Computers and Electronics in Agriculture 115 (2015) 171–179

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data

Kenichi Tatsumi [a,*], Yosuke Yamashiki [b], Miguel Angel Canales Torres [c], Cayo Leonidas Ramos Taipe [c]

[a] Department of Environmental and Agricultural Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan
[b] Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University, Kyoto, Japan
[c] Departamento de Recursos Hídricos, Universidad Nacional Agraria La Molina, La Molina, Peru

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form 16 February 2015
Accepted 2 May 2015
Available online 14 June 2015

Keywords:
Crop classification
Upland field
Random forest
Landsat 7 ETM+
Enhanced vegetation index

ABSTRACT

Crop classification of homogeneous landscapes and phenology is a common requirement to estimate land cover mapping, monitoring, and land use categories accurately. In recent missions, classification methods using medium or high spatial resolution data, which are multi-temporal with multiple frequencies, have become more attractive. A new mode of incorporating spatial and temporal dependence in a homogeneous region was tried using the Random Forest (RF) classifier for crop classification. A time-series of medium spatial resolution enhanced vegetation index (EVI) and its summary statistics obtained from Landsat 7 Enhanced Thematic Mapper Plus (Landsat 7 ETM+) were used to develop a new technique for crop type classification. Eight classes were studied: alfalfa, asparagus, avocado, cotton, grape, maize, mango, and tomato. Evaluation was based on several criteria: sensitivity to training dataset size, the number of variables, and mapping accuracy. Results showed that the training dataset size strongly affects the classifier accuracy, but if the training data increase, the rate of improvement decreases. The RF algorithm yielded overall accuracy of 81% and a Kappa statistic of 0.70, indicating high model performance. Additionally, the variable importance measures demonstrated that the mode and sum of EVI had extremely important variables for crop class separability. RF had computationally good performance. They can be enhanced by choosing an appropriate classifier for multiple statistics and time-series of Landsat imagery. It might be more economical to use no-cost imaging for crop classification using open-source software.

1. Introduction

Earth observation satellites are indispensable for the estimation of crop classification and land-cover monitoring (Asner et al., 2002; Jakubauskas et al., 2002; Rodriguez-Galiano et al., 2012a). Large-scale information related to the spatial distribution of crop types is critical to many applications, from crop modeling and management and estimation of cultivated areas (Loveland, 1991; Mkhabela et al., 2011). Satellite image data have been used more widely than airborne imagery for large-scale crop classification because of their synoptic scale. As described above, classification and area mapping of crop types using image data from satellite sensors have been performed. Several researchers have also applied single-date and multi-date multispectral imagery for profiles of crop phenology and classification (Ryerson et al., 1985; Panigrahy and Sharma, 1997; Oetter et al., 2000; Murakami et al., 2001; Simonneaux et al., 2008; Yang et al., 2011; Mellor et al., 2013). Multi-date, high-resolution, and multi-spectral bands generally yield better classification for crops than single and few spectral signatures, especially similar spectral signatures.

In Peru, GDP grew 42% during 2002–2011. Agriculture accounted for 7% of total GDP in 2011. Peru is also a leading exporter of organic products such as asparagus, coffee, and bananas. However, environmental degradation related to El Niño/La Niña, heavy rain, drought, decrease of groundwater, salt damage, acidification of soil and desertification hinders stable productivity. In considering the national export policy, management and understanding of crop items are important in terms of the crop price, supply, and quality. Furthermore, a direct and strong correlation exists between the change in the cultivated area, crop varieties and crop supply and demand, cultivation technique, and the number of producers in Peru. For this reason, crop classification

* Corresponding author at: Department of Environmental and Agricultural Engineering, Tokyo University of Agriculture and Technology, 3-5-8 Saiwai-cho, Fuchu, Tokyo 183-8509, Japan. Tel./fax: +81 42 367 5679.
E-mail address: ktatsumi@cc.tuat.ac.jp (K. Tatsumi).

http://dx.doi.org/10.1016/j.compag.2015.05.001
0168-1699/

mapping development is an urgent necessity for labor-saving of a field survey related to land use information and grasping cropping information easily and widely. However, several limitations exist in relation to crop classification that must be resolved. First, separating one crop from another is difficult because of variations in moisture, elevation, temperature, soil properties, fertilization, irrigation, planting dates, and tillage practices. Second, many limitations of crop classification are related to the similarity of reflectance of upland field crops, and field-to-field spatial and spectral variations, which are different from the patterns of individual crop phenology. Third, crop classification requires methods that can be interpreted readily and which can be simplified and operated easily with user-defined parameters that are automated to adjust in practice.

An examination of crop classification methods revealed that many available methods have used remotely sensed image data. Traditional unsupervised methods such as ISODATA or K-means, parametric supervised algorithms such as maximum likelihood, machine learning algorithms such as artificial neural networks, support vector machines, decision trees and ensembles of classifiers have been applied to land cover using remote sensing datasets (Foody, 2004; Lippitt et al., 2008; Mathur and Foody, 2008; Rogan et al., 2008; Guerrero et al., 2012; Rodriguez-Galiano et al., 2012a,b). These algorithms were evaluated using multiple remote sensing data and ancillary data from many crop-growing environments. They are effective because they are independent of data distribution assumptions, which can improve classification accuracy. Machine learning techniques, which use ensembles of classifications, have been estimated for many applications recently (Wang et al., 2004; Yang et al., 2011; Zhang et al., 2013; Cracknell and Reading, 2014). Based on the background presented above, Random Forest (RF) classifiers increasingly provide a new means to predict land-cover classification maps that are robust to variations in class reflectance caused by land use or disturbances in regional scale mapping. The RF classifiers are ensemble decision trees developed in the field of machine learning. The classifiers combine bootstrap sampling to construct a large set of classification of individual decision trees (Breiman, 2001; Rodriguez-Galiano et al., 2012a,b; Mellor et al., 2013). In fact, RF has high accuracy for land-cover classification across heterogeneous landscapes. It is more sensitive to noise than the other classification methods (Pal, 2005; Waske and Braun, 2009; Oliveira et al., 2012; Rodriguez-Galiano et al., 2012a,b). Moreover, RF classifier runs efficiently with huge datasets. Recent studies showed that RF can incorporate multiple variables of remote sensing with categorical land use data to improve classification performance and to discriminate between forests and other ground cover (Lawrence et al., 2006; Martinuzzi et al., 2009; Ghimire et al., 2010; Latifi et al., 2010; Guo et al., 2011; Oliveira et al., 2012; Rodriguez-Galiano et al., 2012a,b). In addition, topographic (e.g., elevation, slope and aspect) and bioclimatic (e.g., temperature, precipitation) variables used in combination with spectral image data have been demonstrated to enhance forest/non-forest, habitat and vegetation classification (Franklin, 1995; Fahsi et al., 2000; Joy et al., 2003; Gislason et al., 2006; Sesnie et al., 2008).

Most reports of previous work related to crop identification primarily describe land-cover categories (e.g., urban, grass, water, forest, and crops). Additionally, multi-source remote sensing and ancillary (e.g., topographic, bioclimate) data, which are big data, could also be used to discriminate forest/non-forest using the RF classifier. Few studies have examined the use of multi-date imagery for upland field crops. In this study, to conduct highly accurate and simple analysis related to crop classification, we specifically examined the temporal profiles of crop phenology as attested in the Enhanced Vegetation index (EVI) and its summary statistic obtainable from Landsat 7 ETM+ 30 m resolution data. The objectives of this study were (1) to evaluate the functional performance and availability of RF for classifying eight upland field crops in medium homogeneous study areas of about 23,000 ha using predictor variables obtained solely from Landsat 7 ETM+ across the Ica region in Peru; and (2) to assess the effectiveness and efficiency of the RF classifier for crop classification using open-source software, which is Geographic Resources Analysis Support System (GRASS) (GRASS Development Team, 2012), R 3.1.0 (R Development Core Team, 2011) and freely available data, thereby avoiding the use of ancillary and high-cost data.

2. Materials and methods

2.1. Study site

The Ica region, located in southern Peru at approximately 14°S, 75.7°W (Fig. 1), is the target area of this study. Agriculture is conducted on the flat Ica landscape, which dominates the southwestern parts of this region, with circumjacent barren areas and mountains. Agriculture in the region relies on an aquifer fed by glacial melt water. The aquifer is being rapidly depleted, leading to calls for more efficient irrigation, or addition of reservoirs. Cropland areas occupy about 23,000 ha, with elevations extending from sea level to 500 m. The eastern part of this region has topography that is too rugged for agriculture. The Ica region tends to a desert climate. Temperatures are hot during the summer months and warm during winter months. The main crop classes are cotton, grapes, asparagus, maize, and tomato. The following eight classes were included in this study: alfalfa, asparagus, avocado, cotton, grape, maize, mango and tomato.

2.2. Landsat time-series data

The database used for this study consists of a four-year time series of optical data from Landsat 7 ETM+. Multi-time-series Landsat 7 ETM+ are used to characterize phenological variation in the state of farming crops. Satellite sensor images of Ica, captured by Landsat-7 of worldwide reference system (WRS) during 2008–2011 (path 6, row 70, total 53 scenes), with WGS84 datum, projection UTM18S, have been used for crop classification analyses. Level processing of these data is the Landsat Terrain Corrected Product (L1T), which uses ground control and relief models to attain absolute geodetic accuracy. These data are co-registered and orthorectified. Several studies have demonstrated that a combination of multi-time series images can increase the distinction between spectrally similar covers representing the phenological vegetation condition (Lunetta and Balogh, 1999; Yuan et al., 2005). For these data, we (1) transformed the calibrated digital number (*DN*) of Landsat 7 ETM+ products to at-surface reflectance with simplified atmospheric correction (Tizado, 2013) and (2) removed cloud effects.

Radiometric calibration converts *DN* to at-sensor radiance (*RAD*), *RAD* is defined as

$$RAD = (DN - DN_{\min}) * \frac{(L_{\max} - L_{\min})}{(DN_{\max} - DN_{\min})} + L_{\min} \quad (1)$$

where *RAD* stands for the at-sensor radiance ($W/(m^2 * sr * \mu m)$), $L_{\max}$ and $L_{\min}$ are the calibration constants which is described the Landsat metadata file, and $DN_{\max}$ and $DN_{\min}$ respectively denote the highest and the lowest points of the range of rescaled radiance in a calibrated digital number. The at-surface reflectance REF_{surf} is defined as:

$$S_{\text{rad}} = e^{\left[\frac{t}{\cos(sat_{\text{zenith}})}\right]} * \left[E_{\text{SUN}} * \sin(e) * e^{\left[\frac{t}{\sin(e)}\right]} + \pi * RAD_{\text{dark}}\right] / (\pi * d^2) \quad (2)$$

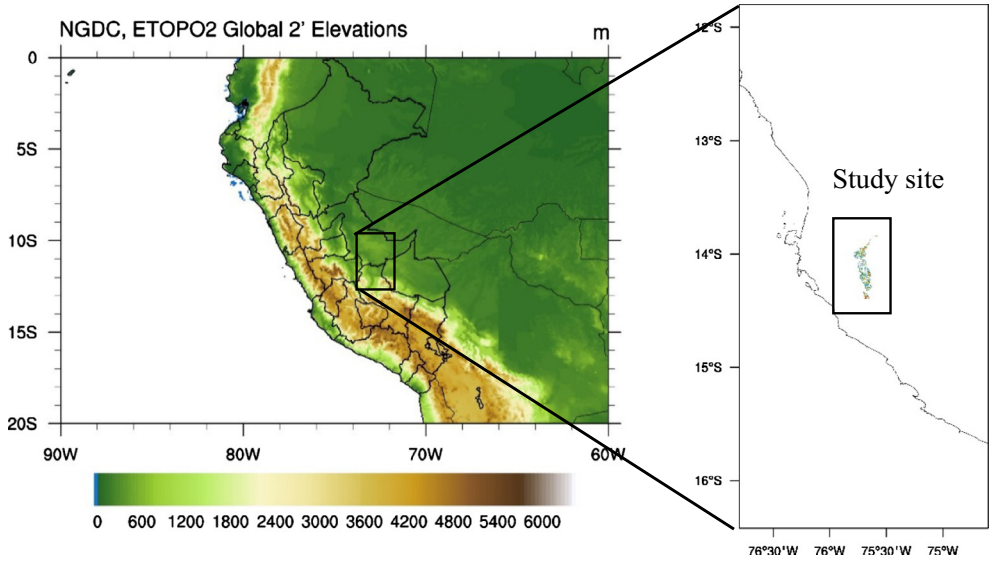

Fig. 1. Study area location.

$$REF_{surf} = (RAD - RAD_{dark} + 0.01 * S_{rad})/S_{rad} \quad (3)$$

where S_{rad} represents the sun radiance, sat_{zenith} denotes the satellite zenith angle, and E_{sun} signifies the mean solar exoatmospheric irradiance ($W/(m^2 * \mu m)$). Furthermore, d is the earth sun distance in astronomical units, e stands for the solar elevation angle, and RAD_{dark} signifies the at-sensor radiance calculated from the darkest object that shows DN with at least 1000 pixels for the entire image. Second, cloud-cover effects are removed from images using an automated cloud-cover assessment algorithm from Irish (2000) with constant values for the pass filter algorithm from Irish et al. (2006). The present study does not use cloud pixels.

The corrected Landsat 7 ETM+ showed a failure of the Scan Line Corrector (SLC) on May 31, 2003. Therefore, without an operating SLC, the ETM+ line of sight now traces a zigzag pattern along the satellite ground track. Consequently, the imaged area is duplicated, with width that increases toward the scene edge. Since that time, images have had wedge-shaped gaps on both sides of each scene, resulting in an estimated 22% of any given scene being lost (USGS, 2013). Therefore, the portion including the data loss is not used in the analysis.

To produce variables for the RF classifier model, EVI is first produced using the Landsat 7 ETM+ surface reflectance band 1 (blue), 3 (red), and 4 (near infrared). The enhanced vegetation index (EVI) is commonly used in studies using remote sensing data because it is optimized to enhance the vegetation signal, yielding improved vegetation monitoring and improved sensitivity in high-biomass regions (deFries et al., 1995; Liu and Huete, 1995). Moreover, EVI can reduce the influences of canopy background and atmospheric variation compared to NDVI (Huete et al., 2002). We proceed by determining the EVI values for each Landsat-7 ETM+ scene with the following:

$$EVI = 2.5 \times \frac{NIR - R}{NIR + 6.0 \times R - 7.5 \times B + 1} \quad (4)$$

Therein, NIR, R, and B respectively denote atmospherically corrected surface reflectance in near-infrared, red, and blue bands. The images are enhanced spectrally using multi-date EVI values. In this study, we make each output cell value a function of the values assigned to the corresponding cells in the input raster map layers for 2008–2011. This procedure produced 7 statistic features: average value (average), most frequently occurring value (mode), lowest/highest value range of values (max/min), sum of values (sum), statistical variance (variance), number of different values within the time-series data (diversity). These are calculated for one-month periods during the calendar year. The "r.series" module in the GRASS is used to produce these raster datasets, and obtained datasets are used as input variables in RF classification.

2.3. Reference data

Land cover maps were obtained from a combination of digital aerial photographic interpretations of color aerial photographs and a detailed field survey of the ground across the study area acquired during 2008–2011 (Fig. 5a). The field survey was conducted by La Molina National Agrarian University. The created polygon data were converted to raster format to align with the 30 m × 30 m pixels of Landsat satellite imagery. These reference data were used for classification and validation procedures.

2.4. Random forest approach

An ensemble classification algorithm, RF, consists of a group of tree-based classifiers $\{h(x, \Theta_k, k = 1,\ldots)\}$, where x is the input vector and Θ_k are independent and identically distributed random vectors (Breiman, 2001; Hastie et al., 2009). RF uses bootstrapping with replacement to enhance the diversity of classification trees, which allocate each pixel to a class in accordance with the maximum number of votes from the collection of trees. This method, although it has shown high accuracy and ability to model complex interactions among variables, is a "black-box" because the individual trees cannot be estimated separately (Prasad et al., 2006).

To run the RF model, it was necessary to define several important adjustable parameters. The primary parameters are the number of predictors at each decision tree node split (*mtry*) and the number of decision trees to run (*ntree*). Liaw and Wiener (2002) report that $mtry = 1$ can give good performance. Rodriguez-Galiano et al. (2012a) showed that reducing *mtry* weakens each tree of the model, but it also reduces the correlation among individual trees, which increases the model accuracy. Oliveira et al. (2012) reported that an increase in values of *mtry* would result in a higher predictive performance of the model and attribution of higher importance to fewer variables. In consideration of these points, it is necessary to optimize the parameters *mtry* and *ntree* to maximize the model accuracy.

First, to evaluate the model performance, all data were divided with stratified random sampling ranging from 10% (11,781 pixels) to 90% (105,994 pixels) in increments of 10% for test data, left out

of the training data. The set of test data, which is an independent validation set, was used only for the model evaluation. Moreover, the remaining training dataset was divided to 75% (training dataset) and 25% (validation dataset) for the sake of a repeated leave-group out cross-validation (LGOCV) strategy. This procedure is repeated 10 times to estimate robust prediction performance. Each datum of the validation data and test data is used to compute accuracies and error rates averaged over all predictions and to estimate each variable's importance in the classification.

To reduce data redundancy and to assist the model interpretation and the absolute values of pairwise correlation coefficients were considered. Predictors with near zero-variance values were removed. If two variables are highly correlated (>0.75), then, the variable with the largest mean absolute correlation is automatically removed from the model.

The kappa coefficient (Cohen, 1960), and the producer and user accuracy are calculated to evaluate the crop identification performance. The kappa coefficient, a statistical measure of inter-rater reliability, is calculated as follows:

$$\text{kappa} = \frac{(P(a) - P(e))}{1 - P(e)} \tag{5}$$

Therein, $P(a)$ denotes the overall percent agreement, which represents the relative observed agreement fraction. $P(e)$ is the hypothetical probability of change agreement, which stands for the expected probability fraction between observed data and the RF predictions. Therefore, kappa = 1 shows that the raters are in complete agreement; kappa = 0 indicates no agreement among raters. Producer accuracy is the proportion of a crop class on the reference ground that is classified correctly in the field. The user accuracy is the proportion of a predicted class on a map, which matches the corresponding class on the reference ground. Producer accuracy stands for the share of ground data that are consistent with classification results obtained using RF prediction, whereas user accuracy measures the percentage of classification results that are classified correctly (see Table 2 footnote). Moreover, the selection of the most relevant variables to include the final RF model is done by ranking the variables according to their importance in all samples. The Random Forest package is included in a statistics package (R 3.1.0; R Development Core Team, 2011) that is an open source language and software environment. It is used for statistical computing.

3. Results and discussion

3.1. Effects of the number of predictor variables, and training data on classification accuracy

The collection of ground truth information in the target area to train the model classifier is a difficult and time-consuming task that is also expensive, especially when processing numerous and similar phenologies of categories, some of which have high intra-class variation. The largest number of training areas possible must be used to represent the entire variation in a category (Pal and Mather, 2003). Nevertheless, it is possible to design an optimum sampling scheme that can operate rapidly and economically with an acceptable classification accuracy level (Lippitt et al., 2008; Rogan et al., 2008; Rodriguez-Galiano et al., 2012a,b). This study examines eight crop species that have high variation and similar phenology. Consequently, it is necessary to use numerous training datasets to construct the RF classifier (Rodriguez-Galiano et al., 2012a,b). The effects of the number of predictor variables (p) (*mtry*), and training data on classification accuracy were evaluated using the overall accuracy (see Table 2 footnote) and kappa coefficient in the classification of the training data as altering the training sets in increments of 10%, from 10% to 90%.

Table 1 shows the relation accuracy and *mtry* equal to the 2, 7 (sqrt(p)), 25, 49 (maximum p), respectively. It showed that the model accuracy depends on *mtry* to some extent. Accuracy results across tuning *mtry* parameters from LGOCV reached a maximum level when *mtry* equals 25. However, these results showed classification accuracy could not be significantly affected by a change in *mtry*

Fig. 2 shows accuracy/kappa when sum of the training and validation data were 11,781 (10% of all dataset) to 105,994 (90% of all dataset). For LGOCV, the training percentage was set to 75% to construct the RF model, with intervals of about 11,700 and *mtry* = 25. As a boundary about 30,000 training sample sizes, the rate of improvement of accuracy/kappa is different. From the 30% threshold, accuracy decreases more abruptly to attain kappa equal to 0.45 (Fig. 2). The ranking of the *mtry* variables in terms of importance did not change significantly with different *mtry*. These results are consistent with those presented by Cutler et al. (2007) and Oliveira et al. (2012), and the final number of variables using RF model may be not important in an improvement performance compared to the number of training dataset in this case. After all, larger numbers of training datasets will increase the accuracy/kappa, but the rate of improvement is not constant.

Based on the evaluation presented above, the final models were bolstered with the 49 predictor variables to estimate many variable importance, and the training data are 79,499 pixel points. *ntree* initial parameter was set to 100 to produce stable results because a tree number of 30 or less strongly affects the classifier accuracy. The steady state is reached at 100 trees or more. Since, the number of *ntree* eventually converge in model, it is not need to be a parameter study.

3.2. Variable importance

RF algorithms accommodate two features of the importance of the variables by the mean decrease in accuracy (MDA) and the mean decrease in Gini index (MDG). Fig. 3 shows the respective contributions of predictor variables to the RF final classification model generated with respect to the monthly time-series EVI variables in terms of MDA and MDG in accuracy, respectively. According to MDA and MDG, the average, mode, sum, and variance of monthly EVI series were shown to be more important variables, with values greater than 0.02 and 2000, respectively. On the other hand, the contributions of max, min, and diversity in classifying the data are less than the others. Especially, averages of February and April, mode and sum of May showed a greater contribution to the RF model accuracy. Higher values of MDA and MDG indicate variables that are more important to the classification the data. Though, it is an important to reduce the initial number of variables with respects to computational costs and efficiency, it should be careful not to have either too few variables or too many variables.

Fig. 4 shows the MDA variable importance measures of each upland crops for the RF classification. It showed that the overall trend related to variable importance is the same as the overall contribution (Figs. 3 and 4). As the overall trend, when the mode and sum have not been used for RF classifier, classification accuracy significantly decreases. Therefore, for each crop, the mode and sum obtained from the EVI time-series have great importance in the class classification of field crops. Especially, the mode and sum of May and July for all crops have a greater importance for the classification. In this period, all crops are under development stage, mid-season stage, and late season stage, and the mode and sum of EVI could have high sensitivity to vegetation signal due to dense vegetation cover.

Table 1
Effects of training set size and the number of variables on classification accuracy.

	Training sample size								
mtry	8839	17,671	26,503	35,336	44,167	53,000	61,833	70,667	79,499
2	0.656	0.681	0.702	0.715	0.726	0.735	0.744	0.753	0.757
7	0.666	0.692	0.715	0.727	0.739	0.748	0.757	0.766	0.771
25	0.678	0.705	0.730	0.743	0.755	0.765	0.774	0.784	0.789
49	0.673	0.700	0.725	0.740	0.750	0.761	0.770	0.778	0.783

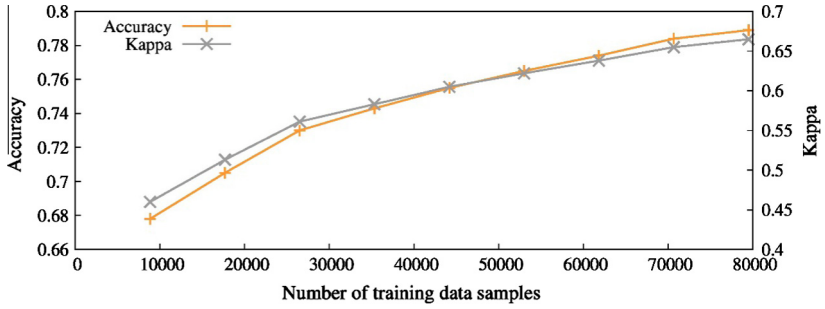

Fig. 2. Variation of accuracy and kappa index with increasing number of training data sizes.

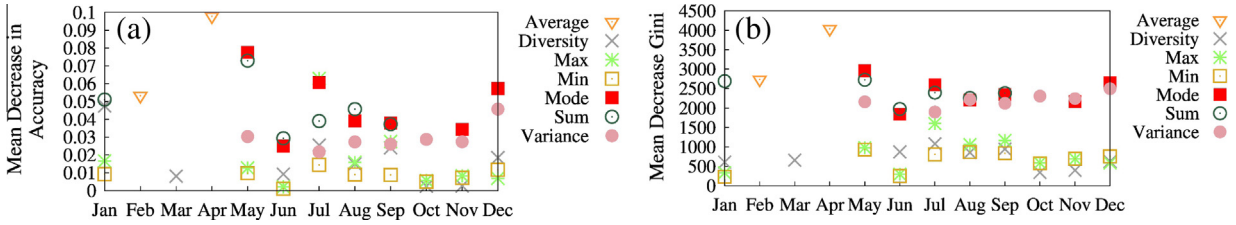

Fig. 3. Average of all crops variable importance with respect to summary statistics of EVI in terms of (a) the mean decrease in accuracy and (b) mean decrease in Gini index.

Peru is the second largest asparagus production country in the world, and the cultivated area in this region is about 11,605 ha. Though asparagus grows year-round in this region, it generally begin to grow from early May after dormant stage (January to March). Therefore, the explanatory variable during winter season may relatively contribute to increased identification of asparagus cultivated area. However, since the Landsat data in the study area were missing through January to March, the importance of contribution to the classification in these months will be future tasks. Avocado, a crop of great economic importance, is being produced increasingly in recent years. The yield in the study area was 9.7 t/ha in 2008, 9.8 t/ha in 2009, 14.2 t/ha in 2010, and 13.2 t/ha in 2011 as reported by La Molina National Agrarian University. Harvest of avocado is mainly performed in winter season (Agro Peruano, 2012), therefore, dense crop cover provided an important predictor variables. However, why the lower values of mean decrease in accuracy in June was unclear. Even though we cannot conclude that important variable obtained from this study is also applicable under other region, it is important to take into account as many variables, which may be the agronomic important.

3.3. Classification accuracy

Fig. 5 presents classification results of the ground truth (a) and those obtained using RF (b). Visual assessment of the classification results shows good performance by RF classifier ensembles, and shows the general structures of the study area. However, classification results obtained using RF include noise even in homogeneous areas. The main reason for the higher accuracy is certainly the underlying assumption of classifier ensembles. The generation of independent classifiers and the performance is influenced directly by this independence. A random selection of input variables features seems well suited for multi-temporal approaches. Moreover, the ensemble performance is increased further by combining this feature selection with random selection (Waske and Braun, 2009).

Error matrices were used to assess accuracy assessment for crop type classification using the RF model. Table 2 shows the accuracy assessment for crop type mapping obtained through the RF final model using an independent sample (test data) of 11,773 pixels. The results of the accuracy assessment through the RF model confirm the generally good performance of classifier ensembles considering the large number of farming crops. The overall accuracy is 81%, with a kappa coefficient of 0.70 for test data. These results showed an almost complete match according to Landis (1977).

User and producer accuracy values of individual classes were, respectively, 89% and 60% on average. However, producer accuracy was lower for alfalfa (44%), avocado (39%), and tomato (39%) classes. In other words, the classification map missed 56%, 61%, and 61% (omission error) of the alfalfa, avocado, and tomato areas on the ground, indicating a tendency for the model to misclassify alfalfa, avocado, and tomato as cotton (11/34, 231/467, 98/222). Therefore, this high omission error was attributable mainly to spectral similarities between alfalfa, avocado, and cotton in fields. Therefore, we think that alfalfa and avocado were typically misclassified within cotton. However, the misclassification between alfalfa and avocado is unclear. Asparagus, grape, maize, and mango were also misclassified within cotton. In this study, the cultivated area of cotton accounts for 50% of the total cultivated area. The mode properties of cotton show similar characteristics to those of other crops. Therefore, crops with a small cultivated area compared to cotton are likely to be misclassified as cotton in pixels. Waske and Braun (2009) show that the different crop types

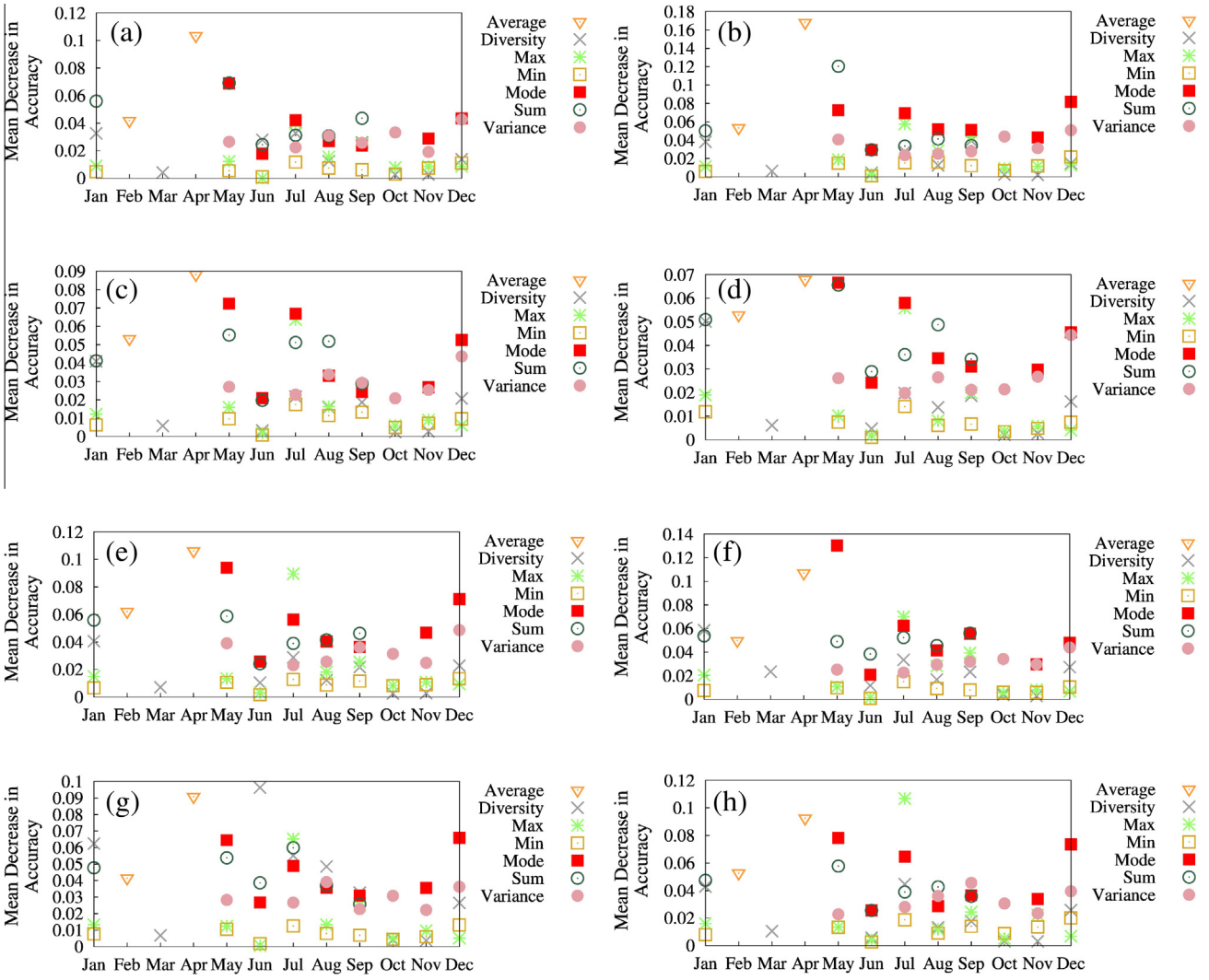

Fig. 4. Variable importance with respect to summary statistics of EVI in terms of the mean decrease in accuracy (a) alfalfa, (b) asparagus, (c) avocado, (d) cotton, (e) grape, (f) maize, (g) mango and (f) tomato.

(cereals, orchards, rapeseed and root crop) and complexity of the classes (e.g., orchards, urban) still yield accurate and stable results, which is extremely promising. Results of our study show that a certain problem remains in accuracy for upland field crops, which has similar phenology and summary statistics.

However, the user accuracy is a rigorous independent validation for all crops (everywhere > 76%). User accuracies were excellent for alfalfa, asparagus, avocado, mango, and tomato (94%, 91%, 95%, 90%, and 98%, respectively) and very good for grape and maize (86% and 83%, respectively). The producer accuracy for cotton was excellent (98%), but the user accuracy was only 76%. In fact, 135/5854 cotton pixels on ground truth were misclassified as other crops (omission error), whereas 1796/7515 cotton pixels stratified on the classification map belonged to the other seven classes (commission error). However, results show that the classification result of the models is correct in general, and that the reliability of the image classification is high. Apart from a few exceptions, results indicate that application of the RF model for agricultural classes as their areal quantification are advancing and might demonstrate the classifier's utility in crop type mapping management actions and yield forecasts.

In this study, the highest producer accuracy was accomplished for cotton (98%), which in the Ica region occupies the largest cultivated area. The lowest producer accuracy was attained for avocado and tomato. It occurred in a small cultivated area, which suggests that crop classification accuracy increases as the validation data become numerous. The categories most difficult to classify were those with the same variable importance ranks with cotton in topside, such as alfalfa and avocado. Our results show something about the findings of other studies using RF, that this ensemble RF classifier is useful to learn multiple crop cover types. In this study, the spatial resolution of the crop fields is the same as the image sensor resolution. Alfalfa, avocado, and tomato have lower producer's classification accuracy and occur in a small cultivated area, but a completely mixel-free classification is unlikely at the field boundary. Therefore, methods applying much finer spatial resolution (e.g., IKONOS, GeoEye, WorldView and SAR data) might greatly enhance the accuracy. Finally, the RF classifier model of EVI time-series data shows considerable promise as a tool for crop-type identification monitoring. Especially, changes in the mode and sum of EVI time-series are indicative of significant land surface changes.

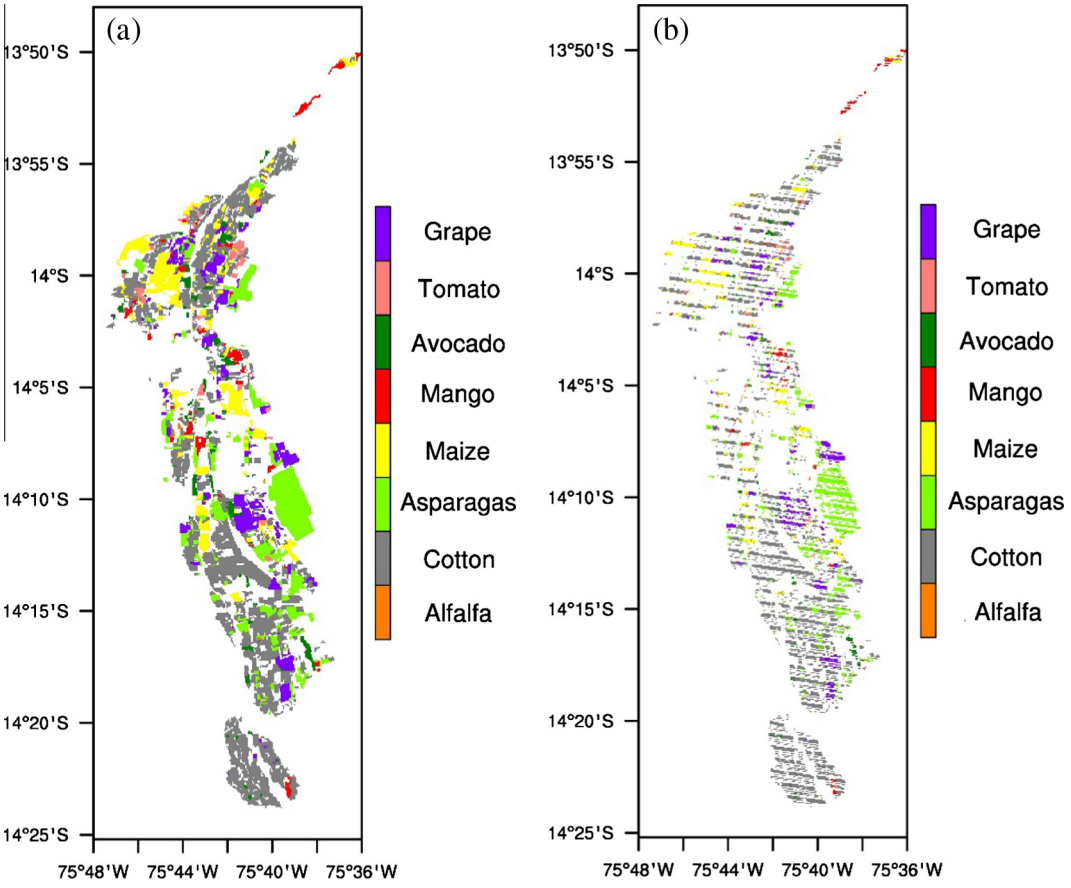

Fig. 5. Classification results maps generated from the Landsat 7 ETM+ image dataset and the corresponding ground truth (a), using RF (b) (diagonal lines of (b) is missed by SLC).

Table 2
Error matrix and accuracy measures based on the RF final model.

Predicted class	Reference class									
	Alfalfa	Asparagus	Avocado	Cotton	Grape	Maize	Mango	Tomato	Total	User accuracy%
Alfalfa	15	0	0	0	0	1	0	0	16	94[a]
Asparagus	1	1733	11	57	51	24	14	4	1895	91
Avocado	0	1	183	4	1	2	0	1	192	95
Cotton	11	494	231	5719	417	402	143	98	7515	76
Grape	1	41	20	29	800	15	4	17	927	86
Maize	2	31	19	37	30	726	12	14	871	83
Mango	4	1	3	7	4	6	242	1	268	90
Tomato	0	0	0	1	0	1	0	87	89	98
Total	34	2301	467	5854	1303	1177	415	222	11,773	
Producer accuracy %	44[a]	75	39	98	61	62	58	39		81[b]

[a] Producer's accuracy of alfalfa = 15/34 = 44%, and User's accuracy of alfalfa = 15/16 = 94%.
[b] Overall accuracy = (15 + 1773 + 183 + 5719 + 800 + 726 + 242 + 87)/11773 = 81%.

4. Conclusions

This study was undertaken to investigate the applicability of the RF classifier model for crop cover identification in a geomorphological homogeneous crop area (23,441 ha) in Ica, Peru. We incorporated a suite of multi-temporal Landsat 7 ETM+ satellite imagery variables into an RF model. The RF performed well in the classifications with eight upland field crops. Moreover, we examined the RF accuracy with the number of variables, training dataset sizes. The number of variables has little effect on the classifier accuracy. The more the training data are increased, the higher the classifier accuracy becomes. Therefore, there is less need for tuning the model parameters. The following are summaries of the findings.

1. RF requires only three user-defined parameters to be set: the number of trees, the number of random split variables, and the training dataset sizes. The training dataset sizes have a strong effect on the classifier accuracy, but if the number of training data increases, the rate of improvement decreases. In addition, the number of variables only slightly influences the classifier accuracy.
2. The RF model can evaluate the variable importance for the overall classification of the crop classification and for the classification of each field crop category by the mean decrease accuracy and mean decrease Gini index. Variable importance measures demonstrate that the mode and sum of EVI have extremely important variables for crop

class separability. The variance and diversity of EVI also have importance for classification in crop categories. It might be more economical to use no-cost imagery for crop classification. However, we have difficulty understanding the precise rules used to generate the conclusive crop classifications because of the many multiple classification trees are created from resampling the same dataset.

3. Results of this study demonstrate that the application of RF classification approaches is useful for crop classification and land management. They can be useful to detect mislabeled areas, even multiple crop species in this area that have similar phenology. The overall accuracy and kappa coefficient for crops were 81% and 0.70, respectively. Therefore, technical challenges exist for crop classification using the RF model in the case of the small training dataset and similarly variable importance rank for crop species (e.g., inclusion of elevation and climatic variables in the model).

The approaches and methods presented in this study use only the no-cost dataset. They can be useful for different crops in medium areas. We must evaluate and compare these approaches and methods using other types of remote sensing data and additional variables.

Acknowledgements

We thank JICA–JSPS for providing funding for this study. We also appreciate La Molina National Agrarian University for field data collection and ground verification, which supported this study.

References

Agro Peruano: Abriendo un nuevo horizonte junto con el Japón, 2012. <http://www.ccipj.org.pe/LINKSforWEB/no11dec2012/LINKS11_especial.pdf> (accessed 20.11.2014).

Asner, G.P., Keller, M., Pereira, R., Zweede, J.C., 2002. Remote sensing of selective logging in Amazonia Assessing limitations based on detailed field observations, Landsat ETM+, and textural analysis. Remote Sens. Environ. 80 (3), 483–496.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Measur. 20 (1), 37–46.

Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Comput. Geosci. 63, 22–33.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88 (11), 2783–2792.

deFries, R., Hansen, M., Townshend, J., 1995. Global discrimination of land cover from metrices derived from AVHRR data sets. Remote Sens. Environ. 54 (3), 209–222.

Fahsi, A., Tsegaye, T., Tadesse, W., Coleman, T., 2000. Incorporation of digital elevation models with Landsat–TM data to improve land cover classification accuracy. For. Ecol. Manage. 128, 57–64.

Foody, G.M., 2004. Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes. Int. J. Remote Sens. 25 (15), 3091–3104.

Franklin, J., 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Prog. Phys. Geogr. 19 (4), 474–499.

Ghimire, B., Rogan, J., Miller, J., 2010. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. Remote Sens. Lett. 1 (1), 45–54.

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. Pattern Recogn. Lett. 27 (4), 294–300.

GRASS Development Team. Geographic Resources Analysis Support System (GRASS) Software; Version 6.4; Open Source Geospatial Foundation Project; 2012. <http://grass.osgeo.org/> (accessed 30.11.2013).

Guerrero, J.M., Pajares, G., Montalvo, M., Romeo, J., Guijarro, R.M., 2012. Support vector machines for crop/weeds identification in maize fields. Expert Syst. Appl. 39 (12), 11149–11155.

Guo, L., Chehata, N., Mallet, C., Boukir, S., 2011. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. ISPRS J. Photogramm. Remote Sens. 66 (1), 56–66.

Hastie, T., Tibshirani, R., Friedman, J., 2009. Random forests. The Elements of Statistical Learning. Springer, New York, pp. 587–604.

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 Sens. Environ. 83 (1–2), 195–213.

Irish, R.R., 2000. Landsat 7 Automatic Cloud Cover Assessment. In Shen, S.S., Descour, M.R. (Eds.): Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI, Proceedings of SPIE, 4049, pp. 348–355.

Irish, R.R., Barker, J.L., Goward, S.N., Arvidson, T., 2006. Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm. Photogramm. Eng. Remote Sens. 72 (10), 1179–1188.

Jakubauskas, M.E., Legates, D.R., Kastens, J.H., 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. Comput. Electron. Agr. 37 (1–3), 127–139.

Joy, S.M., Reich, R.M., Reynolds, R.T., 2003. A non-parametric, supervised classification of vegetation types on the Kaibab National Forest using decision trees. Int. J. Remote Sens. 24 (9), 1835–1852.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Latifi, H., Nothdurft, A., Koch, B., 2010. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. Forestry 83 (4), 395–407.

Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). Remote Sens. Environ. 100 (3), 356–362.

Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R News 2 (3), 18–22.

Lippitt, C.D., Rogan, J., Li, Z., Eastman, J.R., Jones, T.G., 2008. Mapping Selective Logging in Mixed Deciduous Forest: A Comparison of Machine Learning Algorithms. Photogrammetric Engineering & Remote Sensing 74 (10), 1201–1211.

Liu, H., Huete, A., 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Trans. Geosci. Remote Sens. 33 (2), 457–465.

Loveland, T.R., Merchant, J.W., Ohlen, D.O., Brown, J.F., 1991. Development of a land-cover characteristics database for the conterminous US. Photogrammetric Engineering & Remote Sensing 57 (11), 1453–1463.

Lunetta, R.S., Balogh, M.E., 1999. Application of Multi-temporal Landsat 5 TM imagery for Wetland Identification. Photogrammetric Engineering and Remote Sensing 65 (11), 1303–1310.

Martinuzzi, S., Vierling, L.A., Gould, W.A., Falkowski, M.J., Evans, J.S., Hudak, A.T., Vierling, K.T., 2009. Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. Remote Sens. Environ. 113 (12), 2533–2546.

Mathur, A., Foody, G.M., 2008. Crop classification by support vector machine with intelligently selected training data for an operational application. Int. J. Remote Sens. 29 (8), 2227–2240.

Mellor, A., Haywood, A., Stone, C., Jones, S., 2013. The performance of Random Forests in an Operational Setting for Large Area Sclerophyll Forest Classification. Remote Sensing 5 (6), 2838–2856.

Mkhabela, M.S., Bullock, P., Raj, S., Wang, S., Yang, Y., 2011. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agric. For. Meteorol. 151 (3), 385–393.

Murakami, T., Ogawa, S., Ishitsuka, N., Kumagai, K., Saito, G., 2001. Crop discrimination with multitemporal SPOT/HRV data in the Saga Plains. Japan. International Journal of Remote Sensing 22 (7), 1335–1348.

Oetter, D.R., Cohen, W.B., Berterretche, M., Maiersperger, T.K., Kennedy, R.E., 2000. Land cover mapping in an agricultural setting using multiseasonal Thematic Mapper data. Remote Sens. Environ. 76 (2), 139–155.

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., Pereira, J.M.C., 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multi Regression and Random Forest. For. Ecol. Manage. 275, 117–129.

Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26 (1), 217–222.

Pal, M., Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens. Environ. 86 (4), 554–565.

Panigrahy, S., Sharma, S.A., 1997. Mapping of crop rotation using multidate Indian Remote Sensing Satellite digital data. ISPRS J. Photogramm. Remote Sens. 52 (2), 85–91.

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9 (2), 181–199.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Australia. <http://www.r-project.org> (accessed 30.11.2013).

Rodriguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C., 2012a. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sens. Environ. 121, 93–107.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012b. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104.

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., Roberts, D., 2008. Mapping land-cover modifications over large areas: a comparison of machine learning algorithms. Remote Sens. Environ. 112 (5), 2272–2283.

Ryerson, R.A., Dobbins, R.N., Thibault, C., 1985. Timely crop area estimates from Landsat. Photogramm. Eng. Remote Sens. 51 (11), 1735–1743.

Sesnie, S.E., Gessler, P.E., Finegan, B., Thessler, S., 2008. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. Remote Sens. Environ. 112 (5), 2145–2159.

Simonneaux, V., Duchemin, B., Helson, D., Er-Raki, S., Olioso, A., Chehbouni, A.G., 2008. The use of high-resolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central Morocco. Int. J. Remote Sens. 29 (1), 95–116.

Tizado, E.J., 2013. GRASS GIS manual. <http://grass.osgeo.org/grass64/manuals/i.landsat.toar.html> (accessed 28.11.2013).

USGS, 2013. SLC-off Products: Background. <http://landsat.usgs.gov/products_slcoffbackground.php> (accessed 29.11.2013).

Wang, L., Sousa, W.P., Gong, P., Biging, G.S., 2004. Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama. Remote Sens. Environ. 91 (3–4), 432–440.

Waske, B., Braun, M., 2009. Classifier ensembles for land cover mapping using multitemporal SAR imagery. ISPRS J. Photogramm. Remote Sens. 64 (5), 450–457.

Yang, C., Everitt, J.H., Murden, D., 2011. Evaluating high resolution SPOT 5 satellite imagery for crop identification. Comput. Electron. Agr. 75 (2), 347–354.

Yuan, F., Bauer, M.E., Heinert, N.J., Holden, G.R., 2005. Multi-level land cover mapping of the Twin Cities (Minnesota) metropolitan area with multi-seasonal Landsat TM/ETM+ Data. Geocarto Int. 20 (2), 5–13.

Zhang, W., Yang, Y., Wang, Q., 2013. A study on software effort prediction using machine learning techniques. Commun. Comput. Inform. Sci. 275, 1–15.