

機械学習法を用いた可視近赤外波長域鉱物反射スペクトル解析

○坪井伸子¹, 杉田精司¹, 廣井孝弘², 永田賢二¹, 岡田真人¹

¹東大・新領域・複雑理工,²ブラウン大・地球科学

1. Introduction

衝突クレータの形成時に天体内部が掘削されるため、衝突クレータは天体内部の情報を得ることができる場所である。その観測には、リモートセンシングの手法がよく使われる。特に、可視・近赤外波長域は、鉱物に特徴的なスペクトルを示し、可視・近赤外反射スペクトルの分布からは天体表面の鉱物分布を得ることができる。このような観測から得られた天体の鉱物についての情報は、天体の環境や熱的・化学的進化過程を考察できる重要な知見を与えてくれる。しかし、天体表面の可視・近赤外反射スペクトルには、ケイ酸塩鉱物による複雑に重なり合った複合吸収帯が存在するため、その解析は容易ではない。この複合吸収帯を分離する方法として MGM (Modified Gaussian Model) がある (e.g., Sunshine et al, 1990)。MGM は、複合吸収帯を複数のガウス関数で分離することで、連続スペクトルがもつ豊富な情報を効率よく抽出することができる。

月探査衛星かぐやなど近年の探査機からは大量スペクトルデータが得られており、煩雑な手動解析に頼らない手法が求められている。MGM による自動解析については、様々な研究が行われているが、いまだに実現できていない。その理由の一つとして、ガウス関数の数を主観的に決めていたことが挙げられる。MGM 自動解析にとって、ガウス関数の数を決めることは第一に行わなければならない。なぜならもしガウス関数の数が変われば、中心波長位置・強度が変わってしまい、得られる結果に大きな影響を与えるからである。

そこで本研究では、この問題を解決するために、最適なガウス関数の数を自動的・客観的に決めることができる Cross-validation を MGM に取り入れることを提案する。また今回は、代表的なケイ酸塩鉱物である Olivine と Pyroxene の混合物の反射スペクトルを用いて、提案手法の有用性の検証を行った。

2. Method

Cross validation は、最適なモデルを選択する場合に一般的に使われる統計学的手法である (Stone, 1974)。また Cross-validation はモデルの Overfitting を防ぐ手段としても有効である。変数 (ガウス関数)

の数を増やす程、モデルと観測データとの二乗誤差は小さくなる Overfitting という現象がある。そのため、二乗誤差の最小化だけから最適なモデルを選ぶことはできない。そこでデータの一部を評価に使う Cross validation が考え出された。Cross validation は、まず観測データをランダムに test set と validation set に分割する。Test set を使ってモデルを構築し、その推定されたモデルと validation set の RMS error (ここでは validation error と呼ぶ) からモデルを評価する。Validation error は、初めは変数を増やす程減少していくが、ある点から増加していく。これは、validation set には含まれない test set のノイズにまでモデルが無理に合わせてしまったためである (Overfitting)。したがって validation error が最小になる点が最適なモデルであることがわかる。(図 1)

本研究では、この Cross validation を MGM に取り入れる。MGM 解析する際、ガウス関数の数を 1,2,3, … と変化させながら Cross validation を行う。そして validation error がある数から増加したなら、増加する前の数 (つまり validation error が最少になる数) を最適な数とする。このように Cross validation はガウス関数の数を適当に変化させながら最適な数を選ぶので、自動的に最適なガウス関数の数を決めることができ、MGM を自動化する際に有効な手段になる。

3. Olivine and Pyroxene Mixture

惑星を覆っている地殻には Pyroxene が多く含まれる。一方で、Olivine はそれほど頻繁には地殻に産しない。惑星表面に Olivine が露出するのは、塩基性の非常に強い溶岩やマントル物質が露出する場合など地質学的な特徴が明瞭なケースが多い。そのため、Olivine が惑星表面に観測された場合、その発見意義は非常に大きい。しかし、混合した反射スペクトルの見た目から Olivine の存否を正確に判断することは、非常に困難である。なぜなら Olivine と Pyroxene の 1 μm 付近の吸収帯の位置は非常に近接しているため、ガウス関数同士が重なり合い、本来現れるはずの吸収帯が観測できない可能性がある。また、透明な緑の Olivine より、黒色である Pyroxene の方が圧倒的に吸収効率が高い。そのため、Olivine の吸収が相対的に軽微となる。本研究では、Cross-validation

の適用で Pyroxene に富む岩相のスペクトルから Olivine を検出できるか検証を行う。具体的には、Pyroxene と Olivine が混合した試料の反射スペクトルについて Cross-validation を取り入れた MGM を行った。そして各鉱物に特徴的な吸収帯を検出できるか、最適なバンド数に違いがあるか検証を行った。

混合物に本手法を適用する前に各鉱物の最適なバンド数を知る必要がある。Olivine については4つのガウス関数で分離できること(Tsuboi et al., 2009, Lunar Planet. Symp)、Pyroxene については5つのガウス関数で分離できることがわかっている(坪井 他., 2009, 惑星科学会秋季講演会)。

4. Result and Discussion

Pyroxene と Olivine の混合比が 50 : 50 と 75 : 25 で混合した 2 つの反射スペクトルについて本手法を適用した。Pyroxene と Olivine の混合物のスペクトルは、両者の最適なバンド数を足し合わせた 9 つのバンドを持つと予想されるが、結果はそれより少ない 6 つのバンドが最適であることがわかった(図 1)。これは、混合物のスペクトルでは、Olivine と Pyroxene の吸収帯が互いに重なり合ってしまうことが原因であると考えられる。さらに各スペクトルについて、1.05 μm 付近の Olivine に由来する吸収帯 (Sunshine et al. 1998) が観測された(図 2)。この Olivine のみが持つ吸収帯の観測は Olivine の含

有率が 25% と 50% の両方のケースで確認された。したがって Olivine が 25wt% 以上含まれていれば Cross-validation を取り入れた本手法で Olivine に由来する吸収帯を検出できると言える。

さらなる解析として、LCP (Low-Ca pyroxene) と HCP (High-Ca pyroxene) の混合物について検証する。Olivine と HCP の 1 μm 付近の吸収帯の位置は近接している。例えば LCP に混合している鉱物が Olivine か HCP かの区別は難しい。そこで、LCP 50% と HCP 50% で混合したスペクトルに本手法を適用した。そして LCP と HCP の混合物と LCP と Olivine の混合物のスペクトルとでは、最適なバンド数や中心波長位置に違いがあるかについて検証を行った。その結果、LCP と HCP の混合物のスペクトルは、LCP と Olivine の混合物と異なる 5 つのバンドで分離できることがわかった。特に、LCP と HCP の混合物のスペクトルには、1.05 μm 付近の吸収帯は検出されないことがわかった(図 3)。これらの結果から、本手法を用いることで LCP に混合しているのが HCP か Olivine か判別できることがわかった。

Cross-validation を MGM に取り入れることで、混合した試料の反射スペクトルでも、混合前の各鉱物の吸収帯の特徴を検知できることがわかった。このように本手法を用いることで、自動的に鉱物の同定が可能になると期待している。

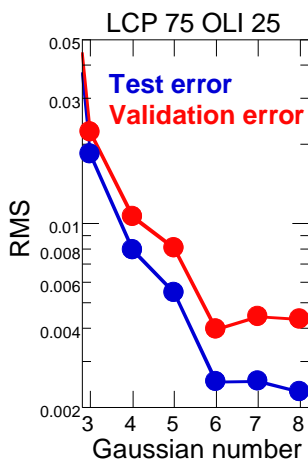


図 1

LCP と Olivine が混合した試料の反射スペクトルに提案手法を適用した結果。ガウス関数の数を増やすほど test error は減少するが、validation error はガウス関数の数が 6 を境に増加する。このことから最適な数は 6 個であることがわかる。

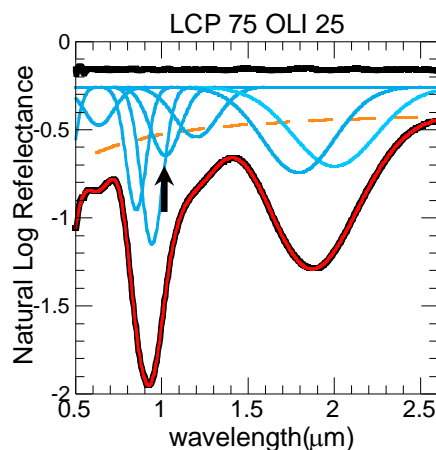


図 2

LCP と Olivine が 75:25 で混合した試料の反射スペクトルに MGM を適用した結果。ガウス関数の数は Cross validation を使って 6 個と決めている。1.05 μm 付近に Olivine に由来する吸収帯 (矢印で示している) があることがわかる。

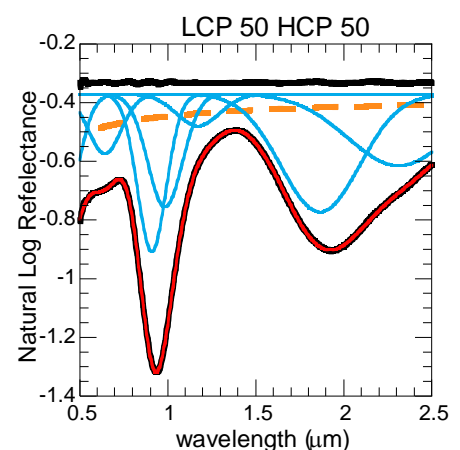


図 3

LCP と HCP が 50:50 で混合した試料の反射スペクトルに MGM を適用した結果。ガウス関数の数は Cross validation を使って 5 個と決めている。図 2 と異なり 1.05 μm 付近に吸収帯は現れない。