



Title	CバンドSARデータを利用した機械学習アルゴリズムによる圃場の作物分類
Author(s)	山谷, 祐貴; 谷, 宏; 王, 秀峰 他
Citation	写真測量とリモートセンシング, 56(4), 143-148 https://doi.org/10.4287/jsprs.56.143
Issue Date	2017-09-08
Doc URL	https://hdl.handle.net/2115/71450
Rights	© 2017 一般社団法人 日本写真測量学会
Type	journal article
File Information	56_143.pdf



Cバンド SAR データを利用した 機械学習アルゴリズムによる圃場の作物分類

Crop Classification by Machine Learning Algorithm Using C-band SAR Data

山谷 祐貴*・谷 宏**・王 秀峰**・菌部 礼***
小林 伸行****・望月貫一郎*****・野田 萌*****

*Yuki YAMAYA, Hiroshi TANI, Xiufeng WANG, Rei SONOBE
Nobuyuki KOBAYASHI, Kan-ichiro MOCHIZUKI and Megumi NODA*

Abstract : This paper presents crop classification using satellite data to establish a mapping method to replace the existing ground survey. We used five scenes of C-band fully polarimetric SAR satellite Radarsat-2 data. Datasets of sigma naught and four polarimetric parameters, Freeman–Durden (FD), Van Zyl (VZ), Yamaguchi (YG), and Cloude–Pottier (CP), were calculated from each image data. We assessed the accuracy of the classification obtained by the random forest machine learning algorithm. Three results are shown. First, the highest accuracy using only one of the five datasets (0.918) was obtained by the VZ parameter dataset. Second, using three datasets, the combination of the sigma naught, VZ parameter, and CP parameter datasets obtained the highest accuracy (0.922). Third, when we used all five datasets, the accuracy (0.918) was not improved. These results confirm that crop classification using Radarsat-2 C-band data is very effective and the use of a combination of sigma naught, VZ parameters, and CP parameters obtained the highest accuracy.

1. 背景と目的

わが国では近年、国や地方自治体における業務の合理化を図るための一環として、統計情報の取得に費やすコストの削減がなされている。したがって、人件費の削減により人手不足となり、従来の人の手による現地調査が行われにくくなるため、農業地域において現状を正確に捉えることが困難となっている。また、農家においては、農地の賃貸借の進展や法人化に伴う集団化などにより、圃場の特性を熟知した管理が難しくなっている現状にある（志賀，2010）。以上の点から、統計情報を取得するためのコストの削減を進める一方

で、圃場ごとの作付状況の詳細を今後も継続的かつ長期的・広域的に把握していく手法が必要であり、これらの両立が日本の農業地域における課題の一つとなっているといえる。この課題を解決するために、従来の現地調査に代わる手法の確立は不可欠である。その中でも衛星リモートセンシング技術は、農業分野における有望な統計情報収集手段の一つとして期待されている。

統計情報を得る方法の一つとして、積み上げ方式がある。個々の圃場の作付状況を把握し、それを積み上げることで対象領域の統計値を得ることができる。これまで実際に、リモートセンシングから作付作物を分類する研究は多くなされてきた（Nordkvist *et al.*, 2012, Zhong *et al.*, 2014など）。しかし、その多くは可視光や近赤外の受動型センサである光学センサを利用したものである。光学センサは夜間の撮像が困難であり、また雲により撮像機会を逸するため、重要な時期のデータ取得が困難となる。さらに、大気や雲などによる反射光の減衰が生じやすいことから、作付作物の分類精度が大幅に低下してしまう恐れがある

* 北海道大学大学院農学院

** 北海道大学大学院農学研究院

*** 静岡大学農学部

**** 株式会社スマートリンク北海道

***** 株式会社パスコ

***** 札幌市

「写真測量とリモートセンシング」VOL. 56, NO. 4, 2017

本研究では、Cバンドの SAR データを使用し、畑作地帯における作付作物の分類を行い、その精度を評価した。分類の手法には、先行研究(Sonobe *et al.*, 2014, 山谷ほか, 2017)と同様に、パターン認識において特に高い精度が期待できる機械学習アルゴリズムを使用した。また、分類を行うにあたり、SAR データから算出される主要な数値の後方散乱係数だけでなく、各種の散乱成分分解から算出される散乱成分も作付作物の分類処理の変数として使用した。その中で、後方散乱係数や散乱成分のどのような組み合わせが作物種の分類において高い精度につながるかに着目して検討した。

表2 算出した各データセットの要素

データセット		算出した要素		
後方散乱係数	HH 偏波 (σ_{HH}^0)	HV 偏波 (σ_{HV}^0)	VV 偏波 (σ_{VV}^0)	
CP parameters	エントロピ (H)	アニソトロピ (A)	アルファ角 (α)	
FD parameters	二回散乱 (Pd)	体積散乱 (Pv)	表面散乱 (Ps)	
VZ parameters	二回散乱 (Pd)	体積散乱 (Pv)	表面散乱 (Ps)	
YG parameters	二回散乱 (Pd)	体積散乱 (Pv)	表面散乱 (Ps)	ヘリックス散乱 (Pc)

$$\sigma^0 = 20 \log_{10}(DN/A2) + 10 \log_{10}(\sin \theta_{loc}) \quad (1)$$

ここで、 θ_{loc} は衛星の入射角を示し、 DN は衛星データのピクセル値を、 $A2$ は入射角と偏波によって異なるスケーリングゲイン値を示している。なお、4 偏波の衛星データにおける HV 偏波と VH 偏波は、モノステティックでの計測の場合は値が理論上等しくなるため (Jiao *et al.*, 2014)、RADARSAT-2 データより算出される σ_{VH}^0 は使用しないこととした。したがって、残りの 3 偏波の後方散乱係数を使用した (表 2)。その他、4 種類の散乱成分分解手法を用い、それぞれから散乱成分を算出した。使用した分解手法は、3 種の散乱モデル分解法と、1 種の固有値解析法である。前者は Freeman-Durden 法 (FD) (Freeman and Durden, 1998)、Van Zyl 法 (VZ) (Van Zyl *et al.*, 2011)、Yamaguchi 法 (YG) (Yamaguchi *et al.*, 2005) であり、後者は Cloude-Pottier 法 (CP) (Cloude and Pottier, 1996) である (表 2)。以上 5 種のデータセットの計 16 要素を、5 時期すべてのデータから算出し、計 80 の変数を分類に使用した。

2.3 解析方法

本研究では、対象地域における圃場境界 GIS データを用い、圃場ポリゴン単位で作付作物の分類を行った。まず、道路や防風林などの混入および影響を防ぐため、圃場ポリゴンの内側にバッファを生成し、その領域を除いた。使用した RADARSAT-2 データのオルソ補正後のピクセルスペーシングが 8 m であることを考慮し、バッファは 10 m に設定した。次に、算出したそれぞれの変数について、1 圃場ごとに平均値を抽出した。その後、分類処理のための教師データとして、先行研究 (Hartfield *et al.*, 2013) に倣い 20% の圃場を無作為に抽出し、残りの 80% をテストデータとした。圃場数は表 3 に示す。このテストデータについて変数の値による教師付き分類を行い、その精度を評価した。

分類には機械学習アルゴリズムであるランダムフォ

表3 各圃場数の内訳

作物	圃場数		
	教師データ	テストデータ	合計
小麦	267	1,167	1,434
豆類	211	835	1,046
馬鈴薯	151	595	746
甜菜	114	444	558
牧草	137	516	653
トウモロコシ	135	504	639
合計	1,015	4,061	5,076

レスト (Breiman, 2001) を使用した。ランダムフォレストとは、変数の一部を使用した決定木を複数作成し、それらの決定木から算出された結果の多数決により、最終的な分類を決定する手法である。本研究では、決定木の数を 200 として分類を行った。なお本アルゴリズムを使用するにあたり、フリー統計ソフトである R (R Core Team, 2014) 及び、randomForest パッケージ (Liaw and Wiener, 2002) を活用した。

2.4 分類処理の流れ

本研究では、RADARSAT-2 データを使用した分類を行い、データセットを組み合わせることで高い分類精度が得られるかについて着目し、検討を行った。まず、RADARSAT-2 データから算定に用いた 5 種のデータセット (σ^0 , CP, FD, YG, VZ) のうち、いずれか 1 セットを使用した分類により精度の比較を行った。次に、散乱モデル分解法として FD, YG, VZ の 3 種類を使用したため、これらのうちいずれか 1 種類のデータセットに、後方散乱係数 σ^0 と固有値解析法である CP を加え、算出される成分の異なる 3 種類のデータセット (後方散乱係数、散乱モデル分解法、固有値解析法) を組み合わせた分類により精度の比較を行った。最後に、5 種類すべてのデータセットを使用した場合で分類結果を比較した。

分類の精度を評価するために使用した指標は、全体精度 (Overall Accuracy, OA) 及びカッパ係数 (κ)

の2種類である。その他、作物ごとの分類精度の評価に、参照データの精度を示すプロデューサーズ精度 (PA) 及び分類結果の精度を示すユーザーズ精度 (UA) の2種類を使用した。また、分類結果の差の有意性には有意差検定統計量 Z (Congalton and Green, 2008) を使用して評価した。Z 値が1.96を超えれば、5 %の有意水準において、2つの分類結果に有意な差があると判断することができる。

3. 結果と考察

3.1 いずれか1種のデータセットを使用した分類

全ての分類結果から OA と κ を図2に示す。1セットのみを使用した分類において、最も高い精度が得られたのは、VZにより算出された要素を使用した場合 (OA=0.918) であった (表4)。したがって、1セットのみを使用して分類する場合、VZの散乱成分が作付作物の分類に最も貢献することが示された。この結果は、先行研究 (山谷ほか, 2017) において同様の場

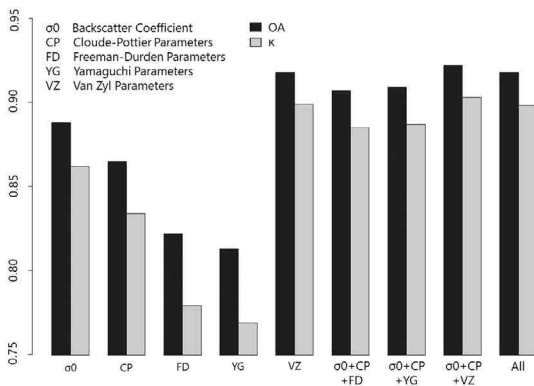


図2 各分類結果の全体精度とカッパ係数

所、時期数、方法で行った TerraSAR-X の X バンドデータによる分類 (OA=0.895) よりも少ないデータセット数で、かつ高い精度であった。このことは、C バンド衛星による作付作物分類の有効性を示唆している。特に本研究では、5月下旬や8月のデータセットも使用していたため、先行研究より高い精度を示したことも考えられる。一方、VZ 以外では0.9を上回る高い全体精度を得ることはできなかった。

3.2 算出される成分の異なる3種のデータセットを使用した分類

算出される要素の異なる3セットを組み合わせることで分類を行った結果、全体精度はそれぞれ0.9を超える高い精度となった。その中でも特に、 σ^0 , CP, VZ の組み合わせで分類を実行した場合、最も高い精度 (OA=0.922) が得られ、カッパ係数も0.9を上回った (表5)。この値は3.1で行った VZ のみを使用した分類結果よりも高い精度であり、算出される要素の異なる3セットを組み合わせることの有用性が示された。類似の研究でも、作付作物の分類において、後方散乱係数・固有値解析法・散乱モデル分解法の3グループで算出される9要素を併用することの有用性は述べられており (Xu *et al.*, 2014)、本研究でも同様な結果が得られたと言える。

しかしこの結果について、3.1の VZ のみを使用した分類との間で Z 検定を行ったところ、Z 値は0.576 (<1.96) となり、5 %の有意水準において有意な差は見られなかった。したがって本研究により、散乱モデル分解法である VZ のみを使用することで、他のデータセットを使用しなくてもある程度高い精度で分類を行えることが示された。

表4 RADARSAT-2 の VZ のみを使用した分類

		参照データ						
分類結果		豆類	甜菜	牧草	トウモロコシ	馬鈴薯	小麦	合計
	豆類	762	16	5	31	11	2	827
	甜菜	6	396	4	1	5	4	416
	牧草	8	2	479	22	8	19	538
	トウモロコシ	34	12	15	413	20	4	498
	馬鈴薯	17	15	4	30	548	7	621
	小麦	8	3	9	7	3	1,131	1,161
	合計	835	444	516	504	595	1,167	4,061
	PA	0.913	0.892	0.928	0.819	0.921	0.969	
全体精度								0.918
カッパ係数								0.899

表5 RADARSAT-2 データの σ , CP, VZ を使用した分類

		参照データ							
分類結果		豆類	甜菜	牧草	トウモロコシ	馬鈴薯	小麦	合計	UA
	豆類	762	19	4	35	8	4	832	0.916
	甜菜	7	401	5	2	3	5	423	0.948
	牧草	9	3	484	13	8	14	531	0.911
	トウモロコシ	32	5	15	415	26	3	496	0.837
	馬鈴薯	17	12	4	32	545	5	615	0.886
	小麦	8	4	4	7	5	1,136	1,164	0.976
	合計	835	444	516	504	595	1,167	4,061	
PA	0.913	0.903	0.938	0.823	0.916	0.973			
全体精度									0.922
カッパ係数									0.903

3.3 すべてのデータセット(5セット)を使用した分類

最後に、5種類すべてのデータセットを組み合わせ使用した分類を行った。その結果、全体精度は0.918となり、 σ^0 , CP, VZの3セットのみを組み合わせただけの場合と有意な差は見られなかっただけでなく ($Z=0.707<1.96$)、むしろ精度は劣っていた。このように、類似した要素を算出する散乱成分分解(FD, VZ, YG)同士の多重共線性により、精度の低下を引き起こしてしまう可能性があるため、併用する必要は低いと結論付けた。

4. 結 論

本研究では、Cバンドの4偏波SARデータに機械学習アルゴリズムのランダムフォレストを適用し、作付作物の分類を行った。その結果、5時期のデータを使用することで、圃場の作付作物を0.9以上の高い全体精度で分類できることが示された。特に、van Zyl法の散乱成分は作付作物の分類に最も有効であり、このデータセットに後方散乱成分や固有値解析法のCloude-Pottier法を加えた3種のデータセットで分類を行うことで、最も高い精度が得られた。また、van Zyl法の散乱成分のみで分類を行っても、3種のデータセットを使用した場合と有意な差はないため、van Zyl法のみで高い精度での分類を行うこともできることが示された。最後に、この3種のデータセットを使用した分類結果のマップを図3に示す。

謝 辞

本研究を遂行するにあたり使用した圃場境界GIS

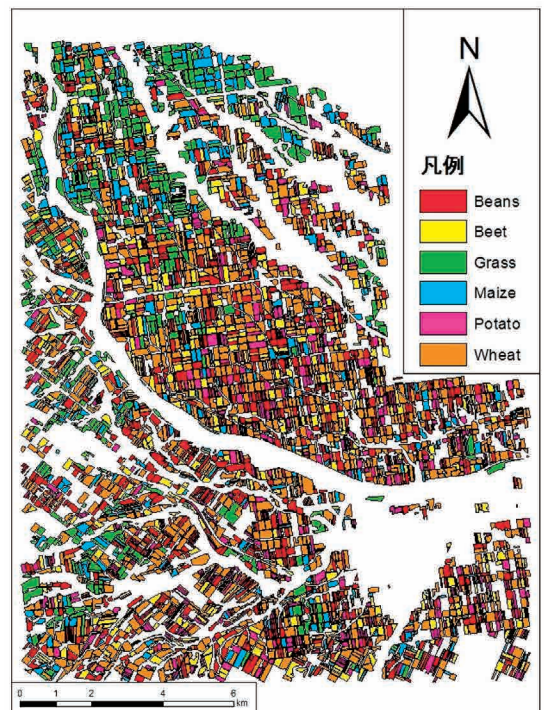


図3 RADARSAT-2 データの σ , CP, VZ を使用した分類マップ

データは、十勝農業共済組合より提供して頂いた。記して謝意を表する。

(受付日2017.7.3, 受理日2017.7.11)

引用文献

国土地理院, 2016. 基盤地図情報ダウンロードサービス, Ibaraki, Japan.
<http://fgd.gsi.go.jp/download> (accessed 7 Mar. 2017)

- 志賀弘行, 2010. 農業分野における衛星画像の利用, Hokkaido, Japan.
http://www.alic.go.jp/joho-s/joho07_000102.html
 (accessed 13 Sep. 2016)
- 山谷祐貴, 藺部 礼, 谷 宏, 王 秀峰, 小林伸行, 望月貫一郎, 2017. TerraSAR-X データを使用したランダムフォレストによる作付作物の分類. 北海道大学農学部紀要, 34(2), pp.1-11.
- Bindlish, R. and A.P. Barros, 2001. Parameterization of vegetation backscatter in radar-based, soil moisture estimation. *Remote Sensing of Environment*, 76, pp.130-137.
- Blaes, X., L. Vanhalle and P. Defourny, 2005. Efficiency of crop identification based on optical and SAR image time series. *Remote Sensing of Environment*, 96(3-4), pp.352-365.
- Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp.5-32.
- Cloude, S.R. and Pottier, E., 1996. A review of target decomposition theorems in radar polarimetry. *IEEE Transactions on Geoscience and Remote Sensing*, 34(2), pp.498-518.
- Congalton, R.G. and K. Green, 2008. Assessing the accuracy of remotely sensed data: Principles and practices. CRC Press, Florida, 137pp.
- Freeman, A. and Durden, S.L., 1998. A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 36(3), pp.963-973.
- Hartfield, K.A., Marsh, S.E., Kirk, C.D. and Carrier, Y., 2013. Contemporary and historical classification of crop types in Arizona. *International Journal of Remote Sensing*, 34(17), pp.6024-6036.
- Jiao, X.F., Kovacs, J.M., Shang, J.L., McNairn, H., Walters, D., Ma, B.L. and Geng, X.Y., 2014. Object-oriented crop mapping and monitoring using multi-temporal polarimetric RADARSAT-2 data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 96, pp.38-46.
- Liaw, A. and M. Wiener, 2002. Classification and Regression by Random Forest. *R News* 2, pp.18-22.
- McNairn, H., Ellis, J., Van Der Sanden, J.J., Hirose, T. and Brown, R.J., 2002. Providing crop information using RADARSAT-1 and satellite optical imagery. *International Journal of Remote Sensing*, 23(5), pp.851-870.
- NGA GEONet Names Server, 2013. Earth Gravitational Model 2008 (EGM2008), United States.
<http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008> (accessed 7 Mar. 2017)
- Nordkvist, K., A. Granholm, J. Holmgren, H. Olsson and M. Nilsson, 2012. Combining optical satellite data and airborne laser scanner data for vegetation classification. *Remote Sensing Letters*, 3(5), pp.393-401.
- R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
<http://www.R-project.org/> (accessed 29 Jan. 2017)
- Sonobe R., Tani H., Wang X., Kobayashi N. and Shimamura H., 2014. Random forest classification of crop type using multi-temporal TerraSAR-X dual-polarimetric data. *Remote Sensing Letters*, 5(2), pp.157-164.
- Van Zyl, J.J., Arii, M. and Kim, Y., 2011. Model-Based Decomposition of Polarimetric SAR Covariance Matrices Constrained for Nonnegative Eigenvalues. *IEEE Transactions on Geoscience and Remote Sensing*, 49(9), pp.3452-3459.
- Xu, J., Li, Z., Tian, B.S., Huang, L., Chen, Q. and Fu, S.T., 2014. Polarimetric analysis of multi-temporal RADARSAT-2 SAR images for wheat monitoring and mapping. *International Journal of Remote Sensing*, 35(10), pp.3840-3858.
- Yamaguchi, Y., Moriyama, T., Ishido, M. and Yamada, H., 2005. Four-component scattering model for polarimetric SAR image decomposition. *IEEE Transactions on Geoscience and Remote Sensing*, 43(8), pp.1699-1706.
- Zhong, L.H., Gong, P. and Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sensing of Environment*, 140, pp.1-13.