Environ Model Assess (2013) 18:547–558
DOI 10.1007/s10666-013-9359-1

Estimation of Vegetation Coverage in Semi-arid Sandy Land Based on Multivariate Statistical Modeling Using Remote Sensing Data

Wei Chen · Tetsuro Sakai · Kazuyuki Moriya · Lina Koyama · Chunxiang Cao

Received: 2 April 2012 / Accepted: 6 February 2013 / Published online: 19 February 2013

Abstract The estimation of vegetation coverage is essential in the monitoring and management of arid and semi-arid sandy lands. But how to estimate vegetation coverage and monitor the environmental change at global and regional scales still remains to be further studied. Here, combined with field vegetation survey, multispectral remote sensing data were used to estimate coverage based on theoretical statistical modeling. First, the remote sensing data were processed and several groups of spectral variables were selected/proposed and calculated, and then statistically correlated to measured vegetation coverage. Both the single- and multiple-variable-based models were established and further analyzed. Among all single-variable-based models, that is based on Normalized Difference Vegetation Index showed the highest R (0.900) and R^2 (0.810) as well as lowest standard estimate error (0.128024). Since the multiple-variable-based model using multiple stepwise regression analysis behaved much better, it was determined as the optimal model for local coverage estimation. Finally, the estimation was conducted based on the optimal model and the result was cross-validated. The coefficient of determination used for validation was 0.867 with a root-mean-squared error (RMSE) of 0.101. The large-scale estimation of vegetation coverage using statistical modeling based on remote sensing data can be helpful for the monitoring and controlling of desertification in arid and semi-arid regions. It could serve for regional ecological management which is of great significance.

Keywords Correlation analysis · Remote sensing · Statistical modeling · Vegetation coverage · Vegetation index · Environment assessment

W. Chen (✉) · T. Sakai · K. Moriya · L. Koyama
Biosphere Informatics Laboratory, Department of Social Informatics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan
e-mail: chenwei@bre.soc.i.kyoto-u.ac.jp

C. Cao
State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications of Chinese Academy of Sciences, Beijing 100101, China

1 Introduction

As an important parameter frequently used to describe the quantity of vegetation on land surface, vegetation coverage is defined as the ratio of upright area of vegetation projected to land surface area [1–3]. It has always been a crucial variable in terrestrial ecosystem monitoring and climate change studies [4], thus playing an important role in the research of various fields, such as climate change, resources survey, as well as environment management and so on [5–7]. Especially in the context of global climate change, desertification is gradually becoming a serious phenomenon in arid and semi-arid regions, especially in western China [2]. As a crucial factor indicating the impacts of many biophysical processes including the evaporation and transpiration as well as photosynthesis, vegetation coverage is attracting wide attention from not only the experts but the public [8].

In western China, there are a large area of arid sandy lands, among which, the Mu Us sandy land ranks the four largest ones. The shrubs and meadow play an essential role in the terrestrial ecosystems of this region [9, 10] and vegetation coverage, which acts as a main factor indicating the land use and land cover, has been closely related with soil and water loss and land desertification in local area. Thus, it is crucial to perform estimation of the vegetation coverage which can serve for the monitoring and improvement of local ecological environment [11, 12].

Traditional ground-based field measurements of vegetation coverage, which are implemented mainly through plot survey [13] or digital photograph classification [14], seem relatively subjective, expensive, time consuming as well as labor intensive, and thus appear unrealistic for the dynamic monitoring of land vegetation status in wide-range, long-term and with high accuracy. As an alternative, the advanced remote sensing techniques, which are greatly developed and widely applied in various fields, have been proved to be able to provide systematic, repetitive and comprehensive observations at scales from local to continental and even global [15, 16]. Recently, a variety of remote-sensing satellites and sensors with different spatial, spectral, and temporal resolutions have been developed rapidly, providing sufficient spatial and temporal coverage of high-quality data at different scales. As a result, the potential of remote sensing techniques as a monitoring tool has been widely recognized [17, 18].

Consequently, remote sensing-data now represent an affordable, cost-effective, and standardized source of environmental information and have been greatly used for environmental surveys. Furthermore, multisource remote-sensing data have also been utilized to monitor the vegetation growth and environmental change in many disciplines [6, 10, 11, 16, 19–23]. The remote-sensing approaches are potentially suited for the effective and efficient estimation of vegetation coverage.

Among all remote-sensing data-based applications, multispectral satellite and aerial images which are well suitable for capturing horizontally distributed characteristics and changes, have been widely used to estimate vegetation coverage especially at moderate resolution and in large scales [16, 24]. Based on these data, the statistical modeling algorithm is one of the most convenient and effective methods. It is generally performed through spectral reflectance or transformations such as vegetation indices (e.g., Normalized Difference Vegetation Index (NDVI)) which are statistically correlated to field measured coverage values [12]. As the statistical modeling approaches show obvious advantages of simple modeling principles and easy calculations, less demanding of input data, thus are particularly suitable for quantitative estimation of vegetation coverage.

In this study, taking the Mu Us sandy land in western China as study site, we proposed to estimate the vegetation coverage based on multivariate statistical models. During the modeling process, several groups of remote-sensing variables had been selected and calculated, some of which were established by ourselves, and then statistically correlated to vegetation coverage data from field survey. Both the single- and multiple-variable relationships between remote-sensing variables and field measured values were further analyzed and final estimation was performed using an optimal model. The estimation result was also cross-validated and can be used to provide overall quantitative descriptions of local vegetation. It could serve for local vegetation monitoring and ecological management which is of great significance.

2 Methods

2.1 Study Site

Our study site is the Mu Us sandy land which is located at the junction of Ningxia, Inner Mongolia autonomous region and Shaanxi province with geographic coordinates ranging from 37°27.5′ to 39°22.5′ N in latitude and from 107°20′ to 110°30′ E in longitude. This region with an area of 4×10^4 km^2 and elevation from 1,100 to 1,300 m, includes the Ordos city in southern Inner Mongolia Autonomous Region, the northern part of Shaanxi Yulin, and the northeast of the Ningxia Hui Autonomous Region. As one of the four largest sandy lands in China, it lies in the transition between the Ordos Plateau and the Loess Plateau [2].

The specific study area ranging from 37°39′ to 39°24′ N in latitude and from 108°17′ to 109°40′ E in longitude locates in the hinterland of the Mu Us sandy land in Wushen Banner, Ordos City, Inner Mongolia Autonomous Region (Fig. 1). The region has a typical temperate continental semi-arid climate with low rainfall, droughts, strong winds and evaporation as well as abundant sunshine. The annual average temperature is 6 to 8 °C with an annual average accumulated temperature of 2,621 °C. The annual evaporation is about 2,592 mm, while the annual average of sunshine hours are 2,860 h which together result in an arid climate. Although the sandy land is seated in a semi-arid steppe area, wind-blown sand accounts for 78.3 % of the total area of Wushen Banner and the soil type of this area includes meadow soil, chestnut soil, saline soil, loess soil, and swamp soil [9]. The shrubs which are dominant in the Mu Us sandy land include the species of *Tamarix*, *Salix psammophila*, *Peking willow*, *Hippophae rhamnoides*, *Scoparium*, *Caragana*, *Artemisia*, *Hedysarum*, and *Salix cheilophila* [7].

2.2 Remote-Sensing Image

In this research, the HJ-1 ("Huan Jing"—Chinese pronunciation of "Environment") charge-coupled device (CCD) data acquired at 2nd July 2009 was used for modeling and analysis. The Small Satellite Constellation A and B satellites for Environment and Disaster Monitoring and Forecasting (HJ-1A/1B satellites) were launched at 11:25 on 6 September 2008 (http://www.cresda.com/n16/n1130/n1582/8384.html). The HJ-1A satellite is equipped with a CCD camera and a hyperspectral image radiometer, while the HJ-

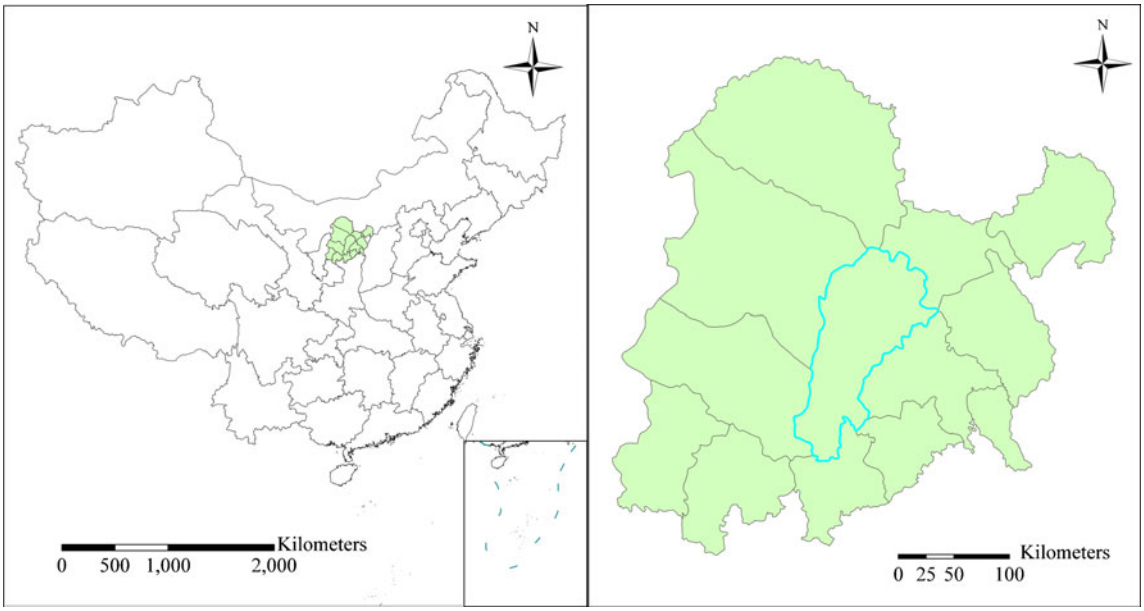

Fig. 1 The location of the Mu Us sandy land in China (*left*) and the specific study area in sandy land (*right*)

1B satellite is equipped with a CCD camera and infrared camera. The CCD cameras on HJ-1A and HJ-1B are of the same design and make complementary observations. They are combined to achieve four-spectrum pushbroom imaging with a ground swath width of 700 km and a ground pixel resolution of 30 m. As only the CCD data are used here, the main characteristics of the CCD cameras are listed in Table 1.

2.3 Data Pre-processing

Based on acquisition of the remote-sensing image, by using the Landsat Thematic Mapper/Enhanced Thematic Mapper plus image of the same area as the reference image and selecting the same points of road intersections, border points of water, as well as other typical features, the HJ-1B image was geometrically corrected through the polynomial model. During the geometric correction, the error was controlled within 0.5 pixels, which met the requirement of the study.

After geometric correction, we calculated the surface reflectance from the digital number (DN) value through radiometric calibration and atmospheric correction. Firstly, we converted the DN value of the raw image to at-satellite radiance through the following equation:

$$L_\lambda = \mathrm{DN}/\mathrm{gain} + \mathrm{offset} \quad (1)$$

Where L_λ is the at-satellite radiance, DN is the digital number of the raw image, and gain and offset indicate the absolute radiometric calibration coefficients of HJ-1B CCD (Table 2), respectively. The unit of converted radiance is W m^{-2} sr^{-1} μm^{-1}.

The atmospheric correction was applied based on moderate resolution atmospheric transmission (MODTRAN) model. The MODTRAN model employs a stored spectral database for H_2O, CO_2, O_3, N_2O, CO, CH_4, O_2, NO, SO_2, NO_2, NH_3, and HNO_3. It considers the effects of scattering (such as Rayleigh and Mie scattering) and allows the user to specify the profiles of temperature, water vapor density, ozone, aerosols, and any other gases that may vary with time [25]. In our study area, the aerosol model was of the rural-based aerosol and the aerosol optical thickness which described the aerosol optical properties came from the product of moderate resolution imaging spectroradiometer (MODIS). MODIS is a key instrument aboard the Terra and Aqua satellites. Terra's orbit around the earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra MODIS and Aqua MODIS are viewing the entire earth's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups of wavelengths. These data will improve our understanding of global dynamics and processes occurring on the land, in the oceans, and in the lower atmosphere (http://modis.gsfc.nasa.gov/index.php).

Through the pre-processing steps of geometric and atmospheric correction, we finally acquired the surface reflectance of the study area.

2.4 Field Survey

In order to investigate the feasibility and practicality of environment monitoring in Mu Us sandy land using HJ-1A/B satellite data, a field survey was planned and achieved in the land area within Wushen Banner, Ordos from 3th to 15th July 2009. As the spatial resolution of HJ-1 satellite was

Table 1 Characteristics of HJ-1A/1B satellite CCD camera

Satellite	Sensor	Bands	Wavelength range (μm)	Spatial resolution (m)	Swath width (km)	Sub-cycle (days)
HJ-1A/B	CCD	1	0.43–0.52	30	360	4
		2	0.52–0.60	30		
		3	0.63–0.69	30		
		4	0.76–0.90	30		

30 m, the size of our sample plots was designed as 30×30 m. The coordinates of four corners and the centre of each plot were measured using a Differential Global Positioning System device with a maximum error of 6 m. Within each plot, the maximum height and the crown radius of each clump of shrubs were measured using altimeter rods, tape measure, etc. There were totally 19 sample plots which were surveyed.

As the vegetation within these plots in the study area were relatively sparse and unevenly distributed, it was unsuited to use the diagonal intercept sampling or visual interpreting methods to measure the vegetation coverage. An alternative method here being used was to calculate the projected area of tree's canopy on the ground based on the canopy structural parameters, which was then divided by the plot's area to ultimately calculate the vegetation coverage.

As the field measured coverage are spatial data, we need to test the magnitude of spatial autocorrelation. Here, we selected the Moran's I index [26] which is one of the most famous and commonly used method to calculate spatial autocorrelation. In general, a Moran's I Index value near + 1.0 indicates clustering while an index value near −1.0 indicates dispersion. The calculation was conducted in the software of ArcGIS 9.3 (Developed by ESRI corporation). The calculation result of "Moran's Index=0.2" and "Z Score= 1.06" indicate that while somewhat clustered, the pattern of data may be due to random chance. It means that the original spatial data of measured coverage do not have significant spatial autocorrelation and can be used to perform linear regression analysis which requiring independent observations. It meets the assumption of data independence.

2.5 Multivariate Selection and Correlation Analysis

Based on the acquisition of pre-processed remote sensing image and the field measured vegetation coverage, we were able to carry out modeling. During the modeling process, we selected and calculated several groups of spectral variables which were then statistically correlated to field measured values. These variables were as follows:

Table 2 The absolute radiometric calibration coefficient of HJ-1B CCD

HJ-1B CCD	$Gain_{(DN/W\ m^{-2}\ sr^{-1}\ mm^{-1})}$	$Offset_{(W\ m^{-2}\ sr^{-1}\ mm^{-1})}$
Band1	0.5782	3.4608
Band2	0.5087	5.8769
Band3	0.6825	8.0069
Band4	0.6468	8.8583

2.5.1 Surface Reflectance of Single Band

From long-term studies, we have found that there is significant difference between the spectral characteristics of vegetation and other surface features. Especially for healthy vegetation, characteristics of spectral curves (Fig. 2) are particularly notable: a small peak of reflectivity of 10–20 % high at the green band with central wavelength of 0.55 μm; two absorption valleys at the blue-purple band of 0.45 μm and red band of 0.65 μm, respectively; a sharp increase shown as a steep on the curve at about 0.7–0.8 μm followed by a significant reflectivity peak of 40 % or larger in the near-infrared band of 0.8–1.3 μm. Additionally, there are another three absorption valleys at the bands of 1.45, 1.95, and 2.6–2.7 μm, respectively. The spectral characteristics of green vegetation is a comprehensive reflection of its own biophysical characteristics and environmental impacts, which can be used to identify and monitor vegetation in various related fields [1, 27].

Based on above analysis, we realize that single-band reflectance could be used to establish regression model with measured coverage values for the coverage estimation at

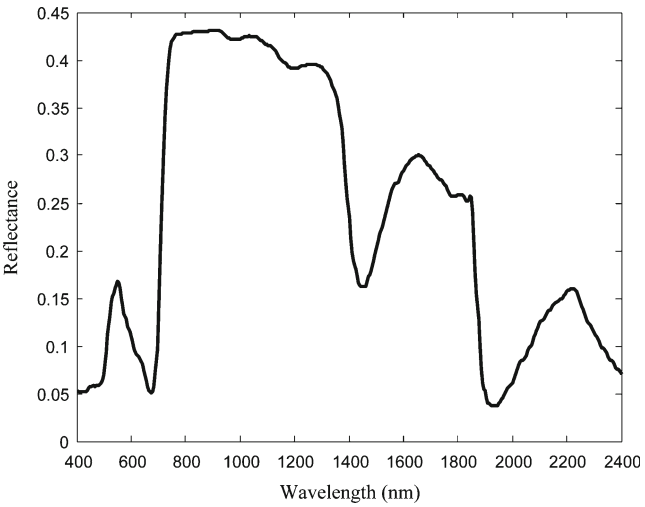

Fig. 2 The spectral curve of healthy vegetation

larger scales. For example, Graetz et al. [28] took use of the reflectance of Landsat Multi-Spectral Scanner (MSS) band 5 and the measured vegetation coverage to build linear regression models and then estimated the coverage of large sparse grassland. Here, we used the four single-band reflectance of HJ-1 image to carry out correlation analysis and establish statistical models. As the Pearson correlation analysis was performed, the correlation coefficient was calculated according to Eq. 2:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (2)$$

Here, the x_i and y_i indicate two groups of variables; $\bar{x}$ and $\bar{y}$ are corresponding mean values, and r is the correlation coefficient. The result of correlation analysis was shown in Table 3.

From Table 3, we could find that the significant probabilities (Sig.) between reflectance of band1, band2, and band3 and measured vegetation coverage were all less than 0.05 which indicates significant correlation between them.

Table 3 The results of Pearson correlation analysis between four groups of variables and measured vegetation coverage

Variables[a]	Pearson correlation coefficient	Sig. (1-tailed)
Single-band reflectance of four bands of HJ-1		
Band1	−0.575	0.005[b]
Band2	−0.700	0.000[b]
Band3	−0.825	0.000[b]
Band4	−0.008	0.486
Vegetation indexes		
DVI	0.729	0.000[b]
RVI	0.896	0.000[b]
NDVI	0.900	0.000[b]
SAVI	0.837	0.000[b]
ARVI	0.875	0.000[b]
EVI	0.845	0.000[b]
The first three components from PCA		
PCA_1	−0.436	0.031[a]
PCA_2	−0.865	0.000[b]
PCA_3	−0.333	0.082
SBI, GVI, and YSI from modified tasseled cap transform		
SBI	−0.671	0.001[b]
GVI	0.776	0.000[b]
YSI	−0.265	0.136

[a] Variables of Pearson correlation analysis, another one is measured vegetation coverage

[a] Significant correlation ($0.01 < \text{Sig.} \leq 0.05$)

[b] Extremely significant correlation ($\text{Sig.} \leq 0.01$)

2.5.2 Vegetation Indexes

Vegetation Index (VI) is a combination of single bands of remote sensing image and can be used as a simple, effective and experienced characterization of ground vegetation state [29–31]. Previous studies have shown that vegetation index has a good correlation with a variety of physiological and ecological parameters of vegetation and thus be widely used to diagnose a range of vegetation biophysical parameters, such as Leaf Area Index, vegetation coverage, biomass and so on [12, 32, 33].

In our study, when establishing the relationships between vegetation coverage and vegetation indexes, we considered the impacts of soil background and atmosphere and then finally selected six indexes for modeling. They are Difference Vegetation Index (DVI) [34], Ratio Vegetation Index (RVI) [35], NDVI [36], Soil-Adjusted Vegetation Index (SAVI) [29], Atmospherically Resistant Vegetation Index (ARVI) [37], and Enhanced Vegetation Index (EVI) [38] which were expressed from Eq. 3 to Eq. 8.

$$\mathrm{DVI} = \rho_{\mathrm{NIR}} - \rho_{\mathrm{Red}} \quad (3)$$

$$\mathrm{RVI} = \frac{\rho_{\mathrm{NIR}}}{\rho_{\mathrm{Red}}} \quad (4)$$

$$\mathrm{NDVI} = \frac{(\rho_{\mathrm{NIR}} - \rho_{\mathrm{Red}})}{(\rho_{\mathrm{NIR}} + \rho_{\mathrm{Red}})} \quad (5)$$

$$\mathrm{SAVI} = \frac{(1 + \mathrm{L}_1) \cdot (\rho_{\mathrm{NIR}} - \rho_{\mathrm{Red}})}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{Red}} + \mathrm{L}_1} \quad (6)$$

$$\mathrm{ARVI} = \frac{(\rho_{\mathrm{NIR}} - \rho_{\mathrm{RB}})}{(\rho_{\mathrm{NIR}} + \rho_{\mathrm{RB}})} \quad (7)$$

$$\mathrm{EVI} = \frac{2.5 \cdot (\rho_{\mathrm{NIR}} - \rho_{\mathrm{Red}})}{\rho_{\mathrm{NIR}} + c_1 \cdot \rho_{\mathrm{Red}} - c_2 \cdot \rho_{\mathrm{Blue}} + \mathrm{L}_2} \quad (8)$$

Here ρ_{NIR}, ρ_{Red}, and ρ_{Blue} indicate the reflectance of the near-infrared and red and blue bands respectively. Then $\rho_{\mathrm{RB}} = \rho_{\mathrm{Red}} - \gamma(\rho_{\mathrm{Blue}} - \rho_{\mathrm{Red}})$ and γ which indicating the radiation correction coefficient of optical path is assumed to the recommended value (by Kaufman) of 1. The parameter L_1 in Eq. 6 indicates a soil adjusted coefficient which is normally assumed to be 0.5 for most areas. In Eq. 8, we set c_1=6.0, c_2= 7.5 and L_2=1 which have been used in the MODIS product.

After acquiring the six vegetation indexes, Pearson correlation analysis was also performed and results were shown in Table 3. From this part, it could be easily found that the

Table 4 The model summary and ANOVA of single-variable-based models and the multiple-variable-based model

Independent variables[a]	R	R^2	SEE	F statistics	Sig.
Single band reflectance of HJ-1[b]					
Band1	−0.575	0.330	0.241	8.386	0.010[c]
Band2	−0.700	0.490	0.210	16.362	0.001[c]
Band3	−0.825	0.681	0.166	36.225	0.000[c]
Vegetation indexes[b]					
DVI	0.729	0.532	0.201	19.300	0.000[c]
RVI	0.896	0.803	0.130	69.425	0.000[c]
NDVI	0.900	0.810	0.128	72.633	0.000[c]
SAVI	0.837	0.701	0.161	39.930	0.000[c]
ARVI	0.875	0.766	0.142	55.593	0.000[c]
EVI	0.845	0.714	0.157	42.419	0.000[c]
The first two components from PCA[b]					
PCA_1	−0.436	0.189	0.265	3.989	0.062
PCA_2	−0.865	0.748	0.147	50.543	0.000[c]
SBI and GVI from modified tasseled cap transform[b]					
SBI	−0.671	0.450	0.218	13.938	0.002[c]
GVI	0.776	0.602	0.186	25.686	0.000[c]
Multiple-variable-based model					
NDVI and GVI	0.945	0.893	0.099	66.962	0.000[c]

SEE standard error of the estimate

[a]Independent variables include constant; dependent variable is measured coverage

[b]Single-variable-based model

[c]Extremely significant correlation (Sig.≤0.01)

Pearson correlation coefficients between all six indexes and measured coverage were larger than 0.7 and corresponding significant probabilities (Sig.) were all 0.000 which indicating extremely significant correlation between them.

2.5.3 Principal Components from Principal Component Analysis

When the concept of principal component was firstly proposed in 1901 by Pearson, it was only for non-random variables. Then its application was extended to random vectors in 1933. Principal Components Analysis is one kind of transformations used to simplify the data structure through dimension reduction. It is usually used to produce uncorrelated output bands, to segregate noise components, and to reduce the dimensionality of data sets. Because multispectral data bands are often highly correlated, the principal components transformation is used to produce uncorrelated output bands. This is done by finding a new set of orthogonal axes that have their origin at the data mean and that are rotated so the data variance is maximized. By principal component analysis (PCA), multiple variables (indicators) were reduced to much less variables (indicators), which still incorporated most of the original information [39]. Its principle could be expressed by following equations:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = (X_1 \quad X_2 \quad \cdots \quad X_P) \qquad (9)$$

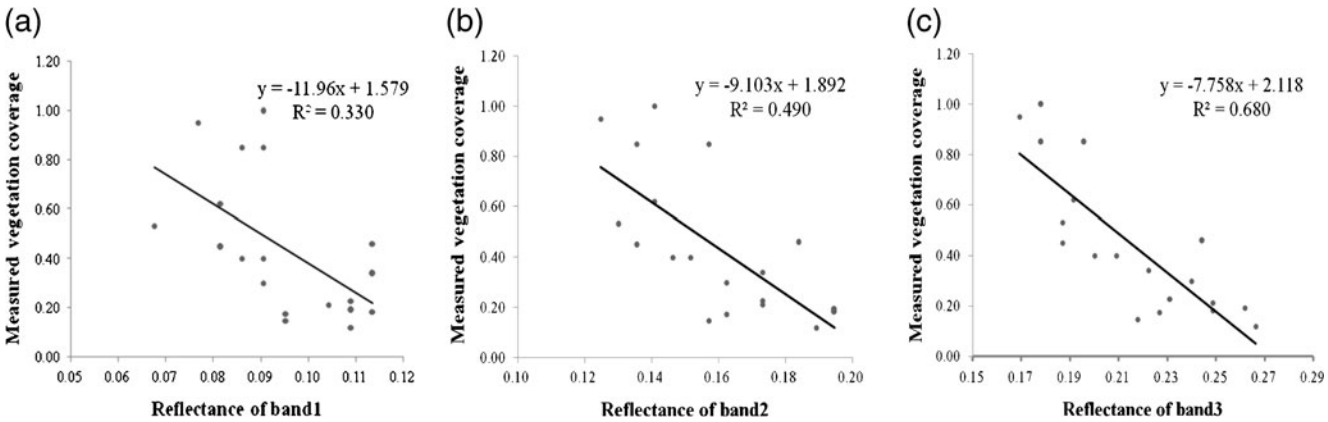

Fig. 3 The models based on single-band surface reflectance

$$\begin{cases} F_1 = a_{11}x_1 + a_{21}x_2 + \cdots + a_{p1}x_p \\ F_2 = a_{12}x_1 + a_{22}x_2 + \cdots + a_{p2}x_p \\ \vdots \\ F_p = a_{1p}x_1 + a_{2p}x_2 + \cdots + a_{pp}x_p \end{cases} \quad (10)$$

$$F_i = a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{pi}x_p \quad i = 1, 2 \cdots p$$

Here $(X_1 \quad X_2 \quad \cdots \quad X_P)$ indicate the original variables and $(F_1 \quad F_2 \quad \cdots \quad F_P)$ indicate the new variables produced from PCA. They meet the two conditions: $cov(F_i, F_j) = 0$ and $var(F_i) \geq var(F_j) \quad i < j$. Generally, the first three components contain more than 85 % of the previous information which are usually used to represent the original image.

Using the first three components from PCA, Pearson correlation analysis was also performed and result was shown in Table 3. From it, we could conclude that the first two components were both significantly correlated with measured coverage (Sig.<0.05) while the third one was not.

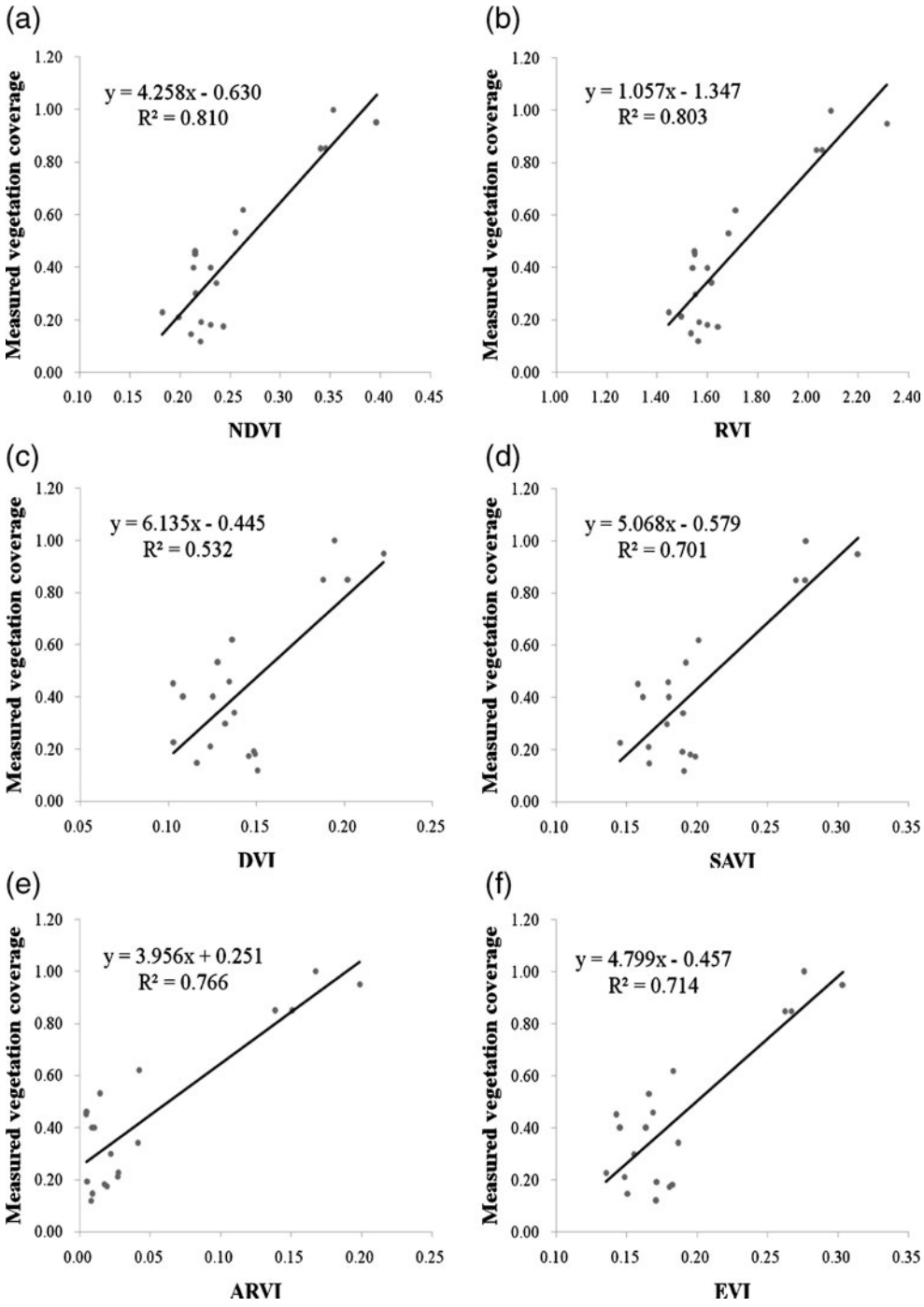

Fig. 4 The models based on vegetation indexes

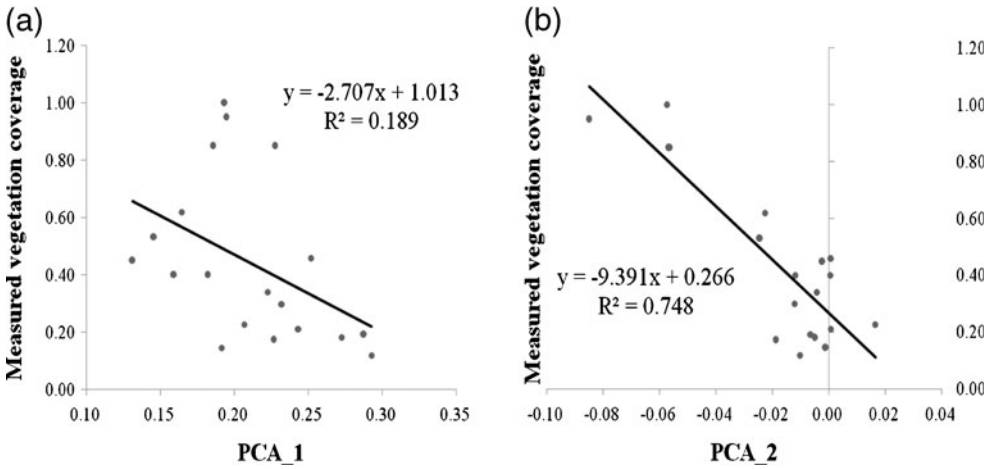

Fig. 5 The models based on two principal components from PCA

2.5.4 Spectral Variables from Tasseled Cap Transformation

Similar with PCA, tasseled cap transformation is also one type of spectral transformations which turn original, highly covariant data into three uncorrelated indices called brightness, greenness, and wetness (the third tasseled cap transformation components for MSS was called yellow stuff). But the difference lies in that it is only designed for Landsat data during a long period of time and be extended to a few other sensors such as IKONOS [40] and MODIS [41] until recently.

At present, there was no existing transform matrix for HJ data which we can use in our study, new transform matrix should be proposed. Thus before performing the tasseled cap transformation, we modified the existing transform matrix designed for Landsat MSS [42] and TM/ETM+ [43] sensors to fit the HJ-1 image according to their similarities in spectral characteristics. The designed transformation matrix was shown as follows:

$$B = \begin{bmatrix} 0.332 & 0.603 & 0.675 & 0.262 \\ -0.285 & -0.244 & -0.544 & 0.724 \\ -0.899 & 0.428 & 0.076 & -0.041 \end{bmatrix} \quad (11)$$

Based on the improved tasseled cap transformation, three variables of SBI (Soil Brightness Index), Green Vegetation Index (GVI) and Yellow Stuff Index (YSI) were got and correlation analysis was performed. From Table 3, we concluded that the SBI and GVI had significant correlation with measured coverage and hence be used to estimate vegetation coverage.

3 Results

3.1 Multivariate Statistical Modeling

We had selected the reflectance of four bands of HJ-1 image, six vegetation indexes, three principal components from PCA and three variables of SBI, GVI, and YSI from tasseled cap transformation and then analyzed the correlation between all variables and measured vegetation coverage. With those showing significant correlation, statistical regression models were established and analysis of variance (ANOVA) were conducted. The results were shown in Table 4. The effective models which were statistically significant could be used for estimation of vegetation coverage at large scales.

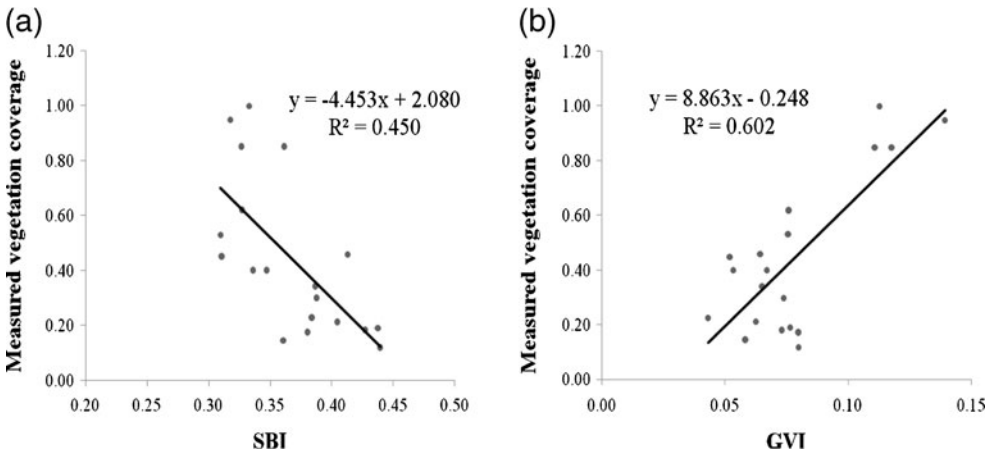

Fig. 6 The models based on SBI and GVI from tasseled cap transformation

3.1.1 Single Band Reflectance of HJ-1 Data

For this group of variables, we used the reflectance of band1, band2 and band3 to establish the statistical regression models. The modeling results were shown in Table 4 and Fig. 3. From them, we could conclude that the three models were all significantly effective (Sig.≤0.01) and could be used to estimate the vegetation coverage individually.

3.1.2 Vegetation Indexes

Based on these vegetation indexes, we were also able to establish the estimation models which were shown in Fig. 4. The model summary and ANOVA result were shown in Table 4. As expressed in the table, the correlation coefficient (R) and determination coefficient (R^2) were all quite high (most R are larger than 0.8 and most R^2 are larger than 0.7) and the standard error of the estimate (SEE) were particularly little. These coefficients together indicated the validity and efficiency of these models. The significant probability (Sig.) of 0.000 was a supplementary indication of the effects.

Compared with the models from single-band reflectance, the advantage of these ones seem especially obvious. This may be due to that the vegetation indexes are combinations of single bands and thus incorporate comprehensive information. Additionally, the indexes remove the impact of soil background and atmosphere which help to improve their accuracy. These also account for the wide applications of vegetation indexes in vegetation and ecological studies.

3.1.3 Principal Components from PCA

As only the first two principal components were significantly correlated with measured coverage, we established models only based on the two variables. The results were shown in Table 4 and Fig. 5 from which we found that the model from PCA_2 was significant and could be used for estimation of coverage individually. However, the model based on PCA_1 act insignificantly (Sig.=0.062>0.05) and thus cannot be used individually.

3.1.4 Spectral Variables from Tasseled Cap Transformation

With the proposed new transformation matrix, three variables were acquired from the tasseled cap transformation and two of them (SBI and GVI) had significant correlation with measured coverage. Based on SBI and GVI, statistical models were established and shown in Fig. 6. The summary and ANOVA of them can be seen in Table 4. From them, we concluded that the two models were both significant and could be used for estimation. In fact, the nature of GVI is close to that of vegetation index.

3.2 Multiple-Variable-Based Modeling and Selection of Optimal Model

All above models are single-variable-based, multiple-variable-based models usually have a better modeling accuracy and estimation effects. Thus, stepwise regression analysis was conducted to estimate vegetation coverage based on SPSS 16, using the reflectance of band1, band2, and band3 of HJ-1 data, six vegetation indexes, two principal components from PCA as well as SBI, GVI from tasseled cap transformation as the input independent variables. The stepwise regression analysis can automatically select the most significant variable(s) and produce the "best" regression model(s). The probability of F test in the model was set for variables to enter at ≤0.05 while to remove at ≥0.10. The mode outputted by stepwise regression analysis was shown in Table 4.

From Table 4, we found that only two variables of NDVI and GVI entered the multiple regression model while others removed because of insignificant impact or collinearity between variables. Here, based on SPSS 16.0, we also performed the test of normality of the error variance using Kolmogorov–Smirnov and Shapiro–Wilk statistics and the test of homoscedasticity using Levene statistic. The testing results showed that the three variables of NDVI, GVI, and measured coverage which were incorporated in the multiple regression model significantly met those assumptions. It made the multiple-variable-based model effective and reliable.

From the comparison and simple analysis of all single-variable-based models, we found that the model based on NDVI had the highest R (0.900) and R^2 (0.810) as well as lowest SEE (0.128), thus showed the best effect. But when taking the multiple-variable-based model into account, we concluded that the

Table 5 The detailed evaluation coefficients of the optimal model

Independent variable	Dependent variable	Unstandardized coefficients		T	Sig.
		B	Standard error		
Constant	Measured coverage	−0.874	0.121	−7.197	0.000[a]
NDVI		8.689	1.314	6.611	0.000[a]
GVI		−11.193	3.174	−3.526	0.003[a]

[a]Extremely significant correlation (Sig.≤0.01)

multiple-variable-based model behaved even much better (higher R and R^2 and lower SEE). Thus finally it was determined as the optimal model for estimation considering all the evaluation coefficients. This model was expressed as Eq. 12 and the detailed evaluation coefficients were shown in Table 5.

$$\text{coverage} = -0.874 + 8.689\ \text{NDVI} - 11.193\ \text{GVI} \quad (12)$$

3.3 Estimation and Validation of Vegetation Coverage

Based on the selected optimal model from multiple stepwise regression analysis, the vegetation coverage of the study area was estimated and the distribution was mapped as shown in Fig. 7. In this figure, the white marks area with no vegetation, which are covered by bare land, water and buildings, and so on. The light green areas indicate the vegetation cover lower than 0.25 while the red areas indicate a value higher than 0.75. By reference to original remote sensing image and Google Earth image of the study area, the estimation results are accordant with the actual situation.

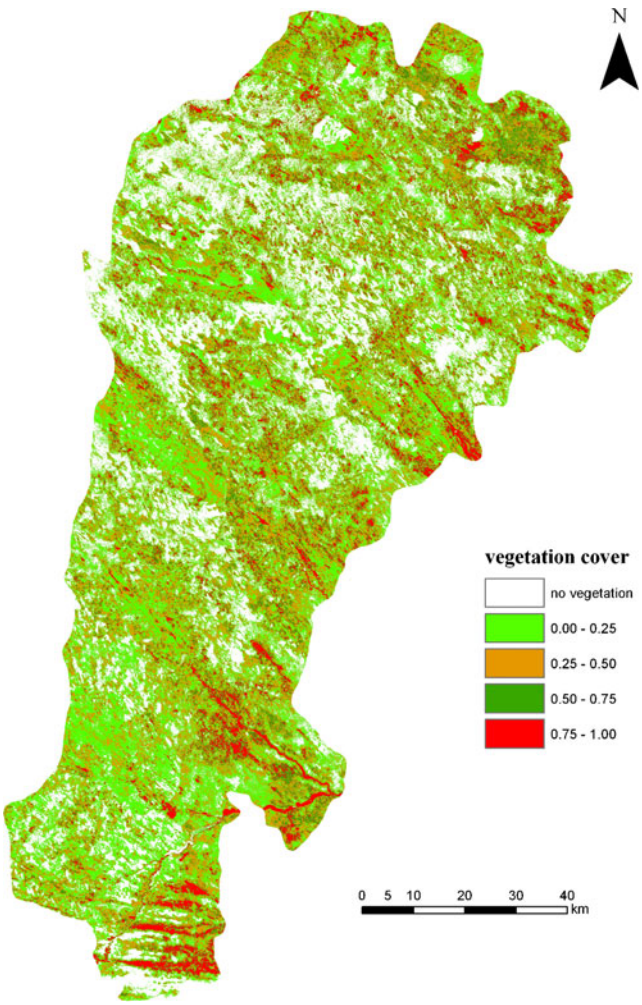

Fig. 7 The distribution map of estimated vegetation coverage in the study area

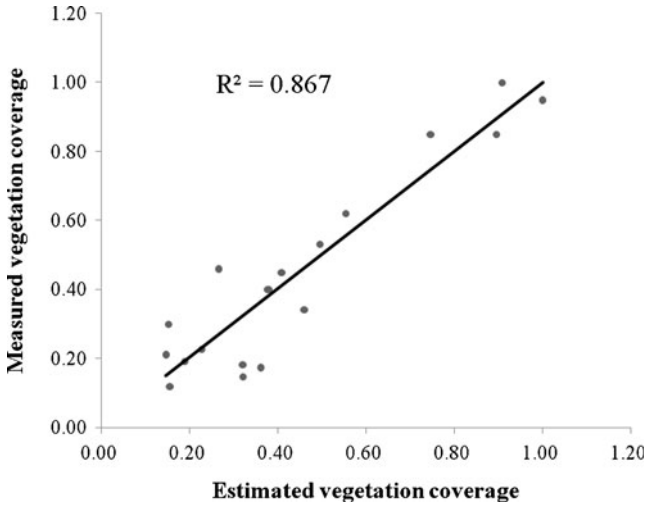

Fig. 8 The cross-validation of estimated vegetation coverage

Additionally, in order to illustrate the effect of the estimation result adequately, quantitative validation was conducted using measured coverage values of 19 field plots. As these data had also been used in modeling process, the approach of cross-validation was used to ensure the validity of the verification. The cross-validation is done through "leave-one-out" method. It involves iteratively removing one plot of vegetation coverage from the full set, fitting a regression equation to the remainder of the plots, and applying the equation to predict the value for the "left-out" plot. This work is repeated for each plot in the sample to produce a set of plot-based estimated vegetation coverage.

The validation result was shown in Fig. 8. The determination coefficient (R^2) of cross-validation was 0.867 with a RMSE of 0.101. For estimation of vegetation coverage, it was a quite low value. Besides, the significant probability of 0.000 also indicated no significant difference between measured and estimated coverage. All these conclusions demonstrated that the selected optimal model is applicable for local coverage estimation and the result was valid and useful for further applications and actual environment management.

4 Discussion

Vegetation coverage, which is a basic and important environmental parameter, serves as a crucial input variable of a large number of ecological and hydrological models and thus attracts much attention. This study provides an example on its estimation based on multivariate statistical modeling using remote sensing data. Although some positive

experience has been acquired, there are still many aspects which could be discussed and improved.

Firstly, during the modeling process, the NDVI performed excellent and entered the stepwise multiple-variable-based model. It was a quite normal phenomenon as NDVI was one of the oldest, most well known, and widely used vegetation indexes. The initiative of modeling lay in the acquisition and using of GVI. Here, based on the modified tasseled cap transformation, a variable of GVI was got from HJ-1 data for the first time. In fact, we were surprised and excited that it had significant correlation with measured coverage and also been incorporated in the optimal model for coverage estimation. Does it fit for other vegetation applications? More cases are in need for further validation in future studies.

Then as current statistical regression models are restricted to a specific study area, we need to adjust the coefficients or try to build physical model for other areas.

Additionally, there were only 19 effective sample plots investigated for studies because of the hot summer weather and hard sampling conditions in sandy land. There is no doubt that we have to add more sample plots for modeling and validation in future work.

Finally, the optical image used in this study is only the HJ-1 multispectral data which has some limitations in the spectral resolution and spatial resolution. In future studies, we need to try other multispectral images such as Landsat TM/ETM+, Advanced Land Observing Satellite as well as hyperspectral data (MODIS, etc.), and explore their potential for the estimation of coverage and many other ecological parameters. As more and more active remote sensing techniques, such as light detection and ranging (LiDAR) and synthetic aperture radar (SAR) have been widely and successfully used in various applications, the synergy estimation combining optical images with LiDAR or SAR data should be considered.

Acknowledgments The authors are grateful to China Centre for Resources Satellite Data and Application (CRESDA) for providing the HJ-1 data. Thanks to professor Mohamed Musbah for providing language editing on the manuscript. The authors also thank all people that have given help on the field survey.

References

1. Zhao, Y. S. (2003). *Remote sensing applications and its principles*. Beijing: Science Press.
2. Ju, C. Y., Cai, T. J., & Yang, X. H. (2008). Topography-based modeling to estimate percent vegetation cover in semi-arid Mu Us sandy land, China. *Computers and Electronics in Agriculture, 64*, 133–139.
3. Zhu, L., Xu, J. F., Huang, J. F., Wang, F. M., Liu, Z. Y., & Wang, Y. (2008). Study on hyperspectral estimation model of crop vegetation cover percentage. *Spectroscopy and Spectral Analysis, 28*, 1827–1831.
4. Barlage, M., & Zeng, X. B. (2004). The effects of observed fractional vegetation cover on the land surface climatology of the community land model. *Journal of Hydrometeorology, 5*, 823–830.
5. Zeng, X. B., Rao, P., Defries, R. S., & Hansen, M. C. (2003). Interannual variability and decadal trend of global fractional vegetation cover from 1982 to 2000. *Journal of Applied Meteorology, 42*, 1525–1530.
6. Jiang, X. G., Wang, D., Tang, L. L., Hu, J., & Xi, X. H. (2008). Analyzing the vegetation cover variation of China from AVHRR-NDVI data. *International Journal of Remote Sensing, 29*, 5301–5311.
7. Chen, W., Cao, C. X., He, Q. S., Guo, H. D., Zhang, H., Li, R. Q., et al. (2010). Quantitative estimation of the shrub canopy LAI from atmosphere-corrected HJ-1 CCD data in Mu Us Sandland. *Science in China-Earth Science, 53*, 26–33.
8. Chen, Y., & Gillieson, D. (2009). Evaluation of Landsat TM vegetation indices for estimating vegetation cover on semi-arid rangelands: a case study from Australia. *Canadian Journal of Remote Sensing, 35*, 435–446.
9. Wu, B., & Ci, L. J. (2001). Temporal and spatial patterns of landscape in the Mu Us Sandland, Northern China. *Acta Ecologica Sinica, 21*, 191–196.
10. Cao, C. X., Bao, Y. F., Xu, M., Chen, W., Zhang, H., He, Q. S., et al. (2012). Retrieval of forest canopy attributes based on Geometric-Optical model using airborne LiDAR and optical remote sensing data. *International Journal of Remote Sensing, 33*, 692–709.
11. Archer, E. R. M. (2004). Beyond the "climate versus grazing" impasse: using remote sensing to investigate the effects of grazing system choice on vegetation cover in the eastern Karoo. *Journal of Arid Environments, 57*, 381–408.
12. Xu, M., Cao, C. X., Tong, Q. X., Zhang, H., He, Q. S., Gao, M. X., et al. (2010). Remote sensing based shrub above-ground biomass and carbon storage mapping in Mu Us desert, China. *Science in China-Technical Science, 53*, 176–183.
13. Liu, B., Yue, Y. M., Li, R., Wang, K. L., Zhang, B., & Tong, Q. X. (2010). Study on the relation between fraction cover and mixed spectra in karst environment. *Spectroscopy and Spectral Analysis, 30*, 2470–2474.
14. Chen, Z. H., Chen, W. J., Leblanc, S. G., & Henry, G. H. R. (2010). Digital photograph analysis for measuring percent plant cover in the Arctic. *Arctic, 63*, 315–326.
15. Pisek, J., & Chen, J. M. (2009). Mapping forest background reflectivity over North America with multi-angle imaging spectroradiometer (MISR) data. *Remote Sensing of Environment, 113*, 2412–2423.
16. Stojanova, D., Panov, P., Gjorgjioski, V., Kohler, A., & Dzeroski, S. (2010). Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecological Informatics, 5*, 256–266.
17. Gillespie, T., Foody, G., Rocchini, D., Giorgi, A., & Saatchi, S. (2008). Measuring and modeling biodiversity from space. *Progress in Physical Geography, 32*, 203–221.
18. Hall, K., Reitalu, T., Sykes, M. T., & Prentice, H. C. (2012). Spectral heterogeneity of QuickBird satellite data is related to fine-scale plant species spatial turnover in semi-natural grasslands. *Applied Vegetation Science, 15, 145–157.*
19. Buyantuyev, A., Wu, J., & Gries, C. (2007). Estimating vegetation cover in an urban environment based on Landsat ETM+ imagery: a case study in Phoenix, USA. *International Journal of Remote Sensing, 28*, 269–291.
20. Levesque, J., & Staenz, K. (2008). Monitoring mine tailings revegetation using multitemporal hyperspectral image data. *Canadian Journal of Remote Sensing, 34*, 5172–5186.
21. Amiri, R., Weng, Q. H., Alimohammadi, A., & Alavipanah, S. K. (2009). Spatial-temporal dynamics of land surface temperature in relation to fractional vegetation cover and land use/cover in the

Tabriz urban area, Iran. *Remote Sensing of Environment, 113*, 2606–2617.
22. Smith, A. M. S., Falkowski, M. J., Hudak, A. T., Evans, J. S., Robinson, A. P., & Steele, C. M. (2009). A cross-comparison of field, spectral, and Lidar estimates of forest canopy cover. *Canadian Journal of Remote Sensing, 35*, 447–459.
23. Rastmanesh, F., Moore, F., Kharrati-Kopaei, M., & Behrouz, M. (2010). Monitoring deterioration of vegetation cover in the vicinity of smelting industry, using statistical methods and TM and ETM+ imageries, Sarcheshmeh copper complex, Central Iran. *Environmental Monitoring and Assessment, 163*, 397–410.
24. Borchard, P., & Eldridge, D. (2012). Vegetation changes associated with cattle (*Bos taurus*) and wombat (*Vombatus ursinus*) activity in a riparian forest. *Applied Vegetation Science, 15*, 62–70.
25. Kotchenova, S. Y., Vermote, E. F., Levy, R., & Lyapustin, A. (2008). Radioactive transfer codes for atmospheric correction and aerosol retrieval: intercomparison study. *Applied Optics, 47*, 2215–2226.
26. Zhang, T. L., & Lin, G. (2007). A decomposition of Moran's I for clustering detection. *Computational Statistics and Data Analysis, 51*, 6123–6137.
27. Chen, W. (2011). *Retrieval and Comparative Analysis of Vegetation Fractional Cover in Mu Us Sandy land*. Master thesis, Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China.
28. Graetz, R. D., Pech, R. R., & Davis, A. W. (1988). The assessment and monitoring of sparsely vegetated rangelands using calibrated Landsat data. *International Journal of Remote Sensing, 7*, 1201–1222.
29. Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*, 295–309.
30. Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment, 62*, 241–252.
31. Guo, N. (2003). Vegetation index and its advances. *Arid Meteorology, 21*, 71–75.
32. Shibayama, M., & Akiyama, T. (1989). seasonal visible, near-infrared and mid-infrared spectra of rice canopies in relation to LAI and above-ground dry biomass. *Remote Sensing of Environment, 27*, 119–127.
33. Chen, J. M., & Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote Sensing of Environment, 55*, 153–162.
34. Jordan, C. F. (1969). Derivation of leaf area index from quality of light on the forest floor. *Ecology, 50*, 663–666.
35. Pearson,R.L. & Miller,L.D. (1972). Remote Mapping of Standing Crop Biomass for Estimation of the Productivity of the Shortgrass Prairie. In: *Proceedings of the Eighth International Symposium on Remote Sensing of Environment* (pp. 1355–1381). Willow Run Laboratories, Environmental Research Institute of Michigan, Ann Arbor, Michigan USA.
36. Rouse, J.W., Haas, R.H., Cehell, J.A, Deering, D.W. & Harlan, J.C. (1974). Monitoring vegetation system in the Great Plains with ERTS. In: *Proceedings of Third Earth Resources Technology Satellite-1 Symposium* (pp. 310–317). Greenblet: NASA/GSFC, Type III, Greenbelt, MD, USA.
37. Kaufman, Y. J., & Tanre, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing, 2*, 261–270.
38. Liu, H. Q., & Huete, A. R. (1995). A feedback based modification of the NDVI to minimize canopy background and atmosphere noise. *IEEE Transactions on Geoscience and Remote Sensing, 33*, 457–465.
39. Yu, X. L., & Ren, X. S. (1999). *Multivariate statistical analysis*. Beijing: China Statistics Press.
40. Horne, J.H. (2003). A tasseled cap transformation for IKONOS images. In: *ASPRS Annual Conference Proceedings*, Alaska, USA.
41. Lobser, S. E., & Cohen, W. B. (2007). MODIS tasseled cap: land cover characteristics expressed through transformed MODIS data. *International Journal of Remote Sensing, 28*, 5079–5101.
42. Perry, C. R., Jr., & Lautenschlager, L. F. (1984). Functional equivalence of spectral vegetation indices. *Remote Sensing of Environment, 14*, 169–182.
43. Crist, E. P., & Cicone, R. C. (1984). Application of the tasseled cap concept to simulated Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing, 50*, 343–352.