

連続波長データに対する パターン展開法の適用可能性

醍 醐 元 正

I は じ め に

最近の衛星センサーは技術の進歩を反映して同時に多くの波長を観測出来るものが多くなっている。多波長データの全ての波長域を利用した解析手法としては, principal components transformation [1], canonical analysis [2] や spectral mixing analysis [3] 等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多波長データを解析する手法としてパターン展開法〔4〕と名付けた手法を研究開発して来た。この手法は元々 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。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。TM データにパターン展開法を適用して解析した結果によると、殆どの土地被覆物からの反射光は三つの基本的なスペクトルパターンの合成で精度良く近似出来る事が判った。すなわち TM の 6 バンドデータの範囲では地上の被覆物は水・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。

この成果をもとに、我がグループではさらに ADEOS-II/GLI センサーの

多チャンネルデータ利用の基礎研究として、GLI データに対するパターン展開法の適用可能性について研究を行った〔5〕。ADEOS-II 衛星は西暦2000年に打ち上げ予定であり、GLI からのデータはまだ利用出来ない。そこで、この調査には地上で測定された放射分光計による反射スペクトルのデータと AMSS のデータを使用した。AMSS は GLI の予備研究の為に開発された、GLI と同等の測定波長域を持つ航空機搭載センサーである。GLI センサーは 36チャンネルを持つが、熱赤外域を除いた 350-2500 nm の波長域では 29チャンネルである。その中から水蒸気や酸素に強く吸収される波長域を除き、又地上の空間分解能が異なりながら波長域が近いチャンネル等を整理して、この解析の為に 19チャンネルを選んだ。この解析で、19チャンネルのデータにおいても地上被覆物のスペクトルパターンは三つの基本スペクトルパターンの合成によって、うまく表現出来る事が判った。

この様に 6, 19チャンネルのデータに対してパターン展開法が適用可能である事が判ると、当然次にはこのパターン展開法の考え方がどの波長域まで成立するかが興味の対象となる。筆者はその点について上記の研究の際に測定された放射分光計データを用いて、可能な限り広い波長域に対してのパターン展開法の適用可能性について検討したので本稿において報告する。まず次節でこれまでの多波長データ解析手法について簡単にまとめ、Ⅲ節でパターン展開法について説明する。そしてⅣ節で解析に使用したデータについて紹介した後、Ⅴ節で解析結果を述べて議論を行う。

Ⅱ これまでの多波長データ解析手法

多波長で測定された画像データの解析手法には始めに述べた様に principal components transformation, canonical analysis や spectral mixing analysis 等がある。

principal components transformation と canonical analysis は多変量解析の手法を画像解析に利用したものである。principal components transformation

は主成分分析法, canonical analysis は外的基準のある場合の主成分分析法を用いる.

これらの解析では n チャンネル画像を解析するのに 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素の分類に適する新しい座標軸を求めるのである. 多変量解析法の画像データへの自然な適用であり, 数学的にも綺麗な方法であると言えるであろう.

しかしこれらの方法で得られた新しい座標軸は, 数学的には最適なものではあろうが, 実体的な意味はない. 即ち我々が普通に考える水・植生・土壤等の分類項目とは無関係であり, それらの分類と関連付けたいならばそれは又別に考えなければならない. また座標軸を求めるのに使う相関行列 (もしくは共分散行列) はそれぞれの画像毎に求めるので, 異なった画像を比較するのは困難である. すなわち経時変化や地域差等に言及するにはそれなりの工夫が必要である.

spectral mixing analysis は end members method と呼ばれ, その名の通り分類したいカテゴリー (end members) 毎に基本のスペクトルを定め, それらの基本スペクトルの線形和で各画素のスペクトルを近似しようとするものである. そして線形和の係数で, 画素がそれぞれのカテゴリーに帰属する割合を表す事になる.

この解析法は地表被覆物の分類に有力であるが, 分類のカテゴリーは各解析毎に定められるので, 分類カテゴリーの異なる解析間の比較は不可能である. また principal components transformation 等と同様に画像間の比較は簡単ではなく, 経時変化や地域間の比較等を行おうとすると, データの校正に多大の労力を必要とする.

Ⅲ パターン展開法

1 パターン展開法とは

パターン展開法は spectral mixing analysis の一種とも考えられる. パター

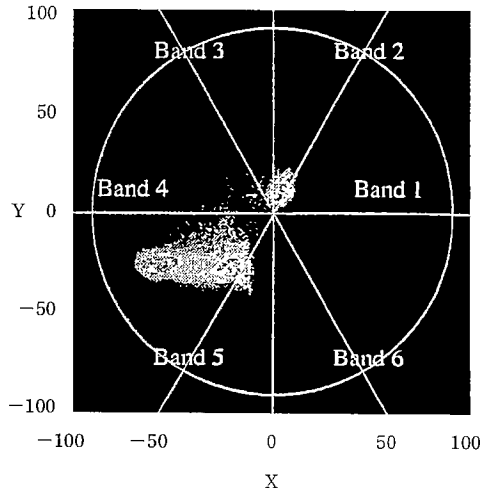


図1 LANDSAT/TM の6次元空間を2次元上に投影した図。
 各バンド軸はそれぞれ60°の角度をなす様に投影されている。
 また各画素のベクトルはパターン展開法の意味で1に規格化されている。

ン展開法でも三つの基本スペクトルパターンを定め、それらの線形和で各画素のスペクトルを表現しようとするからである。ただ spectral mixing analysis では画像や解析者が異なる毎に分解の基本のスペクトルが変わるのに対して、パターン展開法ではその基本スペクトルパターンが固定されているのが異なる。そして基本パターンが固定しているので解析間の比較が可能になる。

この様に基本スペクトルを固定したのは、地球からの反射光のスペクトルパターンがある種の構造を持つという事実由来する。この構造とは、多チャンネル衛星画像データの各画素を表す点を n 次元空間の中で表示すると、殆どの点は空間の三つの部分とそれらを結ぶ部分に集中すると言う事を指す(図1参照)。さらに詳しく調べると、実際のデータは3次元超平面の、それも基本スペクトルパターンのベクトルで構成される三角錐の部分に集中する事が判る。

この事実は一方で principal components transformation 等とも関連するとも

思われる。多分 principal components transformation ではその第 3 軸までで殆どの情報は表現されるのであろう。パターン展開法との違いは principal components transformation はそれらの三つの軸が直交しているのに対して、パターン展開法では、上記の自然の構造を生かす為に、三つの軸が独立ではあるが直交していないと言う事である。この様に軸を定めた事によりパターン展開法の展開係数は水・植生・土壌の分類項目と直接関連付けられる事になるのである。

2 パターン展開法の表式

パターン展開法では各画素のデータは輝度値から反射率に変換して使用される。各バンドの反射率 A_i は次の様にして近似される。

$$A_i \rightarrow C_w P_{iw} + C_v P_{iv} + C_s P_{is}$$

P_{iw}, P_{iv}, P_{is} : 水・植生・土壌の標準スペクトルパターン

$$\text{但し} \left(\sum_{i=1}^n P_{ik} = 1, (k=w, v, s) \right)$$

即ちバンド和を 1 に規格化している

C_w, C_v, C_s : 水・植生・土壌の展開係数

展開係数は最小二乗法で決定される。すなわち各バンドの余り R_i

$$R_i = A_i - (C_w P_{iw} + C_v P_{iv} + C_s P_{is})$$

を使用してその二乗和

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

が最小になる様に定められる。そして全チャンネルの相対誤差 E を

$$E = \sqrt{X} / \sum_{i=1}^n A_i \quad (2)$$

と表す。

ここでフィットは二乗和の最小化を使い、標準スペクトルの規格化には各バンドの線形和を利用している事が気になるかも知れない。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、普通の二乗和の平方根ではなく、各バンドの線形和を使用しているからである。この様にすると三つの展開係数の

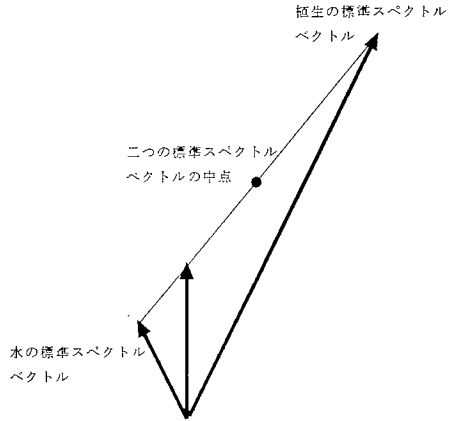


図 2

和が近似されたスペクトルのベクトルの大きさに等しくなるのである。

標準スペクトルパターンを1に規格化しているのはスペクトルの形だけを取り出しているのであるが、もっと実的な意味もある。この三つの標準スペクトルでは水の反射率が際だって低く、スペクトルパターンの各バンドの値も小さい。これをそのまま使用してパターン展開を行うと、例えば水と植生の丁度中間の角度の方向にあるベクトルでは図2の様に、水の展開係数の方が大きくなる。すなわち水に近い画素であると判断されてしまう。それでも数学的には問題はないのであろうが、我々は展開係数を実際の分類に使用したいのであり、空間的に一様に分布した点をパターン展開すると水に近い画素と植生に近い画素がそれぞれ半数程度であって欲しい。そこで三つの標準スペクトルパターンの大きさを1に規格化する事によってこの不都合を回避しているのである。

IV 使用 データ

使用した放射分光計は Field Spec Fr 及び Opto Research 社の MSR7000 であり、どちらも波長 1 nm 間隔のデータが得られる。しかし実際の衛星センサーのバンド幅は 10 nm 程度以上のものが多い。そこで今回の解析でも 1 nm

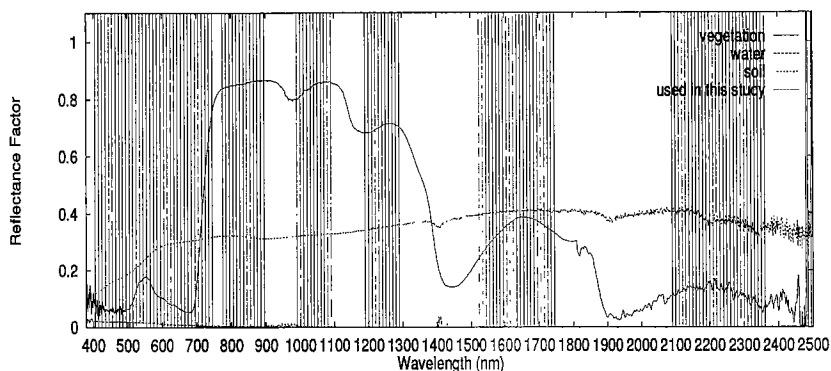


図3 今回の連続波長データの解析に使用したチャンネルを縦線で表示してある。
三つのスペクトルは規格化されていない，標準スペクトルを表している。

間隔のデータの10個の平均を取って10 nm バンド幅のデータとして使用した。

波長領域も大気の透過率が80%以上の図3の領域のみを解析の対象とした。これは実際の衛星データでは大気の透過率の低い波長域は使用しないからである。また水の標準スペクトルパターンの決定が実験室では不可能で，自然の太陽光の下で行う必要があるからでもある。実験室内で水のサンプルの反射率を測定しようとする，どうしても容器の底の反射の影響があり，測定が不正確になる。そして太陽光の下での測定では，大気の透過率の低い波長域は入射光が少なくてやはり測定が不正確になるのである。

この結果，使用したバンド数は121となった。またこの解析に使用したサンプルは約600であった。

図4に今回の解析で使用した水・植生・土壌の規格化した標準スペクトルパターンを示す。標準パターンの為のサンプルは，それぞれ，和歌山県加太港の海水，アラカシの葉を十枚重ねたもの，中国敦煌近辺の砂漠の砂を使用した。

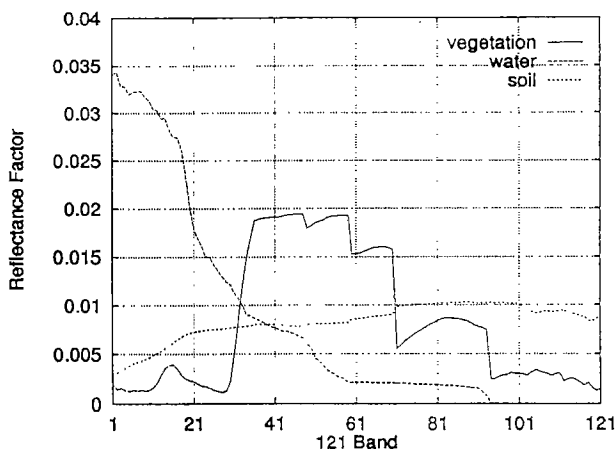


図4 規格化された三つの標準スペクトルパターン

V 解析の結果と議論

幾つかのサンプルデータに対して121全てのチャンネルを使ってパターン展開を適用した結果を図5に示す。この図よりパターン展開の余り R_i は非常に小さい事が判る。図6は全600サンプルに関する相対誤差 E の分布をプロットしたものである。全サンプルの相対誤差の平均は約1.3%となる。

チャンネル数を減少させた時に、相対誤差がどうなるかを見る為に、一つおき、三つおき七つおきにチャンネルを取りだしたデータ等を作って、それらに対してパターン展開法を適用して見た。それらのデータはそれぞれ60チャンネルのデータが2組、30チャンネルのデータが4組と15チャンネルのデータが8組である。

それらのデータの組のパターン展開法での相対誤差を、使用したチャンネル数に対してプロットしたものが図7である。これを見るとチャンネル数が増加するに従って相対誤差が小さくなっている。チャンネル数が増加するとフィットが更にうまくいくと思えるが、実はこれは相対誤差の定義の結果である。

パターン展開法では三つの展開係数を使用するので、チャンネル数 n の

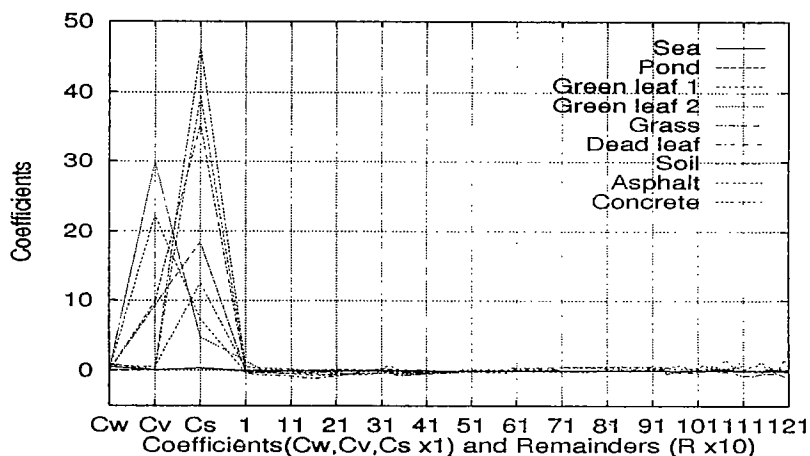


図5 幾つかのサンプルデータにパターン展開法を適用した結果. グラフの左側三つの値がそれぞれ水・植生・土壌の展開係数, その右側から121のチャンネルそれぞれの余りを表示している.

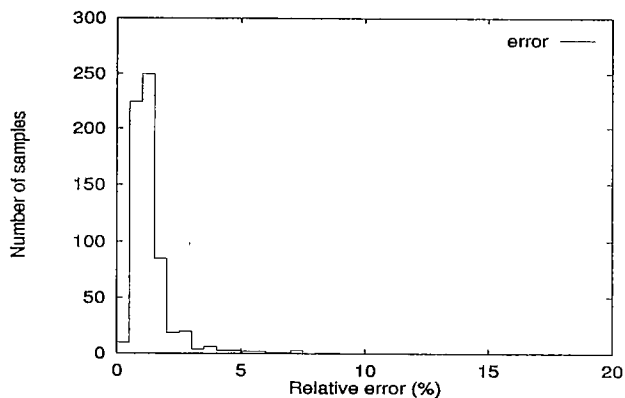


図6 600サンプルに対する相対誤差

データでは自由度は $n-3$ である. もしそれらの自由度に対するフィットが全て独立に行われ得て, かつ同じ程度の分散を持っているとするならば, 式(1)で表される余り R_i の二乗和 X は $n-3$ に比例する. 一方式(2)で表される相

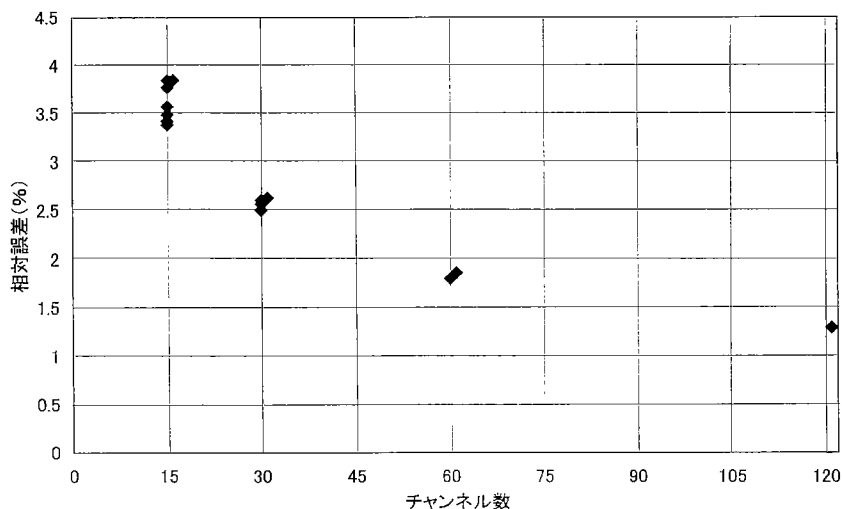


図7 チャンネル数の変化に対する相対誤差の変化のグラフ

対誤差を考えると、チャンネルの線形和である分母が大体 n に比例するとして、分子は $(n-3)^{1/2}$ に比例するから、相対誤差そのものは n が大きくなると $n^{-1/2}$ に従って小さくなって行くのである。

これは相対誤差が、チャンネル数が変化しない時にその誤差の程度を見積もるためのものだからであり、今回の様にチャンネル数、すなわち自由度が変化する時には別の指標を用いるべきなのである。その為の指標として今回は自由度当たりの余りの二乗和

$$K = X/\text{freedom} = \sum_{i=1}^n R_i^2 / (n-3)$$

を用いた。この K をチャンネル数 n に対してプロットしたものが図8である。これらは全てほぼ同じで大体 0.002 という値を持っている。先程の仮定、すなわち

- (1) 全ての自由度が独立
- (2) 同じ分散の値を持っている

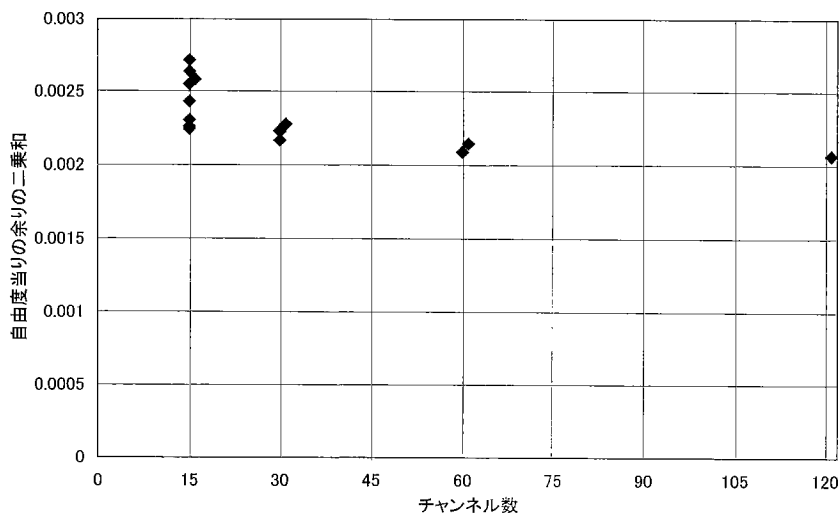


図 8 チャンネル数の変化に対する自由度当たりの余りの二乗和のグラフ

という条件が満たされていれば、先程の 0.002 という値はこの分散に等しくなる。そしてこの値の平方根 0.045 はチャンネル当たりのフィットの誤差であり、その値はチャンネル数が増加してもそれほど変化しないと言う事が判る。

またこの事をベクトル空間上での言葉に直せば、標準スペクトルパターンが構成する三角錐からデータ点までの平均距離は、全ての軸即ちチャンネルにおいて同程度である、と言う事である。

VI ま と め

放射分光計のデータを使用して連続スペクトルデータをパターン展開すると、LANDSAT/TM データや GLI の波長帯のデータと同様に土地被覆物のスペクトルパターンを三つの標準スペクトルパターンに分解する事が出来た。この事実を利用すれば、ある程度以上のバンド数があるセンサーに対しては、パターン展開法を適用して三つの展開係数を求め、これらの展開係数を使用して同じ枠組みの中で利用する事が可能となる。その為にはパターン展開法を用い

るものが全て同じ標準スペクトルパターンを使用する事が重要である。我々のグループはその為の標準スペクトルパターンを定め、現在公開の準備を行っている。

この様にパターン展開法が連続スペクトルデータに対して適用可能であると言う事実はパターン展開法の応用を広げるものであり、大きな意味を持つであろうが、筆者はその事よりもこの地球上に充満する光のスペクトルが殆ど三つのパターンの合成として表現可能であるという事実の方に心を引かれる。特に我々の目がこの世を三つの色で見ているという事との暗合が興味深く、将来何らかの形でこの二つの事実が結びつけられれば面白いのではないかと考えている。

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実験データの利用を許可して頂きました。ここに記して感謝の意を表します。

【参考文献】

- [1] Gonzalez, R. C. & P. Wintz, *Digital Image Processing*, Addison-Wesley, Mass., 1977.
- [2] Merembeck, B. F. & B. J. Turner, "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] Hayashi, A. et. al., "An Algorithm and a New Vegetation Index for ADEOS-II/GLI Data Analysis," *Journal of The Remote Sensing Society of Japan*, Vol. 18, No. 2, 1998, pp. 28-50.