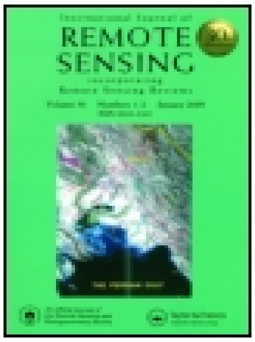

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong

Yuhao Jin, Xiaoping Liu, Yimin Chen & Xun Liang

To cite this article: Yuhao Jin, Xiaoping Liu, Yimin Chen & Xun Liang (2018): Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong, International Journal of Remote Sensing, DOI: 10.1080/01431161.2018.1490976

To link to this article: https://doi.org/10.1080/01431161.2018.1490976

Published online: 23 Aug 2018.

INTERNATIONAL JOURNAL OF REMOTE SENSING
https://doi.org/10.1080/01431161.2018.1490976

Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong

Yuhao Jin[a,b], Xiaoping Liu [a,b], Yimin Chen [c] and Xun Liang[a,b]

[a]School of Geography and Planning, Sun Yat-sen University, Guangzhou, China; [b]Guangdong Key Laboratory for Urbanization and Geo-Simulation, Sun Yat-sen University, Guangzhou, China; [c]School of Geographic Sciences Key Lab. of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai, China

ABSTRACT

Land-cover mapping in complex farming area is a difficult task because of the complex pattern of vegetation and rugged mountains with fast-flowing rivers, and it requires a method for accurate classification of complex land cover. Random Forest classification (RFC) has the advantages of high classification accuracy and the ability to measure variable importance in land-cover mapping. This study evaluates the addition of both normalized difference vegetation index (NDVI) time-series and the Grey Level Co-occurrence Matrix (GLCM) textural variables using the RFC for land-cover mapping in a complex farming region. On this basis, the best classification model is selected to extract the land-cover classification information in Central Shandong. To explore which input variables yield the best accuracy for land-cover classification in complex farming areas, we evaluate the importance of Random Forest variables. The results show that adding not only multi-temporal imagery and topographic variables but also GLCM textural variables and NDVI time-series variables achieved the highest overall accuracy of 89% and kappa coefficient (κ) of 0.81. The assessment of the importance of a Random Forest classifier indicates that the key input variables include the summer NDVI followed by the summer near-infrared band and the elevation, along with the GLCM-mean, GLCM-contrast.

ARTICLE HISTORY
Received 20 November 2017
Accepted 31 May 2018

1. Introduction

Land-cover mapping and monitoring is significant for sustainable development planning, rational exploitation of land resources, and climate change research (Bounoua et al. 2002; Turner Ii, Moss, and Skole 1993; Jung et al. 2006; Liu et al. 2018). It plays an important role in monitoring sustainable forest management, deforestation, agricultural planning, and urban growth (Foley et al. 2011; Gong et al. 2013; Liang et al. 2018; Liu et al. 2017). Remotely sensed data are widely used in land-cover mapping and monitoring. Remote sensing has the characteristics of rapid, macroscopic, and synchronous

CONTACT Xiaoping Liu liuxp3@mail.sysu.edu.cn School of Geography and Planning, Sun Yat-sen University, Guangzhou, 510275, China

monitoring, which provides efficient and quick technical means for extracting land-cover information (Friedl and Brodley 1997). Remotely sensed data have derived many global and regional land-cover data products, for example, the Moderate Resolution Imaging Spectrometer land-cover product (MODIS) (Friedl et al. 2010), the global land-cover map of 2009 (GlobCover2009), the 30 m global land-cover data set (GlobeLand30) (Chen et al. 2015), and finer resolution observation and monitoring of global land cover (Gong et al. 2013). However, there are still several problems with existing land-cover products in complex land-cover areas. For instance, it is difficult to map multiple classes of land-cover classifications in complex terrains and landforms. Thus, several area of natural grassland and improved grassland can be easily classified as forest or cropland without the vegetation phenology information.

The normalized difference vegetation index (NDVI) data are a common data source for land-cover mapping that can be used to obtain phenological information of vegetation (Pittman et al. 2010). Compared with other vegetation types such as grassland and forests, cropland has its own characteristics in the stages of sowing, growth, and harvest. However, it is not easy to distinguish cropland from forest and grassland using a single image. NDVI time series data, including Landsat and MODIS NDVI time series, are of great importance to land-cover mapping in farming areas. These data are usually used to extract vegetation information based on the threshold method. Meanwhile, in order to yield the best accuracy for land-cover classification, not only vegetation phenology information but also the textural features of each category are significant.

It has become a very common method to obtain textural features of remotely sensed data by means of the Grey Level Co-occurrence Matrix (GLCM) (Ozdemir et al. 2008). The textural features obtained by GLCM range from mean to variance and homogeneity, to contrast and entropy, to correlation and dissimilarity. The GLCM has been widely accepted and used for vegetation structure modelling (Castillo-Santiago, Ricker, and de Jong 2010) and land-cover classification (Johansen et al. 2007). However, the selection of the proper algorithm for land-cover classification depends on the ability of the algorithm to cope with the problem, including data errors, the complexity of the study area, and the lack of training data (Rogan et al. 2008).

The classification algorithm is an important part of land-cover mapping, but its efficiency and accuracy are influenced by many factors, such as image resolution and atmospheric conditions during imaging (Rogan et al. 2008). Therefore, there is no general classification algorithm at present and different research purposes need to choose different classification models. For example, unsupervised classification methods including *k*-means or Iterative Selforganizing Data Analysis Techniques Algorithm are often used in low-resolution remote-sensing images (Wagstaff 2001; Dunn 1973). Meanwhile, land-cover classification based on remote-sensing data and machine-learning algorithms have been the focus of academic research. Traditional classification algorithms have gradually been overtaken by machine-learning algorithms, because of their lack of effectiveness and accuracy, such as decision trees (Breiman et al. 1984), artificial neural networks (Mas and Flores 2008), support vector machines (Mountrakis, Im, and Ogole 2011), and Random Forest (Breiman 2001). However, most of the machine-learning algorithms also have their drawbacks. Artificial neural networks are highly non-linear large-scale systems, and the complexity of the artificial neural network makes it impossible to analyse its performance parameters accurately (Foody and Arora 1997). Support vector machines involve the calculation of the number of samples, and the storage and computation of the matrix, will consume a great deal of machine memory and computation time when the number of samples is large (Mountrakis, Im, and Ogole 2011).

The Random Forest algorithm is based on multi-classification or regression decision trees (Breiman 2001). However, the Random Forest has been applied in land-cover classification in complex region mostly using single-temporal imagery and NDVI (Chapman 2010), hyperspectral and multispectral remotely sensed data (Pal 2005; Sesnie et al. 2008), or digital elevation models (DEMs) (Ghimire, Rogan, and Miller 2010). Meanwhile, most studies did not adequately analyse which auxiliary variable features were most relevant in the classification of complex farming area (Chan and Paelinckx 2008). In addition, most studies have not further discussed the sensitivity of the Random Forest model to the training data set reduction (Gislason, Benediktsson, and Sveinsson 2006).

In this study, our goal was to evaluate the addition of both time-series NDVI and the GLCM textural variables using the Random Forest for land-cover mapping in a complex farming region of Central Shandong. We determined the best input variables for the Random Forest from multi-temporal remotely sensed and auxiliary data by the evaluation of the variable importance. The performance of the Random Forest algorithm was also evaluated by the number of trees and predictor variables, as well as the sensitivity of the training data size changes.

2. Study area

Shandong Province is one of the largest agricultural provinces in China. Pingyin County is selected as the study area, which is located in central Shandong (36° 1′–36° 23′ N, 116° 12′–116° 27′ E) (Figure 1). As a transitional zone between Mount Tai and the West Shandong Plain, Pingyin County is one of the typical mountainous farming areas in central Shandong. The county is dominated by hills and mountains, covering an area of 515.16 km^2, approximately 62.3% of the total area. In addition to the area along the Yellow River and the eastern part of the county, the remainder is the low foothill area. The climate is warm temperate zone monsoon, characterized by four seasons: dry and windy in spring, hot and rainy in summer, bright in autumn, and cold and less snow in winter. According to the China Statistical Yearbook for Regional Economy-2011, the area of cropland increased by 3.89 km^2 from 2004 (18.34 km^2) to 2010 (22.23 km^2). The cropland area is dominated by cinnamon soil supporting primary crops of corn, soybeans, wheat, and potatoes. Most of these planting structures are double cropping systems.

3. Data and methods

3.1. *Land-cover classes and reference data*

To reflect the major land types in this area, the classification scheme used in this project was based on land-cover maps developed in 2010 by Institute of Geographic Sciences and Natural Resources Research (Brewster, Allen, and Kopp 1999). Land cover was classified into six types: cropland, forest, grassland, water, residential land, and unused land (see Table 1).

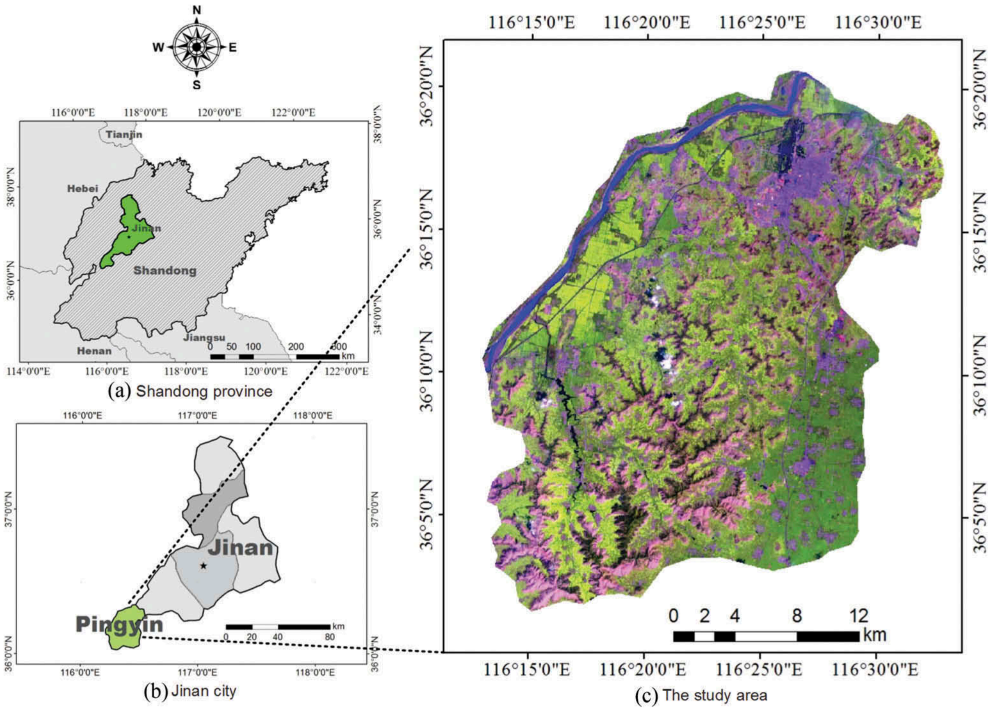

Figure 1. Study area in Pingyin, Shandong. The study area displays the red, near-infrared band and middle infrared band of the Landsat Thematic Mapper (TM) data with blue, green, and red colour.

Table 1. Land-cover classification schemes.

Classification	Description
Cropland	The land planted for crops, including cultivated land, newly opened wasteland, fallow land, fallow land, grass land rotation land
Forest	The land grows shrubs, trees, bamboo, and coastal mangrove forest land, covered by natural or planted forests with a canopy density of over 30% and natural or newly forested or shrub with a canopy density between 10% and 30%
Grassland	Lands mainly covered by herbaceous plants, including the shrub grassland and the grassland with a canopy density of less then 10%, and the grassland with coverage of more than 5%
Water	Reservoirs and ponds, rivers, and flooded lands
Residential land	Land used for townships and rural settlements, also including industrial sites and mining
Unused land	Land refers to the area that is not put into practical use or difficult to use

It was commonly believed that the number of training pixels should at least be equal to 10 times the number of variables used in the classification model (Jensen and Lulla 1987). To achieve better results, other studies have found that machine-learning algorithms demanded a large number of training data (Gislason, Benediktsson, and Sveinsson 2006). However, if the training sample size for each land-cover class is the same, the proportion of some important classes will be small and some of the less important classes will be assigned to a high proportion and be prone to over-fitting problems. In this research, we used a random sampling method that can balance the number of sampling points, which meant that the number of sampling points of each land-cover class was related to the proportion of the total pixels. More specifically, a total of 2000 reference training and test

Table 2. Summary of reference training data and testing data.

Classification	Training points prior to filtering	Final points for model training	Points for accuracy testing
Cropland	862	789	394
Forest	129	115	58
Grassland	82	71	37
Water	40	36	18
Residential land	250	224	112
Unused land	12	10	6
Total	1375	1245	625

data sets were divided randomly to each land-cover class in accordance with their respective proportions (see Table 2). A stratified random sample of 70% of the reference point data for training the Random Forest classifier and 30% of the reference point data for testing the accuracy of the results was used.

To train and validate the land-cover classifications, an extensive reference data set was gathered in our study area. The reference training and test data sets were collected from a time series of Landsat TM images and Google Earth high-resolution images by visual interpretation. Each sample unit was checked according to the common global land-cover validation database, including GlobeLand30 and FROM-GC (Gong et al. 2013). There were a total of 2000 sample units in the study area, including 1256 cropland sample units, 187 forest sample units, 153 grass sample units, and 362 residential land sample units. We filtered the training points based on field photo interpretation, GlobeLand30, and FROM-GC to determine whether they were appropriate references for each class.

3.2. *Satellite and ancillary data*

3.2.1. *Multi-temporal satellite data*

To create land-cover maps in complex farming areas, it is common to use not only a single date of remote-sensing images but also multi-temporal remotely sensed data that is commonly used to characterize phenological changes in vegetation cover states (Yuan et al. 2005). A large number of studies have shown that these multi-temporal remotely sensed data provide the differences between similar spectral coverage types (Chen et al. 2014).

In the research, it is difficult to distinguish cropland from forest and grassland in the hilly area of central Shandong by only using the Landsat imagery of single dates. Therefore, spring, summer, and autumn images have been used for land-cover classifications because these images contain most of the phenological changes (April, August, and October, respectively). These three periods represent the most significant characteristics of the main vegetation types in the study area that are essential for the accurate classification of land-cover types. According to the spectral characteristics of the images, cropland (such as corn, soybeans, wheat, and potatoes) can be confused with temperate deciduous broad-leaved forests in both spring and summer images. In autumn images, highly reflective surfaces, such as residential land, can be confused with unused land. The concept of seasons is related to the Northern Hemisphere.

Three Landsat 5 TM images of central Shandong were selected with acquisition dates corresponding to the 8 April 2009, 30 August 2009, and 17 October 2009, and they were from Path: 122, Row: 035 of the Landsat Worldwide Reference System. The Band 1, band 2,

Table 3. Multi-temporal spectral variables (MTS) used in Random Forest classification.

Season	Data	Source	Ancillary features	Calculation method
			B, G, R, NIR, MIR1, MIR2, TIR	
Summer	8 April 2009		MNDWI	$(R_{Green} - R_{MIR1})/(R_{Green} + R_{MIR1})$
			NDBI	$(R_{MIR1} - R_{NIR})/(R_{MIR1} + R_{NIR})$
Summer	30 August 2009	Landsat 5 TM	Brightness	$0.2937R_{Blue} + 0.2493R_{Green} + 0.4843R_{Red} + 0.5565R_{NIR} + 0.4482R_{MIR1} + 0.1763R_{MIR2}$
Autumn	17 October 2009		Greenness	$-0.2748R_{Blue} - 0.2235R_{Green} - 0.5536R_{Red} + 0.7543R_{NIR} + 0.0840R_{MIR1} - 0.1800R_{MIR2}$
			Wetness	$0.1406R_{Blue} + 0.1843R_{Green} + 0.3659R_{Red} + 0.3306R_{NIR} - 0.6712R_{MIR1} - 0.4122R_{MIR2}$

band 3, and band 4,respectively, denote the blue (B), green (G), red (R), and near-infrared (NIR) bands of the TM images. Band 5, band 7, and band 6 are two mid-infrared (MIR1, MIR2) and thermal infrared (TIR) bands from all images. As NIR, MIR1, MIR2, and TIR are suitable for detecting the water content in plants and soils, we use them for land-cover mapping (see Table 3) (Baker et al. 2006).

To distinguish vegetation from non-vegetation, we used the brightness, greenness, and wetness axes of the Tasselled Cap transformation, which has long been estimated to play an important role in improving assessment of land-cover change, classification results, and assisting in estimates of vegetation structure (Kauth 1975; Crist and Cicone 1984). Other indices were also used for land-cover classification, including the Normalized Difference Built-Up Index (Zha, Gao, and Ni 2003), and the Modified Normalized Difference Water Index (Gao 1996). All of these ancillary features were calculated in the ESRI ENVI 5.1 platform.

3.2.2. *NDVI time-series data*

The seasonal pattern of vegetation provides the foundation for the study of land-cover mapping through time-series remote-sensing images. Compared with other vegetation types such as grassland and forests, cropland has its own characteristics in the stages of sowing, growth, and harvest. At the same time, with one or more narrow peaks, cropland is characterized by more irregular NDVI profiles, whose peak values constantly change substantially compared to the peaks of grasslands and forests.

Our study area has a large number of cloud-free Landsat images, since the rainfall in central Shandong was relatively small. A total of 10 scenes of the Landsat 5 TM (Path 122/Row 035) to 2009 spanning from 7 March (day 66) to 4 December (day 338) were atmospherically corrected for surface reflectance by using the Landsat Ecosystem Disturbance Adaptive Processing System (Masek et al. 2006), then they were used to generate a time-series Landsat NDVI image. Since we selected Landsat imagery with less than 15% clouds, some areas of images acquired on days 82, 274, and 338 were partially contaminated by clouds. However, the cloud-covered areas of these images were not in the study area, and thus they had little impact on the acquisition of the time-series NDVI.

In Figure 2, we demonstrated the NDVI time series profiles of different land-cover types in an area of central Shandong. Over the time from 7 March to 4 December, 272 days in all including seasons from spring to winter, the NDVI values of cropland varied greatly from sowing and growth to harvest. Thus, crop phenology in this region

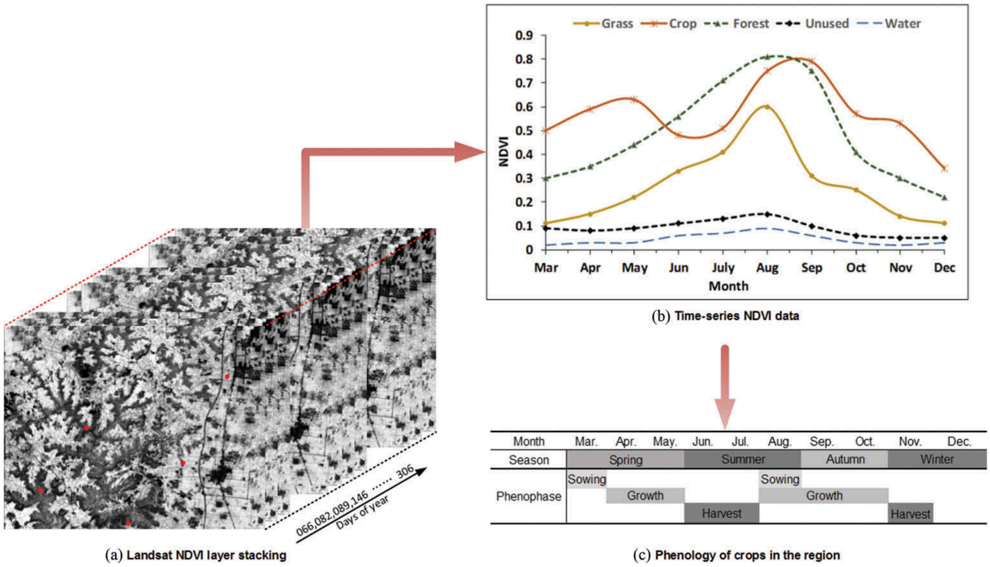

Figure 2. NDVI time series profiles and crop phenology analysis.

was divided into three stages: the sowing (March and August), growth (from April to May and from August to October), and harvest (from June to July and November) seasons. Figure 2 also indicates that the plant system of double cropping is one of the significant variables in the region.

3.2.3. *GLCM textural data*

The use of textural features data in land-cover mapping requires decisions about multiple associated variables, of which each results in different textures that may have different values. In the research, to introduce texture information into remote-sensing image classification, we used the GLCM, which has been widely accepted and used for vegetation structure modelling (Castillo-Santiago, Ricker, and de Jong 2010) and land-cover classification (Johansen et al. 2007).

The GLCM is a matrix of size $N \times N$ where N is the number of levels specified under quantization, which is the relative frequency of a tabular pair of digital numbers (Haralick, Shanmugam, and Dinstein 1973). Each band was segmented separately, and multiple textural variables of Landsat imagery were extracted by GLCM measures (Zhang et al. 2009). We considered a set of eight textural variables that could be obtained from the GLCM calculated in the ENVI 5.1 platform (see Table 4): mean, variance, homogeneity, contrast, dissimilarity, entropy, angular second moment, and correlation.

3.2.4. *Topographic data*

Because of the complicated topography and different soil moisture conditions in central Shandong, topographic variable data are a significant factor affecting land-cover classification (Chen et al. 2017). The National Aeronautics and Space Administration Shuttle Radar Topography Mission (SRTM) provides 90 m resolution DEM that can be used to determine elevation and derive slope and aspect (Moore 1991). We calculated the slope

Table 4. GLCM textural variables (TXT) used in Random Forest classification.

Type	Source	Ancillary features	Calculation method
Texture features (TXT)	Landsat 5 TM	GLCM-mean GLCM-variance GLCM-homogeneity GLCM-contrast GLCM-dissimilarity GLCM-entropy GLCM-angular second moment GLCM-correlation	Calculated in the ENVI 5.1 platform

and aspect information for each pixel based on SRTM and resampled these terrain attributes to the extent of Landsat TM image for subsequent work, supported by the ESRI ArcGIS (v. 10.2).

3.3. *Random Forest classification*

Random Forest has been developed rapidly and has been widely used in many fields such as medicine, economics, and geography during the past twenty years. Breiman (2001) proposed Random Forest, which changes the way the classification or regression tree is constructed. RF can process hundreds of input features without deleting features and estimates the importance of features in the classification. As demonstrated by many researches (Dobra and Gehrke 2001; Schroff, Criminisi, and Zisserman 2008), Random Forest is more robust than a single decision tree, and its generalization performance is much better than traditional learning algorithms.

Random Forest classification (RFC) is an ensemble classification method consisting of many decision tree classification models, $\{h(\boldsymbol{X},\Theta k), k = 1, \ldots\}$, where $\boldsymbol{x}$ is the input vector and $\{\Theta k\}$ are random vectors of independent distribution (Foody and Arora 1997). Hence, some vectors may be trained more than once in a classifier, while other vectors may never be used. Therefore, as it becomes more robust when it faces slight changes in input data and increases classification accuracy, better classifier stability is achieved. A k number of samples was extracted from the original training set using bootstrap sampling, and the sample size of each sample was the same as that of the original training set. A k number of trees was set up for k samples, and a k number of classification results were obtained. In other words, in order to classify new data sets, a constant number of k random variables are used, and each sample of the data set is classified by a k number of trees. According to the classification results, each record was voted on to determine its final classification.

Some experiments have shown that Random Forest unlike other methods based on boosting and bagging is not over-sensitive to overtraining or noise (Briem, Benediktsson, and Sveinsson 2002; Chan and Paelinckx 2008). Each subset for the growth of a single decision tree contains approximately 67% of the calibration data set. Elements not included in this subset are included in another subset called out-of-bag (OOB). Using these elements to estimate the performance of the model is called OOB estimation, and the ratio between OOB elements and misclassification contributes to the unbiased estimation of generalization error (Wolpert and Macready 1999).

To estimate the importance of variables, the OOB error of each decision tree in the Random Forest e_t needs to be calculated from the OOB data. A new OOB error *ei*t is calculated by randomly changing the values of the variable X^i. The mean decrease in accuracy (MDA) score can be expressed as Equation (1).

$$V(X^i) = \frac{1}{N}\sum_{N}^{t=1}(e_t^i - e_t) \tag{1}$$

In addition to the OOB error-based approach MDA, another method of calculating the importance of the variable is based on the mean decrease Gini measure, which is frequently used as one of the important ways for selecting the best split in an RF and measures the impurity of a specified variable. In these two methods, the more reduction in OOB accuracy and Gini index caused by variable changes, the more important the variable is (Foody and Arora 1997).

RFC improves the classification model by constructing different training sets to increase the differences among the classification models, and the sequence $\{h^1(\boldsymbol{X}), h^2(\boldsymbol{X}),\ldots, h^k(\boldsymbol{X})\}$ is obtained through a k number of training times and then used to form a multi-class model system. Thus, the application of RF in remote sensing has many advantages.

- It performs well. It provides top-level accuracy among current popular similar algorithms. It can process big data without feature selection and feature deletion and maintain the classification error balance when the class size distribution is unbalanced.
- It requires little manual intervention. It can determine the characteristics of the data by itself, thus simplifying the Random Forest design process.
- It provides a variety of data characterizations. We can calculate the importance of each classification feature by estimating the generalization error in the process of forest growth.
- It has very fast computing speed. Since the amount of computation in a Random Forest is proportional to the depth of a tree, classification or regression is very rapid on a growing tree. The amount of computation for a Random Forest training is proportional to the number of all its trees, which makes it easier for Random Forests to be parallelized.

4. Results and discussion

4.1. *Influence of the number of trees and predictive variables*

Random Forest is a classification tree-based algorithm considering one or several variables. RF can be divided into single variable algorithms or multiple variables, if the decision is made from a single variable or is coordinated by a great many variables (Breiman 2001). Hence, the additional parameter n represents a subset of n predictive variables, which does not appear in the traditional classification tree. The parameter affects the intensity and the correlation of each tree and affects the accuracy of the classifier and the generalization error.

Breiman proposed that generalization errors always converged and over-training was not a problem when the number of trees was increased. Meanwhile, the reduction in the number of predictive variables (n) decreased the correlation between decision trees,

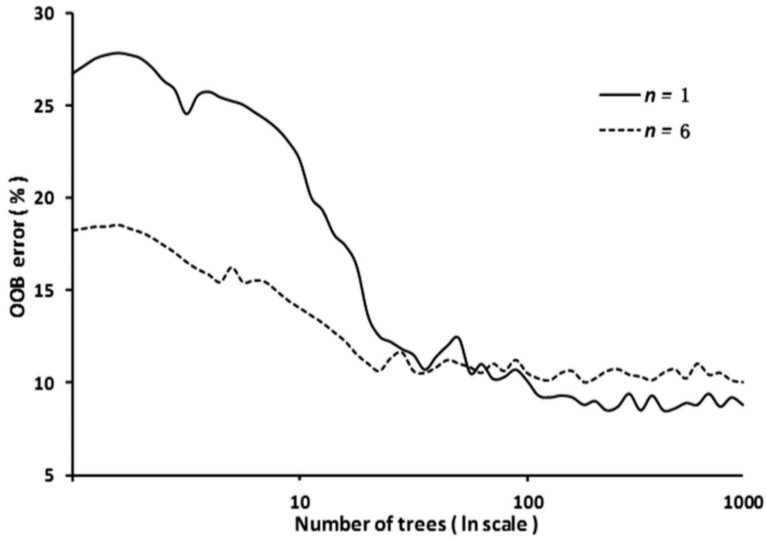

Figure 3. OOB error index of the RF models for different numbers of trees (*k*) and predictive variables (*n*).

thereby increasing the accuracy of the model. Therefore, the optimization of the number of trees (*k*) and predictive variables (*n*) played a significant role in improving the classifier's accuracy.

Figure 3 shows that the OOB error of the model depends on the number of trees and the predictor variables. As seen from the figure, when the number of trees was 100 and the predicted variables were equal to the minimum and maximum (1 and 6), the OOB error converges to 10% and 11%, respectively. By contrast, the OOB errors of the RF models of 1000 trees composed of *n* different predictor variables converged to 9% and 10%. Thus, in Random Forest, it led to lower classification accuracy when the number of trees was very small, but the generalization error did not increase or decrease if we added more trees infinitely, and a larger number of trees led to more stable classification.

When the value of *k* is 1000, the OOB error difference between the minimum and maximum of the predictive variable was approximately less than 1%. Therefore, we can deduce that Random Forest was not sensitive to the number of predictive variables (*n*) once the error convergences were reached.

4.2. *Variable importance*

The information on the importance of the variables in the classification of each category was provided by the distribution of variables in the decision tree (Pal 2005). However, the integration of classifiers based on multiple decision trees was hardly possible for this interpretation. Thus, Random Forest was used to evaluate the importance of variables through the OOB error and the Gini index (Section 3.3).

Meanwhile, as seen in Figure 4, in order to classify the land-cover types in the study area, multi-temporal spectral variables (MTS), topographic variables (DEM), GLCM textural variables (TXT), and NDVI time-series variables (NDVI) were used in Random Forest

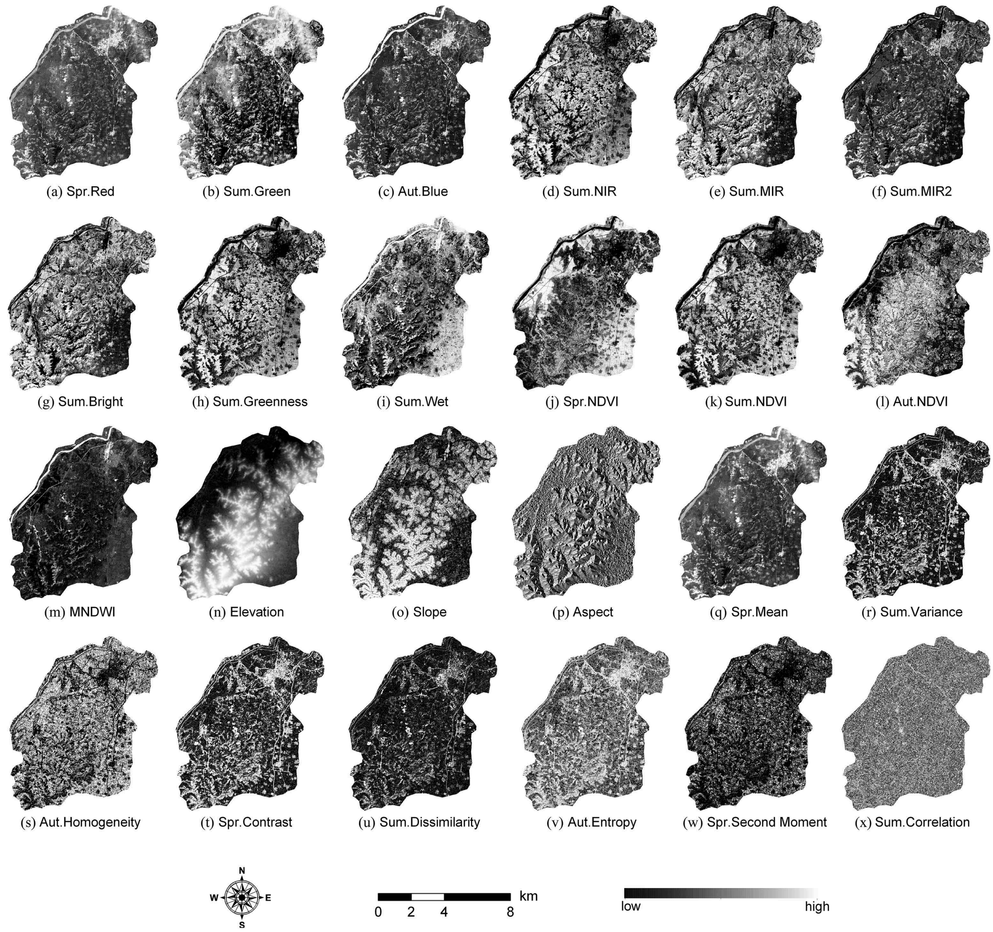

Figure 4. Ancillary variables in models.

and the experiment was divided into four models: MTS + DEM, MTS + DEM + TXT, MTS + DEM + NDVI, MTS + DEM + TXT + NDVI (Section 3.2).

Figure 5 shows that the 10 most important variables of the models, the MTS + DEM, the MTS + DEM + TXT, the MTS + DEM + NDVI, and the MTS + DEM + TXT + NDVI, are listed in terms of the OOB error and the GINI index. MTS variables and NDVI time-series variables (NDVI) are expressed in the form of a combination of seasonal abbreviations and band names, which spr, sum, and aut represent spring, summer, and autumn, respectively, when expressing seasonal abbreviations. As seen in Figure 5, according to both OOB error and the GINI index, the multi-temporal texture of the NIR band, the elevation, and the summer greenness were the three most important variables in the MTS + DEM model. For the MTS + DEM + TXT model, among the eight textural variables used for land-cover classification, mean was the most important, followed by contrast, variance, and entropy. In the terrain variables, elevation played a major role, followed by the slope, and the remaining variables were less significant. Compared to the previous two models, the NDVI time-series variables were used in the MTS + DEM + NDVI model, and they had the most important influence on class-separability. According to the OOB MDA, the NDVI variables of three

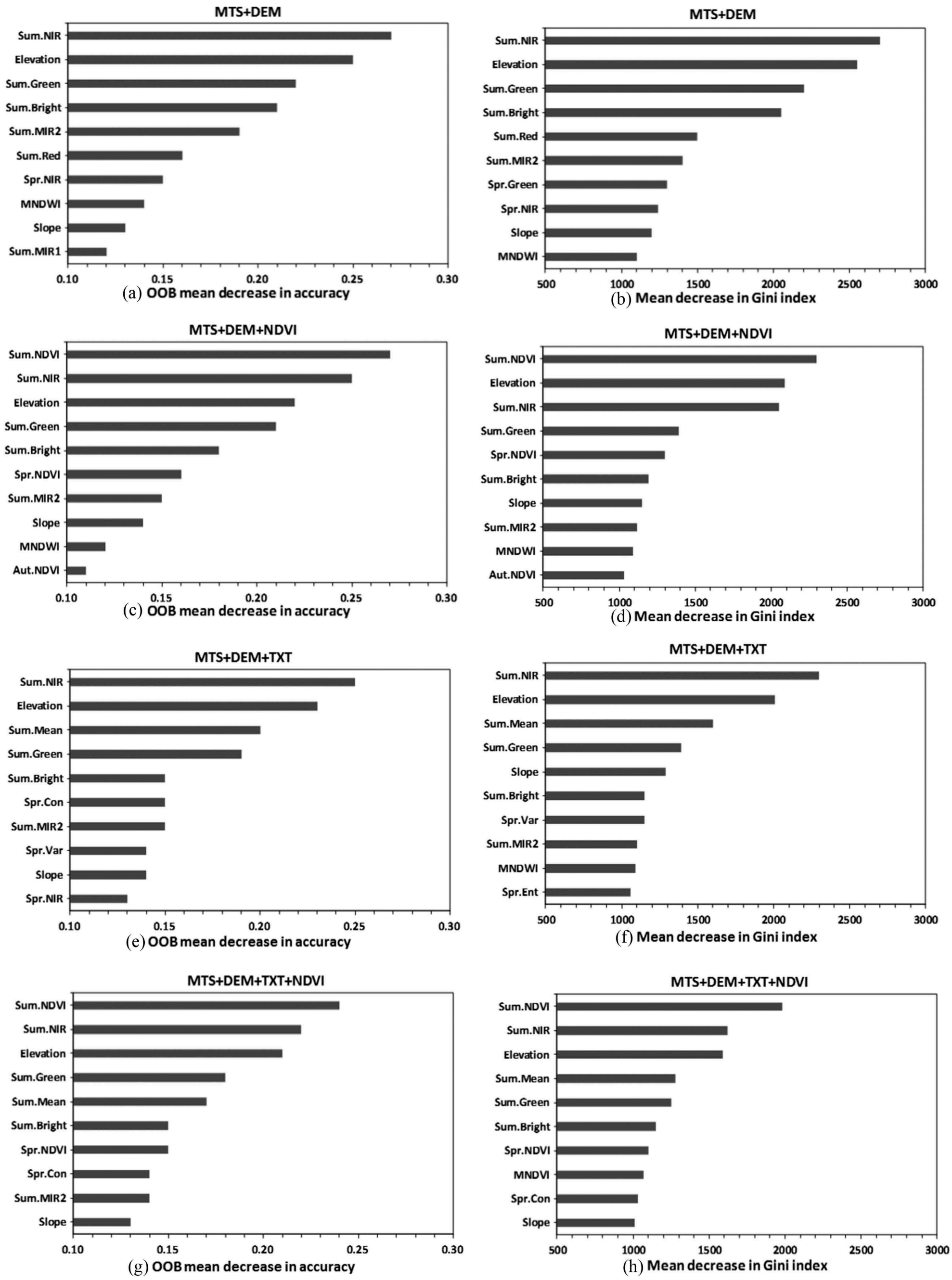

Figure 5. Variable importance based on OOB and Gini measures.

seasons have a significant contribution to land-cover classification, with values equal to 0.27, 0.16, and 0.11. Finally, although the OOB and Gini approaches in the MTS + DEM + TXT + NDVI model showed slightly different variable importance, regardless of the order in which the importance of variables appears, the most important variable was the summer NDVI, followed by the summer NIR band and the elevation, with values greater than 0.9 in the OOB approach and greater than 1800 in the Gini index. In addition, the most important GLCM textural variables were the mean and contrast (OOB MDA equal to 0.18 and 0.14). The brightness, and greenness derived from the Kauth–Thomas transform also have important influence.

After an assessment of the importance of a Random Forest classifier for land-cover classification, the 10 variables that contributed most to the classification were selected in each of the 4 models. Through the evaluation of the importance of variables, unimportant variables were eliminated to reduce variable dimensions, model computation time, and improve efficiency.

4.3. *Map accuracy*

The confusion matrix was used to evaluate classification accuracy. The overall accuracy and the two measures of quantity disagreement and allocation disagreement were useful to summarize a confusion matrix (2011). The confusion matrix was calculated by the classification results of different models after the evaluation of the importance of variables, and the overall accuracy, quantity disagreement and allocation disagreement statistic were obtained. The difference between the results of the four models is shown by the accuracy contrast table (see Table 5).

Table 5 compared the overall accuracy, the quantity disagreement, and allocation disagreement of the MTS + DEM, the MTS + DEM + TXT, the MTS + DEM + NDVI, and the MTS + DEM + TXT + NDVI models and concluded that the overall accuracy was increasing gradually with the addition of different types of variables. Meanwhile, the addition of variables was able to decrease both QD and AD value.

The MTS + DEM + TXT and the MTS + DEM + NDVI models had overall accuracies of 80.2% (total disagreement = 19.8%) and 84.1% (total disagreement = 15.9%), respectively. The MTS + DEM + TXT model incorporated multi-temporal bands, topographic variables, and the GLCM variables. Compared with the overall accuracy of the MTS + DEM model, the result of this model revealed that addition of textural variables increased the accuracy of the model by approximately 7%. On the other hand, instead of adding textural variables, the MTS + DEM + NDVI model incorporated all available multi-temporal bands, topographic variables, and the NDVI time-series data, which greatly increased the accuracy of the model by over 10%. Thus, it can be seen that the influence of the NDVI time-series variables on classification accuracy is greater than the influence of textural variables.

To search the model with the highest accuracy, both textural variables and NDVI time-series variables were added to the MTS + DEM + TXT + NDVI models. As seen in the Table 5, adding not only multi-temporal bands and topographic variables but also textural variables and NDVI time-series variables achieved the highest overall accuracy

Table 5. Comparison of different classification accuracy.

	MTS + DEM	MTS + DEM + TXT	MTS + DEM + NDVI	MTS + DEM + TXT + NDVI
OA (%)	73.9	80.2	84.1	88.9
QD (%)	6.4	4.5	3.1	1.5
AD (%)	19.7	15.3	12.8	9.6

of 88.9%, and the disagreement percentage mainly due to the allocation disagreement, of 9.6%, rather than the quantitative disagreement, of only 1.5%. Compared to the absence of textural and NDVI variables, the addition of the GLCM variables and NDVI time-series variables made the overall accuracies significantly increase, with values equal to 15.0%. Moreover, they led to a reduction of the QD and AD value by 4.9% and 10.1%. These findings show that the Random Forest model improved the overall accuracy and the kappa coefficient (κ) using textural variables and NDVI time-series variables based on MTS and DEM variables.

4.4. *Comparison of different classification*

To test the classification results after variable selection and discuss the influence of different auxiliary variables on the land-cover classification in central Shandong, the MTS + DEM, the MTS + DEM + TXT, the MTS + DEM + NDVI, and the MTS + DEM + TXT + NDVI models were selected for the comparison of different classification results. After the evaluation of the importance of the variables, the 10 most significant variables were selected in each of the MTS + DEM, the MTS + DEM + TXT, the MTS + DEM + NDVI, and the MTS + DEM + TXT + NDVI models (Section 4.2).

In Figure 6, wide misclassifications were observed in the MTS + DEM and the MTS + DEM + TXT, especially in cropland, forest, and grassland. Some pixels of natural grassland and improved grassland were classified as forest and other pixels were misclassified as cropland, such as residential land and grassland. In contrast to Figures 6(b) and (c), cropland actually occupied most of the areas that might be occupied by residential land in the MTS + DEM. Having added the GLCM variables, the MTS + DEM + TXT extracted textural information from the study area much more accurately, including traffic roads, urban land, and rural residential areas. In Figure 6(d), with the introduction of NDVI time-series variables, the wide confusion of forest and grassland observed in Figures 6(b) and (c) was greatly reduced, and forest and grassland were distinctly delineated. In addition, several areas were originally misclassified as cropland that had been divided into grasslands or forest. However, the MTS + DEM + NDVI cannot classify the roads accurately in the absence of textural variables. Thus, Figure 6(e) shows that the MTS + DEM + TXT + NDVI that with added GLCM textural variables (TXT) and NDVI time-series variables (NDVI) into the MTS + DEM was the most accurate Random Forest model, and most of the areas corresponding to cropland, forest, grassland, and residential land had been correctly classified. Water and unused land were still in good agreement with testing data.

As illustrated in Section 4.3, the MTS + DEM + TXT + NDVI model using MTS variables, topographic variables (DEM), and GLCM textural variables (TXT) and along with NDVI time-series variables (NDVI) had the highest classification accuracy.

In Table 6, the user's accuracy for different classes was higher than 80%, and the cropland and the unused land had a higher accuracy (>90%) than other land-cover classes, which were also higher than the overall accuracy (89%). Compared to the accuracy of the cropland, the user's accuracy of the forest and grassland was lower, with values equal to 82% and 81%. The confusion of grassland, forest, and cropland is the main cause of their accuracy errors. Since the grasslands were usually associated with forests, especially in the low foothill area, it was more difficult to accurately distinguish between grasslands and forests. Therefore, the accuracy of grasslands and

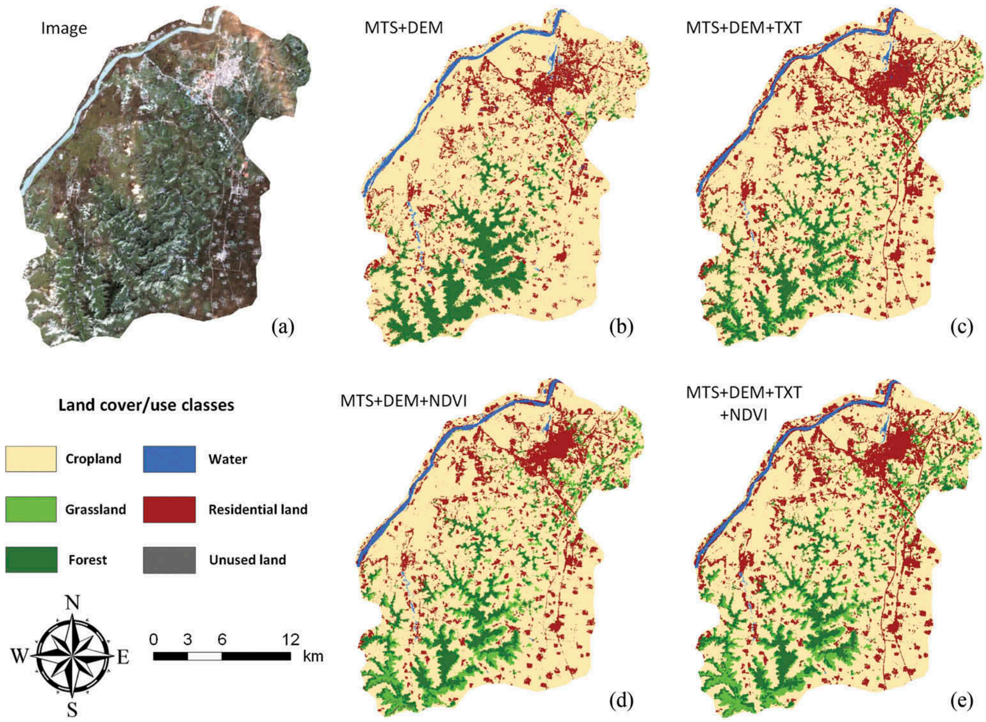

Figure 6. Comparison of different classification results.

Table 6. Comparison of accuracies of different classes.

Classification	Ground truth (pixels)							
	Cropland	Forest	Grassland	Water	Residential land	Unused land	Total	User's accuracy (%)
Cropland	365	6	4	3	15	1	394	92.6
Forest	5	48	5	0	0	0	58	82.8
Grassland	4	3	30	0	0	0	37	81.1
Water	2	0	0	16	0	0	18	88.9
Residential land	13	2	2	3	92	0	112	82.1
Unused land	1	0	0	0	0	5	6	83.3
Total	390	59	41	22	107	6	625	
Producer's accuracy (%)	93.6	81.4	73.2	72.7	86.0	83.3		

Overall accuracy (%) = 88.9; quantity disagreement (%) = 1.5; allocation disagreement (%) = 9.6.

forests in the MTS + DEM + TXT + NDVI model was acceptable because the grassland was confused with the forest less than 8% for the time. In addition, water and testing data showed good consistency, with a user's accuracy of 88%. The residential land was accurate in terms of the user and producer (82% and 85%, respectively), which was confused with cropland by approximately 11%.

Table 7 shows McNemar test results with the number of pixels correctly or wrongly classified by the MTS + DEM, the MTS + DEM + TXT, the MTS + DEM + NDVI, and the MTS + DEM + TXT + NDVI classification models. The symbol, f_{11}, denotes the number of cases wrongly classified by both maps while f_{22} denotes the number of cases correctly classified by both maps, while f_{12} and f_{21} are the cases that are correctly classified by one

Table 7. McNemar test showing the improvement of the MTS + DEM + TXT + NDVI classification over the MTS + DEM, the MTS + DEM+ TXT, the MTS + DEM+ NDVI classifications models.

Classification 1	Classification 2	f_{11}	f_{12}	f_{21}	f_{22}	Chi-square (χ^2)	*p*-Value
MTS + DEM + TXT + NDVI	MTS + DEM	60	15	103	447	64.5	<0.001
MTS + DEM + TXT + NDVI	MTS + DEM + TXT	56	23	81	465	32.3	<0.001
MTS + DEM + TXT + NDVI	MTS + DEM + NDVI	48	25	64	488	17.1	<0.001

*f*11: Number of cases with wrong classification in both maps; classification 1 and classification 2.
*f*12: Denotes number of cases that are wrongly classified by classification 1 but correctly classified by classification 2.
*f*21: Number of cases that are correctly classified by classification 1 but wrongly classified by classification 2.
*f*22: Number of cases with correct classification in both maps.

classifier but wrongly classified by the other. The table indicates the classifiers agree on *f*22 and *f*11 cases but disagree on *f*12 and *f*21 cases. The McNemar test clearly shows significant improvement of the MTS + DEM classification model over MTS + DEM, the MTS + DEM + TXT, the MTS + DEM + NDVI classifications models (Table 7).

4.5. *Effect of reduction in the training data*

It was a time-consuming and laborious task to acquire large-scale training data to train the RF classifier in the classification of complex regions with a great many categories. A large number of studies have shown that the number and area of training data samples were as large as possible to better contain variable conditions in each category (Breiman 2001). However, if the training data were not representative, then the accuracy could be reduced (Ghimire, Rogan, and Miller 2010). Thus, it was necessary to design a scheme to select training a sample that was feasible both in time and economic terms and could also achieve acceptable accuracy (Rogan et al. 2008).

In Table 8, the Random Forest classifier had a low sensitivity to the reduction in training data size. The impact of the reduction of training data size on the Random Forest performance was evaluated in terms of overall accuracy in the classification results of the RF model. The training data size was changed by 5% increments and the training data were reduced from 5% to 70%, while the overall classification accuracy decreased by less than 5%. Then, from the threshold equal to 70%, the accuracy was reduced more abruptly to achieve an overall accuracy equal to approximately 60%, when the training data had been reduced by 95%.

The results show that the classification accuracy of the Random Forest classifier decreased with the reduction in training data set size, but not in a linear pattern. As

Table 8. Impact of training data size reduction in the classification accuracy.

Reduction (%)	Overall accuracy (%)	Reduction (%)	Overall accuracy (%)
0	88.96	50	86.54
5	88.80	55	85.69
10	88.89	60	86.10
15	88.63	65	85.71
20	88.18	70	85.03
25	88.42	75	81.52
30	87.96	80	76.70
35	87.73	85	71.36
40	87.21	90	69.82
45	87.69	95	61.30

the overall accuracy was decreased abruptly after reaching a threshold of 70% in this study, we could draw a conclusion that a multivariate Random Forest classifier did not have too much sensitivity to the reduction of training data.

5. Conclusions

The objective of this study was to evaluate the addition of both time-series NDVI and the GLCM textural variables using Random Forest for land-cover mapping in a complex farming region of Central Shandong. Furthermore, we determined the best choice from various sources of multi-temporal remotely sensed and ancillary data to accurately identify the land-cover map. By incorporating multi-temporal data and auxiliary variables, including MTS variables, topographic variables (DEM), GLCM textural variables (TXT), and NDVI time-series variables (NDVI), the Random Forest model performed well in diverse land-cover classes. We achieved the goal by testing the classification results of models from four data sets and discovered the key variables that contributed most to the classification, including the NDVI, NIR, and MIR bands, greenness and brightness, mean and contrast, elevation and slope. The specific targets of the research were to evaluate the importance of variables through and the OOB error and the Gini index. The results of different classifications and the effect of the reduction of training data set size on the Random Forest performance were evaluated in terms of testing data accuracy.

The over-training was not a problem in the Random Forest model because of the Strong Law of Large Numbers, and the number of trees (k) and predictive variables (n) were two significant parameters it needed to set. The number of trees is proportional to the accuracy of the Random Forest until the number is 100 and the generalization error converges to approximately 10%. Once the error reaches the convergence, the accuracy of the RF model is no longer affected by the number of variables, thus the Random Forest model hardly requires guidance.

In addition, the RF could evaluate the importance of variables for the different land-cover classifications through the OOB error and the Gini index. This assessment was significant for the classifications of complex regions, because a large amount of training data sources with varying variables were needed in farming and mountainous areas. The assessment indicated how the summer NDVI became the most important variable in the Random Forest model in the central Shandong, followed by the summer NIR band and the elevation. In addition, the GLCM textural variables, brightness, and greenness derived from the Kauth–Thomas transform were also of great importance in the land-cover classification.

While the Random Forest model used MTS variables, topographic variables (DEM), GLCM textural variables (TXT), and NDVI time-series variables (NDVI), the results show that it had higher classification accuracy than the classifier without added textural and NDVI time-series variables. Meanwhile, a high user's accuracy for cropland, forest, and grassland was obtained, with values equal to 92%, 82%, and 81%, which verified the feasibility of an effective solution for the confusion of cropland, forest, and grassland in complex farming areas.

In the RF model, the reduction in training data led to a relative increase in overall classification error, but not in a linear pattern. The reduction of the size of the training data had no significant effect on the accuracy of the classifier prior to the threshold of 70%, which represented that a multivariate Random Forest classifier is not sensitive to the reduction in the training set. Some of the reasons for the situation might be because there were some classes of redundant training data, although the training data were decreased.

Acknowledgements

We sincerely thank two editors and two anonymous reviewers for their constructive suggestions that significantly strengthened this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by the National Key R&D Program of China [No. 2017YFA0604404] and the National Natural Science Foundation of China [Grant No. 41671398].

ORCID

Xiaoping Liu http://orcid.org/0000-0003-4242-5392
Yimin Chen http://orcid.org/0000-0003-1976-6041

References

Baker, C., R. Lawrence, C. Montagne, and D. Patten. 2006. “MAPPING WETLANDS and RIPARIAN AREAS USING LANDSAT ETM+ IMAGERY and DECISION-TREE-BASED MODELS.” *Wetlands* 26: 465–474. doi:10.1672/0277-5212(2006)26[465:MWARAU]2.0.CO;2.

Bounoua, L., R. DeFries, G. J. Collatz, P. Sellers, and H. Khan. 2002. “Effects of Land Cover Conversion on Surface Climate.” *Climatic Change* 52: 29–64. doi:10.1023/A:1013051420309.

Breiman, L. 2001. “Random Forests.” *Machine Learning* 45: 5–32. doi:10.1023/A:1010933404324.

Breiman, L., J. Friedman, R. Ohlsen, and C. Stone. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.

Brewster, C. C., J. C. Allen, and D. D. Kopp. 1999. “IPM from Space: Using Satellite Imagery to Construct Regional Crop Maps for Studying Crop—Insect Interaction.” *American Entomologist* 45: 105–117. doi:10.1093/ae/45.2.105.

Briem, G. J., J. A. Benediktsson, and J. R. Sveinsson. 2002. “Multiple Classifiers Applied to Multisource Remote Sensing Data.” *IEEE Transactions on Geoscience and Remote Sensing* 40: 2291–2299. doi:10.1109/TGRS.2002.802476.

Castillo-Santiago, M. A., M. Ricker, and B. H. J. de Jong. 2010. “Estimation of Tropical Forest Structure from SPOT-5 Satellite Images.” *International Journal of Remote Sensing* 31: 2767–2782. doi:10.1080/01431160903095460.

Chan, J. C.-W., and D. Paelinckx. 2008. “Evaluation of Random Forest and Adaboost Tree-Based Ensemble Classification and Spectral Band Selection for Ecotope Mapping Using Airborne Hyperspectral Imagery.” *Remote Sensing of Environment* 112: 2999–3011. doi:10.1016/j.rse.2008.02.011.

Chapman, D. S. 2010. "Random Forest Characterization of Upland Vegetation and Management Burning from Aerial Imagery." *Journal of Biogeography* 37 (1): 37–46. doi:10.1111/j.1365-2699.2009.02186.x.

Chen, J., J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, et al. 2015. "Global Land Cover Mapping at 30m Resolution: A POK-based Operational Approach." *ISPRS Journal of Photogrammetry and Remote Sensing* 103: 7–27. doi:10.1016/j.isprsjprs.2014.09.002.

Chen, Y., X. Li, X. Liu, and B. Ai. 2014. "Modeling Urban Land-Use Dynamics in a Fast Developing City Using the Modified Logistic Cellular Automaton with a Patch-Based Simulation Strategy." *International Journal of Geographical Information Science* 28: 234–255. doi:10.1080/13658816.2013.831868.

Chen, Y., X. Liu, X. Li, X. Liu, Y. Yao, G. Hu, X. Xu, and F. Pei. 2017. "Delineating Urban Functional Areas with Building-Level Social Media Data: A Dynamic Time Warping (DTW) Distance Based K-Medoids Method." *Landscape & Urban Planning* 160: 48–60. doi:10.1016/j.landurbplan.2016.12.001.

Crist, E. P., and R. C. Cicone. 1984. "A Physically-Based Transformation of Thematic Mapper Data—The TM Tasseled Cap." *IEEE Transactions on Geoscience and Remote Sensing* GE-22: 256–263. doi:10.1109/TGRS.1984.350619.

Dobra, A., and J. Gehrke. 2001. Bias Correction in Classification Tree Construction. In *Proceedings of the Seventeenth International Conference on Machine Learning, Williams College,* edited by. C. E. Brodley and A. P. Danyluk. Williamstown, MA: Williams College.

Dunn, J. C. 1973. "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters." *Journal of Cybernetics* 3: 32–57. doi:10.1080/01969727308546046.

Foley, J. A., N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, et al. 2011. "Solutions for a Cultivated Planet." *Nature* 478: 337–342. doi:10.1038/nature10452.

Foody, G. M., and M. K. Arora. 1997. "An Evaluation of Some Factors Affecting the Accuracy of Classification by an Artificial Neural Network." *International Journal of Remote Sensing* 18: 799–810. doi:10.1080/014311697218764.

Friedl, M. A., and C. E. Brodley. 1997. "Decision Tree Classification of Land Cover from Remotely Sensed Data." *Remote Sensing of Environment* 61: 399–409. doi:10.1016/S0034-4257(97)00049-7.

Friedl, M. A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang. 2010. "MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets." *Remote Sensing of Environment* 114: 168–182. doi:10.1016/j.rse.2009.08.016.

Gao, B. C. 1996. "NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space." *Remote Sensing of Environment* 58: 257–266. doi:10.1016/S0034-4257(96)00067-3.

Ghimire, B., J. Rogan, and J. Miller. 2010. "Contextual Land-Cover Classification: Incorporating Spatial Dependence in Land-Cover Classification Models Using Random Forests and the Getis Statistic." *Remote Sensing Letters* 1: 45–54. doi:10.1080/01431160903252327.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. 2006. "Random Forests for Land Cover Classification." *Pattern Recognition Letters* 27: 294–300. doi:10.1016/j.patrec.2005.08.011.

Gong, P., J. Wang, L. Yu, Y. Zhao, Y. Zhao, L. Liang, Z. Niu, et al. 2013. "Finer Resolution Observation and Monitoring of Global Land Cover: First Mapping Results with Landsat TM and ETM+ Data." *International Journal of Remote Sensing* 34: 2607–2654. doi:10.1080/01431161.2012.748992.

Haralick, R. M., K. Shanmugam, and I. Dinstein. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3: 610–621. doi:10.1109/TSMC.1973.4309314.

Jensen, J. R., and K. Lulla. 1987. "Introductory Digital Image Processing: A Remote Sensing Perspective." *Geocarto International* 2: 65. doi:10.1080/10106048709354084.

Johansen, K., N. C. Coops, S. E. Gergel, and Y. Stange. 2007. "Application of High Spatial Resolution Satellite Imagery for Riparian and Forest Ecosystem Classification." *Remote Sensing of Environment* 110: 29–44. doi:10.1016/j.rse.2007.02.014.

Jung, M., K. Henkel, M. Herold, and G. Churkina. 2006. "Exploiting Synergies of Global Land Cover Products for Carbon Cycle Modeling." *Remote Sensing of Environment* 101: 534–553. doi:10.1016/j.rse.2006.01.020.

Kauth, R. J. 1975. "The Tasseled Cap-A Graphic Description of Spectral-Temporal Development of Agricultural Crops as Seen by Landsat." In *Proceedings Sympposium on Machine Processing of Remotely Sensed Data*, 4B41–4B51. West Lafayette, MA: Purdue University.

Liang, X., X. Liu, X. Li, Y. Chen, H. Tian, and Y. Yao. 2018. "Delineating Multi-scenario Urban Growth Boundaries with a CA-Based FLUS Model and Morphological Method." *Landscape and Urban Planning* 177: 47–63.

Liu, X., G. Hu, Y. Chen, X. Li, X. Xu, S. Li, F. Pei, and S. Wang. 2018. "High-Resolution Multi-Temporal Mapping of Global Urban Land Using Landsat Images Based on the Google Earth Engine Platform." *Remote Sensing of Environment* 209: 227–239.

Liu, X., X. Liang, X. Li, and X. Xu. 2017. "A Future Land Use Simulation Model (FLUS) for Simulating Multiple Land Use Scenarios by Coupling Human and Natural Effects." *Landscape and Urban Planning* 168: 94–116.

Mas, J. F., and J. J. Flores. 2008. "The Application of Artificial Neural Networks to the Analysis of Remotely Sensed Data." *International Journal of Remote Sensing* 29: 617–663. doi:10.1080/01431160701352154.

Masek, J. G., E. F. Vermote, N. E. Saleous, R. Wolfe, F. G. Hall, K. F. Huemmrich, G. Feng, J. Kutler, and L. Teng-Kui. 2006. "A Landsat Surface Reflectance Dataset for North America, 1990-2000." *IEEE Geoscience and Remote Sensing Letters* 3: 68–72. doi:10.1109/LGRS.2005.857030.

Moore, I. D. 1991. "Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications." *Hydrol Processing* 5: 3–30. doi:10.1002/hyp.3360050103.

Mountrakis, G., J. Im, and C. Ogole. 2011. "Support Vector Machines in Remote Sensing: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 66: 247–259. doi:10.1016/j.isprsjprs.2010.11.001.

Ozdemir, I., A. D. Norton, Y. U. Ozkan, A. Mert, and O. Senturk. 2008. "Estimation of Tree Size Diversity Using Object Oriented Texture Analysis and Aster Imagery." *Sensors* 8. doi:10.3390/s8010464.

Pal, M. 2005. "Random Forest Classifier for Remote Sensing Classification." *International Journal of Remote Sensing* 26: 217–222. doi:10.1080/01431160412331269698.

Pittman, K., M. C. Hansen, I. Becker-Reshef, P. V. Potapov, and C. O. Justice. 2010. "Estimating Global Cropland Extent with Multi-Year MODIS Data." *Remote Sensing* 2: 1844–1863. doi:10.3390/rs2071844.

Rogan, J., J. Franklin, D. Stow, J. Miller, C. Woodcock, and D. Roberts. 2008. "Mapping Land-Cover Modifications over Large Areas: A Comparison of Machine Learning Algorithms." *Remote Sensing of Environment* 112: 2272–2283. doi:10.1016/j.rse.2007.10.004.

Schroff, F., A. Criminisi, and A. Zisserman. 2008. *Object Class Segmentation Using Random Forests*. Leeds, UK: BMVC.

Sesnie, S. E., P. E. Gessler, B. Finegan, and S. Thessler. 2008. "Integrating Landsat TM and SRTM-DEM Derived Variables with Decision Trees for Habitat Classification and Change Detection in Complex Neotropical Environments." *Remote Sensing of Environment* 112: 2145–2159. doi:10.1016/j.rse.2007.08.025.

Turner Ii, B. L., R. H. Moss, and D. L. Skole. 1993. *Relating Land Use and Global Land-Cover Change*. Barcelona: Human Dimensions of Global Environmental Change Programme.

Wagstaff, K. 2001. "Constrained K-Means Clustering with Background Knowledge." In *Proceedings of the 18th International Conference on Machine Learning (ICML 2001)*, 577–584. Williamstown, MA: Williams College.

Wolpert, D. H., and W. G. Macready. 1999. "An Efficient Method to Estimate Bagging's Generalization Error." *Machine Learning* 35: 41–55. doi:10.1023/A:1007519102914.

Yuan, F., M. E. Bauer, N. J. Heinert, and G. R. Holden. 2005. "Multi-Level Land Cover Mapping of the Twin Cities (Minnesota) Metropolitan Area with Multi-Seasonal Landsat TM/ETM+ Data." *Geocarto International* 20: 5–13. doi:10.1080/10106040508542340.

Zha, Y., J. Gao, and S. Ni. 2003. "Use of Normalized Difference Built-Up Index in Automatically Mapping Urban Areas from TM Imagery." *International Journal of Remote Sensing* 24: 583–594. doi:10.1080/01431160304987.

Zhang, Y., L. Wu, N. Neggaz, S. Wang, and G. Wei. 2009. "Remote-Sensing Image Classification Based on an Improved Probabilistic Neural Network." *Sensors* 9: 7516–7539. doi:10.3390/s90907516.