VIETNAM NATIONAL UNIVERSITY, HANOI

UNIVERSITY OF ENGINEERING AND TECHNOLOGY

MAN DUC CHUC

RESEARCH ON LAND-COVER CLASSIFICATION METHODOLOGIES FOR OPTICAL SATELLITE IMAGES

MASTER THESIS IN COMPUTER SCIENCE

Hanoi – 2017

VIETNAM NATIONAL UNIVERSITY, HANOI

UNIVERSITY OF ENGINEERING AND TECHNOLOGY

MAN DUC CHUC

RESEARCH ON LAND-COVER CLASSIFICATION METHODOLOGIES FOR OPTICAL SATELLITE IMAGES

DEPARTMENT: COMPUTER SCIENCE

MAJOR: COMPUTER SCIENCE

CODE: 60480101

MASTER THESIS IN COMPUTER SCIENCE

SUPERVISOR: Dr. NGUYEN THI NHAT THANH

Hanoi – 2017

PLEDGE

I hereby undertake that the content of the thesis: "**Research on Land-Cover classification methodologies for optical satellite images**" is the research I have conducted under the supervision of Dr. Nguyen Thi Nhat Thanh. In the whole content of the dissertation, what is presented is what I learned and developed from the previous studies. All of the references are legible and legally quoted.

I am responsible for my assurance.

Hanoi, day month year 2017
Thesis's author

Man Duc Chuc

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisor, Dr. Nguyen Thi Nhat Thanh. She has given me the opportunity to pursue research in my favorite field. During the dissertation, she has given me valuable suggestions on the subject, and useful advices so that I could finish my dissertation.

I also sincerely thank the lecturers in the Faculty of Information Technology, University of Engineering and Technology - Vietnam National University Hanoi, and FIMO Center for teaching me valuable knowledge and experience during my research.

Finally, I would like to thank my family, my friends, and those who have supported and encouraged me.

This work was supported by the Space Technology Program of Vietnam under Grant VT-UD/06/16-20.

Hanoi, day month year 2017

Man Duc Chuc

Content

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

In this chapter, I briefly present an introduction to remote sensing images and its applications in different research areas. Furthermore, the problem of land cover classification is also presented. Current progress and challenges in land cover classification are discussed. Finally, motivations and problem statement of the research are shown in the end of the chapter.

1.1. Motivation

Remotely-sensed images have been used for a long time in both military and civilization applications. The images could be collected from satellites, airborne platforms or Unmanned Aerial Vehicles (UAVs). Among the three, satellite images have gained popularity due to large coverage, available data and so on. In general, remotely-sensed images store information about Earth object's reflectance of lights, i.e. Sun's light in passive remote sensing [1]. Therefore, the images contain itself lots of valuable information of the Earth's surface or even under the surface.

Applications of remotely-sensed images are diverse. For example, satellite images could be used in agriculture, forestry, geology, hydrology, sea ice, land cover mapping, ocean and coastal [1]. In agriculture, two important tasks are crop type mapping and crop monitoring. Crop type mapping is the process of identification crops and its distribution over an area. This is the first step to crop monitoring which includes crop yield estimation, crop condition assessment, and so on. To these aims, satellite images are efficient and reliable means to derive the required information [1]. In forestry, potential applications could be deforestation mapping, species identification and forest fire mapping. In the forest where human access is restricted, satellite imagery is an unique source of information for management and monitoring purposes. In geology, satellite images could be used for structural mapping and terrain analysis. In hydrology, some possible applications cloud be flood delineation and mapping, river change detection, irrigation canal leakage detection, wetlands mapping and monitoring, soil moisture monitoring, and a lot of other researches. Iceberg detection and tracking is also done via satellite data. Furthermore, air pollution and meteorological monitoring

could be possible from satellite perspective. In general, many of the applications more or less relate to land cover mapping, i.e. agriculture, flood mapping, forest mapping, sea ice mapping, and so on.

Land cover (LC) is a term that refers to the material that lies above the surface of the Earth. Some examples of land covers are: plants, buildings, water and clouds. Land cover is the thing that reflects or radiates the Sun's lights which then be captured by the satellite's sensors. Land use and land cover classification (LULCC) has been considering as one of the most traditional and important applications in remote sensing since LULCC products are essential for a variety of environmental applications [2]. Figure 1 shows a land cover map for Mekong river delta, Vietnam in 2012 derived from MODIS images [3]. This map shows distribution of rice lands in the region.

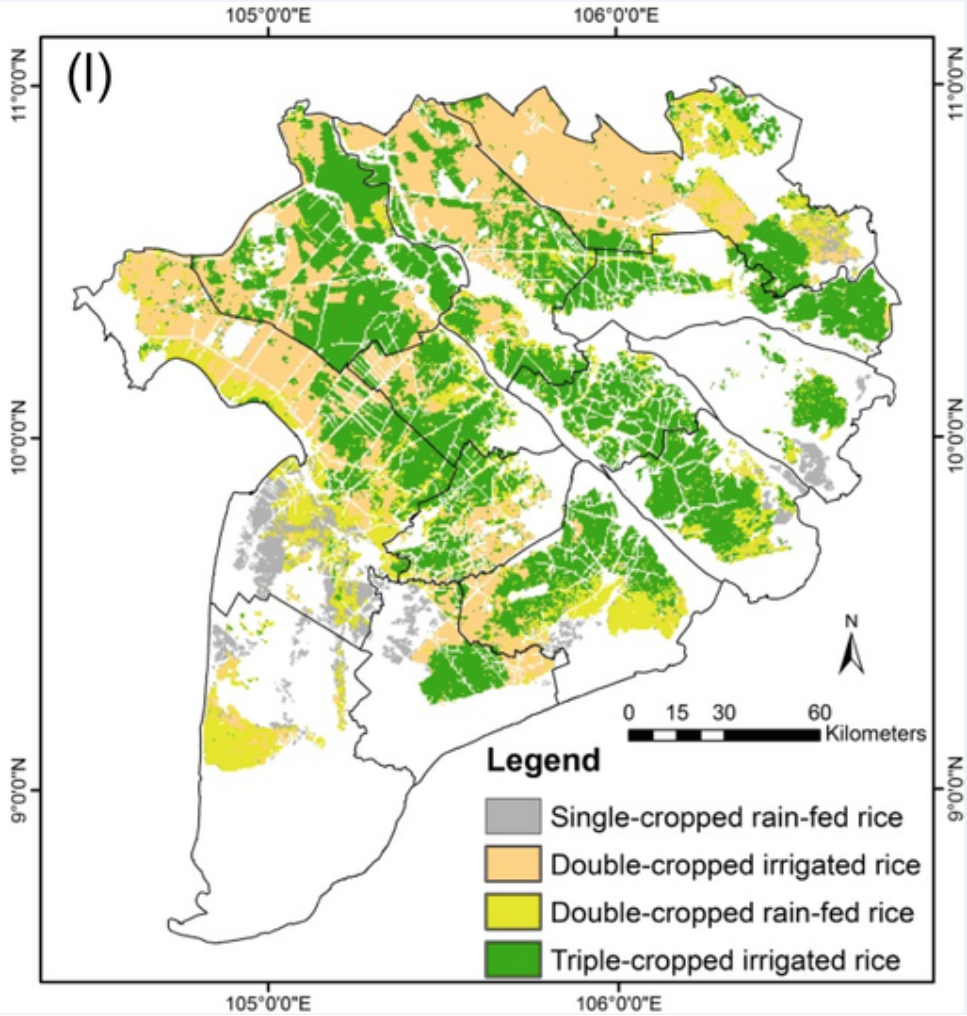

Figure 1. Rice covers map of Mekong river delta, Vietnam in 2012.

Regarding land cover classification (LCC), there are currently many researches around the world. These researches could be categorized by several criteria such as geographical scale of classification, multiple land covers classification or single land

cover classification. For the former, LCC can be classified into regional or global studies. Regional studies focus on investigating LCC methods for one or more specific regions. Global studies concern classification at global scale. There are currently some already published global land-cover datasets as presented in Table 1.

Table 1. Description of seven global land-cover datasets.

	GLCC	**UMD**	**GLC2000**	**MODIS LC**	**GlobCover**	**GLCNMO**	**FROM-GLC**
Sensor	AVHRR	AVHRR	SPOT-4 VEGETATION	MODIS	MERIS	MODIS	LANDSAT
Acquisition time	04/1992–03/1993	04/1992–03/1993	11/1999–12/2000	01/2001–12/2002	12/2004 – 06/2006	01/2008 – 12/2008	01/2010 – 12/2010
Spatial resolution	1 km	1 km	1 km	500 m	300 m	500 m	30 m
Input data	IGBP 1-km AVHRR 10-day composite, DEM data, Ecoregions data, Maps data.	41 metrics derived from NDVI and AVHRR bands 1–5, EROS urban, MODIS water mask	Daily mosaics of 4 spectral channels and NDVI of SPOT, JERS-1 and ERS radar data, DMSP data, DEM	Monthly MODIS L2/L3 composite, EOS land/water mask, MODIS 16-day EVI, MODIS 8-day DEM	MERIS L1B data, MERIS mosaics	16-day composite of MODIS 2008 Data MOD44W and SRTM DEM	Landsat TM/ETM+ (30 meter), MODIS EVI time series (250 meter) Bioclimatic variables (1km) global DEM (1km)
Classification method	Classification with post-classification refinement Unsupervised	Decision tree	Unsupervised classification	Decision tree, Neural networks	Unsupervised classification	Combined method of supervised classification and individual mapping	Maximum likelihood (MLC), J4.8 Decision tree, Random forests and Support

							vector machine
LC class	17 classes	14 classes	23 classes	17 classes	22 classes	20 classes	10 classes
Validation data	Landsat TM and SPOT images	Other digital datasets	High resolution satellite data, and ancillary information	High resolution land cover information	SPOT-VEGETATION NDVI, and Virtual/Google Earth	Integrated potential map, Google Earth image, MODIS images	MODIS vegetatio, DEM and soil-water condition maps
Reported accuracy	Globally 66.9%	Globally 69%	Globally 68.6 ± 5%	Globally 75%	Globally 67.1%	Globally 77.9%	Globally 64.9%

Although there are many efforts to map land covers globally, the LC accuracies are still much lower than regional LC maps. This is understandable as there are many challenges in LCC at global scale including diversity of land-cover types, lack of ground-truth data, and so on [4]. In regional studies, the difficulties are more or less reduced, thus resulting in more accurate LC maps. Some typical regional LC studies could be mentioned, i.e. Hannes et al. investigated Landsat time series (2009 - 2012) for separating cropland and pasture in a heterogeneous Brazilian savannah landscape using random forest classifier and achieved and overall accuracy of 93% [5]. Xiaoping Zhang et al. used Landsat data to monitor impervious surface dynamics at Zhoushan islands from 2006 to 2011 and achieved overall accuracies of 86-88% [6]. Arvor et al. classified five crops in the state of Mato Grosso, Brazil using MODIS EVI time series and their OAs ranged from 74 – 85.5% [7].

Although land-cover classification (LCC) mapping at medium to high spatial resolution is now easier due to availability of medium/high spatial resolution imagery such as Landsat 5/7/8 [8], in cloud-prone areas, deriving high resolution LCC maps from optical imagery is challenging because of infrequent satellite revisits and lack of cloud-free data. This is even more pronounced in land cover with high temporal dynamics, i.e. paddy rice or seasonal crops, which require observation of key growing stages to correctly identify [9], [10]. Vietnam is located in a tropical monsoon climate frequently covered by cloud [11], [12]. Some studies used high temporal resolution but low spatial resolution images (MODIS) [13]. Some studies employed single-image classifications [14]. However, common challenges of mono-temporal approaches include misclassification between bare land or impervious surface and vegetation cover type [15]. Whereas land cover classification using cloud-free Landsat scenes may lack enough observations to capture temporal dynamics of land-cover types.

1.2. Objectives, contributions and thesis structure

To date, land cover classification in cloud-prone areas is challenging. Furthermore, efficient LC methods for the regions, especially for areas with high temporal dynamics of land covers, are still limited. In this thesis, the aim is to propose a classification method for cloud-prone areas with high temporal dynamics of land-cover types. It is also the main contribution of the research to current development of land cover classification. To assess its classification performance, the proposed method is first tested in Hanoi, the capital city of Vietnam. Hanoi is one of the cloudiest areas on Earth and has diverse land covers. In particular, the results of this thesis could be applicable to other cloudy regions worldwide and to clearer ones also.

This thesis is organized into five chapters. In chapter 1, I give an introduction to remotely-sensed data and its application in various domains. A problem statement is also presented. Theoretical backgrounds in remote sensing, compositing methods and land cover classification methods are introduced in Chapter 2. Proposed method is presented in Chapter 3. Chapter 4 details experiments and results. Finally, some conclusions of my thesis are drawn in Chapter 5.

CHAPTER 2. THEORETICAL BACKGROUND

This chapter reviews necessary concepts used in this thesis. Basic knowledge of remote sensing science is presented in section 2.1. Section 2.2 introduces satellite images and details of Landsat 8 data. Compositing methods for satellite images are summarised in section 2.4. Finally, machine learning methods in land cover classification are discussed in section 2.5.

2.1. Remote sensing concepts

2.1.1. General introduction

Remote sensing is a science and art that acquires information about an object, an area or a phenomenon through the analysis of material obtained by specialized devices. These devices do not have a direct contact with the subject, area, or studied phenomena (Figure 2) [1].

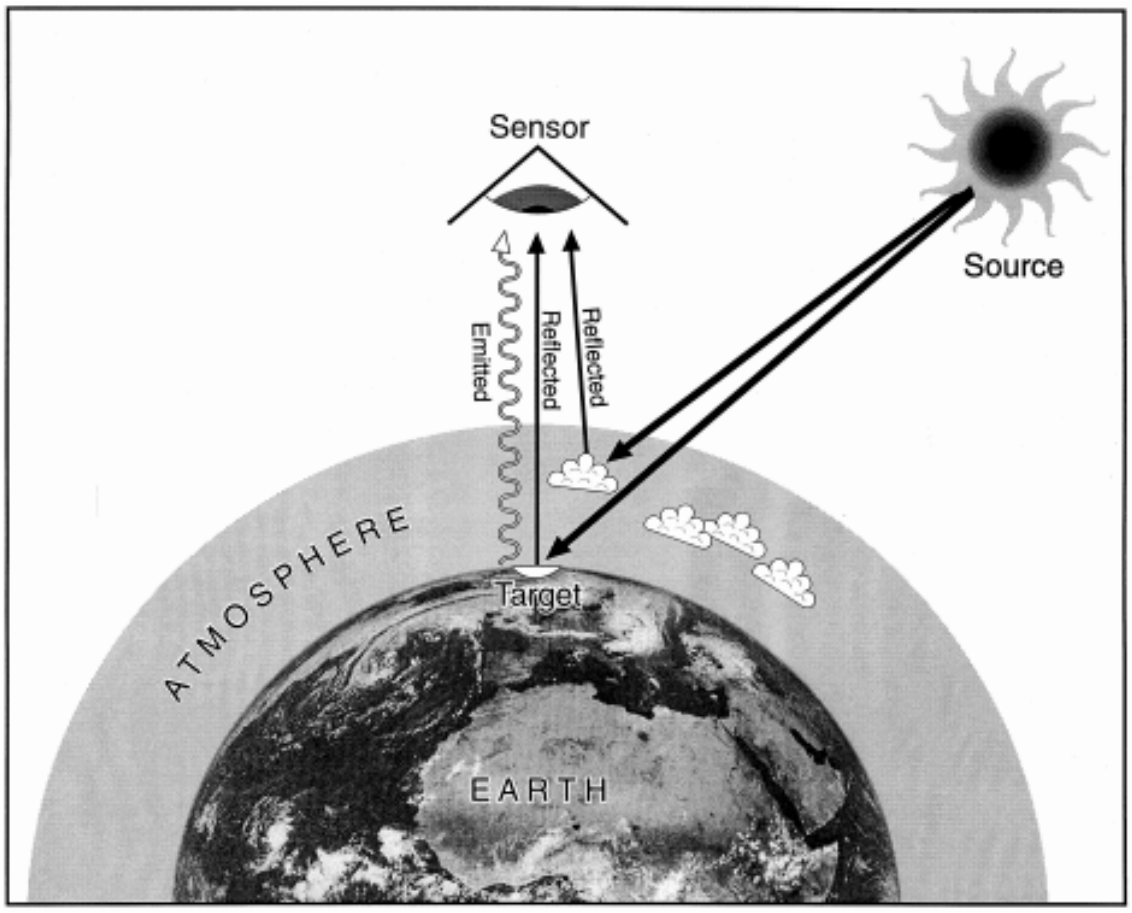

Figure 2. The acquisition of data in remote sensing[1].

Electromagnetic waves that are reflected or radiated from an object are the main source of information in remote sensing. A remote sensing image provides information about the objects in form of radiated energy in recorded wavelengths. Measurements and analyses of the spectral reflectance allow extraction of useful information of the ground. Equipments used to sense the electromagnetic waves are called sensor. Sensors are cameras or scanners mounted on carrying platforms. Platforms carrying sensors are called carrier, which can be airplanes, balloons, shuttles, or satellites. Figure 1 shows a typical scheme for remote sensing image acquisition. The main source of energy used in remote sensing is solar radiation. The electromagnetic waves are sensed by the sensor on the receiving carrier. Information about the reflected energy could be processed and applied in many fields such as agriculture, forestry, geology, meteorology, environments and so on.

A remote sensing system works in the following model: a beam of light, emitted by the sun/the satellite itself, firstly reaches the Earth surface. It is then partially absorbed, reflected and radiated back to the atmosphere. In the atmosphere, the beam may also be

[1] http://tutor.nmmu.ac.za/uniGISRegisteredArea/intake13/Remote%20Sensing%20and%20GIS/sect2pr.pdf

absorbed, reflected or radiated for another time. On the sky, the satellite's sensor will pick up the beam that is reflected back to it. After that it is the process of transmitting, receiving, processing and converting the radiated energy into image data. Finally, interpretation and analysis of the image is done to apply in real-life applications. Figure 3 illustrates typical components of a remote sensing system [1].

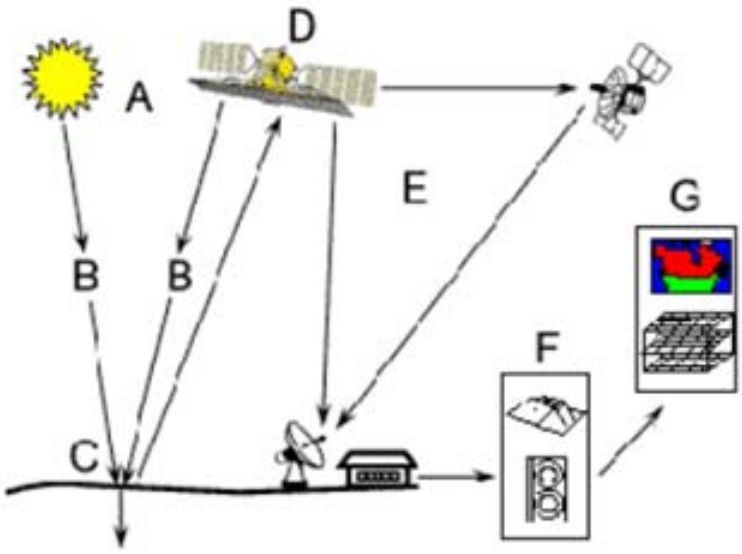

Figure 3. Introduction of a typical remote sensing system.

Symbols:

- A: energy source.
- B: incoming source.
- C: the ground target.
- D: satellite.
- E: receiving system.
- F: image analysis system.
- G: application system.

2.1.2. Classification of remote sensing systems

Remote sensing systems can be classified by following criterias: energy source, satellite's orbit, spectrum of the receiver, etc [1].

Classification based on energy source: passive and active remote sensing systems (Figure 4).

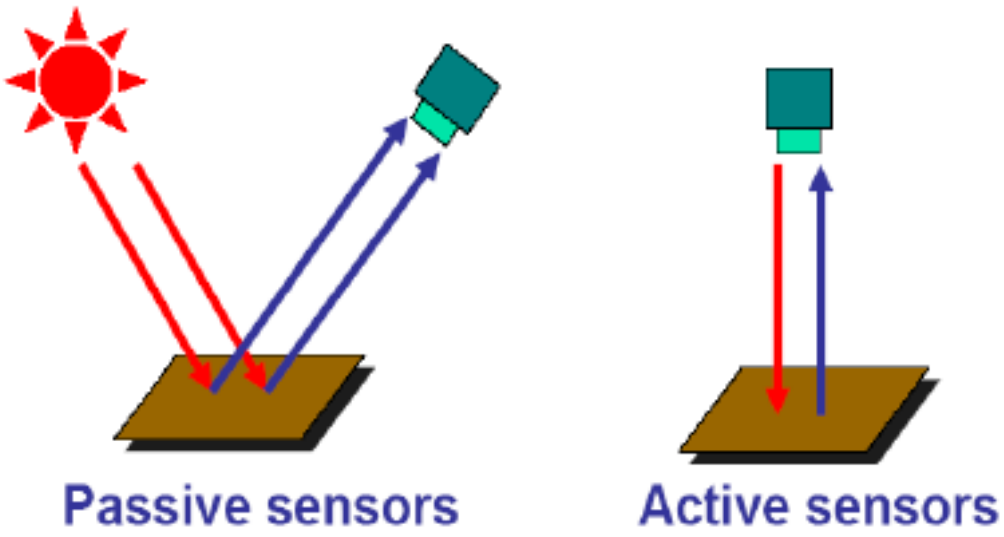

Figure 4. Passive (left) and active (right) remote sensing systems.

- **Active remote sensing system**: the source energy is the light emitted by an artificial device, usually the transmitter placed on the flying equipment.
- **Passive remote sensing system**: the source energy is the Sun's light.

Classification based on orbit (Figure 5**)**:

- **Geostationary satellite**: is a satellite with a rotational speed equal to the rotational speed of the earth. Relative position of the satellite as compared to the earth is stationary.
- **Polar orbital satellite:** is a satellite with orbital plane which is perpendicular or near perpendicular to the equatorial plane of the earth. The satellite's rotation speed is different from the rotation speed of the earth. It is designed so that the recording time on a particular region is the same as the local time. And the revisit time for a particular satellite is also fixed. For example, Landsat 8 has a revisit time of 16 days[2].

[2] https://landsat.usgs.gov/landsat-8

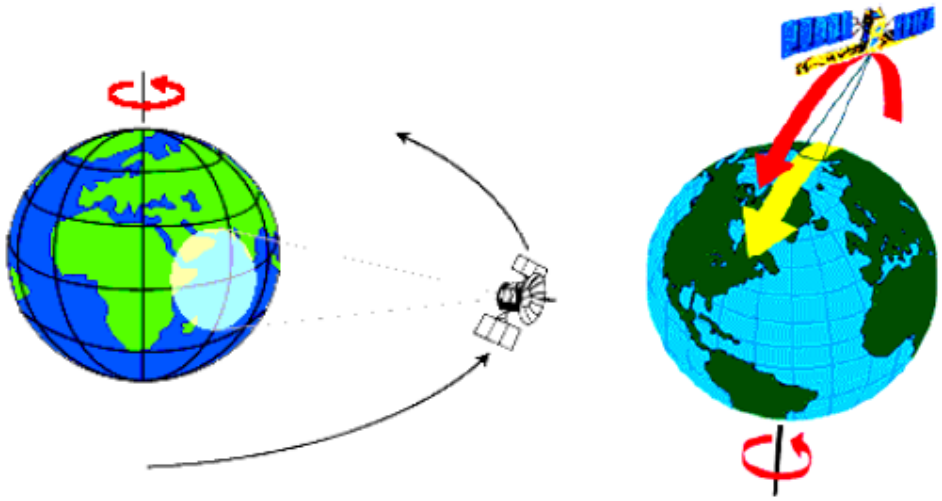

Figure 5. Geostationary satellite (left) and Polar orbital satellite (right).

Classification by receiving spectrum: visible spectrum, thermal infrared, microwave,….

The sun is the main source of energy for remote sensing in visible and infrared bands. Earth surface objects can also emit their energy in thermal infrared spectrum. Microwave remote sensing uses ultra-high frequency radiation with a wavelength of one to several centimeters. The energy used for active remote sensing is actively generated from the transmitter. Radar technology is a type of active remote sensing. Active radar emits energy to objects, then captures the radiation which is scattered or reflected from the object.

2.1.3. Typical spectrum used in remote sensing systems

In fact, there are many different types of light. However, only a few spectral bands are used in remote sensing (Figure 6). The following are frequently used.

- **Visible light**: are lights whose wavelengths are between 0.4 and 0.76 microns. The energy provided by these wave bands plays an important role in remote sensing.
- **Near Infrared**: are lights whose wavelengths are between 0.77 and 1.34 microns.
- **Middle Infrared**: are lights whose wavelengths are between 1.55 and 2.4 microns.

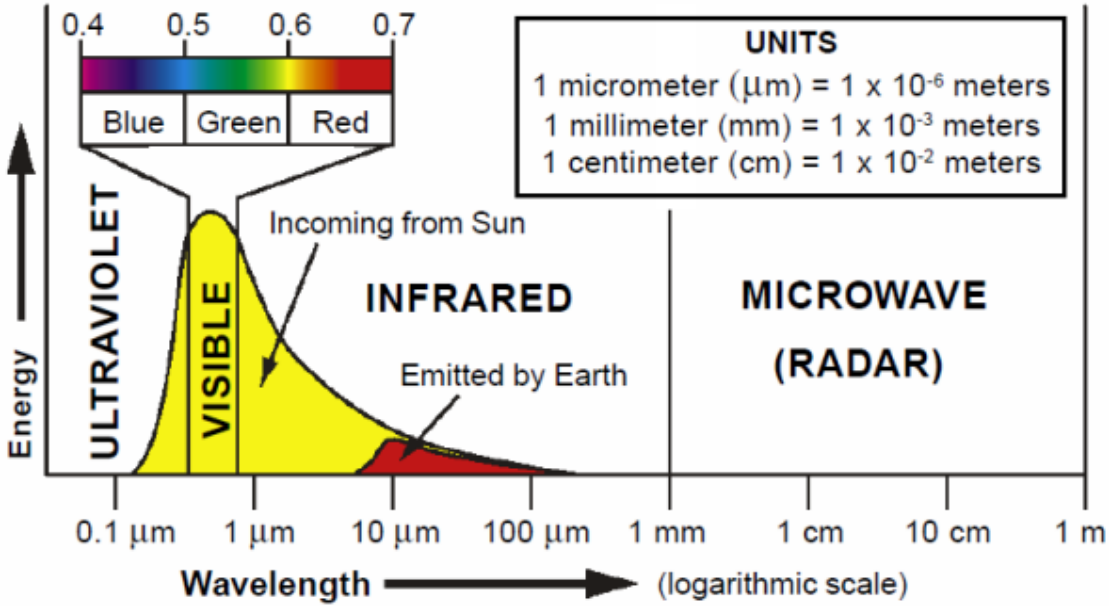

Figure 6. Typical wavelengths used in remote sensing[3].

- **Thermal Infrared**: are lights whose wavelengths are between 3 and 22 microns.
- **Microwave**: are lights whose wavelengths are between 1 and 30 microns. Atmosphere does not strongly absorb wavelengths greater than 2 centimeters which allows day and night energy intake, without the effects of clouds, fog or rain.

2.2. Satellite images

2.2.1. Introduction

Satellite images are images of Earth or other planets collected by observation satellites. The satellites are often operated by governmental agencies or businesses around the world. There are currently many Earth observation satellites and they have common characteristics including spatial resolution, spectral resolution, radiometric resolution and temporal resolution. A detailed description of each resolution is shown below [1].

- **Spatial resolution**: refers to the instantaneous field of view (IFOV) which is the area on the ground viewed by the satellite's sensor. For example, the Landsat 8 satellite has 30-meter spatial resolution which means that a Landsat 8's pixel covers an area on the Earth's surface of 30m x 30m.
- **Spectral resoalution**: spectral resolution describes the ability of the sensor to

[3] http://www.remote-sensing.net/concepts.html

receive the Sun's light. If conventional cameras on the phone can only obtain wavelengths in the visible range including red, green and blue lights, many satellite sensors have possibility to sense many other wavelengths such as near infrared, short-wave infrared, and so on. For example, the TIRS sensor mounted on Landsat 8 satellite can receive wavelengths ranging from 10.6 to 12.51 micrometers.

- **Radiometric resolution**: the radiometric resolution of a sensor describes the ability to distinguish very small differences in light energy. A better radiometric resolution can detect small differences in reflection or energy output.
- **Temporal resolution**: temporal resolution of a satellite is the time interval between two successive observations over the same area on the Earth's surface. For example, the temporal resolution of Landsat 8 satellite is 16 days.

There are currently many Earth observation satellites having different spatial resolutions, temporal resolutions, radiometric resolutions and spectral resolutions. Table 2 compares these resolutions of some well-known satellites.

Table 2. Some featured satellite images

	Satellite image	**Type**	**Typical spatial resolution**	**Spectral resolution (exclude panchromatic)**	**Radiometric resolution**	**Temporal resolution**
1	MODIS	Optical	250 – 1000m	36 bands	12 bits	Daily
2	SPOT 5	Optical	10m	4 bands (Green, Red, Near IR, SWIR)	8 bits	2-3 days, depending on latitude
3	Landsat 8	Optical	30m	10 bands (Coastal -> TIRS2)	12 bits	16 days
4	Sentinel 2A	Optical	10 – 20m	12 bands (Coastal -> SWIR)	12 bits	10 days

2.2.2. Landsat 8 images

The 8th Landsat satellite - Landsat 8 (Figure 7) was successfully launched into orbit on February 12, 2013. This is a joint project between NASA and the US Geological Survey. Landsat 8 satellite provides medium resolution images (from 15 to 100 meters), with polar coverage.

Figure 7. Landsat 8 images[4]

Landsat 8 satellite has two sensors: Operational Land Imager (OLI) and Thermal InfraRed Sensor (TIRS). These two sensors provide images at a spatial resolution of 30 meters for visible/near infrared/infrared bands, 100 meters for thermal bands and 15 meters for panchromatic band. For the thermal bands, the manufacturer increased their spatial resolution up to 30m through a resampling procedure. The ground coverage of a Landsat 8 image is limited to 185km x 180km. Satellite altitude reaches 705 km.

A comparison of Landsat 7 and Landsat 8 bands is provided in Figure 8:

[4] NASA's Goddard Space Flight Center

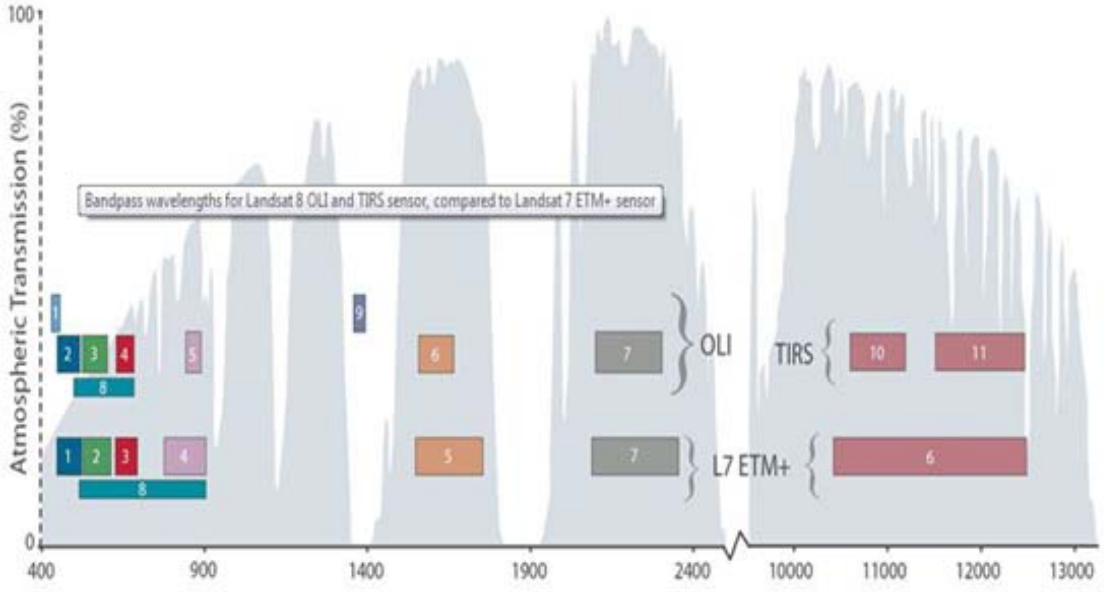

Figure 8. Landsat 7 and Landsat 8 bands[5]

Landsat 8 is programmed to fly around the Earth for 99 minutes, covers the entire surface of the Earth for 16 days. With about 400 images acquired per day, Landsat 8 satellite provides a more accurate view of Earth's variations within 10 years of its life.

Landsat 8 images are provided to users via the Internet. Each image product is a compressed file containing 12 TIFF image files and a metadata file. Landsat 8 images are stored in raster format, which means that they are made up of pixels. Each image is a grid of pixels. Among the 12 TIFF files, 11 files are numbered from 1 to 11 indicating the band number. Each of the files stores energy values that the sensors receive in 16-bit integer format which is also known as digital numbers (DN) (Table 3). The remaining file is a BQA file added by the manufacturer.

Table 3. Landsat 8 bands[6].

Band	**Name**	**Central wavelength (μm)**	**Spectral range (μm)**
1	Coastal Aerosol (OLI)	0.443	0.433-0.453
2	Blue (OLI)	0.482	0.450-0.515
3	Green (OLI)	0.562	0.525-0.600
4	Red (OLI)	0.655	0.630-0.680

[5] Website http://www.imagico.de/map/landsat8.php
[6] Website http://landsat.gsfc.nasa.gov/?page_id=5377

5	NIR (OLI)	0.865	0.845-0.885
6	SWIR 1 (OLI)	1.610	1.560-1.660
7	SWIR 2 (OLI)	2.200	2.100-2.300
8	Panchromatic (OLI)	0.590	0.500-0.680
9	Cirrus (OLI)	1.375	1.360-1.390
10	Thermal 1	10.8	10.3-11.3
11	Thermal 2	12.0	11.5-12.5

In this study, I used Landsat 8 Surface Reflectance images (Figure 8). Landsat 8 Surface Reflectance data are generated from the Landsat Surface Reflectance Code (LaSRC), which makes use of the coastal aerosol band to perform aerosol inversion tests [16]. LaSRC has a unique radiative transfer model and it also uses auxiliary climate data from MODIS sensor. Figure 9 shows a Landsat 8 image before and after atmospheric correction. In the uncorrected image (left), it could be clearly seen impacts of atmosphere in blurred areas (exclude cloudy areas). This impact is significantly reduced in the corrected image (right).

Currently, Landsat 8 SR data product contains seven bands including Coastal Aerosol, Blue, Green, Red, NIR, SWIR1, SWIR2. Besides, there are also cloud mask bands, and some ancillary data.

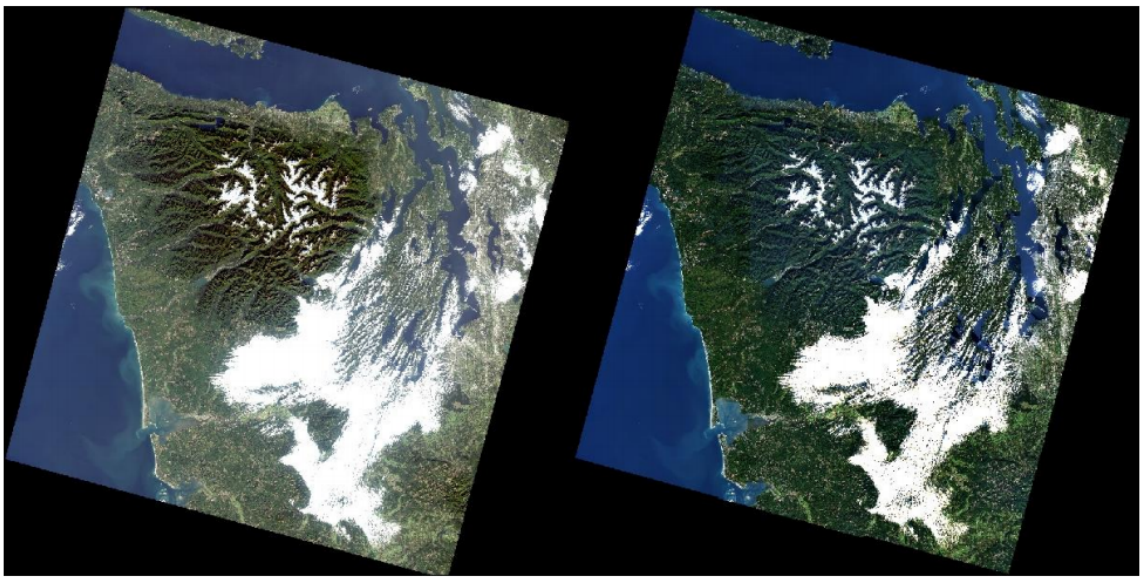

Figure 9. Comparison of Landsat 8 OLI (left) and SR (right) images.

2.3. Compositing methods

Optical satellite images have a big drawback. In particular, they are heavily impacted by clouds. If a region is covered by clouds during its satellite passing time, the recorded data is considered lost. Therefore, methods for tackling clouds in optical satellite images have been studied by many researchers. Pixel-based image compositing is a paradigm in remote sensing science that focuses on creating cloud-free, radiometrically and phenologically consistent image composites. The image composites are spatially contiguous over large areas [17]. In the past, some compositing methods for low spatial resolution images (i.e. 500x500m or greater) were developed [18], [19]. Those methods were used primarily to reduce the impacts of clouds, aerosol contamination, data volume and view angle effects which are inherent in the images. Due to high temporal resolution of the satellites, the compositing methods were relatively simple, i.e. use maximum Normalized Difference Vegetation Index (NDVI) or minimum view angle to pick an appropriate observation for a target pixel. Since the opening of the Landsat archive, compositing methods for Landsat images have been developed and benefitted by pre-existing approaches for MODIS and AVHRR data.

Recently, a number of best-available-pixel compositing (BAP) methods have been proposed for medium/high satellite images. Generally, BAP methods replace cloudy pixels with best-quality pixels from a set of candidates through rule-based procedures. Selection rules are based on spectral-related information, that is, maximum normalized difference vegetation index (NDVI) [20] and median near-infrared (NIR) [21]. On another approach, Griffiths et al. proposed a BAP method ranking candidate pixels by score set such as distance to cloud/cloud shadow, year, and day-of-year (DOY) [22]. This method was improved by incorporating new scores for atmospheric opacity and sensor types [17]. Gómez et al. recently offered a review emphasizing BAP potential for monitoring in cloud-persistent areas [23], which includes applications in forest biomass, recovery and species mapping [24], [25], [26], change detection applications [27], and general land-cover applications [28].

A summary of several compositing methods is presented in Table 4.

Table 4. Review of compositing methods for satellite images.

	Study	Satellite images	Method
1	Hansen et al. 2008 [29]	Landsat 5, 7	$Pixel_{selected} = \min\{P_{cloud\&shadow}^{1st_{candidate}}, \dots, P_{cloud\&shadow}^{nth_{candidate}}\}$ Where: $Pixel_{selected}$: the candidate pixel selected for composition

			$P_{cloud\&shadow}^{nth_{candidate}}$: probability of cloud/cloud shadow of the same pixel in nth candidate image If two or more candidate pixels have equal $P_{cloud\&shadow}$, then choose the pixel value closest to a forest reference value (100)
2	Roy et al. 2010 [20]	Landsat 5, 7	$Pixel_{selected}$ $= \begin{cases} \max\{NDVI^{1st_{candidate}}, \dots, NDVI^{nth_{candidate}}\}, & if\ vegetated\ pixels \\ \max\{BTEM^{1st_{candidate}}, \dots, BTEM^{nth_{candidate}}\}, & if\ unvegetated\ pixels \end{cases}$ Where: NDVI: Normalized Difference Vegetation Index BTEM: Brightness Temperature Eligible candidate pixels must be of minimal cloud, snow, and atmospheric contamination
3	Potapov et al. 2011 [21]	Landsat 7	$Pixel_{selected} = \min \text{abs}\{NIR^{1st_{candidate}} - NIR^{median}, \dots, NIR^{nth_{candidate}} - NIR^{median}\}$ Where: NIR: near infrared spectral band Only satellite images acquired in growing seasons are eligible for ranking procedure
4	White et al. 2014 [17]	Landsat 5, 7	$Pixel_{selected} = \max\{sum\ of\ scores\ (opacity, cloud\&shadow, doy, sensor)^{1st_{candidate}}, \dots\}$

2.4. Machine learning methods in land cover study

Basically, LC classification is a type of classification on image data. Therefore, machine learning classifiers are also applicable to LC classification. In fact, there existed a huge amount of researches on machine learning classifiers in LCC. These methods range from simple thresholding to more advanced approaches such as maximum likelihood, logistic regression, decision tree (ID3, C4.5, C5), random forest, support vector machine (SVM), artificial neuron network (ANN) and so on [30], [31], [32], [33], [34]. Some well-known classifiers are presented below.

2.4.1. Logistic Regression

Logistic regression is a generalized linear model which is often used for classification. Suppose the training data represented by $\{x_i, y_i\}$, $i = 1, \dots, k$, where $x \in$

R^n is a n-dimensional space vector and $y \in \{1, -1\}$ is a class label. A logistic regression model could be written as:

$$f(x) = y' = \theta(w^T x) \tag{1}$$

Where w is the weights vector, θ is sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

To train a logistic regression model, a cost function is defined as:

$$J(w, x_i, y_i) = -\sum_{k=1}^{N} (y_i log(y'_i) + (1 - y_i)\log(1 - y'_i)) \tag{3}$$

To optimize weights, gradient descent algorithm is used which incrementally adjust weights based on gradient direction of cost function at each training step. Finally, the weight vector is updated as follow:

$$w = w + \eta(y_i - y'_i)\, x_i \tag{4}$$

Where η is learning rate.

To extend logistic regression from binary classification to multiclass classification, one can employ one-vs-all strategy. In this case, each class is trained against other classes. A new sample x is assigned to class i if probability of $y_x = i$ is the largest of all classes.

2.4.2. Support Vector Machine

Support Vector Machines (SVM) is a group of supervised learning methods as introduced in [35]. SVMs seeks to find the decision boundary that gives the best generalization – also known as the optimal separating hyperplane in multi-dimensional space.

Suppose the training data represented by $\{x_i, y_i\}$, $i = 1, \ldots, k$, where $x \in R^n$ is a n-dimensional space vector and $y \in \{1, -1\}$ is a class label. This set of training data can be separated by a hyperplane if there exists a vector $w = (w_1, \ldots, w_k)$ and a scalar b satisfying the following inequality:

$$y_i(wx_i + b) - 1 + \xi_i \geq 0 \qquad \forall y = \{+1, -1\} \tag{5}$$

Where ξi is a slack variable which indicates the distance the data sample is from the optimal hyperplane. The objective function can be written as following:

$$\frac{1}{2}||w||^2 + C\sum_{i=1}^{k}\xi_i \quad (6)$$

C is a constant used to control the degree of the penalty associated with training samples that occur on the wrong side of the optimal separating hyperplane. C should be considered closely for each individual classification task. The optimal hyperplane can be identified by minimizing the objective function in Eq. (2) under the constraint in Eq. (1). This can be done by utilizing Lagrange multipliers and quadratic programming methods.

The basic approach to SVM classification may be extended to allow for nonlinear decision boundaries by mapping the input data into higher-dimensional space H so that in the new space, data can be linearly separated. To do this, a kernel function is introduced: $K(x_i, x_j) = (\phi(x_i), \phi(x_j))$, where an input data sample x can be represented as $\phi(x)$ in the space H. This kernel function allows computing the inner product $(\phi(x_i), \phi(x_j))$ without knowing exactly the representation of the data samples xi and xj in the higher space. There exist several kernel function types including polynominal-based and radial basis function (RBF) kernels, etc.

Because SVM was developed as a binary classifier, it is necessary to adapt this method to multiclass classification problems. There are two common approaches for solving the problem. The first is known as one-against-one method and the second is one-against-all method.

2.4.3. Artificial Neural Network

In machine learning, Artificial Neural Networks (ANNs) are a group of statistical learning models that are inspired by biological neural networks in the human brain [31]. In general, an ANN often consists of an interconnected group of neural nodes that correspond to the neurons of the human brain. Various types of neural network have been developed over the previous decades. The most widely used model is multilayer perceptron (MLP), a feed-forward neural network, due to its simplicity to understand and interpret. Backpropagation learning algorithm introduced by Rumelhart et al. is the most popular algorithm used to train a MLP [36]. Figure 10 presents a three-layer perceptron with three inputs, two outputs, and one hidden layer assembled by five neurons:

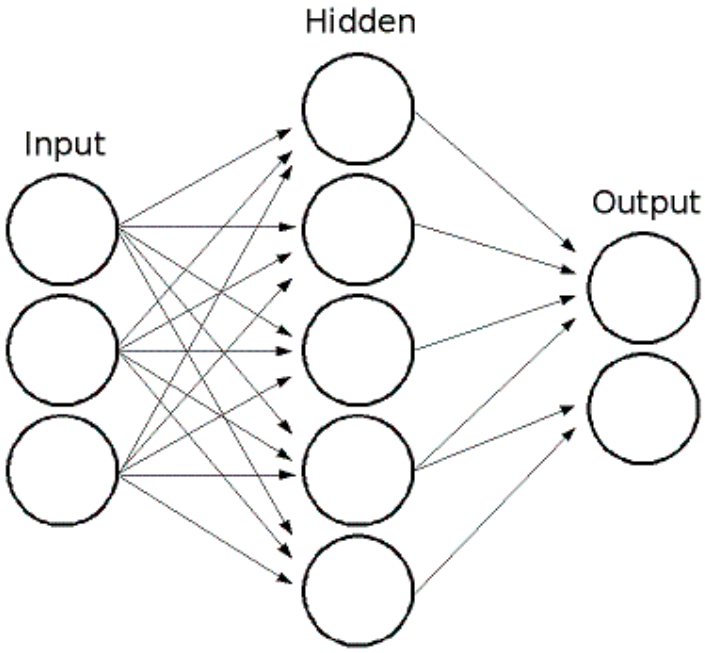

Figure 10. An example of MLP.

Each neuron has several input links. The inputs are output values from neurons in the previous layer. At a particular neuron, the inputs are summed up with certain weights, plus a bias term. The sum is then transformed using an activation function *f* which may be different for different neurons. In other words, given the inputs x_j of the layer, the output y_j of the layer $n + 1$ are computed as:

$$z_i = \sum_j \left(w_{i,j}^{n+1} \times x_j\right) + w_{i,bias}^{n+1}$$
$$y_i = f(z) \tag{7}$$

There are different activation functions. Some standard functions are:

Identify function:

$$f(x) = x \tag{8}$$

Rectified linear unit (ReLU) function:

$$f(x) = \begin{cases} 0, for\ x < 0 \\ x, for\ x \geq 0 \end{cases} \tag{9}$$

Sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{10}$$

Tanh function:

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (11)$$

Many studies have concentrated on using ANNs for land-cover discrimination and their experiments have suggested these types of models are more accurate than traditional statistical methods, i.e. MLC, because ANNs make no prior assumption about the normal distribution of the data and can construct complex decision boundaries [33].

2.4.4. eXtreme Gradient Boosting

eXtreme Gradient Boosting (XGBoost) is a new classifier as introduced by Tianqi Chen [37]. XGBoost is now emerging as a powerful tool in many classification domains. This classifier has been proved to be very powerful in Kaggle competitions. It is an optimized version of Gradient Boosting Machines (GBM). In gradient boosting, trees are build sequentially and each new model uses gradient descent algorithm. The final model is a result of an addition of multiple functions from the beginning. In XGBoost, trees can be built in parallel. It also has an inner paradigm for handing sparse input data. Thus it may work in some cases where image data is missing due to clouds. The principle difference between boosting methods and conventinal methods is that optimization is held on function space in the former methods [38].

A XGBoost model could be written as in Eq. (12):

$$y_i' = \Phi(x_i) = \sum_{k=1}^{K} f_k(x_j), \; f_k \in F \quad (12)$$

Where F is the function space of base learners, x_i is a n-dimensional array input vector, y_i' is the function's prediction. To build the set of functions used in the model, the following objective function is used:

$$L(\Phi) = \sum_i l(y'_i, y_i) + \sum_k \Omega(f_k) \quad (13)$$

$\Omega(f) = \gamma T + \frac{1}{2}\lambda||w||^2$ is the regularization term where T is the number of leaves, w is leaf weights

However, the objective function considers functions as its parameter thus cannot be trained using traditional methods in Euclide space. Therefore, it is trained in an additive manner which fixes what was learned and add one new tree at a time. Furthermore, different loss functions could be used in XGBoost, i.e. Gaussian L_2, Laplace L_1, Binomial, Adaboost, Huber, … or any custom function defined by users. Base learners could also be specified on purpose.

2.4.5. Ensemble methods

Ensembling is an efficient way to improve classification. Random Forest

implements a typical ensemble method. In Random Forest, individual tree prediction outputs are used for majority voting to derive final prediction. Thus, trees in Random Forest are independent to each other. Stacking is another ensemble method. Stacking models means there exists several layers of models. The output of the previous layer is used to train the next layer. Besides, there are also other ensemble methods, i.e. boosting (XGBoost). In this research, I investigate stacking and majority voting methods for ensembling of different classifiers. And only the best result is reported.

2.4.6. Other promising methods

Deep learning is another powerful tool for classification. It emerged from traditional ANN through a series of innovations in training of ANN (activation function, big data, increased computing power, and so on). These innovations help to create very large neural network architectures. In remote sensing, deep learning is already applied in some classifications such as object detection in very high resolution optical images/UAV images, roof detection, ship detection [39], [40] where labelling process could be done in an easier manner as compared to coarser satellite images. Deep learning architectures (i.e. Convolutional Neural Network) often require a large amount of labelled training data in order to achieve good classification performance as compared to traditional classifiers. This requirement is not an easy task in many applications. Therefore, deep learning for LCC using medium/high satellite images is still limited.

CHAPTER 3. PROPOSED LAND COVER CLASSIFICATION METHOD

This chapter presents the proposed land cover classification method. Firstly, I introduce the study area in section 3.1. Next, the data used and preprocessing methods are presented in section 3.2. Section 3.3 details the proposed method and every of its components. Finally, section 3.4 summarizes the evaluation metrics used to assess classification performance.

3.1. Study area

In this study, I chose Hanoi city as study area (Figure 11). Hanoi is the capital of Vietnam, the country's second largest city covering approximately 3,300 km^2, located in the centre of Red River Delta (RRD). Hanoi has three basic kinds of terrain including a fertile delta, midland region and mountainous zone. The city is mainly divided into agricultural area (56.6%) and non-agricultural area (40.6%) in 2010 [41]. In agricultural areas, paddy rice is dominant (60.9%) followed by other crops such as maize as well as various vegetable crops. Paddy rice is planted two times per year, while crops are grown in other dedicated areas. Occasionally, short-season vegetable crops or aquaculture are grown before the start of the first rice season. Non-agricultural areas are mostly covered by impervious surfaces and mosaicked natural landscape. Accordingly, I investigate seven LC classes for Hanoi including paddy rice, cropland, grass/shrub, trees, bare land, impervious area and water body.

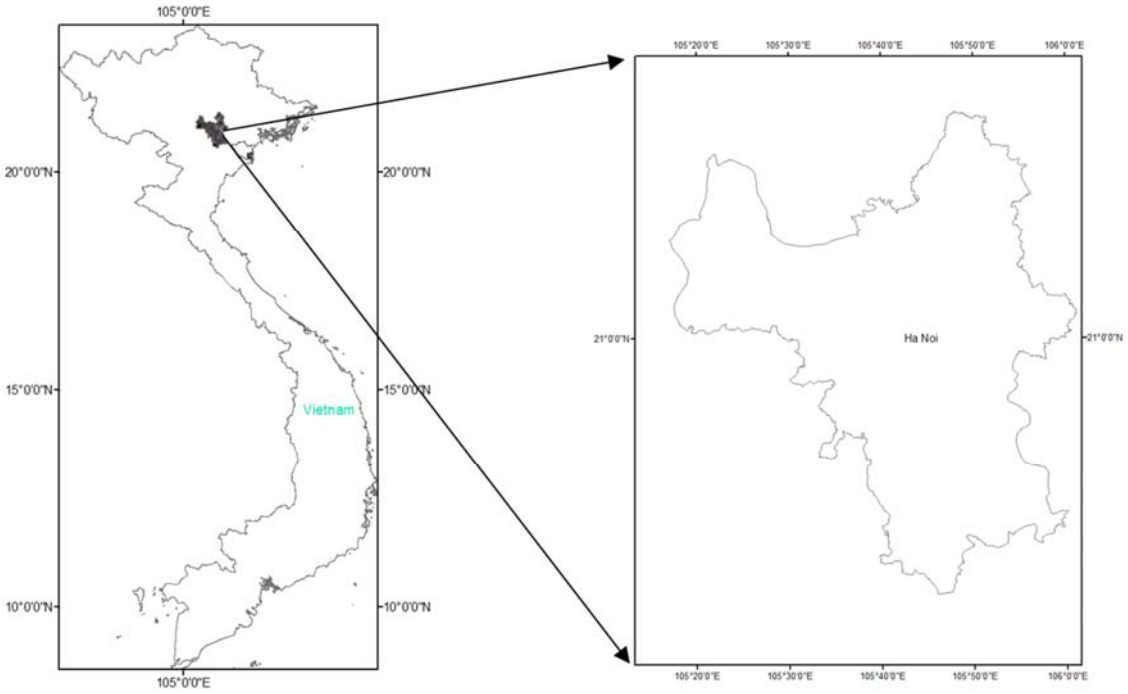

Figure 11. Hanoi city, study area of this study.

3.2. Data collection

3.2.1. Reference data

Official land-use data from Hanoi Environment and Natural Resources Department is used for training and testing data selection [42]. The selection procedure is based on stratified random sampling method. This is done separately for training and testing data. And these datasets are guaranteed to share no same point on the ground. Since different land uses may contain the same land-cover types, I therefore generated 11 strata labelled as bare area, long-term crops, short-term crops, forest, grass, impervious area, mudflats, rice, water, others and overlap areas of the land use strata. Training and testing data are randomly sampled from the strata and then labelled into 7 classes using high resolution images of Google Earth and field data (Figure 12). Total numbers of training and testing data are 5079 and 2748 points with the training and testing point distribution as presented in Table 5.

Table 5. Training and testing data.

	Training	**Testing**
Crop	596	331
Bare land	75	56
Paddy rice	1195	646

Water body	773	446
Tree	997	491
Impervious area	1049	523
Grass/Shrub	394	255
Total	**5079**	**2748**

It is observed that the classes have different numbers of training and testing data. For example, in the training data, impervious area and paddy rice have 1049 and 1195 points respectively. In contrast, bare land has only 75 points. This is due to areal difference of these classes in Hanoi. While impervious area and paddy rice are easily seen everywhere, bare land is rather rare. However, one important thing for the testing data is that the inclusion probability of every individual point is known. When randomizing data points, each point has an inclusion probability as presented in equation (1):

$$P_{T,i}^{inclusion} = \frac{\text{number of randomized points of strata T}}{\text{total points of strata T}} \quad (14)$$

In which, $P_{T,i}^{inclusion}$ is inclusion probability of point i of strata T. In stratified random sampling, inclusion probability of a point is determined by the strata to which it belongs only, regardless of size of the study area. Inclusion probability is then used to weigh individual points in subsequent calculations including producer accuracy, user accuracy, overall accuracy and F1-score. This is necessary to ensure that these statistics are representative for the study area. Point's weight is calculated as inverse of inclusion probability:

$$W_i = \frac{1}{Pi} \quad (15)$$

Figure 11 shows an example of randomized points in Google Earth. Each point is represented by a square covering an area of 30x30 meters, which is equal to Landsat 8 spatial resolution, on the ground. By using Google Earth's very high resolution images, the labelling process is cost-saving, easier and also reliable.

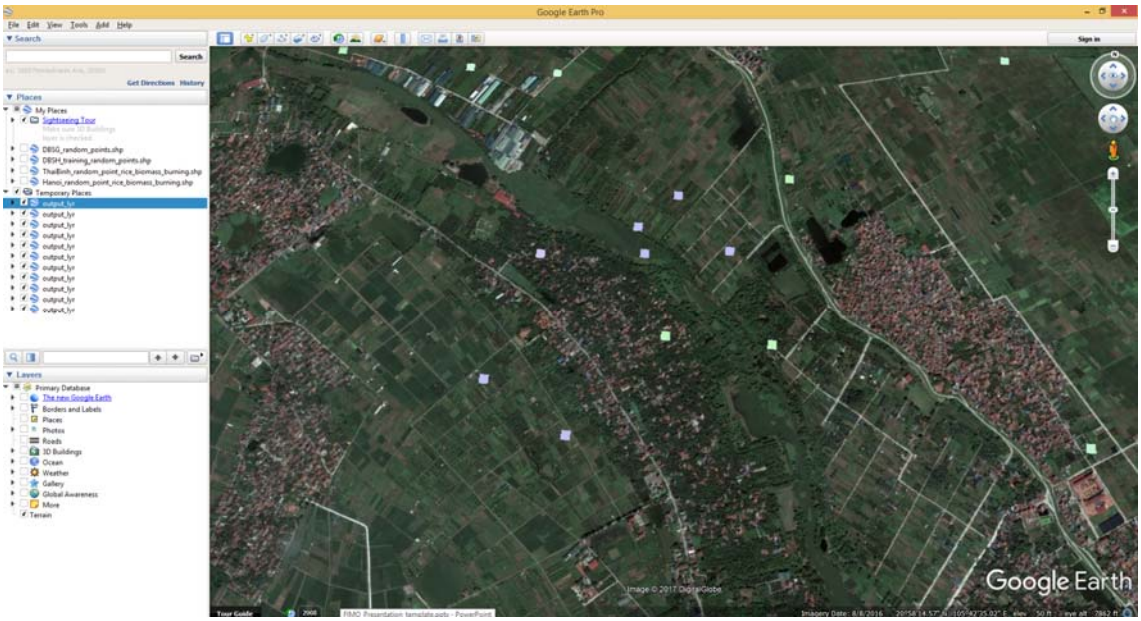

Figure 12. Examples of experimental data shown in Google Earth, sampled points are represented by while-colored squares over the Google Earth base images.

3.2.2. *Landsat 8 SR data*

To prepare imagery for the 2016 Hanoi land-cover map, all Landsat 8 Surface Reflectance (L8SR) images acquired over Hanoi city from 2013 to 2016 are collected from the USGS Earth Explorer (https://earthexplorer.usgs.gov/). There are 54 available L8SR scenes which are not 100% cloud-contaminated. As Hanoi is covered by two consecutive L8SR scenes per revisit (Figure 13), the consecutive images are mosaicked, resulting in 27 final images covering entire Hanoi city area (Figure 14).

Figure 13. Landsat 8 footprints over Hanoi.

After that, I performed some statistics on the images. Figure 12 shows distribution

of the images by month and year (Figure 12a), and some cloud cover's statistics (Figure 12b). Due to infrequent satellite revisits and frequent cloud covers, there are at least 0 and at most 2 images per month. Most of the months have 0-1 image only.

The problem of cloud covers is even more pronounced when it comes to analyse at per-image level. From Figure 14b, it is clearly observed that 15 out of 27 images are at least 50% covered by cloud. There are only five almost clear images over the 4-year period in Hanoi city.

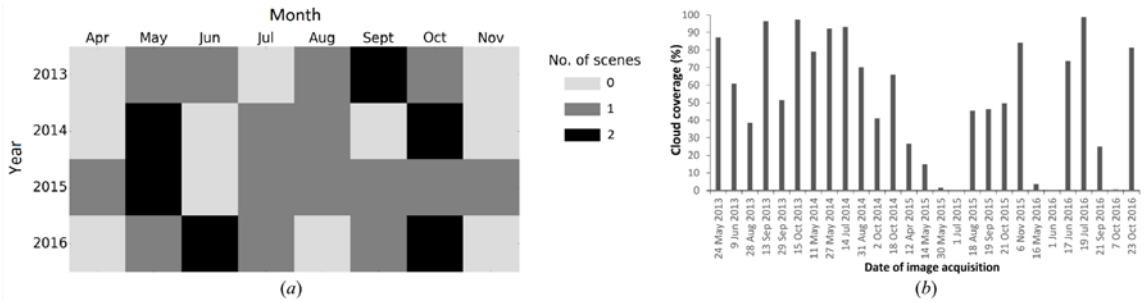

Figure 14. Statistics of Landsat 8 SR images over Hanoi, (a) number of images by year and month, (b) cloud coverage percentage per image

3.2.3. *Ancillary data*

Another ancillary data in this study is rice area statistics in 2016 produced by Hanoi Statistics Office (http://thongkehanoi.gov.vn/). This statistics include rice planting area at provincial level. The official rice area is used to compare with satellite-derived rice areas.

3.3. Proposed method

Overall flowchart of the method is displayed in Figure 15. The proposed method includes four main parts. Firstly, all Landsat 8 SR images are fed to compositing process to create a dense time series of cloud-free Landsat 8 images, i.e up to five images which is distributed across classification year (2016). After that, the composited images are used to extract spectral-temporal features. There will be three independent classifications. The first is classification using single image only (single-image classification), the second classification uses the whole time-series images with a single classifier (XGBoost), last classification is an improved version of the second classification with an addition of more features and ensemble of more strong classifiers. Finally, those classification models are validated against the testing data and statistical data as presented in previous sections.

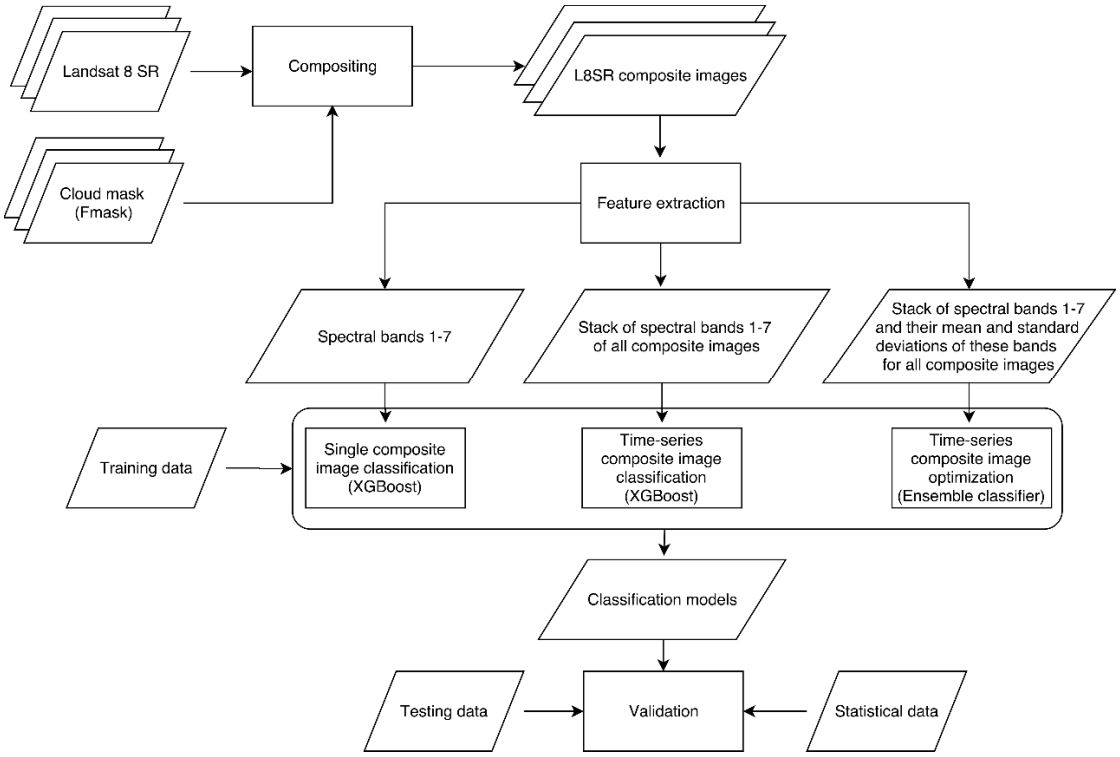

Figure 15. Overall flowchart of the method

3.3.1. Generation of composite images

The purpose of this step is to generate a dense, cloud-free time series to capture major spectral variations for 2016 land cover classification. The target images for compositing were the 5 clearest L8SR images from: 16th May 2016 (DOY 137), 1st June 2016 (DOY 153), 17th June 2016 (DOY 169), 21st September 2016 (DOY 265), and 7th October 2016 (DOY 281). These images were the targets for the compositing process which replaces their own cloud/cloud shadow pixels with best quality pixels from the above potential candidate images based on a scoring method as described below.

For each target image, clear pixels remain while cloudy pixels are replaced by a clear observation selected from the candidates. I combine two BAP methods proposed in [17], [22] and modify the year score and opacity score for compatibility with the problem. For each clear pixel in a candidate image, a score is computed based on 4 sub-scores: year score, DOY score, opacity score and distance from cloud/cloud shadow pixel. Those scores could be categorized into: image-level scores and pixel-level scores. Year and DOY are image-level scores. This means that every pixel in a particulate candidate image shares the same grade. Opacity and distance to cloud/cloud shadow are pixel-level scores. Each pixel has its own grade based on the score's ranking strategy. Year score is extended from [22] for 3 support years instead of 2 in original method. Year score decrease with distance from target year (2016) to support years (2015, 2014,

2013). DOY score reflects ranges of target day and support days following Gaussian distribution. DOY score is computed as follow:

$$\text{Score}_{\text{DOY}} = = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{xi-\mu}{\sigma})^2} \quad (16)$$

Where σ is the DOY standard deviation, μ is the target DOY and x_i is DOY of a given candidate image. DOY score concerns only difference in acquisition date regardless of acquisition year. The closer candidate DOY to target DOY, the higher DOY grade the candidate image possess.

Distance to cloud/cloud shadow is calculated by a Sigmoid function of distances from the pixel to cloud/cloud shadow, obtained from the file sr_cfmask [43], in radius of 50 pixels around. This score is computed as follow:

$$\text{Score}_{\text{Cloud/Shadow\_Distance}} = \frac{1}{1+ e^{(-0.2*(\min\{Di,Dreq\}-\left(\frac{Dreq-Dmin}{2}\right)))}} \quad (17)$$

Where D_i is a given pixels distance to clouds/cloud shadows, D_{req} is pre-defined minimum required distance (50 pixels), D_{min} is the minimum distance of the given pixel observations. D_{min} is 0 in this study. The closer to clouds/cloud shadows is the candidate pixel, the lower is its grade.

The opacity score requires an aerosol image as input [44], but L8SR provides only discrete aerosol information (i.e. 4 aerosol levels) in the sr_cloud files. Therefore, I assigned opacity scores to the aerosol levels using a sigmoid function. Finally, a pixel's total score is derived by summing up the four sub-scores. The candidate pixel owning the greatest total score is chosen to replace the clouded pixel in the target image (Table 6).

Table 6. Summary of Year score, DOY score, Opacity score and Distance to cloud/cloud shadow for L8SR composition

Year		**DOY**	**Opacity**		**Distance to cloud/cloud shadow**
Score	**Description**		**Score**	**Description**	
1.00	2016	Constraint is ± 30 days from target day. Scoring by a Gaussian function.	0.023	High aerosol content	Constraint is 50 pixel radius from considered pixel.
0.68	2015		0.223	Average aerosol content	
0.42	2014		0.777	Low aerosol content	
0.22	2013		0.977	Climatology-level aerosol content	

					Scoring by a Sigmoid function.

Figure 16 shows clear observation count maps for four target images DOY 137, 169, 265 and 281. It is clear that pixels of DOY 137 and DOY 265 have more cloud-free candidates than those of DOY 169 and DOY 281. On average, each clouded pixel in the target images has 4-5 candidates for compositing process.

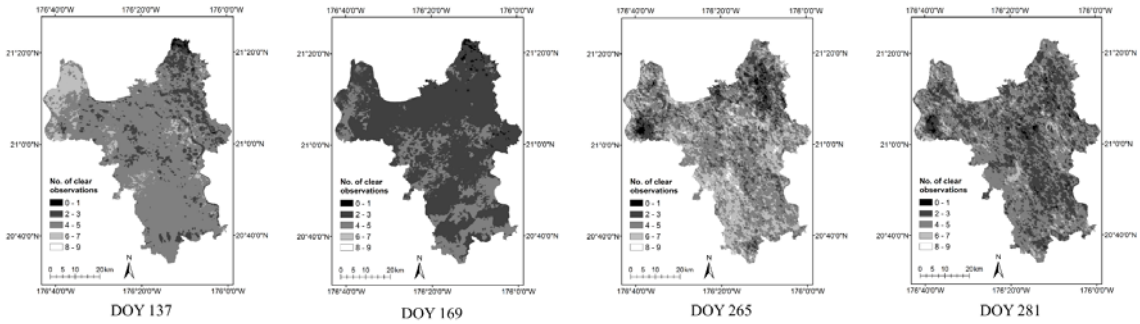

Figure 16. Clear observation count maps for each image used in the compositing process (DOY 137, 169, 265, 281)

3.3.2. Land cover classification

Three classification methods are investigated as in Figure 13. First, an XGBoost classifier is applied on 7 spectral bands of each composite image to obtain 5 LC maps for 2016. The second is time-series classification using XGBoost classifier on stack of 7 spectral bands of 5 composites (i.e. 35 spectral-temporal features). After that, they are compared to assess if a time-series of composites is better than individual composites for classification. The third improves the time-series composite classification by adding Mean Standard Deviations (MSDs) of each band calculated from the composites. Five single classifiers (XGBoost, LR, SVM-RBF, SVM-Linear and MLP) and an ensemble model using majority voting (i.e. predicted class labels are voted by five classifiers having the same weight) are compared. The selection of these classifiers is due to wide applications for LCC using SVM and MLP [31], [33] and LR [32] reported in literature. Additionally, XGBoost is investigated due to novelty (Chen and Guestrin 2016) and current lack of LCC applications. Besides these classifiers, Random Forest (RF) and K-nearest neighbours (KNN) are also investigated in ensemble method. However, RF and KNN achieved lower accuracy as compared to the other classifiers in both single-classifier tests and ensemble test. Therefore, RF and KNN are excluded in reported experiments.

All of these classifiers have specific hyper-parameters that require tuning for the best classification performance. Specifically, SVM-RBF's hyper-parameters are penalty (C) and gamma. SVM-Linear requires penalty (C) only. Important hyper-parameters

forming a base architecture of MLP include activation function (activation), number of hidden layers (hidden layers) and number of hidden nodes in individual hidden layers (hidden nodes). Similar to SVM, LR also has a regularization parameter (C) for individual training data importance (Hackeling 2017). XGBoost has many hyper-parameters in which the three most important ones are the number of boosted trees (n_estimators) and two others for over-fitting prevention: maximum tree depth (max_depth) and minimum sum of weights of all observations required in a child (min_child_weight).

All classifications were performed on the same training and testing points. To select best hyper-parameters for each classifier, 10-fold cross validation and grid-search techniques on the training set are used. .Then all training data is used to train classifiers with best parameters. Testing sets are separated from training sets to assess trained classifiers. Scikit-learn implementation of the classifiers is employed in the experiments (http://scikit-learn.org). Scikit-learn is a python-based machine learning library with robust tools and easy-to-use interface. It is also built on Python thus is appropriate for satellite processing as it can be written with other image processing libraries such as Numpy and GDAL.

3.4. Metrics for classification assessment

Overall accuracy (OA), kappa coefficient, producer accuracy (PA), user accuracy (UA) and F1 score (F1) are used as evaluation metrics in this study [45], [46]. OA and kappa coefficient are computed for classification level. PA, UA and F1 are class specific. Formula of the metrics are presented below.

$$OA = N_{correct} / N_{total} \tag{18}$$

$$UA = N^{T}_{correct} / N^{T}_{classified} \tag{19}$$

$$PA = N^{T\_ref}_{correct} / N^{T\_ref} \tag{20}$$

$$F1 = 2\frac{PA*UA}{PA+UA} \tag{21}$$

In which:

$N_{correct}$: number of correct classified points.

N_{total}: total number of points.

$N^{T}_{correct}$: number of correctly classified points in a given class.

$N^{T}_{classified}$: number of classified points in a given class.

$N^{T\_ref}_{correct}$: number of correctly classified in reference data of a given class.

N^{T_ref}: number of points in reference data in a given class.

Additionally, classification maps are validated against statistical data and visually examined.

CHAPTER 4. EXPERIMENTS AND RESULTS

This chapter presents results of the method including compositing results (section

4.1), land cover classification results based on point validation (section 4.2), land cover classification results based on map validation (section 4.3). Conclusion of advantages and disadvantages of the method are discussed at the end of the chapter.

4.1. Compositing results

Before composition, the average cloud percentage over 5 target images is 20.54% where image at DOY 169 is cloudiest with 73.63% cloud pixels. After compositing, all images are at least 99.78% clear (i.e. DOY 265). However, there are remaining cloudy pixels without replacement candidates. 2015 data mostly contributes to composition with 72.36%, followed by 2013 (22.04%), 2014 (5.55%) and 2016 (0.05%) data.

NDVI and Bare Soil Index (BSI) temporal profiles of seven land cover classes are presented in Figure 15. NDVI and BSI are spectral indices which are calculated from Landsat 8 spectral bands. NDVI is an index of plant "greenness". BSI is sensitive to soil content on the ground. Formulas of NDVI and BSI are presented below.

$$NDVI = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \tag{22}$$

$$BSI = \frac{(\text{SWIR} + \text{Red}) - (\text{NIR} + \text{Blue})}{(\text{SWIR} + \text{Red}) - (\text{NIR} + \text{Blue})} \tag{23}$$

Where NIR is near infrared band (Band 4), Red is red band (Band 3), SWIR is short-wave infrared band (Band 6) and Blue is blue band (Band 2).

From Figure 17, it could be seen that seven classes can be divided into four distinct groups: (impervious area, bare land), paddy rice, water, and (tree, crop, grass and shrub). Due to cultivation practices, paddy rice's NDVI and BSI temporal profile varies across the year.

Although pixel candidates are carefully selected by BAP, they are still spectrally different from neighbouring pixels of other candidate images. For example, for DOY 265 in Figure 18b, composite pixels over a rice planting area show different colour blocks. Some cloudy pixels are replaced by vegetated observations while others are replaced by flooded observations. This indicates selection of appropriate images has significant impact on BAP composites for areas with a high temporal dynamic of land-cover types, especially rice and agricultural areas. Thus, knowledge of local agricultural calendar could improve image selection for spectrally-uniform BAP composites.

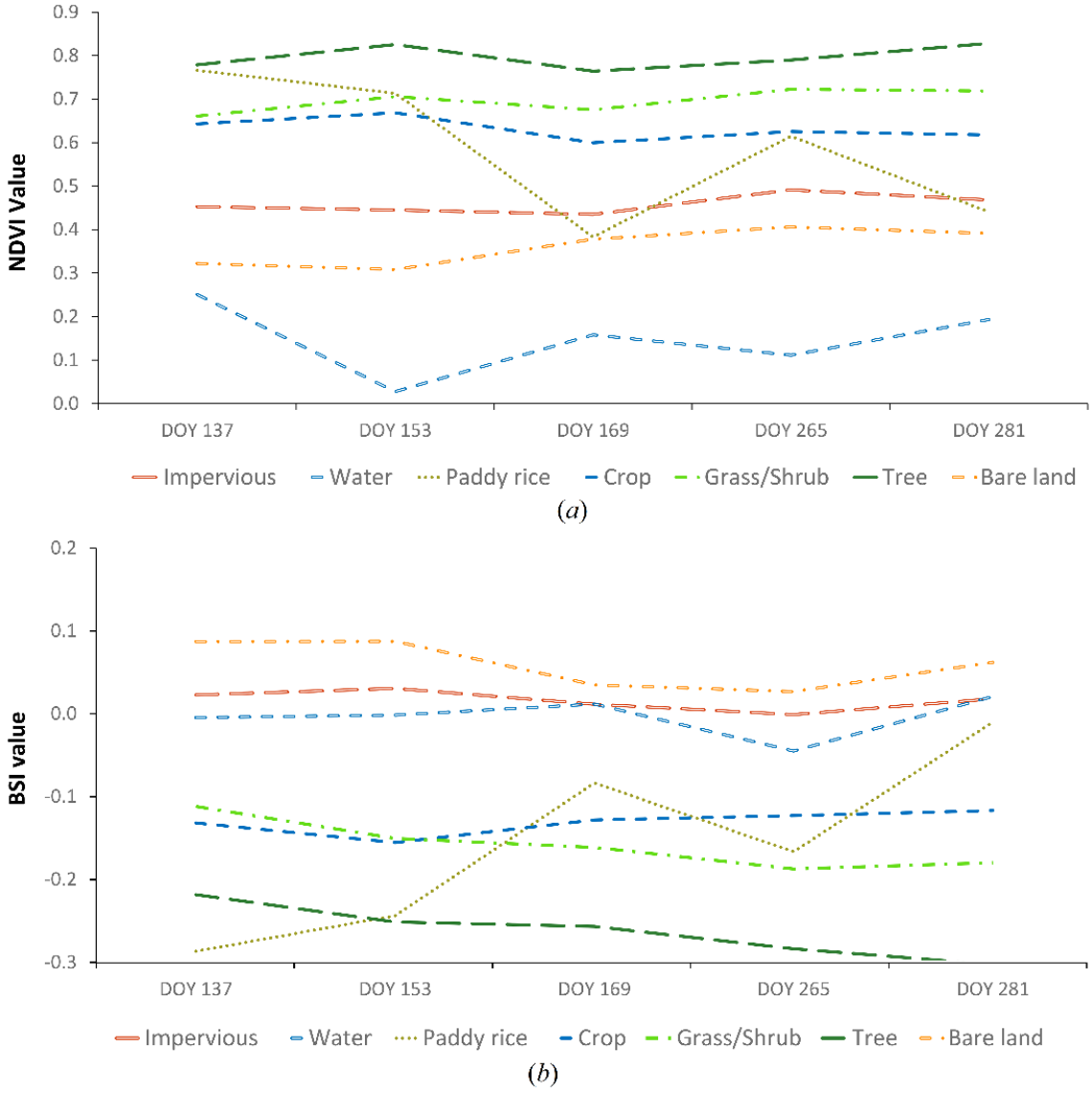

Figure 17. NDVI (above) and BSI (below) temporal profile of land-cover class

4.2. Assessment of land-cover classification based on point validation

4.2.1. Yearly single composite classification versus yearly time-series composite classification

Test set validation results are provided in Table 7. It is found that classifications using time-series composites outperformed all single-image classifications with 10.03% higher OA and 0.13 higher kappa coefficient on average. Single-image classification is also unstable as the results ranging from 68.43 – 76.38% for OA, 0.59 – 0.68 for kappa coefficient. Three out of five single-image classifications achieved greater than 72% OA, except for the DOY 169 and DOY 265, which have higher BAP pixels included, with 73.60% and 24.76% OA respectively.

Table 7. F1 score, F1 score average, OA and kappa coefficient for 7 land cover classes of six classification cases obtained using XGBoost. Best classification cases are written in bold.

	DOY 137	**DOY 153**	**DOY 169**	**DOY 265**	**DOY 281**	**Time series**
Crop	0.50	0.39	0.36	0.33	0.40	**0.58**
Bare land	0.06	**0.26**	0.04	0.17	0.14	0.22
Paddy rice	0.87	0.84	0.81	0.73	0.80	**0.91**
Water	0.85	0.86	0.73	0.81	0.83	**0.91**
Tree	0.67	0.70	0.66	0.65	0.74	**0.80**
Impervious area	0.84	0.87	0.78	0.83	0.86	**0.90**
Grass/Shrub	0.36	0.29	0.30	0.27	0.28	**0.44**
F1 score average	0.76	0.74	0.69	0.68	0.73	**0.82**
OA (%)	76.4	75.7	69.7	68.4	73.6	**82.8**
kappa coefficient	0.68	0.68	0.61	0.59	0.66	**0.77**

Considering per-class accuracy, classification of vegetation classes are significantly improved with time series classification, as those classes have high temporal dynamics which are best captured by multiple observations. From the results, rice in green stage in DOYs of 137, 153, 265 is most confused with crop and grass/shrub (see Figure 18c). In DOY 169, rice fields are flooded, thus resulting in confusion of rice and water. In the last image, DOY 281, harvested rice is confused with bare land and impervious area (Figure 18c). By integrating all confusing information in time-series classification, rice are better separated from other vegetation classes with F1=0.91 (Figure 18d).

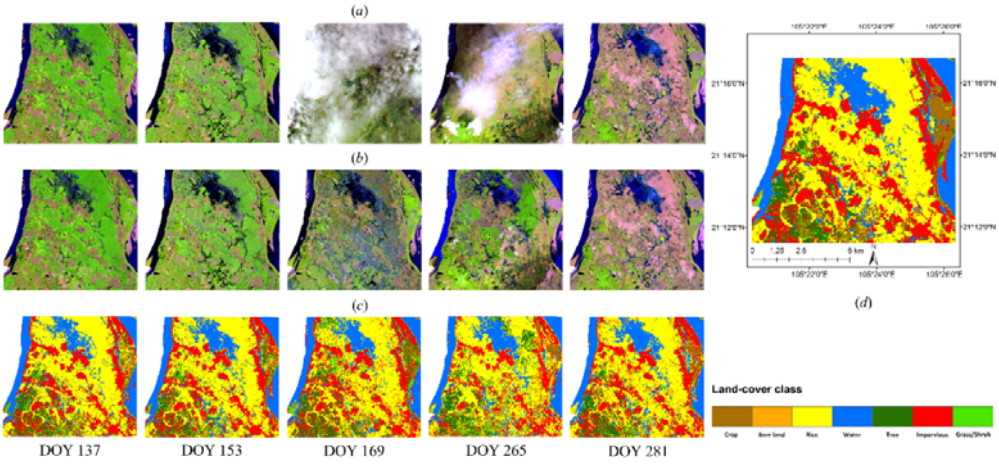

Figure 18. (a) Original surface reflectance images, (b) composite images, (c) classification maps for each image, and (d) classified map obtained from time-series composite images.

Although most LC classes are better identified in time-series classification, bare land had confusion with impervious area (maximum F1=0.26, the time-series F1=0.22). This is attributed to the two classes having spectrally similar and stable reflectance through time. Crop and grass/shrub are occasionally misclassified due to similar spectral signals and mixed pixels. Water is separable from other classes due to its unique spectral properties, but some water bodies are seasonally vegetated, leading to misclassification of water and vegetation. Thus, water also benefits from multiple image observations.

4.2.2. Improvement of ensemble model against single-classifier model

For ensemble classification, the following single models with their optimized parameters are employed: i) XGBoost with n_estimators=1000, max_depth=5, min_child_weight=1; ii) LR with C=1; iii) SVM-RBF with C=10, gamma=0.03125; iv) SVM-Linear with C=8; v) MLP with activation=tank, hidden layers=1, and hidden nodes=40. Classifiers perform on a stack of 35 spectral temporal features and 7 MSDs of spectral bands. Majority voting technique is employed for the ensemble model.

Table 8. OA, kappa coefficient, F1 score average for each single-classifier and ensemble model. Best classification cases are written in bold.

Measure	**Classifier**					
	XGBoost	**LR**	**SVM-RBF**	**SVM-Linear**	**MLP**	**Ensemble**
OA (%)	83.2	82.6	82.9	81.9	83.1	**84.0**
kappa coefficient	0.77	0.77	0.78	0.77	0.78	**0.79**
F1 score average	0.82	0.82	0.83	0.83	0.83	**0.84**

Using an ensemble of supervised classifiers improves the classification (Table 8). It is seen that individual models have similar accuracies with SVM-Linear is the lowest at 81.94% OA and XGBoost is the highest with 83.23% OA. The ensemble model is better than all individual models with OA=83.96% and kappa coefficient=0.79. Per-class accuracies of the ensemble model filter the best results from all single-classifier models. Classifier F1 score performance is presented in Figure 19.

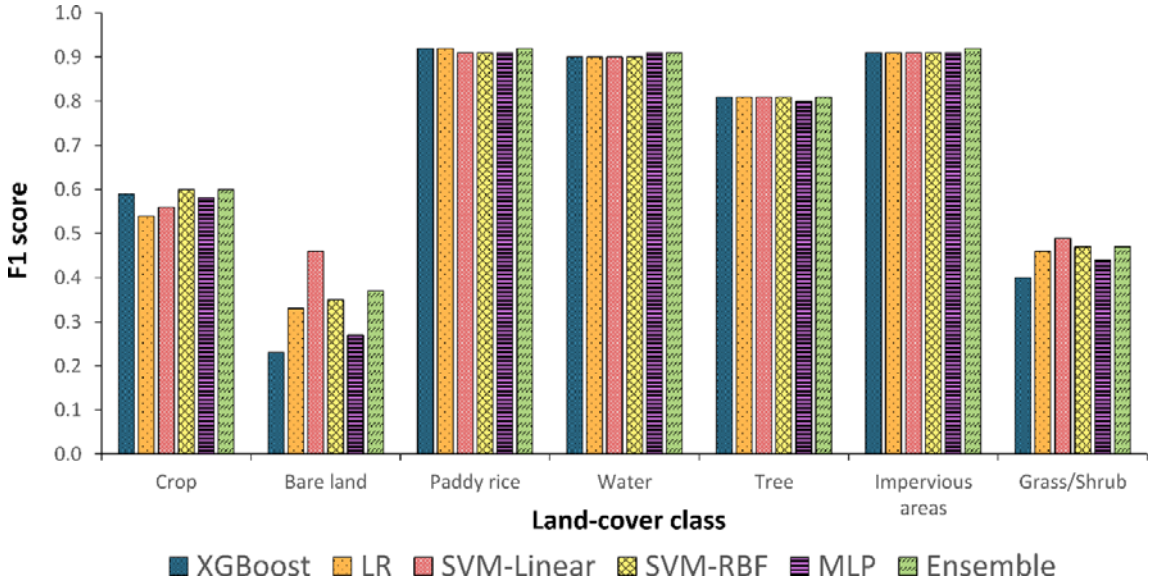

Figure 19. F1 score for land-cover class obtained using multiple classifiers.

XGBoost is not effective at classifying bare land (F1=0.23) and grass/shrub (F1=0.4), but this disadvantage is overcome by SVM-RBF and SVM-Linear with F1 of 0.35, 0.46 for bare land and 0.47, 0.49 for grass/shrub respectively. SVM-RBF and SVM-Linear are generally high performing. Paddy rice, impervious area, water and tree have similar accuracies between classifiers which could be explained as the classes are quite separable in this time-series domain. MLP is overall good compared to other classifiers, but it performs poorly on bare land (F1 = 0.27). Ensemble model achieved similar accuracies of paddy rice, water, tree and impervious areas as compared to other classifiers. However, for crop, grass/shrub and bare land which are easily confused with other classes (Figure 19), ensemble model generally achieved better classification accuracies than any single-classifier model. By integrating models, individual strengths remain, while weaknesses are reduced. Table 9 presents confusion matrix of the ensemble model with User Accuracy (UA) and Producer Accuracy (PA) for each class.

Table 9. Confusion matrix of ensemble model.

	Crop	**Bare land**	**Rice**	**Water**	**Tree**	**Imper vious**	**Grass/ Shrub**	**Reference total**	**UA (%)**
Crop	**222**	3	25	4	24	22	31	331	66.1
Bare land	6	**22**	1	1	0	22	4	56	33.5
Rice	37	0	**581**	16	2	3	7	646	91.6
Water	5	0	11	**411**	4	11	4	446	90.9
Tree	26	2	3	2	**433**	8	17	491	83.2
Impervious	19	6	4	3	5	**485**	1	523	93.1
Grass/Shrub	56	7	12	5	47	11	**117**	255	38.9

Classification total	371	40	637	442	515	562	181	**2748**	**OA (%)**
PA (%)	55.1	41.0	92.8	92.0	79.3	90.5	59.8	**OA (%)**	**84.0**

4.3. Assessment of land-cover classification results based on map validation

The LC map of the ensemble model is displayed in Figure 20. It is observed that paddy rice and impervious area are the dominant classes.

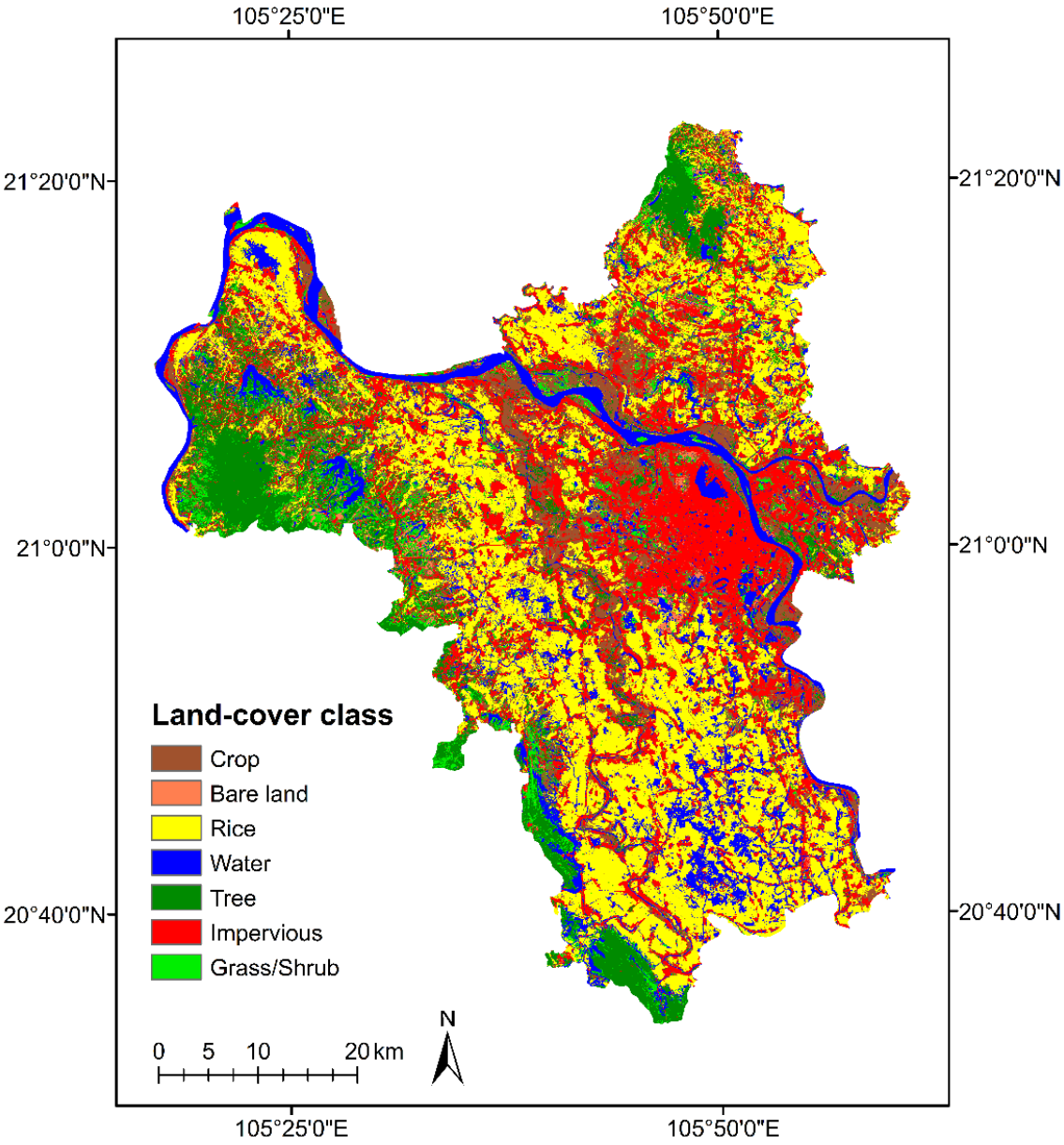

Figure 20. 2016 Land-cover map for Hanoi based on the most accurate classification using time-series composite imagery and the ensemble of five classifiers.

According to Hanoi Statistic Office, rice area in Hanoi for the spring-summer season

is approximately 99,454 ha [47]. Rice area is computed for the classification maps and compared to the official statistic. The ensemble rice map is closest to the official number, and slightly overestimates by 4,764 ha (4.79%). Additional classifiers are shown in (Table 10).

Table 10. Error (ha and %) of rice mapped area for different classification scenarios.

Classification	**DOY 137 composite**	**DOY 153 composite**	**DOY 169 composite**	**DOY 265 composite**	**DOY 281 composite**	**Time series of composites**	**Time series of composites with optimization**
Error(ha)	+8,657	+13,510	+15,674	+16,785	+8,990	+7,811	+4,764
Error(%)	8.70	13.58	15.76	16.88	9.04	7.85	4.79

To summary, the best land-cover map using the ensemble model achieved 83.91% OA with kappa coefficient of 0.79. This is in comparison to 72% OA using the unmodified compositing algorithm in a slightly larger region and a few additional land cover types [22]. Additional regional land cover mapping studies had generally good accuracy with: 89% OA for forest/non-forest cover maps [21], 90% OA for urban landscape with dense time-series stack [48], 89% OA for land cover map in a less-cloudy region with automated pre-processing and random forest [49], 89.42% OA in a recent rice/non-rice cover study over Red River Delta with dense Landsat 8 time-series stack [50], and 84% OA in a recent land cover study over Hanoi employing radar to overcome clouds [51].

Multi-year composition increases cloud-free pixels in composites, especially over cloud-persistent areas such as Hanoi, Vietnam. A time-series composites with over 99% cloud-free pixels was developed. One disadvantage of this compositing is that it does not account for intra-annual vegetation phenology. However, using time-series composites still improves classification performance in comparison with any single composite classification. This is attributed to the effective representation of seasonal temporal dynamics of land-cover types. Among the top supervised classifiers, XGBoost performed best for land cover mapping. However, an ensemble model still improved classification results by promoting individual strengths and reducing weaknesses. This ensemble model is especially effective for confusing classes (bare land, crop, grass/shrub) but not already well-separated classes (paddy rice, water). In the future, image composition accounting for phenology could improve composite quality and classification accuracy for improved mapping of land cover types with high temporal dynamics.

CHAPTER 5. CONCLUSION

In this thesis, I have conducted a research on land cover classification using Landsat 8 satellite images. Specifically, I have presented in this thesis: (i) fundamental concepts of remote sensing sciences, (ii) satellite images and its applications in various domains, (iii) land cover classification problems. A comprehensive review of land cover classification methods has been conducted to address its current developments. LCC is a traditional application in remote sensing. Many LCC studies have been conducted in different places on Earth. However, LCC using optical satellite images in cloud-prone areas with high temporal dynamics of land covers is still challenging due to lack of cloud-free data. In this thesis, I have proposed a LCC method for these areas. The result of this research is also published in the *International Journal of Remote Sensing (Taylor & Francis)* in a paper entitled "*Improvement of land-cover classification over frequently cloud-covered areas using Landsat 8 time-series composites and an ensemble of supervised classifiers*".

In this thesis, I have proposed a LCC method for these areas. Firstly, a dense time-series of composite images was constructed from all available multi-year Landsat 8 images over the study area. A modified compositing method was proposed for the compositing process using Landsat 8 SR images. The result images are almost cloud-free thus are ready for feature extraction. An ensemble of five experimentally strongest supervised classifiers in the experiments was built to classify a stack of composite images and additional features (Mean Standard Deviations). The best land-cover map achieved 83.91% OA with kappa coefficient of 0.79. Some conclusions could be drawn from the research including: (i) multi-year composition increases cloud-free pixels in composites, especially over cloud-persistent areas such as Hanoi, Vietnam; (ii) accurate land cover maps could be derived from time-series composite images; (iii) ensemble learning could slightly improve classification as compared to any single-classifier model, however, significant improvements are observed for confusing classes as in single model, but not for well-separated classes.

There are also some remaining problems including: (i) The compositing method does not account for intra-annual vegetation phenology thus may not be good enough for some land covers like paddy rice; (ii) there are still significant confusions between bare land/impervious surface, grass/crops/trees due to their similar spectral characteristics, even in temporal domain. Therefore, future researches could be placed on improvement of compositing methods for high temporal dynamics land covers. And

development of LCC methods for better separating of bare land/impervious surface, grass/crops/trees.

Reference

1. *Fundamentals of Remote Sensing*, .
2. Hibbard K., Janetos A., van Vuuren D.P., et al. (2010). Research priorities in land use and land-cover change for the Earth system and integrated assessment modelling. *Int J Climatol*, **30**(**13**), 2118–2128.
3. Son N.T., Chen C.F., Chen C.R., et al. (2013). A phenology-based classification of time-series MODIS data for rice crop monitoring in Mekong Delta, Vietnam. *Remote Sens*, **6**(**1**), 135–156.
4. Kuemmerle T., Erb K., Meyfroidt P., et al. (2013). Challenges and opportunities in mapping land use intensity globally. *Curr Opin Environ Sustain*, **5**(**5**), 484–493.
5. Müller H., Rufin P., Griffiths P., et al. (2015). Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. *Remote Sens Environ*, **156**, 490–499.
6. Zhang X., Pan D., Chen J., et al. (2013). Using long time series of Landsat data to monitor impervious surface dynamics: a case study in the Zhoushan Islands. *J Appl Remote Sens*, **7**(**1**), 73515.
7. Arvor D., Jonathan M., Meirelles M.S.P., et al. (2011). Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *Int J Remote Sens*, **32**(**22**), 7847–7871.
8. Wulder M.A., Masek J.G., Cohen W.B., et al. (2012). Remote Sensing of Environment Opening the archive : How free data has enabled the science and monitoring promise of Landsat. *Remote Sens Environ*, 1–9.
9. Le Toan T., Ribbes F., Li-Fang Wang, et al. (1997). Rice crop mapping and monitoring using ERS-1 data based on experiment and modeling results. *IEEE Trans Geosci Remote Sens*, **35**(**1**), 41–56.
10. Kontgis C., Schneider A., and Ozdogan M. (2015). Mapping rice paddy extent and intensification in the Vietnamese Mekong River Delta with dense time stacks of Landsat data. *Remote Sens Environ*, **169**, 255–269.

11. Whitcraft A.K., Vermote E.F., Becker-Reshef I., et al. (2015). Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. *Remote Sens Environ*, **156**, 438–447.
12. Nguyen T.T.N., Bui H.Q., Pham H. V, et al. (2015). Particulate matter concentration mapping from MODIS satellite data: a Vietnamese case study. *Environ Res Lett*, **10**(**9**), 95016.
13. Duong N.D. (2003). Study of Land Cover Change in Vietnam for the Period 2001-2003 Using Modis 32 Days Composite. (**May**).
14. Ngo L.T., Mai D.S., and Pedrycz W. (2015). Semi-supervising Interval Type-2 Fuzzy C-Means clustering with spatial information for multi-spectral satellite image classification and change detection. *Comput Geosci*, **83**, 1–16.
15. Henits L., Jürgens C., and Mucsi L. (2016). Seasonal multitemporal land-cover classification and change detection analysis of Bochum, Germany, using multitemporal Landsat TM data. *Int J Remote Sens*, 1–16.
16. USGS Provisional Landsat 8 Surface Reflectance Data Available | Landsat Missions. <https://landsat.usgs.gov/provisional-landsat-8-surface-reflectance-data-available>, accessed: 11/22/2017.
17. White J.C., Wulder M.A., Hobart G.W., et al. (2014). Pixel-based image compositing for large-area dense time series applications and science. *Can J Remote Sens*, **40**(**3**), 192–212.
18. Cihlar J., Manak D., and D'Iorio M. (1994). Evaluation of compositing algorithms for AVHRR data over land. *IEEE Trans Geosci Remote Sens*, **32**(**2**), 427–437.
19. (2002). An overview of MODIS Land data processing and product status. *Remote Sens Environ*, **83**(**1**–**2**), 3–15.
20. Roy D.P., Ju J., Kline K., et al. (2010). Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sens Environ*, **114**(**1**), 35–49.
21. Potapov P., Turubanova S., and Hansen M.C. (2011). Remote Sensing of Environment Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. *Remote Sens Environ*, **115**(**2**), 548–561.
22. Griffiths P., Van Der Linden S., Kuemmerle T., et al. (2013). A pixel-based

landsat compositing algorithm for large area land cover mapping. *IEEE J Sel Top Appl Earth Obs Remote Sens*, **6**(**5**), 2088–2101.

23. Gómez C., White J.C., and Wulder M.A. (2016). Optical remotely sensed time series data for land cover classification : A review. *ISPRS J Photogramm Remote Sens*, **116**, 55–72.
24. Zald H.S.J., Wulder M.A., White J.C., et al. (2016). Integrating Landsat pixel composites and change metrics with lidar plots to predictively map forest structure and aboveground biomass in Saskatchewan, Canada. *Remote Sens Environ*, **176**, 188–201.
25. Thompson S.D., Nelson T.A., White J.C., et al. (2015). Mapping Dominant Tree Species over Large Forested Areas Using Landsat Best-Available-Pixel Image Composites. *Can J Remote Sens*, **41**(**3**), 203–218.
26. Pickell P.D., Hermosilla T., J. Frazier R., et al. (2015). Forest recovery trends derived from Landsat time series for North American boreal forests. *Int J Remote Sens*, **37**(**1**), 138–149.
27. Hermosilla T., Wulder M.A., White J.C., et al. (2015). An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sens Environ*, **158**, 220–234.
28. Franklin S.E., Ahmed O.S., Wulder M.A., et al. (2015). Large Area Mapping of Annual Land Cover Dynamics Using Multitemporal Change Detection and Classification of Landsat Time Series Data. *Can J Remote Sens*, **41**(**4**), 293–314.
29. Hansen M.C., Roy D.P., Lindquist E., et al. (2008). A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. **112**, 2495–2513.
30. Li G., Lu D., Moran E., et al. (2012). Comparative analysis of classification algorithms and multiple sensor data for land use/land cover classification in the Brazilian Amazon. *J Appl Remote Sens*, **6**(**1**), 61706.
31. Foody G.M. and Mathur A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Trans Geosci Remote Sens*, **42**(**6**), 1335–1343.
32. Mallinis G. and Koutsias N. (2008). Spectral and Spatial-Based Classification for Broad-Scale Land Cover Mapping Based on Logistic Regression. *Sensors*, 8067–

8085.

33. Kavzoglu T. and Mather P.M. (2003). The use of backpropagating artificial neural networks in land cover classification. *Int J Remote Sens*, (**December 2014**), 37–41.
34. Pal M. and Mather P.M. (2006). Support vector machines for classification in remote sensing. *Int J Remote Sens*, (**March 2013**), 37–41.
35. Cortes C. and Vapnik V. (1995). Support-Vector Networks. *Mach Learn*, **20**(**3**), 273–297.
36. Rumelhart D.E., Hinton G.E., and Williams R.J. (1986). Learining Internal Representations by Error Propagation. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 1, 318–362.
37. Chen T. and Guestrin C. (2016). XGBoost: A Scalable Tree Boosting System. *arXiv:160302754v3*.
38. Natekin A. and Knoll A. (2013). Gradient boosting machines, a tutorial. *Front Neurorobot*, **7**(**DEC**).
39. Liu Y., Cui H., Kuang Z., et al. (2017). Ship Detection and Classification on Optical Remote Sensing Images Using Deep Learning. **5012**.
40. Ammour N., Alhichri H., Bazi Y., et al. (2017). Deep learning approach for car detection in UAV imagery. *Remote Sens*, **9**(**4**).
41. Government of Vietnam (2013). Resolution on landuse planning from 2011-2015 and by 2020 for Hanoi. .
42. Hanoi Environment and Natural Resources Department (2010). Land use statistics of Hanoi. <http://qhkhsdd.hanoi.gov.vn>.
43. Zhu Z., Wang S., and Woodcock C.E. (2015). Improvement and expansion of the Fmask algorithm : cloud , cloud shadow , and snow detection for Landsats 4 – 7 , 8 , and Sentinel 2 images. *Remote Sens Environ*.
44. White J.C., Hostert P., Griffiths P., et al. (2014). Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science. *Can J Remote Sens*, (**August 2014**).
45. Congalton R.G. and Green K. (2008), *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, CRC Press, Taylor & Francis Group, Boca Raton.

46. Powers D.M.W. (2011). Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation. *J Mach Learn Technol*, **2**(**1**), 37–63.
47. Office H.S. (2016). Report of socio-economic survey. <http://thongkehanoi.gov.vn/>.
48. Castrence M., Nong D., Tran C., et al. (2014). Mapping Urban Transitions Using Multi-Temporal Landsat and DMSP-OLS Night-Time Lights Imagery of the Red River Delta in Vietnam. *Land*, **3**(**1**), 148–166.
49. Mack B., Leinenkugel P., Kuenzer C., et al. (2017). A semi-automated approach for the generation of a new land use and land cover product for Germany based on Landsat time-series and Lucas *in-situ* data. *Remote Sens Lett*, **8**(**3**), 244–253.
50. Man Duc Chuc, Nguyen Hoang Anh, Nguyen Thanh Thuy, Bui Quang Hung N.T.N.T. (2017). Paddy Rice Mapping in Red River Delta region Using Landsat 8 Images : Preliminary results. *9th Int Conf Knowl Syst Eng (KSE 2017)*.
51. Nguyen D., Wagner W., Naeimi V., et al. (2015). Rice-planted area extraction by time series analysis of ENVISAT ASAR WS data using a phenology-based classification approach: A case study for Red River Delta, Vietnam. *Proceedings of the International Archives Photogrammetry, Remote Sensing and Spatial Information Science, Berlin, Germany*.