



Title	Mapping crop cover using multi-temporal Landsat 8 OLI imagery
Author(s)	Sonobe, Rei; Yamaya, Yuki; Tani, Hiroshi; Wang, Xiufeng; Kobayashi, Nobuyuki; Mochizuki, Kan-ichiro
Citation	International Journal of Remote Sensing, 38(15), 4348-4361 https://doi.org/10.1080/01431161.2017.1323286
Issue Date	2017-05-18
Doc URL	http://hdl.handle.net/2115/70679
Rights	This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Remote Sensing on 18 May 2017, available online: http://www.tandfonline.com/doi/abs/10.1080/01431161.2017.1323286 .
Type	article (author version)
File Information	Manuscript_TRES-PAP-2016-1123.docx.pdf



[Instructions for use](#)

Mapping crop cover using multi-temporal Landsat 8 OLI imagery

Crop classification maps are useful for estimating amounts of crops harvested, which could help address challenges in food security. Remote sensing techniques are useful tools for generating crop maps. Optical remote sensing is one of the most attractive options because it offers vegetation indices with frequent revisits, adequate spatial and spectral resolution and some data has been distributed free of charge. However, sufficient consideration has not been given to the potential of vegetation indices calculated from Landsat 8 Operational Land Imager (OLI) data. This paper describes the use of Landsat 8 OLI data for the classification of crops in Hokkaido, Japan. In addition to reflectance, vegetation indices calculated from simple formulas that consisted of combinations of two or more reflectance wavebands were evaluated, as well as the six components of the Kauth-Thomas transform. The vegetation indices based on shortwave infrared bands (bands 6 or 7) improved classification accuracy and using a combination of all derived data from Landsat 8 OLI data resulted in an overall accuracy of 94.5% (allocation disagreement=4.492 and quantity disagreement=1.017).

Keywords: Bayesian optimisation; crop; Kauth-Thomas transform; Landsat-8; random forests; vegetation indices.

1. Introduction

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 and Egbert, 2008).

Therefore, mapping cropland is needed to carry out social policies as well as estimating the amount and type of crops harvested in a certain area. In Japan, some local governments are documenting field properties including crop types and locations by site investigations. Reducing labour costs and developing easier methods such as the techniques based on satellite remote sensing are required due to high cost of existing methods.

Many studies have focused on land use and land cover classification based on remote sensing techniques. The Landsat series data are suitable for land characterisation activities, owing in large part to the spatial, spectral and radiometric qualities of the data (Wulder *et al.*, 2008, Vogelmann *et al.*, 2016). The data acquired from Landsat satellites are also useful for assessing the condition of vegetation. Vegetation indices (VIs) derived from reflectance data acquired from optical sensors have been widely used to assess variations in the physiological state and biophysical properties of vegetation (Bankestad and Wik, 2016, de Klerk *et al.*, 2016, Sonobe and Wang, 2017).

VIs are calculated from simple formulas that consist of a combination of two or more reflectance wavebands calculated from various satellite data. Specifically, the Normalized Difference Vegetation Index (NDVI), calculated from Landsat data, has been applied for monitoring vegetation systems or ecological responses to

environmental change (Holden and Woodcock, 2016). The Soil-Adjusted Vegetation Index (SAVI) (Huete, 1988), Enhanced Vegetation Index (EVI) (Huete *et al.*, 2002) and Ratio Vegetation Index (RVI) (Tucker, 1979, Sellers, 1985) have also been used to evaluate the abilities of monitoring vegetation (She *et al.*, 2015, Karan *et al.*, 2016). However, few indices have considered the abilities of shortwave infrared bands, which are mostly influenced by plant constituents including pigments, leaf water content and biochemicals, such as lignin and cellulose (Asner, 1998). In this study, indices based on differences (D), simple ratios (SR) and normalised differences (ND) were considered and all possible combinations of bands for the index types were evaluated for their contributions to classifications. Furthermore, the Kauth-Thomas (KT) transform is a useful tool for understanding crop development in spectral space (Kauth and Thomas, 1976) and it was modified for Landsat 8 Operational Land Imager (OLI) (Baig *et al.*, 2014). In this study, the use of VIs and KT components were evaluated for identifying crop species as well as reflectance acquired from Landsat 8 OLI data.

Continued algorithm development has led to improved classification accuracy and reduced computational cost for land cover classifications, and machine learning approaches have been widely used in image classification over the past two decades (Pal *et al.*, 2013). Some earlier studies using optical imagery showed that random forests (RF) was superior to the maximum likelihood classifier for classification

(Rodriguez-Galiano *et al.*, 2012). Furthermore, it can efficiently handle large sample sizes, run with a vast numbers of input variables, quantify each input variable based on an importance measure, and is robust with irregular samples, such as outliers and noise (Breiman, 2001, Cutler *et al.*, 2007). In this study, RF modified by Ishwaran (Ishwaran and Kogalur, 2007, Ishwaran *et al.*, 2008) was used for classification since it is useful for evaluating the variable importance (VIMP), which is calculated from training data (Ishwaran, 2007).

The main targets of this study were to (1) evaluate the accuracy of crop classification using Landsat 8 OLI data and (2) identify the best combination of the information derived from Landsat 8 OLI data.

2. Materials and Methods

2.1. Study area

The study area is located in the western Tokachi plain, Hokkaido, Japan (Figure 1, 142°55'12" to 143°05'51" E, 42°52'48" to 43°02'42" N)(Sonobe *et al.*, 2015b). The area has a humid continental climate characterised by warm summers and cold winters, with an average annual temperature of 6°C and an annual precipitation of 920 mm. The dominant crops are beans, beets, grasses, maize, potatoes and winter wheat.

<Figure 1>

Winter wheat was seeded in September, beets and potatoes were seeded or transplanted in late April to early May and maize and beans were seeded in mid-May. The harvesting periods were late July to early August for winter wheat, late August to September for potatoes and maize, late September to early November for beans and November for beets. Grasses were harvested twice a year; the first harvest was in late June to early July and the second was in late August.

2.2. Satellite data

The characteristics of the satellite images used in this study are shown in Table 1. The Landsat-8 OLI data were acquired on 20 May, 7 July and 24 August 2016. The data, which had been spatially orthorectified by the U.S. Geological Survey, were downloaded from EarthExplorer (<http://earthexplorer.usgs.gov/>). To correct the imagery for atmospheric effects, the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm (Matthew *et al.*, 2000) was applied using ENVI software (Version 5.3) and the reflectance values of bands 1-7 were used for classification in this study. Several studies have attempted to develop spectral indices based on the metrics of the reflectance to conduct non-destructive evaluations of physiological states and biophysical properties. Among them, indices based on difference (D type) (Sonobe and Wang, 2016), simple ratio (SR) (Blackburn, 1998) and normalized differences (ND type, NDVI belongs to this type) (Tucker, 1979, Blackburn,

1998, Gandia *et al.*, 2005) have been widely applied. Three following types of indices were calculated based on the reflectance acquired from OLI data;

$$D(\lambda_1, \lambda_2) = R_{\lambda_1} - R_{\lambda_2} \quad (1)$$

$$SR(\lambda_1, \lambda_2) = R_{\lambda_1} / R_{\lambda_2} \quad (2)$$

$$ND(\lambda_1, \lambda_2) = (R_{\lambda_1} - R_{\lambda_2}) / (R_{\lambda_1} + R_{\lambda_2}) \quad (3)$$

where R is reflectance at a given band (λ_1 or λ_2) and all 21 (7C_2 , the number of ways to choose a sample of 2 bands from a set of 7 bands where order does not matter and replacements are not allowed) combinations were calculated for each type and OLI image. SAVI (Huete, 1988) and EVI (Huete *et al.*, 2002) were also calculated.

$$SAVI = 1.5 \times (R_{\text{Band5}} - R_{\text{Band4}}) / (R_{\text{Band5}} + R_{\text{Band4}} + 0.5) \quad (4)$$

$$EVI = 2.5 \times (R_{\text{Band5}} - R_{\text{Band4}}) / (R_{\text{Band5}} + 6 \times R_{\text{Band4}} - 7.5 \times R_{\text{Band2}} + 1) \quad (5)$$

The earlier studies have shown that the data of the KT transform, also known as the Tasseled-Cap transform, improved the accuracy of the crop classification (Galvao *et al.*, 2015, Schultz *et al.*, 2016). Table 2 shows the coefficients of KT data for OLI (Baig *et al.*, 2014) and KT data was generated using these coefficients.

<Table 1>

<Table 2>

2.3. GIS data

Reference data were provided by Tokachi Nosai (located in Obihiro city, Hokkaido).

Nosai conducts an Agricultural Insurance, which supports farmers after natural disasters and contributes to the growth of Japanese agriculture. Reference data were provided as a polygon shape file, which included attribute data, such as crop type. The field size ranged from 0.1 to 17.5 ha with a median of 2.0 ha. The position of the centre of the fields was calculated and the reflectance values of all bands were extracted using these points. The process was applied to avoid selecting training pixels from the edge of a field, which would create a mixed signal and affect the accuracy assessment.

2.4. Classification algorithm

A total of 4719 fields (981 beans fields, 569 beet fields, 640 grasslands, 317 maize fields, 783 potato fields and 1,429 winter wheat fields) covered the area in 2016. In this data-rich situation the best approach for model selection and model assessment is to randomly divide the dataset into three parts: a training set, which is used to fit the models; a validation set, which is used to estimate prediction error for model selection and a test set, which is used for assessment of the generalisation error of the final chosen model (Hastie *et al.*, 2009). A stratified random-sampling approach was applied to select the fields used for training (50%), validation (25%) and test data (25%). We conducted classifications using the following four different datasets; Case 1: reflectance

data, Case 2: KT data, Case 3: VIs, Case 4: reflectance + VIs data and Case 5:

reflectance + KT + VIs data.

RF is an ensemble learning technique that builds multiple trees (the classification and regression tree, CART) based on random bootstrapped samples of the training data. Each tree is built using a different subset from the original training data, containing approximately two-thirds of the cases, and the nodes are split using the best variable out of a group of randomly selected variables. This strategy provides robustness to over-fitting and can handle thousands of dependent and independent input variables without variable deletion. The output is determined by a majority vote of the trees. The two user-defined parameters are the number of trees (ntree) and the number of variables used to split the nodes (mtry). RF differs from CART as they are grown nondeterministically in order to decorrelate the trees and reduce variance using a two-stage randomisation procedure related to a bootstrap sample and random variable selection. Prior to the construction of each decision tree, some samples are extracted at random with replacement from the original training dataset and these samples are taken into account in the tree building. Then splits are performed based on Gini index over mtry directions chosen uniformly at random. When ntree is increased, the generalisation error always converges, and over-training is not a problem. As a result, a tree of RF is grown as deeply as possible under the constraint that each terminal node must contain

no fewer than $\text{nodesize} \geq 1$ case. It is also generally believed that randomising the splitting rule can improve performance for ensembles (Ishwaran, 2015). So, RF modified by Ishwaran (Ishwaran and Kogalur, 2007, Ishwaran *et al.*, 2008) was used for classification and the contributions of each variable were assessed based on the variable importance (VIMP), which is calculated from training data and defined as the difference between prediction error when the variable is noisy versus prediction error caused by randomly permuting it (Ishwaran, 2007). Therefore, the value of VIMP indicates degradation if that variable was not used for classification and we can reveal which variable is effective for identifying crop types. As a result, three additional hyperparameters were considered, including the minimum number of unique cases in a terminal node (nodesize), the maximum depth to which a tree should be grown (nodedepth) and the number of random splitting (nsplit).

Grid search strategies have been applied for optimising the hyperparameters of machine learning algorithms (Puertas *et al.*, 2013, Sonobe *et al.*, 2015a). However, grid search strategies were shown to be a poor choice for configuring algorithms for new data sets, and this suggested the use of the Gaussian process, Bayesian optimisation (Bergstra and Bengio, 2012). Bayesian optimisation is a framework used to optimise hyperparameters of noisy, expansive black-box functions and it defines a principled approach to modelling uncertainty. Depending on its characteristics, such as simplicity

and flexibility, the Gaussian process (GP) has commonly been used for Bayesian optimisation (Snoek *et al.*, 2015). Therefore, the five hyperparameters were selected through the framework of Bayesian optimisation, in which a learning algorithm's generalisation performance is modelled as a sample from a GP (Snoek *et al.*, 2012). The parameter spaces were 3 to 1000 for *ntree*, 1 to the total of variables (i.e. 21 for Case 1, 18 for Case 2, 195 for Case 3, 216 for Case 4 and 234 for Case 5) for *mtry*, 1 to 20 for *nodesize*, 1 to 50 for *nodedepth* and 1 to 20 for *nsplit*.

2.5. Accuracy assessment

The classification maps were evaluated for their overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA) and the two simple measures of quantity disagreement (QD) and allocation disagreement (AD), which provides an effective summary of a cross-tabulation matrix (Pontius and Millones, 2011). OA is the total classification accuracy. PA is obtained by dividing the number of correctly classified fields in each crop type by the number of reference fields. This value represents how well the reference fields of the crop cover type are classified. UA is computed by dividing the number of correctly classified fields in each crop type by the total number of fields classified in that crop type. It represents the probability that a field classified into a given crop type actually represents that crop type. QD is defined as the difference between the reference data and the classified data based on a mismatch of class

proportions. AD is considered the difference between the classified data and the reference data due to incorrect spatial allocations of pixels in the classification. The total disagreement is the sum of QD and AD (Pontius and Millones, 2011). AD indicates the difference between the classified data and the reference data misled by incorrect spatial allocations, and QD provides information on the error in the relative proportion of the classes in the classification results.

To compare the accuracy of the classification methods, we used McNemar's test (McNemar, 1947). McNemar's test considers the lack of independent samples by focusing on how each point was either correctly or incorrectly classified in the two classifications being compared. A χ^2 value ≥ 3.84 indicates a significant difference between the two methods at the 95% level of significance.

3. Results and discussion

3.1. Reflectance obtained from Landsat OLI data for six crops

The mean reflectance of each crop is shown in Figure 2. There were small plants (less than 7 cm in height) in the fields of beans, beet, maize and potato on 20 May; therefore, their reflectance features were similar to bare soil and their differences were not visible. In contrast, the fields of wheat and grass were covered with vegetation, and a high reflectance at near infrared (band 5) was confirmed.

The reflectance in the peak of the growing season (7 July) showed similar patterns in the visible domain (bands 1-4), but the patterns at near and shortwave near infrared bands (bands 5-7) were different among them and could be divided into three groups. The first group had a specific peak at band 5 and beet, potato and wheat belonged to this group; the second showed a gradual decrease with the wavelength and grass and maize belonged to this group, and the last group had a gradual increase from band 5 to band 6 and beans belonged to this group.

On 24 August, the reflectance of wheat was similar to bare soil with peak at band 6 since the wheat had been cultivated. The characteristic reflectance features related to wilt caused by chemical treatments were also confirmed in the potatoes. Other crops had similar spectral patterns with peaks at band 5. Especially the date belonged to the late growing season or the early in the ripening period of beans and beet. As the result, the band 5 of these crops was bigger than that observed on 7 July.

<Figure 2>

3.2. Accuracy Validation

The optimal values of the combinations of the five parameters (ntree, mtry, nodesize, nodedepth, nsplit) were (543, 13, 1, 48, 10) for Case 1; (858, 14, 2, 37, 11) for Case 2; (394, 185, 2, 11, 2) for Case 3; (478, 155, 3, 16, 2) for Case 4 and (827, 80, 2, 42, 8) for Case 5. It was conceivable that these negative effects related to large sets were

relatively low since the values of $mtry$ were higher than the default value which was proposed by Breiman (2001). The corresponding confusion matrices of classifications are given in Table 3 and the results of McNemar's test are shown in Table 4. The best PAs and the best UAs of beans, grass and maize were confirmed by Case 5, the best UAs of beet and potato were confirmed by Case 4 (0.986 and 0.931, respectively) and the best UA of wheat was confirmed by Case 1 (0.983). The classification results were different from each other except for the relationship between Case 3 and 4 ($p < 0.05$ based on McNemar's test). Case 5 was the best based on OA and AD+QD and Figure 3 shows the crop classification map generated from Case 5. Thus, reflectance data were not needed when VIs were applied for classification.

<Table 3>

<Table 4>

<Figure 3>

Figure 4 shows the relative importance of the contribution to the classification model for the combination of reflectance data, KT data and VIs (Case 5). According to the VIMP, the metrics derived from bands 6 or 7 were most significant for classification. The difference between bands 6 and 7 on 7 July was useful to identify grass, potato and wheat. However, some metrics could be used instead of the reflectance values at these bands, since there was no significant difference between the classification result of

Cases 3 and 4 (Table 4). Previous studies have shown that reflectance at shortwave infrared contributed to the estimation of photosynthetic pigments, water, nitrogen, cellulose, lignin, phenols and leaf mass per area (LMA), which accurately estimated canopy traits (Asner *et al.*, 2011). Therefore, the reflectance at bands 6 and 7 in the growing season periods may be useful to detect differences in vegetation properties, such as biomass or photosynthetic characteristics among crops. However, the data acquired on 20 May contributed to the identification of grass since the first harvest of grass was conducted in late June to early July and the second was conducted in late August. The data also contributed to the identification of potato, though there were small plants in the fields because of the pronounced furrow ridges (30-35 cm in height).

Although the reflectance of bands 3 and 6 on 7 July was useful for identifying wheat, the performances of classification was improved by using the metrics calculated from two bands (not restricted to the published indices). The differential type of index using reflectance at band 3 and band 4 on 20 May was also useful due to the differences in crop phenologies although the other metrics were not contributed to identify wheat. The contributions of the KT data, which is calculated from the reflectance of six bands, were relatively low. However, the KT data helped to improve the classification accuracy because there was a significant difference between Cases 4 and 5 (Table 4).

<Figure 4>

3.3. Comparison with traditional random forests

The traditional RF method (Breiman, 2001) has widely been used in previous studies with high accuracies (Pal, 2005, Kavzoglu and Colkesen, 2013, Sonobe *et al.*, 2014).

The comparison between modified RF and traditional RF was conducted (Table 5) and the significant differences were confirmed for all cases. The result for Case 5 was also best for traditional RF and OA of 0.943, AD of 4.492 and QD of 1.186 were confirmed. Although the differences were quite small (less than 1%), the difference was significant ($p < 0.05$, χ^2 of 4.00) and the advantage of the modified RF was confirmed. However, the traditional RF still had an OA of more than 0.9 for Cases 1-4 and it could be an attractive option for crop classification using Landsat 8 OLI imagery.

4. Conclusions

Three Landsat 8 OLI images, acquired on 20 May, 7 July and 24 August, were used and the reflectance, KT transform data and vegetation indices including D, SR and ND were calculated. Classifications were made using four different datasets: reflectance data (Case 1), KT data (Case 2), VIs (Case 3), reflectance + VIs (Case 4) and reflectance + KT + VIs data (Case 5). When all data were combined (Case 5), the best performance, in terms of OA and QD+AD was confirmed, and an OA of 94.5% was achieved.

Furthermore, the VIs derived from bands 6 or 7 were important for crop classification based on the VIMP values, and this study demonstrated that adding

vegetation indices improved the classification accuracy. Collecting training data is still a big problem and saving some manual labour related to this task are required. The potential application for the cross-year training and classification (cross-year classification) should be analysed for future work.

Acknowledgments

The authors would like to thank Tokachi Nosai for providing the field data.

Funding

This work was supported by JSPS KAKENHI [grant number 15H06249].

Disclosure statement

No potential conflicts of interest are reported by the authors.

References

- Asner, G.P., 1998. "Biophysical and biochemical sources of variability in canopy reflectance." *Remote Sensing of Environment*, 64 (3): 234-253. doi:10.1016/S0034-4257(98)00014-5.
- Asner, G.P., Martin, R.E., Knapp, D.E., Tupayachi, R., Anderson, C., Carranza, L., Martinez, P., Houcheime, M., Sinca, F. and Weiss, P., 2011. "Spectroscopy of canopy chemicals in humid tropical forests." *Remote Sensing of Environment*, 115 (12): 3587-3598. doi:10.1016/j.rse.2011.08.020.
- Baig, M.H.A., Zhang, L.F., Shuai, T. and Tong, Q.X., 2014. "Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance." *Remote Sensing Letters*, 5 (5): 423-431. doi:10.1080/2150704X.2014.915434.
- Bånkestad, D. and Wik, T., 2016. "Growth tracking of basil by proximal remote sensing of chlorophyll fluorescence in growth chamber and greenhouse environments." *Computers and Electronics in Agriculture*, 128, 77-86. doi: 10.1016/j.compag.2016.08.004.

- Bergstra, J. and Bengio, Y., 2012. "Random search for hyper-parameter optimization." *Journal of Machine Learning Research*, 13 (1): 281-305.
- Blackburn, G.A., 1998. "Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves." *International Journal of Remote Sensing*, 19(4): 657-675. doi:10.1080/014311698215919.
- Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5-32. doi:10.1023/A:1010933404324.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A. and Hess, K.T., 2007. "Random forests for classification in ecology." *Ecology*, 88 (11): 2783-2792. doi:10.1890/07-0539.1.
- De Klerk, H.M., Gilbertson, J., Luck-Vogel, M., Kemp, J. and Munch, Z., 2016. "Using remote sensing in support of environmental management: A framework for selecting products, algorithms and methods." *Journal of Environmental Management*, 182, 564-573. doi:10.1016/j.jenvman.2016.07.073.
- Galvao, L.S., Dos Santos, J.R., Da Silva, R.D., Da Silva, C.V., Moura, Y.M. and Breunig, F.M., 2015. "Following a site-specific secondary succession in the Amazon using the Landsat CDR product and field inventory data." *International Journal of Remote Sensing*, 36(2): 574-596. doi:10.1080/01431161.2014.999879.
- Gandía, S., Fernández, G. and Moreno, J., 2004. "Retrieval of Vegetation Biophysical Variables from CHRIS/PROBA Data in the SPARC Campaign." In *Proceedings of the 2nd CHRIS/Proba Workshop, Frascati, Italy, 28-30 April, 2004*, 40-48. Frascati, Italy: ESA/ESRIN.
- Hastie, T., Tibshirani, R. and Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition*. New York, United States of America: Springer-Verlag.
- Holden, C.E. and Woodcock, C.E., 2016. "An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations." *Remote Sensing of Environment*, 185: 16-36. doi:10.1016/j.rse.2016.02.052.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. and Ferreira, L.G., 2002. "Overview of the radiometric and biophysical performance of the MODIS vegetation indices." *Remote Sensing of Environment*, 83(1-2): 195-213. doi:10.1016/S0034-4257(02)00096-2.
- Huete, A.R., 1988. "A Soil-Adjusted Vegetation Index (SAVI)." *Remote Sensing of Environment*, 25(3): 295-309. doi:10.1016/0034-4257(88)90106-X.
- Ishwaran, H., 2007. "Variable importance in binary regression trees and forests." *Electronic Journal of Statistics*, 1: 519-537. doi:10.1214/07-EJS039.
- Ishwaran, H., 2015. "The effect of splitting on random forests." *Machine Learning*, 99: 75-118. doi:10.1007/s10994-014-5451-2.

- Ishwaran, H. and Kogalur, U.B., 2007. "Random survival forests for R." *R News*, 7: 25-31.
- Ishwaran, H., Kogalur, U.B., Blackstone, E.H. and S., L.M., 2008. "Random survival forests." *The Annals of Applied Statistics*, 2(3): 841-860. doi:10.1214/08-AOAS169.
- Karan, S.K., Samadder, S.R. and Maiti, S.K., 2016. "Assessment of the capability of remote sensing and GIS techniques for monitoring reclamation success in coal mine degraded lands." *Journal of Environmental Management*, 182: 272-283. doi:10.1016/j.jenvman.2016.07.070.
- Kauth, R.J. and Thomas, G.S., 1976. "The tasselled Cap - A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat." *In Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, West Lafayette, Indiana, 29 June - 1 July, 1976, 4B-41 to 4B-51. New York: Institute of Electrical and Electronics Engineers, Inc.
- Kavzoglu, T. and Colkesen, I., 2013. "An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping." *International Journal of Remote Sensing*, 34(12): 4224-4241. doi:10.1080/01431161.2013.774099.
- Matthew, M.W., Adler-Golden, S.M., Berk, A., Richtsmeier, S.C., Levine, R.Y., Bernstein, L.S., Acharya, P.K., Anderson, G.P., Felde, G.W., Hoke, M.L., Ratkowski, A.J., Burke, H.-H.K., Kaiser, R.D. and Miller, D.P., 2000. "Status of atmospheric correction using a MODTRAN4-based algorithm." *In* S.S. Shen, Descour, Michael R. (ed.) *SPIE 4049, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*. Orlando, FL: SPIE-The International Society for Optical Engineering. , 410341.
- McNemar, Q., 1947. "Note on the sampling error of the difference between correlated proportions or percentages." *Psychometrika*, 12(2): 153-157. doi:10.1007/BF02295996.
- Pal, M. 2005. "Random Forest Classifier for Remote Sensing Classification." *International Journal of Remote Sensing*, 26: 217-222. doi:10.1080/01431160412331269698.
- Pal, M., Maxwell, A.E. and Warner, T.A., 2013. "Kernel-based extreme learning machine for remote-sensing image classification." *Remote Sensing Letters*, 4(9): 853-862. doi:10.1080/2150704X.2013.805279.
- Pontius, R. G., and M. Millones. 2011. "Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment." *International Journal of Remote Sensing*, 32 (15): 4407-4429. doi:10.1080/01431161.2011.552923.
- Puertas, O., Brenning, A. and 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. doi: 10.1016/j.rse.2013.06.003.
- Rodriguez-Galiano, V., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. and 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. doi:10.1016/j.rse.2011.12.003.
- Schultz, M., Clevers, J., Carter, S., Verbesselt, J., Avitabile, V., Quang, H.V. and Herold, M., 2016. "Performance of vegetation indices from Landsat time series in deforestation monitoring." *International Journal of Applied Earth Observation and Geoinformation*, 52: 318-327. doi:10.1016/j.jag.2016.06.020.
- Sellers, P.J., 1985. "Canopy reflectance, photosynthesis and transpiration." *International Journal of Remote Sensing*, 6(8): 1335-1372. doi:10.1080/01431168508948283.
- She, X.J., Zhang, L.F., Cen, Y., Wu, T.X., Huang, C.P. and Baig, M.H.A., 2015. "Comparison of the Continuity of Vegetation Indices Derived from Landsat 8 OLI and Landsat 7 ETM+ Data among Different Vegetation Types." *Remote Sensing*, 7(10): 13485-13506. doi:10.3390/rs71013485.
- Snoek, J., Larochelle, H. and Adams, R.P., 2012. "Practical Bayesian optimization of machine learning algorithms." In: P. Bartlett, ed. *Advances in Neural Information Processing Systems 25*, Lake Tahoe, Nevada, USA, 2960-2968.
- Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., Patwary, M.M.A., Prabhat and Adams, R.P., 2015. "Scalable Bayesian optimization using deep neural networks." In: F. Bach & D. Blei, eds. *the 32nd International Conference on Machine Learning (ICML)*, Paris, 2171–2180.
- Sonobe, R., Tani, H., Wang, X., Kobayashi, N. and Shimamura, H., 2014. "Random forest classification of crop type using multi- temporal TerraSAR- X dual-polarimetric data." *Remote Sensing Letters*, 5(2): 157-164. doi:10.1080/2150704X.2014.889863.
- Sonobe, R., Tani, H., Wang, X., Kobayashi, N. and Shimamura, H., 2015a. "Discrimination of crop types with TerraSAR-X-derived information." *Physics and Chemistry of the Earth*, 83-84: 2-13. doi:10.1016/j.pce.2014.11.001.
- Sonobe, R., Tani, H., Wang, X., Kojima, Y. and Kobayashi, N., 2015b. "Extreme Learning Machine-based Crop Classification using ALOS/PALSAR Images." *Japan Agricultural Research Quarterly*, 49(4): 377-381. doi:10.6090/jarq.49.377.
- Sonobe, R. and Wang, Q., 2016. "Assessing the xanthophyll cycle in natural beech leaves with hyperspectral reflectance." *Functional Plant Biology*, 43(5): 438-447. doi:10.1071/FP15325.

- Sonobe, R. and Wang, Q., 2017. "Hyperspectral indices for quantifying leaf chlorophyll concentrations performed differently with different leaf types in deciduous forests." *Ecological Informatics*, 37: 1-9. doi:10.1016/j.ecoinf.2016.11.007.
- Tucker, C.J., 1979. "Red and photographic infrared linear combinations for monitoring vegetation." *Remote Sensing of Environment*, 8(2): 127–150. doi:10.1016/0034-4257(79)90013-0.
- Vogelmann, J.E., Gallant, A.L., Shi, H. and Zhu, Z., 2016. "Perspectives on monitoring gradual change across the continuity of Landsat sensors using time-series data." *Remote Sensing of Environment*, 185: 258-270. doi:10.1016/j.rse.2016.02.060.
- Wardlow, B.D. and 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(3): 1096-1116. doi:10.1016/j.rse.2007.07.019.
- Wulder, M.A., White, J.C., Goward, S.N., Masek, J.G., Irons, J.R., Herold, M., Cohen, W.B., Loveland, T.R. and Woodcock, C.E., 2008. "Landsat continuity: Issues and opportunities for land cover monitoring." *Remote Sensing of Environment*, 112(3): 955-969. doi:10.1016/j.rse.2007.07.004.

Tables

Table 1. Characteristics of satellite data

Date	Scene ID	Path	Row
20 May 2016	LC81070302016141LGN00	107	30
7 July 2016	LC81070302016189LGN00	107	30
24 August 2016	LC81070302016237LGN00	107	30

Table 2. Kauth-Thomas (KT) transform coefficients for Landsat 8 OLI.

Component	Blue	Green	Red	NIR	SWIR1	SWIR2
Brightness	0.3029	0.2786	0.4733	0.5599	0.5080	0.1872
Greenness	-0.2941	-0.2430	-0.5424	0.7276	0.0713	-0.1608
Wetness	0.1511	0.1973	0.3283	0.3407	-0.7117	-0.4559
Component 4	-0.8239	0.0849	0.4396	-0.0580	0.2013	-0.2773
Component 5	-0.3294	0.0557	0.1056	0.1855	-0.4349	0.8085
Component 6	0.1079	-0.9023	0.4119	0.0575	-0.0259	0.0252

Table 3. Accuracy results for the classifications

	Case 1	Case 2	Case 3	Case 4	Case 5
PA					
Beans	0.939	0.918	0.939	0.935	0.943
Beet	0.958	0.951	0.965	0.965	0.965
Grassland	0.938	0.881	0.938	0.931	0.938
Maize	0.700	0.675	0.750	0.750	0.750
Potato	0.964	0.933	0.959	0.964	0.964
Wheat	0.969	0.958	0.975	0.975	0.975
UA					
Beans	0.913	0.879	0.924	0.920	0.928
Beet	0.971	0.944	0.979	0.986	0.979
Grassland	0.955	0.953	0.974	0.974	0.974
Maize	0.778	0.667	0.759	0.750	0.779
Potato	0.913	0.910	0.930	0.931	0.926
Wheat	0.983	0.974	0.978	0.978	0.978
OA	0.938	0.915	0.943	0.942	0.945
AD	4.661	6.949	4.831	4.831	4.492
QD	1.525	1.525	0.847	0.932	1.017

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

Table 4. McNemar's test performed on four combinations.

	Case 1	Case 2	Case 3	Case 4	Case 5
Case 1		14.96	14.08	15.79	10.44
Case 2			20.11	18.29	21.30
Case 3				2.67	4.33
Case 4					7.00
Case 5					

Note: A χ^2 value ≥ 3.84 indicates a significant difference between two datasets in the classification results ($p < 0.05$).

Table 5. McNemar's test results obtained between RF modified by Ishwaran (Ishwaran and Kogalur, 2007, Ishwaran et al., 2008) and traditional RF.

Case 1	Case 2	Case 3	Case 4	Case 5
6.13	10.48	5.00	11.29	4.00

Note: A χ^2 value ≥ 3.84 indicates a significant difference between two datasets in the classification results ($p < 0.05$).

Figures

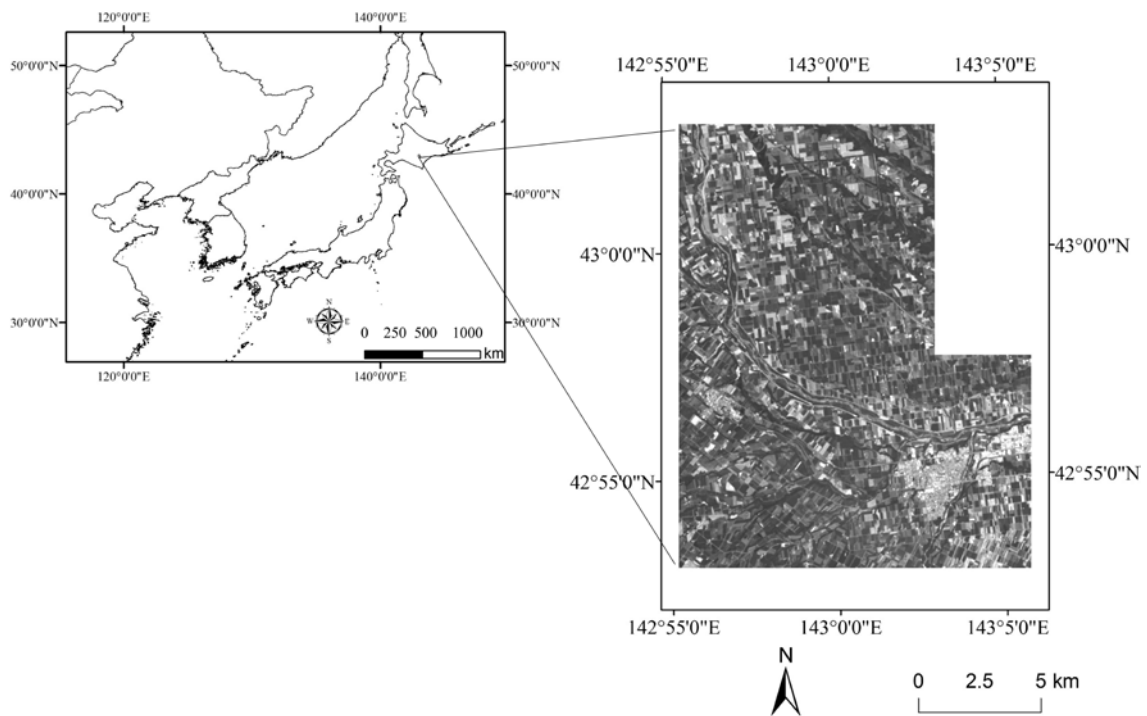


Figure 1. The study area (background map shows Landsat 8 OLI data obtained on 7 July 2016).

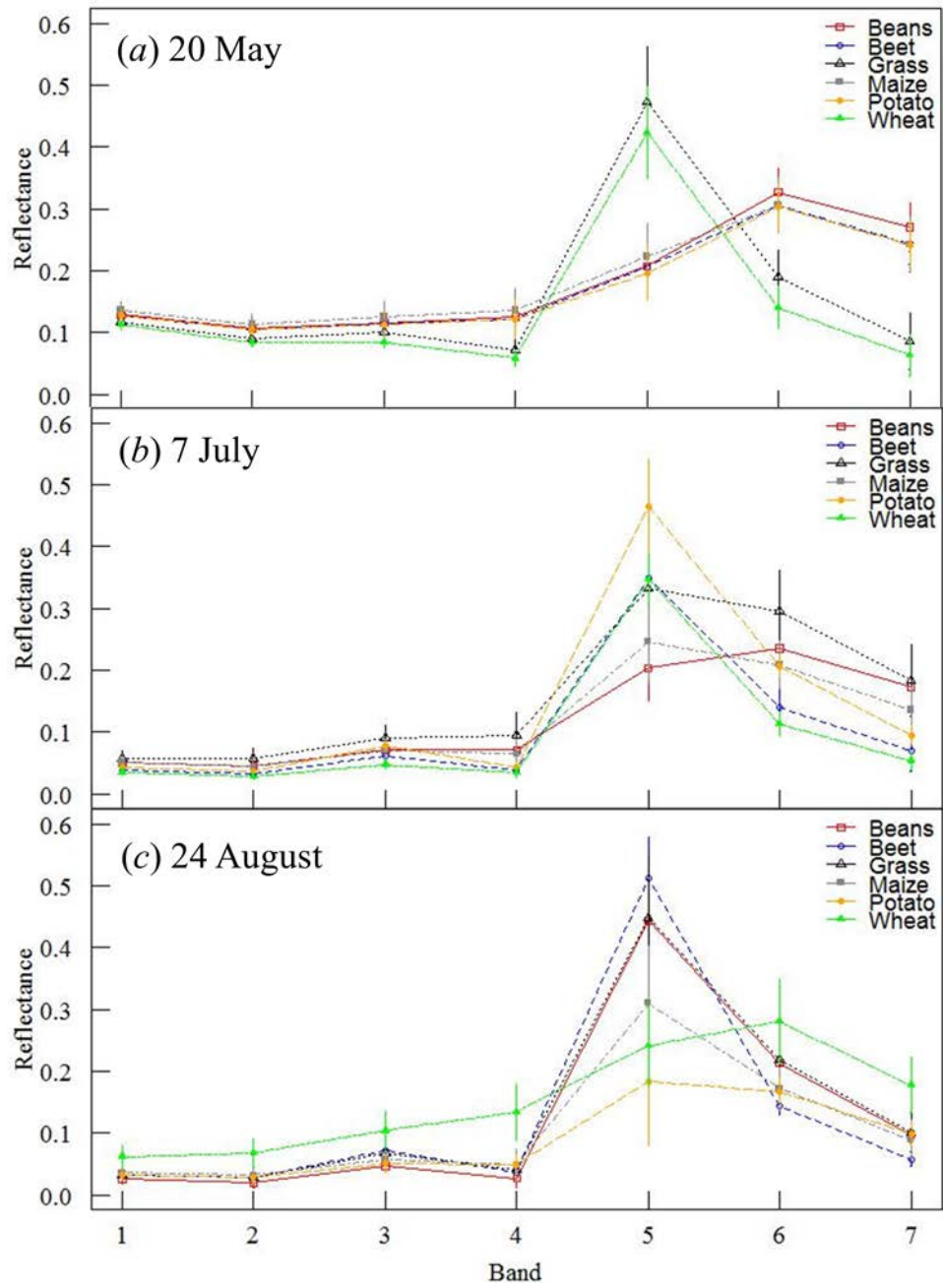


Figure 2. Mean reflectance spectra and standard deviations (a) on 20 May, (b) on 7 July, and (c) on 24 August.

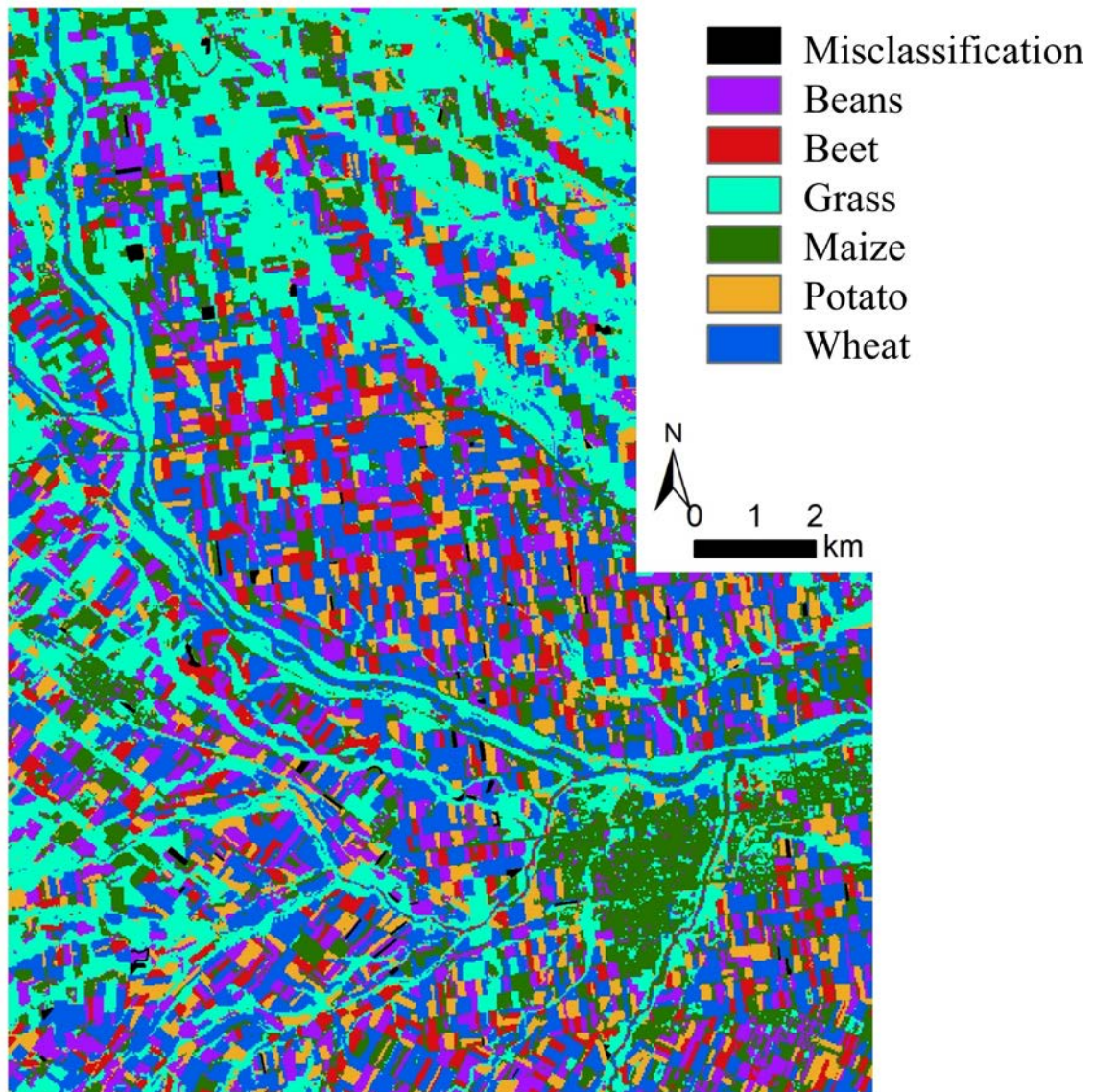


Figure 3. Crop classification map with misclassified fields printed in black.

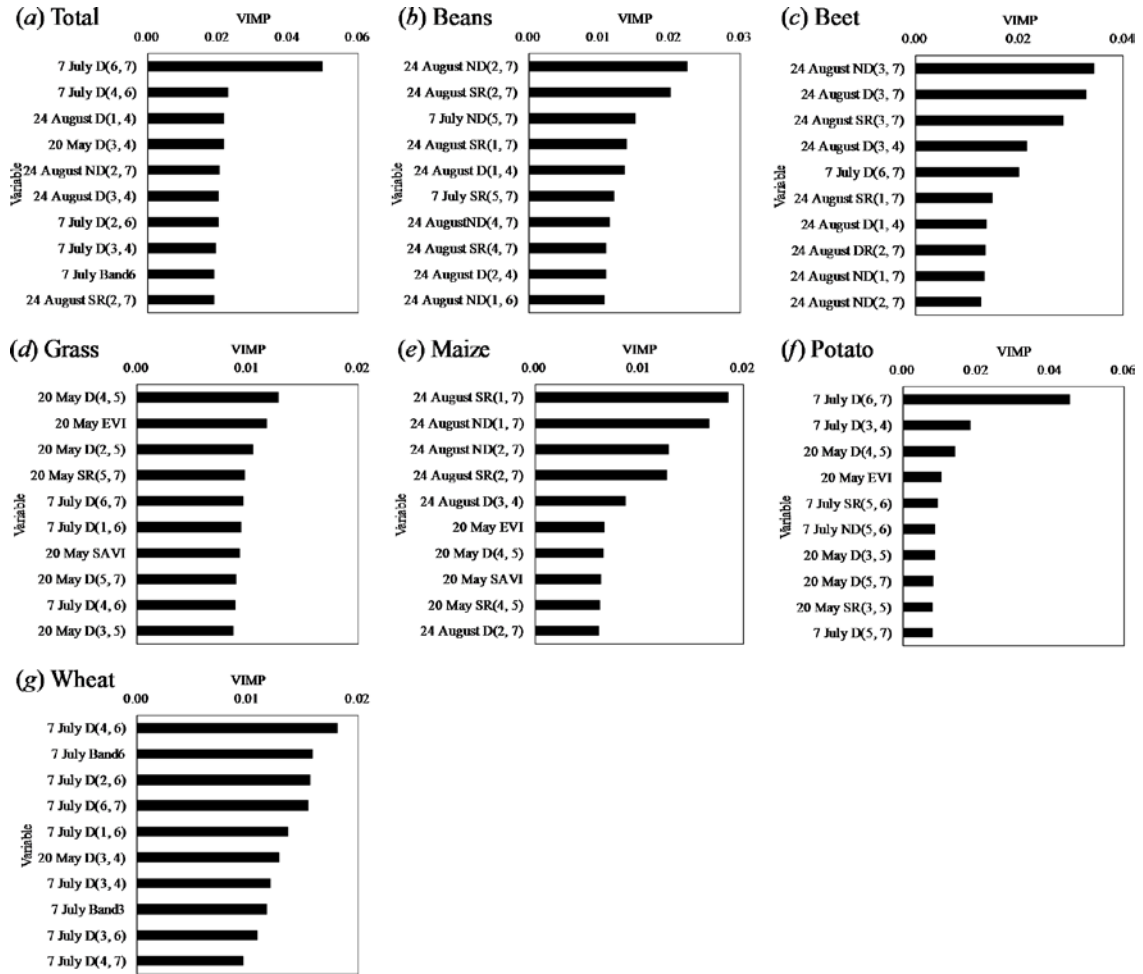


Figure 4. Variable importance of the top 10 useful variables (Case 5) for identification of each crop; (a) totally, (b) beans, (c) beet, (d) grass, (e) maize, (f) potato and (g) wheat.

VIMP: variable importance, which is calculated from training data (Ishwaran, 2007).

$$D(\lambda_1, \lambda_2) = R_{\lambda_1} - R_{\lambda_2}, SR(\lambda_1, \lambda_2) = R_{\lambda_1}/R_{\lambda_2} \text{ and } ND(\lambda_1, \lambda_2) = (R_{\lambda_1} - R_{\lambda_2})/(R_{\lambda_1} + R_{\lambda_2})$$

where R is reflectance at a given Landsat band number (λ_1 or λ_2).