

Crop classification using spectral indices derived from Sentinel-2A imagery

Nobuyuki Kobayashi, Hiroshi Tani, Xiufeng Wang & Rei Sonobe

To cite this article: Nobuyuki Kobayashi, Hiroshi Tani, Xiufeng Wang & Rei Sonobe (2020) Crop classification using spectral indices derived from Sentinel-2A imagery, Journal of Information and Telecommunication, 4:1, 67-90, DOI: [10.1080/24751839.2019.1694765](https://doi.org/10.1080/24751839.2019.1694765)

To link to this article: <https://doi.org/10.1080/24751839.2019.1694765>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 24 Nov 2019.



Submit your article to this journal [↗](#)



Article views: 9116



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 36 View citing articles [↗](#)



Crop classification using spectral indices derived from Sentinel-2A imagery

Nobuyuki Kobayashi^a, Hiroshi Tani ^b, Xiufeng Wang^b and Rei Sonobe ^c

^aSmart Link Hokkaido, Iwamizawa, Japan; ^bResearch Faculty of Agriculture, Hokkaido University, Sapporo, Japan; ^cFaculty of Agriculture, Shizuoka University, Shizuoka, Japan

ABSTRACT

Optical remote sensing is one of the most attractive options for generating crop cover maps because it enables computation of vegetation indices, which are useful for assessing the condition of vegetation. The Sentinel-2A Multispectral Instrument (MSI), which is a multispectral sensor with 13 bands covering the visible, near infrared and short-wave infrared (SWIR) wavelength regions, offers a vast number of vegetation indices. Spectral indices, which are combinations of spectral measurements at different wavelengths, have been used in the previous studies and they sometimes contributed to improve classification accuracies. In this study, 91 published spectral indices were calculated from the MSI data. Additionally, classification algorithms are essential for generating accurate maps and the random forests classifier is one of which possesses the five hyperparameters were applied. The improvements in classification accuracies were confirmed achieving an overall accuracy of 93.1% based on the reflectance at 4 bands and 8 spectral indices.

ARTICLE HISTORY

Received 25 April 2018



Accepted 15 November 2019

KEYWORDS

Crop; random forests; spectral indices; Sentinel-2A; variable selection

1. Introduction

Crop cover maps with accurate spatiotemporal information are required, since agricultural fields are developed and managed through a variety of social actions or policies, which can hugely impact biogeochemical and hydrologic cycles, climate, ecosystem functions, the economy and human health (Wardlow & Egbert, 2008). In order to carry out social policies as well as estimating the amount and type of crops harvested in a certain area, some local governments apply manual labour to document field properties, such as crop type and location (Sonobe, 2019). However, labour is expensive and it is necessary to develop more efficient techniques. Quite a few studies have focused on the production of thematic maps using remote sensing data (Avci & Sunar, 2015; Foody, 2002) and they reported that remote sensing provides effective support to management of agricultural fields by providing spatial and temporal information for vegetation monitoring over large areas (Gao et al., 2017). Therefore, mapping crops

CONTACT Rei Sonobe  reysnb@gmail.com  Faculty of Agriculture, Shizuoka University, 836 Ohya, Shizuoka, Japan, 422-8529

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

based on remote sensing data can be an alternative to the routine collection of annual agricultural surveys.

Optical remote sensing is one of the most attractive options for obtaining crop information at local to global scales (Chen, Son, & Chang, 2012; Lv & Liu, 2010; Seiler, Kogan, Wei, & Vinocur, 2007). Spectral indices, which are combinations of spectral measurements at different wavelengths, have been used to extract vegetation phenology or quantify biophysical parameters, such as leaf area index (Castillo, Apan, Maraseni, & Salmo, 2017; Prasad, Sarkar, Singh, & Kafatos, 2007; Sonobe, Miura, Sano, & Horie, 2018; Xie et al., 2015). The Normalised Difference Vegetation Index is popular and has been widely applied to evaluate biomass (Thimsuwan, Eiumnoh, Honda, & Tingsanchali, 2000) and drought forecasting (Martyniak, Dabrowska-Zielinska, Szymezyk, & Gruszczynska, 2007). Some indices have been developed for mineral mapping in mountains (Rowan & Mars, 2003) or environmental change detection (Escadafal, Belghith, & Ben Moussa, 1994). Furthermore, the use of spectral indices improves classification accuracy (Sonobe et al., 2017a). Indeed, many earth observation satellites like Moderate Resolution Imaging Spectro-radiometer (MODIS) or Landsat offer great opportunities to carry out research on global environmental issues and some spectral indices have been widely used for land cover classification (Bansal, Katyal, & Garg, 2017; Siachalou, Mallinis, & Tsakiri-Strati, 2017; Song, Woodcock, Seto, Lenney, & Macomber, 2001). Besides them, the great potential of hyperspectral-derived spectral vegetation indices for species classification has been shown and (Liu, Coops, Aven, & Pang, 2017) reported that classification accuracy would be improved when a combination of structural and spectral information was used. However, most published spectral indices have not been considered for crop classification. Especially, few indices are based on shortwave infrared bands (SWIR) even though SWIR are mostly influenced by plant constituents including pigments, leaf water content and biochemical (Asner, 1998; Pena, Liao, & Brenning, 2017).

The Sentinel-2A satellite was launched on 23 June 2015 and collects multispectral data including 13 bands covering the visible and SWIR wavelength regions, and includes three atmospheric bands (Band 1, Band 9 and Band 10). Furthermore, Sentinel-2B was launched on 7 March 2017 and provides great opportunities for monitoring agricultural fields. Even excluding the three atmospheric bands, various spectral indices can be extracted. Therefore, in this research we calculated 91 spectral indices from Sentinel-2A data and their importance in crop type identification was evaluated.

Geographic object-based image analysis (GEOBIA) has been paid most attention now and that's useful to clarify the borders of fields. However, very fine resolutions of less than 1 m are required to make good use of remote sensing data in GEOBIA (Baker, Warner, Conley, & Mcneil, 2013). Besides, (Hartfield, Marsh, Kirk, & Carriere, 2013) reported the classification accuracy based on pixel-based classification using classification and regression tree (CART) was superior to that based on object-oriented classifiers using Landsat data. The random forests (RF) is an improved algorithm based on an ensemble learning technique that builds multiple CART (Breiman, 2001). Indeed, RF has been extremely successful as a general-purpose classification and regression method (Biau & Scornet, 2016). RF fits many classification trees to training datasets, and then combines the predictions from all the trees for a final decision. Therefore, RF was applied in this study. Furthermore, it is easy to evaluate the variable importance (VIMP), which is calculated when a classification model is developed and shows how much worse the prediction would be

without that variable when RF is applied (Ishwaran, 2007). The original RF has two hyperparameters: the number of trees (*ntree*) and the number of variables used to split the nodes (*mtry*). However, tuning of the hyperparameters related to nodes sometimes improves classification accuracies, therefore, RF as modified by Ishwaran (Ishwaran & Kogalur, 2007; Ishwaran, Kogalur, Blackstone, & Lauer, 2008) was applied in this study.

Although the grid search has widely been used for tuning hyperparameters of machine learning algorithms, it could be a poor choice for configuring algorithms for new data sets (Bergstra & Bengio, 2012). On the other hand, Bayesian optimization is useful to obtain effective combinations of hyperparameters and has been applied in previous studies (Sonobe et al., 2017b), therefore, Bayesian optimization was applied in this study. The main objectives of this study were (1) to evaluate the potential of metrics extracted from Sentinel-2A data for crop classification and (2) to identify which spectral indices are effective for identification of crop types.

2. Materials and methods

2.1. Study area

The study area was in Hokkaido, Japan (142°55'12" to 143°05'51"E, 42°52'48" to 43°02'42"N) at an elevation between 50 and 230 m. The location has warm summers and cold winters, with an average annual temperature of 6°C and an annual precipitation of 920 mm. A total of 4719 crop fields were in the study area, and the field size ranged from 0.1 to 17.5 ha with a mean of 2.6 ha, a skewness of 2.2 and a standard deviation of 1.7 ha. The dominant crops were beans (981 fields, *Vigna angularis* and *Glycine max*), beetroots (569 fields, *Beta vulgaris*), grass (640 fields, *Phleum pratense* and *Dactylis glomerata*), maize (317 fields, *Zea mays*), potatoes (783 fields, *Solanum tuberosum*) and winter wheat (1429 fields, *Triticum aestivum*). Rice is cultivated mainly in the central parts in Hokkaido (i.e. Ishikari, Sorachi and Kamikawa). In the northern and eastern parts of Hokkaido (i.e. Kitami, Abashiri and Tokachi) rice cultivation could not be established because of very cool climate. The crop species chosen in this study have mainly been cultivated in Tokachi (Figure 1).

The 2016 crop map was provided by Tokachi Nosai (located in Obihiro City, Hokkaido. Nosai provides agricultural insurance, which helps stabilize farmers suffering from damage caused by natural disasters and contributes to the growth of Japanese agriculture). The 2016 crop map was received as a polygon shape file that included attribute data, such as crop types from April to August. In 2016, seeding of winter wheat was conducted after September, therefore the crop cover did not change between the date recorded and August 2016.

2.2. Satellite data

Sentinel-2A is designed to measure reflectance of blue, green, red and near-infrared-1 bands at 10 m resolution; red edge 1–3, near-infrared-2, and short-wave infrared 1 and 2 at 20 m; and 3 atmospheric bands (Band 1, Band 9 and Band 10) at 60 m. However, the three atmospheric bands were not used during this study because they are mainly dedicated to atmospheric corrections and cloud screening (Drusch et al., 2012). In

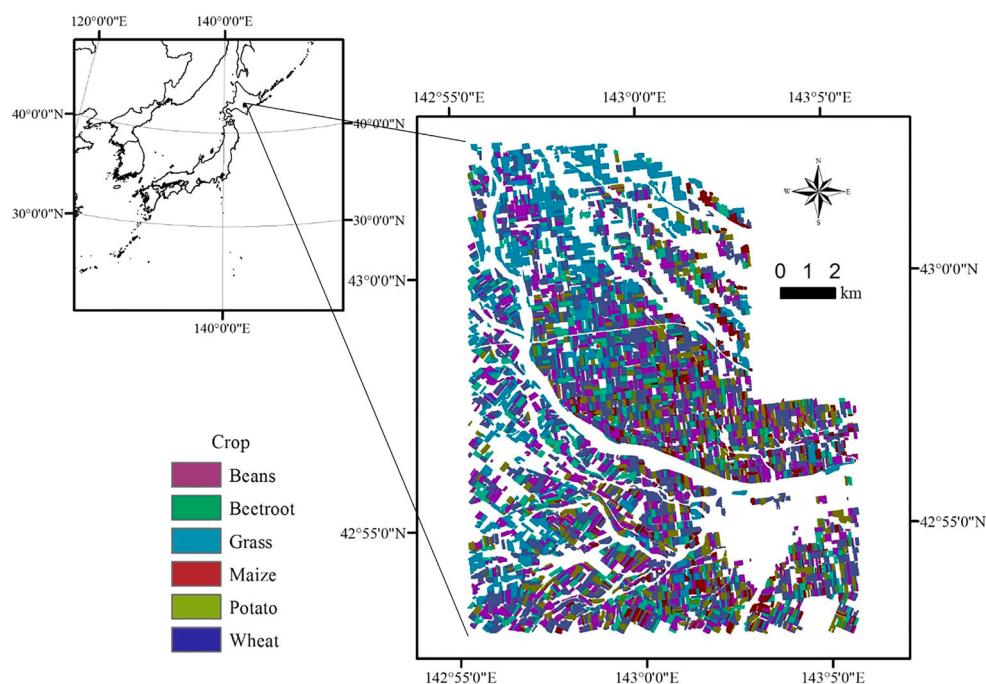


Figure 1. Crop cover map generated from reference data.

previous studies, the satellite data showed great potential for agricultural purposes including crop and grass chlorophyll and nitrogen content (Clevers & Gitelson, 2013), vegetation state in grasslands and savannah (Hill, 2013), and potato crop yield (Al-Gaadi et al., 2016).

The seven observations were conducted by Sentinel-2A from May to September 2016 for the whole site. However, only one image date (11 August) during the growing season was used due to better quality for classification purposes; the other images were covered with cloud. On 11 August, the mean temperature was 20.7°C, the duration of sunshine was 8.2 h, the atmospheric pressure was 999.6 hPa and the relative humidity was 66%.

In this study, the Level 1C data acquired on 11 August 2016 were downloaded from EarthExplorer (<https://earthexplorer.usgs.gov/>). All bands are converted to 10 m resolution using Sentinel-2 Toolbox version 5.0.4. Then, average reflectance values of each band were calculated for the fields using field polygons to compensate for spatial variability and to avoid problems related to uncertainty in georeferencing. Beside them, 91 published spectral indices were evaluated for their potential for crop classification (Table 1). The wavelength difference (D), such as Green Difference Vegetation Index (GDVI) (Tucker, 1979a), and the normalized differences (ND), such as Green Normalized Difference Vegetation Index (GNDVI) (Gitelson et al., 1996), Normalized difference near infrared/shortwave infrared normalized Burn Ratio (NBR) (Key & Benson, 2006) and Normalized Difference Infrared Index (NDII) (Zarco-Tejada et al., 2001), are the most common forms of spectral indices. Gitelson et al. (2009, 2001a) proposed inverse reflectance differences, such as Anthocyanin Reflectance Index (ARI) or Carotenoid Reflectance Index (CRI). However, most indices are based on reflectance in more than three bands. Some indices, such as Soil Adjusted

Table 1. Spectral indices used in this study.

Index	Full name	Reference	Formula
AFRI1.6	Aerosol free vegetation index 1.6	Karnieli, Kaufman, Remer, and Wald (2001)	$\frac{\text{Band}8a - 0.66*\text{Band}11}{\text{Band}8a + 0.66*\text{Band}11}$
AFRI2.1	Aerosol free vegetation index 2.1		$\frac{\text{Band}8a - 0.5*\text{Band}12}{\text{Band}8a + 0.5*\text{Band}12}$
ARI	Anthocyanin reflectance index	Gitelson, Chivkunova, and Merzlyak (2009)	$\frac{1}{\text{Band}3} - \frac{1}{\text{Band}5}$
ARVI	Atmospherically resistant vegetation index	Kaufman and Tanre (1992)	$\frac{(\text{Band}8 - (\text{Band}4 - \gamma(\text{Band}2 - \text{Band}4)))}{(\text{Band}8 + (\text{Band}4 - \gamma(\text{Band}2 - \text{Band}4)))}$ <p>The γ is a weighting function that depends on aerosol type. In this study, a value of 1 for γ.</p>
ARVI2	Atmospherically resistant vegetation index 2		$-0.18 + 1.17*\left(\frac{\text{Band}8 - \text{Band}4}{\text{Band}8 + \text{Band}4}\right)$
ATSAVI	Adjusted transformed soil-adjusted vegetation index	Baret and Guyot (1991)	$\frac{a*(\text{Band}8 - a*\text{Band}4 - b)}{\text{Band}8 + \text{Band}4 - ab + X(1 + a^2)}$ <p>$a = 1.22, b = 0.03, X = 0.08$</p>
AVI	Ashburn vegetation index	Ashburn (1978)	$2*\text{Band}8a - \text{Band}4$
BNDVI	Blue-normalized difference vegetation index	Yang, Everitt, and Bradford (2007)	$(\text{Band}8 - \text{Band}2)/(\text{Band}8 + \text{Band}2)$
BRI	Browning reflectance index	Chivkunova et al. (2001)	$\frac{1/\text{Band}3 - 1/\text{Band}5}{\text{Band}6}$
BWDRVI	Blue-wide dynamic range vegetation index	Hancock and Dougherty (2007)	$\frac{0.1*\text{Band}7 - \text{Band}2}{0.1*\text{Band}7 + \text{Band}2}$
CARI	Chlorophyll absorption ratio index	Kim, Daughtry, Chappelle, McMurtrey, and Walthall (1994)	$\frac{\text{Band}5*\sqrt{(a*\text{Band}4 + \text{Band}4 + b)^2}}{\text{Band}4}*(a^2 + 1)^{0.5}$ <p>$a = (\text{Band}5 - \text{Band}3)/150$</p> <p>$b = \text{Band}3*550*a$</p>

(Continued)

Table 1. Continued.

Index	Full name	Reference	Formula
CCCI	Canopy chlorophyll content index	El-Shikha et al. (2008)	$\frac{(\text{Band8} - \text{Band5})}{(\text{Band8} + \text{Band5})}$
CRI550	Carotenoid reflectance index 550	Gitelson, Merzlyak, and Chivkunova (2001a)	$\frac{(\text{Band8} - \text{Band4})}{(\text{Band8} + \text{Band4})}$
CRI700	Carotenoid reflectance index 700		$\frac{1}{\text{Band2}} - \frac{1}{\text{Band3}}$
CVI	Chlorophyll vegetation index	Hunt, Daughtry, Eitel, and Long (2011)	$\frac{1}{\text{Band2}} - \frac{1}{\text{Band5}}$
Datt1	Vegetation index proposed by Datt 1	Datt (1999)	$\frac{\text{Band8} * \text{Band4}}{(\text{Band3})^2}$
Datt2	Vegetation index proposed by Datt 2	Datt (1998)	$\frac{\text{Band8} - \text{Band5}}{\text{Band8} - \text{Band4}}$
Datt3	Vegetation index proposed by Datt 3		$\frac{\text{Band4}}{\text{Band3} * \text{Band5}}$
Datt3	Vegetation index proposed by Datt 3		$\frac{\text{Band8a}}{\text{Band3} * \text{Band5}}$
DVI	Differenced vegetation index	Richardson and Wiegand (1978)	$2.4 * \text{Band8} - \text{Band4}$
EPICar	Eucalyptus pigment index for carotenoid	Datt (1998)	$0.0049 * \left(\frac{\text{Band4}}{\text{Band3} * \text{Band5}} \right)^{0.7488}$
EPICla	Eucalyptus pigment index for chlorophyll a		$0.0161 * \left(\frac{\text{Band4}}{\text{Band3} * \text{Band5}} \right)^{0.7784}$
EPIClab	Eucalyptus pigment index for chlorophyll a + b		$0.0236 * \left(\frac{\text{Band4}}{\text{Band3} * \text{Band5}} \right)^{0.7954}$
EPIClb	Eucalyptus pigment index for chlorophyll b		$0.0337 * \left(\frac{\text{Band4}}{\text{Band3}} \right)^{1.8695}$
EVI	Enhanced vegetation index	Huete et al. (2002)	$2.5 * \frac{\text{Band8} - \text{Band4}}{\text{Band8} + 6 * \text{Band4} - 7.5 * \text{Band2} + 1}$

EVI2	Enhanced vegetation index 2	Miura, Yoshioka, Fujiwara, and Yamamoto (2008)	$2.4 * \frac{\text{Band8} - \text{Band4}}{\text{Band8} + \text{Band4} + 1}$
EVI2.2	Enhanced vegetation index 2.2	Jiang, Huete, Didan, and Miura (2008)	$2.5 * \frac{\text{Band8} - \text{Band4}}{\text{Band8} + 2.4 * \text{Band4} + 1}$
Fe2	Ferric iron index	Rowan and Mars (2003)	$\frac{\text{Band12}}{\text{Band8}} + \frac{\text{Band3}}{\text{Band4}}$
GARI	Green atmospherically resistant vegetation index	Gitelson, Kaufman, and Merzlyak (1996)	$\frac{\text{Band8} - (\text{Band3} - (\text{Band2} - \text{Band4}))}{\text{Band8} - (\text{Band3} + (\text{Band2} - \text{Band4}))}$
GBNDVI	Green-Blue normalized difference vegetation index	Wang, Huang, and Chen (2010)	$\frac{\text{Band8} - (\text{Band3} + \text{Band2})}{\text{Band8} + (\text{Band3} + \text{Band2})}$
GDVI	Green difference vegetation index	Tucker (1979a)	$\text{Band8} - \text{Band3}$
GEMI	Global environment monitoring index	Pinty and Verstraete (1992)	$\frac{n * (1 - 0.25 * n) - \text{Band4} - 0.125}{1 - \text{Band4}}$
			$n = \frac{2 * \text{Band5}^2 - \text{Band4}^2 + 1.5 * \text{Band5} + 0.5 * \text{Band4}}{\text{Band5} + \text{Band4} + 0.5}$
GLI	Green leaf index	Gobron, Pinty, Verstraete, and Widlowski (2000)	$\frac{2 * \text{Band3} - \text{Band5} - \text{Band2}}{2 * \text{Band3} + \text{Band5} + \text{Band2}}$
GNDVI	Green normalized difference vegetation index	Gitelson et al. (1996)	$\frac{\text{Band8} - \text{Band3}}{\text{Band8} + \text{Band3}}$
GNDVI2	Green normalized difference vegetation index 2		$\frac{\text{Band7} - \text{Band3}}{\text{Band7} + \text{Band3}}$
GOSAVI	Green optimized soil adjusted vegetation index	Rondeaux, Steven, and Baret (1996)	$\frac{\text{Band8} - \text{Band3}}{\text{Band8} + \text{Band3} + 0.16}$
GRNDVI	Green-Red normalized difference vegetation index	Wang, Huang, Tang, and Wang (2007)	$\frac{\text{Band8} - (\text{Band3} + \text{Band5})}{\text{Band8} + (\text{Band3} + \text{Band5})}$
GSAVI	Global vegetation moisture index	Tian, Zhu, and Cao (2005)	$1.5 * \frac{\text{Band8} - \text{Band3}}{(\text{Band8} + \text{Band3} + 0.5)}$

(Continued)

Table 1. Continued.

Index	Full name	Reference	Formula
GVMI	Global vegetation moisture index	Glenn, Nagler, and Huete (2010)	$\frac{(\text{Band8} + 0.1) - (\text{Band12} + 0.02)}{(\text{Band8} + 0.1) + (\text{Band12} + 0.02)}$
Hue	Hue	Escadafal et al. (1994)	$\text{atan}\left(\frac{2*\text{Band5} - \text{Band3} - \text{Band2}}{30.5} * (\text{Band3} - \text{Band2})\right)$
I	Intensity		$(\text{Band5} + \text{Band3} + \text{Band})/30.5$
IF	IF proposed by Escadafal		$2 * \frac{\text{Band5} + \text{Band3} - \text{Band2}}{\text{Band3} - \text{Band2}}$
IPVI	Infrared percentage vegetation index	Crippen (1990)	$\frac{\text{Band8}}{\frac{\text{Band8} + \text{Band5}}{2}} \left(\frac{\text{Band5} - \text{Band3}}{\text{Band5} + \text{Band5}} + 1 \right)$
IR550	Inverse reflectance at 550 nm	Gitelson, Merzlyak, Zur, Stark, and Gritz (2001b)	$1/\text{Band3}$
IR700	Inverse reflectance at 700 nm		$1/\text{Band5}$
LCI	Leaf chlorophyll index	Datt (1999)	$\frac{\text{Band8} - \text{Band5}}{\text{Band8} + \text{Band4}}$
Maccioni	Vegetation index proposed by Maccioni	Maccioni, Agati, and Mazzinghi (2001)	$\frac{\text{Band7} - \text{Band5}}{\text{Band7} - \text{Band4}}$
MCARI	Modified chlorophyll absorption in reflectance index	Daughtry, Walthall, Kim, De Colstoun, and McMurtrey (2000)	$((\text{Band5} - \text{Band4}) - 0.2 * (\text{Band5} - \text{Band3})) * \frac{\text{Band5}}{\text{Band4}}$
MCARI/ MTVI2	MCARI/MTVI2	Eitel, Long, Gessler, and Smith (2007)	MCARI/MTVI2
MCARI/ OSAVI	MCARI/OSAVI	Haboudane, Miller, Pattey, Zarco-Tejada, and Strachan (2004)	MCARI/OSAVI
MCARI1	Modified chlorophyll absorption in reflectance index 1		$1.2 * (2.5 * (\text{Band8} - \text{Band4}) - 1.3 * (\text{Band8} - \text{Band3}))$
MCARI2	Modified chlorophyll absorption in reflectance index 2		$1.5 * \frac{2.5 * (\text{Band8} - \text{Band4}) - 1.3 * (\text{Band8} - \text{Band3})}{\sqrt{(2 * \text{Band8} + 1)^2 - (6 * \text{Band8} - 5 * \sqrt{\text{Band4}}) - 0.5}}$
MGVI	Green vegetation index proposed by Misra	Misra, Wheeler, and Oliver (1977)	$-0.386 * \text{Band3} - 0.530 * \text{Band4} + 0.535 * \text{Band6} + 0.532 * \text{Band8}$

mNDVI	Modified normalized difference vegetation index	Main et al. (2011)	$\frac{\text{Band8} - \text{Band4}}{\text{Band8} + \text{Band4} - 2*\text{Band2}}$
MNSI	Non-such index proposed by Misra	Misra et al. (1977)	$0.404*\text{Band3} + 0.039*\text{Band4} - 0.505*\text{Band6} + 0.762*\text{Band8}$
MSAVI	Modified soil adjusted vegetation index	Qi, Chehbouni, Huete, Kerr, and Sorooshian (1994)	$\frac{2*\text{Band8} + 1 - \sqrt{(2*\text{Band8} + 1)^2 - 8*(\text{Band8} - \text{Band5})}}{2}$
MSAVI2	Modified soil adjusted vegetation index 2		$\frac{2*\text{Band8} + 1 - \sqrt{(2*\text{Band8} + 1)^2 - 8*(\text{Band8} - \text{Band4})}}{2}$
MSBI	Soil brightness index proposed by Misra	Misra et al. (1977)	$0.406*\text{Band3} + 0.600*\text{Band4} + 0.645*\text{Band6} + 0.243*\text{Band8}$
mSR	Modified simple ratio 670/800	Sims and Gamon (2002)	$\frac{\text{Band8} - \text{Band2}}{\text{Band4} - \text{Band2}}$
MSR670	Modified Simple Ratio 670/800	Chen (1996)	$\frac{\frac{\text{Band8}}{\text{Band4}} - 1}{\sqrt{\frac{\text{Band8}}{\text{Band4}} + 1}}$
MSRNir/Red	Modified simple ratio Nir/Red	Chen and Cihlar (1996)	$\frac{\frac{\text{Band8}}{\text{Band5}} - 1}{\sqrt{\frac{\text{Band8}}{\text{Band5}} + 1}}$
MTVI2	Modified triangular vegetation index 2	Haboudane et al. (2004)	$1.5*\frac{1.2*(\text{Band8} - \text{Band3}) - 2.5*(\text{Band4} - \text{Band3})}{\sqrt{(2*\text{Band8} + 1)^2 - (6*\text{Band8} - 5*\sqrt{\text{Band4}}) - 0.5}}$
MYVI	Misra Yellow Vegetation Index	Misra et al. (1977)	$0.723*\text{Band3} - 0.597*\text{Band4} + 0.206*\text{Band6} - 0.278*\text{Band8}$
NBR	Normalized difference Nir/Swir normalized burn ratio	Key and Benson (2006)	$\frac{\text{Band8} - \text{Band12}}{\text{Band8} + \text{Band12}}$
ND774/677	Normalized difference 774/677	Zarco-Tejada, Miller, Noland, Mohammed, and Sampson (2001)	$\frac{\text{Band7} - \text{Band4}}{\text{Band7} + \text{Band4}}$
NDII	Normalized difference infrared index	Hardisky, Klemas, and Smart (1983)	$\frac{\text{Band8} - \text{Band11}}{\text{Band8} + \text{Band11}}$

(Continued)

Table 1. Continued.

Index	Full name	Reference	Formula
NDRE	Nnormalized difference Red-edge	Barnes et al. (2000)	$\frac{\text{Band7} - \text{Band5}}{\text{Band7} + \text{Band5}}$
NDSI	Normalized difference salinity index	Dehni and Lounis (2012)	$\frac{\text{Band11} - \text{Band12}}{\text{Band11} + \text{Band12}}$
NDVI	Normalized difference vegetation index	Tucker (1979b)	$\frac{\text{Band8} - \text{Band4}}{\text{Band8} + \text{Band4}}$
NDVI2	Normalized difference vegetation index 2	Tucker (1979a)	$\frac{\text{Band12} - \text{Band8}}{\text{Band12} + \text{Band8}}$
NGRDI	Normalized green red difference index	Zarco-Tejada et al. (2001)	$\frac{\text{Band3} - \text{Band5}}{\text{Band3} + \text{Band5}}$
OSAVI	Optimized soil adjusted vegetation index	Rondeaux et al. (1996); Haboudane, Miller, Tremblay, Zarco-Tejada, and Dextraze (2002)	$1.16 * \frac{\text{Band8} - \text{Band4}}{\text{Band8} + \text{Band4} + 0.16}$
PNDVI	Pan normalized difference vegetation index	Wang et al. (2007)	$\frac{\text{Band8} - (\text{Band3} + \text{Band5} + \text{Band2})}{\text{Band8} + (\text{Band3} + \text{Band5} + \text{Band2})}$
PPR	Plant pigment ratio	Metternicht (2003)	$\frac{\text{Band3} - \text{Band2}}{\text{Band3} + \text{Band2}}$
PVR	Photosynthetic vigour ratio		$\frac{\text{Band3} - \text{Band4}}{\text{Band3} + \text{Band4}}$
RBNDVI	Red-Blue normalized difference vegetation index	Wang et al. (2007)	$\frac{\text{Band8} - (\text{Band4} + \text{Band2})}{\text{Band8} + (\text{Band4} + \text{Band2})}$
RDVI	Renormalized difference vegetation index	Broge and Leblanc (2001)	$\frac{\text{Band8} - \text{Band4}}{\sqrt{\text{Band8} + \text{Band4}}}$
REIP	Red-edge inflection point	Herrmann et al. (2011)	$700 + 40 * \left(\frac{\left(\frac{\text{Band4} + \text{Band7}}{2} \right) - \text{Band5}}{\text{Band6} - \text{Band5}} \right)$

Rre	Reflectance at the inflexion point	Clevers et al. (2002)	$\frac{\text{Band4} + \text{Band7}}{2}$
SAVI	Soil adjusted vegetation index	Huete (1988)	$1.5 * \frac{\text{Band8} - \text{Band4}}{\text{Band8} + \text{Band4} + 0.5}$
SBL	Soil background line	Richardson and Wiegand (1978)	$\text{Band8} - 2.4 * \text{Band4}$
SIPI	Structure intensive pigment index	Penuelas, Baret, and Filella (1995)	$\frac{\text{Band8} - \text{Band2}}{\text{Band8} - \text{Band4}}$
SIWSI	Shortwave infrared water stress index	Fensholt and Sandholt (2003)	$\frac{\text{Band8a} - \text{Band11}}{\text{Band8a} + \text{Band11}}$
SLAVI	Specific leaf area vegetation index	Lymburner, Beggs, and Jacobson (2000)	$\frac{\text{Band8}}{\text{Band4} + \text{Band12}}$
TCARI	Transformed chlorophyll absorption Ratio	Daughtry et al. (2000)	$3 * \left((\text{Band5} - \text{Band4}) - 0.2 * (\text{Band5} - \text{Band3}) \left(\frac{\text{Band5}}{\text{Band4}} \right) \right)$
TCARI/ OSAVI	TCARI/OSAVI	Haboudane et al. (2002)	TCARI/OSAVI
TCI	Triangular chlorophyll index	Haboudane, Tremblay, Miller, and Vigneault (2008); Hunt et al. (2011)	$1.2 * (\text{Band5} - \text{Band3}) - 1.5 * (\text{Band4} - \text{Band3}) * \sqrt{\frac{\text{Band5}}{\text{Band4}}}$
TVI	Transformed vegetation index	Rouse, Haas, Schell, and Deering (1974)	$\sqrt{\text{NDVI} + 0.5}$
VARI700	Visible atmospherically resistant index 700	Gitelson et al. (2001b)	$\frac{\text{Band5} - 1.7 * \text{Band4} + 0.7 * \text{Band2}}{\text{Band5} + 2.3 * \text{Band4} - 1.3 * \text{Band2}}$
VARIgreen	Visible atmospherically resistant index green		$\frac{\text{Band3} - \text{Band4}}{\text{Band3} + \text{Band4} - \text{Band2}}$
VI700	Vegetation index 700	Gitelson, Kaufman, Stark, and Rundquist (2002)	$\frac{\text{Band5} - \text{Band4}}{\text{Band5} + \text{Band4}}$
WDRVI	Wide dynamic range vegetation index	Gitelson (2004)	$\frac{0.1 * \text{Band8} - \text{Band4}}{0.1 * \text{Band8} + \text{Band4}}$

Vegetation Index (SAVI) (Huete, 1988) and Chlorophyll Absorption Ratio Index (CARI) (Kim et al., 1994), are based on a transformation technique to minimize the influence of soil brightness or the effects of non-photosynthetic materials, and they have very complex forms. Moreover, some integrated indices, such as the ratio between Transformed Chlorophyll Absorption Ratio (TCARI) and Optimized Soil Adjusted Vegetation Index (OSAVI) (Haboudane et al., 2002), were also considered in this study.

2.3. Classification algorithm

We used a stratified random sampling approach (Foody, 2009) to select the fields used for developing (50%), hyperparameter tuning (25%) and evaluation (25%) of the classification models to prevent overfitting of classification models (Hastie, Tibshirani, & Friedman, 2009). Crop classifications were conducted using the following three different datasets; Case 1: reflectance data, Case 2: spectral indices data and Case 3: a combination of reflectance and spectral indices in order to evaluate if adding vegetation indices improved the classification accuracy.

RF has demonstrated its ability to yield accurate land cover maps (Rodriguez-Galiano, Chica-Olmo, Abarca-Hernandez, Atkinson, & Jeganathan, 2012; Sonobe, Tani, Wang, Kobayashi, & Shimamura, 2014). Furthermore, RF carries out classifications with a shorter computing time; although the number and quality of variables has a great influence on computational time (Pelletier, Valero, Inglada, Champion, & Dedieu, 2016). In RF, multiple decision trees are generated based on random bootstrapped samples from the training data (Breiman, 2001) and nodes are split using the best split variable from a group of randomly selected variables (Liaw & Wiener, 2002). Finally, the output is determined by a majority vote from the trees. The original RF has two hyperparameters including the number of trees (*ntree*) and the number of variables used to split the nodes (*mtry*). However, the best split for a node can increase classification accuracy (Ishwaran et al., 2008; Ishwaran & Kogalur, 2007; Sonobe et al., 2017a). Thus, three hyperparameters related to nodes were also considered: the minimum number of unique cases in a terminal node (*nodesize*), the maximum depth of tree growth (*nodedepth*) and the number of random splits (*nsplit*).

Although the grid search is one of the most common strategies to optimize the hyperparameters of machine learning algorithms (Puertas, Brenning, & Meza, 2013), it could be a poor choice for configuring algorithms for new data sets (Bergstra & Bengio, 2012). Therefore, Tuning these hyperparameters was conducted using the Gaussian process, Bayesian optimization, which is superior to the grid search strategy for configuring algorithms from new data sets (Bergstra & Bengio, 2012). In Bayesian optimization, the black-box function (such as machine learning algorithms), is sampled from a Gaussian process and its posterior distribution is maintained. To select the hyperparameters of the next experiment, the expected improvement was used in this study.

2.4. Accuracy assessment

The kappa statistic (Cohen, 1960) had been used as an accuracy measure, however, it has fundamental conceptual flaws, such as being undefined even for simple cases, or having no useful interpretation (Pontius & Millones, 2011). Therefore. The classification accuracies

were assessed using quantity disagreement (QD) and allocation disagreement (AD), which provide an effective summary of the cross-tabulation matrix (Pontius & Millones, 2011). Each cell of the confusion matrix was calculated using the following equation;

$$P_{ij} = W_i \frac{n_{ij}}{n_{i+}} \quad (1)$$

where P_{ij} denotes the proportion of fields in class i according to the classification results, and class j according to the reference data (i.e. the polygon shape file provided by Tokachi Nosai). W_i is the fields classified as class i , n_{ij} is the number of fields classified as class i according to the classification results, and class j according to the reference data. n_{i+} is the row totals. AD and QD were calculated as follows;

$$AD_i = 2 \min(p_{i+}, p_{+i}) - 2p_{ii} \quad (2)$$

$$AD = \frac{1}{2} \sum_{i=1}^{N_c} AD_i \quad (3)$$

$$QD_i = |p_{i+} - p_{+i}| \quad (4)$$

$$QD = \frac{1}{2} \sum_{i=1}^{N_c} QD_i \quad (5)$$

where N_c is the number of classes, p_{i+} and p_{+i} denote the row and column totals, AD_i is the allocation disagreement for class i and QD_i is the quantity disagreement for class i . QD is defined as the difference between the correct data and the classified data based on a mismatch of class proportions, while AD is the difference between the classified data and the correct data due to incorrect spatial allocations of fields in the classification. The sum of QD and AD indicates total disagreement (Pontius & Millones, 2011). Overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA), which are commonly used for accuracy assessments, were also used.

$$OA = \sum_{i=1}^N p_{ii} / N \quad (6)$$

$$PA = p_{ii} / R_i \quad (7)$$

$$UA = p_{ii} / C_i \quad (8)$$

where N is the number of fields, R_i and C_i denote the total number of classes i in the reference data and the total number of the classification results, and p_{ii} represents the number of correctly classified fields. We analysed whether there were significant differences between the two classification results based on McNemar's test (McNemar, 1947). In this test, a chi-square value greater than 3.84 indicates a significant difference between the two classification results at a 95% significance level.

The data-based sensitivity analysis (DSA), which is a simple method performs a pure black box use of the fitted models by querying the fitted models with sensitivity samples and recording their responses, were used to evaluate the sensitivity of the classification models. Although DSA is similar to a computationally efficient one-dimensional

sensitivity analysis (Kewley, Embrechts, & Breneman, 2000), this method uses several training samples instead of a baseline vector (Cortez & Embrechts, 2013).

3. Results

3.1. Reflectance obtained from Sentinel-2A

The mean reflectance acquired from Sentinel-2A on 11 August is shown in Figure 2. For wheat, the reflectance at Bands 6–8a was lower than for other crops, while the reflectance at Bands 2–5 was higher. In this period, wheat was in the late grain-filling stage, but it had not yet been harvested. Wheat had very different reflectance features the other five living crops. On the other hand, the date corresponded to the late growing season of beetroot and grass and the early ripening period of beans; therefore, the reflectance features of vegetation that are characterized by low reflectance in the visible domain and high reflectance in the infra-red were quite clear in these crops. Regarding potatoes, a chemical treatment to kill potato vines, which facilitates harvest, was used in the study area. As the result, a relatively high reflectance at Bands 2–5 was confirmed in the potatoes as the features of senile plants. For maize, the only C_4 plant, the reflectance at Bands 2–5 was the lowest and at Bands 6–8a was similar to that of potato.

3.2. Accuracy assessment

The best combinations of the five parameters (*ntree*, *mtry*, *nodesize*, *nodedepth*, *nsplit*) were (465, 6, 2, 13, 20) for Case 1; (567, 49, 3, 40, 4) for Case 2 and (82, 45, 4, 43, 7) for Case 3. Although it is usually recommended to set *nodesize* = 1 (Liaw & Wiener, 2002), this value was not selected in any case. The value of *nodedepth* became higher with the increase in the number of variables, while the lowest *nsplit* was confirmed for Case 2 and the highest one was confirmed for Case 1.

The results of accuracy assessments and McNemar's test are given in Tables 2 and 3, respectively. The best OA was confirmed for Case 2 and the worst was confirmed for Case 1 and their differences in classification were significant ($p < 0.05$), even though the statistical differences between these 3 cases are minor.

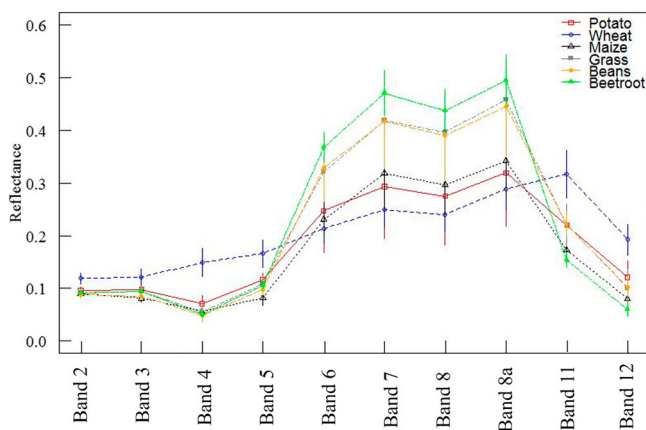


Figure 2. Mean reflectance spectra and standard deviations acquired from Sentinel-2A MSI.

Table 2. Results of accuracy assessment.

	Case 1 (Reflectance)	Case 2 (Spectral indices)	Case 3 (Reflectance+ Spectral indices)
PA			
Beans	0.848	0.874	0.874
Beetroot	0.984	0.943	0.959
Grassland	0.886	0.913	0.913
Maize	0.811	0.887	0.868
Potato	0.854	0.892	0.862
Wheat	0.978	0.990	0.981
UA			
Beans	0.931	0.928	0.908
Beetroot	0.992	1.000	1.000
Grassland	0.795	0.872	0.845
Maize	0.896	0.940	0.939
Potato	0.822	0.841	0.848
Wheat	0.975	0.972	0.975
OA	0.912	0.930	0.924
AD	6.367	4.802	5.950
QD	2.401	2.192	1.670

Note: PA: producer's accuracy; UA: user's accuracy; OA: overall accuracy; QD: quantity disagreement; AD: allocation disagreement.

Table 3. Results of McNemar's test.

	Case 1 (Reflectance)	Case 2 (Spectral indices)	Case 3 (Reflectance+ Spectral indices)
Case 1 (Reflectance)	X	16.90	14.22
Case 2 (Spectral indices)		X	10.62
Case 3 (Reflectance+ Spectral indices)			X

Although Case 1 showed the highest PA of beans, misclassification sometimes occurred with grass or potato. On the other hand, misclassifications between beans and grass were reduced using spectral indices (Cases 2 and 3). However, a few beetroot fields were misclassified as grass or potato for Cases 2 and 3, while all fields classified as beetroot were correct. Although Case 2 had the worst UA of wheat, the use of spectral indices improved classification accuracies for other crops.

3.3. Importance for identifying each crop types

The importance of each index for identifying each crop type was evaluated using the value of VIMP, which indicates degradation if that variable was not used for classification and we can reveal which variable is effective. Figure 4 shows VIMP values of contribution of Case 3 to the classification model. SIPI, which contributed to the identification of maize and wheat, was the most important variable occupying 18.8%. PVR, which contributed to the identification of beans, beetroot, grass and wheat, was second and occupied 14.6%. VARlgreen, which contributed to the identification of beans and wheat, was third; although, it was relatively correlated with PVR (Figure 3).

4. Discussion

The classification results were different from each other ($p < 0.05$ based on McNemar's test) and Case 2 was superior to Case 1 even though the statistical differences between

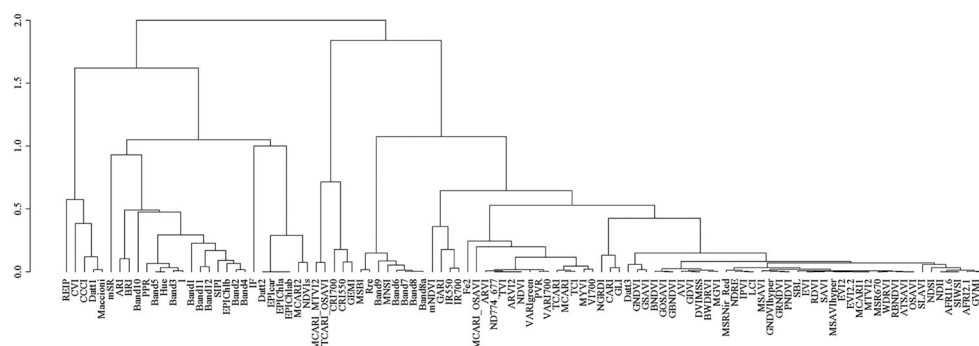


Figure 3. Clustering of correlation relationships among reflectance and spectral indices.

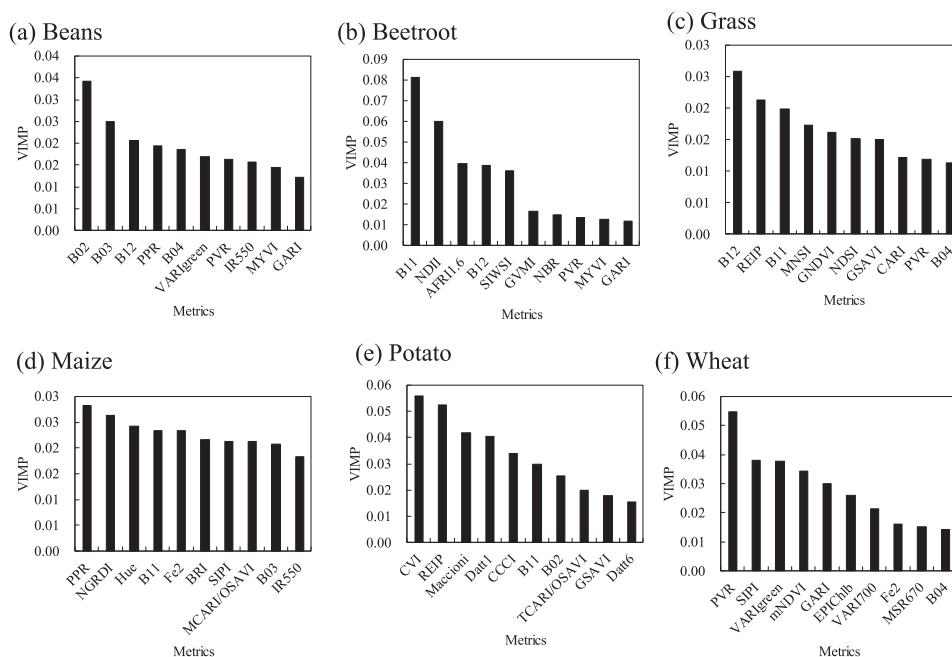


Figure 4. Variable importance of the top 10 useful variables (Case 3) for identification of each crop; (a) beans, (b) beetroot, (c) grass, (d) maize, (e) potato and (f) wheat.

these 3 cases are minor. It was assumed that the use of spectral indices would improve classification accuracy; however, the classification accuracy of Case 3, which was the integration of reflectance and spectral indices, was inferior to Case 2. [Figure 3](#) shows the clustering of the correlation relationship among reflectance and spectral indices. In RF, there is the hypothesis that each variable is independent of the response variable as well as from all other predictors. Clearly, there are large sets of correlated variables (e.g. strong correlations were confirmed between Bands 5 and Hue), which may cause the negative effects (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, [2008](#)).

Next, DSA was applied to clarify which variables contributed to the high accuracy. The importance of identifying each crop type or improving classification accuracies based on

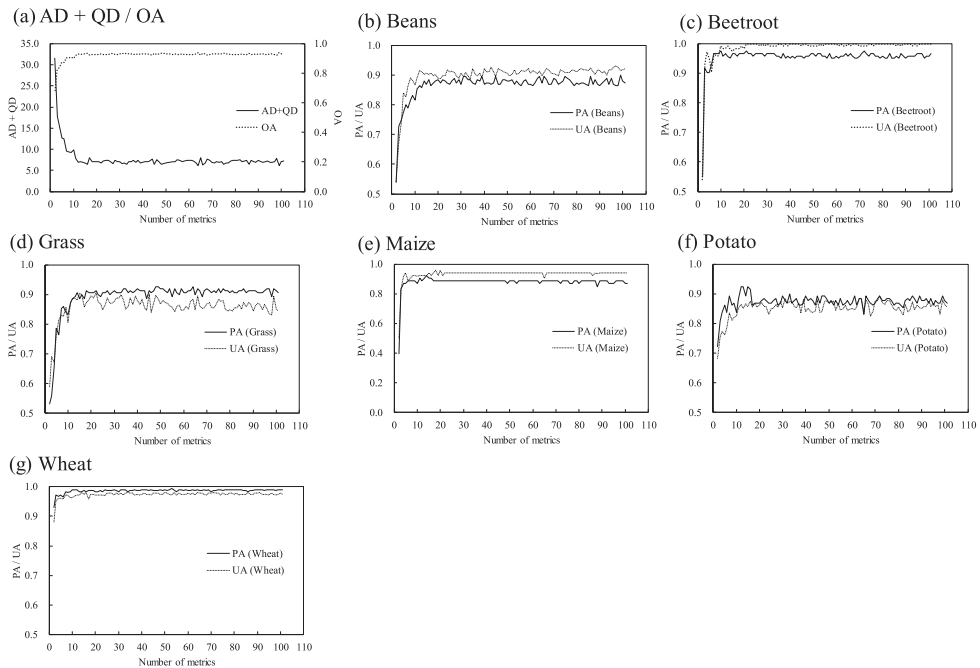


Figure 5. Relationships between accuracies and number of variables.

AD + QD or OA were ranked based on the results of DSA when Case 3 was applied. The classification contribution of different spectral indices and bands were evaluated by adding the matrices of inferior rank one by one (Figure 5). Using only 12 metrics including PVR (normalized difference between Band 3 and Band 4), Band 11, VARlgreen, SIPI, GARI, mNDVI, Band 12, REIP, EPIClb, Band 2, Band 4 and NDII, an OA of 93.1% (AD + QD = 6.89) was achieved and at least 6 metrics were required to achieve an OA over 90%. Figure 6 shows the crop classification map generated when the 12 metrics were used. Especially, using the top 2 matrices, PAs and UAs more than 0.5 with an overall accuracy of 68.4% were achieved. On 11 August, wheat was in the late grain-filling stage and the difference between Band 3 and Band 4 was negative in wheat fields while those over other fields were positive, so identifying wheat could be carried out using only PVR. Besides that, in potato fields, the difference between Band 3 and Band 4 was smaller than those of beans, beet, grass and maize fields because of discolouring by a chemical treatment to kill potato vines (Figure 2). Furthermore, the differences in photosynthetic capacities among crop species and the reflectance at Band 11 is sensitive to non-structural carbohydrates of the leaf (Asner & Martin, 2015). It has a positive correlation with leaf nitrogen content (Bartlett, Ollinger, Hollinger, Wicklein, & Richardson, 2011) and it could be useful to estimate the contents of photosynthetic pigments, water, nitrogen, cellulose, lignin, phenols, and leaf mass per area (Asner et al., 2011).

Adding VARlgreen, which is similar to PVR but Band 2 was deducted in its denominator, was useful for distinguishing beans and beetroot fields. Although the reflectance at Band 4 was almost the same between beans and beetroot, the difference between Band 3 and Band 2 was positive for beetroot and negative for beans. VARlgreen and GARI could emphasize the differences in these spectral features. While the reflectances at Bands 2

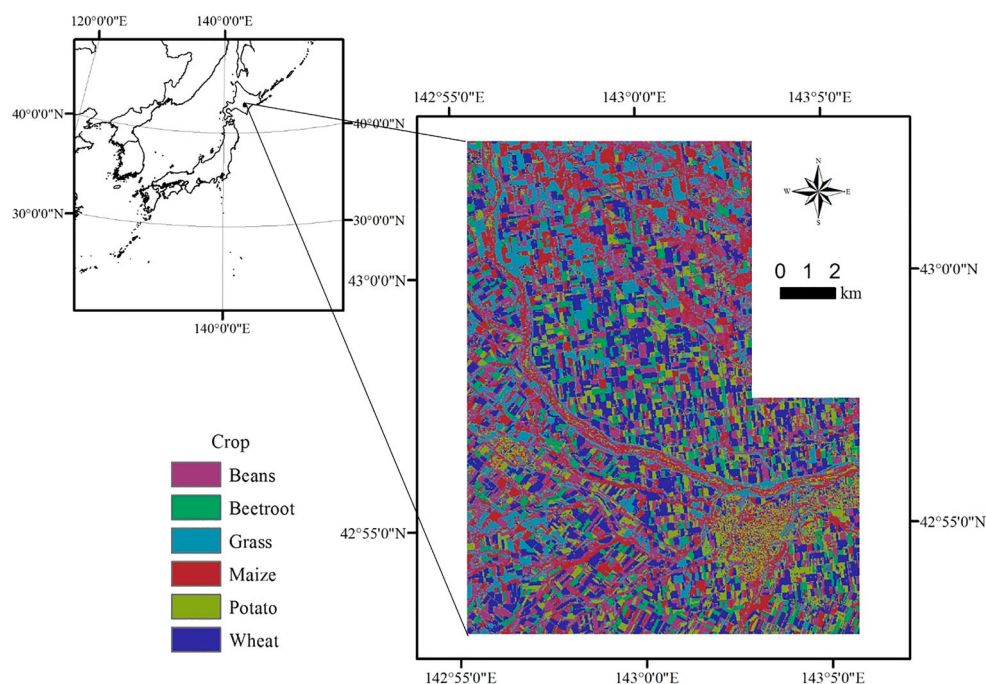


Figure 6. Crop classification map.

and 4 were almost similar among beetroot, grass and maize, the reflectance at Band 8 was markedly different. After identifying beans using the three metrics, SIPI and GARI contributed to improvement of identifying grasslands. However, the differences at Bands 3 and 4, which are considered in GARI, were obscure in Figure 2. This could be the reason that GARI was inferior to SIPI in this study. mNDVI helped to improve the PAs of beans, beetroot and maize, and the UA of grassland by emphasizing the difference in Band 8 using Bands 2 and 4 like VARlgreen and an OA over 90% was achieved. However, high PA or UA of potato was not achieved without Band 4 and NDII. Band 4 of potato was smaller than wheat but larger than the other 4 crops, while Band 8 of potato was larger than wheat but smaller than the other 4 crops. These relationships helped with identification of potato.

The cost sensitive learning was used to deal with the imbalanced data classification problem following (Chen, Liaw, & Breiman, 2004) and penalties were imposed on misclassifying the minority class since RF tends to be biased towards the majority class. When the 12 metrics were used, UAs of beetroot and potato were slightly improved (98.3% to 100.0% for beetroot and 86.3% to 87.9% for potato) compared with the results when this strategy was not applied, however, other indicators were not improved and thus it revealed that the influences on the imbalanced data classification problem were so weak in this study.

5. Conclusions

In this study, only one image from Sentinel-2A MSI was used to generate a classification map. In addition to the original reflectance, 91 published spectral indices were evaluated

for their crop classification performance. The use of spectral indices improved the classification accuracy (achieving an overall accuracy of 93.0%), but the integrated use of reflectance and spectral indices enhanced the negative effects related to large sets of correlated variables and decreased the accuracy (OA = 92.4%).

Next, the importance of identifying each crop type or improving classification accuracies were evaluated based on the results of DSA and the classification contribution of different spectral indices and bands were evaluated by adding the matrices of inferior rank one by one. As the results, the classification based on the reflectance including Bands 2, 4, 11 and 12, and 8 spectral indices including PVR, VARIGreen, SIPI, GARI, mNDVI, REIP, EPIChlb and NDII was useful for identification of each crop and achieved an OA of 93.1%.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Nobuyuki Kobayashi received the M.S. degree from Hokkaido University, Japan. He is currently with Smart Link Hokkaido, Japan. His research mainly focuses on smart agriculture.

Hiroshi Tani received the Ph.D. degree in Agriculture from Hokkaido University, Japan. He is an Associate Professor of the Research Faculty of Agriculture, Hokkaido University. His research interests include agricultural meteorology and remote sensing of environment.

Xiufeng Wang received the B.S. degree from Northeast Agricultural University, Harbin, China, in 1982 and the Ph.D. degree from Hokkaido University, Sapporo, Japan, in 1994. She is an Associate Professor with the Research Faculty of Agriculture, Hokkaido University. Her main research interests are in the study of the analysis methods and its applications about satellite data for agricultural meteorology.

Rei Sonobe received the Ph.D. degree in Agriculture from Hokkaido University, Japan. He is currently an Assistant Professor and a Faculty Member with the Faculty of Agriculture, Shizuoka University, Shizuoka, Japan. His research interests include remote sensing applications for agriculture and machine learning.

ORCID

Hiroshi Tani  <http://orcid.org/0000-0002-3025-7611>

Rei Sonobe  <http://orcid.org/0000-0002-8330-3730>

References

- Al-Gaadi, K. A., Hassaballa, A. A., Tola, E., Kayad, A. G., Madugundu, R., Alblewi, B., & Assiri, F. (2016). Prediction of potato crop yield using precision agriculture techniques. *Plos One*, 11, 16.
- Ashburn, P., 1978. The vegetative index number and crop identification. *The LACIE Symposium Proceedings of the Technical Session* (pp. 843–855).
- Asner, G. P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, 64, 234–253.
- Asner, G. P., & Martin, R. E. (2015). Spectroscopic remote sensing of non-structural carbohydrates in forest canopies. *Remote Sensing*, 7, 3526–3547.

- Asner, G. P., Martin, R. E., Knapp, D. E., Tupayachi, R., Anderson, C., Carranza, L., ... Weiss, P. (2011). Spectroscopy of canopy chemicals in humid tropical forests. *Remote Sensing of Environment*, 115, 3587–3598.
- Avci, Z. D. U., & Sunar, F. (2015). Process-based image analysis for agricultural mapping: A case study in Turkgedli region, Turkey. *Advances in Space Research*, 56, 1635–1644.
- Baker, B. A., Warner, T. A., Conley, J. F., & Mcneil, B. E. (2013). Does spatial resolution matter? A multi-scale comparison of object-based and pixel-based methods for detecting change associated with gas well drilling operations. *International Journal of Remote Sensing*, 34, 1633–1651.
- Bansal, S., Katyal, D., & Garg, J. K. (2017). A novel strategy for wetland area extraction using multispectral MODIS data. *Remote Sensing of Environment*, 200, 183–205.
- Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indexes for Lai and Apar assessment. *Remote Sensing of Environment*, 35, 161–173.
- Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberl, J., Kostrzewski, M., ... Moran, M. S. (2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. (Eds.), *the fifth international conference on precision agriculture and other resource management*. Madison, WI: ASA-CSSA-SSSA.
- Bartlett, M. K., Ollinger, S. V., Hollinger, D. Y., Wicklein, H. F., & Richardson, A. D. (2011). Canopy-scale relationships between foliar nitrogen and albedo are not observed in leaf reflectance and transmittance within temperate deciduous tree species. *Botany-Botanique*, 89, 491–497.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281–305.
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25, 197–227.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Broge, N. H., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 76, 156–172.
- Castillo, J. A. A., Apan, A. A., Maraseni, T. N., & Salmo, S. G. (2017). Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 134, 70–85.
- Chen, J. M. (1996). Evaluation of vegetation indices and a Modified simple ratio for Boreal Applications. *Canadian Journal of Remote Sensing*, 22, 229–242.
- Chen, J. M., & Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using landsat TM images. *Remote Sensing of Environment*, 55, 153–162.
- Chen, C., Liaw, A., & Breiman, L. (2004). *Using random forest to learn imbalanced data*.
- Chen, C. F., Son, N. T., & Chang, L. Y. (2012). Monitoring of rice cropping intensity in the upper Mekong Delta, Vietnam using time-series MODIS data. *Advances in Space Research*, 49, 292–301.
- Chivkunova, O. B., Solovchenko, A. E., Sokolova, S. G., Merzlyak, M. N., Reshetnikova, I. V., & Gitelson, A. A. (2001). Reflectance spectral features and detection of superficial scald -induced browning in storing apple fruit. *Journal of Russian Phytopathological Society*, 2, 73–77.
- Clevers, J., De Jong, S. M., Epema, G. F., Van Der Meer, F. D., Bakker, W. H., Skidmore, A. K., & Scholte, K. H. (2002). Derivation of the red edge index using the MERIS standard band setting. *International Journal of Remote Sensing*, 23, 3169–3184.
- Clevers, J. G. P. W., & Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation*, 23, 344–351.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cortez, P., & Embrechts, M. J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, 225, 1–17.
- Crippen, R. E. (1990). Calculating the vegetation index faster. *Remote Sensing of Environment*, 34, 71–73.
- Datt, B. (1998). Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a + b, and total carotenoid content in eucalyptus leaves. *Remote Sensing of Environment*, 66, 111–121.
- Datt, B. (1999). Remote sensing of water content in Eucalyptus leaves. *Australian Journal of Botany*, 47, 909–923.

- Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B., & McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74, 229–239.
- Dehni, A., & Lounis, M. (2012). Remote sensing techniques for Salt Affected soil mapping: Application to the Oran Region of Algeria. *Iswee'11*, 33, 188–198.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36.
- Eitel, J. U. H., Long, D. S., Gessler, P. E., & Smith, A. M. S. (2007). Using in-situ measurements to evaluate the new RapidEye (TM) satellite series for prediction of wheat nitrogen status. *International Journal of Remote Sensing*, 28, 4183–4190.
- El-Shikha, D. M., Barnes, E. M., Clarke, T. R., Hunsaker, D. J., Haberland, J. A., Pinter, P. J., ... Thompson, T. L. (2008). Remote sensing of cotton nitrogen status using the Canopy chlorophyll content index (CCCI). *Transactions of the Asabe*, 51, 73–82.
- Escadafal, R., Belghith, A., & Ben Moussa, H. (1994). Indices spectraux pour la degradation des milieux naturels en Tunisie arideed. (Eds.), *6e Symposium international sur les mesures physiques et signatures en teledetection* (pp. 253–259). Val d'Isere, France.
- Fensholt, R., & Sandholt, I. (2003). Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment. *Remote Sensing of Environment*, 87, 111–121.
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, 185–201.
- Foody, G.M. (2009). lassification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sensing of Environment*, 113, 1658–1663.
- Gao, F., Anderson, M. C., Zhang, X. Y., Yang, Z. W., Alfieri, J. G., Kustas, W. P., ... Prueger, J. H. (2017). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment*, 188, 9–25.
- Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology*, 161, 165–173.
- Gitelson, A. A., Chivkunova, O. B., & Merzlyak, M. N. (2009). Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. *American Journal of Botany*, 96, 1861–1868.
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58, 289–298.
- Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, 80, 76–87.
- Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001a). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology*, 74, 38–45.
- Gitelson, A. A., Merzlyak, M. N., Zur, Y., Stark, R., & Gritz, U. (2001b). Non-destructive and remote sensing techniques for estimation of vegetation statused. (Eds.), *Third European conference on precision agriculture* (pp. 301–306). Montpellier, France.
- Glenn, E. P., Nagler, P. L., & Huete, A. R. (2010). Vegetation index methods for estimating evapotranspiration by remote sensing. *Surveys in Geophysics*, 31, 531–555.
- Gobron, N., Pinty, B., Verstraete, M. M., & Widlowski, J. L. (2000). Advanced vegetation indices optimized for up-coming sensors: Design, performance, and applications. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 2489–2505.
- Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90, 337–352.
- Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, 81, 416–426.

- Haboudane, D., Tremblay, N., Miller, J. R., & Vigneault, P. (2008). Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data. *Ieee Transactions on Geoscience and Remote Sensing*, 46, 423–437.
- Hancock, D. W., & Dougherty, C. T. (2007). Relationships between blue- and red-based vegetation indices and leaf area and yield of alfalfa. *Crop Science*, 47, 2547–2556.
- Hardisky, M. A., Klemas, V., & Smart, R. M. (1983). The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of *Spartina alterniflora* canopies. *Photogrammetric Engineering & Remote Sensing*, 49, 77–83.
- Hartfield, K., Marsh, S., Kirk, C., & Carriere, Y. (2013). Contemporary and historical classification of crop types in Arizona. *International Journal of Remote Sensing*, 34, 6024–6036.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer-Verlag.
- Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V., & Bonfil, D. J. (2011). LAI assessment of wheat and potato crops by VEN mu S and Sentinel-2 bands. *Remote Sensing of Environment*, 115, 2141–2151.
- Hill, M. J. (2013). Vegetation index suites as indicators of vegetation state in grassland and savanna: An analysis with simulated SENTINEL 2 data for a North American transect. *Remote Sensing of Environment*, 137, 94–111.
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25, 295–309.
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, 195–213.
- Hunt, E. R., Daughtry, C. S. T., Eitel, J. U. H., & Long, D. S. (2011). Remote sensing leaf chlorophyll content using a visible Band index. *Agronomy Journal*, 103, 1090–1099.
- Ishwaran, H. (2007). Variable importance in binary regression trees and forests. *Electronic Journal of Statistics*, 1, 519–537.
- Ishwaran, H., & Kogalur, U. B. (2007). Random survival forests for R. *R News*, 7, 25–31.
- Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, 2, 841–860.
- Jiang, Z. Y., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112, 3833–3845.
- Karnieli, A., Kaufman, Y. J., Remer, L., & Wald, A. (2001). AFRI – aerosol free vegetation index. *Remote Sensing of Environment*, 77, 10–21.
- Kaufman, Y. J., & Tanre, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 261–270.
- Kewley, R. H., Embrechts, M. J., & Breneman, C. (2000). Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks*, 11, 668–679.
- Key, C., & Benson, N. (2006). Landscape assessment: Ground measure of severity; the composite burn index, and remote sensing of severity, the normalized burn index. Landscape assessment: Ground measure of severity; the composite burn index, and remote sensing of severity, the normalized burn index. In D. Lutes, R. Keane, J. Caratti, C. Key, N. Benson, S. Sutherland, & L. Gangi (Eds.), *FIREMON: Fire effects monitoring and inventory system* (pp. 1–51). Fort Collins, CO: Rocky Mountains Research Station, USDA Forest Service.
- Kim, M. S., Daughtry, C. S. T., Chappelle, E. W., McMurtrey, J. E. & Walthall, C. L. (1994). The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (A par). In *6th international symposium on physical measurements and signatures in remote sensing* (pp. 299–306). Val D'Isere.
- Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R News*, 2, 18–22.
- Liu, L. X., Coops, N. C., Aven, N. W., & Pang, Y. (2017). Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. *Remote Sensing of Environment*, 200, 170–182.
- Lv, T. T., & Liu, C. (2010). Study on extraction of crop information using time-series MODIS data in the Chao Phraya Basin of Thailand. *Advances in Space Research*, 45, 775–784.

- Lymburner, L., Beggs, P. J., & Jacobson, C. R. (2000). Estimation of canopy-average surface-specific leaf area using Landsat TM data. *Photogrammetric Engineering and Remote Sensing*, 66, 183–191.
- Maccioni, A., Agati, G., & Mazzinghi, P. (2001). New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra. *Journal of Photochemistry and Photobiology B-Biology*, 61, 52–61.
- Main, R., Cho, M. A., Mathieu, R., O'kenedy, M. M., Ramoelo, A., & Koch, S. (2011). An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 751–761.
- Martyniak, L., Dabrowska-Zielinska, K., Szymczyk, R., & Gruszczynska, M. (2007). Validation of satellite-derived soil-vegetation indices for prognosis of spring cereals yield reduction under drought conditions – case study from central-western Poland. *Advances in Space Research*, 39, 67–72.
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12, 153–157.
- Metternicht, G. (2003). Vegetation indices derived from high-resolution airborne videography for precision crop management. *International Journal of Remote Sensing*, 24, 2855–2877.
- Misra, P. N., Wheeler, S. G., & Oliver, R. E. (1977). Kauth-Thomas brightness and greenness axes. *Contract NASA 9-14350*, 23–46.
- Miura, T., Yoshioka, H., Fujiwara, K., & Yamamoto, H. (2008). Inter-comparison of ASTER and MODIS surface reflectance and vegetation index products for synergistic applications to natural resource monitoring. *Sensors*, 8, 2480–2499.
- Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016). Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187, 156–168.
- Pena, M. A., Liao, R., & Brenning, A. (2017). Using spectrottemporal indices to improve the fruit-tree crop classification accuracy. *Isprs Journal of Photogrammetry and Remote Sensing*, 128, 158–169.
- Penuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll-a ratio from leaf spectral reflectance. *Photosynthetica*, 31, 221–230.
- Pinty, B., & Verstraete, M. M. (1992). GEMI: A non-linear index to monitor global vegetation from satellites. *Vegetatio*, 101, 15–20.
- Pontius, R., & Millones, M. (2011). Death to kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32, 4407–4429.
- Prasad, A. K., Sarkar, S., Singh, R. P., & Kafatos, M. (2007). Inter-annual variability of vegetation cover and rainfall over India. *Advances in Space Research*, 39, 79–87.
- Puertas, O., Brenning, A., & Meza, F. (2013). Balancing misclassification errors of land cover classification maps using support vector machines and Landsat imagery in the Maipo river basin (Central Chile, 1975–2010). *Remote Sensing of Environment*, 137, 112–123.
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48, 119–126.
- Richardson, A. J., & Wiegand, C. L. (1978). Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, 43, 1541–1552.
- Rodriguez-Galiano, V., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P., & Jeganathan, C. (2012). Random forest classification of mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, 121, 93–107.
- Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55, 95–107.
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the great plains with ERTS. In S. C. Freden, E. P. Mercanti, & M. A. Becker (Eds.), *Third earth resources technology satellite-1 Symposium* (pp. 309–317). Washington, DC: NASA.
- Rowan, L. C., & Mars, J. C. (2003). Lithologic mapping in the Mountain Pass, California area using advanced spaceborne thermal emission and reflection radiometer (ASTER) data. *Remote Sensing of Environment*, 84, 350–366.
- Seiler, R. A., Kogan, F., Wei, G., & Vinocur, M. (2007). Seasonal and interannual responses of the vegetation and production of crops in Cordoba-Argentina assessed by AVHRR derived vegetation indices. *Advances in Space Research*, 39, 88–94.

- Siachalou, S., Mallinis, G., & Tsakiri-Strati, M. (2017). Analysis of time-series spectral index data to enhance crop identification over a mediterranean rural landscape. *Ieee Geoscience and Remote Sensing Letters*, 14, 1508–1512.
- Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81, 337–354.
- Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and change detection using Landsat TM data: When and how to correct atmospheric effects? *Remote Sensing of Environment*, 75, 230–244.
- Sonobe, R. (2019). Combining ASNARO-2 XSAR HH and Sentinel-1 C-SAR VH/VV polarization data for improved crop mapping. *Remote Sensing*, 11, 1920.
- Sonobe, R., Miura, Y., Sano, T., & Horie, H. (2018). Estimating leaf carotenoid contents of shade grown tea using hyperspectral indices and PROSPECT-D inversion. *International Journal of Remote Sensing*, 39, 1306–1320.
- Sonobe, R., Tani, H., Wang, X., Kobayashi, N., & Shimamura, H. (2014). Random forest classification of crop type using multi-temporal TerraSAR-X dual-polarimetric data. *Remote Sensing Letters*, 5, 157–164.
- Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., & Mochizuki, K.-I. (2017a). Mapping crop cover using multi-temporal Landsat 8 OLI imagery. *International Journal of Remote Sensing*, 38, 4348–4361.
- Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., & Mochizuki, K.-I. (2017b). Assessing the Suitability of data from Sentinel-1A and 2A for crop classification. *GIScience & Remote Sensing*, 54, 918–938.
- Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *Bmc Bioinformatics*, 9, 307.
- Thimsuwan, Y., Eiumnoh, A., Honda, K., & Tingsanchali, T. (2000). Estimation of methane emission from a deep-water rice field using Landsat TM and NOAA AVHRR: A case study of Bangkok Plain. *Imaging Science Journal*, 48, 77–85.
- Tian, Y., Zhu, Y., & Cao, W. (2005). Monitoring soluble sugar, total nitrogen & its ratio in wheat leaves with canopy spectral reflectance. *Acta Agronomica Sinica*, 31, 355–360.
- Tucker, C. J. (1979a). Monitoring corn and soybean crop development with hand-held radiometer spectral data. *Remote Sensing of Environment*, 8, 237–248.
- Tucker, C. J. (1979b). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127–150.
- Wang, F., Huang, J.-F., & Chen, L. (2010). New vegetation index and its application in estimating leaf area index of rice. *Journal of Plant Nutrition*, 33, 328–333.
- Wang, F.-M., Huang, J.-F., Tang, Y.-L., & Wang, X.-Z. (2007). New vegetation index and its application in estimating leaf area index of rice. *Rice Science*, 14, 195–203.
- Wardlow, B. D., & Egbert, S. L. (2008). Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the US central great plains. *Remote Sensing of Environment*, 112, 1096–1116.
- Xie, Q. Y., Huang, W. J., Dash, J., Song, X. Y., Huang, L. S., Zhao, J. L., & Wang, R. H. (2015). Evaluating the potential of vegetation indices for winter wheat LAI estimation under different fertilization and water conditions. *Advances in Space Research*, 56, 2365–2373.
- Yang, C. G., Everitt, J. H., & Bradford, J. M. (2007). Airborne hyperspectral imagery and linear spectral unmixing for mapping variation in crop yield. *Precision Agriculture*, 8, 279–296.
- Zarco-Tejada, P. J., Miller, J. R., Noland, T. L., Mohammed, G. H., & Sampson, P. H. (2001). Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1491–1507.