

# Improvement of Countrywide Vegetation Mapping over Japan and Comparison to Existing Maps

Ram C. Sharma<sup>1</sup>, Keitarou Hara<sup>1</sup>, Hidetake Hirayama<sup>2</sup>

<sup>1</sup>Department of Informatics, Tokyo University of Information Sciences, Chiba, Japan

<sup>2</sup>Graduate School of Informatics, Tokyo University of Information Sciences, Chiba, Japan

Email: ramcsharma01@gmail.com

**How to cite this paper:** Sharma, R.C., Hara, K. and Hirayama, H. (2018) Improvement of Countrywide Vegetation Mapping over Japan and Comparison to Existing Maps. *Advances in Remote Sensing*, 7, 163-170.

<https://doi.org/10.4236/ars.2018.73011>

**Received:** August 13, 2018

**Accepted:** September 2, 2018

**Published:** September 5, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

This paper presents an improved classification and mapping of vegetation types for all of Japan by utilizing the Moderate-resolution Imaging Spectroradiometer (MODIS) data. The Nadir BRDF-Adjusted Reflectance (MCD43A4 product) data were compared to the conventional Surface Reflectance (MOD09A1/MOY09A1 products) data for the classification of vegetation types: evergreen coniferous forest, evergreen broadleaf forest, deciduous coniferous forest, deciduous broadleaf forest, shrubs, herbaceous, arable; and non-vegetation. Very rich spectral and temporal features were prepared from MCD43A4 and MOD09A1/MOY09A1 products. Random Forests classifier was employed for the classification of vegetation types with the support of ground truth data prepared in the research. Accuracy metrics—confusion matrix, overall accuracy, and kappa coefficient calculated through 10-fold cross-validation approach—were used for quantitative comparison of MCD43A4 and MOD09A1/MOY09A1 products. The cross-validation results indicated better performance of the MCD43A4 (Overall accuracy = 0.73; Kappa coefficient = 0.69) product than conventional MOD09A1/MOY09A1 products (Overall accuracy = 0.70; Kappa coefficient = 0.66) for the classification. McNemar's test was also used to confirm a significant difference ( $p$ -value = 0.0003) between MCD43A4 and MOD09A1/MOY09A1 products. Based on these results, by utilizing the MCD43A4 features, a new vegetation map was produced for all of Japan. The newly produced map showed better accuracy than the extant, MODIS Land Cover Type product (MCD12Q1) and Global Land Cover by National Mapping Organizations (GLCNMO) product in Japan.

## Keywords

Vegetation, Classification, MODIS, Random Forests, Mapping, Japan

## 1. Introduction

Satellite data are a major source of information for classification and mapping of vegetation types. However, classification and mapping of vegetation types using satellite data at broad scale is a challenging field. Supervised classification of remotely sensed data is a commonly used technique for vegetation classification and mapping on a timely basis. A number of supervised classifiers such as decision trees, random forests, support vector machines, and neural networks have been employed for this purpose [1]-[8].

Moderate Resolution Imaging Spectroradiometer (MODIS) based Land Cover Type Product (MCD12Q1), and Global Land Cover by National Mapping Organizations (GLCNMO) are recently available global land cover maps from which vegetation types information can be obtained. The MCD12Q1 product utilizes an ensemble-based supervised classification algorithm (decision trees) complemented by training data from 1860 sites distributed across the Earth's land areas [9]. However, the MCD12Q1 product, which is basically the land cover product, has not proven effective for discriminating the vegetation types in Japan [10]. The Global Land Cover by National Mapping Organizations (GLCNMO) has also provided the global scale land cover map from which information on vegetation types can be obtained. The vegetation types in the GLCNMO product have been mapped based on the decision tree method with the support of ground truth data [11]. However, to the best of our knowledge, accuracy of the GLCNMO map for classification of vegetation types is not assessed so far in Japan.

The objectives of the research were: 1) to compare MODIS Nadir BRDF-Adjusted Reflectance (MCD43A4) product and conventional surface reflectance (MOD09A1/MOY09A1) products for vegetation classification, 2) to produce improved vegetation map at a national scale using better MODIS product, and 3) to compare newly produced map to extant moderate resolution land cover maps. This research deals with classification and mapping of seven vegetation types: evergreen coniferous forest, evergreen broadleaf forest, deciduous coniferous forest, deciduous broadleaf forest, shrubs, herbaceous, arable; and non-vegetation.

## 2. Methodology

### 2.1. Processing of Satellite Data

We processed MCD43A4 product over Japan of 2013. Six bands (Red, Near Infrared, Blue, Green, Mid Infrared, and Shortwave Infrared) were used and three spectral indices—Normalized Difference Vegetation Index [12], Superfine Water Index [13], and Urban Built-Up Index [14]—were calculated. These 8-day data were composited using monthly medians (January to December) and eleven percentiles (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100) methods. In this way, rich features (in total 207) were prepared. We also processed MOD09A1/MYD09A1 products of 2013 over Japan. The MOD09A1/MYD09A1 products provide an es-

estimate of the surface spectral reflectance as it would be measured at ground level in the absence of atmospheric scattering and absorption. Altogether, 207 features were prepared in the manner similar to the MCD43A4 product.

## 2.2. Machine Learning, Mapping and Comparison

Ground truth data, relied on previous study [10], were further increased and refined by visual interpretation of the Google Earth imagery. A total of 500 ground truth points, located all over Japan, were established for each class.

We compared the performance of MCD43A4 features to MOD09A1/MYD09A1 features for the classification of vegetation types. Random Forests classifier-based cross-validation approach was employed for the quantitative comparison of MCD43A4 and MOD09A1/MYD09A1 features with the support of ground truth data. The ground truth data were used for the purpose of supervised classification. Random forests classifier uses bootstrap aggregating (bagging) to form an ensemble of trees by searching random subspaces from the given data (features) and the best splitting of the nodes by minimizing the correlation between the trees [15]. The performance of each dataset was compared by 10-fold cross-validation approach. Machine learning was carried out only on nine folds whereas the remaining fold was used for validation [16]. Optimum number of features was retrieved from inbuilt feature importance module available in the Random Forests algorithm [17]. Hyper-parameters of the Random Forests classifier were tuned by repeated hit and trial method with reference to the validation metrics.

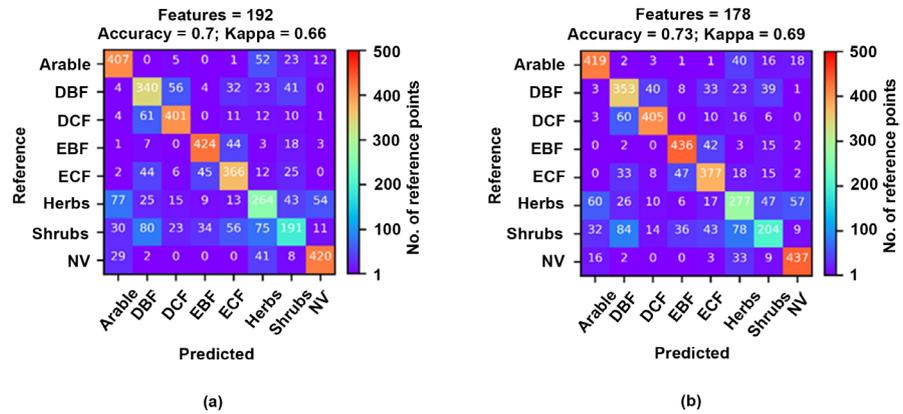
Random Forests classifier was employed for producing the seamless country-wide vegetation map. The newly produced map was compared to the extant MODIS Land Cover Type product (MCD12Q1) and Global Land Cover by National Mapping Organizations (GLCNMO) of 2013. For the comparison, the extant maps (MCD12Q1 and GLCNMO) were remapped according to the legends established in the research. Then, the accuracy metrics (overall accuracy and kappa coefficient) were calculated for all maps using the ground truth data prepared in the research.

## 3. Results and Discussion

### 3.1. MOD09A1/MYD09A1 versus MCD43A4

The confusion matrices obtained with the optimum number of features using Random Forests-based 10-fold cross-validation approach are plotted in **Figure 1**.

The MCD43A4 features indicated better cross-validation results (Overall accuracy = 0.73; Kappa coefficient = 0.69) than the MOD09A1/MYD09A1 features (Overall accuracy = 0.70; Kappa coefficient = 0.66). As seen in **Figure 1**, the MCD43A4 features have also discriminated all land cover types slightly better than the MOD09A1/MYD09A1 features. The cross-validation results obtained from different datasets (MCD43A4 versus MOD09A1/MYD09A1) are also summarized in **Table 1**.



**Figure 1.** Confusion matrices computed: (a) in the case of MOD09A1/MYD09A1 features, (b) in the case of MCD43A4 features.

**Table 1.** Comparison of MCD43A4 and MOD09A1/MYD09A1.

Products	Accuracy assessment		McNemar’s test	
	Overall accuracy	Kappa coefficient	Statistic	p-value
MOD09A1/MYD09A1	0.70	0.66	13.02	0.0003
MCD43A4	0.73	0.69		

We carried out the McNemar’s test to confirm whether the classification results are significantly different or not. McNemar’s test is a statistical test used on paired nominal data, to determine whether the row and column marginal frequencies are equal [18]. It has been effectively used for comparing accuracy assessments of image classifications [19] [20] [21]. The McNemar’s test (Table 1) showed a significant difference (p-value < 0.05) in the classifications obtained from different datasets. The cross-validation results and McNemar’s test indicated superiority of the MCD43A4 features for the classification of vegetation types over the MOD09A1/MYD09A1 features.

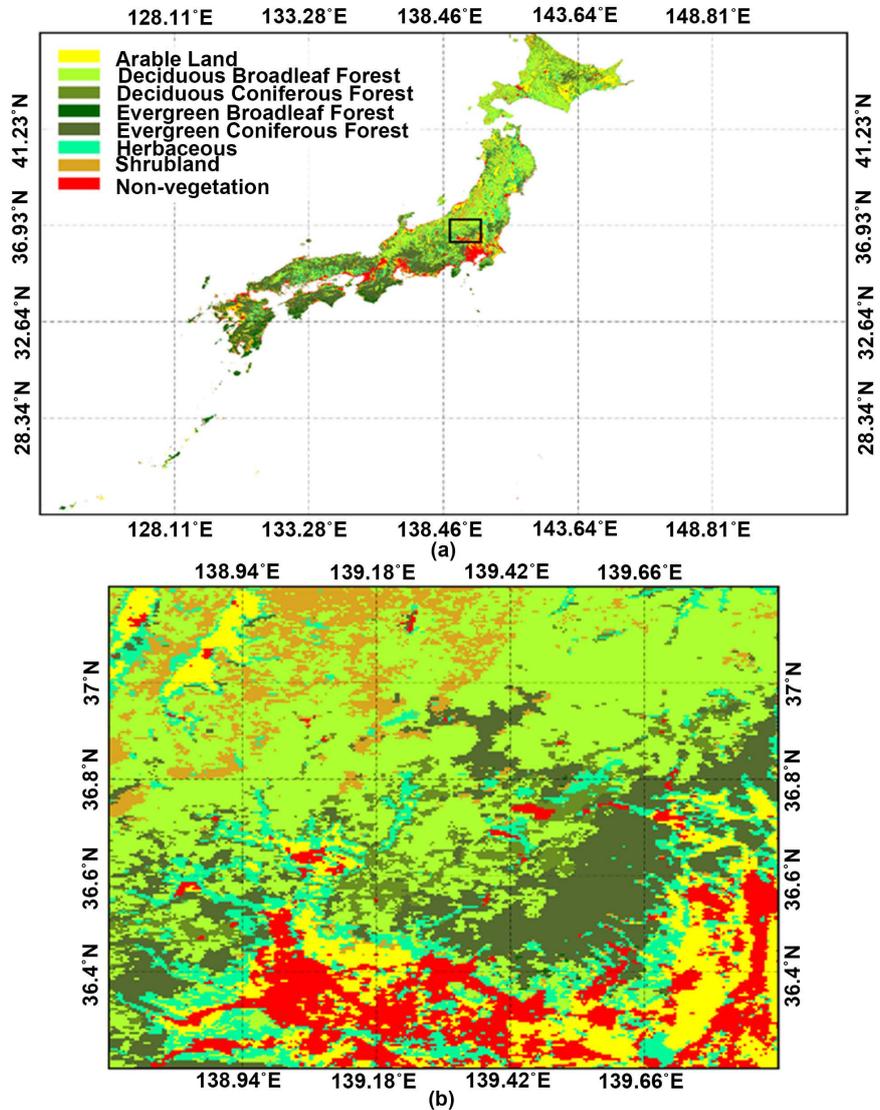
### 3.2. Vegetation Map and Comparison Results

The Random Forests model established with MCD43A4 features was used for the production of a countrywide vegetation map. The newly produced map is displayed in Figure 2.

Comparison results of the newly produced vegetation map with the extant MODIS Land Cover Type product (MCD12Q1) and Global Land Cover by National Mapping Organizations (GLCNMO) are shown in Table 2. This comparison is based on the ground truth data prepared in the research. The newly produced map using the MCD43A4 features by employing Random Forests classifier showed better accuracy than the extant maps in Japan.

## 4. Conclusion

The cross-validation results indicated better performance of the MCD43A4 product than the MOD09A1/MYD09A1 products for the classification of vegetation



**Figure 2.** Countrywide vegetation map produced through the research: (a) Display over the national territory; (b) Zoomed in over the black polygon region in (a). The national boundary is based on Global Administrative Areas database (GADM) version 3.6, May 2018.

**Table 2.** Performance of newly produced vegetation map over the extant maps.

Maps	Overall accuracy	Kappa coefficient
MCD12Q1	0.43	0.37
GLCNMO	0.39	0.31
This study	0.93	0.92

types. The MOD09A1/MOY09A1 products do not account for the dependence of observed reflectance on the view and solar angles, but instead select the best observation data from the eight-day cycle based on high-observation coverage, low-view angle, and the absence of clouds or aerosols. On the contrary, the

MCD43A4 product provides Nadir Bidirectional Reflectance Distribution Function (BRDF)-adjusted reflectance by modeling the anisotropic characteristics of the scattering using multi-date (16 days) quality observations. Therefore, in contrast to the previous study [10], MCD43A4 features were used to produce new vegetation map for all of Japan in this research. The mapping was carried out by employing Random Forests-based supervised classification approach with the support of ground truth data. The newly produced map demonstrated better accuracy than extant maps (MCD12Q1 and GLCNMO). However, moderate resolution satellite data, such as MODIS have limited ability to map fragmented vegetation. In this situation, mapping with high spatial resolution satellite data is suggested.

### Acknowledgements

This research was supported by Japan Society for the Promotion of Science (JSPS) grant-in-aid for scientific research (No. P17F17109). MODIS data used in the research were available from the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Brown de Colstoun, E. (2003) National Park Vegetation Mapping Using Multitemporal Landsat 7 Data and a Decision Tree Classifier. *Remote Sensing of Environment*, **85**, 316-327. [https://doi.org/10.1016/S0034-4257\(03\)00010-5](https://doi.org/10.1016/S0034-4257(03)00010-5)
- [2] Chan, J. C.-W. and Paelinckx, D. (2008) Evaluation of Random Forest and Ada-boost Tree-Based Ensemble Classification and Spectral Band Selection for Ecotope Mapping Using Airborne Hyperspectral Imagery. *Remote Sensing of Environment*, **112**, 2999-3011. <https://doi.org/10.1016/j.rse.2008.02.011>
- [3] Na, X., Zhang, S., Li, X., Yu, H. and Liu, C. (2010) Improved Land Cover Mapping using Random Forests Combined with Landsat Thematic Mapper Imagery and Ancillary Geographic Data. *Photogrammetric Engineering & Remote Sensing*, **76**, 833-840. <https://doi.org/10.14358/PERS.76.7.833>
- [4] Adam, E., Mutanga, O., Odindi, J. and Abdel-Rahman, E.M. (2014) Land-Use/Cover Classification in a Heterogeneous Coastal Landscape Using RapidEye Imagery: Evaluating the Performance of Random Forest and Support Vector Machines Classifiers. *International Journal of Remote Sensing*, **35**, 3440-3458. <https://doi.org/10.1080/01431161.2014.903435>
- [5] van Beijma, S., Comber, A. and Lamb, A. (2014) Random Forest Classification of Salt Marsh Vegetation Habitats Using Quad-Polarimetric Airborne SAR, Elevation and Optical RS Data. *Remote Sensing of Environment*, **149**, 118-129. <https://doi.org/10.1016/j.rse.2014.04.010>
- [6] Feng, Q., Liu, J. and Gong, J. (2015) UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis. *Remote Sensing*, **7**, 1074-1094.

- <https://doi.org/10.3390/rs70101074>
- [7] Omer, G., Mutanga, O., Abdel-Rahman, E.M. and Adam, E. (2015) Performance of Support Vector Machines and Artificial Neural Network for Mapping Endangered Tree Species Using WorldView-2 Data in Dukuduku Forest, South Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **8**, 4825-4840. <https://doi.org/10.1109/JSTARS.2015.2461136>
- [8] Bai, X., Sharma, R.C., Tateishi, R., Kondoh, A., Wuliangha, B. and Tana, G. (2017) A Detailed and High-Resolution Land Use and Land Cover Change Analysis over the Past 16 Years in the Horqin Sandy Land, Inner Mongolia. *Mathematical Problems in Engineering*, **2017**, 1-13. <https://doi.org/10.1155/2017/1316505>
- [9] Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A. and Huang, X. (2010) MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets. *Remote Sensing of Environment*, **114**, 168-182. <https://doi.org/10.1016/j.rse.2009.08.016>
- [10] Sharma, R.C., Hara, K., Hirayama, H., Harada, I., Hasegawa, D., Tomita, M., Tateishi, R., et al. (2017) Production of Multi-Features Driven Nationwide Vegetation Physiognomic Map and Comparison to MODIS Land Cover Type Product. *Advances in Remote Sensing*, **6**, 54-65. <https://doi.org/10.4236/ars.2017.61004>
- [11] Kobayashi, T., Tateishi, R., Alsaadeh, B., Sharma, R.C., Wakaizumi, T., Miyamoto, D., Phong, D.X., et al. (2017) Production of Global Land Cover Data—LCNMO 2013. *Journal of Geography and Geology*, **9**, 1. <https://doi.org/10.5539/jgg.v9n3p1>
- [12] Tucker, C.J. (1979) Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, **8**, 127-150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- [13] Sharma, R., Tateishi, R., Hara, K. and Nguyen, L. (2015) Developing Superfine Water Index (SWI) for Global Water Cover Mapping Using MODIS Data. *Remote Sensing*, **7**, 13807-13841. <https://doi.org/10.3390/rs71013807>
- [14] Sharma, R.C., Tateishi, R., Hara, K., Gharechelou, S. and Iizuka, K. (2016) Global Mapping of Urban Built-Up Areas of Year 2014 by Combining MODIS Multispectral Data with VIIRS Nighttime Light Data. *International Journal of Digital Earth*, **9**, 1004-1020. <https://doi.org/10.1080/17538947.2016.1168879>
- [15] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [16] Sharma, R.C., Hara, K. and Hirayama, H. (2017) A Machine Learning and Cross-Validation Approach for the Discrimination of Vegetation Physiognomic Types Using Satellite Based Multispectral and Multitemporal Data. *Scientifica*, **2017**, Article ID: 9806479. <https://doi.org/10.1155/2017/9806479>
- [17] Sharma, R., Hara, K. and Tateishi, R. (2017) High-Resolution Vegetation Mapping in Japan by Combining Sentinel-2 and Landsat 8 Based Multi-Temporal Datasets through Machine Learning and Cross-Validation Approach. *Land*, **6**, 50. <https://doi.org/10.3390/land6030050>
- [18] McNemar, Q. (1947) Note on the Sampling Error of the Difference between Correlated Proportions or Percentages. *Psychometrika*, **12**, 153-157. <https://doi.org/10.1007/BF02295996>
- [19] de Leeuw, J., Jia, H., Yang, L., Liu, X., Schmidt, K. and Skidmore, A.K. (2006) Comparing Accuracy Assessments to Infer Superiority of Image Classification Methods. *International Journal of Remote Sensing*, **27**, 223-232. <https://doi.org/10.1080/01431160500275762>
- [20] Petropoulos, G.P., Kalaitzidis, C. and Prasad Vadrevu, K. (2012) Support Vector

Machines and Object-Based Classification for Obtaining Land-Use/Cover Cartography from Hyperion Hyperspectral Imagery. *Computers & Geosciences*, **41**, 99-107. <https://doi.org/10.1016/j.cageo.2011.08.019>

- [21] Akar, Ö. (2018) The Rotation Forest Algorithm and Object-Based Classification Method for Land Use Mapping through UAV Images. *Geocarto International*, **33**, 538-553. <https://doi.org/10.1080/10106049.2016.1277273>