

博士論文（要約）

正則化を用いたハイパースペクトルデータによる  
森林モニタリング

東京大学大学院工学系研究科先端学際工学専攻  
高山 泰一

森林の伐採や火災等による攪乱で排出される CO<sub>2</sub> は最も注視すべき温室効果ガスであり、そして気候変動においても重要な主要因子である。この森林由来の温室効果ガスの排出を抑制する活動のためには、吸排出量を正確に算定するための測定方法の確立が必要である。最も正確なモニタリング方法は、踏査による地上調査であるが、世界中における本活動に充てられる資金及び人的、時間的コストも限られているため、高精度且つ低コストに森林モニタリングを実現することが必要となる。そのため、広域を一度にモニタリングが可能なりモートセンシングデータの活用に期待が集められており、実際にすでに、マルチスペクトルセンサや SAR、そして LiDAR などを用いて様々な研究が行われており、近年では航空機搭載ハイパースペクトルセンサを用いた研究も行われ始めている。このハイパースペクトルセンサとは、多波長および高スペクトル分解能のデータが得られる光学センサであり、このセンサデータを用いたバイオマス推定等の回帰問題や樹種分類等の分類問題への適用事例が存在する。回帰および分類の双方ともに十分な精度の推定モデルを構築するためには、ハイパースペクトルセンサデータと地上調査により作成されたデータセットを用いて教師つき学習を行うことが一般的である。この地上調査データの取得には、実際にハイパースペクトルセンサで撮像したエリアへの踏査が必要であるが、実際の森林モニタリングでは地上調査にかけることができる資金、時間、そして人的リソースには制限がある上、ハイパースペクトルセンサで撮像した時期から、森林の攪乱を含めた土地被覆の状況が変化する前に地上調査を実施せねばならず、結果として統計的な観点から十分な数の地上調査地点を確保することは困難である。さらに、通常 100 バンド以上を持つハイパースペクトルセンサを推定モデルに使用する場合、推定モデルの説明変数となるハイパースペクトルセンサのバンド数に対して、地上調査地点数が非常に少なくなる現象が起りやすくなる。この「説明変数の数 ≧ 地上調査地点数」という状況下では、既存の教師つき学習による手法では学習過程において過学習が起りやすくなり、結果として得られる推定モデルが教師データ以外のデータにはフィットしない、つまり推定精度が低下する **Small-sample-size problem (SSSP)** と呼ばれる状況に陥る。ここで起こる過学習を回避し、一般性を担保した推定モデルを作成するためには、「次元の遷移」、「次元圧縮」、そして「元の次元空間を維持したままでの特徴選択 (Feature selection)」という大きく 3 つのアプローチが存在する。本研究では、回帰問題そして分類問題のそれぞれにおいて正則化を用いることで「元の次元空間を維持したままでの Feature selection」を行い、SSSP 下における過学習を回避する手法の提案および推定精度の評価を行った。推定モデルの評価には、2011 年に航空機搭載ハイパースペクトルセンサにより撮像されたハイパースペクトルデータおよび 2011~2013 年に実施された地上調査データを使用した。

まず、回帰問題については、ハイパースペクトルデータのスペクトル情報から対象ピクセルのバイオマス推定を行う回帰モデルの構築を行った。既存研究では、次元圧縮を行う **Partial Least Squares (PLS)** 回帰や次元の遷移を行う **Support Vector Regression (SVR)** が主に用いられており、これらは過学習の影響緩和のアプローチに沿った手法ではあるが、得られる回帰モデルにはすべてのバンドが寄与することになる。実際にセンサで得られるハイパースペクトルデータには、大気による吸収を受ける波長帯や森林のコンディションの違いに反応する波長帯とただノイズにより分散が激しい波長帯など、バイオマス推定に寄与する波長帯とそうでない波長帯が存在する。PLS や SVR のようにすべてのバンドが回帰モデルに使用される場合、SSSP 下ではバイオマス推定に寄与しないバンドの影響を排除することは難しく、結果として推定精度の劣化に至るリスクが存在する。従って、バイオマス推定

に必要なバンド以外は回帰モデルから排除することが、SSSP 下の回帰モデル構築において重要性が高いと考えられる。代表的な Feature selection としては、まずステップワイズ法が挙げられるが、推定にとって重要なバンドを一つずつ増やしていく過程、特に最初の説明変数が少ない段階では推定精度自体が悪く信頼性が低いいため、そのバンド選定過程においても信頼性が低くなる可能性がある。そこで、推定モデルの構築とバンド削減を同時に行うことを目的に、本研究では正則化項を持つ回帰モデルの適用を検討した。具体的には、L2 正則化項を持つリッジ回帰では理論上、バンドを減らす方向に向かいづらいため、L1 正則化項を持つ Lasso 回帰を使用した。結果として、Lasso 回帰により過学習のリスクを抑えることができ既存手法よりも高い推定精度が得られた。しかしながら、Lasso 回帰や他の既存手法では、回帰モデルへのバンド選定において隣接バンドの関係性を加味しないため、ハイパースペクトルデータの特徴であるスペクトル曲線の連続性を考慮しない点や、実際のハイパースペクトルセンサで起こるランダムノイズやピークシフトと呼ばれる取得したスペクトルデータの波長が全体的にずれる現象に対して、十分な頑強性を持たない。従って、本研究では、Lasso 回帰のバンド選定に関する長所を生かしつつ、離散的ではなく連続的にバンド選択を行う Fused lasso 回帰を採用した。交差検証による精度検証の結果、Fused lasso 回帰は他の手法よりも高い推定精度が得られ、さらに選定されたバンドは樹木の葉に含まれる水分やリグニンの量に反応する波長帯であったことから、植物学的な見地からも本手法がバイオマス推定において有望であることが示された。

次に、分類問題としては森林タイプの分類を実施した。分類モデルとして汎用的に用いられている SVM においては分類器に全バンドが使用されるため、SSSP 下では過学習による一般性の喪失のリスクが想定される。この過学習を回避する手法としては、次元削減もしくは Feature selection が代表的である。次元削減については、SVM を実施する前に主成分分析 (PCA: Principal component analysis) を適用するモデルが幅広く用いられている。一方で、Feature selection についても近年いくつかの手法が提案されているが、本研究では正則化項を持つ Lasso 回帰をマルチクラスの分類器に拡張したスパース判別分析 (SDA: Sparse discriminant analysis) および前述の Fused lasso を同様に分類に拡張したモデルとして Fused SDA と名づけた分類モデルを作成したが、両者ともにバンド選定の後の判別分析の過程で極度に次元圧縮を行うため、結果として分類精度が SVM よりも悪化する結果となった。しかしながら、バンド選定自体は過学習や分類にとっては有効な結果であることから、SDA および Fused SDA の過程で選定されたバンドに対して SVM もしくは PCA を適用した後に SVM を実施することで、SSSP 下において他の分類モデルより高い分類精度を持つ結果が得られた。特に Fused SDA による選定バンドに PCA および SVM を実施することで最も高い精度を得ることができた。さらに、Fused SDA では連続的かつ同じ重み係数をもつバンドが選ばれることから、それらのバンドを統合することにより更なる次元削減を行った Fused Integrated SDA+PCA+SVM を提案した。精度検証の結果、本手法は Fused SDA+PCA+SVM よりも高い精度を示す結果となり、つまり本研究では SSSP 下の森林タイプ分類で最も有望な分類器ということが示された。

以上の結果を踏まえ、ハイパースペクトルセンサデータを用いた本研究における提案手法に関するアプリケーションへの展開について、「本手法を基にした森林モニタリングにおける MRV への展開」、「バンド選定がもたらすデータサイズの低減による衛星から地上へのデータ転送容量への貢献」、「森林モニタリングに必要なバンドのみを持つ小型センサの開発への寄与」という 3 つの観点から提案および考察を行った。

既存のハイパースペクトルセンサを用いた回帰問題や分類問題などの研究の多くが推定精度向上に力点を置いている一方で、実用レベルの衛星搭載もしくは国際宇宙ステーション搭載のハイパースペクトルセンサの運用開始が近づいており、ハイパースペクトルセンサの実世界への貢献が本格的に求められる段階に来ているといえる。そのため、実際の運用で想定される問題の解決について、今後、重要性が増してくると考えられる。本研究では特に森林モニタリングへの利用を想定した上で問題となる **SSSP** 下での過学習に対して、正則化を用いての影響を緩和する推定モデルの提案および検証を行うことで、より実際の運用環境で性能を発揮できる手法を明らかにしたものと言える。