The Egyptian Journal of Remote Sensing and Space Sciences 23 (2020) 11–19

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

The Egyptian Journal of Remote Sensing and Space Sciences

journal homepage: www.sciencedirect.com

Monitoring and mapping rural urbanization and land use changes using Landsat data in the northeast subtropical region of Vietnam

Tuyen V. Ha [a,b,*], Mike Tuohy [a], Matthew Irwin [a], Pham V. Tuan [b]

[a] Institute of Agriculture and Environment, Massey University, Palmerston North 4474, New Zealand
[b] Faculty of Natural Resource Management, Thai Nguyen University of Agriculture and Forestry, Thai Nguyen City 25000, Viet Nam

A R T I C L E I N F O

Article history:
Received 31 October 2017
Revised 21 May 2018
Accepted 5 July 2018
Available online 8 December 2018

Keywords:
Rural land cover
Remote sensing
GIS
R programming

A B S T R A C T

Rapid land use change has taken place in many neighboring provinces of the capital of Vietnam such as Thai Nguyen province over the past 2 decades due to urbanization and industrialization. Deriving accurate and updated land cover and land-use change information is essential for the environmental monitoring, evaluation and management. In this study, a robust classification algorithm, Random Forest (RF) was employed in R programming to map and monitor temporal and spatial characteristics of urban expansion and land-use change in Thai Nguyen province, Vietnam. The results showed that there has been a substantial and uneven urban growth and a significant loss of forest and cropland between 2000 and 2016. Most of the conversion of agriculture and forest into built-up and mining uses were largely detected in rural regions and suburbs of Thai Nguyen. Further GIS analysis revealed that rapid urban and industrial expansion was spatially occurred in the southern rural portions and central area of the province. This study also demonstrates the potential of Landsat data and combination of R programming language and GIS to provide a timely, accurate and economical means to map and analyze temporal land cover and land use changes for future national and local land development planning.

1. Introduction

After the adoption of the national economic reform (Doi Moi) in 1986, Vietnam has been witnessed a significantly industrialized and urbanized transition process. The rapid growth of urbanization and industrialization in rural and suburb regions has long been considered a sign of national and local economic prosperity. However, the tremendous changes in the economic system negatively influenced changes in spatial structure and patterns of land cover and land use (Quang and Kammeier, 2002). The rapid expansion of urban and industrial zones on former cropland, forestland and grassland has resulted in the loss of a substantial amount of agricultural land and rural landscapes. This issue is of great significance for Vietnam because majority of its population live in rural areas and are dependent upon farming activities (General Statistics Department of Vietnam, 2015). Therefore, in recent years, Vietnam government had strict control over the conversion of agricultural uses to non-agricultural purposes, particularly growing rice areas (Vietnam Government, 2013).

Thai Nguyen, as an economic, educational and political center in the northeast region of Vietnam and connecting to Hanoi, has experienced rapid development over the past two decades. The temporally cumulative change in land cover/land-use has significantly changed the local landscape and environment such as forest loss (Giap et al., 2003) and landslides (Hieu et al., 2014). Given the scale and potential impact of land use change, there is an urgent need for evaluating the magnitude, pattern and type of land cover/land-use changes. However, little research has been done to determine the rate, trend and amount of land cover changes as well as the potential of satellite data for monitoring and mapping land cover in Thai Nguyen region, Vietnam. For this purpose, using free-cost satellite remote sensing data such as Landsat and Sentinel data to derive land cover/land-use change information is critically important in providing rapidly useful information for local and national land use development planning.

Remote sensing, with recent advances in the technology and an open-access data policy, together with increased temporal acquisition of data (e.g., few days to few weeks), can provide useful land cover/land-use information at a lower cost than traditional ground

Peer review under responsibility of National Authority for Remote Sensing and Space Sciences.
* Corresponding author at: Institute of Agriculture and Environment, Massey University, Palmerston North 4474, New Zealand.
E-mail address: tuyenmassey@gmail.com (T.V. Ha).

https://doi.org/10.1016/j.ejrs.2018.07.001

survey approach (Szuster et al., 2011). Earth observation satellite provides spatially consistent datasets over large areas with both high spatial detail and temporal frequency (Xiao et al., 2006). Landsat program, for example, has continuously provided high-quality imagery data for land science researchers and scientists to monitor and map temporal dynamics of global landscape changes since the 1970s (Hansen and Loveland, 2012). The analysis of these data can offer better understanding of the subtropical landscape patterns and interactions between human activities and natural ecosystems (Rawat and Kumar, 2015).

Since the first lunch of Landsat program in the 1970s, many different classification techniques (e.g., random forest and maximum likelihood classifiers) have been developed and extensively applied to derive land cover/land-use from remotely sensed data. This study chose random forests classifier because of its high accuracy and low risk of overfitting in comparison with traditional classification techniques (Ghimire et al., 2012; Pal, 2005; Rodriguez-Galiano et al., 2012a,b), and this approach has widely used in land cover/land-use mapping and analysis by international scholars in recent years (Gislason et al., 2006; Rodriguez-Galiano et al., 2012a,b). In Vietnam, however, the use of random forests method has received little interests, but the application of remotely sensed data has been increasingly grown. Seto and Fragkias (2007) adopted the remote sensing approach to map out land cover/land-use in two national areas of protected significance, while Beland et al. (2006) assessed the land cover/land-use changes on shrimp aquaculture development in Gia Thuy, Vietnam using Landsat data.

This study aims to provide quantitative maps of temporal land cover changes that have taken place in Thai Nguyen over the last two decades; to analyze its change patterns and major causes that influence land cover conversion; and to evaluate the accuracy of the RF algorithm for land cover classification in subtropical regions. The derived land cover maps and its temporal change analysis can be useful for the national and local land use planning and environment resource conservation missions.

2. Study area

The study area is Thai Nguyen province, which lays in the northeast region of Vietnam and covers 9 local districts and parts of the Tam Dao national park with a total area of 3534 km^2 (Fig. 1). The study area is prominently characterized by subtropical climates (e.g., warm to hot in summers and cool to mild in winters) with relatively strong inter-seasonal variability. The mean annual temperature and precipitation are comparably high, approximately 25 °C and 2250 mm respectively, while elevations range from 4 m to 1587 m above the sea level. The climatic and topographic variabilities produced complex characteristics of subtropical landscapes such as semi-deciduous, scrubs and mountain ranges (Phuong, 2007). At present, forest and scrubs are predominant and cover most of the north and northwest, while agriculture and urban/built-up primarily characterize lower areas and drainage corridors in the southern and central province.

The historical vegetation of the study area was described as dense and diverse, with a prominence of the eucalyptus trees and shrubs (Hoang Ngoc Ha, 2008). Clement and Amezaga (2008) revealed that forest restoration programs implemented between 1985 and 2005 has significantly increased forest cover from 29.2% to 37.6% respectively. However, the recent expansion of industrial and urban zones and other infrastructure networks has increased pressure on local land-use. Conversion of agriculture and forest was largely occurred to serve industrial and urban developments in the southern areas and suburb zones of districts and towns as well as other rural areas.

The study area is home to nearly 1.2 million people in 2016, which represents roughly 1.3% of the Vietnamese population. The local population increased by 5.6% between 1997 and 2016 (Hao Ho, 2015), and the percentage of rural population against urban population remained high. For example, the rural population in 2010 was approximately 74%, while urban residents only accounted for 26% (General Statistics Department of Vietnam, 2011). However, the movement of residential and commercial land use to rural areas at the periphery of urban areas has significantly increased, and this trend is expected to continue occurring. Especially in the last 10 years of investment incentive and priority policies implementation; as a result, urban and industrial expansion has been very substantial. The local government estimated that urban residents will account for 48% of the total population by 2025, whereas urban areas are expected to increase by 9% between 2015 and 2025 (Thai Nguyen People's Committee, 2009)

3. Materials and methods

This section describes the data acquisition and preprocessing procedures and analysis methodologies used. Two independent datasets, namely DEM and Landsat, were primarily used to derive predictor variables for land cover classification. The flow diagram in Fig. 2 outlines the classification analysis and processing steps.

3.1. Data preparation

In this study, several satellite data and local maps (Table 1) were collected from various sources such as U.S Geological Survey, DIVA-GIS and Thai Nguyen Department of Natural Resources and Environment. The Landsat Enhanced Thematic Mapper (ETM+) images were acquired in during winter time (4th October 2000), while Landsat Operational Land Imager (OLI) captured the study area in early winter (6th October 2016). These satellite sensors were designed to collect reflected radiation from the earth's surface features in the visible, near-infrared and thermal near-infrared portions of the electronic spectrum. Both acquired images are nearly cloud free (<10% cloud cover). Thai Nguyen province is entirely contained within Landsat ETM+ and Landsat OLI path 127 and row 45 with Landsat Worldwide Reference System-2 (WRS-2).

These images were pre-processed and performed in ENVI version 5.0 for clipping the study area by taking geo-referenced boundary of the Thai Nguyen map as Area of Interest (AOI). Landsat and DEM data were geometrically corrected using an independent set of ground control points (Zanter, 2015), while radiometric, atmospheric and color composites were implemented in ENVI software to produce a consistent form of the datasets. Also, Landsat level-1 imagery was processed to standard parameters included co-registered images and referenced coordinate systems.

Reference data were collected for assessing the classification accuracy of the Landsat 8 and Landsat 7. The location of the 200 ground control points was recorded in geographic coordinate system using a GPS receiver between December 2016 and January 2017 for the accuracy assessment of classified Landsat 8. The size of a given ground control point was visually estimated with four Landsat pixel cells (3600 m^2) to represent each land cover class of the study area (Fig. 3). In addition, during the field trip we carried out informal interviews with land owners and village leaders to obtain information on previous land cover and land use. This information was complemented with Google Earth's historical imagery and existing topographic map to produce 300 reference points for the accuracy evaluation of the Landsat 7. The training polygons were selected for each land cover class with the support of a 3-m spatial resolution PlanetScope imagery (Planet Team,

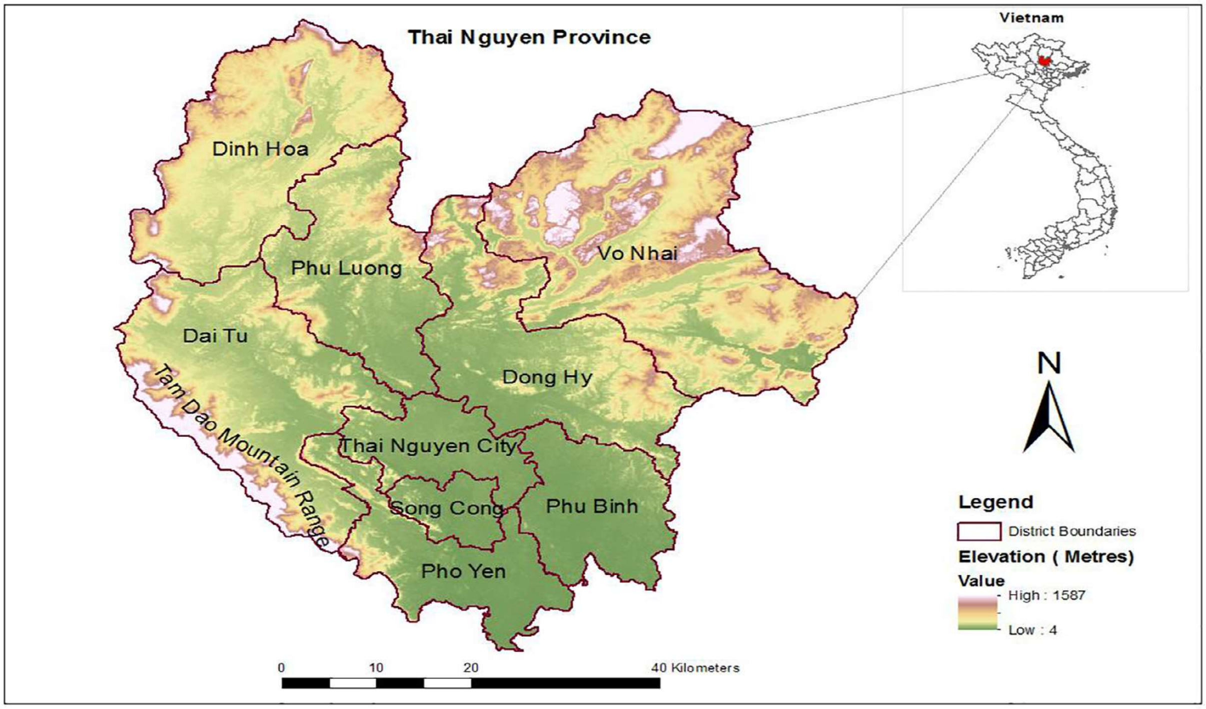

Fig. 1. A map of the study area with 9 districts.

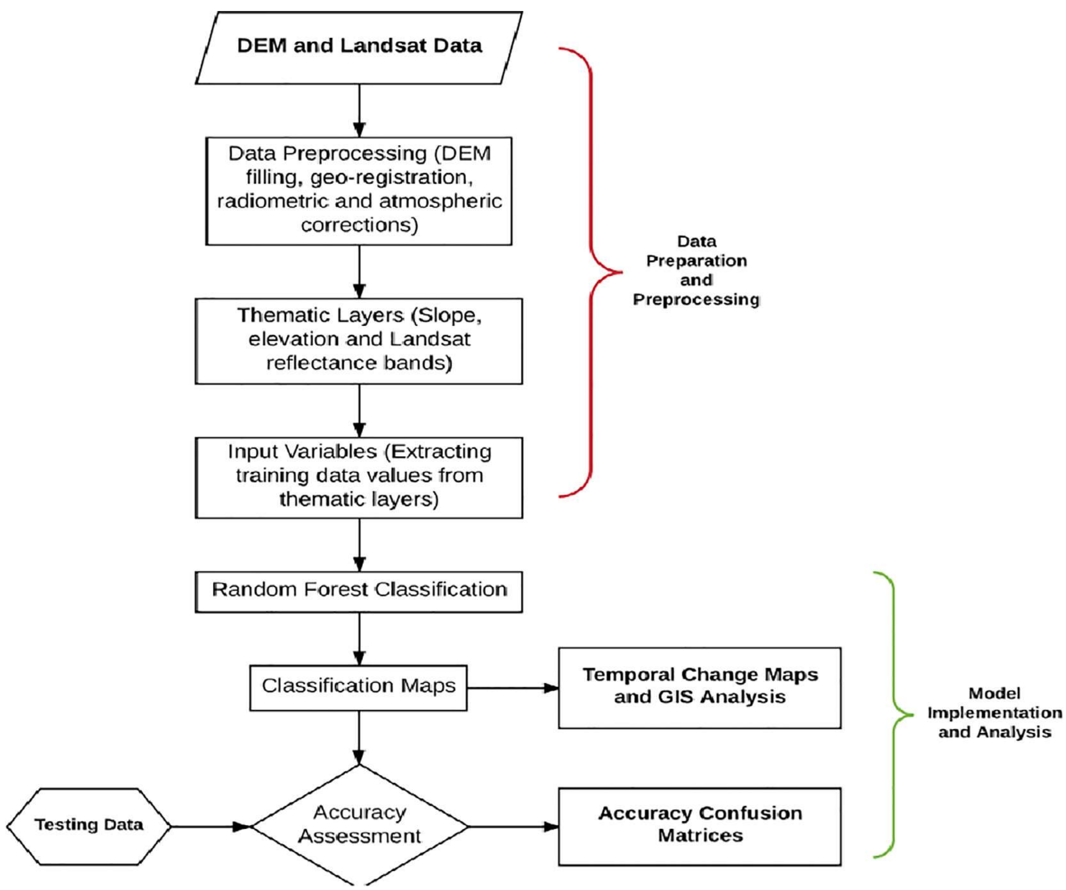

Fig. 2. A flowchart describing the process of classifying and analyzing multi-source remote sensing data used the random forest algorithm (Data preprocessing and temporal change mapping was carried out using ENVI and ArcMap respectively, while R programming was used to extract input variables, train the model and evaluate its accuracy).

Table 1
Data layers are used in this study.

Data	Data Category	Spatial Resolution (m)	Data of Acquisition	Data Source
Landsat ETM+	Imagery	30	4/10/2000	USGS[a]
Landsat OLI/TIR	Imagery	30	6/10/2016	USGS[a]
Road and railway	Vector			DIVA-GIS[b]
Topographic map				DNRE[c]
DEM	ASTER GDEM	30		USGS[a]

[a] U.S. Geological Survey.
[b] DIVA-GIS.
[c] Thai Nguyen Department of Natural Resources and Environment.

Fig. 3. Examples of agriculture; harvested rice field (left) and corn fields (right).

2017) and Google Earth's historical imagery for Landsat 8 and Landsat 7 respectively. Approximately 3000 training pixels for each class were recorded in ArcMap for model training of Landsat 8 and Landsat 7. To increase the accuracy of land cover mapping of the two images, slope and elevation variables were included in the model. These two variables were extracted from slope and elevation data using the above-mentioned training data points in R programming, and are incorporated with all other variables to train the model.

3.2. Classification schemes

This study initially identified seven land cover classes (built-up land, agricultural land, forest land, water bodies, mining extraction, shrubs land and tea plantation) based on unsupervised classifications, band composites and NDVI calculation. However, during the field trip it was observed that the area of tea plantation accounted for only small portion and noticeably mixed spectrally, while shrubs and forest are not clearly distinguished on the ground. Tea plantations are usually grown at the foot of mountains or under the canopy of other tall trees such as fruit trees and timber species (Fig. 4), which make it challenging to separate forest, shrubs and tea plantation. Similarly, the separation of shrubs and forest by the classification algorithm is not easy due to similar spectral characteristics of features. Therefore, tea plantations and shrubs were treated as forest in the classification scheme. Five land cover/land-use categories were finally identified in the study area, namely agriculture, forest, mining, urban/built-up and water (Table 2).

Fig. 4. Tea plantations at the foot of mountain (left) and tea intercropping (right).

Table 2
Description of land cover/land-use categories in Thai Nguyen.

Classification Scheme	Description
Urban/built-up land	Rural houses and urban buildings
	Road network and utilities
	Industrial zones, commercial complexes and current construction areas
	Planned areas for urban and industrial construction
Agriculture	Crop and pasture
	Orchards, groves, nurseries and horticultural areas
	Other farming land
Forest	Deciduous, evergreen forests, shrubs and mixed forests
Water	Streams, rivers, canals, and reservoirs
	Lakes and ponds
Mining	Iron and coal mines
	Bare exposed rocks, transitional areas and mixed barren land
	Open forest and dead tree areas

3.3. Random forest classification

The RF algorithm is an ensemble learning classifier, which has received the increasing interests over the past decades because of its high overall accuracy and robust to noise and overfitting (Breiman, 2001; Rodriguez-Galiano, et al., 2012a,b). This approach are also robust to massive input variables dataset and the variable importance of the classification (Rodriguez-Galiano, et al., 2012a,b).

Random forest classification is based on decision tree approach (Breiman, 2001), where each tree is generated by randomly choosing a subset of sample from original training dataset with replacement known as a bootstrap approach. Each tree will give a vote to the most frequent class of land cover/land-use to the training variables. The building of different trees from various subsets of the original dataset has advantages to increasing the stability and accuracy of the model. This bootstrapping technique makes each individual tree grow containing a certain proportion of the selected training samples. The sample, which is not used in the training sample, is used to test the accuracy of the model called out of bag error (OOB). It is noted that each tree produces an OOB for any given training samples using bootstrap technique. Further reading of random forest model (bootstrap approach) can be found from a study of Breiman (2001) and (Rodriguez-Galiano, et al., 2012a,b).

In this study, the random forests are performed with a combination of several decision trees (200 trees). Each tree takes approximately two thirds of the training samples to train the model and fully grown without any pruning, while the remaining ones are used for internally testing its performance (OOB). By growing the forest, the number of trees (*Ntree*) and input variables (*Mtry*) for the best split at each node needs to be specified (Belgiu and Drăguţ, 2016). This study chose 200 trees and kept *Mtry* as the square root of input variables after many trials. This chosen model produced a stable accuracy rate and reduced computational complexity and correlation between trees. The importance of variables can be estimated internally by permuting the value of out of bag samples (samples not used for training) for a certain class using Mean Decrease Gini. This indicator measures how much a variable decreases the Gini Impurity metric in a certain class (Belgiu and Drăguţ, 2016). Also, the time for running the RF model can be computed as:

$$T\sqrt{MNlog(N)}$$

where T is the number of trees, M is the number of features used for splitting at each node, and N is the number of observation samples (Breiman, 2001).

The classifications were performed using a "randomForest" package (Breiman et al., 2011), which is freely available in R programing language with a great number of built-in functions.

3.4. Accuracy assessment and change detection

Accuracy assessment involves identifying correctly classified pixels that are identically identified in the ground and by the algorithm. A confusion matrix was chosen for this study due to its popularity and simplicity in the field of remote sensing (Foody, 2002). The confusion matrix was computed to indicate how many pixels are correctly classified in each class. Based on the information from the contingency table, many measures can be derived to assess the fitness of a model to particular context, including an overall accuracy, Kappa statistic and producers' and users' accuracy. Overall accuracy (OA) for a particular classified image was calculated by summing the number of correctly classified pixels and dividing the total number of pixels are located along the upper-left to lower-right diagonal of the confusion matrix (Story and Congalton, 1986).

$$OA = \frac{\sum_i^p N_i}{T}$$

where p is the number of classes, N_i is the sum of correctly classified pixels, and T is the total number of reference pixels. The kappa coefficient (k) measures the agreement between classified map and reference values. A kappa value of + 1 represents perfect agreement, while a value of 0 represents no agreement. The kappa coefficient is computed as follows:

$$k = 1 - \frac{1 - p_0}{1 - p_e}$$

where p_0 represents the proportion of correct agreement in the test dataset and p_e is the proportion of agreement that is expected by chance. Confusion matrices were produced using the "caret" package in R programming language (Kuhn, 2008).

Post-classification change detection technique was applied to monitor and map land cover change and to compare changes in land cover of the two classified images. According to (Mas, 1999; Shalaby and Tateishi, 2007), the comparison of two independently classified images with distinct dates using post-classification detection was probably the most effective technique as it can minimize the problem of normalizing atmospheric and sensor variability. The post-classification change detection also provides specific nature of changes between the classified images and "from-to" change information (El-Kawy et al., 2011). In this study, a matrix of land cover/land-use change was produced using ArcMap.

4. Results and discussion

4.1. Land cover/land use change

Land cover classified maps were produced for two years (Fig. 5), 2000 and 2016 respectively, and the statistics of individual class area and its temporal change was also derived (Table 3).

Overall, there is a substantial change in land cover/land-use between 2000 and 2016 in the study area, and these changes largely occurred in southern rural regions and suburbs. To specify, built-up area experienced the highest growth, from only 22.0 km^2 to 292.3 km^2 (an increase of 1228.8%) between 2000 and 2016 respectively, while mining land increased about 31 km^2 (734.1%). By contrast, forest, agriculture and water all decreased. While forest and agriculture decreased by 183.0 km^2 (10.3%) and 116.5 km^2 (7.8%) respectively, water declined approximately 7.2 km^2 (14.1%). Unexpectedly, water area comparably decreased

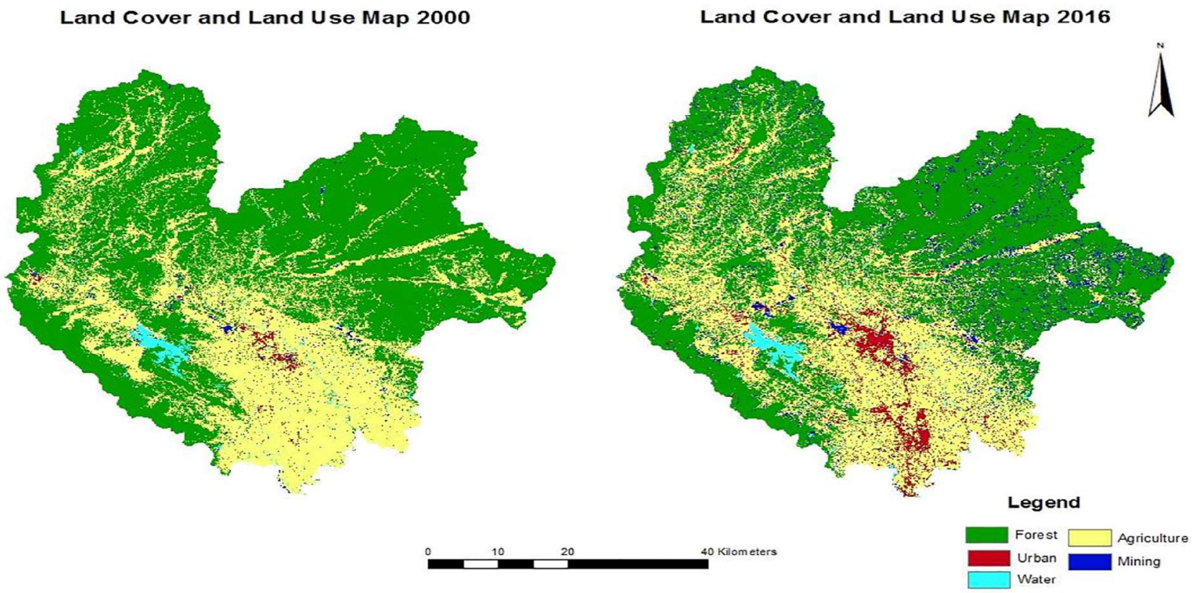

Fig. 5. Land cover classification products derived from 2000 and 2016 Landsat data for Thai Nguyen province.

Table 3
Summary of classification areas for 2000 and 2016.

Land Cover Class	2000		2016		Temporal Change (2000–2016)	
	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%
Agriculture	1128.7	32.1	1012.2	28.7	−116.5	−10.3
Built-up	22.0	0.6	292.3	8.3	270.3	1228.8
Mining land	4.2	0.1	35.2	1.0	31.0	734.1
Forest	2320.1	65.8	2137.1	60.7	−183.0	−7.8
Water	51.3	1.5	44.1	1.3	−7.2	−14.1

over the period. This is believed due to variations in precipitation, water levels of lakes and possible classification errors. However, this finding reflected the current status of surface water shortages reported by Ministry of Natural Resource and Environment (2014), and therefore classification errors and water variations are not likely to account for this decrease.

4.2. Classification accuracy assessment

Overall accuracy, Kappa statistic, producers' accuracy and users' accuracies were derived from confusion matrix to assess the accuracy of classified maps from Landsat data and are summarized in Table 4. According to (Gislason et al., 2006), the RF classifier was commonly reported as an alternative for land cover classification because of its high accuracy derived products. In this study, the overall accuracies for 2000 and 2016 are very high, at 93.4% and 94.1% and Kappa coefficients 0.91 and 0.92 respectively. Producers' accuracies of individual classes for 2000 are relatively high, except for mining extraction categories with low accuracy of 73.2%, while users' accuracies are also high for all classes with a range from 87.2% to 98.0%. Similarly, producers' and users' accuracies of individual classes in 2016 are very high, ranging from 89% to 100%.

The overall accuracy for land cover/land-use maps is often regarded as acceptable above 85% (Anderson, 1976; Thomlinson et al., 1999), while Lins and Kleckner (1996) set a higher standard requiring 90% accuracy. In this study, the overall accuracies for both 2000 and 2016 derived images using the RF were even better. The result of this study in terms of overall accuracy was compatible with previous land cover classification research using the RF technique (Erbek et al., 2004; Kavzoglu and Mather, 2003; Paola and Schowengerdt, 1995). Furthermore, the RF also has advantage of identifying the importance of variables and the effect tree sizes on the model. Fig. 6 showed that elevation and Landsat OLI band 5 (near-infrared band) contributed the most to the model, while Landsat OLI bands 3 and 4 had least contribution to the model. The model performance was relatively stable between 100 and 200 decision trees (Fig. 7). The results of this study showed that adding more decision trees may not make any significant difference in the accuracy of the model, but it can increase computational power and complexity (Rodriguez-Galiano et al., 2012a,b). Especially the Fig. 7 showed that there is a small difference in the model accuracy between 100 and 200 trees.

Table 4
Summary of Landsat ETM+ and Landsat OLI/TIR classification accuracies (%).

Land cover class	Landsat data			
	2000		2016	
	Producer's	User's	Producer's	User's
Forest	98.2	95.3	94.0	100
Agriculture	95.4	88.7	93.2	92.0
Built-up Area	87.3	96.9	97.9	90.3
Extraction	73.2	87.2	88.9	92.0
Water	99.5	97.7	95.6	100
Overall accuracy	93.4		94.1	
Kappa statistic	0.91		0.92	

It is also important to note that although overall accuracies of the two classified images was high, the low individual accuracies for mining land and built-up still occurred for classified map 2000, at 73.21% and 87%. This could be explained by the fact that

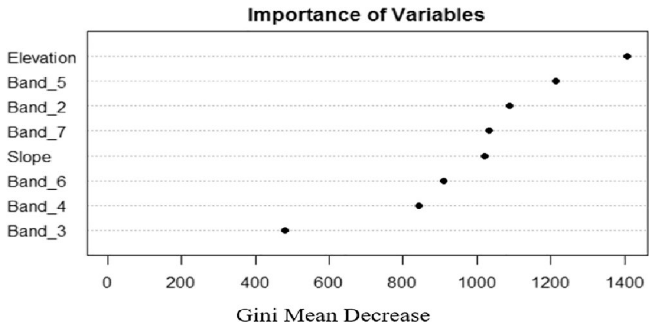

Fig. 6. The importance of variables identified by the RF model.

spectral signature of coal mine land is likely confused with that of old building roofs, while harvested fields surrounding the city were resulted in spectrally confused with built-up area.

4.3. Land cover change patterns

The advantages of products derived from satellite remote sensing include the calculation of the statistics and capability to display the distribution of temporal changes. In this study, a matrix of land cover/land-use changes (Table 5) and a map displaying the spatial distribution of these changes (Fig. 8) were created. Fig. 7 shows that built-up/urban uses increased remarkably between 2000 and 2016 in Thai Nguyen central city and southern rural areas (e.g., Pho Yen and Song Cong districts), while the landscape of north and east regions still stay relatively stabilized, agriculture and forest loss was mapped. The rapid growth of urban and mining uses come from both agriculture and forest areas. For example, around 231 km^2 of agriculture and 42 km^2 of forest was converted into built-up land, while mining land encroached approximately 10 km^2 of agriculture and 21 km^2 of forest during 16 years. In addition, GIS analysis revealed a strong relationship between newly-developed area expansion and proximity to highways. Approximately 69.6% (100.2 km^2) of newly-built areas in this land cover classification occurred within 2 km from main roads (Fig. 8), while nearly 96% (137.6 km^2) of newly built-up area was detected within a 5-km radius from the centre of Thai Nguyen city.

Almost all changes in land cover/land-use in Thai Nguyen have taken place in suburbs (middle) and the southern rural regions (Song Cong and Pho Yen districts). Interestingly, growth was largely concentrated in a strip from the southern perimeter following the Hanoi-Thai Nguyen national highway QL3. This highway is a major connector between Hanoi capital, Thai Nguyen and other northeast provinces of Vietnam. This pattern clearly reflected the recent developments of the province, which focused on urbanization and industrialization forwards to southern rural area. Between 2010 and 2016, industrial zones and infrastructure construction have been significant in Pho Yen district, which was previously rural landscapes. Samsung Electronics Vietnam Thai Nguyen Company Limited occupied about 150 ha, which was converted from agriculture and forest, while more than 50 other industrial zones also occupied in formerly rural land (Business Forum, 2016). Also, the rapid growth of urban population and university infrastructure in Thai Nguyen city has resulted in the expansion of road network and residential areas.

Long-term development policies of the province (Vietnam Government, 2007) were targeted to transform Thai Nguyen into a modernized and industrialized province by 2020. The industry, construction and service activities were targeted to account for about 87% of province's GDP, while agriculture, fishery and forest will be accounted for approximately 13% by 2020. This policy will

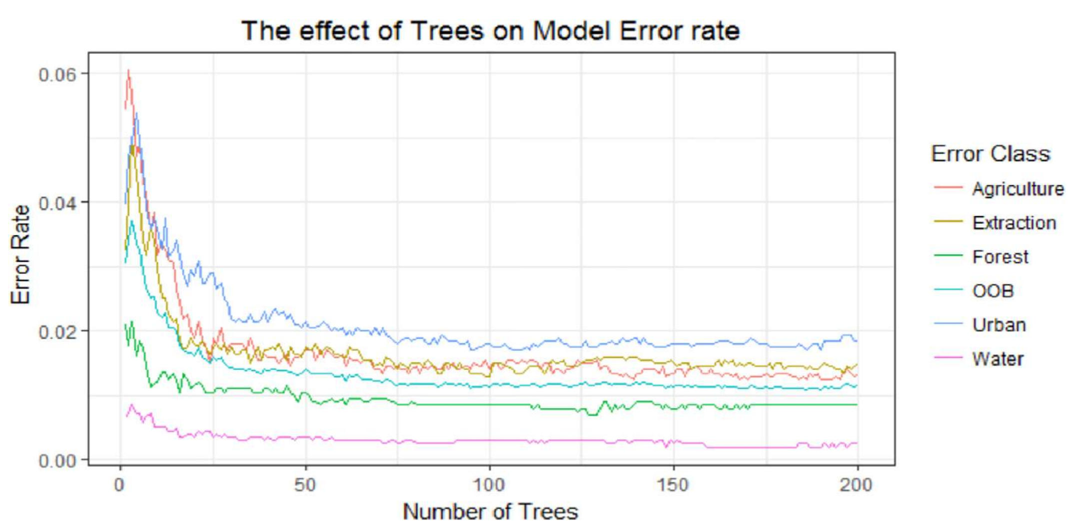

Fig. 7. The effect of tree size on the RF model accuracy.

Table 5
Matrix of land cover/land-use and changes (area in km^2) between 2000 and 2016.

2016	2000					Class Total
	Agriculture	Built-up	Mining Land	Forest	Water	
Agriculture	726.8	3.7	1.5	271.4	8.7	1012.2
Built-up	231.3	16.9	0.7	42.1	1.2	292.3
Mining Land	10.1	0.7	1.3	21.1	1.9	35.2
Forest	155.3	0.2	0.2	1979.2	2.1	2137.1
Water	4.3	0.2	0.3	2.1	37.2	44.1
Class Total	1127.8	21.9	4.2	2315.8	51.1	3521.1

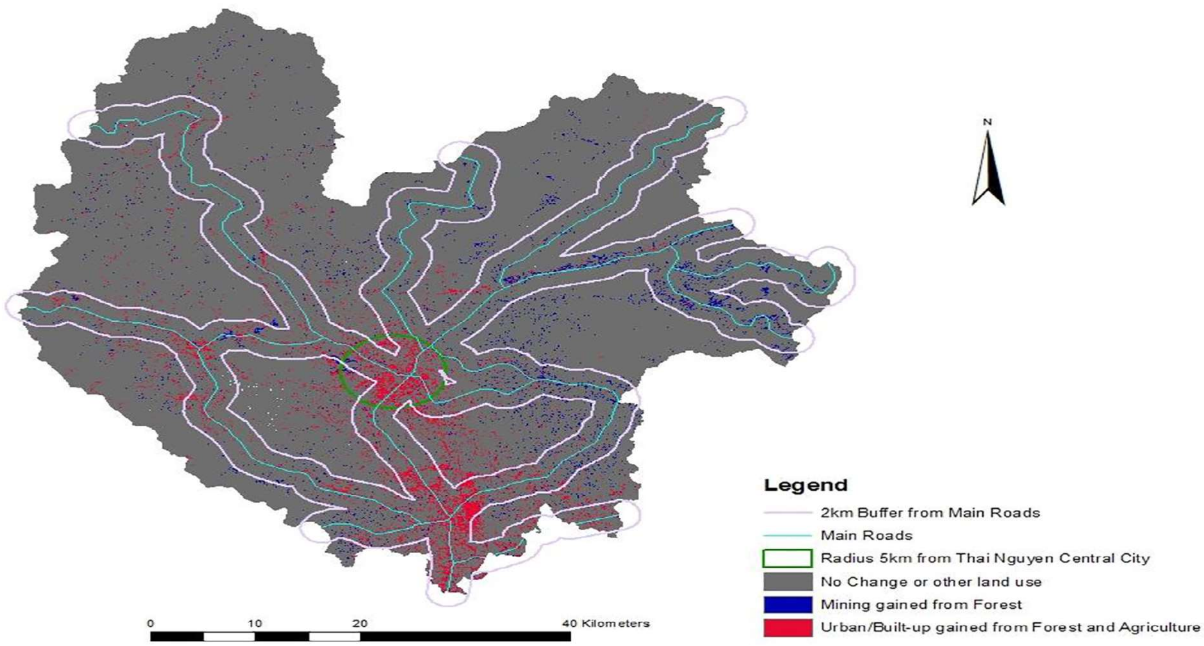

Fig. 8. Major changes in three land cover/land-use categories between 2000 and 2016.

likely continue to facilitate the conversion of agriculture and forest land into built-up area (e.g., industrial zones and urbanized infrastructure) if land cover/land-use information is not adequate or taken into account. This explains the importance of integrating GIS, R programming and remote sensing in monitoring and mapping land cover/land-use changes to provide timely and accurate information for sustainable land use development.

5. Conclusions

The combination of satellite remote sensing, R programming and GIS demonstrated the potential of accurate classification of remote sensing data over the large areas and the informative display of spatial changes to provide timely and accurate land cover information for efficient land management and policy decisions. The RF classifier has advantages of producing more accurate and stable overall and individual accuracies, and therefore offers the opportunity for better resource management and sustainable land use development.

The derived land cover products from Landsat data indicated that Landsat data can be used to successfully to map and monitor land cover/land-use changes with a high accuracy. Overall, a major change in land cover/land-use categories has taken place in Thai Nguyen province, particularly in southern rural regions over last 16 years. Agriculture has been converted into built-up land, mining extraction has expanded into forest land while water had little change. The main causes of land cover changes are due to recent development patterns of expansion of mining activities, industrial and residential zones in formerly agriculture and forest lands.

Conflict of interest

None.

Acknowledgements

We would like to thank Mr. Hoang and Hung for assisting our fieldwork and Dr. Gabor for his comments and thoughts to this manuscript.

Funding information

This study was supported by New Zealand ASEAN Scholars Awards (NZAid Scholarship).

References

Anderson, J.R., 1976. A Land Use and land Cover Classification System for Use with Remote Sensor Data. US Government Printing Office.
Beland, M., Goita, K., Bonn, F., Pham, T., 2006. Assessment of land-cover changes related to shrimp aquaculture using remote sensing data: a case study in the Giao Thuy District, Vietnam. Int. J. Remote Sens. 27 (8), 1491–1510.
Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.
Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.
Breiman, L., Cutler, A., Liaw, A., Wiener, M., 2011. Package'randomForest'. software available at URL: <http://stat-www. berkeley. edu/users/breiman/RandomForests>.
Business Forum, 2016. Thai Nguyen Socio-economic Development Planning.
Clement, F., Amezaga, J.M., 2008. Linking reforestation policies with land use change in northern Vietnam: why local factors matter. Geoforum 39 (1), 265–277.
El-Kawy, O.A., Rød, J., Ismail, H., Suliman, A., 2011. Land use and land cover change detection in the western Nile delta of Egypt using remote sensing data. Appl. Geogr. 31 (2), 483–494.
Erbek, F.S., Özkan, C., Taberner, M., 2004. Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. Int. J. Remote Sens. 25 (9), 1733–1748.
Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80 (1), 185–201.
General Statistics Department of Vietnam, 2015. Population in rural and urban areas.
General Statistics Department of Vietnam, 2011. Population and Employment. Vietnam: Statistical Publishing House Retrieved from <http://www.gso.gov.vn/default.aspx?tabid=512&ItemID=11973>.
Ghimire, B., Rogan, J., Galiano, V.R., Panday, P., Neeti, N., 2012. An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA. GIScience Remote Sens. 49 (5), 623–643.
Giap, D.H., Yi, Y., Cuong, N.X., Luu, L.T., Diana, J.S., Lin, C.K., 2003. Application of GIS and remote sensing for assessing watershed ponds for aquaculture development in Thai Nguyen, Vietnam. Paper presented at the Map Asia Conference.
Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random forests for land cover classification. Pattern Recogn. Lett. 27 (4), 294–300.
Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sens. Environ. 122, 66–74.
Hieu, N.M., Sassa, K., Hamasaki, E., Dang, K., Miyagi, T., 2014. Analysis of a deep-seated landslide in the Phan Me coal mining dump site, Thai Nguyen Province, Vietnam. In: Landslide Science for a Safer Geoenvironment. Springer, pp. 373–377.

Hao Ho, 2015. Local Geography of Thai Nguyen. Retrieved from <https://sites.google.com/site/dialitinhthainguyen/home/dan-so>.

Hoang Ngoc Ha, 2008. Status and Solutions for forest development in Hoa Binh Commune, Dong Hy District, Thai Nguyen.

Kavzoglu, T., Mather, P.M., 2003. The use of backpropagating artificial neural networks in land cover classification. Int. J. Remote Sens. 24 (23), 4907–4938.

Kuhn, M., 2008. Caret package. J. Stat. Software 28 (5), 1–26.

Lins, K., Kleckner, R., 1996. Land cover mapping: An overview and history of the concepts. Gap analysis: A landscape approach to biodiversity planning, 57–65.

Mas, J.-F., 1999. Monitoring land-cover changes: a comparison of change detection techniques. Int. J. Remote Sens. 20 (1), 139–152.

Ministry of Natural Resource and Environment, 2014. Thai Nguyen: Surface Water Distribution and Protection.

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

Paola, J.D., Schowengerdt, R.A., 1995. A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. IEEE Trans. Geosci. Remote Sens. 33 (4), 981–996.

Phuong, V.T., 2007. Forest environment of Vietnam: features of forest vegetation and soils. For. Environ. Mekong River Basin, 189–200.

Planet Team, 2017. Planet Application Program Interface: In Space for Life on Earth.

Quang, N., Kammeier, H.D., 2002. Changes in the political economy of Vietnam and their impacts on the built environment of Hanoi. Cities 19 (6), 373–388.

Rawat, J., Kumar, M., 2015. Monitoring land use/cover change using remote sensing and GIS techniques: a case study of Hawalbagh block, district Almora, Uttarakhand, India. Egypt. J. Remote Sens. Space Sci. 18 (1), 77–84.

Rodriguez-Galiano, V., 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.

Seto, K.C., Fragkias, M., 2007. Mangrove conversion and aquaculture development in Vietnam: a remote sensing-based approach for evaluating the Ramsar Convention on Wetlands. Global Environ. Change 17 (3–4), 486–500.

Shalaby, A., Tateishi, R., 2007. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. Appl. Geogr. 27 (1), 28–41.

Story, M., Congalton, R.G., 1986. Accuracy assessment: a user's perspective. Photogramm. Eng. Remote Sens. 52 (3), 397–399.

Szuster, B.W., Chen, Q., Borger, M., 2011. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. Appl. Geogr. 31 (2), 525–532.

Thai Nguyen People's Committee, 2009. Decision on the approval of Urban System Development Planning. Thai Nguyen, Vietnam: Thai Nguyen Portal.

Thomlinson, J.R., Bolstad, P.V., Cohen, W.B., 1999. Coordinating methodologies for scaling landcover classifications from site-specific to global: steps toward validating global map products. Remote Sens. Environ. 70 (1), 16–28.

Vietnam Government, 2007. Approval of Socio-economic Development Strategy for Thai Nguyen province till 2020. Vietnam Retrieved from <http://www.chinhphu.vn/portal/page/portal/chinhphu/hethongvanban?class_id=1&mode=detail&document_id=23936>.

Vietnam Government, 2013. Land Law. Vietnam.

Xiao, J., Shen, Y., Ge, J., Tateishi, R., Tang, C., Liang, Y., Huang, Z., 2006. Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. Landscape Urban Plann. 75 (1), 69–80.

Zanter, K., 2015. Landsat 8 (L8) Data Users Handbook Survey. Department of the Interior US Geological.