

## 九州地方における衛星データ, 数値標高モデルおよび気象データを併用した森林分布図

## 作成手法の検討

田中真哉・高橋興明・西園朋広・家原敏郎・齋藤英樹(森林総合研究所)・粟屋善雄(岐阜大学)

要旨:九州地方を対象に Landsat ETM+データ, 数値標高モデル, メッシュ気候値を用いて森林/非森林の分類を行った。ETM+データの6つの波長帯における地表面反射率に加え, 4つの地形に関する変数および4つの気象に関する変数により総当たり法によってサポートベクターマシンによる分類を行った。ETM+データのみを用いた場合でも分類精度は96.05%以上と高かったものの, さらに数値標高モデルやメッシュ気候値からの変数を追加することで精度の向上が見られ, 最大で97.46%の分類精度が得られた。森林/非森林の分類精度を向上させるために, スペクトル情報に加え, とりわけ地形条件を考慮することが我が国における効果的な方法と考えられた。

キーワード: 分類, 数値標高モデル, 森林/非森林, ランドサット, メッシュ気候値

**Abstract:** Forest/non-forest classification using by support vector machine classifier was performed in Kyushu region. All variable combinations from six surface reflectance data, four topographic variables and four climatic variables obtained from Landsat ETM+ data, digital elevation model and mesh climatic data were evaluated by best-subset selection procedure. When using surface reflectance data only, overall accuracy was 96.05%. The overall accuracy increased to 97.46% when using best variable combination obtained from topographic and climatic variables with surface reflectance data. For forest/non-forest classification, consideration of topographic condition would be an effective strategy to improve classification accuracy in Japan.

**Keywords:** classification, DEM, forest/non-forest, Landsat ETM+, mesh climatic data

## I はじめに

地域レベルや国レベルなど広域の森林分布を推定する際に衛星データ, とりわけ中空間分解能の衛星データの利用は費用効果的な選択肢である。そのため, 分類精度を高めるために分類手法やデータの補正方法, 観測時期による分類精度の違いなど様々な面から検討が行われてきた。簡便に分類精度を改善するアプローチの一つとして地被クラスごとのスペクトルの季節変化の違いを利用する, すなわち多時期の衛星データを利用する方法があり, 我が国においても多時期の衛星データを利用することにより分類精度が向上することが明らかになっている(5, 6)。一方, 多時期の衛星データを用いることは増やした衛星データの数だけ雲や霧のために解析ができない範囲が増加する。そのため, 分類精度と効率性はトレードオフの関係にある。

地形や気候は植生の分布を規定することはよく知られている(3)。そのため, 景観生態学やGISの研究分野では原植生や現存植生の推定に地形や気象データがしばしば利用される(1, 2)。このようなデータの利用は森林/

非森林の分類についても同様に有効と考えられるほか, 衛星データが観測された時点における日射量やフェノロジーの差異を考慮する上でも有効であろう。そのため, これらを考慮して解析することにより森林/非森林の分類精度を向上できる可能性がある。また, 地形や気象データは日本全域において数値標高モデル(DEM)やメッシュ気候値として整備されており, 安価に取得可能である。したがって, これらのデータの利用は多時期の衛星データを利用する場合のように解析可能範囲が減少しないことも利点である。このようなことから, 本研究では森林/非森林の分類において衛星データとDEMおよびメッシュ気候値から得られる変数を併用することがどの程度有効かについて検討を行い, 以下の2点について知見を得た。(1)分類精度の高い変数の組み合わせは何か。(2)変数の数を増やした時に分類精度はどのように変化するか。なお, 本研究は農林水産省プロジェクト研究「農林水産分野における地球温暖化対策のための緩和及び適応技術の開発」により実施した。

Shinya TANAKA, Tomoaki TAKAHASHI, Tomohiro NISHIZONO, Toshiro IEHARA, Hideki SAITO (FFPRI, 1 Matsunosato, Tsukuba, Ibaraki 305-8687), and Yoshio AWAYA (Gifu University), Study on forest/non-forest classification using by Landsat ETM+, digital elevation model and mesh climatic data in Kyushu region

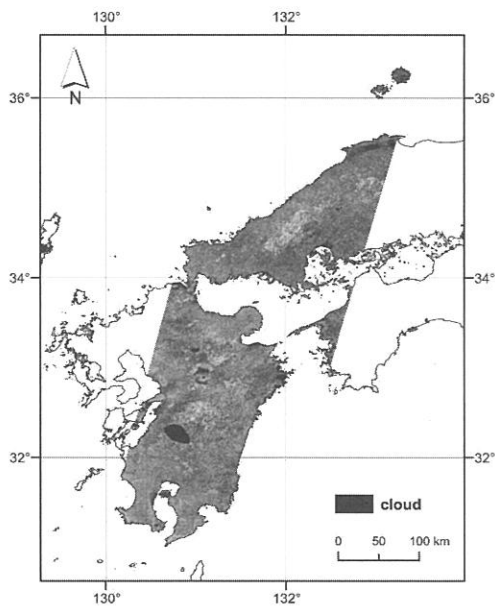


図-1. 研究対象地と解析に用いた TM データ  
Fig.1 Location of study area and TM data

## II データと方法

研究対象地は島根から鹿児島にかけての地域、WRS-2 のパス 112 ロウ 35-38 の範囲とした (図-1)。解析に用いた衛星データは米国地質調査所から配布された Landsat ETM+データ (レベル 1T, オルソ補正済み, 3次畳み込み内挿による再配列, 以後 TM データと記す) である。雲量が少ないこと, 太陽高度が高いこと, そしてスペクトル情報のみ用いた予備解析において分類精度が高かったことから 2002 年 5 月 25 日に観測された TM データを解析に用いた。TM データはバンド 6 を除外した後で 4 つのロウに分割されたデータをモザイク処理した。雲やヘイズのため解析に適さないエリアは目視により除外し, 位置合わせの誤差が 0.5 ピクセル未満になるように精密幾何補正を行った。海域は海岸線の GIS データを利用して除外した。TM データのデジタルナンバーは Dark Object Subtraction 法を用いてバースラディアンズを除いた後, 地表面反射率 (R, 以後バンド#における反射率を R#と記す) に換算し, 以後の解析で用いた。国土地理院から配布された 10m メッシュ DEM は共一次内挿法によってピクセルサイズを 30m にしたうえで, 標高 (Elev) に加え, 傾斜角 (Slope), 相対的な直達日射量 (太陽入射角の余弦;  $\cos\beta$ ), 湿性指数 (TWI) を算出した。気象庁発行のメッシュ気候値 2000 (4) からは暖かさの指数 (WI), 寒さの指数 (CI), 夏季降水量 (SRF), 冬季降水量 (WRF) を算出した。

解析で用いる森林/非森林の参照データ (教師および

検証データ) は空中写真の目視判読により作成した。位置ズレやリサンプリングによる平滑化を考慮し, GIS 上で 1990 年前後の白黒空中写真と重ねあわせ, システムチックに発生させたポイントを中心とした半径 50m の円形範囲を判読した。円形範囲の中心を基準とし, 円形範囲の 50%以上を中心と同一のクラスが占有するという条件を満たす場合にのみ, そのポイントのクラスをピュアなもののみなして以後の解析に利用した。また, 空中写真の撮影時期との時間的なずれから生じる被覆変化を考慮するために, TM データも同時に判読を行い, 宅地や農地への土地利用変化や直近に収穫や攪乱のあったポイントのデータは解析から除外した。上記の結果, 森林 1805 ポイント, 非森林 675 ポイント, 合計 2480 ポイントの参照データを得た。

分類手法は予備解析の結果で分類精度の高かったサポートベクターマシン (SVM) とした。上記により作成した変数の中から R3-R5, R7 の 4 変数は基礎変数として必ず解析で用いることとし, 残りの R1, R2,  $\cos\beta$ , Elev, Slope, TWI, WI, CI, SRF, および WRF から変数を選択する全ての組み合わせ全 1024 通りを計算した。R1 と R2 は波長が短く, ノイズの影響が大きいことから基礎変数からは除外して組み合わせを検討する対象として扱った。SVM で用いるガウシアンパラメータ  $\sigma$  および正規化パラメータ C は組み合わせごとに 2 段階の格子探索による総当たりによって総合精度が最大となる値を用いた。分類精度は 10 分割交差検証による総合精度とカッパ係数に基づいて評価した。

## III 結果

森林/非森林の分類において SVM の分類精度は極めて高く, TM データから得られる反射率データのみを用いた場合でも 96%前後の総合精度が得られた (表-1)。DEM およびメッシュ気候値から得られる変数を利用することで分類精度は向上し, 基礎変数に Slope を追加するだけで約 1%の精度向上が見られた。また, 1 変数を追加する時に上位となった  $\cos\beta$ , Slope, TWI の 3 変数を基礎変数に追加した場合の総合精度は 97.34%で全 1024 組み合わせ中の 18 位となり, 森林/非森林の分類においては DEM から得られるこの 3 変数のみでも大きな効果があった。最も高い分類精度が得られた変数の組み合わせは基礎変数に R1, R2, Elev, Slope, SRF, WRF を加えた場合でこの時の総合精度は 97.46%となった。変数の増加に対する精度の向上傾向は基礎変数に 2, 3 以上の変数を追加した時点ではほぼ頭打ちとなり, それ以上に変数を追加しても大幅な改善は見られなかった (図-2)。

IV 考察

TM データから得られた R に DEM やメッシュ気候値から得られる変数を追加することによって分類精度の向上が認められた。基礎変数に1つの変数を追加した時に上位となった変数は DEM から求められる Slope, TWI,  $\cos\beta$  であった。Slope や TWI は反射率データのみからではわからない平地と山岳地の違いを表現しており、現存植生として山岳地に森林が成立しているという経験的な知見からもこれらの変数の利用は支持できる。また、衛星データの地形補正に用いられ、経験的に有効と考えられる  $\cos\beta$  は、太陽高度の高い時期の衛星データであっても  $\cos\beta$  有効であった。DEM から求められる変数が上位に選択されていること、これらの3変数の利用により全組み合わせ中の18位と高い分類精度が得られたことから森林/非森林の分類においては特に地形を考慮することが重要と考えられた。

既報(6)では北海道東部を対象とした森林/非森林の分類において2時期の衛星データの利用により1時期と比較して平均1.2%の精度向上が見られた。本研究ではこのケースと同等以上に分類精度が向上し、既述の通り変数を追加することによるデメリットは生じない。地形条件や気象条件、とりわけ地形条件の考慮は分類精度と効率性の両面から有効な方法であると考えられた。

本研究は温暖な九州地方を対象として研究を行ったことから積雪や雪崩により成林が阻害されるなどの点について最大積雪深などの変数を利用するなどから明示的に考慮はしていない。しかし、対象地域やスケールなどの条件が異なれば、本研究で考慮しなかった変数が有効になる可能性がある。本研究で得られた DEM やメッシュ気候値を併用した衛星データの分類が有効であるという

知見はおそらくどの地域にも当てはまると考えられるが、変数の選定などについては複数の地域を対象とした比較・検討や対象とするスケールに応じた試行錯誤が必要になると思われる。

本研究で使用した TM データは米国地質調査所から提供されたものである。また、海岸線の GIS データは国土交通省国土計画局から提供された。ここに記して関係各位に謝意を表す。

引用文献

- (1) BRZEZIECKI, B., KIENAST F., and WILDI, O. (1993) A simulated map of the potential natural forest vegetation of Switzerland. *Journal of Vegetation Science* 4: 499-508.
- (2) 池口仁・竹内和彦 (1993) 数値地理情報を用いた日本列島の潜在自然植生の推定. 造園雑誌 56: 343-348.
- (3) 菊池多賀夫 (2001) 地形植生誌. 東京大学出版会, 220pp, 東京.
- (4) 気象庁 (2002) メッシュ気候値 2000. 気象業務支援センター. 東京. (CD-ROM)
- (5) 酒井徹・粟屋善雄・高橋與明・家原敏郎・松本光朗 (2009) 土地被覆分類の精度に影響を及ぼす大気地形補正と多時期データの利用. 日本リモートセンシング学会誌 29: 567-578.
- (6) 田中真哉・高橋與明・家原敏郎・齋藤英樹・粟屋善雄 (2011) 多時期の衛星データとサポートベクターマシンを用いた森林/非森林の分類精度の比較. 関東森林研究 62: 79-82.

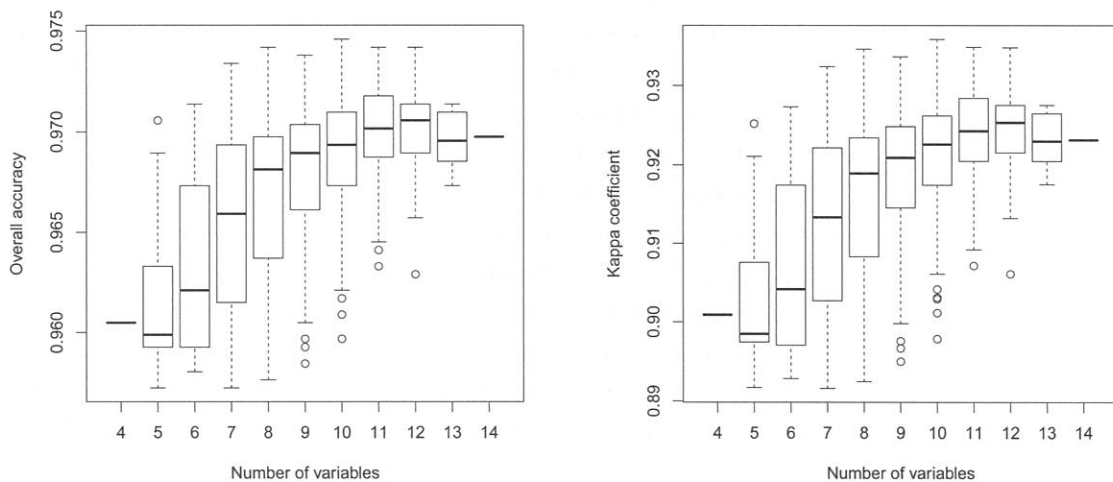


図-2. 変数の数と総合精度(a)およびカッパ係数(b)の関係

Fig.2 Relationship between number of variables and classification accuracy ((a) overall accuracy, (b) Kappa coefficient)

表-1. 総当たり法による SVM による分類精度

Table 1. Classification accuracy of SVM classifiers by best-subset selection procedure

分類精度の 順位*	変数の数	用いた変数	総合精度 (%)	カッパ係数
基礎変数				
961	4	R3, R4, R5, R7	96.05	0.9009
基礎変数と R1 および R2 の組み合わせ				
984	5	+ R1	95.93	0.8974
986	5	+ R2	95.93	0.8970
1007	6	+ R1, R2	95.89	0.8955
基礎変数とそれ以外の変数の組み合わせ**				
250	5	+ Slope	97.06	0.9252
503	5	+ TWI	96.90	0.9211
855	5	+ cos $\beta$	96.33	0.9076
147	6	+ cos $\beta$ , Slope	97.14	0.9273
155	6	+ Slope, TWI	97.14	0.9271
195	6	+ R1, Slope	97.10	0.9262
18	7	+ cos $\beta$ , Slope, TWI	97.34	0.9324
74	7	+ Elev, Slope, SRF	97.22	0.9293
101	7	+ Slope, CI, SRF	97.18	0.9284
7	8	+ R2, cos $\beta$ , Slope, CI	97.42	0.9346
13	8	+ cos $\beta$ , Slope, TWI, CI	97.34	0.9326
17	8	+ R2, Slope, CI, WRF	97.34	0.9325
9	9	+ R2, cos $\beta$ , Slope, CI, WRF	97.38	0.9336
12	9	+ R1, Elev, Slope, SRF, WRF	97.34	0.9328
14	9	+ R2, Elev, Slope, SRF, WRF	97.34	0.9326
1	10	+ R1, R2, Elev, Slope, SRF, WRF	97.46	0.9359
3	10	+ R2, cos $\beta$ , Slope, WI, CI, WRF	97.42	0.9348
6	10	+ R2, cos $\beta$ , Slope, CI, SRF, WRF	97.42	0.9347
2	11	+ R1, R2, cos $\beta$ , Slope, CI, SRF, WRF	97.42	0.9349
4	11	+ R1, R2, cos $\beta$ , Elev, Slope, SRF, WRF	97.42	0.9348
8	11	+ R1, R2, cos $\beta$ , Slope, WI, CI, WRF	97.38	0.9338
5	12	+ R1, R2, cos $\beta$ , Elev, Slope, CI, SRF, WRF	97.42	0.9348
21	12	+ R1, R2, cos $\beta$ , Slope, WI, CI, SRF, WRF	97.30	0.9317
27	12	+ R1, R2, cos $\beta$ , Elev, Slope, TWI, SRF, WRF	97.30	0.9316
129	13	+ R1, R2, cos $\beta$ , Elev, Slope, WI, CI, SRF, WRF	97.14	0.9275
184	13	+ R1, R2, Elev, Slope, TWI, WI, CI, SRF, WRF	97.10	0.9264
185	13	+ R1, R2, cos $\beta$ , Elev, Slope, TWI, WI, CI, WRF	97.10	0.9264
393	14***	+ R1, R2, cos $\beta$ , Elev, Slope, TWI, WI, CI, SRF, WRF	96.98	0.9231

\*総合精度, カッパ係数の順で順位づけし, それでも同じ場合は変数の数が少ないものを上位とした。

\*\* 紙面の都合で上位3組のみを表示した。

\*\*\* 14変数(すべてのデータを利用)の組み合わせは1組である。