

# パターン展開係数が保持する 情報量について

醍 醐 元 正

## I はじめに

最近の多波長衛星データの全ての波長域を利用した解析手法には幾つかの手法が存在する。principal components transformation<sup>[1]</sup>, canonical analysis<sup>[2]</sup>などの多変量解析法を利用したものや spectral mixing analysis<sup>[3]</sup>などが良く知られている。

筆者と共同研究している奈良女子大学の研究グループでは新しくパターン展開法<sup>[4]</sup>と名付けた多波長衛星データを解析する手法を研究開発して来た。パターン展開法は元々 LANDSAT/TM センサーで得られるデータのうち熱赤外域以外の六つの測定波長域を利用した解析手法として開発されたが、その後の研究によって広い範囲の波長について適用可能である事が解った<sup>[5]</sup>。

パターン展開法は地上被覆物の反射スペクトルを三つの基本的なスペクトルパターンによって展開する解析方法である。そしてこれら三つの基本スペクトルパターンは主要な土地被覆物である水・植生・土壌のスペクトルパターンを代表しているのである。前稿<sup>[5]</sup>の結論からわかる様に、広範囲の波長のデータについてほとんどの土地被覆物からの反射光は三つの基本スペクトルパターンの合成で精度良く近似出来る事がわかった。その時の解析では波長として 350 nm から 2500 nm までのうち大気の透過度が80%以上の部分についてバンド幅 10 nm の121バンドのデータを利用した。この様な広範囲の多バンドデー

タに対しても地上の被覆物からの反射光は水・植生・土壌のスペクトルパターンの合成で表現出来るのである。

多波長データを三つのスペクトルパターンの合成で近似する時に問題となるのはその精度である。前稿においても  $\chi^2$  分布を仮定してその近似の精度を見積もった。本稿においては更に情報量という観点から、パターン展開法で三つのスペクトルパターンに分解した場合に三つのパターン展開係数によって元のデータからそのまま保持される情報量について考察したので報告する。先ず次節では前稿で得た  $\chi^2$  分布の仮定からの結論をもう少し詳しく見る。次に良く知られた多変量解析法から導かれる情報量を使つての考察を行ったのでそれについて述べることにする。

## II $\chi^2$ 分布を仮定しての近似誤差

前稿においてデータのバンド数の変化に対してパターン展開法によるフィットでの誤差の程度を見積もった。それによるとバンド数15以上においては自由度当たりの  $\chi^2$  は 0.002 でそれ程変化しないことが判った。以下ではその後に調べたバンド数15以下での  $\chi^2$  の振舞について述べる。

まずこの解析の前提条件について少しまとめておく。パターン展開法では各画素のスペクトルパターンを三つの基本スペクトルパターン（水・植生・土壌）の合成によって近似する。各画素のスペクトルが  $n$  バンドのデータとして表されるとき、各バンドのフィットの余り  $R_i$  の二乗和

$$X = \sum_{i=1}^n R_i^2$$

を最小にするようにパターン展開係数を定める。

ここでバンド数が変化する場合にフィットの誤差の程度を見積もる指標として自由度当たりの余りの二乗和

$$\chi_{n-3}^2 = \sum_{i=1}^n R_i^2 / n - 3$$

がよい事が前回の議論の結論であった。

これは結局衛星データの各バンドの値が統計的に独立であり、同程度の分散の値を持っているという仮定がかなりよく当てはまると言うことを意味する。そして上記の  $X$  は自由度  $n-3$  の  $\chi^2$  分布に従う事になる。

今回の議論においても同様の仮定が正しいとして上記の  $\chi_{n-3}^2$  を誤差の指標として用いる。

この解析に使用したデータは前稿と同じ分光放射計のデータ595サンプルであり、その他の条件も前回と全く同じである。すなわち、データはバンド幅を 10 nm として平均化し、また大気の透過率が80%以上の領域を使用することでバンド数は121となった。

前稿と同様にバンド数を変化させた時に、 $\chi_{n-3}^2$  がどうなるかを見る為に、一つおき、三つおき七つおきなどにチャンネルを取りだしたデータを作って、それらに対してパターン展開法を適用してみた。それらのデータはそれぞれ60, 61, 31, 16バンドのデータが1組、30バンドのデータが3組と15バンドのデータが7組、8バンドのデータが9組、7バンドのデータが8組、6バンドのデータが19組と4バンドのデータが25組である。

これらのデータの組を使って  $\chi_{n-3}^2$  を計算し、バンド数  $n$  に対してプロットしたものが図1である。この図のバンド数15以上の値は当然前回の値と同じであるからほぼ同じで大体 0.002 という値を持っている。しかしバンド数15以下では  $n$  が小さくなると  $\chi_{n-3}^2$  はかなり大きくなると言わなければならない。

この  $n$  が小さいときの  $\chi_{n-3}^2$  の上昇はパターン展開係数に対する非負条件によるところが大きい。前稿では述べなかったがパターン展開法ではパターン展開係数に非負の制約条件を付けてフィットをおこなっている。これはパターン展開法では展開係数をそれぞれ水・植生・土壌の地上成分がその画素内に存在する割合を表すと考えるからであり、存在の割合は負にはなり得ないからこのような制約条件を付加するのである。

しかし数学的には非負条件なしでのフィットの方がきれいである。実際にも

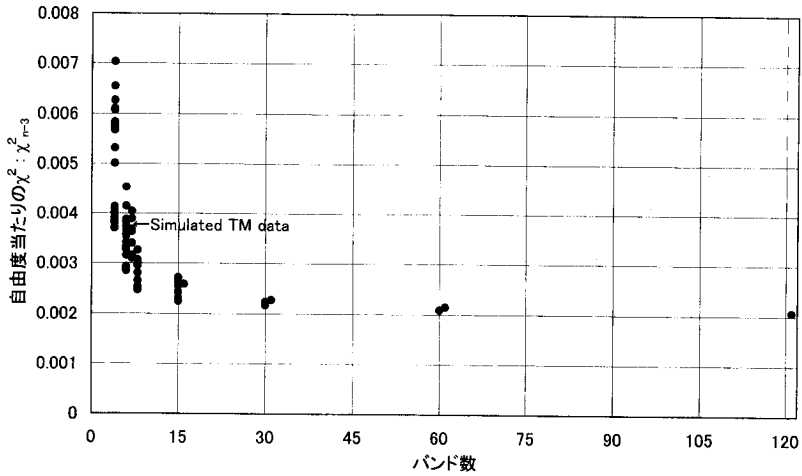


図1 バンド数の変化に対する  $\chi^2_{n-3}$  のグラフ

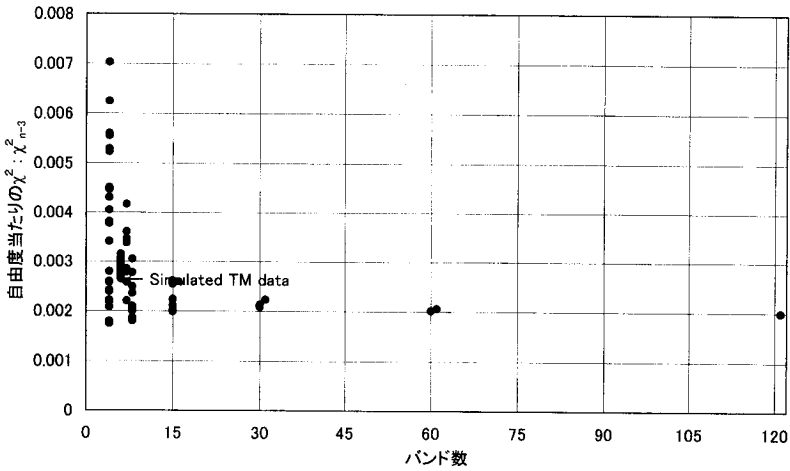


図2 バンド数の変化に対する非負条件を付けないパターン展開法で計算した  $\chi^2_{n-3}$  のグラフ

非負条件なしでフィットしたときの  $\chi^2_{n-3}$  をバンド数に対してプロットすると図2のようになり、 $n$  が小さくなくてもあまり  $\chi^2_{n-3}$  は上昇しなくなる。

このように非負条件なしのパターン展開法においてはフィットしたときの各バンドの誤差は自由度  $n-3$  の  $\chi^2$  分布に従うという仮定がよく成立するといえる。そして各自由度の分散はすべて同程度で約 0.002 となる。ただしこの結論は非負条件なしのパターン展開法に対するものである。我々が画像解析に実際に使用するパターン展開法は非負条件を付けて展開係数を求めるので上記の仮定は必ずしも成立しなくなり、誤差も上記の値より多少は大きくなることに注意が必要である。

### Ⅲ 主成分分析法から求める情報量

#### 1. 主成分分析法における情報量の求め方

多波長の画像データを解析する代表的な手法に多変量解析の手法を画像解析に応用した principal components transformation (主成分変換法) がある。主成分変換法は多変量解析のうちの主成分分析法を用いる。主成分分析法では寄与率によって各主成分が含む情報量を表す考え方が<sup>6)</sup>ある。本稿ではこの手法を用いて主成分変換法やパターン展開法で保持される情報量について考察する。そこで主成分分析法での情報量の考え方について以下で簡単に説明を行う。

主成分分析法を  $n$  バンド画像データに適用するには  $n$  次元ベクトル空間を考え、画像の各画素 (ピクセル) をその空間内の一つの点で表す。一つの画像は空間内の点の集合であり、その集合に対して主成分分析を施して画像の分類に適する新しい座標系を求める。

新しい座標系は上記の  $n$  次元空間内の画素集合の分布から得られた誤差行列 (相関行列または分散共分散行列) の固有値と固有ベクトルを求める事によって定める。この時新しい座標軸の方向は固有ベクトルの方向として与えられるので当然座標系は直交系となる。そして新しい座標系での誤差行列は非対角要素がゼロの対角行列となるので、得られた座標系での変数は互いに無相関である。またそれぞれの固有値は各固有ベクトルの方向でのデータの分散を表すことになる。そして新しい座標系での各座標軸は主成分と呼ばれる。

ここである主成分の分散が大きいと言うことはそのデータを測定して得られる情報が大きいと言うことである。逆に分散が小さいと言うことは元々そのデータの値が限られた範囲に入っていて、その値が得られても得るところが少ないと言うことを表している。すなわちある主成分の分散はその主成分の測定値が持つ情報量を表していると考えることが出来る。

そこで主成分分析では分散の総和に対する各主成分の分散の割合をその軸の寄与率と呼ぶ。分散の総和は先程の話から固有値の総和、すなわち誤差行列のトレースに等しいことが解る。よって各主成分の寄与率  $C_i$  は

$$C_i = \frac{\lambda_i}{tr(\Sigma)}$$

で与えられる。ここで  $\lambda_i$  は第  $i$  主成分の固有値すなわち分散、 $tr(\Sigma)$  は誤差行列のトレースである。そして1番目から  $j$  番目までの主成分の寄与率の和

$$Pr = \sum_{i=1}^j C_i$$

を第  $j$  成分までの累積寄与率という。

この様に主成分分析では各座標軸の持つ情報量の割合という概念が存在する。そこで以下ではこの考えを衛星データに適用して、各主成分の持つ情報量、そしてパターン展開係数の持つ情報量について考察してみる。

ここでもう一つ注意しておきたいことは、規格化についてである。前稿で述べたようにパターン展開法では基本パターンの規格化には各バンドの線形和を利用している。これは、パターン展開法では展開係数とスペクトルの大きさの線形性を保つ為に、スペクトルを表すベクトルの大きさに、普通の二乗和のルートではなく、各バンドの線形和を使用しているからである。

しかし今回の解析は主成分分析法を用いるのであり、これは普通のベクトル空間の概念を用いている。このことから以下の解析においては主成分分析法の概念を正しく利用するためにベクトルの大きさに普通の二乗和の表現を用いているのである。

## 2. 利用したデータと解析の結果

この解析においては行列の固有値方程式を解かなければならない。この場合分光放射計のデータは121次元であり、121次の固有値方程式を解かなければならない。今回は使用したワークステーションの性能と時間の関係から、分光放射計のデータに関しては相関行列を用いた主成分分析のみを行った。

分光放射計のデータ以外にこの解析では実際の衛星データを利用した解析も行った。この解析に使用したデータは LANDSAT-5/TM データである。場所は北近畿 (path: 110, low: 35) で1986年10月14日のデータを使用した。この

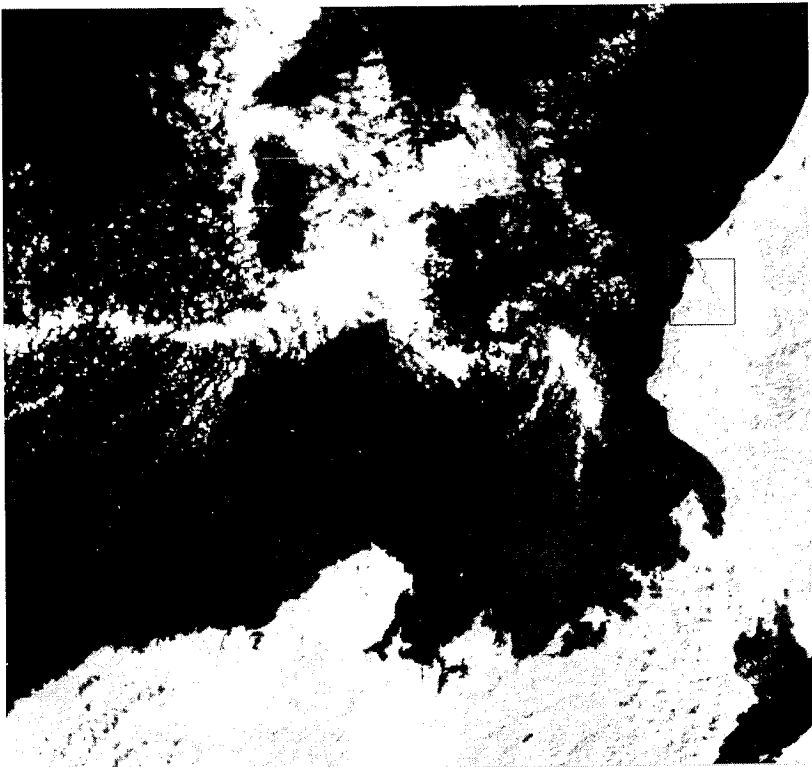


図3-(a) 主成分分析法による解析に使用したシーンの全体

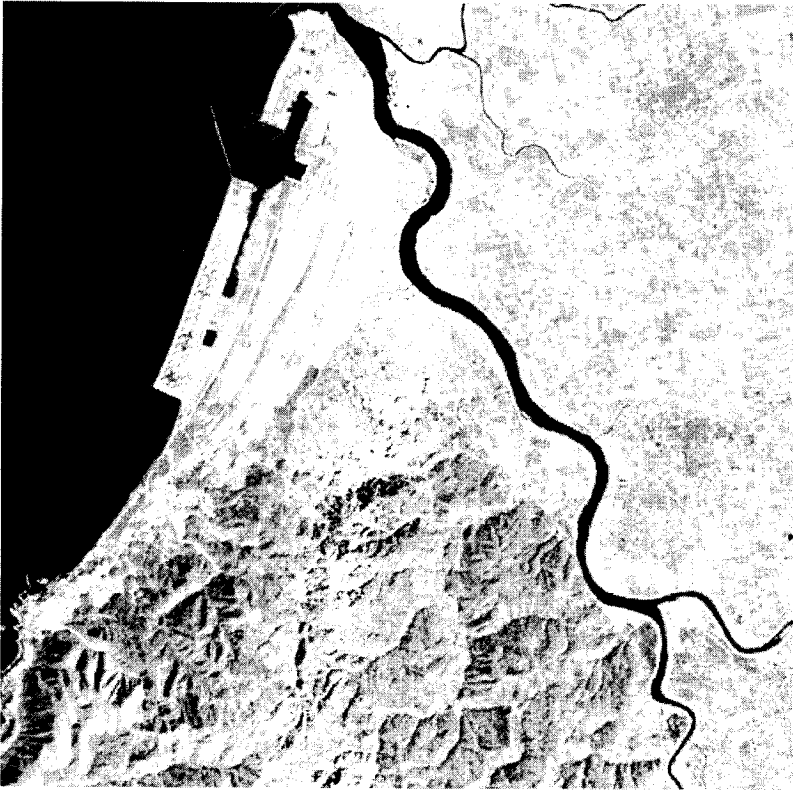


図 3-(b) 実際に使用した512×512画素の領域

シーン全体は近畿北部から福井，金沢あたりまでを含んでいるが，今回はその中で福井近郊，正確には九頭竜川河口から福井市の近郊までの512×512ピクセルの領域を使用した。この領域では目視により水・植生・土壌の部分がそれぞれ1:1:1であるように選択した。使用したシーン全域と実際に使った部分を図3に示す。

実際の解析結果であるが，分光放射計から平均して121バンドにした595サンプルのデータに対して相関行列による主成分分析を行った結果は表1の通りである。この表でははじめの10個の主成分の固有値，寄与率と累積寄与率が表示



表1 分光放射計データに対する相関行列を用いた主成分分析の結果

主成分の番号	固有値	寄与率	累積寄与率
1	77.97	0.6444	0.6444
2	24.77	0.2047	0.8491
3	12.05	0.09956	0.9486
4	2.053	0.01697	0.9656
5	1.323	0.01093	0.9765
6	0.8542	0.007060	0.9836
7	0.3236	0.002674	0.9863
8	0.2814	0.002325	0.9886
9	0.1594	0.001317	0.9899
10	0.1354	0.001119	0.9910

最初の10個の主成分に対する固有値と寄与率を表示してある。

表2 分光放射計データに対する相関行列を用いた主成分分析の結果

	水	植生	土壌
正規化された標準 パターンの射影長	0.9376	0.9936	0.9931
主成分の番号	水の方向余弦	植生の方向余弦	土壌の方向余弦
1	0.5771	0.7602	0.9872
2	0.2466	0.6354	-0.04453
3	0.6965	0.07480	-0.09889
4	0.2718	0.02607	-0.08635
5	0.2076	-0.07920	-0.06280
6	-0.03412	0.04334	0.01120
7	-0.008761	0.03727	0.02541
8	0.01271	-0.02738	-0.008851
9	0.003201	-0.02574	-0.02828
10	-0.01734	-0.01011	-0.0192

正規化された三基本パターンの射影長と方向余弦を表す。

されている。

はじめの三つの主成分は残りの主成分に比べて寄与率が高いことは明らかである。三番目の主成分までの累積寄与率は94.9%であり、最初の三つの主成分によって構成される三次元サブスペースに殆どの情報が含まれていることが判る。

この三次元サブスペースと我々のパターン展開法で用いる三つの基本スペク

トルパターンとの位置関係を見るために、基本スペクトルパターンと同じ方向の単位ベクトルを上記の三次元サブスペースに射影した長さを計算してみると表2のようになった。表の上の部分は三基本パターンのサブスペースへの射影長で、下の部分は基本パターンと10主成分軸との方向余弦を表す。

表2から水・植生・土壌の基本パターンのサブスペースへの射影長はそれぞれ0.938, 0.994, および0.993となることが判る。水の射影長は少し小さく角度に直すと20度程になるが、後二つのパターンは角度に直して6度程度である。このことから主成分分析法から得られた最初の三つの主成分軸で構成された三次元サブスペースと我々の三基本パターンによって構成される三次元サブスペースは121次元空間の中でお互いに近い位置関係にあるとすることが出来るであろう。

次に同じ解析手法を実際の衛星データに適用してみた。この場合サンプル数は262,144と多くなるが、バンド数は6である。

まずは相関行列を用いて主成分分析を行うと、固有値・寄与率は表3の通りとなる。この時は3番目の主成分の寄与率は4番目のその倍程度であまり違いがないと言えるが、3番目の主成分までの累積寄与率は97.7%と大きな値を持つことにはかわりはない。そして表4の射影長を見ると、一番小さな射影長を持つ水でさえその値は0.980で、角度に直すと11度と大変小さな値となる。すなわちこの場合は主成分分析で得られた最も情報を多く持つ三次元サブスペースと基本パターンで構成されるサブスペースは殆ど同じ位置にあるといえる。

次に分散共分散行列を用いて主成分分析を行うと、固有値・寄与率は表5の通りとなる。そして3番目の主成分の寄与率は4番目のその5倍程度で、第3主成分までの累積寄与率は99%と殆どの情報は第三主成分までに含まれることになる。表6の射影長を見ると、水のそれは0.86とかなり小さな値を持つ。これを角度に直すと30度となり、基本パターンからのサブスペースと分散共分散行列による主成分分析からのサブスペースとは少しだけ離れていると

表3 実際の衛星データに対する相関行列を用いた主成分分析の結果

主成分の番号	固有値	寄与率	累積寄与率
1	4.511	0.7519	0.7519
2	1.153	0.1922	0.9441
3	0.1941	0.03236	0.9765
4	0.1021	0.01701	0.9935
5	0.02219	0.003698	0.9972
6	0.01687	0.002811	1.0

各主成分に対する固有値と寄与率を表示してある。

表4 実際の衛星データに対する相関行列を用いた主成分分析の結果

	水	植生	土壌
正規化された標準 パターンの射影長	0.9800	0.9892	0.9996
主成分の番号	水の方向余弦	植生の方向余弦	土壌の方向余弦
1	0.7706	0.5953	0.9284
2	0.4254	-0.6985	-0.2993
3	0.4308	0.3690	-0.2182
4	0.1978	0.1366	0.02562
5	-0.005796	-0.02307	-0.01178
6	-0.02179	0.04744	-0.008488

正規化された三基本パターンの射影長と方向余弦を表す。

考えられる。

上記の主成分分析による解析をまとめると、衛星データの情報は主成分分析法で得られた第三主成分までに95%以上含まれていることになる。そして我々のパターン展開法で使用する基本三パターンで構成される三次元空間はその第三主成分までで構成される三次元空間とほぼ一致する。このことから基本三パターンによる三次元空間もやはり衛星データの95%以上を保持していると言えるであろう。そしてこのことはデータのバンド数が121でも成立し、将来を含め我々が得られる全ての衛星データについて成立することになるのである。

表 5 実際の衛星データに対する分散共分散行列を用いた主成分分析の結果

主成分の番号	固有値	寄与率	累積寄与率
1	10806.	0.7899	0.7899
2	2408.	0.1761	0.9660
3	336.5	0.02460	0.9906
4	79.87	0.005839	0.9964
5	38.48	0.002813	0.9992
6	10.57	0.0007726	1.0

各主成分に対する固有値と寄与率を表示してある。

表 6 実際の衛星データに対する分散共分散行列を用いた主成分分析の結果

	水	植 生	土 壤
正規化された標準 パターンの射影長	0.8618	0.9956	0.9949
主成分の番号	水の方向余弦	植生の方向余弦	土壌の方向余弦
1	0.3789	0.9234	0.9002
2	0.4246	-0.3404	0.4168
3	0.6472	0.1504	-0.07509
4	0.3231	0.0605	0.1005
5	0.3768	0.06939	-0.004046
6	0.1041	-0.01884	-0.009855

正規化された三基本パターンの射影長と方向余弦を表す。

## ま と め

放射分光計のデータと実際の LANDSAT/TM のデータを解析することにより、パターン展開法で求められる展開係数は三つのパラメータで元のデータを大変精度良く再現できることが判った。またパターン展開法は主成分分析法とは全く独立に開発されたものであるにもかかわらず、その三基本パターンが構成する三次元空間は主成分分析法で得られた主要三主成分が構成する三次元空間と極めて近い関係にある。これはパターン展開係数が三主成分が保持する 95%と同程度の情報を保持できている事を表す。

これはパターン展開法が衛星データが元々持っていた性質をうまく利用でき

ていることを表すと言えるであろう。そして主成分分析法で得られた主成分自身には固有の実体的な意味がない一方で、パターン展開法で利用する基本三パターンにはそれぞれ水・植生・土壌と言う固有の意味があり、展開係数は直接その画素の成分割合として解釈できるという特徴がある。

このようにパターン展開係数は元の情報の殆どを保持出来ていると言えるであろうが、これまでの解析は  $\chi^2$  フィットにしる主成分分析法にしる、データの分布が正規分布であると言うことを仮定している。しかしながら実際の衛星データの分布は正規分布ではない事は明らかである。そこで現在、Shanon の情報量の理論を利用して正規分布でない衛星データの情報量を解析する方法を考察中である。

#### 謝 辞

この研究の為に奈良女子大学情報科学科自然情報学講座の皆様には分光放射計の実験データの利用を許可して頂きました。また、奈良女子大学の浦部利佳氏には有益な意見や実際の計算実行等において御助力頂きました。北近畿の LANDSAT/TM 衛星データ購入は文部省学術フロンティア推進事業 (ワールドワイドビジネス研究センター) の補助金によるものです。またこのデータの所有・提供はアメリカ合衆国政府・宇宙開発事業団によります。ここに記して感謝の意を表します。

#### 【参考文献】

- [1] Gonzalez, R. C. & Wintz, P., *Digital Image Processing*, Addison-Wesley, Mass., 1977.
- [2] Merembeck, B. F. & Turner, B. J., "Directed Canonical Analysis and the Performance of Classifiers under its Associated Linear Transformation," *IEEE Trans. Geoscience and Remote Sensing*, GE-18, 1980, pp. 190-196.
- [3] Adams, J. B. et al., "Classification of Multispectral Images Based on Fractions of Endmembers: Application to Land-Cover Change in the Brazilian Amazon," *Remote Sens. Environ.*, Vol. 52, 1995, pp. 137-154
- [4] 藤原昇, 他「衛星データ解析のためのパターン展開法の開発」『日本リモートセンシング学会誌』第16巻 第3号, 1996年, 17-34ページ.

- [5] 醍醐元正「連続波長データに対するパターン展開法の適用可能性」『経済学論叢 (同志社大学)』第50巻 第4号, 1999年, 138-149ページ.
- [6] 河口至商『多変量解析入門』I, 森北出版, 1973年.