

# Landsat 地表面反射率データを用いた 富士山周辺域の土地被覆分類における地形補正の影響

杉田幹夫

(2018 年 1 月 30 日受付 2018 年 2 月 19 日受理)

## Impact of topographic correction on land cover classification of Mt. Fuji and surrounding area using the Landsat surface reflectance data products

Mikio SUGITA

### 要 旨

Landsat 衛星データの高次プロダクトである地表面反射率データを用いて、富士山周辺域の土地被覆分類を行った。2000 年、2008 年および 2016 年に観測された計 8 シーンの Landsat 衛星データによる地表面反射率を用いた。グランドトゥールースデータは、正解情報と仮定した既存の土地被覆分類図からサンプル抽出した。分類モデルには、高い精度での画像分類の実績がある機械学習手法から、Random Forest と Support Vector Machine を使用した。既存の土地被覆分類図に準じて 10 個の土地被覆クラスを設定し、Landsat の 6 つの観測波長帯の地表面反射率、反射率から計算される指標、標高、傾斜など計 14 個を特徴量として設定した。解析範囲に山岳地が含まれるため、分類精度の向上を見込んで地形補正適用前後の地表面反射率をそれぞれ用いた場合の分類精度を比較評価した。その結果、地形補正適用前の地表面反射率データの分類精度は 70% 前後と得られ、地形補正適用後の反射率を使用した場合の分類精度は使用した 8 つの衛星データ観測日いずれでも低下した。地形補正適用による分類精度低下の理由については今後詳細な検証の必要がある。解析対象の地形分布や土地被覆構成などの条件によっては地表面反射率データへの地形補正適用が適さない場合があると考えられ、本研究で用いた方法以外の地形補正方法についても同様の分類精度評価を行うことが課題となる。また、地表面反射率データは一般に多時期、時系列での解析に有効と考えられていることから、複数時期のデータを利用した土地被覆分類の精度に対して地形補正などの補正処理がどのような影響を及ぼすのか評価し、最適な利用法を検討することが求められる。

キーワード：Landsat、地表面反射率、土地被覆分類、地形補正、機械学習

### I 緒言

自然環境などの定期的かつ広域のモニタリングでは、衛星画像などのリモートセンシングデータから土地被覆情報を得ることが重要である (Townshend, 1992)。土地被覆情報を地図として表現した土地被覆分類図は地表の物理形態や利用形態ごとに分類したものであり、衛星画像から二次的に得られる情報の中でも利用価値が高い。衛星画像からの土地被覆分類には様々な統計解析手法が用いられてきており、近年のリモートセンシングデータの画像分類においては機械

学習による分類が有望視されている (Gislason et al., 2006)。

一方、従来からリモートセンシングで用いられてきた Landsat 衛星のデータ配布が USGS (United States Geological Survey, アメリカ地質調査所) によって無料化 (Woodcock et al., 2008) された 2008 年以降、時系列 Landsat 画像を用いた時系列土地被覆解析が増加してきている。例えば、Hansen et al. (2013) は、毎年 Landsat 画像に対して決定木を用いることで、2000 年から 2012 年までの全世界の毎年の森

林被覆変化を推定し、世界全体では高精度で森林変化推定できたことを示している。加えて2017年には、過去のデータ群を一層専門的に比較研究できるように、地表面反射率等の量に変換した Landsat Surface Reflectance Level-2 Science Products (Schmidt et al., 2013; USGS 2017) が、過去の観測データまで遡ってオンデマンドで利用できるようになり、システム補正された Level 1 プロダクトのみならず、高次プロダクトである大気上端反射率データや地表面反射率データなどが USGS より公開されている。

このようなリモートセンシングで得られた衛星画像や反射率データを時系列データとして土地被覆モニタリングに活用する上で、特に山岳地域において、地形効果の影響が制約となる。地形効果とは、地表面の傾斜角および方位角の違いが衛星観測時の太陽の天頂角および方位角の違いと組み合わさって生じる反射率の変動である (Holben and Justice, 1980)。地形効果は同一の土地被覆・土地利用が地形に応じて異なる反射率を示す原因となり、地形の険しい場所での反射率変動増大に寄与している (Hantson and Chuvieco, 2011)。この反射率変動は分類処理において土地被覆の誤分類を引き起こす。これまでも、多時期の画像分類処理の前処理として地形効果の補正処理を実施することにより、特に地形の険しい山岳地域において分類精度が大幅に改善することが報告されている (Riano et al., 2003)。

本研究では、衛星観測データに基づく地表反射率データと既存土地被覆図によるランドトゥールースデータを利用して機械学習手法による土地被覆分類を行い、地形補正によって分類精度にどのような変化があるのか評価を行った。

## II 使用データと解析方法

### 1. 対象領域とランドトゥールースデータ

解析対象領域は、富士山頂、富士五湖、青木ヶ原樹海を含む、面積約 1600km<sup>2</sup> の範囲である。対象領域の最低標高は 167m、最高標高は富士山山頂の 3776m であり、代表的な土地被覆クラスである森林が約 8 割の地表を覆っている。

教師あり学習による土地被覆分類を行うためには、どの地点がどの土地被覆クラスに属するかを示すランドトゥールースデータが必要となる。ランドトゥールースデータとは衛星データとの比較評価用の地上参照情報であり、教師付き分類の教師データとなるほか、分類結果の検証時には検証対象として

用いる。本研究では、既存の土地被覆分類図として、JAXA 日本域高解像度土地利用土地被覆図 2016 年 9 月リリース版 ([http://www.eorc.jaxa.jp/ALOS/lulc/jlulc\\_jpn.htm](http://www.eorc.jaxa.jp/ALOS/lulc/jlulc_jpn.htm)) を使用した。これは、陸域観測技術衛星「だいち」搭載の光学センサ (AVNIR-2) が 2006 年から 2011 年の期間に観測したデータ等を用いて作成されており、総合精度 (Overall Accuracy; OA) が 78.0%、空間解像度 10m、分類クラスを 10 個として作成された日本全域の高解像度土地被覆分類図である。

土地被覆分類の入力データとして 30m 空間解像度の地表面反射率データを用いるので、ランドトゥールースデータも同じ解像度の空間代表性を有する情報源であることが望ましい。しかし本研究では指定した緯度経度に対応する地点の土地被覆クラスを 10m 解像度の日本域高解像度土地利用土地被覆図 (以下、LULC) から読み取り、これを 30m 解像度の土地被覆分類処理のランドトゥールースデータとして用いた。

### 2. 衛星データ

富士山およびその周辺域の土地被覆把握のため、2000 年、2008 年および 2016 年に観測された Landsat 衛星シリーズの観測データを使用した (図 1)。Landsat 衛星シリーズはリモートセンシングの分野において、都市、森林、海洋などの環境変化モニタリングをはじめ、様々な解析に用いられている。本研究では、USGS から提供されている Landsat Surface Reflectance Level-2 Science Products の地表面反射率データを、EarthExplorer ツール (<https://earthexplorer.usgs.gov/>) を通じて入手し用いた。このプロダクトでは、高い位置精度でオルソ幾何補正されたプロダクト L1T に大気補正を施して得られる地表面反射率データを提供している。地表面反射率は、衛星観測による 1 次情報である観測輝度値とは異なり、観測条件 (季節、太陽高度など) の影響が抑制されていて、時系列解析に適したデータである。ただし、使用した地表面反射率データは、放射伝達モデルに基づく大気補正が適用されているが、地形効果についての補正は行われていない (Schmidt et al., 2013; USGS 2017)。

また、このプロダクトには地表面反射率のほかに観測画像の各画素に対応する水、雪、雲などについての品質情報が含まれている。この画素品質情報は、衛星データから解析不能領域を除外する前処理に有



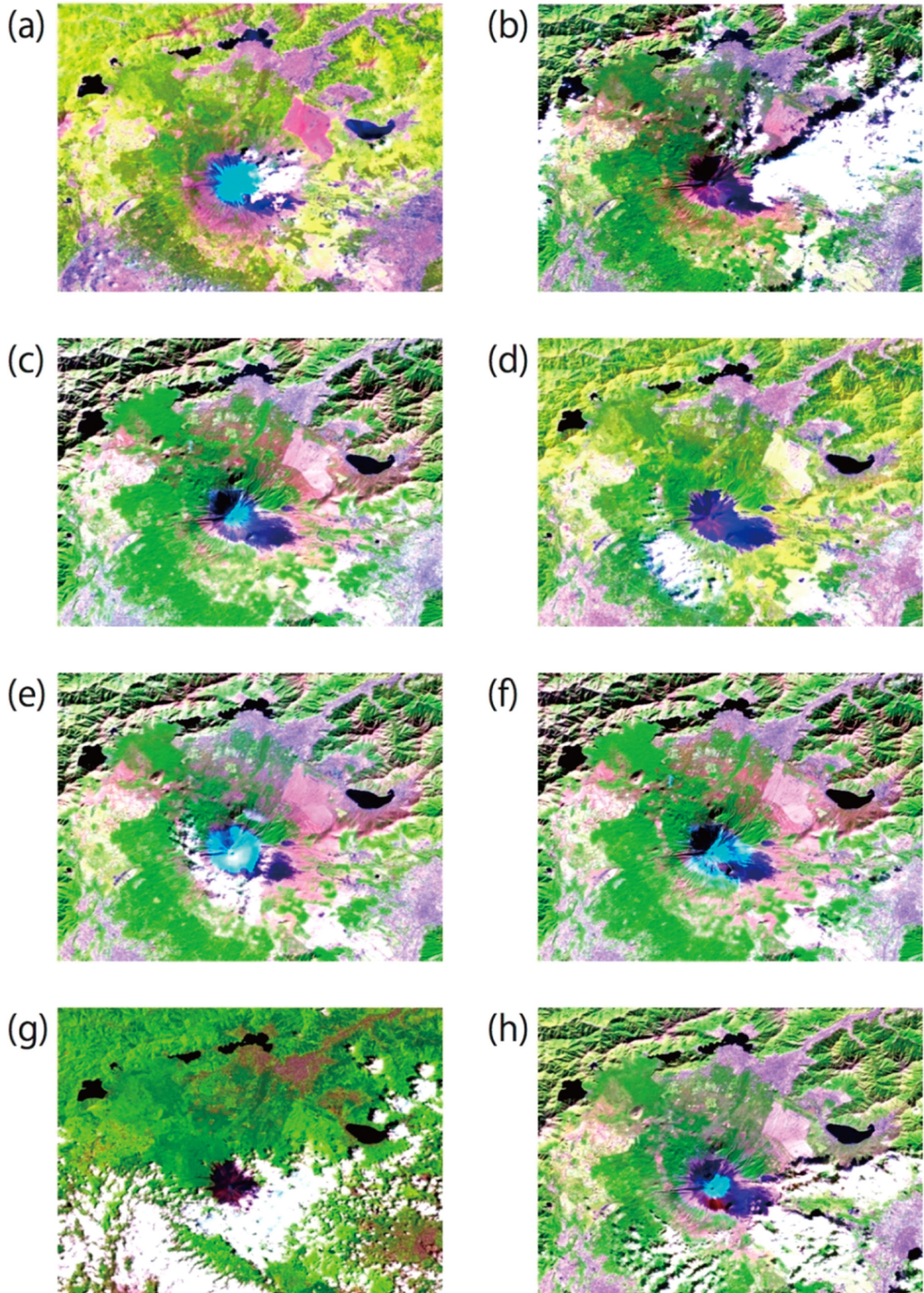


図1 使用した Landsat 衛星データ。フォールスカラー合成画像 (Landsat-5, Landsat-7: RGB 543、Landsat-8: RGB 654)。 (a) 2000.5.23 (ETM+)、 (b) 2000.11.7 (TM)、 (c) 2000.12.9 (TM)、 (d) 2008.9.10 (TM)、 (e) 2008.11.13 (TM)、 (f) 2016.1.4 (OLI)、 (g) 2016.7.14 (OLI)、 (h) 2016.11.3 (OLI)。 Landsat Surface Reflectance products courtesy of the U.S. Geological Survey Earth Resources Observation and Science Center.

効である。使用する衛星観測日によっては解析領域内に雲や雪が地表を覆っている範囲が存在するため、画素品質情報をもとに雲および雲の影と雪面を、その観測日の分類処理から除外した。

表 1 に使用衛星データの衛星、センサ、観測日を示す。ここに示した 3 つの Landsat 衛星 (5 号、7 号、8 号) はいずれも可視バンド、近赤外バンド、中間赤外バンドの観測波長帯を持ち、空間解像度が 30m の中空間分解能衛星である。

表 1 使用衛星データ

衛星	センサ	観測日
Landsat-7	ETM+	2000.5.23
Landsat-5	TM	2000.11.7
Landsat-5	TM	2000.12.9
Landsat-5	TM	2008.9.10
Landsat-5	TM	2008.11.13
Landsat-8	OLI	2016.1.4
Landsat-8	OLI	2016.7.14
Landsat-8	OLI	2016.11.3

TM: Thematic Mapper

ETM+: Enhanced Thematic Mapper Plus

OLI: Operational Land Imager

### 3. 分類手法・評価手法

土地被覆分類クラスは、LULC の分類クラス 10 個 (表 2) をそのまま使用した。解析領域の範囲内

表 2 土地被覆分類クラス

ID	土地被覆分類クラス
1	水域 (water)
2	都市 (urban)
3	水田 (rice paddy)
4	畑地 (crop)
5	草地 (grass)
6	落葉広葉樹 (DBF)
7	落葉針葉樹 (DNF)
8	常緑広葉樹 (EBF)
9	常緑針葉樹 (ENF)
10	裸地 (bareland)

で無作為に 3000 地点を抽出し、各地点について LULC の対応画素から土地被覆クラスを機械的に読み出すことで「緯度、経度、土地被覆クラス」の組を作り、ランドトゥールースデータとした。

土地被覆分類には、ある程度の数のサンプルデータ集合を対象に解析を行い、そのデータから有用な規則や判断基準などを抽出する機械学習 (Mitchell, 1997) を用いた。代表的な機械学習手法の中から Random Forest と Support Vector Machine (SVM) を取り上げた。

Random Forest は、2001 年に Leo Breiman が提案した機械学習アルゴリズムである (Breiman, 2001)。Random Forest では、ブートストラップサンプルから複数の決定木で学習し、それぞれの決定木による分類結果を統合することで最終的な分類結果とする。この手法の特徴として、学習が高速であること、外れ値やノイズに対して相対的に頑健であり、分類問題における性能が相対的に良いことが挙げられている (Breiman, 2001; 杉本ほか, 2007)。Random Forest による分類には、R 3.4.2 の randomForest パッケージを採用した。

Support Vector Machine (SVM) は 1995 年頃に Vladimir N. Vapnik が提案したパターン識別用の機械学習のアルゴリズムである (Vapnik, 1998)。与えられた学習データに対して識別境界面を設定し、クラス間のマージンを最大化する識別境界面を構成することで、未知データがどちらに分類されるかを判定する。この手法の特徴として、データの次元数が大きい場合でも分類の性能が高いことや設定するパラメータが比較的少ないことが挙げられている。SVM による分類には、R 3.4.2 の e1071 パッケージを採用した。

Landsat 衛星データによる地表面反射率 (可視 3 バンド、近赤外 1 バンド、短波長赤外 2 バンドの計 6 バンド)、反射率から計算される指標である正規化植生指標 NDVI (Normalized Difference Vegetation Index) および正規化水域指標 NDWI (Normalized Difference Water Index)、NDVI 画像の肌理を表すテクスチャ特徴量 4 種 (ASM、コントラスト、エントロピー、分散)、標高、傾斜角度による計 14 個の特徴量を、分類モデルへの入力変数とした。

土地被覆分類結果の分類精度評価のため、交差検証法で分類精度を求めた。すなわち、使用したランドトゥールースデータを N 個のグループに分け、そ



のうちの1つをテストデータ、残る  $N - 1$  個を学習データとすることで、 $N$  個のグループそれぞれをテスト事例として  $N$  回分類モデルを作成して総合精度 OA を計算し、得られた  $N$  回分の OA を平均したものを分類精度とした。

#### 4. 地形補正

Landsat 衛星などの受動的光学センサで観測されたデータは、地表の傾斜角および斜面方位と太陽高度および太陽方位との相対的な位置関係に応じて変動する。このような斜面と太陽の位置関係に起因する観測値の変動は「地形効果」と呼ばれる。多時期、時系列の衛星データを用いた解析を行う場合には地形効果を補正することが求められる。本研究で使用する地表面反射率データは上述のように地形効果の補正が行われていない。このため、地形効果補正の前後で分類精度に変化があるか評価する。

多数提案されている地形補正法の中から、本研究では C 補正法 (Teillet et al, 1982) を適用して地形補正を行った。地形補正処理には、斜面傾斜角、斜面方位角、衛星データ観測時の太陽位置情報 (天頂角、方位角) が必要となる。太陽位置情報は使用した衛星データのメタデータに記録されている値を利用した。斜面の傾斜角および方位角は標高データから計算して用いた。本研究では、JAXA が整備し公開している 30m メッシュ版全球 DSM データセット AW3D30 (Tadono et al., 2015) を標高データとして用いた。

### III 結果

使用した地表面反射率データに地形補正を行わず土地被覆分類を行った場合、衛星データ観測日ごとの分類精度を表 3 に示す。衛星データにおいて雲、雲の影および雪面である領域は、画素品質情報を用いて解析から除外しており、実際に機械学習分類に使用したグランドトゥルースの地点数は表 3 に示した通りとなった。得られた分類精度は 64.6% ~

72.8% であり、最も良い分類精度が得られたのは、2000.12.9 の観測データに対して Random Forest で分類した組み合わせであった。使用した 8 シーンのデータ全体では、Random Forest と SVM の手法間で分類精度に系統的な優劣は見られなかった。分類精度が 67% 未満となったのは、相対的に太陽高度の高い 5 月と 9 月の観測データに対する分類結果であった。

次に、反射率データに対して地形補正を適用したのち土地被覆分類を行った場合、衛星データ観測日ごとの分類精度を表 4 に示す。得られた分類精度は 62.3% ~ 68.4% であり、最も良い分類精度が得られたのは、2008.11.13 の観測データに対して SVM で分類した組み合わせであった。地形補正適用前の分類精度と比較すると、いずれの観測日、分類手法についても、一様に分類精度が低下した。また、地形補正前と同じく Random Forest と SVM の手法間で分類精度に系統的な優劣は見られなかった。

### IV 考察

本研究では、富士山周辺を解析対象領域としたケーススタディとして、Random Forest および SVM を分類モデルとして土地被覆分類を行い、地表面反射率データからどの程度の精度で土地被覆分類が可能であるのか評価した。また多くの報告で分類精度向上のために必要とされる地形補正の適用後に分類精度がどう変化するのか、評価を試みた。

まず、LULC と同一の分類クラスを設定した単時期衛星データの土地被覆分類による分類精度は 70% 前後であった。正解情報とした LULC は 10m 解像度で総合精度が 78% であることから、影響要因を精査しない状況による結果としては妥当な分類精度が得られたと考える。リモートセンシングデータからの土地被覆分類において実用のために必要とされる分類精度に関しては総合精度 OA が 85.0% 以上、各クラスの分類精度が 70.0% を基準とする報告 (Thomlinson 1999) があり、総合精度 85% 以

表 3 地形補正を行わない場合の分類精度

衛星	観測日	GT地点数	分類精度 (%)	
			Random Forest	SVM
Landsat-7	2000.5.23	2855	66.9	66.5
Landsat-5	2000.11.7	2357	68.3	68.2
Landsat-5	2000.12.9	2964	72.8	72.3
Landsat-5	2008.9.10	2888	68.8	70.5
Landsat-5	2008.11.13	2790	71.8	72.0
Landsat-8	2016.1.4	2861	71.1	71.2
Landsat-8	2016.7.14	1767	66.3	64.6
Landsat-8	2016.11.3	2552	67.1	67.7

GT: グランドトゥルース

表 4 地形補正適用後の分類精度

衛星	観測日	GT地点数	分類精度 (%)	
			Random Forest	SVM
Landsat-7	2000.5.23	2855	63.6	62.7
Landsat-5	2000.11.7	2351	63.6	66.4
Landsat-5	2000.12.9	2860	67.5	67.8
Landsat-5	2008.9.10	2873	66.0	66.6
Landsat-5	2008.11.13	2739	68.2	68.4
Landsat-8	2016.1.4	2780	67.4	67.8
Landsat-8	2016.7.14	1733	63.2	62.6
Landsat-8	2016.11.3	2473	63.8	62.3

GT: グランドトゥルース

上という基準が分類精度の目標の一つとなり得ると考える。

次に、地表面反射率データに地形補正を適用した場合の分類精度が、地形補正適用前の地表面反射率を学習データとした場合の分類精度よりも3%程度低下した。地形補正前後の衛星画像（図2）を目視比較したところでは、地形補正が一定の効果を認められる。既往研究では、特に山岳地において地形補正による精度向上が報告されている（例えば、Riano et al., 2003）。しかし、本研究の結果は既往の研究報告とは逆に、地形補正後の分類結果は補正前の分類結果より精度が劣る結果となった。まず、本研究の対象領域は、北側に山岳地があるものの、全体に対する割合は小さく、地形補正の効果が限定的であったことが考えられる。この点については、対象領域中の地形の険しい領域と平坦な領域、それぞれで地形効果による分類精度の変化を評価す

地形補正適用前



地形補正適用後



図2 地形補正適用前後の衛星画像（2000.12.9 観測 Landsat-5/TM、RGB321）

るなどして検証する必要がある。結果によっては、地形を区分してそれぞれに地形補正を適用するなどの手順を前処理とする必要があるだろう。次に、本研究では分類手法に2つの機械学習手法を使用している。学習のための特徴量の中に地表面反射率のバンド間演算で正規化された指標が2つ含まれる。この正規化指標は地形効果が軽減された数値となるので、このことが分類精度に関係している可能性がある。さらに、本研究では多数提案されている手法の一つだけを用いた結果であり、他の手法では異なる結果が得られる可能性がある。しかしながら、本研究で使用した8つの観測日のいずれでも地形補正適用後の分類精度が低下しており、地形補正処理が土地被覆分類処理に対して何らかの悪影響を及ぼした可能性が考えられる。この点は検証の必要があるだろう。

ところで、本研究では機械学習手法間の性能比較を目的としていないが、使用した衛星データとグランドトゥールースデータに対する分類精度で評価する限り Random Forest と SVM の手法間で優劣は確認できなかった。これは学習に使用した特徴量の数や、サンプルであるグランドトゥールースデータの数、2つの手法の間で差の生じにくい条件であったことが考えられる。機械学習を用いた画像分類の既往研究では、使用する特徴量が多次元の場合は Random Forest が有効で、SVM では使用する特徴量が多くなると精度が低下するとの報告（望月・村上, 2016）もあり、土地被覆分類により適したモデルの模索には、今後詳細な検討が必要である。また、Random Forest の出力結果から確認できる特徴量の重要度によると、本研究で学習に使用した14個の特徴量のうち、テクスチャ特徴量4種の重要度が他の特徴量より相対的に低くなることが確認できている。30m 分解能の衛星データの解析ではテクスチャ情報は特徴量として適切でないことが示唆される。

## V 結論

本研究では、地表面反射率データへの地形補正適用により、機械学習モデルによる分類精度が低下する結果となった。土地被覆分類では、グランドトゥールースデータ、分類クラスの数、地形補正の有無、土地被覆分類手法、教師あり分類での教師データ数など数々の要因が分類精度に影響する（Lu and Weng 2007）ので、今後、さらに詳細な検討が必



要であると考え。特に、現地調査、高解像度の衛星画像やオルソ空中写真などからデータ収集することにより、正確で信頼性の高いグラントゥールズデータを整備することが重要である。

本研究で使用した Landsat Surface Reflectance Level-2 Science Products の地表面反射率データは、富士山周辺域の解析を行った限りでは、土地被覆分類の前処理として地形補正を実施しない方が、地形補正適用ケースよりも良好な分類精度を得る結果となった。このことは、解析対象の地形分布や土地被覆構成の条件によっては、季節によらず地表面反射率データへの地形補正適用が適さない場合があると考えられ、本研究で用いた C 補正法以外の地形補正手法についても同様の分類精度評価が今後の課題となる。本研究で用いた地表面反射率データは大気補正が適用されているため、一般に多時期、時系列での解析に有効と考えられることから、複数時期のデータを利用した土地被覆分類の精度に対して地形補正などの前処理がどのような影響を及ぼすのかを評価することが今後の課題となる。

## VI 謝辞

本研究で使用した Landsat 地表面反射率データは USGS サイトから入手したものである。日本域高解像度土地利用土地被覆図および 30m メッシュ版全球 DSM データセット AW3D30 は、JAXA サイトより取得して利用した。ここに感謝の意を表す。

## VII 引用文献

- Breiman, L. (2001). Random forests, *Machine Learning*, 40, 532.
- Gislason P.O., Benediktsson J.A., Sveinsson J.R. (2006), Random forest for land cover classification, *Pattern Recognition Letters*, 27, 294-300.
- Hansen M.C., Potapov P.V., Moore R., Hancher M., Turubanova S.A., Tyukavina A., Thau D., Stehman S.V., Goetz S.J., Loveland T.R., Kommareddy A., Egorov A., Chini L., Justice C.O., Townshend J.R.G. (2013) High-resolution global maps of 21-st-century forest cover change. *Science*, 342, 850-853.
- Hantson, S., Chuvieco, E. (2011) Evaluation of different topographic correction methods for Landsat imagery. *International Journal of Applied Earth Observation and Geoinformation*, 13(5): 691-700.
- Holben, B., Justice, C. (1980) The topographic effect on spectral response from nadir-pointing. *Photogrammetric Engineering and Remote Sensing*, 46(9):1191-1200.
- Lu, D., Weng, Q. (2007) A survey of image classification methods and techniques for improving classification performance, *International Journal of Remote Sensing*, 28(5):823-870.
- Mitchell, T. (1997). *Machine Learning*, McGraw-Hill, New York. 352p.
- 望月翔太, 村上拓彦 (2016) 機械学習法を用いた SPOT5/HRG データの土地被覆分類とその精度比較. *統計数理* 64(1): 93-103.
- Riano, D., Chuvieco, E., Salas, F. J., Aguado, I. (2003) Assessment of Different Topographic Corrections in LandsatTM Data for Mapping Vegetation Types. *IEEE Transactions on Geoscience and Remote Sensing* 41: 1056-1061.
- Schmidt, G.L., Jenkerson, C.B., Masek, J., Vermote, E., Gao, F. (2013) Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description: U.S. Geological Survey Open-File Report 2013 1057, 17 p.
- 杉本知之, 下川敏雄, 後藤昌司 (2007) 樹木構造接近法と最近の発展. *計算機統計学* 18: 123-164.
- Tadono, T., Takaku, J., Tsutsui, K., Oda, F., Nagai, H. (2015) Status of ALOS World 3D (AW3D) global DSM generation. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). 3822-3825.
- Teillet, P. M., Guindon, B., Goodenough, D.G. (1982) On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, 8(2): 84-106.
- Thomlinson, J., R., Bolstad, P., V., Cohen, W., B. (1999) Coordinating Methodologies for Scaling Landcover Classifications from Site-Specific to Global: Steps toward Validating Global Map Products: *Remote Sensing of Environment* 70(1):16-28.

- Townshend, J.R.G. (1992) Land cover, International Journal of Remote Sensing, 13, 1319-1328.
- United States Geological Survey (USGS).(2017) Product Guide. Provisional Landsat 8 Surface Reflectance Code (LaSRC) Product. Version 4.2 Department of the Interior: Reston, VA, USA, 39p.
- Vapnik, V.N.(1998) Statistical learning theory, Vol. 1. Wiley New York, 1998. 768p.
- Woodcock, C.E., Allen R., Anderson M., Belward A., Bindschadler R., Cohen W., Gao F., Goward S.N., Helder D., Helmer E., Nemani R., Oreopoulos L., Schott J., Thenkabail P.S., Vermote E.F., Vogelmann J., Wulder M.A., Wynne R.(2008) Free access to landsat imagery, Science 80(320):1011.