

## リモートセンシングデータのデジタル処理 (第3報)

— 地球資源衛星データを用いた首都圏の緑地比率分析 —

Digital Processing of Remote Sensing Data (III)

— An Analysis on Vegetation Coverage by using LANDSAT MSS Data —

村井俊治\*・建石隆太郎\*・河崎和明\*

Shunji MURAI, Ryutaro TATEISHI and Kazuaki KAWASAKI

## 1. はじめに

都市および都市近郊の緑地環境は、都市の健全な姿を監視するための一つの指標となっている。昭和48年には建設省が「首都50Km圏緑地保存整備計画調査」を行ない、首都圏の緑地現況図および緑地環境に関する500メートルメッシュの数直データを作成した。この調査においては航空写真が基礎資料として用いられ、多大の費用と作業が必要とされた。

本研究では、地球資源衛星(LANDSAT)のMSSデジタルデータのスペクトル特性と広域同時性に着目して、LANDSATデータから首都圏の緑地比率を求めるための数学的モデルを作成しようとした。本研究の成果によれば、緑地比率は、LANDSATデータから実用的に十分な精度で求めることが可能であることがわかった。

## 1. 緑地比率を推定するためのモデル

1.1 テスト地域 図1に示すように5万分の1の地図の「八王子」および「東京西南部」をあわせた地域で東西約45Km、南北約18Km、面積で約800Km<sup>2</sup>の地域をテスト地域とした。この地域の西部は高尾および津久井の山岳部を含み、東部は品川および川崎の市街地を含み、中間部には丘陵や田畑あるいは都市近郊などを含んでおり、首都圏の緑地環境を検討するのに適当な内容と規模をもっている。

1.2 緑地比率の観測データ まえにも述べたように昭和48年の建設省による「首都50Km圏緑地保存整備計画調査」では、表1に示されるような16項目の緑地現況区分のデータとして計測された。ここでいう500メートルメッシュとは、5万分の1の地図(経度間隔15分および緯度間隔10分)を縦横ともに40等分して得られる長方形のメッシュである。「八王子」および「東京西南部」の地域をあわせると、3,200個の500メートルメッシュがあるが、一部に120個の欠落データがあるので、3,080個の緑地比率データが用いられた。

建設省が定義した緑地比率は、表1に示されるように16項目のうちから「市街地」、「自然破壊地」、「並木」、「その他」が除かれたものの占める面積率をいう。しかし、この中には、「水部」が含まれており、LANDSATデータのスペクトル特性から、クロロフィルムを含まない。「水部」を緑地として扱うのは適当でない。そこで本研究では、緑地比率データとして、表1に示すようにつぎの三つのものが用いられた。

(1) 第1補正緑地比率(記号 $G_1$ ): 建設省の定義した緑地比率から「水部」のデータを除いて求められた緑地比率

(2) 第2補正緑地比率(記号 $G_2$ ): 緑被率30%以上の「市街地」には、少なくとも30%の緑地があると仮定して、 $G_1$ にこの分をつけ加えた緑地比率



図1 テスト地区(5万分の1地図「八王子」と「東京西南部」, 左下の斜線部

は緑地比率データの観測値の欠落しているところを示す)

\* 東京大学生産技術研究所 第5部

表1 緑地現況区分と緑地比率の定義(係数を乗じた総合の面積率を緑地比率とする)

| 緑地現況区分       | 建設省<br>緑地比率 | 本 研 究          |                |                |
|--------------|-------------|----------------|----------------|----------------|
|              |             | 第1補正緑地比率 $G_1$ | 第2補正緑地比率 $G_2$ | 第3補正緑地比率 $G_3$ |
| (1)針葉樹林(人工)  | 1           | 1              | 1              | 1              |
| (2)針葉樹林(天然)  | 1           | 1              | 1              | 1              |
| (3)広葉樹林      | 1           | 1              | 1              | 1              |
| (4)針・広混交林    | 1           | 1              | 1              | 1              |
| (5)竹林        | 1           | 1              | 1              | 1              |
| (6)公園緑地      | 1           | 1              | 1              | 1              |
| (7)水田        | 1           | 1              | 1              | 1              |
| (8)畑         | 1           | 1              | 1              | 1              |
| (9)園芸樹畑      | 1           | 1              | 1              | 1              |
| (10)緑の多い市街地  | 0           | 0              | 0.3            | 0.3            |
| (11)緑の少ない市街地 | 0           | 0              | 0              | 0              |
| (12)並木       | 0           | 0              | 0              | 0              |
| (13)水部       | 1           | 0              | 0              | 0              |
| (14)荒地       | 1           | 1              | 1              | 0.5            |
| (15)自然破壊地    | 0           | 0              | 0              | 0              |
| (16)その他      | 0           | 0              | 0              | 0              |

注 水部を含む 水部除去 緑の多い市街地の補正 荒地の補正

(3) 第3補正緑地比率(記号 $G_3$ ):「荒地」の中には緑地がないものもあるので、荒地の平均線被率が50%と仮定して $G_2$ をさらに補正して求められた緑地比率

1.3 数学的モデル 緑地比率を求めるための数学的モデルには、つぎの二つのモデルが用いられた。

(1) 重回帰モデル

LANDSAT MSSの4バンドからなるLANDSATデータ $X(x_1, x_2, x_3, x_4)$ を、緑地比率データの500メートルメッシュに対応させておく。また、緑地比率データを $y$ とすると、つぎの線形結合による重回帰式のモデルを求めることとする。

$$y = f(\bar{X}) = a_1 \bar{x}_1 + a_2 \bar{x}_2 + a_3 \bar{x}_3 + a_4 \bar{x}_4 + a_0$$

(2) パラメトリックモデル

LANDSATデータから導かれた変数 $X$ が500メートルメッシュ内に $m$ 個あるとき、変数 $X$ がパラメータ $k$ より大きい値を示す $n$ 個が緑地であると仮定し、緑地比率 $n/m$ を求める。このときパラメータ $k$ の値を変えてみて、緑地比率の観測データ $G$ と、このモデルから求まる緑地比率の推定値 $n/m$ との残差の標準偏差が最小になるような $k$ の値を求め、このときの $k$ の値が与える緑地比率を最良推定値とする。本研究では、このようなモデルをパラメトリックモデルとよぶことにする。

LANDSATデータから導かれる変数 $X$ は、緑地の検出に有効であるとされているつぎの八つの変数がそれぞれ用いられた。

- (1)  $X_1 = x_3/x_1$
- (2)  $X_2 = x_4/x_1$
- (3)  $X_3 = x_3/x_2$
- (4)  $X_4 = x_4/x_2$

$$(5) X_5 = (x_3 + x_4) / (x_1 + x_2)$$

$$(6) X_6 = (x_1 + x_2 - x_3 - x_4) / (x_1 + x_2 + x_3 + x_4)$$

$$(7) X_7 = a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_0$$

$$a_1^* x_1^* + a_2^* x_2^* + a_3^* x_3^* + a_4^* x_4^*$$

(重回帰モデル)

$$(8) X_8 = d_1 x_1^* + d_2 x_2^* + d_3 x_3^* + d_4 x_4^* \quad (\text{第2主成分})$$

ここで、 $x_1$  = LANDSAT MSS 4バンドのデータ(緑バンド)

$x_2$  = LANDSAT MSS 5バンドのデータ(赤バンド)

$x_3$  = LANDSAT MSS 6バンドのデータ(近赤外バンド)

$x_4$  = LANDSAT MSS 7バンドのデータ(赤外バンド)

$x_i^*$  = 平均値 $\bar{x}_i$ 、標準偏差 $\sigma_i$ のまわりに標準化されたデータ:  $x_i^* = (\bar{x}_i - x_i) / \sigma_i$

上の(1)から(6)までの変数は、赤外領域と可視領域のバンド比または差と和の比であり、緑地と相関があるといわれている。(7)は、重回帰モデルにより求められた重回式である。(8)は、主成分分析をしたとき得られる第2主成分のスコアであり、緑地と相関が強いと想定される。

使用されたLANDSATデータは、第2報までのデータと同じ日時のもので、昭和47年11月26日のLANDSAT-1号のマルチスペクトルスカナ(MSS)デジタルデータである。

2. 重回帰分析による緑地比率の推定

2.1 テストケース 表1に示された三種類の緑地比率の定義に対して重回帰分析が行われた。ケース1は、被説明変数として表1に示した $G_1$ を用い、ケース2は $G_2$ 、ケース3は $G_3$ を用いた場合の重回帰分析に対応している。

2.2 重回帰分析 表2は、重回帰分析により求められた係数と重相関係数(観測値と推定値の相関係数)および残差の標準偏差(観測値と推定値の残差の標準偏差)を示す。下段の数字は標準化されたLANDSATデータに対する標準化された係数を示す。重相関係数が最も良かったのはケース3で、 $r = 0.871$ であった。また残差の標準偏差が最も小さかったのはケース3の場合、緑地比率の単位(%)で±15.53%であった。いずれのケースにおいても重相関係数は0.84以上で、緑地比率の推定が可能であることがわかった。

表2 重回帰分析の結果

| 分析結果<br>ケース | 回帰式の係数(下段は標準化係数) |                  |                |                |        | 精 度   |       |
|-------------|------------------|------------------|----------------|----------------|--------|-------|-------|
|             | $a_1$            | $a_2$            | $a_3$          | $a_4$          | $a_0$  | 重相関係数 | 標準偏差  |
| ケー1         | -3.844<br>-0.242 | -4.839<br>-0.378 | 3.697<br>0.331 | 6.002<br>0.377 | 81.368 | 0.845 | 17.82 |
| ケー2         | -3.722<br>-0.237 | -5.140<br>-0.406 | 3.874<br>0.351 | 5.781<br>0.367 | 85.638 | 0.856 | 16.92 |
| ケー3         | -3.340<br>-0.220 | -5.229<br>-0.428 | 3.241<br>0.304 | 6.338<br>0.417 | 83.417 | 0.871 | 15.53 |

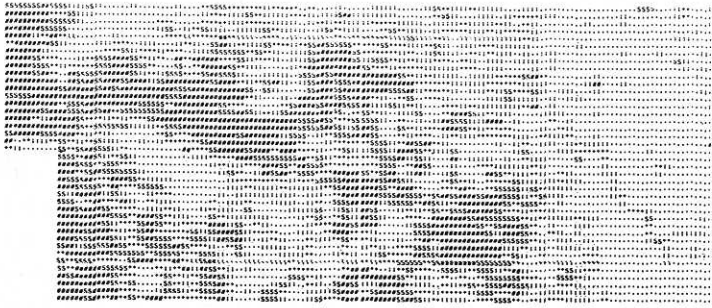


図2(a) 緑地比率観測値 (・0~20%, ; 21~40%,  
++41~60%, SS61~80%, ##81~100%)

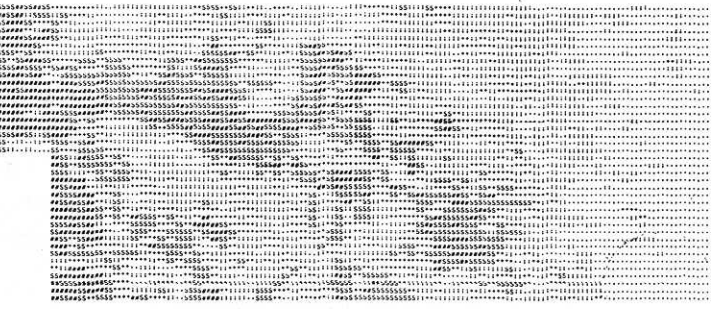


図2(b) 重回帰分析による緑地比率の推定

2.3 精度の検討

図2は、ケース3に対する緑地比率の推定値と観測値を20%毎の5段階にわけて示したものである。

重回帰分析の全体の精度は、重相関数で0.85程度以上あることがわかったが、残差の大きい箇所をしらべてみると、つぎのことが指適される。

(1) 荒地・公園・田畑・山岳部では、推定値が観測値より小さい値を示す。これは、荒地・公園・田畑においては、緑がない部分があるにもかかわらず、緑地比率の定義上緑地に算入しているためである。山岳部においては、山の影がスペクトルデータに影響を与えているために、緑として認識されなかったためである。

(2) 市街地および自然破壊地においては、推定値が観測値より大きい値を示す。これは、市街地および自然破壊地には若干の緑があるにもかかわらず、緑地比率の定義上緑地に算入されなかったためである。

(3) 以上のべたように、LANDSAT データから推定された緑地比率の誤差の原因は説明がつくものであり、その原因はいずれも被説明変数として用いられた緑地比率の観測値の定義に起因している。

(4) したがって、LANDSAT データから求められる緑地比率の方がむしろ緑地の現況に近い緑地比率を再現し

ているといえる。

3. パラメトリックモデルによる緑地比率の推定

3.1 パラメトリックモデルのテスト  
 ケース 被説明変数として用いられた緑地比率データは、 $G_1, G_2, G_3$  の3種類とした。これはそれぞれ、重回帰分析の3つのテストと同じである。

3.2 最良パラメータの算定  
 まえに示した8種類の変数Xに対して、つぎの条件を満足する最良パラメータkを求めた。

m個からなる500メートルメッシュのLANDSAT データのうち、 $X \geq k$  (XとGが正の相関のとき)または $X \leq k$  (XとGが負の相関のとき)を満足するデータがn個あったとき、 $n/m$ を緑地比率とし、観測値Gと推定値 $n/m$ との残差の標準偏差が最小になるようなkを最良パラメータとする。

表3は、八つの変数に対する最良パラメータkの値とそのときの残差の標準偏差を示している。

八つの変数のうち重回帰分析による変数 $X_7$ は、つぎの式が用いられた。

$$X_7 = -3.14x_1 - 5.73x_2 + 4.34x_3 + 5.76x_4 + 71.42 \\ = -0.196x_1^* - 0.453x_2^* + 0.384x_3^* + 0.352x_4^*$$

また、主成分分析によって求められた第2主成分による変数 $X_8$ は、つぎの式が用いられた。

$$X_8 = -0.511x_1^* - 0.414x_2^* + 0.367x_3^* + 0.658x_4^*$$

表3の結果から緑地比率の推定を行うのに最も良い精度を示す変数は $X_3 = x_3/x_2 = \text{MSS 6バンド (近赤外バンド) / MSS5バンド (赤バンド)}$ であり、最良パラメータkの値およびそのときの残差の標準偏差 $\sigma$ は、それぞれケース1で $k=1.16, \sigma=17.27$ 、ケース2で $k=1.14, \sigma=16.74$ 、ケース3で $k=1.14, \sigma=15.36$ であった。

図3は、パラメトリックモデルによって求められたケース3に対する緑地比率図を20%毎の5段階に分けて出力したものである。

3.3 重回帰モデルとパラメトリックモデルの比較

つぎに重回帰モデルとパラメトリックモデルの特徴やその相違を考察してみよう。

(1) 重回帰モデルにおいては、500メートルメッシュのデータとして平均値を用いて緑地比率との重回帰分析

表3 パラメトリックモデルによる緑地比率の推定結果

| 変 量   | パラメータと標準偏差 |          | ケース1 |          | ケース2 |          | ケース3 |          |
|---|------------|----------|------|----------|------|----------|------|----------|
|   | $k$        | $\sigma$ | $k$  | $\sigma$ | $k$  | $\sigma$ | $k$  | $\sigma$ |
| $X_1 = x_3 / x_1$   | 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    |      |          |
| $X_7 = a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_0$       | 49         | 17.58    | 47   | 17.36    | 48   | 16.08    |      |          |
| $X_8 = d_1 x_1^* + d_2 x_2^* + d_3 x_3^* + d_4 x_4^*$     | -0.1       | 17.85    | -0.2 | 17.69    | -0.1 | 16.58    |      |          |

注  $k$ : 最良パラメータ,  $\sigma$ : 残差の標準偏差,  $X_7$ : 重回帰式,  $X_8$ : 第2主成分

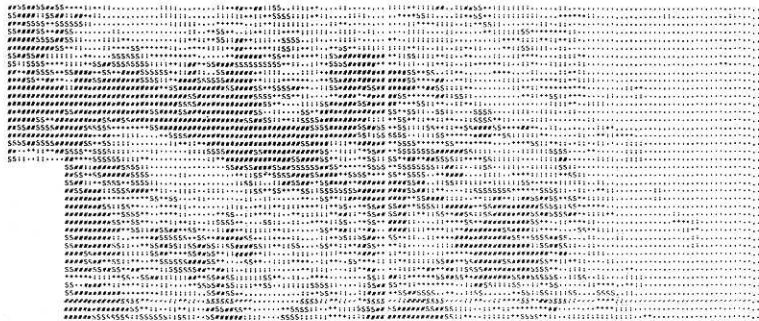


図3 パラメトリック法による緑地比率の推定 (・: 0 ~ 20%, ;: 21 ~ 40%, +: 41 ~ 60%, SS61 ~ 80%, ## 81 ~ 100%)

を行ったのに対し、パラメトリックモデルでは、500メートルメッシュ内のLANDATデータの1点1点ごとに緑地かを判別した累計が、観測値に最も近似するような変量とパラメータを求めようとした。したがって、重回帰モデルは、平均値を用いた統計学的考え方にもとづいており、パラメトリックモデルは、1点1点のスペクトル特性の累積を考慮することを可能にしている。

(2) 重回帰分析よりパラメトリックモデルによる緑地比率の推定がわずかに良い。しかし両者の間にそれ程大きな差はない。また残差の大きいところの原因は、両者ともに同じ傾向であった。これはまえにものべたように、緑地比率の定義からくる誤差が原因となっている。

(3) パラメトリックモデルにおいて求められた最良の変量  $X_3 = x_3 / x_2$  とパラメータ  $k = 1.14$  は、第1報 (生産研究 Vol 29, No 2) でのべたパターン分類における植物パターンの分類基準とはほぼ合致することがわかった。すなわち、第1報でのべたパターン分類では、LANDSAT デジタルデータを放射輝度レベルにしたとき得られるスペクトル曲線のパターンのうち、植物パターンは、MSS 5バンド ( $x_2$  赤バンド) とMSS 6バンド ( $x_3$  近赤外バンド) の2つの放射輝度レベル ( $2.00 \times x_2$  と  $1.76 \times x_3$ ) が正の傾きをもつことが分類基準となっていた。この基準は、 $1.76x_3 / 2.00x_2 \geq 1.0$ 、 $x_3 / x_2 \geq 2.00 / 1.76$

= 1.14となり、パラメトリックモデルの  $k = 1.14$  に相当する。

この結果、緑地か否かを判別するのに、放射輝度レベルでのスペクトル曲線パターンを用いて良いことが確認された。

(4) 重回帰モデルは、簡便に緑地比率を推定するのに便利であるが、その質的内容を論じる柔軟性はパラメトリックモデルに比べて少ない。パラメトリックモデルは、作業量は多いが将来他の環境指標を求めるのに拡張して応用しうる。たとえば、水部や都市構造物などの環境要素を判別するのに応用できる。

(5) いずれの方法も、第2報 (Vol 29, No 4) でのべた“混合問題”の解決に有効である。重回帰モデルが平均値を用いた統計学的解決方法であるのに対し、パラメトリックモデルは、メッシュ内の面積率を用いた解決法である。

ま と め

(1) LANDSAT データから首都圏の緑地比率を重回帰係数で約0.85、標準偏差で±15%の精度で推定することができる。

(2) 緑地比率の推定には、重回帰モデルおよびパラメトリックモデルのいずれも利用可能である。

(3) パラメトリックモデルは、応用面において柔軟性がある。重回帰モデルは簡便である。

(4) LANDSAT データから求められた緑地比率は、緑地の現況をきわめて良く表現している。

(5) 緑地比率の定義を再考する必要がある。リモートセンシングデータの有効利用をはかる上では、スペクトル特性の上から緑地を定義することを提言したい。すなわち、本研究によれば、放射輝度レベルで赤バンド (MSS 5バンド) と近赤外バンド (MSS 6バンド) の傾きが正となるパターンを有する点を緑地と認めてよい。これは緑地の有するクロフィルの赤バンドにおける吸収と近赤外バンドにおける反射の特性からも十分合理性を有していると考えられる。

(1977年3月24日受理)