Journal of
Applied Remote Sensing

Timely and accurate national-scale mapping of urban land in China using Defense Meteorological Satellite Program's Operational Linescan System nighttime stable light data

Yang Yang
Chunyang He
Qiaofeng Zhang
Lijian Han
Shiqiang Du

SPIE

Timely and accurate national-scale mapping of urban land in China using Defense Meteorological Satellite Program's Operational Linescan System nighttime stable light data

Yang Yang,[a,b] Chunyang He,[a] Qiaofeng Zhang,[c] Lijian Han,[d] and Shiqiang Du[a]

[a]Beijing Normal University, State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing 100875, China
hcy@bnu.edu.cn

[b]Ocean University of China, School of Law & Political Science, Qingdao 266100, China

[c]Murray State University, Department of Geosciences, Murray, Kentucky 42071

[d]Chinese Academy of Sciences, Research Center for Eco-Environmental Sciences, Beijing 100085, China

Abstract. Urban land accounts for a small fraction of the Earth's surface area but rapid increases in urban land have a disproportionate influence on the environment. China is a living laboratory in urbanization and has witnessed fast urban growth in recent decades. The timely and accurate mapping of urban land in China is an urgent and basic issue toward clarifying the urbanization process and revealing its environmental impacts. Nighttime stable light (NSL) data obtained by the Defense Meteorological Satellite Program's Operational Linescan System (DMSP/OLS) can provide an economical way to map urban land nationwide. However, it is difficult to apply existing methods to accurately extract urban land from DMSP/OLS NSL data covering the entirety of China due to China's large area and substantial regional variation. A stratified support vector machine (SSVM)-based method used to map the urban land in China in 2008 at a national scale using DMSP/OLS NSL and SPOT normalized difference vegetation index data is presented. The results show that measurement of urban land in China in 2008 using SSVM achieves an average overall accuracy of 90% and an average Kappa of 0.69. The success of this research demonstrates the great potential of SSVM for clarifying the urbanization process in continental and global research. [DOI: 10.1117/1.JRS.7.073535]

Keywords: urban land; Defense Meteorological Satellite Program's Operational Linescan System; nighttime stable light; stratified support vector machine; normalized difference vegetation index; national scale; China.

Paper 12213 received Jul. 18, 2012; revised manuscript received Jun. 9, 2013; accepted for publication Jun. 12, 2013; published online Jul. 15, 2013.

1 Introduction

Urban land accounts for a small fraction of the Earth's surface area. However, a rapid increase in urban land exerts a disproportionate influence on the environment in terms of energy and resource fluxes, biogeochemical cycles, and ecological and climatic changes at local to global scales.[1–4] As a living laboratory in urbanization, China has witnessed an unprecedented urban expansion process over the last three decades, which has greatly influenced food security and the ecological health of the nation.[5,6] The timely and accurate mapping of urban land within China is of great importance to clarify the urbanization process and to help ensure more sustainable future urbanization.[7,8]

Remote sensing technology in recent years has proved to be of great importance in mapping urban land due to its ability to allow quick, periodic revisits and large-area coverage.[9] Satellite

0091-3286/2013/$25.00 © 2013 SPIE

images have been used in numerous studies to map urban land.[10–14] Images with high spatial resolution are required to satisfy several important and detailed urban information requirements in individual cities,[15] but they fall short in examining urban patterns at the national scale.[16] Moderate resolution Landsat imageries (about 30 m) are widely employed for land cover mapping at continental scales, such as in the National Land-use/Land-Cover datasets (NLCD) developed by the Chinese Academy of Sciences, which can accurately represent the land cover condition in China.[16–21] However, due to the limitations of Landsat images, they require a large amount of human and computational resources to map urban land for the whole nation.[17,18,21] For example, to develop the NLCD1995, eight research institutions and about 100 scientists were organized, and scenes from a total of 520 Landsat thematic mapper (TM) images were used.[18] Therefore, it is difficult to map urban land at a large spatial scale quickly using Landsat images with their limited geographic coverage.

The Defense Meteorological Satellite Program's Operational Linescan System (DMSP/OLS) nighttime stable lights (NSLs) data (city lights) are unique among remote sensing data sources, as it offers a uniquely "human" view of the Earth's surface. It is often used to map urban land and has been demonstrated to be a unique and valuable resource for mapping urban land in a timely and economical manner at national, continental, and global scales due to its appropriate spatial and temporal resolution.[22–30] The DMSP/OLS instrument has a low-light imaging capability that can detect lighting at the Earth's surface at night. The OLS sensor offers a wide "city lights" view of the Earth's surface, scanning a swath as wide as 3000 km of land surface in one pass.[31] The DMSP/OLS nighttime light data are suitable for providing urban information at a large scale due to their appropriate resolution. Numerous studies on mapping urban land at a national scale have used the DMSP/OLS nighttime light data. For example, Imhoff et al. mapped urban land in the continental United States using the DMSP/OLS nighttime light data in 1994/1995[23]; He et al. also mapped urban land in mainland China in 1992, 1996, and 1998 using the DMSP/OLS nighttime light data.[7]

At present, the methods for mapping urban land using DMSP/OLS nighttime light data can be generally grouped into two broad types: those based on threshold techniques and those based on classification. Threshold approaches often extract urban land data from DMSP/OLS nighttime light data by isolating pixels that have a brightness value above a threshold. Generally, the threshold can be determined empirically according to related background knowledge,[32] supported by mutation detection using the DMSP/OLS dataset alone,[23] or predefined with the help of ancillary materials by comparing the results against reference information.[7,33,34] The threshold techniques have been widely used to extract urban land from DMSP/OLS nighttime light data in previous studies. However, threshold-based approaches tend to exaggerate the spatial extent of metropolitan areas (the "blooming effect"); conversely, they may omit large numbers of small cities with low light brightness. Additionally, the 6-bit DMSP/OLS data [digital number (DN) ranges between 0 and 63] often resulted in data saturation in the urban landscape.[28] Hence, the accuracy and reliability of results extracted using thresholds are usually compromised when applied to a large area, so the use of the threshold method is limited in multiscale applications. Different remotely sensed data have their own characteristics, and using them in combination could provide more information than their individual use.[35] The previous research has identified a strong negative correlation between the vegetation index, or vegetation abundance, and urban land.[36,37] More recently, image classification principles have been adopted to distinguish each lit pixel as urban or nonurban with the help of auxiliary data, such as the normalized difference vegetation index (NDVI) data.[35,38] The complementary characteristics between DMSP/OLS and NDVI data made the combined new datasets more informative than the DMSP/OLS data alone.[35] This classification method can overcome the inherent problems of the threshold method by extracting urban land at the pixel level.[38] It has been successfully applied to map urban land at local and regional scales.[35,38]

The support vector machine (SVM) method has been widely used in recent studies of image classification.[39] The SVM is a nonparametric method based on statistical learning theory.[40] The classification by SVM is based on fitting an optimal-separating hyperplane between classes by focusing on the training samples that lie at the edge of the class distributions and between the class centroids, the support vectors.[41,42] Comparative studies have shown that SVM can provide higher accuracy than other widely used pattern recognition techniques, such as neural networks and decision trees, as well as conventional probabilistic classifiers, such as the maximum likelihood classification.[41,43–45]

Cao et al.[38] proposed an original support vector machine (OSVM)-based method of image classification and successfully applied it to part of eastern China using DMSP/OLS nighttime light and satellite pour l'observation de la Terre (SPOT) NDVI data. In the whole study area, their OSVM approach involved using several simple criteria to build one initial training set of urban and nonurban pixels, and they designed one iterative classification procedure that was capable of running continuous training for the SVM classifier to deal with the transitional features of urban land. The OSVM method effectively improved the accuracy of urban land extraction, especially in the transitional zone between urban and rural areas. However, the OSVM method had three apparent drawbacks when applied at the national scale: (1) grouping the entire procedure together under one head, so that the heterogeneity among regions in China was ignored; (2) inflexibility in selecting training samples, especially using only one subjective threshold to choose the urban training samples in the whole nation; and (3) unitary postclassification processing in the whole nation, which relied on only a single empirical NDVI threshold to eliminate pseudourban pixels.[38] The first drawback meant that the regional variation within the nation was ignored because a single initial training set and a single iterative classification procedure were inadequate to manage all subregions of this large area. The second drawback meant that urban training samples were often missed in smaller cities with lower light brightness and exaggerated in larger cities with higher light brightness. The third limitation could lead to the inclusion of some pseudourban pixels in areas of lower NDVI and the exclusion of some urban pixels in areas of higher NDVI. These drawbacks could result in significant error in urban land extraction and thereby limit the accuracy and reliability of the OSVM approach at the national scale.

Digital OLS data were not made available by the US Department of Defense and consequently were not archived by the National Oceanic and Atmospheric Administration's National Geophysical Data Center (NOAA-NGDC) from the time of declassification until 1992. Since digital archiving began, a number of products have been released.[46] Presently, there are four different types of imagery associated with the DMSP/OLS dataset: (1) frequency-detected DMSP/OLS data, which only record the frequency of nighttime lights but without intensity information;[23,47] (2) radiance-calibrated DMSP/OLS data, which need manual control on the gain of the sensor;[48] (3) nonradiance-calibrated, which only covered Asia in 1992, 1996, and 1998;[7,49,50] and (4) average DN DMSP/OLS data, which come in the form of average DN values and are the latest, most extensively released and widely used type of DMSP/OLS nighttime lights data.[30,35,38,51–54] The last type of data has four different versions. The newest, version 4 DMSP/OLS nighttime lights time-series dataset from 1992 to 2010, was recently published by NGDC. This dataset included three kinds of data: cloud-free coverage data, nighttime light data with no further filtering, and NSL data. The DMSP/OLS NSL data contain a DN that indicates the average nighttime light intensity observed within each year. A number of constraints were considered to identify the best quality nighttime light data to produce the DMSP/OLS NSL data.[30,51] Hence, the DMSP/OLS NSL data enable timely and reliable mapping of urban land at the national scale.

To effectively map urban land in China at the national scale using DMSP/OLS NSL data, this article will (1) propose a stratified support vector machine (SSVM)-based method that can effectively map national urban land from DMSP/OLS NSL data and (2) map the urban land in China in 2008 at the national scale using DMSP/OLS NSL and SPOT DNVI datasets acquired in 2008.

2 Materials and Preprocessing

The research study area was the whole of China. We used DMSP/OLS NSL annual composite data (Version 4) in 2008, which was collected by the US Air Force Weather Agency and processed by NOAA-NGDC. The data are available for downloading from http://mapserver.ngdc.noaa.gov/cgi-bin/public/ms/gcv4/dl (accessed 24 April 2013). The data are in 30-arc sec grids (the spatial resolution is approximately 1 km at the equator and 0.8 km at 40°N), spanning −180 to 180 degrees in longitude and −65 to 75 degrees in latitude.[55] The data values signify the light brightness levels and range from 1 to 63, whereas background and noise are recorded as zero. The data encompass lights from cities, towns, and other sites with persistent lighting. A number of constraints are used to exclude sunlight, glare, moonlight, observations of clouds with

lighting, and aurora features. Ephemeral events such as fires have been discarded. After obtaining the global DMSP/OLS NSL data for 2008, NSL data were extracted according to Chinese administrative boundaries. Finally, the data were projected using the Lambert Azimuthal Equal Area projection and resampled to a pixel size of 1 km to facilitate calculation [Fig. 1(a)].

The SPOT/VGT S10 dataset was also included in data collection. Data were radiometrically calibrated, precisely georeferenced and corrected for atmospheric effects when distributed. S10 data have a spatial resolution of 1 km, providing 10-day maximum value combined NDVI, and are available for download from: http://free.vgt.vito.be (accessed 24 April 2013). The global data from April to September 2008 (the growing season) were acquired and processed by maximum value composition to create a yearly maximal NDVI image. After extracting the data for China, the SPOT NDVI image was reprojected to Lambert Azimuthal Equal Area projection for matching with DMSP/OLS NSL data [Fig. 1(b)].

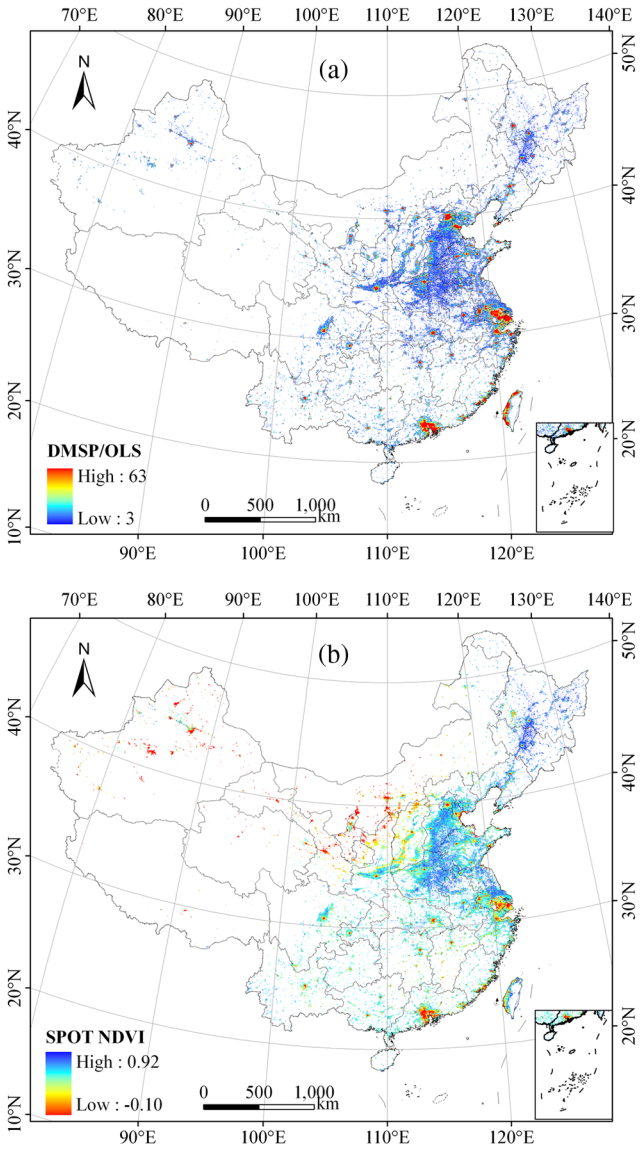

Fig. 1 Study area and data. (a) Defense Meteorological Satellite Program's Operational Linescan System (DMSP/OLS) nighttime stable light (NSL) data of China in 2008. (b) SPOT normalized difference vegetation index (NDVI) maximum value composition of China during April to September 2008. Note: the DMSP/OLS NSL data are digital values of lighted pixels. The white background pixels, recorded as zero in DMSP/OLS NSL data, were considered nonurban land. NDVI values show only within areas identified as lighted.

Eight good quality Landsat ETM+images of scenes in 2008 were obtained from the United States Geological Survey: http://earthexplorer.usgs.gov/ (accessed 24 April 2013) to validate the urban land mapped by DMSP/OLS NSL and SPOT NDVI data. They mainly covered eight cities: Harbin (path/row: 118/28, 8 September 2008), Beijing (path/row: 123/32, 11 September 2008), Taipei (path/row: 117/43, 26 April 2008), Nantong (path/row: 119/38, 18 November 2008), Xi'an (path/row: 127/36, 22 August 2008), Jingdezhen (path/row: 121/40, 15 October 2008), Chengdu (path/row: 129/39, 30 April 2008), and Urumchi (path/row: 143/29, 9 October 2008.

We also retrieved a point dataset of urban centers in China, derived from a map at 1:4,000,000 scale distributed by the National Geomatics Center of China: http://ngcc.sbsm.gov.cn (accessed 24 April 2013), as auxiliary data.

Urban population and gross domestic product (GDP) data for each province of China in 2008 were collected from the China Statistical Yearbook.[56] Urban land statistical data for major cities of China in 2008 were collected from the China City Statistical Yearbook.[57]

3 Methodology

Many studies have shown that the stratification approach is effective for obtaining accurate mapping of a large area with great regional variation.[58,59] For example, for unsupervised change detection in large sized multitemporal images, Bovolo and Bruzzone[58] developed a split-based approach, which splits the large size images into a set of user-defined nonoverlapping subimages. To facilitate processing and classification in mapping global urban areas, a global stratification of the Earth's land surface was developed in the work of Schneider et al.[59] Therefore, in our study, the idea of SSVM is to classify unknown pixels at different levels of urban development using a region-specific processing procedure based on stratification.

First, all of China were stratified into several smaller regions according to regional variation in level of urban development. Second, training samples representative of the status of each region were selected using a self-adaptation procedure according to the regional statistic feature of DMSP/OLS NSL and SPOT NDVI data. Third, a SVM-based iterative classification and training procedure were adopted to identify the urban pixels in each region separately. Finally, adaptive postclassification processes according to the regional statistic feature of DMSP/OLS NSL and SPOT NDVI data were carried out for each region to obtain more accurate mapping results of urban land for all of China (Fig. 2).

3.1 *Stratification of All of China According to Regional Discrepancy in Urban Development*

Stratification is the first step of SSVM. The principle of stratification is to split the whole country into subregions based on the "divide and conquer" strategy such that significant heterogeneity exists among regions and relative homogeneity within each region. Based on this stratification, training samples, image classification, and postclassification in SSVM were selected for each subregion separately. So the stratification was the basis on which SSVM reflected regional variations in urban development in China.

The Coordinated Regional Development Strategy and Policy Reports, published by the Development Research Center of the State Council of China, divide China into eight economic regions.[60,61] Analysis of socio-economic and remote sensing data revealed that these eight economic regions reasonably reflect the regional variation in urban development in China. The regions are Northeast China (NEC), Northern Coastal China (NCC), Southern Coastal China (SCC), Eastern Coastal China (ECC), Middle Reaches of the Yellow River (MRYLR), Middle Reaches of the Yangtze River (MRYTR), Southwest China (SWC), and Northwest China (NWC; Fig. 3).

Variations in economic development among the eight regions were apparent both in GDP and per capital GDP [Fig. 4(a), Table 1]. The GDP of SCC accounts for 24.85% of the nation's total and is the highest of all the regions, about 10 times that of NWC, which is the lowest of all the

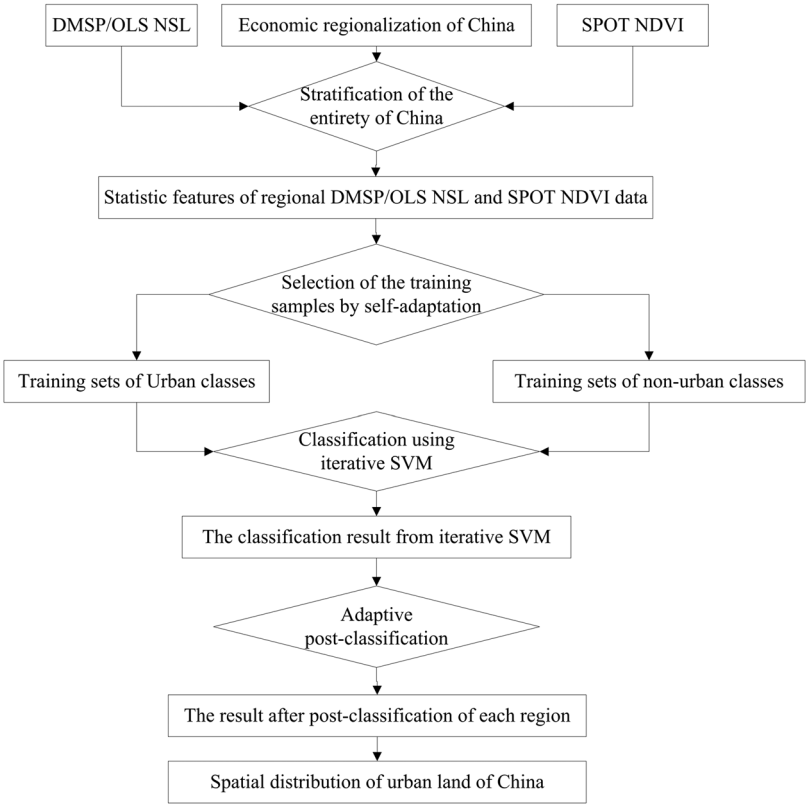

Fig. 2 Flow chart of stratified support vector machine (SSVM).

regions and only accounts for 2.65% of the national total. Additionally, per capital GDPs in SCC and SWC are 66.25% higher and 22.65% lower than the national average, respectively.

Meanwhile, DMSP/OLS NSL and SPOT NDVI data also reveal noticeable differences among the eight regions [Fig. 4(b), Table 1]. Since the Reform and Open Policy of the late 1970s, the government of China has given priority to development of coastal areas, so the level of development of coastal areas is far ahead of that in other regions. DMSP/OLS NSL

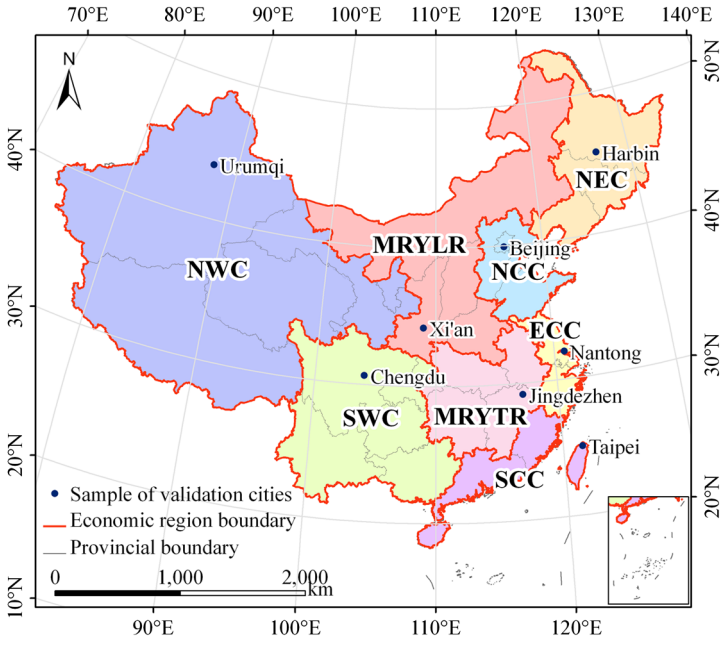

Fig. 3 The map of economic regions in China.

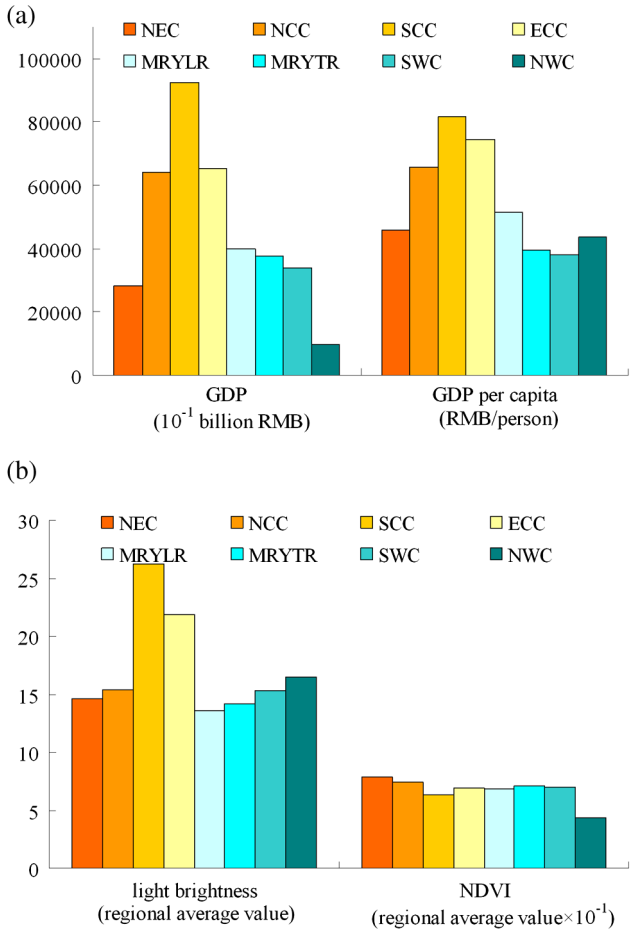

Fig. 4 Discrepancies among the eight economic regions of China. (a) Economic statistical data. (b) DMSP/OLS NSL and SPOT NDVI data. Note: 1 RMB = 0.1618 US dollar (the exchange rate was last updated on 24 April 2013).

Table 1 Discrepancies among the eight economic regions of China.

	GDP	GDP/capita	DMSP/OLS NSL		SPOT NDVI	
Economic regions	(10^8 RMB)	(RMB/person)	$\bar{x}_{ols}$	S_{ols}	$\bar{x}_{ndvi}$	S_{ndvi}
Northeast China	28,195.63	45,735.00	14.59	14.24	0.79	0.12
Northern Coastal China	64,103.08	65,691.50	15.38	14.11	0.74	0.12
Southern Coastal China	92,328.24	81,612.15	26.26	17.27	0.63	0.16
Eastern Coastal China	65,497.68	74,504.79	21.88	17.20	0.68	0.15
Middle Reaches of the Yellow River	39,959.63	51,443.06	13.55	12.23	0.68	0.15
Middle Reaches of the Yangtze River	37,841.52	39,520.92	14.15	12.52	0.71	0.12
Southwest China	33,807.99	37,970.00	15.34	11.65	0.70	0.10
Northwest China	9,835.47	43,529.79	16.50	13.52	0.43	0.23
China	371,569.24	49,089.59	16.64	14.65	0.69	0.16

Note: $\bar{x}_{ols}$ and S_{ols} are the arithmetic mean value and standard deviation of light brightness, and $\bar{x}_{ndvi}$ and S_{ndvi} are the arithmetic mean value and standard deviation of NDVI in each region, respectively. 1 RMB = 0.1618 U.S. dollar (the exchange rate was last updated on 24 April 2013).

data reveal that the average light brightness in the coastal areas of China, such as SCC and ECC, is much higher than other regions in inland China. Due to striking differences in local climate, soil properties, and the availability of water among the regions, regional variation in NDVI is also remarkable. SPOT NDVI data indicate that the average NDVI in NEC (0.79), which ranks first among all regions, is significantly higher than that in NWC (0.43), which ranks last.

3.2 *Selection of Training Samples Using a Self-Adaptation Approach*

This section presents a novel procedure of self-adaptation in SSVM for selecting training samples to overcome the inflexibility of OSVM, which selects training samples using somewhat arbitrary thresholds. The self-adaptation procedure for selecting samples that represent urban and nonurban lands was based on the statistical features of DMSP/OLS NSL and SPOT NDVI data in the corresponding regions. The flowchart is shown in Fig. 5. It involves four steps: (1) acquisition of the statistical features of NSL and NDVI data for each economic region; (2) determination of potential urban patches; (3) selection of the urban training samples; and (4) selection of the nonurban training samples.

3.2.1 *Acquisition of the statistical features of NSL and NDVI data for each economic region*

We selected the arithmetic mean and standard deviation of DMSP/OLS NSL and SPOT NDVI data in each economic region as the statistical features. The average value and standard deviation of light brightness are expressed as $\bar{x}_{\text{ols}}$ and S_{ols}, and the average value and standard deviation of NDVI are expressed as $\bar{x}_{\text{ndvi}}$ and S_{ndvi}. These were calculated from the digital values of all lit pixels in each economic region.

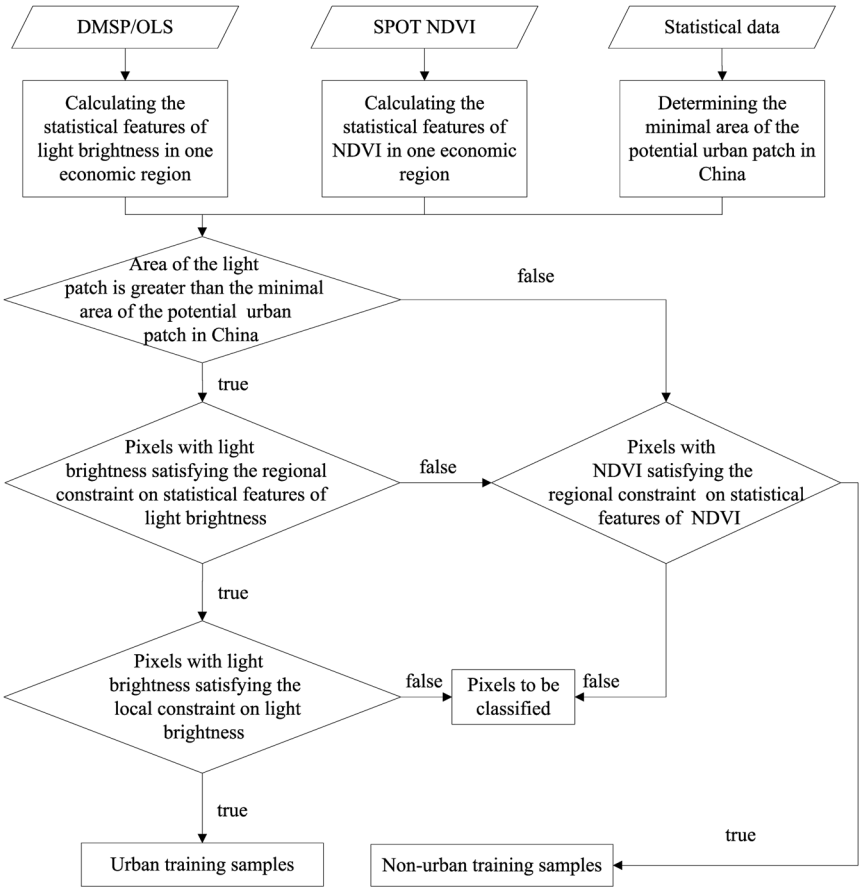

Fig. 5 Flowchart showing self-adaptive training sample selection.

3.2.2 *Determination of potential urban patches*

Potential urban patches were identified as clusters of lit pixels with a minimum area of 4 km^2 in the DMSP/OSL NSL data. According to the China City Statistical Yearbook, Ryonan city in Gansu province had the smallest area of urban land (8 km^2) among 287 major cities of China in 2008.[57] However, we selected a lower threshold of 4 km^2 to avoid missing smaller cities. It should be noted that 4 km^2 was set as the threshold value of the potential urban patches, which meant that only the clusters of lit pixels bigger than 4 km^2 were candidates for designation as urban land. However, this was not the final outcome of the process; instead, it simply eliminated the pixels that had no light and the light patches smaller than 4 km^2.

3.2.3 *Selection of the urban training samples*

First, preliminary urban training samples, which were chosen from the potential urban patches, were selected as the pixels that had light brightness values higher than one standard deviation above the average value in each economic region. Furthermore, a preliminary training sample was kept as an urban training sample if it had the maximum brightness value in a local 3×3-pixel window.

3.2.4 *Selection of the nonurban training samples*

Candidate nonurban training samples included all lit pixels except the urban training samples. A candidate was kept as a nonurban training sample if it had an NDVI value greater than the average value in each economic region.

3.3 *Image Classification Using SVM Based on the Divide-and-Conquer Strategy*

The SVM classifier was applied to the composite data of DMSP/OLS NSL and SPOT NDVI using the urban and nonurban training samples. However, considering the great regional variation within China, we applied the SVM classifier in each economic region separately based on the divide-and-conquer strategy rather than classifying the whole nation as a unit.

Furthermore, the SVM classifier was applied in an iterative manner. Specifically, in the first step of the loop procedure, the initial urban training samples were defined as urban seeds, and all the unknown pixels within a 3×3 window for each seed were classified by the SVM classifier. Then, the pixels identified as urban land in the previous step were assigned as new urban training samples and seeds. Next, those pixels were inputs into the SVM classifier for a new classification. Finally, the loop stopped when the number of newly identified urban pixels was zero.[38]

3.4 *Implementation of Adaptive Postclassification Processing*

Adaptive postclassification processes were introduced in SSVM to overcome the limitation of using a single NDVI threshold in OSVM. In keeping with the regional statistic features of DMSP/OLS NSL and SPOT NDVI data, adaptive postclassification processes were carried out on the classification result described previously in Sec. 3.3 for each economic region. Some pseudourban pixels with NDVI higher than one standard deviation above the average value in each economic region or light brightness value more than one standard deviation below the average value in each economic region were relabeled as nonurban pixels. The result of iterative SVM classification for a pixel is defined as Y_1, and the result of the pixel after postclassification Y_2 is described as follows:

$$Y_2 = \begin{cases} 0 & N' > (\bar{x}_{\text{ndvi}} + S_{\text{ndvi}}) | L' \le (\bar{x}_{\text{ols}} + S_{\text{ols}}) \\ Y_1 & \text{otherwise} \end{cases},$$

where 0 is the label of the nonurban class, and N' and L' are the NDVI and light brightness of the pixel, respectively.

4 Results and Accuracy Assessment

Figure 6 presents the final urban land map of China in 2008 using the SSVM procedures described in Sec. 3. In 2008, there were 55,731 km^2 of urban land in the whole of China, accounting for about 0.59% of the total land area in China (Table 2).

The variations in urban land among the eight economic regions are noticeable both in absolute area and in percentage of total land (Fig. 7, Table 2). The area of urban land in the coastal areas of China, such as SCC (11,149 km^2), ECC (10,280 km^2), and NCC (9,059 km^2), were much greater than those in regions in inland China. The area of urban land in SCC accounted for

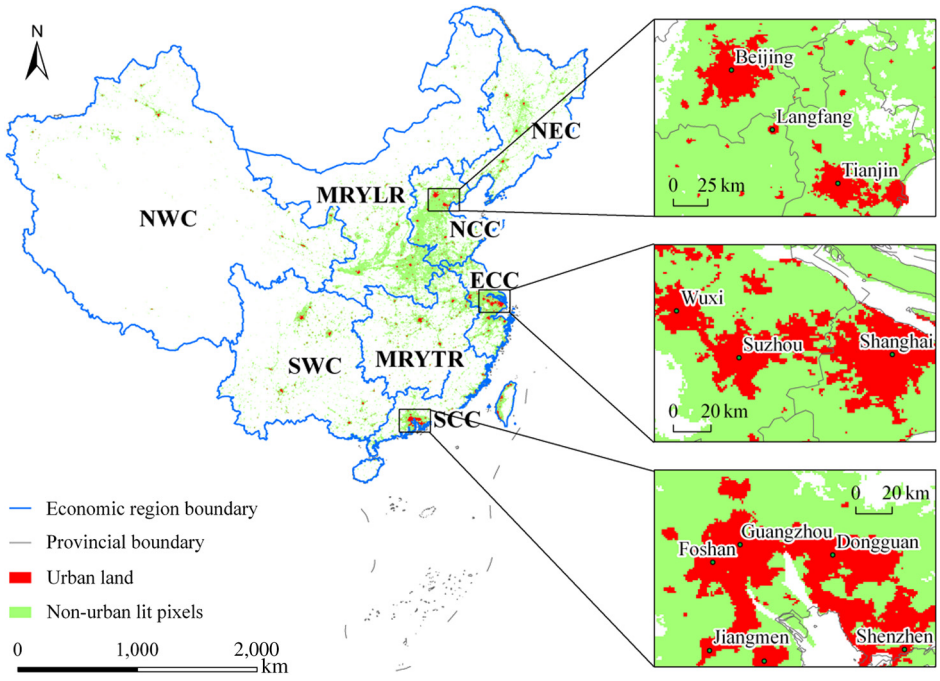

Fig. 6 Resultant land within China in 2008 using SSVM. Note: The white background pixels, recorded as zero in DMSP/OLS NSL data, were considered nonurban land.

Table 2 Urban land extraction results of eight economic regions.

Economic regions	Area of urban land (km^2)	Total area of land (km^2)	Percentage urban land (%)
Northeast China	3928	790,506	0.50
Northern Coastal China	9059	370,953	2.44
Southern Coastal China	11,149	370,251	3.01
Eastern Coastal China	10,280	209,540	4.91
Middle Reaches of the Yellow River	5918	1,670,950	0.35
Middle Reaches of the Yangtze River	6690	704,756	0.95
Southwest China	5648	1,364,178	0.41
Northwest China	3059	4,000,311	0.08
China	55,731	9,481,445	0.59

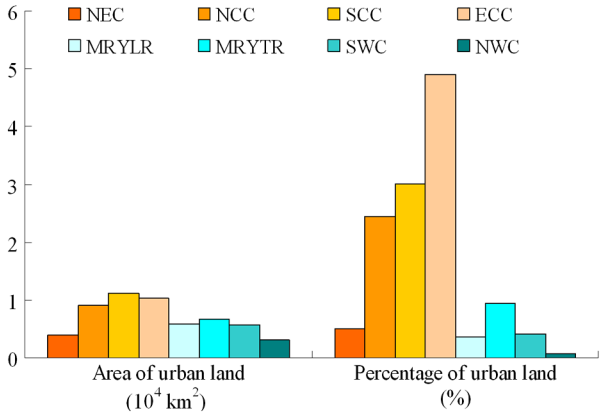

Fig. 7 Area of urban land and percentage of urban land in the eight economic regions.

about one-fifth of the nation's total and is the largest of all the regions, about four times that of NWC (3059 km^2). Meanwhile, the percentages of urban land in ECC (4.91%), SCC (3.01%), and NCC (2.44%) in the coastal areas of China were much higher than those in other regions in inland China, such as NWC (0.08%), MRYLR (0.35%), SWC (0.41%), NEC (0.50%), and MRYTR (0.95%).

To assess the performance of SSVM, we also applied OSVM using the procedure set out by Cao et al.[38] to extract urban land for China using the same DMSP/OLS NSL and SPOT NDVI data.

Elvidge et al.[47,48] and Sutton[25] demonstrated that nighttime light brightness correlated well with population size and level of economic development. We evaluated the results from SSVM and OSVM against urban population and GDP data by province.[56] Evaluation results revealed that the urban land in China in 2008 calculated by SSVM correlated better with statistical data about urban populations and GDP than the results of OSVM did (Fig. 8). The correlation coefficients between statistical data and urban land derived using SSVM ($R > 0.90$, $p < 0.0001$) were almost twice those for OSVM ($R < 0.48$, $p < 0.05$).

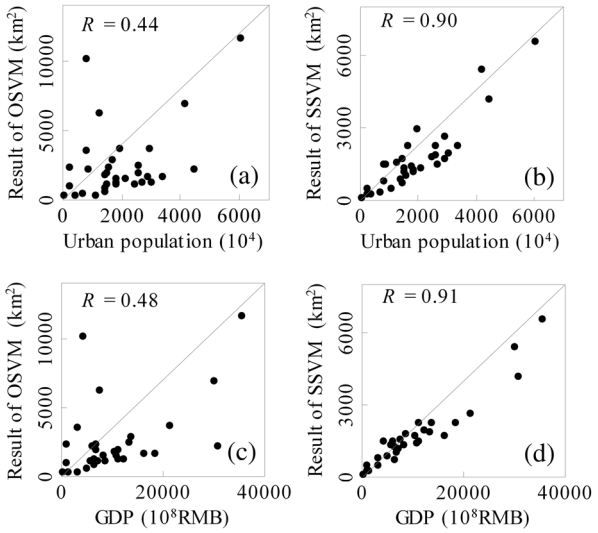

Fig. 8 Correlation analysis between the results from original support vector machine (OSVM)/SSVM and statistical data. (a) Urban land from OSVM and urban population. (b) Urban land from SSVM and urban population. (c) Urban land from OSVM and gross domestic product (GDP). (d) Urban land from SSVM and GDP.

At present, the methods for validation of thematic maps can be grouped into two broad types: those based on field surveys and those based on finer resolution images. For coarse-resolution thematic maps, the field surveys method is complicated by geolocation uncertainties, time and cost requirements, the limited number of on-the-ground samples, field instrument calibration, and sampling errors.[62] As an alternative approach to field surveys, fine-resolution images and derived land cover maps have commonly been used for validation of thematic maps.[63–65]

In this study, we used eight Landsat ETM+images (30-m spatial resolution), which were acquired in 2008, to validate the urban land of China mapped by SSVM. Note that the images were carefully selected to cover one city sample in each economic region (Fig. 2). Meanwhile, to fully consider the urban development discrepancies in Chinese cities, eight validation cities with diverse urban sizes were carefully chosen to represent different levels of urban development in China. The fine-resolution urban land maps produced by the SVM classifier based on training data extracted from typical urban land were resampled to 1-km spatial resolution and compared with the coarse-resolution DMSP/OLS-based urban land map from OSVM and SSVM. Considering that the spatial resolution of Landsat ETM+data is much finer than that of DMSP/OLS NSL data, the classification results of ETM+were accurate enough to be used as reference maps for validation.[7,28,33,38] The evaluation results revealed that SSVM effectively improved the accuracy of extracted urban land compared with OSVM. The overall accuracy (OA) and Kappa improved from 81 to 90% and from 0.52 to 0.69, respectively (Fig. 9). The accuracy of results for Xi'an in MRYLR and Urumchi in NWC were particularly improved; OA improved from 65 to 89% and from 49 to 89%, and Kappa improved from 0.00 to 0.76 and from 0.17 and 0.67, respectively. The accuracy of results for selected cities in ECC, NEC, SCC, and MRYTR also improved to a certain extent, but the accuracy of results from SSVM and OSVM in cities in NCC and SWC did not improve significantly (Fig. 9).

In summary, compared with OSVM, SSVM can extract urban land more accurately for all of China, which has an obvious imbalance in economic development. Three main factors are

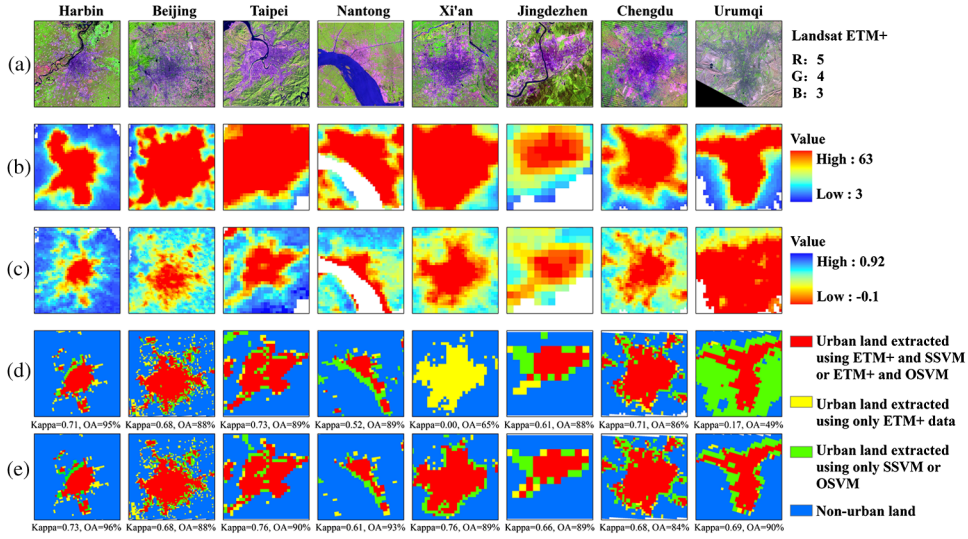

Fig. 9 Accuracy comparisons of the results from OSVM and SSVM using Landsat ETM+imagery. (a) Landsat ETM+imagery. (b) DMSP/OLS NSL data. (c) SPOT NDVI data. (d) Comparison of the results from OSVM and Landsat ETM+classification. The red color is urban land simultaneously extracted from OSVM based on NSL and classification based on ETM+, and the green color is the urban land only extracted from OSVM based on NSL. (e) Comparison of the extracted results from SSVM and Landsat ETM+classification. The red color is urban land simultaneously extracted from SSVM based on NSL and classification based on ETM+, and the green color is the urban land only extracted from SSVM based on NSL. Note: (1) The white background pixels, recorded as zero in DMSP/OLS NSL data, were considered nonurban land. (2) In both (d) and (e), the yellow color is urban land only extracted from ETM+, and the blue color is nonurban land.

responsible for this improved accuracy: (1) SSVM adequately reflects regional variations in urban development in China by stratifying all of China into several smaller regions. The eight economic regions of China exhibit clear differences in socio-economic and remote sensing data, so classification is more efficient when the heterogeneity among regions and relative homogeneity within one region are incorporated. (2) The self-adaptation approach to the training sample selection in SSVM, which effectively selected samples considering the regional statistical features of remote sensing data, overcame OSVM's limitations in selecting training samples. Figure 10 shows how valid urban training samples, such as those in Xi'an and Xianyang, were omitted in OSVM but successfully selected in SSVM. Better training samples improve the accuracy of urban mapping in an economic region. (3) The adaptive postclassification processes successfully enhanced the performance of SSVM. Figure 11 shows how the pseudourban pixels surrounding Urumchi and Karamay were relabeled as nonurban pixels more accurately in SSVM than in OSVM, thereby efficiently decreasing commission errors in results for that region.

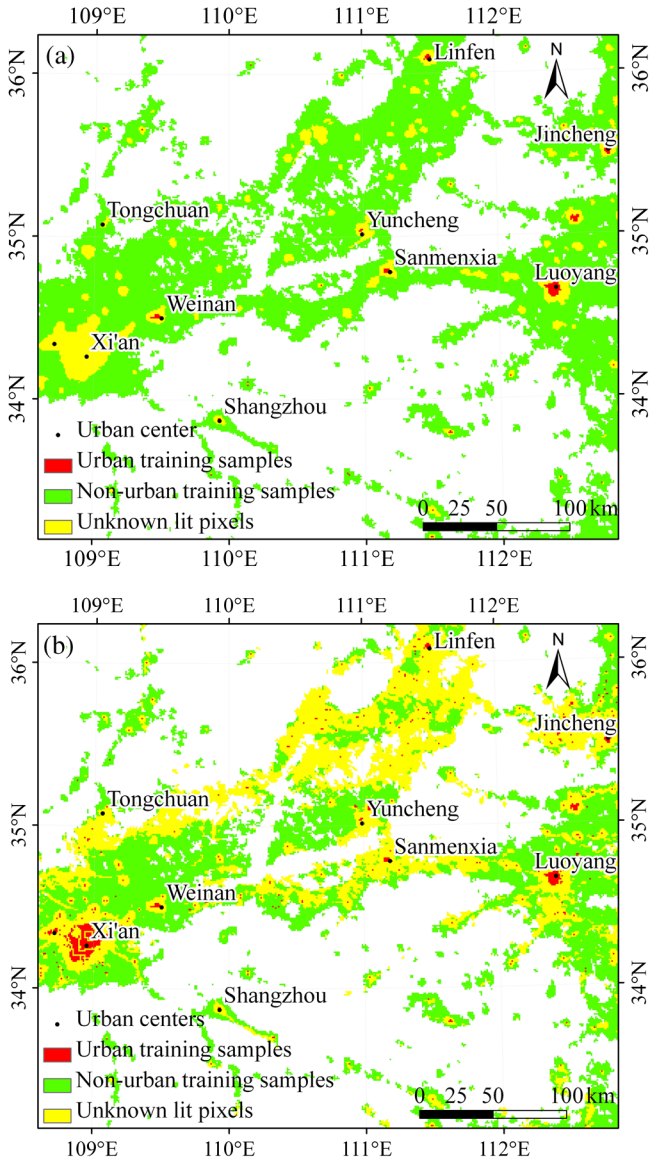

Fig. 10 Comparison of training samples in the middle reaches of the Yellow River (MRYLR) region. (a) OSVM and (b) SSVM. Note: the white background pixels, recorded as zero in DMSP/OLS NSL data, were considered nonurban land.

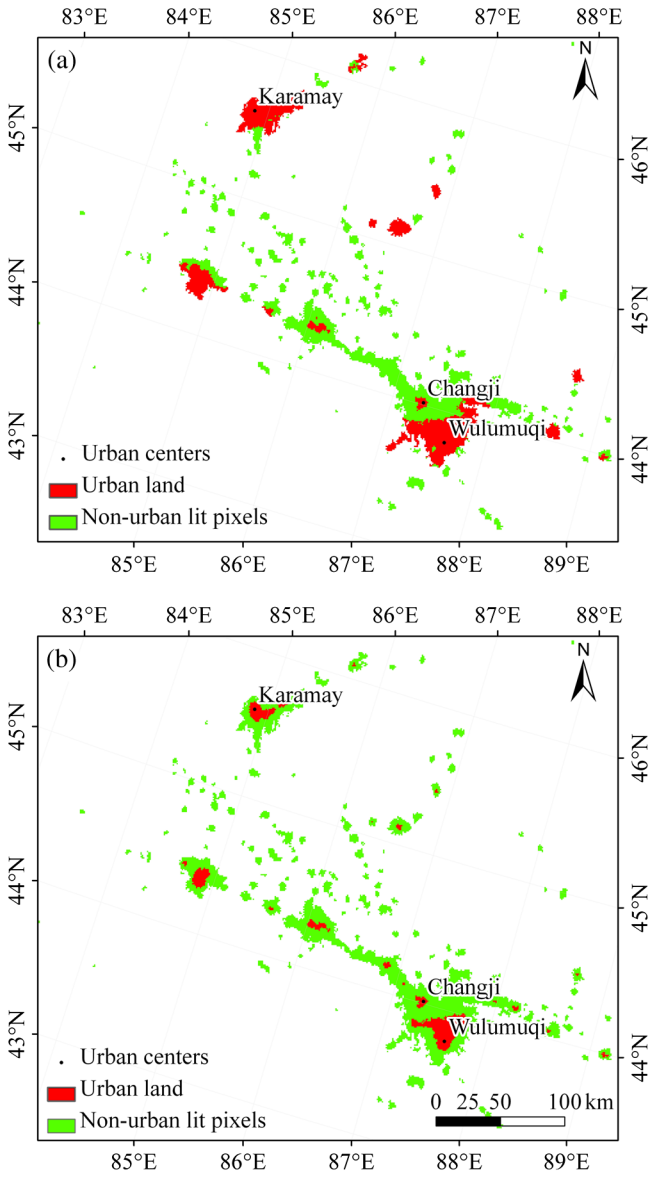

Fig. 11 Comparison of postclassification results in the Northwest China (NWC) region. (a) OSVM and (b) SSVM. Note: the white background pixels, recorded as zero in DMSP/OLS NSL data, were considered nonurban land.

5 Conclusions and Discussion

As a living laboratory of urbanization, China has witnessed fast urban growth in recent decades. This rapid increase in urban land has a great impact on food security and the ecological health of China. The timely and accurate mapping of urban land within China is the first step toward clarifying the urbanization process and revealing its environmental impacts. To effectively map urban land in China at the national scale using DMSP/OLS NSL data, this article proposed a SSVM method as an improvement over the OSVM approach and mapped the urban land within China in 2008 at the national scale.

The result showed that the area of urban land within China in 2008 was 55,731 km^2, which accounted for about 0.59% of the total land area in China. The urban lands in the coastal areas of China were much greater than in other regions in inland China in both area and percentage. For example, the area of urban land in SCC was about four times as high as that in NWC, and the percentage of urban land in ECC was about 64 times as high as that in NWC. This may be because the government of China has given priority to urban development in coastal areas

since the Reform and Open Policy in the late 1970s. The urban land mapped in this study can provide some useful information for recognizing the pattern of urban land in China and developing effective urban planning strategies to reduce regional disparity on urban development in the future.

OSVM is an effective method to extract urban land from DMSP/OLS NSL data at local and regional scales and has performed well for parts of eastern China, but it cannot be effectively applied to all of China, which has a large area and great regional variation. The SSVM method developed in this research can extract all of the urban land of China by effectively using DMSP/OLS NSL data. Basically, SSVM divides an entire study area into several smaller regions according to regional variations in level of urban development. Next, it selects training samples through a self-adaptation process that effectively reflects the urban development status in each economic region. Finally, based on results from iterative classification using SVM for each region, it carries out diverse post-classification processes for each region to capture urban land more accurately. SSVM improves on OSVM due to its regional stratification, self-adaptive training sample selection, separate SVM, and region-specific postclassification. These aspects make SSVM an effective method for extracting urban land within China, which has great regional disparity in urban development.

A comparison of the accuracy of results for urban land within China in 2008 extracted from DMSP/OLS NSL and SPOT/VGT NDVI data revealed that SSVM clearly decreased the omission and commission errors of OSVM. The correlation coefficients between relevant statistical data and urban land derived from SSVM ($R > 0.90$, $p < 0.0001$) were almost twice those from OSVM ($R < 0.48$, $p < 0.05$). Compared with OSVM, the OA of SSVM improved from 81 to 90%, and Kappa improved from 0.52 to 0.69, using reference data derived from Landsat ETM+imagery.

The proposed SSVM method expands on the existing advantages of OSVM for urban land extraction using DMSP/OLS NSL data. The success of this research in its application of SSVM to a large scale area such as the whole of China demonstrates its potential in continental and global research to clarify the urbanization process. However, the parameters set for selecting the training samples and the postclassification of SSVM in this article basically only involved regional variations in socio-economic and remote sensing data of China. Further calibration will be required when the method is applied to other areas. Additionally, nighttime light data with finer spatial resolution[65] are also needed to enhance the feasibility, applicability, and efficiency of SSVM.

Acknowledgments

The authors would like to acknowledge the support from the National Basic Research Program of China (Grant No. 2010CB950901), and the State Key Laboratory of Earth Surface Processes and Resource Ecology (Grant No. 2010-KF-02). The authors would also like to thank the anonymous reviewers for their insightful suggestions.

References

1. E. F. Lambin and H. Geist, "Global land use and land cover change: What have we learned so far?," *Global Change Newsl.* **46**, 27–30 (2001).
2. E. Kalnay and M. Cai, "Impact of urbanization and land-use change on climate," *Nature* **423**(6939), 528–531 (2003), http://dx.doi.org/10.1038/nature01675.
3. J. A. Foley et al., "Global consequences of land use," *Science* **309**(5734), 570–574 (2005), http://dx.doi.org/10.1126/science.1111772.
4. N. B. Grimm et al., "Global change and the ecology of cities," *Science* **319**(5864), 756–760 (2008), http://dx.doi.org/10.1126/science.1150195.
5. D. Normile, "China's living laboratory in urbanization," *Science* **319**(5864), 740–743 (2008), http://dx.doi.org/10.1126/science.319.5864.740.
6. P. L. Knox and L. McCarthy, *Urbanization: An Introduction to Urban Geography*, 2nd ed., Prentice Hall, New Jersey (2005).
7. C. Y. He, "Restoring urbanization process in China in the 1990s by using non-radiance calibrated DMSP/OLS nighttime light imagery and statistical data," *Chin. Sci. Bull.* **51**(13), 1614–1620 (2006), http://dx.doi.org/10.1007/s11434-006-2006-3.

8. J. X. Zhang and Y. H. Zhang, "Remote sensing research issues of the National Land Use Change Program of China," *ISPRS J. Photogramm.* **62**(6), 461–472 (2007), http://dx.doi.org/10.1016/j.isprsjprs.2007.07.002.
9. T. N. Carlson, "Applications of remote sensing to urban problems," *Rem. Sens. Environ.* **86**(3), 273–274 (2003), http://dx.doi.org/10.1016/S0034-4257(03)00073-7.
10. X. Li and A. G. O. Yeh, "Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the Pearl River Delta," *Int. J. Rem. Sens.* **19**(8), 1501–1518 (1998), http://dx.doi.org/10.1080/014311698215315.
11. X. Yang and C. P. Lo, "Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area," *Int. J. Rem. Sens.* **23**(9), 1775–1798 (2002), http://dx.doi.org/10.1080/01431160110075802.
12. W. Ji et al., "Characterizing urban sprawl using multi-stage remote sensing images and landscape metrics," *Comput. Environ. Urban Syst.* **30**(6), 861–879 (2006), http://dx.doi.org/10.1016/j.compenvurbsys.2005.09.002.
13. F. Pacifici, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," *Rem. Sens. Environ.* **113**(6), 1276–1292 (2009), http://dx.doi.org/10.1016/j.rse.2009.02.014.
14. T. Van de Voorde, "Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data," *Landsc. Urban Plan.* **102**(3), 143–155 (2011), http://dx.doi.org/10.1016/j.landurbplan.2011.03.017.
15. J. R. Jensen and D. C. Cowen, "Remote sensing of urban/suburban infrastructure and socio-economic attributes," *Photogramm. Eng. Rem. Sens.* **65**(5), 611–622 (1999).
16. G. J. Tian et al., "Analysis of spatio-temporal dynamic pattern and driving forces of urban land in China in 1990s using TM images and GIS," *Cities* **22**(6), 400–410 (2005), http://dx.doi.org/10.1016/j.cities.2005.05.009.
17. J. Y. Liu et al., "Study on spatial pattern of land-use change in China during 1995-2000," *Sci. China Ser. D.* **46**(4), 373–384 (2003).
18. J. Y. Liu et al., "Spatial and temporal patterns in China's cropland during 1990-2000: an analysis based on Landsat TM data," *Rem. Sens. Environ.* **98**(4), 442–456 (2005), http://dx.doi.org/10.1016/j.rse.2005.08.012.
19. M. H. Tan, X. B. Li, and C. H. Lu, "Urban land expansion and arable land loss of the major cities in China in the 1990s," *Sci. China Ser. D.* **48**(9), 1492–1500 (2005), http://dx.doi.org/10.1360/03yd0374.
20. G. J. Tian, Z. F. Yang, and Y. Q. Zhang, "The spatio-temporal dynamic pattern of rural residential land in China in the 1990s using Landsat TM images and GIS," *Environ. Manage.* **40**(5), 803–813 (2007), http://dx.doi.org/10.1007/s00267-006-0048-6.
21. J. Y. Liu et al., "Spatial patterns and driving forces of land use change in China during the early 21st century," *J. Geogr. Sci.* **20**(4), 483–494 (2010), http://dx.doi.org/10.1007/s11442-010-0483-4.
22. C. D. Elvidge et al., "Satellite inventory of human settlements using nocturnal radiation emissions: a contribution for the global tool chest," *Global Change Biol.* **3**(5), 387–395 (1997), http://dx.doi.org/10.1046/j.1365-2486.1997.00115.x.
23. M. L. Imhoff et al., "A technique for using composite DMSP/OLS "City Lights" satellite data to accurately map urban areas," *Rem. Sens. Environ.* **61**(3), 361–370 (1997), http://dx.doi.org/10.1016/S0034-4257(97)00046-1.
24. T. W. Owen et al., "Using DMSP-OLS light frequency data to categorize urban environments associated with US climate observing stations," *Int. J. Rem. Sens.* **19**(17), 3451–3456 (1998), http://dx.doi.org/10.1080/014311698214127.
25. P. C. Sutton, "A scale-adjusted measure of 'urban sprawl' using nighttime satellite imagery," *Rem. Sens. Environ.* **86**(3), 353–369 (2003), http://dx.doi.org/10.1016/S0034-4257(03)00078-6.
26. A. Schneider, M. A. Friedl, and C. E. Woodcock, "Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data," *Photogramm. Eng. Rem. Sens.* **69**(12), 1377–1386 (2003).
27. H. D. Eva et al., "A land cover map of South America," *Global Change Biol.* **10**(5), 731–744 (2004), http://dx.doi.org/10.1111/gcb.2004.10.issue-5.

28. C. Small, F. Pozzi, and C. D. Elvidge, "Spatial analysis of global urban extent from DMSP-OLS night lights," *Rem. Sens. Environ.* **96**(3–4), 277–291 (2005), http://dx.doi.org/10.1016/j.rse.2005.02.002.
29. C. D. Elvidge et al., "Global distribution and density of constructed impervious surfaces," *Sensors* **7**(9), 1962–1979 (2007), http://dx.doi.org/10.3390/s7091962.
30. Q. Zhang and K. C. Seto, "Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data," *Rem. Sens. Environ.* **115**(9), 2320–2329 (2011), http://dx.doi.org/10.1016/j.rse.2011.04.032.
31. C. D. Elvidge et al., "Mapping of city lights using DMSP Operational Linescan System data," *Photogramm. Eng. Rem. Sens.* **63**(6), 727–734 (1997).
32. T. Croft, "Nighttime images of the earth from space," *Sci Am.* **239**(6), 68–79 (1978).
33. M. Henderson et al., "Validation of urban boundaries derived from global night-time satellite imagery," *Int. J. Rem. Sens.* **24**(3), 595–609 (2003), http://dx.doi.org/10.1080/01431160304982.
34. Z. F. Liu et al., "Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008," *Landsc. Urban Plan.* **106**(1), 62–72 (2012), http://dx.doi.org/10.1016/j.landurbplan.2012.02.013.
35. D. S. Lu et al., "Regional mapping of human settlements in southeastern China with multisensor remotely sensed data," *Rem. Sens. Environ.* **112**(9), 3668–3679 (2008), http://dx.doi.org/10.1016/j.rse.2008.05.009.
36. Q. Weng, D. S. Lu, and J. Schubring, "Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies," *Rem. Sens. Environ.* **89**(4), 467–483 (2004), http://dx.doi.org/10.1016/j.rse.2003.11.005.
37. R. R. Gillies et al., "Effects of urbanization on the aquatic fauna of the Line Creek watershed, Atlanta—a satellite perspective," *Rem. Sens. Environ.* **86**(3), 411–422 (2003), http://dx.doi.org/10.1016/S0034-4257(03)00082-8.
38. X. Cao et al., "A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data," *Rem. Sens. Environ.* **113**(10), 2205–2209 (2009), http://dx.doi.org/10.1016/j.rse.2009.06.001.
39. G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*, John Wiley & Sons, Chichester, UK (2009).
40. V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York (2000).
41. G. M. Foody and A. Mathur, "Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification," *Rem. Sens. Environ.* **93**(1–2), 107–117 (2004), http://dx.doi.org/10.1016/j.rse.2004.06.017.
42. G. M. Foody and A. Mathur, "The use of small training sets containing mixed pixels for accurate hard image classification: training on mixed spectral responses for classification by a SVM," *Rem. Sens. Environ.* **103**(2), 179–189 (2006), http://dx.doi.org/10.1016/j.rse.2006.04.001.
43. F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Rem. Sens.* **42**(8), 1778–1790 (2004), http://dx.doi.org/10.1109/TGRS.2004.831865.
44. P. Mantero, G. Moser, and S. B. Serpico, "Partially supervised classification of remote sensing images through SVM-based probability density estimation," *IEEE Trans. Geosci. Rem. Sens.* **43**(3), 559–570 (2005), http://dx.doi.org/10.1109/TGRS.2004.842022.
45. M. Fauvel et al., "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Rem. Sens.* **46**(11), 3804–3814 (2008), http://dx.doi.org/10.1109/TGRS.2008.922034.
46. C. N. H. Doll, "CIESIN thematic guide to night-time light remote sensing and its applications," Center for International Earth Science Information Network of Columbia University, Palisades, New York, http://sedac.ciesin.columbia.edu/guides (12 July 2012).
47. C. D. Elvidge et al., "Night-time lights of the world: 1994-1995," *ISPRS J. Photogramm.* **56**(2), 81–99 (2001), http://dx.doi.org/10.1016/S0924-2716(01)00040-5.
48. C. D. Elvidge et al., "Radiance calibration of DMSP-OLS low-light imaging data of human settlements," *Rem. Sens. Environ.* **68**(1), 77–88 (1999), http://dx.doi.org/10.1016/S0034-4257(98)00098-4.

49. J. Chen et al., "A study of the urbanization process in China based on DMSP/OLS data: development of a light index for urbanization level estimation," *J. Rem. Sens.* **7**(3), 168–175 (2003), in Chinese.
50. L. Zhuo et al., "Application of compound night light index derived from DMSP/OLS data to urbanization analysis in China in the 1990s," *Acta Geogr. Sin.* **58**(6), 893–902 (2003), in Chinese.
51. C. D. Elvidge et al., "A fifteen-year record of global natural gas flaring derived from satellite data," *Energies* **2**(3), 595–622 (2009), http://dx.doi.org/10.3390/en20300595.
52. Z. Q. Xie et al., "Accelerated human activities affecting the spatial pattern of temperature in the Yangtze River Delta," *Quat. Int.* **226**(1–2), 112–121 (2010), http://dx.doi.org/10.1016/j.quaint.2010.04.027.
53. C. N. H. Doll and S. Pachauri, "Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery," *Energy Policy* **38**(10), 5661–5670 (2010).
54. C. Small et al., "Spatial scaling of stable night lights," *Rem. Sens. Environ.* **115**(2), 269–280 (2011), http://dx.doi.org/10.1016/j.rse.2010.08.021.
55. T. Ma et al., "Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: a comparative case study from China's cities," *Rem. Sens. Environ.* **124**(9), 99–107 (2012), http://dx.doi.org/10.1016/j.rse.2012.04.018.
56. National Bureau of Statistics of China, *China Statistical Yearbook-2009*, China Statistics Press, Beijing, China (2009).
57. National Bureau of Statistics of China, *China City Statistical Yearbook-2009*, China Statistics Press, Beijing, China (2010).
58. F. Bovolo and L. Bruzzone, "A split-based approach to unsupervised change detection in large-size multitemporal images: application to tsunami-damage assessment," *IEEE Trans. Geosci. Rem. Sens.* **45**(6), 1658–1670 (2007), http://dx.doi.org/10.1109/TGRS.2007.895835.
59. A. Schneider, M. A. Friedl, and D. Potere, "Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'," *Rem. Sens. Environ.* **114**(8), 1733–1746 (2010), http://dx.doi.org/10.1016/j.rse.2010.03.003.
60. Development Research Center of the State Council of China, "Coordinated Regional Development Strategy and Policy Reports," 27 April 2005, http://www.cqvip.com/QK/85199X/200565/15595896.html (12 July 2012).
61. P. L. Fan and J. G. Qi, "Assessing the sustainability of major cities in China," *Sustain Sci.* **5**(1), 51–68 (2010), http://dx.doi.org/10.1007/s11625-009-0096-y.
62. S. Ganguly et al., "Generating vegetation leaf area index earth system data record from multiple sensors. Part 2: implementation, analysis and validation," *Rem. Sens. Environ.* **112**(12), 4318–4332 (2008), http://dx.doi.org/10.1016/j.rse.2008.07.013.
63. X. M. Xiao et al., "Characterization of forest types in Northeastern China, using multi-temporal SPOT-4 VEGETATION sensor data," *Rem. Sens. Environ.* **82**(2–3), 335–348 (2002), http://dx.doi.org/10.1016/S0034-4257(02)00051-2.
64. J. M. N. Silva, A. C. L. Sá, and J. M. C. Pereira, "Comparison of burned area estimates derived from SPOT/VEGETATION and Landsat ETM+ data in Africa: influence of spatial pattern and vegetation type," *Rem. Sens. Environ.* **96**(2), 188–201 (2005), http://dx.doi.org/10.1016/j.rse.2005.02.004.
65. N. Levin and Y. Duke, "High spatial resolution night-time light images for demographic and socio-economic studies," *Rem. Sens. Environ.* **119**, 1–10 (2012), http://dx.doi.org/10.1016/j.rse.2011.12.005.

Biographies and photographs of the authors are not available.