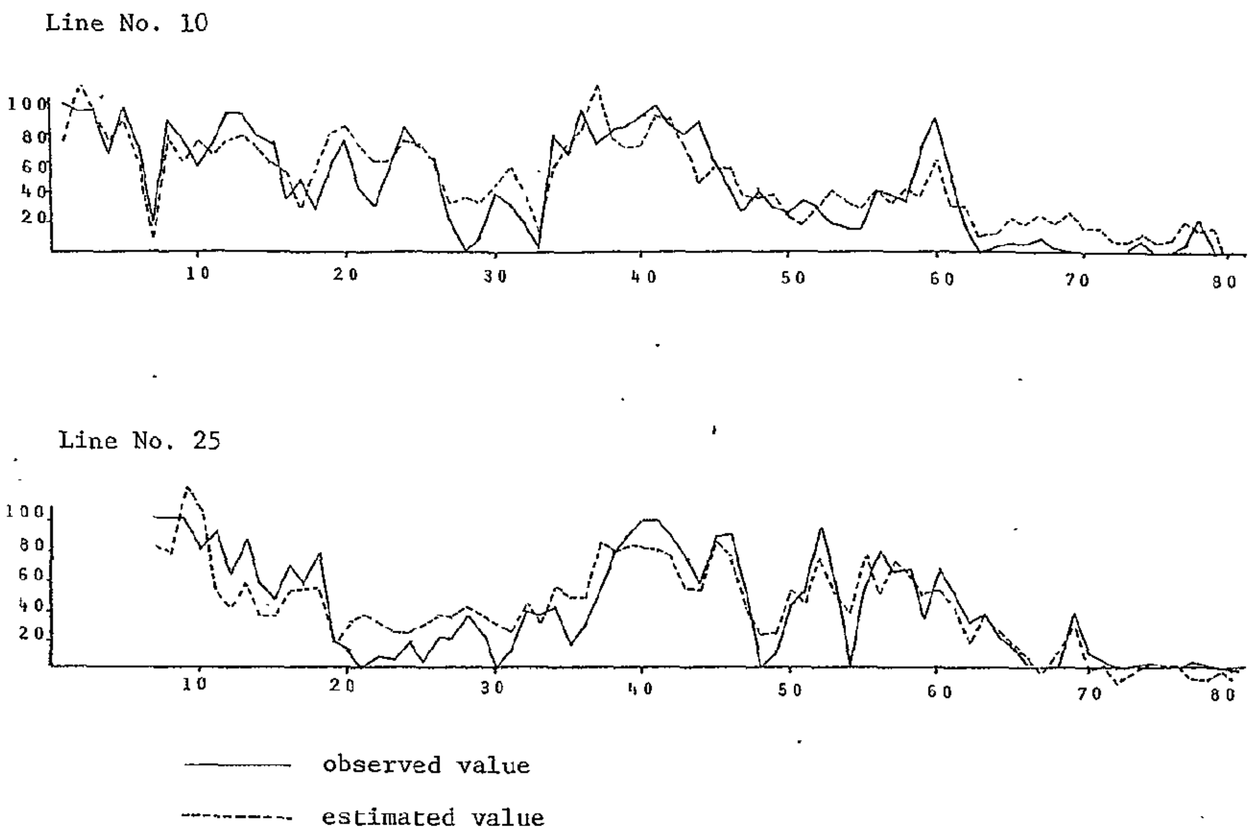

Fig. 2 Cross Sections of Vegetation Cover for Comparison between Observed Value and Estimated Value

OBSERVATION OF VEGETATION COVERAGE G 3 N1542511A

OBSERVATION OF VEGETATION COVERAGE G 3 N1542507A

FIG. 3 A OBSERVED VALUE OF VEGETATION COVER

. 0- 20 % ; 20- 40 % + 40- 60 % s 60- 80 % # 80- 100 %

FIG. 3 B VEGETATION COVER ESTIMATED BY MULTI-REGRESSION ANALYSIS

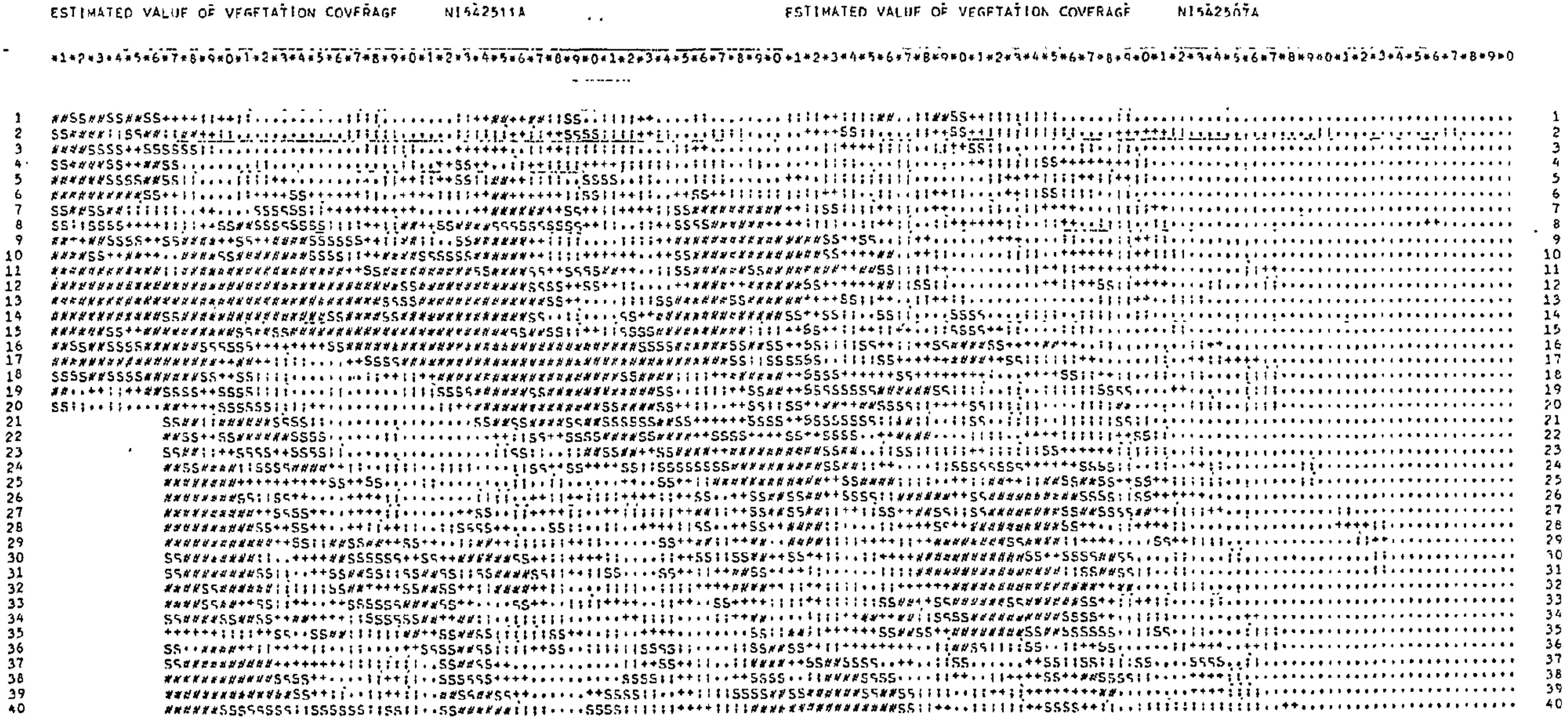

FIG. 3 C VEGETATION COVER ESTIMATED BY PARAMETRIC MODEL

FIG.3 B VEGETATION COVER ESTIMATED BY MULTI-REGRESSION ANALYSIS--COLOR PRINTED