EXAMINATION OF TIME-SERIES GENERATION METHOD OF SATELLITE DATA FOR VEGETATION MAPPING IN A CLOUDY REGION

Hidetake Hirayama [1], Ram C. Sharma [1], Keitarou Hara [1]

[1] Department of Informatics, Tokyo University of Information Sciences,
4-1 Onaridai, Wakaba-ku, Chiba 265-8501, Japan

Email: hh207501@rsch.tuis.ac.jp; sharma@rsch.tuis.ac.jp; hara@rsch.tuis.ac.jp

KEY WORDS: Temporal resolution, Vegetation map, Sentinel-2, Complement method, Tadami Biosphere Reserve

ABSTRACT: With the recent developments in satellite remote sensing technology, the spatial resolution of observation data has increased, and detailed spectral information of the land surface has become available. The high resolution satellite images have been utilized frequently for producing vegetation maps in wide areas by applying machine learning techniques to their analysis. However, cloudy and outer values are major obstacles in the production of vegetation maps with higher accuracy from the satellite images. In the conventional method, cloudy pixels are removed based on a quality assessment that estimates the probability of clouds, and composite images are generated on a monthly or seasonal basis. The monthly or seasonal composite images generated by the conventional method can be useful to the discrimination of vegetation physiognomic types such as deciduous broad-leaved forests. However, the monthly or seasonal composite images might be lacking the temporal information required for distinguishing vegetation types at a community level such as beech forests. The purpose of this research is to generate time-series data that maintains high temporal resolution without losing the temporal information while removing the clouds and outer values, as a prerequisite for producing improved vegetation maps in a cloudy region. This research was conducted in Tadami Biosphere Reserve located in Fukushima prefecture in the north-east of Japan. The Tadami Biosphere Reserve is a mountainous and cloudy region with plenty of snowfall in winter. This research was implemented for 2020 by acquiring and processing Sentinel-2/Multispectral Instrument (MSI) data (95 scenes in total) from November, 2019 to February, 2021 that included the scenes of 60 days before and after the study period. In the newly proposed methodology in this research, cloudy and outlier pixels were detected and extracted by quantile analysis in a 60-day moving window period, and the time-series data were generated by applying movement statistics in a 30-day period. In comparison to the conventional method that depends on the quality of cloud flags and is susceptible to the loss of temporal information in the course of generating monthly or seasonal composite images, the newly proposed methodology in this research could generate a smooth, continuous and stable time-series data by retaining the temporal resolution.

1. INTRODUCTION

Vegetation mapping is the process of delineating geographical distribution and landscape patterns of vegetation types. Vegetation map is essential as a basic data for understanding ecosystems, their monitoring and management (Cawsey et al., 2002; Liu et al., 2019). Ecosystem management needs to be carried out in a long-term and wide area basis (Brussard et al., 1998; CBD, 2007). In recent years, vegetation mapping using satellite remote sensing technology has been actively studied. Furthermore, it is becoming highly accurate by employing machine learning and deep learning techniques (Boudewijn et al., 2020; Hirayama et al., 2018; Ramezan et al., 2021; Sharma et al., 2017). More recently, Sharma (2021) has envisioned a Genus-Physiognomy system of classifying plant communities such as Fagus deciduous forests, Quercus evergreen forests, Quercus shrubs, etc. from satellite images for vegetation mapping in wide areas.

Cloudy and outer values are major obstacles in the production of vegetation maps with higher accuracy from the satellite images (Nazarova et al., 2020). In the conventional method, cloudy pixels are removed based on a quality assessment or specific conditions that estimates the probability of clouds, and composite images are generated on a monthly or seasonal basis (Kollert et al., 2021; Zekoll et al., 2021). The monthly or seasonal composite images generated by the conventional method can be useful to the discrimination of vegetation physiognomic types such as deciduous broad-leaved forests (Griffiths et al., 2019; LUO et al., 2021; Xie et al., 2019). However, the monthly or seasonal composite images might be lacking the temporal information required for distinguishing vegetation types at a community level such as beech forests. It is necessary to generate a continuous time-series data in order to improve the accuracy of vegetation maps and to upgrade the vegetation classes from physiognomy level to community level. In recent years, a method for generating time-series data by synthesizing multiple satellites has also been devised (Tewes et al., 2015; Wu et al., 2018). However, these methods may be difficult to implement as they require some additional data and advanced computations. In this context, a useful and practical method is needed for generating time-series data for widespread application.

2. PURPOSE OF THE STUDY

The purpose of this study is to develop a time-series generation method of satellite data that can be applied for improving vegetation mapping from high-resolution satellite images.

3. METHODOLOGY

3.1 Target area and satellite data

The target area selected in the research was Tadami Biosphere Reserve (hereafter referred to as the "Tadami BR") located in Fukushima prefecture in the north-east of Japan. The Tadami BR is categorized as a humid subtropical climate in the Köppen climate classification system, and belongs to the Japan Sea area climate and is an area of high frequency rainfall and heavy snowfall (Tadami Beech Center, 2013). Almost the entire area of Tadami BR is a mountainous area of 1,000 m above sea level. Deciduous broadleaf forests (*Fagus crenata* and *Quercus mongolica*) are widely distributed in this area.

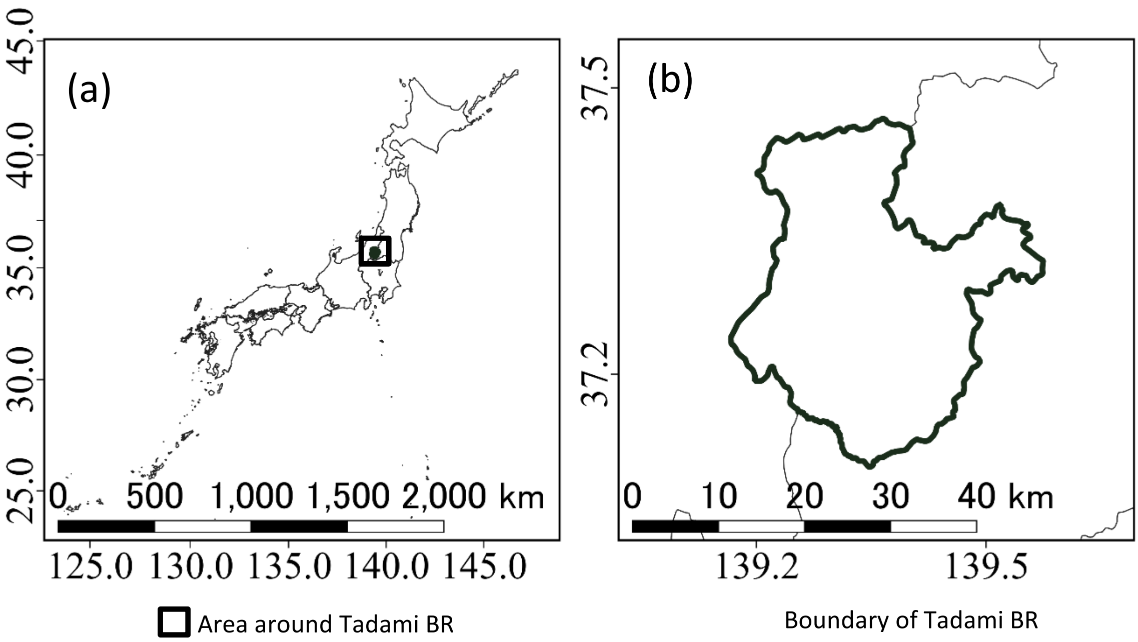

Figure 1. Location of target area. The black box shows the area surrounding the Tadami BR (a). The black line shows the boundary of Tadami BR (b).

Sentinel-2/Multispectral Instrument (MSI) images available at spatial resolution of 10-60 m were utilized in this research. We acquired and processed Sentinel-2/MSI images (95 scenes in total) from November, 2019 to February, 2021 that included the scenes of 60 days before and after the study year of 2020 (Table 1). All the collected scenes were converted to Level 2A product using sen2cor software (ESA, 2019).

Table 1. Number of collected scenes

Target period	Number of scenes
2019/11/04 ~ 2019/12/31	11
2020/01/01 ~ 2020/12/31	72
2021/01/01 ~ 2021/02/26	12
Total	95

3.2 Conventional method using quality assessment data

Quality assessment data are provided for the presence of the cloud probability (high and medium), cirrus, cloud shadows, and so on (ESA, 2018). The presence of these elements is provided as a pixel-by-pixel flag so that noisy and invalid data can be masked out on the basis of flags. We applied seven types of flags (no data, dark area pixels, cloud shadows, cloud medium probability, cloud high probability, cirrus, should never happen) and corresponding pixels were masked out as missing values.

In a conventional method, seasonal or monthly composite images are created after applying the data flags. On the other hand, in the proposed method, monthly composite images are not created in order to retain the temporal resolution of the satellite data. Finally, the Normalized Difference Vegetation Index (NDVI; Rouse Jr et al., 1974) was calculated from the time-series data.

3.3 Proposed method for generating time-series data

The proposed method consists of 3 steps, (i) calculate the NDVI, (ii) remove outliers value and (iii) complement them using the values before and after. First, NDVI values were calculated from the Sentinel-2/MSI images. Second, quantile analysis was used to remove outliers. Quantile analysis was performed based on the values of 60 days (30 days before and 30 days after the target time). For the target period, values more than 95 % and less than 5 % were defined as outliers. This outlier value removal process was executed for the whole series of data. Finally, movement statistics was used for complementing the values. The 30 days moving average procedure complemented the outlier values removed by the quantile analysis.

3.4 Evaluation method of time-series data

Time-series data were generated for eight land cover classes based on the physiognomic levels (Table 2). By calculating statistical values of the vegetation index, it was judged whether seasonal changes could be detected for each class. In addition, DBF was further separated into five community levels (Fagus DBF, Quercus DBF, Alnus DBF, Carpinus DBF and Juglans DBF) as shown in Table 3.

Distribution data of each class were prepared as GIS polygons based on field survey information and visual interpretation of the Google Earth images. Using this distribution data, corresponding pixel values of the satellite images were extracted, and 100 samples were prepared for each class. Then, the graph was plotted based on the average value of 100 samples.

Table 2. Eight land cover and physiognomic classes

Class name	Abbreviation
Deciduous broadleaf forest	DBF
Deciduous conifer forest	DCF
Evergreen conifer forest	ECF
Shrub	Shrub
Herb	Herb
Cultivated	Cultivated
Water	Water
Barren	Barren
Urban built-up	Urban built-up

Table 3. Five DBF classes at community level

Class name
Fagus DBF
Quercus DBF
Alnus DBF
Carpinus DBF
Juglans DBF

4. RESULTS AND DISCUSSION

4.1 Complement to the physiognomic levels

As a result of removing clouds with the quality assessment data flag (Figure 2), many missing values were detected from June to July, when the frequency of rainfall was high, especially in Japan. Thus, NDVI values fluctuated greatly for each scene. Figure 3 shows the trend of NDVI values for each season obtained from the complement of data by the proposed method. It shows that DBF, DCF, Herb, Shrub, and Cultivated classes had an increase in NDVI values from June to July, and a decrease from September to November, similar to the general vegetation trend. In addition, ECF class showed a constant NDVI value throughout the year compared to other physiognomic classes.

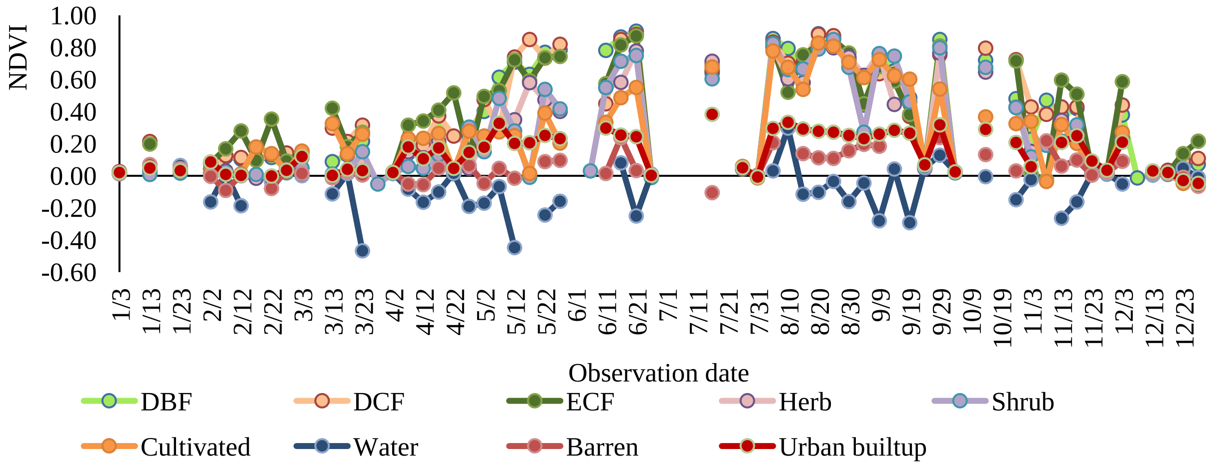

Figure 2. Time-series data for physiognomic levels from conventional method

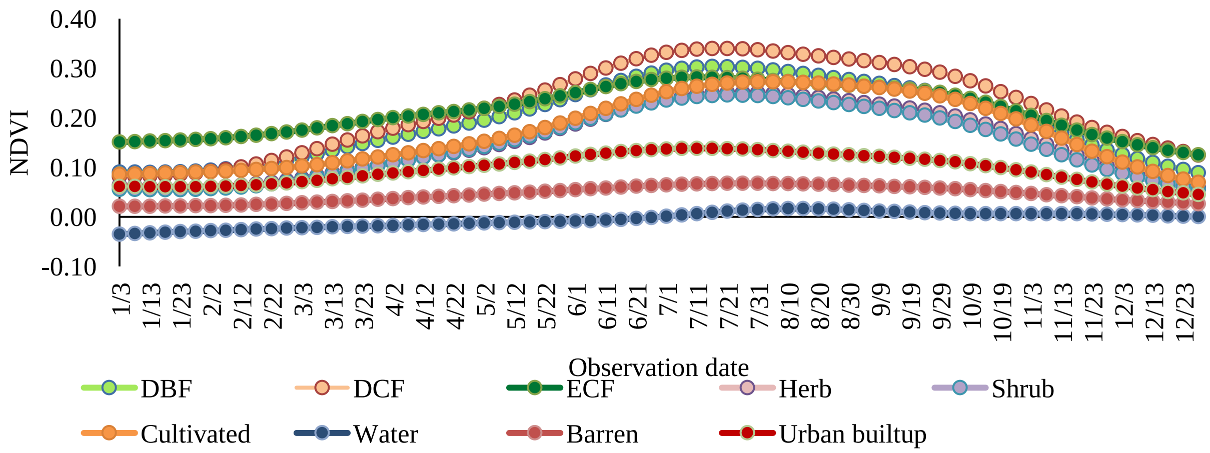

Figure 3. Time-series data for physiognomic levels from proposed method

The statistical values from 100 samples prepared for each class (Table 4) showed that the coefficient of variation (CV) of the data generated from the proposed method was overwhelmingly low. This means that there is less variability within the class in the proposed method. In addition, the maximum and minimum NDVI values of the conventional method were 1 and -1, which quite far from the general NDVI values of vegetation.

Table 4. Statistics of time-series data for physiognomic levels

Class	Conventional method				Proposed method			
	Max	Min	Mean	CV	Max	Min	Mean	CV
DBF	0.999	-0.355	0.061	1.038	0.455	-0.019	0.054	0.061
DCF	0.999	-0.428	0.184	1.043	0.430	0.033	0.172	0.067
ECF	1.000	-1.000	0.114	1.012	0.428	0.005	0.100	0.037
Herb	0.976	-0.972	0.052	0.988	0.408	-0.046	0.037	0.150
Shrub	0.999	-0.324	0.044	1.058	0.414	-0.004	0.039	0.070
Cultivated	0.999	-0.225	0.044	0.979	0.379	0.017	0.042	0.057
Water	1.000	-1.000	-0.100	-0.401	0.156	-0.152	-0.072	-0.091
Barren	1.000	-1.000	-0.023	-60.394	0.236	-0.150	-0.016	-23.021
Urban built-up	0.860	-0.501	0.042	0.872	0.367	-0.005	0.038	0.106

4.2 Complement to the community levels

Similar to section 4.1, the conventional method shows that the NDVI value fluctuates up and down; whereas the proposed method shows stable seasonal changes (Figure 4, 5). In the conventional method, it was difficult to distinguish between DBF communities as lines of the graph crossed at several points. On the other hand, according to the proposed method (Figure 5), DBF could be clustered into three groups: Fagus DBF- Quercus DBF, Carpinus DBF-Juglans DBF, and Alnus DBF. Each of them was found to have similar height and minimum NDVI values (Table 5) and NDVI shapes (Figure 5).

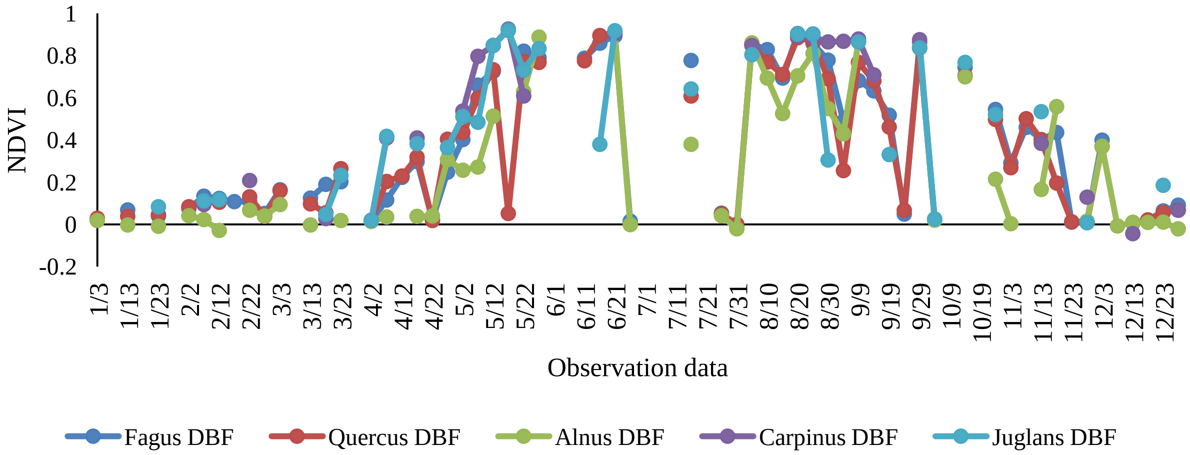

Figure 4. Time-series data for community levels from conventional method

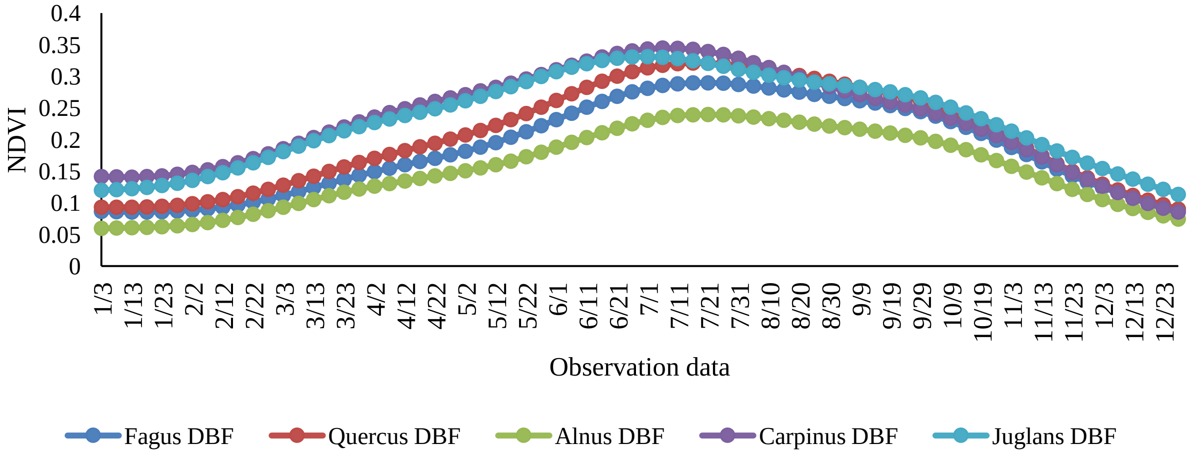

Figure 5. Time-series data for community levels from proposed method

Table 5. Statistics of time-series data for community levels

Class	Conventional method				Proposed method			
	Max	Min	Mean	CV	Max	Min	Mean	CV
Fagus DBF	0.987	-0.295	0.112	1.076	0.448	0.015	0.088	0.061
Quercus DBF	0.999	-0.312	0.071	1.038	0.455	0.014	0.056	0.103
Alnus DBF	0.932	-0.355	0.085	1.153	0.428	0.001	0.062	0.110
Carpinus DBF	0.944	-0.196	0.218	1.124	0.410	0.045	0.209	0.220
Juglans DBF	0.958	-0.289	0.184	1.106	0.396	0.047	0.170	0.112

5. CONCLUSION

This research proposed a method to generate time-series data by identifying outliers of satellite data by quantile analysis and complementing them by implementing a moving average method. It is expected that the methodology presented in the research is effective and useful for producing highly accurate vegetation maps in wide areas. Moreover, it should be an effective satellite data maintenance technique for upgrading the classification of vegetation types from physiognomic to community level.

Acknowledgement

This research was supported by the academic research grant of "Natural Capital, Tadami" in the 2020 - 2021. The research was also partially supported by JSPS KAKENHI (Grants-in-Aid for Scientific Research) JP19H04320.

References

Boudewijn, van L., Tobak, Z., Kovács, F., 2020. Machine Learning Techniques for Land Use/Land Cover Classification of Medium Resolution Optical Satellite Imagery Focusing on Temporary Inundated Areas. J. Environ. Geogr. 13, 43–52. https://doi.org/10.2478/jengeo-2020-0005

Brussard, P.F., Reed, J.M., Tracy, C.R., 1998. Ecosystem management: What is it really? Landsc. Urban Plan. 40, 9–20. https://doi.org/10.1016/S0169-2046(97)00094-7

Cawsey, E.M., Austin, M.P., Baker, B.L., 2002. Regional vegetation mapping in Australia: A case study in the practical use of statistical modelling. Biodivers. Conserv. 11, 2239–2274. https://doi.org/10.1023/A:1021350813586

CBD, 2007. CBD Decision V/6: The ecosystem approach [WWW Document]. Conv. Biol. Divers. URL https://www.cbd.int/ecosystem/principles.shtml (accessed 8.8.21).

ESA, 2019. Sen2Cor: Configuration and User Manual [WWW Document]. ESA. URL http://step.esa.int/thirdparties/sen2cor/2.8.0/docs/S2-PDGS-MPC-L2A-SUM-V2.8.pdf (accessed 1.15.20).

ESA, 2018. Sentinel-2 Level-2A Algorithm Overview [WWW Document]. ESA. URL https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm (accessed 1.15.20).

Griffiths, P., Nendel, C., Hostert, P., 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. Remote Sens. Environ. 220, 135–151. https://doi.org/10.1016/j.rse.2018.10.031

Hirayama, H., Sharma, R.C., Tomita, M., Hara, K., 2018. Evaluating multiple classifier system for the reduction of salt-and-pepper noise in the classification of very-high-resolution satellite images. Int. J. Remote Sens. 40, 2542–2557. https://doi.org/10.1080/01431161.2018.1528400

Kollert, A., Bremer, M., Löw, M., Rutzinger, M., 2021. Exploring the potential of land surface phenology and seasonal cloud free composites of one year of Sentinel-2 imagery for tree species mapping in a mountainous region. Int. J. Appl. Earth Obs. Geoinf. 94, 102208. https://doi.org/10.1016/j.jag.2020.102208

Liu, Y., Lü, Y., Fu, B., Harris, P., Wu, L., 2019. Quantifying the spatio-temporal drivers of planned vegetation restoration on ecosystem services at a regional scale. Sci. Total Environ. 650, 1029–1040. https://doi.org/10.1016/j.scitotenv.2018.09.082

LUO, C., LIU, H. jun, LU, L. ping, LIU, Z. rong, KONG, F. chang, ZHANG, X. le, 2021. Monthly composites from Sentinel-1 and Sentinel-2 images for regional major crop mapping with Google Earth Engine. J. Integr. Agric. 20, 1944–1957. https://doi.org/10.1016/S2095-3119(20)63329-9

Nazarova, T., Martin, P., Giuliani, G., 2020. Monitoring vegetation change in the presence of high cloud cover with sentinel-2 in a lowland tropical forest region in Brazil. Remote Sens. 12, 1–20. https://doi.org/10.3390/rs12111829

Ramezan, C.A., Warner, T.A., Maxwell, A.E., Price, B.S., 2021. Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data. Remote Sens. 13, 1–27. https://doi.org/10.3390/rs13030368

Rouse Jr, J., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec. Publ. 351, 309.

Sharma, R.C., 2021. Genus-Physiognomy-Ecosystem (GPE) System for Satellite-Based Classification of Plant Communities. Ecologies 2, 203–213. https://doi.org/10.3390/ecologies2020012

Sharma, R.C., Hara, K., 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 (Cairo). 2017, 1–8. https://doi.org/10.1155/2017/9806479

Tadami Beech Center, 2013. Bio sphere reserve nomination form -Tadami Biosphere Reserve- [WWW Document]. URL http://tadami-br.jp/Tadamibr-nomination_JPN.pdf (accessed 8.8.21).

Tewes, A., Thonfeld, F., Schmidt, M., Oomen, R.J., Zhu, X., Dubovyk, O., Menz, G., Schellberg, J., 2015. Using

RapidEye and MODIS data fusion to monitor vegetation dynamics in semi-arid rangelands in South Africa. Remote Sens. 7, 6510–6534. https://doi.org/10.3390/rs70606510

Wu, M., Yang, C., Song, X., Hoffmann, W.C., Huang, W., Niu, Z., Wang, C., Li, W., Yu, B., 2018. Monitoring cotton root rot by synthetic Sentinel-2 NDVI time series using improved spatial and temporal data fusion. Sci. Rep. 8, 1–12. https://doi.org/10.1038/s41598-018-20156-z

Xie, S., Liu, L., Zhang, X., Yang, J., Chen, X., Gao, Y., 2019. Automatic land-cover mapping using landsat time-series data based on google earth engine. Remote Sens. 11. https://doi.org/10.3390/rs11243023

Zekoll, V., Main-Knorn, M., Louis, J., Frantz, D., Richter, R., Pflug, B., 2021. Comparison of masking algorithms for sentinel-2 imagery. Remote Sens. 13, 1–21. https://doi.org/10.3390/rs13010137