



Title	ランドサット MSS データによる作物識別の精度について
Author(s)	堀口, 郁夫; 林田, 朋之; 谷, 宏
Citation	北海道大学農学部邦文紀要, 14(4), 330-336
Issue Date	1985-12-28
Doc URL	http://hdl.handle.net/2115/12031
Type	bulletin (article)
File Information	14(4)_p330-336.pdf



[Instructions for use](#)

ランドサット MSS データによる作物識別の精度について

堀口 郁夫・林田 朋之・谷 宏

(北海道大学農学部農業物理教室)

(昭和60年1月8日受理)

Accuracy of Classification for Crops Using Landsat MSS Data

Ikuko Horiguchi, Tomoyuki Hayashida
and Hiroshi Tani

(Laboratory of Agricultural Physics, Faculty of Agriculture,
Hokkaido University, Sapporo, Japan)

I. 序 論

人工衛星によるリモートセンシングの特徴は、地上における観測のように二次元(点や線)の観測でなく、広い面積の二次元の観測であり、三次元のデータも得ることが出来ることである。

1972年に地球資源探査衛星(現在のランドサット(Landsat)がうち上げられて約10年過ぎ、衛星からのデータを利用した地球上の資源状況の解析も、飛躍的に発展した。農業部門への人工衛星データの利用は、すでにアメリカにおいて実用化の段階まで進んでいる。すなわち、栽培作物の識別、耕地の面積計算、収量予想、病虫害の発生状況の把握、土壌水分・作物水分状況によるかんがい時期の決定など、広く利用され始めている。

一方、わが国においては、1979年からランドサットデータの直接受信を行ない、データの提供を開始しているが、農業方面の利用は一般にまだ普及していない。その理由は、

- ① 耕地が狭く、地上における情報が密で人工衛星データに情報を依存する必要が少ない。
- ② 耕地の一区画当たりの面積が小さく、人工衛星から十分な精度の情報が得られない。
- ③ アジアモンスーン地帯に属し、雲が多いためデータの取得出来る回数がきわめて少ない。
- ④ 解析ソフトの技術者が少なく、この方面の研究が遅れている。

などが考えられる。しかし、北海道は比較的耕地面積が広く、地上の情報も粗であるなど、ランドサットデータの利用に比較的適している地域である。また、現在デー

タを送信しているランドサット5号は、高性能のTM(Thematic Mapper)を搭載しており、地上における分解能も向上している。したがって、今後農業方面の利用も増大するものと思われる。

この研究は、ランドサットのMSS(Multispectral scanner)データが、どの程度の精度で作物の識別が出来るかを調べたものである。

II. 解析方法

解析に使用したデータは、ランドサット3号の1982年7月8日のもので、パス(path)116、ロウ(row)30、シーン(scene)札幌のものである。

解析手順は、

- ① 地上グラントルース(ground truth)データの作成
 - ② 栗沢トレーニングエリア(training area)における解析
 - ③ 南幌テストエリア(test area)における解析と精度の比較
 - ④ 石狩平野への適用
- の4段階からなっている。

II-1 地上グラントルースデータの作成

地上のグラントルースデータの取得場所として、Fig. 1に示す北海道南幌町・栗沢町・江別市の一部を選んだ。この地区の水田および畑の一区画ごとの作付状況を、市町村役場のデータで調べ、不明の場所は現地調査した。このデータをランドサットデータと同じ大きさのます目にメッシュ化して、グラントルースデータとし

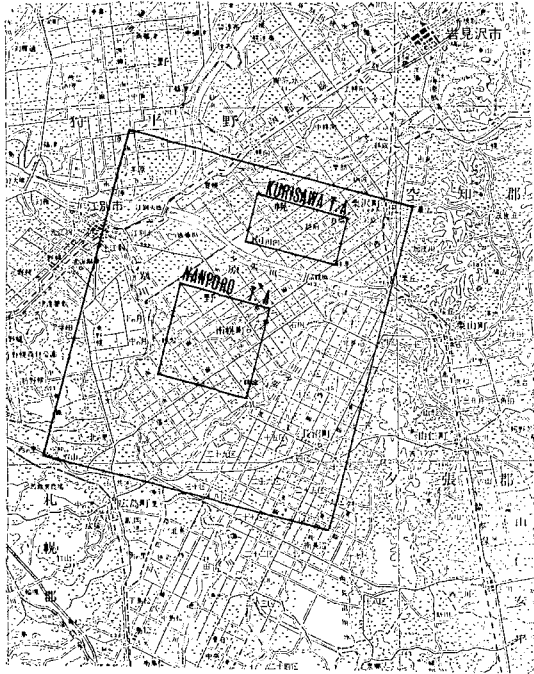


Fig. 1. Area of ground truth measurement.

た。この場合、一メッシュが同一作物で占められるものだけをその作物のメッシュに分類し、2つ以上の作物が入る場合は、その他の作物として分類して、それぞれ符号化してデータを作成した。

解析は Fig. 1 に示す地域の内、夕張川以北の地区（以下、栗沢トレーニングエリアと称する）について行ない、この解析結果を、夕張川以南の地区（以下、南幌テストエリアと称する）にあてはめ、作物判別の精度の検討を行なった。なお、栗沢トレーニングエリアは、50 ライン×90 ピクセル、南幌テストエリアは100 ライン×100 ピクセルである。また、(4) の石狩平野への適用には、グラントルース全調査地区をトレーニングエリア（300 ライン×300 ピクセル）として解析を行ない、その結果を石狩平野に適用した。

II・2 栗沢トレーニングエリアにおける解析

今回研究に使用したランドサットの全データは4バンド750,000 ピクセル (Pixel) で膨大な量である。一方、ランドサット MSS データは、バンド4とバンド5およびバンド6とバンド7が相関が強いことが知られているため、全データの4バンド(4変量)を主成分分析で2変量に圧縮した。この2変量について、南幌トレーニングエリアの作物識別を以下に示す判別分析と最尤法で解析

を試みた。

i) 判別分析

主成分分析で2変量に圧縮されたデータ $(x_{1:n}, x_{2:n}; n=1, \dots, N$ でデータ数)を G_1 作物と G_2 作物に分類する場合、データはそれぞれ G_1 と G_2 の2の群をなして存在すると考えられる。そして、2つの群の境界付近では、 $G_1 \cdot G_2$ は重複して存在する。この重複する部分をどちらかの作物に判別するのであるが、誤まって判別される確率を出るだけ小さくするような関数を計算し、この関数(判別関数)によって作物の種類を識別する方法を判別分析という。

判別関数は(1)式で定義される。

$$z = x'S^{-1}(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 + \mu_2)'S^{-1}(\mu_1 - \mu_2) \dots \dots \dots (1)$$

ただし、 μ_1, μ_2 は各変量 $x_{1:n}, x_{2:n}$ の平均値を要素とするベクトル、 S は分散共分散行列

ここで

- $z \geq \log|K|$ ならば x は G_1 に属する
- $z < \log|K|$ ならば x は G_2 に属する

K は G_1 と G_2 の先験確率 $P(1), P(2)$ と誤判断によって生じる損失 $C(1|2), C(2|1)$ (未知の時 $C(1|2)=1, C(2|1)=0$)とすると(2)式より計算出来る。

$$K = C(1|2) \cdot P(2) / C(2|1) \cdot P(1) \dots \dots \dots (2)$$

ii) 最尤法

データを作物の種類別に分類する場合、未知のデータを(3)式で表わされる距離によって分類する方法である。

$$L = \sum_i^2 \sum_j^2 (x_i - \mu_i)(x_j - \mu_j) |S_{ij}| + \log|S_{ij}| - \log|q| \dots \dots \dots (3)$$

ただし、 L : 距離(小さいほどカテゴリーへの尤度が高い)

- x_i, x_j : 未知点のデータ値
- μ_i, μ_j : 各作物ごとの平均値
- S_{ij} : 各作物ごとの分散共分散行列
- q : 各作物ごとの先験確率

今回の解析として、水田、秋小麦、春小麦、エン麦、牧草、小豆、ビート、その他の8項目をえらんだ。

II・3 南幌テストエリアにおける解析と精度の比較

南幌テストエリアについては、II・2に示した判別分析と最尤法の解析結果を適用して作物の識別を行なうとともに、グラントルースがない場合の計算法であるクラ

スター分析も行なった。さらに、判別分析・最尤法・クラスタ分析による作物識別の精度を調べた。

i) 南幌テストエリアのクラスタ分析

クラスタ分析は、データを構成している測定値を何らかの基準によって分類する方法である。今回はデータ数が膨大のため非階層的手段である K-mean 法によって行なった。解析手法は次に示すとおりである。

(1) 2次元空間(2変量)において、 q 個のクラスターの中心(seed-point)をある初期条件によって設定する。また、クラスター間の距離の最小値 T ($T=35.0$ とした)を決めておく。

(2) 観測値を最も近いクラスターに分類し、それぞれ重心を計算する。

(3) クラスタ間での重心の距離が最小値 T より小さければ、その2つのクラスターを融合して、重心を計算しなおす。

(4) このようにしてクラスターを融合し、すべてのクラスター間の重心が距離の最小値 T 以上になった時、最終的な重心をそのクラスターの中心として固定する。

(5) すべてのデータは、再び最も近い中心に割りあてられ分類が行なわれる。

ii) 解析精度の計算

判別分析・最尤法・クラスタ分析の解析結果の精度 k は、(4)式で行なった。すなわち、Fig. 2 に示す A 面積を地上調査ではある G_1 作物と分類されたが、計算では G_1 作物と識別されなかった部分、 C 面積を地上調査でも計算でも G_1 作物と識別された部分、 B 面積を計算でのみ G_1 作物と識別された部分で、精度は、 C 面積の全

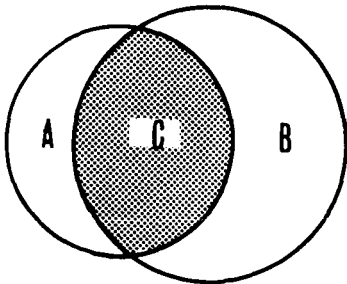


Fig. 2. Accuracy defined by ground truth and calculation.

- A: The area discriminated as G crop by ground truth only.
- B: The area discriminated as G crop by calculation only.
- C: The area discriminated as G crop by ground truth and calculation.

体の面積に対する百分率で表わした。

$$k = \frac{C}{A+B+C} \times 100 \dots\dots\dots (4)$$

II・4 石狩平野への適用

後に示す III・3 の結果、最尤法による識別が最も精度がよかったため、地上調査したグラントルース全データ(300ライン×300ピクセル)について最尤法の計算を行なった。さらに、この結果を石狩平野に適用し、土地被覆状況の分類を試みた。分類項目は、水田、森林、畑地、市街地、裸地、河川などの水および海の7つである。

III. 解析結果及び考察

III・1 地上におけるグラントルースデータの作成

地上調査したデータをランドサットデータと同じようにメッシュ化して表わしたのが Photo 1 である。また、栗沢トレーニングエリアと南幌テストエリアの各作物のピクセル数は、Table 1 に示すとおりである。両エリアとも水田と小麦畑が大部分を占め他の作物畑はわずかである。

Table 1. Number of classified pixels for ground truth data at Kurisawa Training Area and Nanporo Test Area

Field	Number of pixels at Kurisawa T.A. (km ²)	Number of pixels at Nanporo T.A. (km ²)
Paddy	1417 (603.08)	3981 (1694.31)
W. Wheat	1499 (637.97)	1313 (558.81)
S. Wheat	285 (121.30)	1209 (514.55)
Grass	62 (26.39)	138 (58.73)
Beet	57 (24.26)	113 (48.09)
Red Bean	78 (33.20)	95 (40.43)
Oats	226 (96.19)	90 (38.30)
Others	876 (372.83)	3061 (1302.76)
Total	4500 (1915.20)	10000 (4256.00)

III・2 栗沢トレーニングエリアにおける解析

栗沢トレーニングエリアの作物の識別を行なう前に全データについて主成分分析を行ない、4バンドを2変量に圧縮した。第1主成分 z_1 と第2主成分 z_2 は次式で表わされる。

$$z_1 = -0.013 B_4 + 0.013 B_5 + 0.618 B_6 + 0.786 B_7 \dots\dots\dots (5)$$

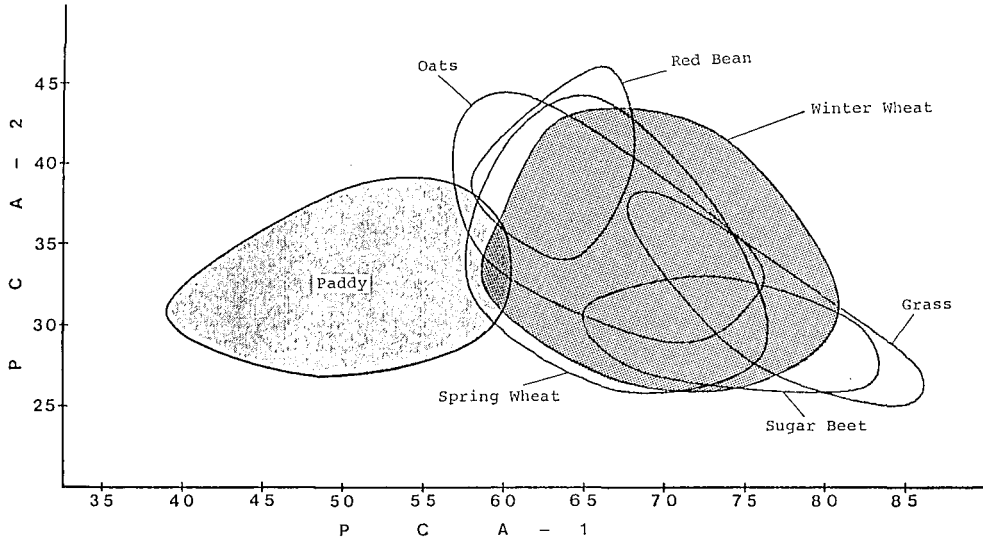


Fig. 3. The distribution range of first principal component and second principal component for each crop.

PCA-1: First principal component.
 PCA-2: Second principal component.

$$z_2 = 0.372 B_4 + 0.893 B_5 + 0.195 B_6 - 0.162 B_7 \dots (6)$$

なお、第2主成分までの累積寄与率は94.9%である。

i) 判別分析

x軸に第1主成分の値を、y軸に第2主成分の値をプロットした作物ごとの範囲をFig. 3に示す。Fig. 3によると畑作物は互いに重複した位置に分布して作物間の識別は困難であるが、水田と畑地の識別は出来ると思われる。水田とピクセル数の多い冬小麦畑との判別式は(7)式で表わされた。この式によって識別された水田・冬小麦畑と、地上調査によって得られたグラントールとを比較した精度は、水田で93.7%であり、冬小麦畑は93.9%であった。

$$z_2 = 66.8 z_1 - 366.6 \dots (7)$$

ii) 最尤法

最尤法では、8種類の作物について識別を行なった。その結果をTable 2に示す。Table 2から明らかなくとく、面積が広い水田の識別精度は、82.3%であるが、他の作物については精度が悪く、ほとんど識別不可能である。これは、Fig. 3で示したごとく、畑作物の2変量の分布がほとんど同じ位置に分布しているからである。また、判別関数による識別精度と比較すると、やや悪くなっている。これは、判別分析は、2作物間の判別であ

Table 2. Analysis of field classification by the Most Likelihood Method for Kurisawa Training Area

Field	No. of pixels for ground truth	No. of calculated pixels	Percentage of coincidence (%)
Paddy	1417	1575	82.3
W. Wheat	1499	2191	52.6
S. Wheat	285	0	0.0
Grass	62	116	17.1
Beet	57	6	0.0
Red Bean	78	108	6.9
Oats	226	225	13.3
Others	876	279	8.1
Total	4500	4500	—

ったが、最尤法は8種類の作物について同時に行なったためと思われる。

III・3 南幌テストエリアにおける解析と精度の比較

栗沢トレーニングエリアにおける判別関数・最尤法の解析結果を南幌テストエリアに適用した結果と、南幌テストエリアのクラスター分析結果をTable 3に示す。栗沢トレーニングエリアの解析において畑の作物間の

Table 3. Comparison of accuracy between three classification methods for paddy field at Nanporo Test Area

Method	No. of pixels for ground truth	No. of calculated pixels	Percentage of coincidence (%)
Most Likelihood Method	3981	4229	85.2
Discriminant function method	3981	4675	83.2
Cluster Analysis	3981	3451	77.8

識別は困難であったため、水田のみの識別を行なった。Table 3 には、解析結果と(4)式で計算したグランドトルースとの精度も示してある。

3種類の解析の結果、最も精度よく解析出来たのは、最尤法である。しかし、判別分析との精度の差はあまり大きくない。クラスター分析は地上のグランドトルースがない場合に使用される解析法で3種の解析法のうち、精度が最も悪かった。

III・4 石狩平野への適用

識別精度が最もすぐれていた最尤法について石狩平野全域 (Photo 2) に適用した。トレーニングデータは、地上調査したグランドトルース全体 (300 ライン×300 ピクセル) を使用した。この識別結果を Table 4 に、また画像の出力結果を Photo 2 に示す。Photo 2 によると、札幌市の市街地の中に水田と分類されている部分がある。この原因は、グランドトルースに市街地のデータ

Table 4. Analysis of land cover for Ishikari Plain used by the Moist Likelihood Method

Land cover	No. of calculated pixels	Area (km ²)	Percentage (%)
Paddy	127,513	542.7	17.0
Forest	324,047	1,379.1	43.2
Field crops or grass	166,541	538.9	22.2
City	94,071	400.4	12.5
Bare soil	10,127	43.1	1.4
Water	8,353	35.5	1.1
Sea	19,349	82.3	2.6
Total	750,000	3,192.0	100.0

が少ないことによる誤差と思われる。

IV. 結 語

ランドサット MSS データ (1982年7月8日) を用いて作物の識別を試みた。その結果次のことが明らかになった。

i. 水田と畑地の識別は比較的高い精度で可能であったが、畑作物間の識別精度は悪く実際上識別は不可能である。

ii. 解析法では、判別分析、最尤法、クラスター分析のうち、最尤法が最も精度が良かった。

iii. 最尤法による水田の識別精度は、85.2%であった。

今回の解析は、1ピクセルが完全に同一の作物で占められている場合の解析で、1ピクセルに2作物以上が入った場合は除外されている。解析地域のピクセルの大きさは約56m×76mであるため、この2作物以上が含まれているピクセルの解析法を確立するとさらに精度が向上するものと思われる。しかし、今回の識別精度は、最良のもので85.2%の精度である。グランドトルースデータを多くとると精度が若干良くなるかもしれないが、この程度の精度では、まだ実用に供するのは困難であると思われる。また、畑作物間の識別には、時期の異なる2つ以上のデータによって可能になると思われるが、希望の時期のデータが、取得出来るかどうかが問題である。

現在、データを送信しているランドサット5号は、TMを搭載しておりこの地上の分解能は約30m×30mである。このデータを利用することによって、より解析精度が向上出来れば種々実用に供せるようになるかもしれない。今後は、TMデータを使用した作物の識別を行なう予定である。

なお、この研究の一部に、57年度北海道科学研究費(個人研究)を使用した。関係各位に謝意を表す。

V. 参 考 文 献

1. 河口至商：多変量解析入門 I, p. 35-51・79-91. 森北出版 1973
2. 河口至商：多変量解析入門 II, p. 26-44. 森北出版 1973
3. KENDALL, S. H., (奥野忠一, 大橋靖雄訳): 多変量解析, p. 15-53・165-195. 培風館 1981
4. 反野義興, 西原秀廉, 由井美雄: 最尤法識別分類及びキュービック・コンボリューション補間用ハードウェア内蔵したカラーイメージディスプレイ装置, リ

モートセンシング学会第2回学術講演会論文集, 63-66. 1982

Summary

Crops were classified by using Landsat MSS data and its accuracy was studied.

As the ground truth, Nanporo-town, Kurisawa-town and part of Ebetsu-city (Fig. 1) of Hokkaido were selected to study the classification of paddy fields as well as other planting fields, and mesh data of the ground truth was made. Among these areas, Kurisawa was designated as a training area (4500 pixels), crop classification was carried out by the Discriminant analysis (Eq. 1) and the Most Likelihood Method.

The result of the Discriminant Analysis shows that the Discriminant function between paddy field and winter wheat field is given by Eq. 7. The result of the Most Likelihood Method is shown in Table 2. These results were applied to Nanporo Test Area. Also the classification for

crops was carried out by using the Cluster Analysis. The result is shown in Table 3. In addition, the accuracy in Table 3 is a figure calculated by Eq. 4. Table 3 shows that the Most Likelihood Method proved to be the most accurate and the classification for paddy field was 85.2%. This method was applied to Ishikari-Plain, and the results are shown in Table 4 and Photo 2.

From the above studies, the following observations were made.

- 1) Paddy field and crops field classification was carried out with a comparatively high degree of accuracy, however, classification among crops was practically impossible to do because of low degree of accuracy.
- 2) Of the three methods, Discriminant Analysis, the Most Likelihood Method, and Cluster Analysis, the Most Likelihood Method was proved to be the most accurate.
- 3) The accuracy of classification for paddy field by using the Most Likelihood Method was 85.2%.

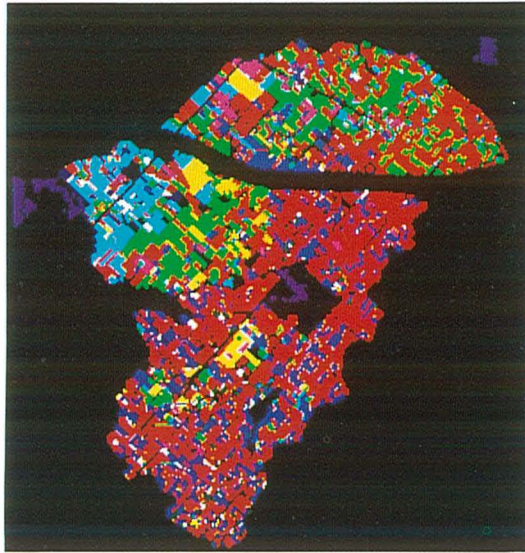


Photo 1. Mesh-diagram of ground truth.
(red: paddy field, green: winter wheat, blue: spring wheat, sky blue: grass
dark green: beet, pink: red bean, white: oats, black: others)

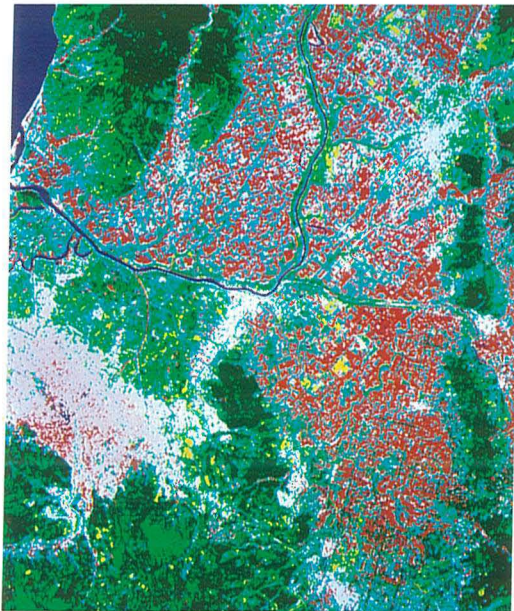


Photo 2. Analysis of land cover for Ishikari Plain used
by the Most Likelihood Method.
(red: paddy field, blue: field crops, green: forest, white: city)