An easily implemented method to estimate impervious surface area on a large scale from MODIS time-series and improved DMSP-OLS nighttime light data

An easily implemented method to estimate impervious surface area on a large scale from MODIS time-series and improved DMSP-OLS nighttime light data

Sophak Pok[1], Bunkei Matsushita[2]*, Takehiko Fukushima[3]

[1] Graduate School of Life and Environmental Sciences, University of Tsukuba, Tennoudai 1-1-1, Tsukuba, Ibaraki, 305-8572, Japan

[2] Faculty of Life and Environmental Sciences, University of Tsukuba, Tennoudai 1-1-1, Tsukuba, Ibaraki, 305-8572, Japan

[3] Kasumigaura Environmental Science Center, Okijuku-machi 1853, Tsuchiura, Ibaraki, 300-0023, Japan

*Corresponding Author

Phone: 81-29-853-7190

Fax: 81-29-853-7190

Email: matsushita.bunkei.gn@u.tsukuba.ac.jp

An easily implemented method to estimate impervious surface area on a large scale from MODIS time-series and improved DMSP-OLS nighttime light data

Abstract

It is important for researchers and policy-makers to frequently update the amount and spatial distribution of impervious surface area (ISA) on earth, because the level of imperviousness not only indicates urbanization, but is also a key indicator of ecological conditions. In this study, we developed an easily implemented method for estimating the ISA percentage (ISA%) from vegetation index data obtained from a moderate resolution imaging spectroradiometer (MODIS) and nighttime light data obtained from the Defense Meteorological Satellite Program's Operational Line-scan System (DMSP-OLS). The proposed method consists of four major steps. First, a non-vegetation fraction map was generated from 16-day composited time-series MODIS normalized difference vegetation index data using the temporal mixture analysis method. Second, the enhanced-vegetation-index-adjusted nighttime light index (EANTLI) was used to overcome the saturation problem and blooming effects in the original DMSP-OLS data. Third, the relationship between ISA% and EANTLI was derived based on a statistical analysis of the non-vegetation fraction image and the EANTLI image to obtain a preliminary ISA% map. Finally, the final ISA% map was obtained by selecting smaller values from the preliminary ISA% map and non-vegetation fraction map for each pixel. The validation results showed that the developed method has promising accuracy for estimating the ISA% in our study area (mainly consisting of four Southeast Asian countries: Thailand, Laos, Cambodia, and Vietnam), with a root mean square error value of 0.111, a systematic error value of 0.061, and a determination coefficient of 0.87. Another important finding is that there are two relationships between ISA% and improved nighttime light (i.e., EANTLI): the natural logarithmic function is suitable for ISA% values between 0% and 50%, and the quadratic

polynomial function should be used for ISA% values larger than 50%. The developed method has high potential for application to the generation of global ISA% maps with frequent updates due to its easy implementation and the ready availability of input data.

Keywords: impervious surface area, temporal mixture analysis, EANTLI, MODIS, DMSP-OLS

1. Introduction

The term *impervious surfaces* describes all the man-made types of land cover on the Earth's surface, including rooftops, roads, sidewalks, parking lots and so on. This man-made land cover prevents water from infiltrating into the soil, and thus has different surface characteristics from natural land cover. Increases in the impervious surface area (ISA) in a watershed have a major impact on the hydrological and thermal characteristics of the watershed. Therefore, the ISA is considered a useful index for evaluating urbanization and watershed function (e.g., Arnold and Gibbons, 1996) as well as impact of environmental factors on climatology stations (e.g., Gallo and Xian, 2016).

Remote sensing techniques have been widely used to estimate ISA since the 2000s (see the extensive review by Weng, 2012). One challenge in applying these techniques is that many mixed pixels exist in satellite imagery, especially imagery with medium to coarse resolution. To solve this problem, a large number of studies on estimating ISA at a sub-pixel level have been conducted, and numerous methods have been developed (see Table 1 for a summary of the major existing methods). Among them, the mixture analysis technique is one of the main approaches for remotely estimating ISA with medium to coarse resolution imagery (Weng, 2012). Generally, the mixture analysis technique can be further separated into two sub-categories: spectral mixture analysis (SMA) and

temporal mixture analysis (TMA). The main difference between SMA and TMA is that the former uses spectral information while the latter uses temporal information for each endmember (i.e., pure land cover type). Because it uses temporal information, the TMA can only be applied to satellite imagery with coarse resolution, since the revisit time is longer for satellite imagery with medium to high spatial resolution. For example, Wu and Murray (2003) obtained the reflectance spectra of four endmembers (i.e., vegetation, soil, impervious surface with low albedo, and impervious surface with high albedo) from a Landsat/TM image to estimate the ISA in an American city (an example of the use of SMA); Yang et al. (2012) obtained the temporal profiles of three endmembers (i.e., forest, crop, and ISA) from the moderate resolution imaging spectroradiometer (MODIS) normalized difference vegetation index (NDVI) time-series datasets to estimate the ISA for all of Japan (an example of the use of TMA). Compared to SMA, TMA can efficiently reduce endmember variability, which is one of the largest sources of error in mixture analysis, and thus achieve improved ISA estimation (Somers et al., 2011; Yang et al., 2012). However, TMA has usually suffered from bare land effects when it has been used to estimate the ISA in a large area that contains more than one type of non-vegetation land cover, leading to an overestimation of ISA (Yang et al., 2014).

Table 1 Major methods for estimating ISA percent at sub-pixel level from satellite sensors with medium to coarse spatial resolution

Category	Sub-category	Approach Descriptions	Major Limitations	References for Example
Empirical method	Vegetation distribution-based	Build a relationship between ISA percent and vegetation indices (or greenness) by considering the inverse correlation between impervious surface and vegetation cover in urban areas	Model performacne strongly depend on calibration data	Bauer et al., 2008; Lu et al., 2008
	ANN	Build a relationship between ISA percent and reflectance spectra using neural network techniques	Model performance strongly depend on training data	Lee and Lathrop, 2006; Weng and Hu, 2008
	Nighttime light-based	Build a relationship between ISA percent and DMSP-OLS data	Model performacne strongly depend on calibration data	Elvidge et al., 2007; Lu et al., 2008
Mixture analysis method	SMA	Estimate ISA percent using mixture analysis techniques based on spectral information obtained from satellite data	Suffer from spectral similarity among non-vegetation land cover types and spectral variability within endmembers	Wu and Murray, 2003; Wu, 2004; Lu and Weng, 2006; Powell et al., 2007; Yang et al., 2010
	TMA	Estimate ISA percent using mixture analysis techniques based on temporal information obtained from satellite data	Suffer from temporal similarity among non-vegetation land cover types, and only suitable for coarse resolution imagery	Knight and Voth, 2011; Yang et al, 2012; Yang et al., 2014; Li and Wu, 2015

ANN: artificial neural network; SMA: spectral mixture analysis; TMA: temporal mixture analysis

Nighttime light (NTL), which can be obtained from the Defense Meteorological Satellite Program's Operational Line-scan System (DMSP-OLS), is another data source that is widely used for estimating ISA (e.g., Elvidge et al., 2007) or improving ISA estimation (e.g., Lu et al., 2008; Yang et al., 2014). However, pixel saturation in city cores and pixel blooming in suburbs are two obstacles to the successful application of DMSP-OLS data. Recently, several attempts have been made to address this issue (Lu et al., 2008; Zhang et al., 2013; Zhuo et al., 2015; Liu et al., 2015). The common concept underlying these efforts is the generation of a new index using a combination of the DMSP-OLS NTL and MODIS data-based indices. For example, Lu et al. (2008) proposed a human settlement index (HSI) determined through the combined use of a normalized DMSP-OLS NTL and the maximum MODIS-NDVI between April and October in 2000; Zhang et al. (2013)

modified the HSI to develop a vegetation-adjusted NTL urban index (VANUI) to overcome the drawbacks of the HSI; Zhuo et al. (2015) developed another index by combining the DMSP-OLS NTL with the MODIS enhanced vegetation index (EVI), which was named the EVI-adjusted NTL index (EANTLI), to further improve the performance of the HSI and VANUI; in the same year, Liu et al. (2015) also reported a new index by combining the DMSP-OLS NTL with the MODIS EVI and normalized difference water index (NDWI) and named it the normalized urban areas composite index (NUACI), which also showed better performance than the HSI and VANUI.

Coarse spatial resolution images are suitable for estimating ISA on a large scale (regional or global) because their processing is not time or labor intensive and thus the product can be frequently updated (Xian and Homer, 2010). Another important point is that the use of coarse spatial resolution images (usually having high temporal resolution) allows users to benefit from the merits of the TMA. In addition, it can be considered that the saturation problems and blooming effects in the original DMSP-OLS NTL data will limit its usefulness, while the improved DMSP-OLS NTL data (i.e., EANTLI) will help us to more efficiently mitigate the bare land effects in TMA-based ISA estimation. Consequently, the objective of this study is to propose an easily implemented method that can accurately estimate ISA on a large scale from MODIS time-series and the improved DMSP-OLS NTL data.

2. Study area

The study area mainly consisted of four southeast Asian countries (Thailand, Laos, Cambodia and Vietnam; Fig. 1) and contained several metropolitan areas, including Bangkok (Thailand), Hanoi (Vietnam), Ho Chi Minh City (Vietnam), Vientiane (Laos), and Phnom Penh (Cambodia). Six MODIS tiles (h27v6, h27v7, h27v8, h28v6, h28v7 and h28v8) were needed to cover the study area.

The study area is characterized by a monsoon climate with a dry season and wet season of more or less equal length. The distribution of mean annual rainfall over this area is highly variable, ranging from more than 3000 mm in north-central Laos to less than 1000 mm in the much drier areas of northeast Thailand (MRC, 2011). A part of the study area annually experiences flooding, with inundation lasting between two and six months. Three broad land use types were identified in the region, i.e., paddy fields, mixed forest, and land cultivated for cash crops (MRC, 2011).
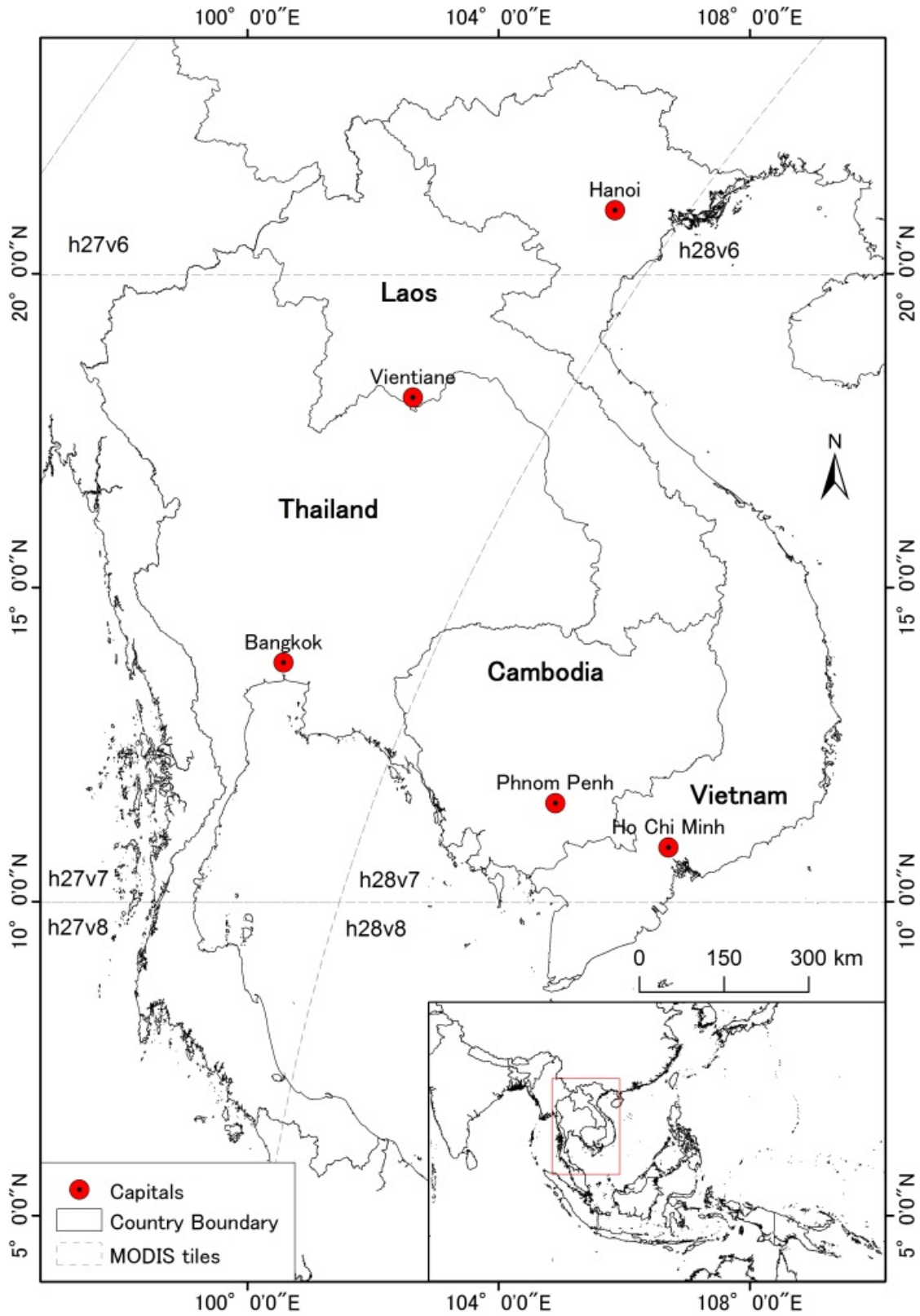
Fig. 1. Location of the study area and the six MODIS tiles used in this research.

3. Data sources

3.1. MODIS data

Three MODIS products were used in this study. These products were all acquired from the U.S. Geological Survey website (http://glovis.usgs.gov/), and reprojected using the MODIS Reprojection Tool to meet the projection of the DMSP-OLS NTL data (i.e., geographic projection with datum of WGS84). The first product was the MODIS 16-day composited NDVI with a spatial resolution of 1 km (MOD13A2) for 2001. There are 23 elements in total for one year which can provide temporal information about the land surface. The downloaded MODIS NDVI temporal profiles were stacked as one multi-layer file and further smoothed using a Savitzky-Golay filter-based method to minimize the remaining cloud/noise effects and/or effects caused by atmospheric variability and imperfect sensor calibration (Chen et al., 2004). The smoothed NDVI data were then sorted in ascending order, and the last 12 maximum NDVI elements were extracted. This is because previous studies have shown that the sorted NDVI temporal profiles (especially for high NDVI value zones) can reduce endmember variability and thus result in more accurate estimation of ISA (Yang et al., 2012; Yang et al., 2014).

The second product was the MODIS monthly composited EVI with a spatial resolution of 1 km (MOD13A3). This product was used together with the DMSP-OLS NTL data to calculate the EANTLI (Zhuo et al., 2015). The third product was the MODIS Land Water Mask with a 250-m spatial resolution (MOD44W). This product was used to mask water pixels before the ISA estimation was performed.

3.2. DMSP-OLS NTL data

The global DMSP-OLS NTL stable light product of 2001 (version 4; geographic projection with datum of WGS84; spatial resolution: 1 km) from DMSP satellite F15 was downloaded from the NOAA website (http://ngdc.noaa.gov/eog/index.html). The area corresponding to the study area was then cut out. The effects of moonlight, stray light, clouds, ephemeral fires, and gas flares have been removed from the product to ensure that the lights are almost all from human settlements (Elvidge et al., 2007). The light saturation problem in urban centers and the blooming effects in suburban areas were mitigated by combining the data with MODIS EVI data (Zhuo et al., 2015).

3.3. Historical images from Google Earth

The reference ISA data for accuracy assessment were collected from high-resolution Google Earth images of 2001-2004. In total, 97 stratified random samples were collected in order to evaluate the accuracy of the ISA estimation in a full dynamic range (i.e., ISA between 0% and 100%). For each reference sample, a 3 km x 3 km sampling unit (corresponding to 3 x 3 pixels of MODIS products or DMSP-OLS NTL data) was utilized to reduce the impact of geometric errors associated with different datasets. The 97 extracted sample polygons (3 km x 3 km) were then classified as ISA or non-ISA using the unsupervised classification method with visual interpretation or the visually digitizing method in ENVI 5.2. Finally, the ISA fraction was calculated for each reference sample.

4. Development of an easily implemented method for estimating ISA from MODIS and DMSP-OLS NTL data

A flowchart of the proposed method is shown in Fig. 2. The method contains four major steps,

which will be described in subsections 4.1 (Step 1 in Fig. 2), 4.2 (Step 2 in Fig. 2), and 4.3 (Steps 3 and 4 in Fig. 2).
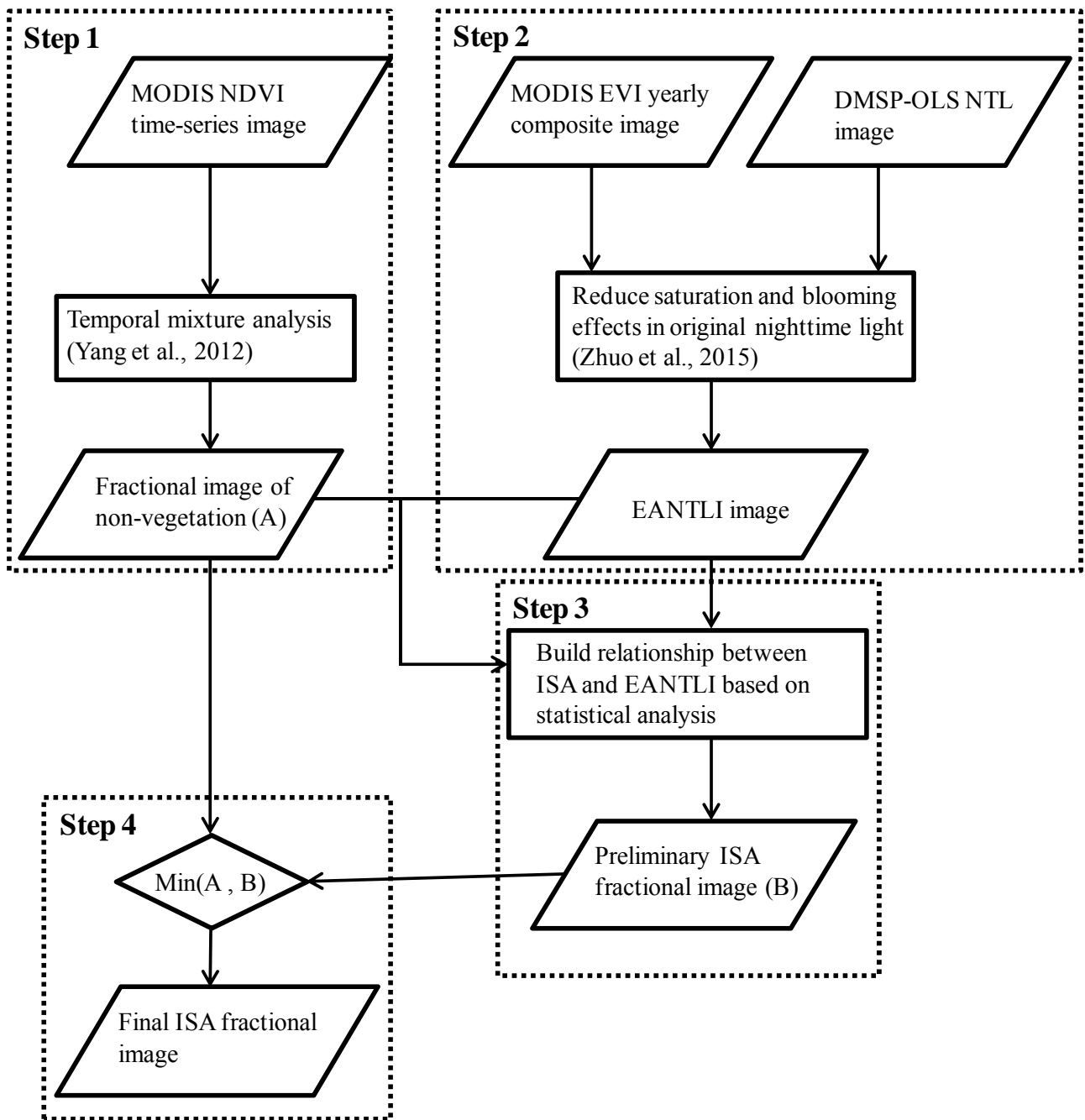

Fig. 2. Flowchart of the proposed method. Min (A, B) represents the selection of smaller value between images A and B.

4.1. Estimating the non-vegetation fraction from time-series MODIS NDVI data using the TMA method

In the TMA method, the NDVI value of a mixed pixel is considered to be the linear combination of the NDVI values of the endmembers and can be mathematically written as (Yang et al., 2012):

$$\text{NDVI}_{\text{mix}} = \sum_{i=1}^{n} f_i \text{NDVI}_i + \varepsilon \tag{1}$$

where NDVI_{mix} is the temporal profile of the NDVI of the target pixel, NDVI_i is the temporal NDVI signature of endmember i, f_i is the fraction of endmember i, and ε is the residual representing the model error. The fractions of the endmembers are commonly constrained by:

$$\sum_{i=1}^{n} f_i = 1 \text{ and } f_i \geq 0 \tag{2}$$

In this study, the fraction of each endmember was estimated using the constrained_min function in ver. 8.4 of the Interactive Data Language (IDL; Exelis Visual Information Solutions Inc., Boulder, CO).

A minimum noise fraction (MNF) transform was carried out for the last 12 maximum NDVI values in sorted NDVI temporal profiles to facilitate the selection of endmembers. The scatter plots of the first three MNF components were used to guide the selection of the endmembers (Fig. 3).

From Fig. 3, four endmembers were identified. They were forest, multi-crop, single-crop and non-vegetation (ISA and/or bare land). The temporal profiles of the selected endmembers are shown in Fig. 4. It can be seen that the NDVI temporal profile of forest showed almost no change at high NDVI values, whereas the NDVI temporal profile of non-vegetation showed almost no change, but showed low NDVI values. In contrast, the NDVI temporal profiles of crops (both single- and

multi-crop) showed increasing trends in the NDVI values, which indicated the presence of phenological characteristics different from those of the other two endmembers (i.e., forest and non-vegetation).
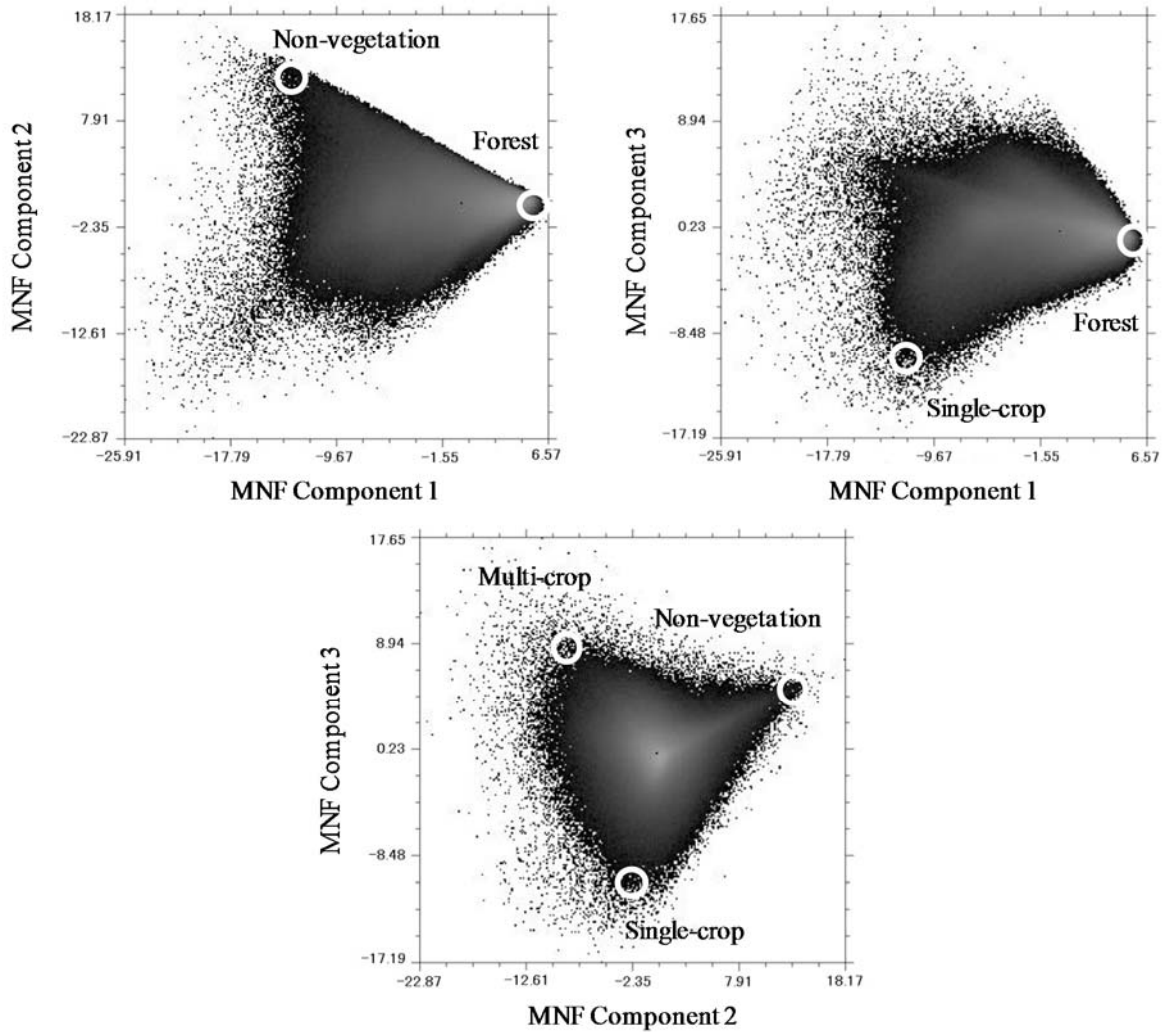

Fig. 3. Feature space representations of the first three minimum noise fraction (MNF) components. Four endmembers were identified: forest, multi-crop, single-crop and non-vegetation (including impervious surface and bare land).
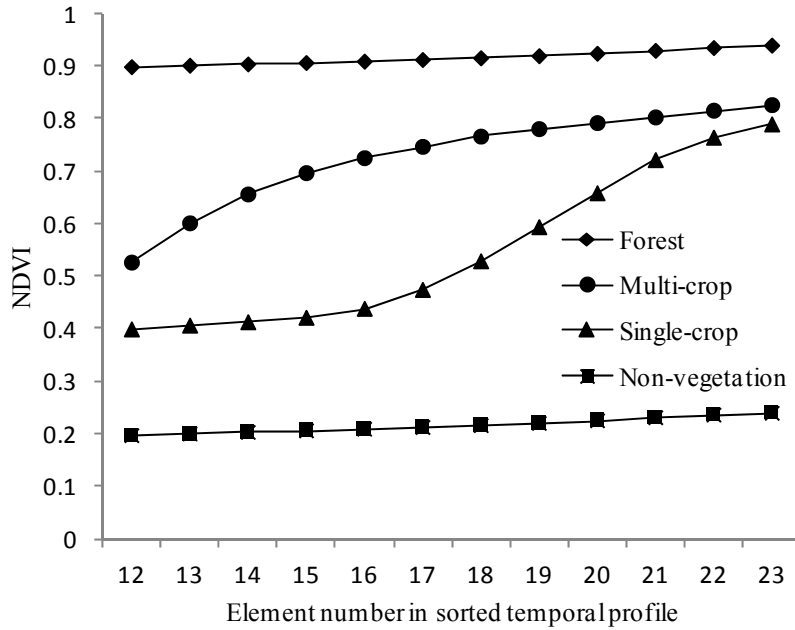
Fig.4. NDVI temporal profiles of endmembers for temporal mixture analysis.

The fraction image of the non-vegetation endmember was calculated by Eq. (1) using the constraints of Eq. (2). In addition, since the purpose of this study was to estimate ISA, the pure vegetation pixels (i.e., pixels with NDVI values larger than 0.8 for all 12 elements) were identified and masked out before the fraction estimation to save computation time.

4.2. *Calculating EANTLI from MODIS EVI and DMSP-OLS NTL data*

The EVI-adjusted NTL index (EANTLI) developed by Zhuo et al. (2015) is used to overcome the two main problems in the original DMSP-OLS NTL data (i.e., light saturation in urban centers and the blooming effect in suburban regions). It is known that the light saturation in urban centers conceals light intensity variations and spatial details, while the blooming effect leads to the overestimation of the lighted area in suburban regions (Lu et al., 2008; Zhang et al., 2013; Zhuo et al., 2015). Mathematically, the EANTLI can be written as (Zhuo et al., 2015):

14

$$\text{EANTLI} = \frac{1 + \big((\text{NTL})_{\text{norm}} - (\text{EVI})\big)}{1 - \big((\text{NTL})_{\text{norm}} - (\text{EVI})\big)} (\text{NTL}) \tag{3}$$

where $(\text{NTL})_{\text{norm}}$ represents the normalized DMSP-OLS NTL value (divided by the maximum value of 63), (NTL) is the original DMSP-OLS NTL value, and (EVI) is derived from the monthly MODIS-EVI products. According to Eq. (3), it is clear that the EANTLI can enlarge the original NTL digital number (DN) values when $(\text{NTL})_{\text{norm}}$ is larger than (EVI) (i.e., the potential saturated area) and reduce the NTL DN values when $(\text{NTL})_{\text{norm}}$ is smaller than (EVI) (i.e., the potential blooming area), and thus corrects the light saturation in urban centers and mitigates the blooming effects in suburban regions. The details of the rationality and performance of EANTLI can be found in Zhuo et al. (2015).

The EANTLI image was generated from the MODIS EVI product and DMSP-OLS NTL data using Eq. (3).

4.3. Building a relationship between ISA and EANTLI

Figure 5 shows the results of a statistical analysis of the original and improved (i.e., EANTLI) DMSP-OLS NTL for 10 non-vegetation fraction groups ranging from 1% to 100% divided into deciles (i.e., 1-10%, 11-20%, 21-30%, 31-40%, 41-50%, 51-60%, 61-70%, 71-80%, 81-90% and 91-100%). The non-vegetation fractions were obtained from the results of Section 4.1. The NTL or EANTLI values from the lower to the upper limit at the 5th, 25th, 50th, 75th, and 95th percentiles are shown for each non-vegetation fraction group. Figure 5 shows that: (1) very low NTL or EANTLI values appeared even for some pixels with a higher non-vegetation fraction (e.g., larger than 50%); (2) a light saturation problem appeared in Fig. 5a when the non-vegetation fraction was larger than

50%; (3) the light saturation problem disappeared if EANTLI was used instead of NTL (Fig. 5b); (4) higher NTL values can be observed even for pixels with a very low non-vegetation fraction (e.g., less than 20%); (5) this problem was mitigated if EANTLI was used instead of NTL; (6) the EANTLI values at the 95th percentile line are in direct proportion to the non-vegetation fractions. The first finding indicates that the TMA method suffered from the problem of the mixture of ISA and bare land in a MODIS pixel. The second to fifth findings suggest the improvements obtained by using EANTLI instead of NTL. The sixth finding provides a hint as to how to build a relationship between ISA and EANTLI.

Generally, NTL should become brighter with the increase of the ISA percentage (ISA%) because both the ISA and the NTL are good indicators of urbanization (Elvidge et al., 2007). From Fig. 5, it can be seen that only the EANTLI values at the 95th percentile line can maintain the positive correlation of ISA% with NTL. In addition, the pixels with maximum EANTLI values in each non-vegetation fraction group should probably only include ISA as non-vegetation land cover (i.e., the non-vegetation fraction can be considered to be the ISA fraction for these pixels), whereas the pixels with smaller EANTLI values probably include both ISA and bare land as non-vegetation land cover. In this study, the EANTLI value at the 95th percentile line was used instead of the maximum EANTLI value in each group to build the relationship between ISA% and EANTLI. This was done for two reasons. First, the EANTLI values probably varied slightly even for the same ISA% (e.g., they varied from the 95th percentile line to the maximum). Second, the maximum EANTLI values were probably due to the imperfect removal of ephemeral lights from fires from those pixels (e.g., gas flares and forest fires; Elvidge et al., 2007). The sensitive analysis results shown in Fig. 6 also indicate that the use of EANTLI values at the 95th percentile line is reasonable. From Fig. 6, it can be seen that the highest correlation coefficient and the smallest error index among the tested percentile lines occurred at the 95th percentile, except for the case of a

16

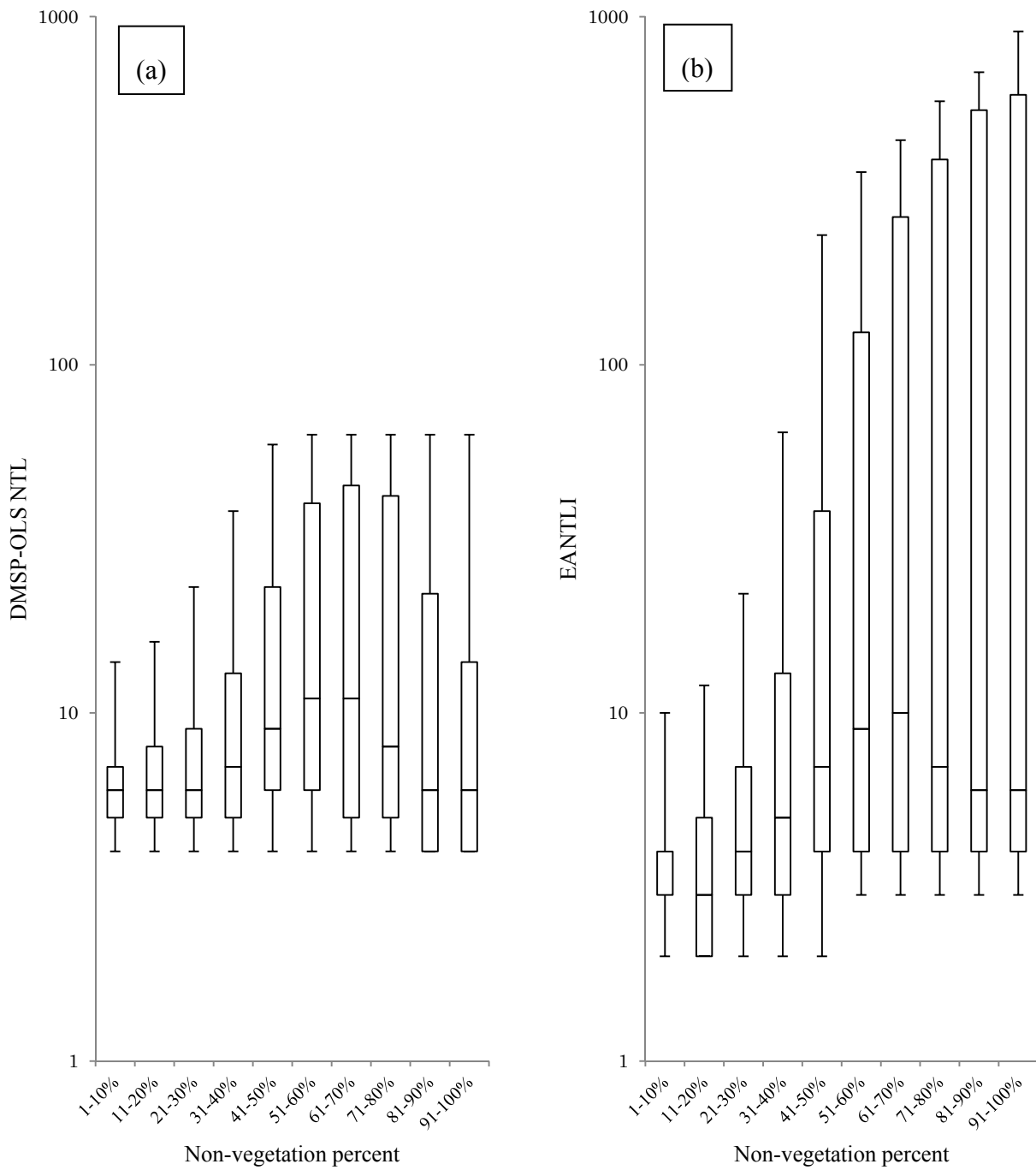non-vegetation fraction of less than 50% (which occurred at the 100th percentile line).

Fig. 5. Statistics of DMSP-OLS NTL values (a) and EANTLI values (b) for 10 non-vegetation fraction groups ranging from 1% to 100% divided into deciles. For each group, the 5%, 25%, 50%, 75%, 95% percentiles are shown.
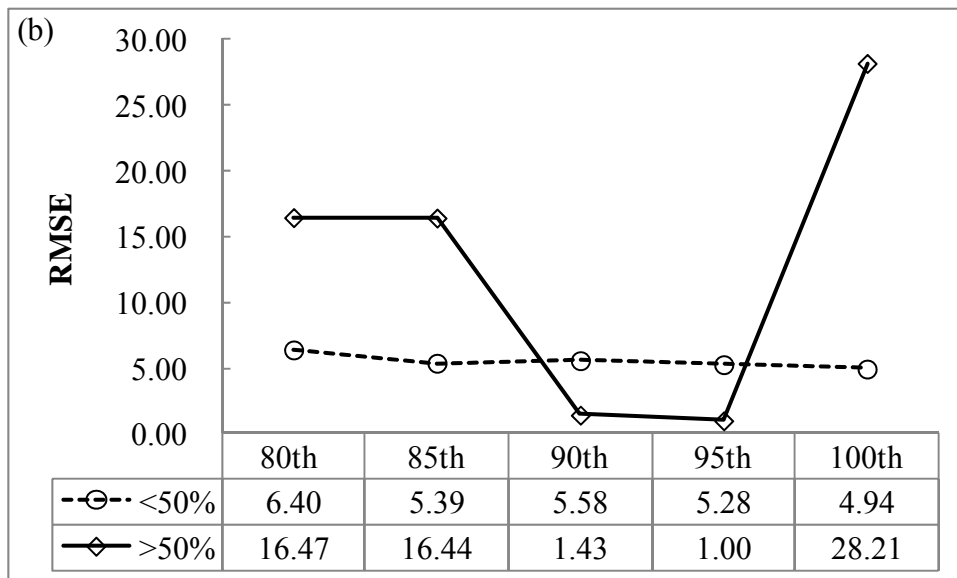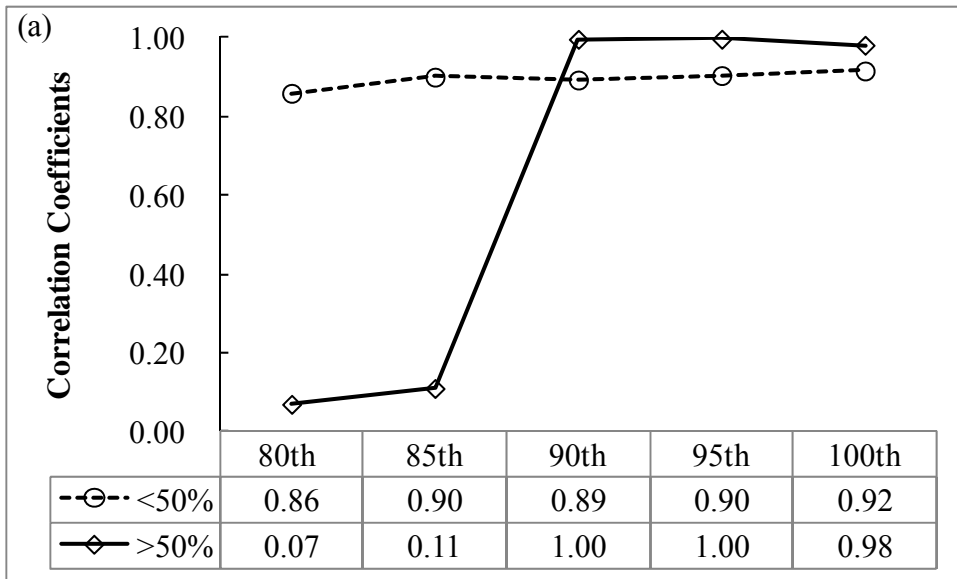
Fig. 6. Sensitivity of relationship between ISA% and EANTLI values to the use of EANTLI values at different percentile lines. (a) Correlation coefficients between ISA% and EANTLI values at each tested percentile line. (b) Root mean square error (RMSE) of estimated ISA% using the relationship at each tested percentile line. <50%: the case for ISA% less than 50%; >50%: the case for ISA% larger than 50%.

Based on the above considerations, the scatter plots of the EANTLI values at the 95th percentile line and the corresponding ISA% values for each group are shown in Fig. 7. It was found

that the EANTLI values increased slowly with the ISA% values when the EANTLI values were less than 64 (corresponding to an ISA% value of 40%), but there appears to be a breakpoint at the EANTLI value of 236 (corresponding to an ISA% value of 50%) where the curve trend changed (i.e., above 236, the EANTLI values increased dramatically with the increase of ISA% values). This finding indicates that there exist two relationships between the EANTLI and ISA% values. The first relationship can be represented by a natural logarithmic function for the EANTLI range between 0 and 236, and the second relationship can be represented by a polynomial function for EANTLI values larger than 236. The equations for the two relationships can be expressed as:

$$ISA\% = 8.5651 \cdot \ln(EANTLI) + 1.0063 \qquad (EANTLI<236) \qquad (4)$$

$$ISA\% = -0.00005 \cdot EANTLI^2 + 0.1329 \cdot EANTLI + 19.224 \qquad (EANTLI \geq 236) \qquad (5)$$

The constant term in Eq. (5) was slightly adjusted from 20.464 to 19.224 to allow the two equations to give the same ISA% values at the breakpoint of EANTLI=236.

A preliminary ISA% map was then generated from the EANTLI image using Equations (4) and (5). The preliminary ISA% map was further compared to the non-vegetation fraction map, and the lower value was selected as the final ISA% value for each pixel. This is because non-vegetation land cover contains both ISA and bare land, and thus the ISA% should be equal to or smaller than the fraction of non-vegetation estimated using the TMA method.
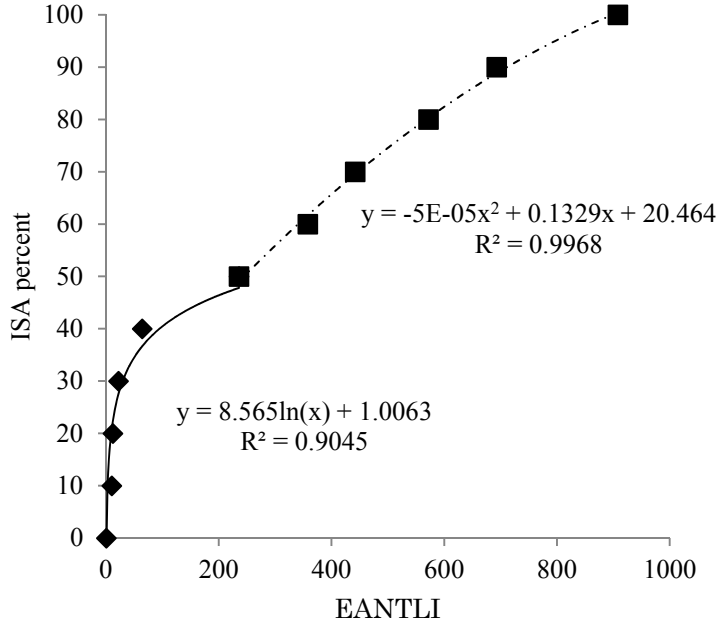
19

Fig. 7. Two relationships between the EANTLI and ISA% values.

4.4. Accuracy assessment

Two indices, namely the root mean square error (RMSE) and systematic error (SE), were used in accuracy assessment. These indices are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}\left(\text{ISA}_{\text{est,i}} - \text{ISA}_{\text{ref,i}}\right)^2} \qquad (6)$$

$$\text{SE} = \frac{1}{n}\sum_{i=1}^{n}\left(\text{ISA}_{\text{est,i}} - \text{ISA}_{\text{ref,i}}\right) \qquad (7)$$

where $\text{ISA}_{\text{est,i}}$ is the estimated fraction of the ISA from the MODIS and DMSP-OLS data, $\text{ISA}_{\text{ref,i}}$ is the referenced data in the sampling window i, and n is the total number of sampling windows for validation (a total of 97 samples were collected in this study). The RMSE denotes the sample standard deviation of the differences between the estimated values and the reference/observed values, while the SE denotes the average bias in the estimation. The coefficient of determination

20

(R^2) between $ISA_{est,i}$ and $ISA_{ref,i}$ was also calculated.

5. Performance of the proposed method

Figure 8 shows the fraction map of non-vegetation based on the TMA method (Fig. 8a), the preliminary fraction map of ISA obtained using Equations (4) and (5) (Fig. 8b), and the final fraction map of ISA obtained through selection of the smaller fraction values between the above two maps (Fig. 8c). The EANTLI map is shown in Fig. 8d for reference. From Fig. 8a, it can be clearly seen that bare lands made quite a large contribution to the non-vegetation fraction, especially for the areas without any NTL (also see Fig. 8d). The effects of bare lands on ISA estimation were greatly reduced through the combination of the non-vegetation fraction map and the EANTLI image (Fig. 8b). These improvements occurred not only in areas without NTL (e.g., areas along the Mekong River denoted as Area 1 in Fig. 8a; also see Figs. 9a-9c for a comparison of the details), but also in areas with NTL (e.g., the areas surrounding Bangkok, Hanoi, Ho Chi Minh City, Vientiane, and Phnom Penh; also see Figs. 9d-9f for an enlargement of Area 2 in Fig. 8a showing Phnom Penh city and its surrounding areas). Some slight changes were found when the final ISA fraction map (Fig. 8c) was compared to the preliminary ISA fraction map (Fig. 8b), such as in the areas in or surrounding Bangkok and Hanoi (also see Figs. 9g-9i for an enlargement of Area 3 in Fig. 8a and the Google Earth image in Fig. 9j for three special cases in Bangkok; the colors changed from yellow or red to green for these three cases).

To quantitatively evaluate the performance of the proposed method, 97 samples (3 km × 3 km for each sample) from the historical images in Google Earth were interpreted and compared with the estimated ISA% values (Fig. 10a). The corresponding 97 values for the non-vegetation fraction (TMA-based results, Yang et al., 2012) were also compared with the Google-Earth-image-based

reference values for ISA% (Fig. 10b). The RMSE of the proposed method was 0.111, with slight overestimation (SE=0.061). In addition, 87% of the variance in the ISA% estimations determined by the proposed method corresponded with the estimations from Google Earth images (R^2=0.87). In contrast, the TMA-based results showed larger scatter (RMSE=0.263) and overestimation (SE=0.203), especially in low ISA% areas (i.e., less developed areas). The coefficient of determination was 0.58, which was also lower than that of the proposed method.

Figure 11 provides a further demonstration of the proposed method using a selected sample from Fig. 10. The Google Earth image was collected on February 12, 2001 from a rural area near Bangkok, Thailand (Fig. 11a). From the Google Earth image, it can be seen that the major land use type of this sample is cropland with several small areas of ISA (e.g., some houses and roads, ISA%= 4.44% by visual interpretation). However, the TMA-based result showed a larger fraction of non-vegetation due to the presence of large amounts of bare land (Fig. 11b, N-V%=40.78%). Although the overestimation of ISA% from the TMA-based method was largely improved by the proposed method, a slight underestimation resulted because 5 of the 9 pixels had no lights but contained a small portion of ISA (gray pixels in Fig. 11c, ISA%=3.58%).
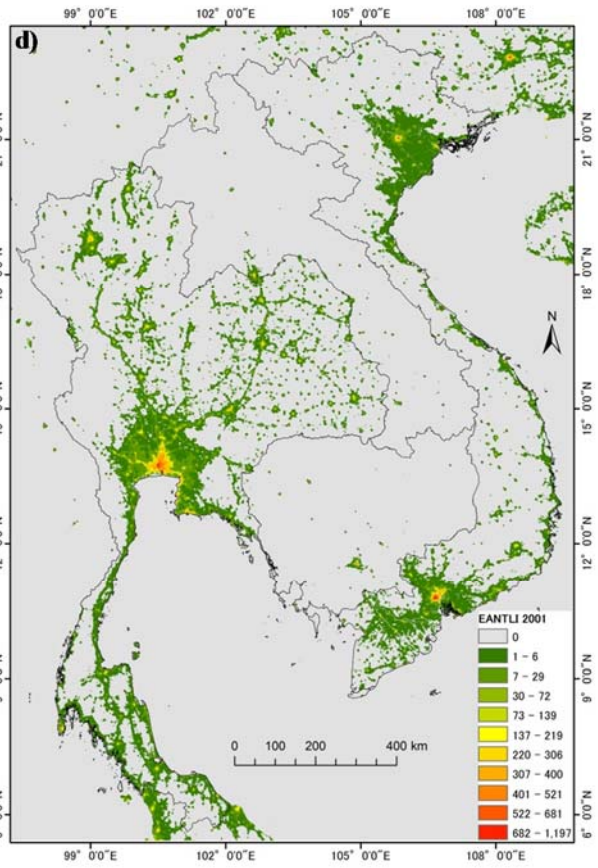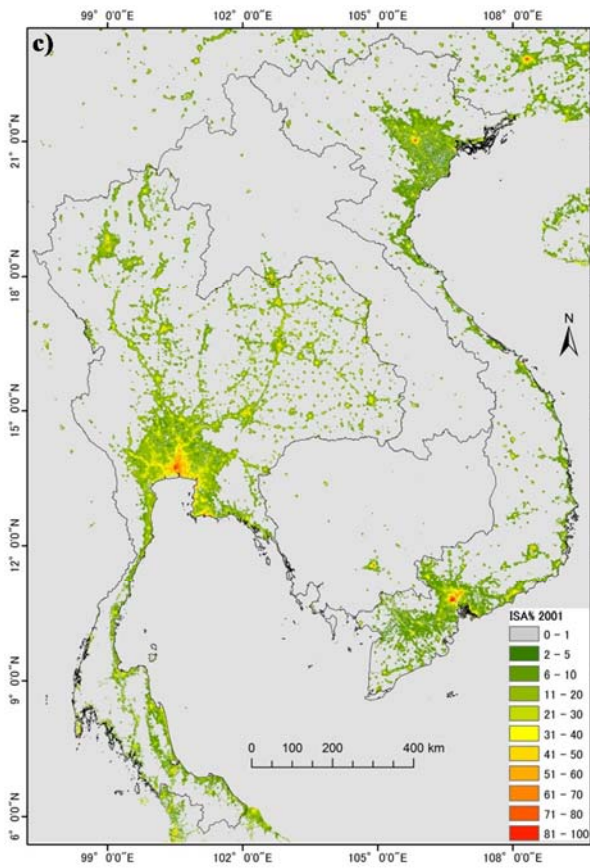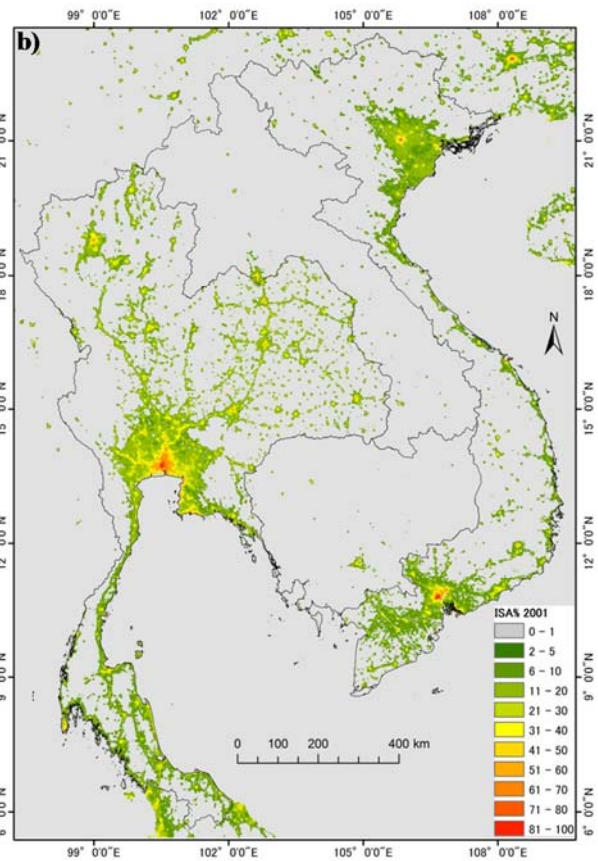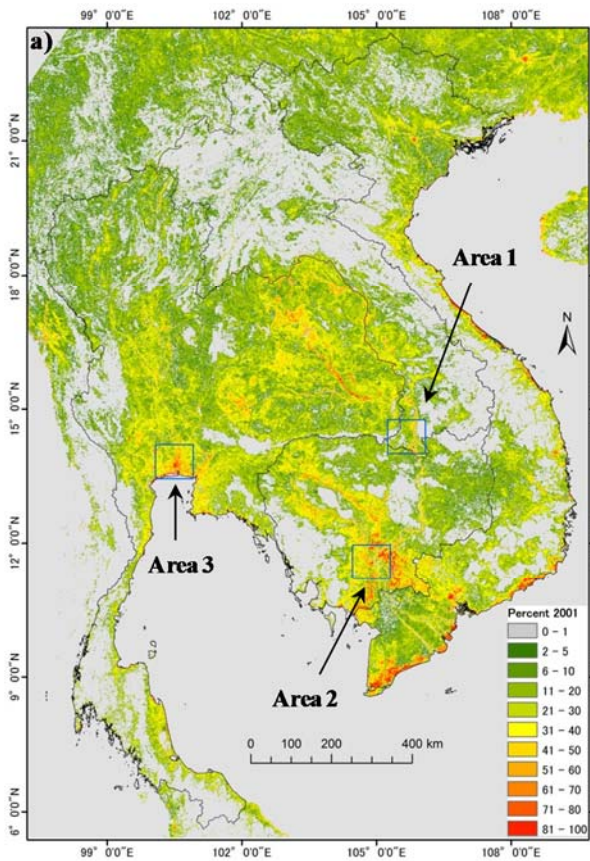
Fig. 8. Distribution maps of (a) non-vegetation fraction based on temporal mixture analysis method; (b) preliminary ISA fraction estimated using Equations (4) and (5); (c) final ISA fraction obtained through the selection of the smaller fraction values between (a) and (b); (d) EANTLI calculated using Equation (3). Three example areas (Area 1 to Area 3) were enlarged in Fig. 8 for comparison of the details.

Non-vegetation (TMA)   Preliminary ISA   EANTLI

Area 1

(a)   (b)   (c)

Non-vegetation (TMA)   Preliminary ISA   EANTLI

Area 2

(d)   (e)   (f)

Preliminary ISA   Final ISA   EANTLI

Area 3

(g)   (h)   (i)

20 km   N

Google Earth Image

(j)

ISA% 2001   EANTLI 2001

0 − 1   0
2 − 5   1 − 6
6 − 10   7 − 29
11 − 20   30 − 72
21 − 30   73 − 139
31 − 40   137 − 219
41 − 50   220 − 306
51 − 60   307 − 400
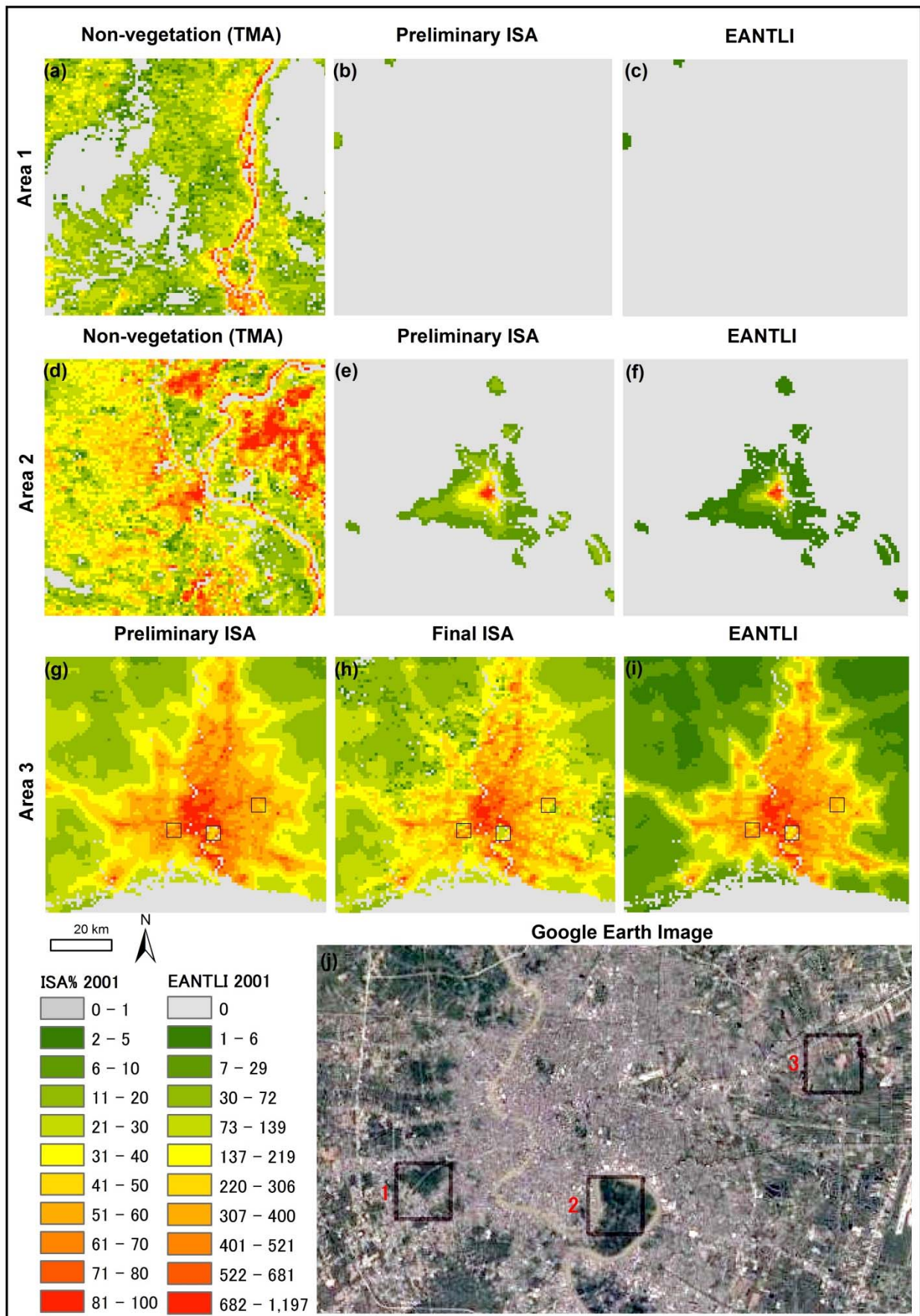61 − 70   401 − 521
71 − 80   522 − 681
81 − 100   682 − 1,197

Fig. 9. Comparison of non-vegetation fraction map, preliminary ISA fraction map, final ISA fraction map, and EANTLI map for three example areas shown in Fig. 8a. One Google Earth image was also shown for further visual comparisons of three small areas (black boxes).
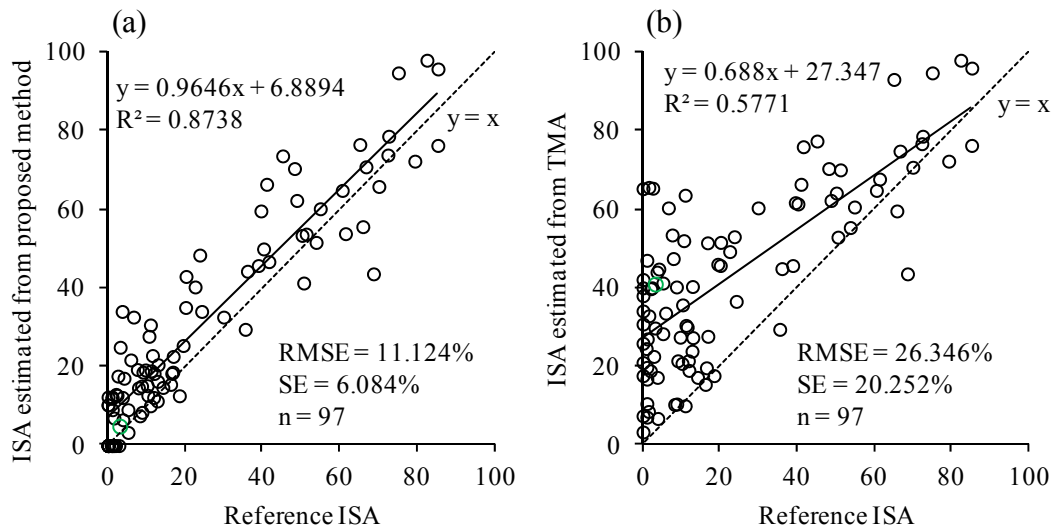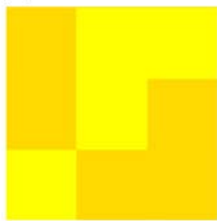

Fig. 10. Accuracy assessment. (a) Estimated ISA% by the proposed method are assessed by interpretation of the corresponding Google Earth images; (b) Estimated non-vegetation fraction based on TMA method (Yang et al., 2012) are assessed by interpretation of the corresponding Google Earth images.

(a) ISA %= 4.44%

(b) N-V%= 40.78% (c) ISA %= 3.58%

Perecnt 2001		
0 – 1		41 – 50
2 – 5		51 – 60
6 – 10		61 – 70
11 – 20		71 – 80
21 – 30		81 – 100
31 – 40		

Fig. 11. Distribution maps for a sample selected from Fig. 10 for visual comparison. (a) Reference image from Google Earth; (b) non-vegetation fraction (N-V%) estimated using the TMA method; (c) ISA% estimated using the method proposed in this study. The area of each map is 3 km by 3 km.

6. Discussion

Building the relationship between ISA% and NTL is a key element in the proposed method. To do this, reference datasets of the ISA% and corresponding NTL values are necessary. Landsat

TM/ETM+ data have generally been used to generate a reference dataset of ISA% in previous studies (Elvidge et al., 2007; Lu et al., 2008; Liu et al., 2015; Huang et al., 2016). However, there are some drawbacks to the use of Landsat data for this purpose. First, the Landsat data still suffer from the mixed pixel problem in the ISA% estimation. Even though an SMA-based method was used to estimate ISA% from Landsat data, it remains a challenge to efficiently reduce endmember variability in the SMA-based method to provide a more accurate ISA% estimation due to the spectral complexity of each endmember (Yang et al., 2010; Somers et al., 2011). Second, additional time and labor are needed for processing Landsat data. This drawback will limit the application of this method for estimating ISA% on a large scale (e.g., regional or global). Furthermore, since Landsat data processing is time-consuming, it is not easy to frequently update a new regional or global ISA% product.

In this study, ISA% reference data were obtained by statistically analyzing the non-vegetation fraction and corresponding EANTLI data to avoid the need for Landsat data (Fig. 5). The non-vegetation fractions were estimated from MODIS NDVI time-series data using the TMA method. The term "non-vegetation" was used in place of the term "ISA" in this step, because ISA and bare land have similar NDVI temporal profiles and thus can be considered the same endmember (Yang et al., 2014). It can be assumed that, in each non-vegetation fraction group, some pixels with the maximum EANTLI value only include ISA as the non-vegetation land cover (i.e., without bare land effects). According to this assumption, the non-vegetation fractions for these pixels can be considered ISA% reference data. The validation results shown in Fig. 10 indicate that this assumption was reasonable in our study area. In addition, there are two advantages to using the TMA method instead of the SMA method for estimating the non-vegetation fraction: (1) the TMA method provides an efficient reduction of endmember variability; and (2) this method is suitable for application on a large scale over a long term because it can be applied to MODIS time-series data

(Yang et al., 2012).

For NTL, it is well known that two large obstacles (i.e., the light saturation problem and the blooming effects) have emerged in the application of the original DMSP-OLS data (Lu et al., 2008; Zhang et al., 2013; Zhuo et al., 2015). From Fig. 5a, it can also be seen that the highest light intensities do not vary when the non-vegetation fraction is larger than 50%, and the higher light intensities are still found even when the non-vegetation fraction is smaller than 20%. These two problems can be solved through the use of the EANTLI image instead of the original DMSP-OLS image (Fig. 5b). As mentioned in section 4.3, the EANTLI values at the 95th percentile line were used in this study instead of the maximum EANTLI values. This led to 5% of the pixels in each non-vegetation fraction group with EANTLI values being larger than the EANTLI values used for building the relationships (i.e., Eq. (4) and Eq. (5)), and thus the ISA% values of these pixels in the preliminary ISA fraction map will be larger than the values of the corresponding non-vegetation fraction. In addition, imperfectly removed blooming effects can also cause overestimations in the preliminary ISA% map. These overestimations can be mitigated through the addition of the fourth step (e.g., Area 3 in Figs. 8a-8c; also Figs. 9g-9j).

Lu et al. (2008) reported a natural-logarithmic-function-based regression model for estimating ISA% (i.e., referred to as "fractional settlements" in their paper) from the HSI values. However, the natural logarithmic function was only found to be suitable for the range of ISA% values between 0% and 50% in our study area (Fig. 7). For ISA% values larger than 50%, the light intensities increased quickly with the increase of ISA%, and thus a quadratic-polynomial-function-based model was used to replace the natural-logarithmic-function-based regression model (Eq. 5). Similar relationships between the ISA% and EANTLI values were also found by analyzing MODIS and DMSP-OLS data for 2006 and 2012 using the same method (results not shown). These results probably indicate that different light use efficiencies existed between the developed and

29

undeveloped regions in our study area.

Since global MODIS NDVI/EVI products are available for the year 2000 up to the present, a yearly non-vegetation fraction map can be generated on a global scale for the same period based on the TMA method. On the other hand, yearly DMSP-OLS datasets are available from 1992 to 2013, and new NTL data (Suomi NPP-VIIRS: Suomi National Polar-Orbiting Partnership – Visible Infrared Imaging Radiometer Suite) have been collected since 2011. Therefore, it is possible to generate a yearly distribution map of ISA% values from 2000 to the present on a global scale using the proposed method. However, the size of the area that is the most appropriate for building the best relationship between ISA% and NTL should be further studied, for two reasons. First, the existence of a sufficient number of pixels with only ISA as a non-vegetation land cover is required for each of the 10 non-vegetation fraction groups (defined by deciles from 1% to 100%). Therefore, a larger area is necessary to meet this requirement. Second, the light use efficiency probably differs among countries or regions of the world due to different levels of economic development. Therefore, it is better to use different relationships to obtain more accurate ISA% estimations for each country or region, and thus a smaller study area should be selected. These two considerations are contradictory. In addition, although the MODIS NDVI/EVI products are only available from year 2000, the NDVI time-series data can be extended to the 1980s by using the data from the Advanced Very High Resolution Radiometer of the NOAA (NOAA/AVHRR). Therefore, it is possible to obtain a global ISA% map since 1992 if the NDVI is used to replace EVI in Equation (3). However, how the replacement influences the performance of NTL data should be further investigated in a future study. Moreover, inter-annual correction should also be implemented to reduce temporal inconsistency between spatially enhanced DMSP-OLS time-series data (i.e., EANTLI in this study, Xie and Weng, 2017).

7. Conclusions

In this study, an easily implemented method was developed for estimating ISA% on a large scale. The developed method includes four major steps and requires three major products from two satellite sensors. The four major steps are: (1) estimate the non-vegetation fraction from MODIS NDVI time-series products using the TMA method; (2) obtain improved NTL data by calculating EANTLI from the monthly MODIS EVI product and DMSP-OLS NTL data; (3) generate a preliminary ISA% map by building a relationship between ISA% and EANTLI based on the statistical analysis of the maps from the first and second steps; (4) obtain a final ISA% map by comparing the preliminary ISA% map to the non-vegetation fraction map and selecting the smaller values. The validation results showed that the developed method has promising accuracy for estimating the ISA% in our study area (which mainly consists of four Southeast Asian countries: Thailand, Laos, Cambodia, and Vietnam), with an RMSE value of 0.111, an SE value of 0.061, and a determination coefficient of 0.87. Another important finding is that there are two relationships between ISA% and the improved NTL (i.e., EANTLI): the natural logarithmic function is suitable for ISA% values between 0% and 50%, and the quadratic polynomial function should be used for ISA% values larger than 50%. The developed method shows high potential for use in generating a global ISA% map with frequent updates because of its easy implementation and readily available input data.

References

Arnold, C.L., Gibbons, C.J., 1996. Impervious Surface Coverage: The Emergence of a Key Environmental Indicator. Journal of the American Planning Association 62 (2), 243-258.

Bauer, M.E., Loffelholz, B.C., Wilson, B., 2008. Estimating and mapping impervious surface area by regression analysis of Landsat imagery. In Remote Sensing of Impervious Surfaces; Weng, Q., Ed.; CRC Press: Boca Raton, FL, USA, pp. 3–19.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. Remote Sensing of Environment 91 (3–4), 332-344.

Elvidge, D.C., Tuttle, T.B., Sutton, C.P., Baugh, E.K., Howard, T.A., Milesi, C., et al., 2007. Global Distribution and Density of Constructed Impervious Surfaces. Sensors 7, 1962-1979.

Gallo, K., Xian, G., 2016. Changes in satellite-derived impervious surface area at US historical climatology network stations. ISPRS Journal of Photogrammetry and Remote Sensing 120, 77-83.

Huang, X., Schneider, A., Friedl, M.A., 2016. Mapping sub-pixel urban expansion in China using MODIS and DMSP/OLS nighttime lights. Remote Sensing of Environment 175, 92-108.

Knight, J., Voth, M., 2011. Mapping Impervious Cover Using Multi-Temporal MODIS NDVI Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 4 (2), 303-309.

Lee, S., Lathrop, R.G., 2006. Subpixel analysis of Landsat ETM+ using Self-Organizing Map (SOM) neural networks for urban land cover characterization. IEEE Transactions on Geoscience and Remote Sensing 44 (6), 1642-1654.

Li, W., Wu, C., 2015. Incorporating land use land cover probability information into endmember class selections for temporal mixture analysis. ISPRS Journal of Photogrammetry and Remote Sensing 101, 163-173.

Liu, X., Hu, G., Ai, B., Li, X., Shi, Q., 2015. A Normalized Urban Areas Composite Index (NUACI) Based on Combination of DMSP-OLS and MODIS for Mapping Impervious Surface Area. Remote Sensing 7 (12), 17168-17189..

Lu, D., Tian, H., Zhou, G., Ge, H., 2008. Regional mapping of human settlements in southeastern China with multisensor remotely sensed data. Remote Sensing of Environment 112 (9), 3668-3679.

Lu, D., Weng, Q., 2006. Use of impervious surface in urban land-use classification. Remote Sensing of Environment 102 (1–2), 146-160.

MRC (Mekong River Commission) (2011). Planning Atlas of the Lower Mekong River Basin. Online accessed:

http://www.mrcmekong.org/assets/Publications/basin-reports/BDP-Atlas-Final-2011.pdf.

Powell, R.L., Roberts, D.A., Dennison, P.E., Hess, L.L., 2007. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. Remote Sensing of Environment 106 (2), 253-267.

Somers, B., Asner, G.P., Tits, L., Coppin, P., 2011. Endmember variability in Spectral Mixture Analysis: A review. Remote Sensing of Environment 115 (7), 1603-1616.

Weng, Q., 2012. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. Remote Sensing of Environment 117, 34-49.

Weng, Q., Hu, X., 2008. Medium spatial resolution satellite imagery for estimating and mapping urban impervious surfaces using LSMA and ANN. IEEE Transactions on Geoscience and Remote Sensing, 46 (8), 2397-2406.

Wu, C., 2004. Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery. Remote Sensing of Environment 93 (4), 480-492.

Wu, C., Murray, A.T., 2003. Estimating impervious surface distribution by spectral mixture analysis. Remote Sensing of Environment 84 (4), 493-505.

Xian, G., Homer, C., 2010. Updating the 2001 National Land Cover Database Impervious Surface Products to 2006 using Landsat Imagery Change Detection Methods. Remote Sensing of Environment 114 (8), 1676-1686.

Xie, Y., Weng, Q., 2017. Spatiotemporally enhancing time-series DMSP/OLS nighttime light imagery for assessing large-scale urban dynamics. ISPRS Journal of Photogrammetry and Remote Sensing 128, 1-15.

Yang, F., Matsushita, B., Fukushima, T., 2010. A pre-screened and normalized multiple endmember spectral mixture analysis for mapping impervious surface area in Lake Kasumigaura Basin, Japan. ISPRS Journal of Photogrammetry and Remote Sensing 65 (5), 479-490.

Yang, F., Matsushita, B., Fukushima, T., Yang, W., 2012. Temporal mixture analysis for estimating impervious surface area from multi-temporal MODIS NDVI data in Japan. ISPRS Journal of Photogrammetry and Remote Sensing 72, 90-98.

Yang, F., Matsushita, B., Yang, W., Fukushima, T., 2014. Mapping the human footprint from satellite measurements in Japan. ISPRS Journal of Photogrammetry and Remote Sensing 88, 80-90.

Zhang, Q., Schaaf, C., Seto, K.C., 2013. The Vegetation Adjusted NTL Urban Index: A new approach to reduce saturation and increase variation in nighttime luminosity. Remote Sensing of Environment 129, 32-41.

Zhuo, L., Zheng, J., Zhang, X., Li, J., Liu, L., 2015. An improved method of night-time light saturation reduction based on EVI. International Journal of Remote Sensing 36 (16), 4114-4130.