

Accuracy of land use change detection using support vector machine and maximum likelihood techniques for open-cast coal mining areas

Shivesh Kishore Karan · Sukha Ranjan Samadder

Received: 25 December 2015 / Accepted: 19 July 2016 / Published online: 27 July 2016
© Springer International Publishing Switzerland 2016

Abstract One objective of the present study was to evaluate the performance of support vector machine (SVM)-based image classification technique with the maximum likelihood classification (MLC) technique for a rapidly changing landscape of an open-cast mine. The other objective was to assess the change in land use pattern due to coal mining from 2006 to 2016. Assessing the change in land use pattern accurately is important for the development and monitoring of coalfields in conjunction with sustainable development. For the present study, Landsat 5 Thematic Mapper (TM) data of 2006 and Landsat 8 Operational Land Imager (OLI)/Thermal Infrared Sensor (TIRS) data of 2016 of a part of Jharia Coalfield, Dhanbad, India, were used. The SVM classification technique provided greater overall classification accuracy when compared to the MLC technique in classifying heterogeneous landscape with limited training dataset. SVM exceeded MLC in handling a difficult challenge of classifying features having near similar reflectance on the mean signature plot, an improvement of over 11 % was observed in classification of built-up area, and an improvement of 24 % was observed in classification of surface water using SVM; similarly, the SVM technique improved the overall land use

classification accuracy by almost 6 and 3 % for Landsat 5 and Landsat 8 images, respectively. Results indicated that land degradation increased significantly from 2006 to 2016 in the study area. This study will help in quantifying the changes and can also serve as a basis for further decision support system studies aiding a variety of purposes such as planning and management of mines and environmental impact assessment.

Keywords Change detection · Coal mining · High-resolution images · Land degradation · Maximum likelihood classification · Support vector machines

Introduction

Due to its low cost and abundance in the nature when compared to other fuels, especially in the case of electricity generation, coal remains one of the most important energy sources (Smith 1997). However, coal mining has created a number of environmental challenges as it causes irreversible damage to the surrounding topography including the water regime. Coal mining causes subsidence, air pollution, and contamination of soil (WCA 2014). Land degradation due to land cover change in the vicinity of mining areas is one of the most important factors affecting ecological systems at the local scale. Monitoring these changes in land cover is essential in mapping the latest developments, especially in case of open-cast mines, where topographical changes are frequent. For effective planning and management of

S. K. Karan · S. R. Samadder (✉)
Department of Environmental Science and Engineering, Indian
School of Mines, Dhanbad 826004, India
e-mail: sukh_samadder@yahoo.co.in

S. K. Karan
e-mail: shivesh.karan@gmail.com

operations, mining areas require thorough monitoring of open-cast pits and degraded lands; remote sensing techniques can be highly instrumental in this regard (Demirel et al. 2011). Sustainable development prioritizing environmental accountability is the need of the hour.

Remote sensing and GIS tools have been used extensively in the mining industry for various purposes such as mineral exploration, modelling and monitoring, mine planning, and environmental impact assessment (Van der Meer et al. 2012; Karan and Samadder 2016). These tools provide a quick and cost-efficient means of mapping large geographic areas (Demirel et al. 2011). With the recent development in sensor and satellite technologies, the latest geospatial information has become more economical and readily available for users to exploit. Image classification is an elaborate procedure that may be affected by several factors, such as landscape, data type, image pre-processing and classification approaches (Lu and Weng 2007). Several image classification techniques are available, each with their own advantages and disadvantages. Support vector machine (SVM) is a recent non-parametric supervised statistical machine learning technique that aims to find an optimal hyperplane (Cortes and Vapnik 1995), which separates the multispectral feature data into discrete predefined clusters consistent with training datasets. SVM-based techniques are distinctly attractive in remote sensing due to their ability to produce higher classification accuracy even with limited training dataset (Mantero et al. 2005).

Demirel et al. (2011) employed SVM to identify, quantify, and analyse the spatial response of landscape change due to mining activities in Turkey. Their results indicated that SVM could be used efficiently in monitoring environmental impacts of mining with several constraints like remote locations in mountainous region and cloud cover. Otukei and Blaschke (2010) evaluated the performance of three different classification algorithms, decision trees, support vector machines and maximum likelihood in assessing the land cover change of Pallisa District, Eastern Uganda. It was reported in their study that although all the three techniques performed relatively well, SVM revealed an improvement of classification accuracy which was probably due to the simplification of vector space needed for development of hyperplanes. Hernandez et al. (2007) evaluated the use of SVM

classification in mapping priority habitat. Their results revealed an improvement of about 28 % in the overall classification accuracy when compared to maximum likelihood classification (MLC). They also reported that only a small training dataset was required for the definition of optimal separating hyperplane for isolating a class of interest from other classes. Size of training data plays an important role in the accuracy of classification and it varies for different types of satellite images. Selection of the optimum number of training samples may increase the overall classification accuracy (Foody et al. 2006). Several other studies have reported the effectiveness of SVM classification over other pixel-based techniques for variety of purposes like biophysical tasks, land use and land cover tasks, and geomorphological tasks (Melgani 2006, Tang et al. 2008; Andermann and Gloaguen 2009; Cao et al. 2009a; Knorn et al. 2009; Knudby and LeDrew, 2010).

In the present study, we evaluated the effectiveness of support vector machine-based land use classification over a conventional classification mechanism. The present study area is densely populated around the mining sites; this presents a problem while classifying land cover accurately due to the presence of built-up area. The other objective was to study the change in land use pattern over a period of 10 years due to extensive open-cast coal mining. Accurate assessment of this land use change in mining areas is important for planning of operations and mine closure.

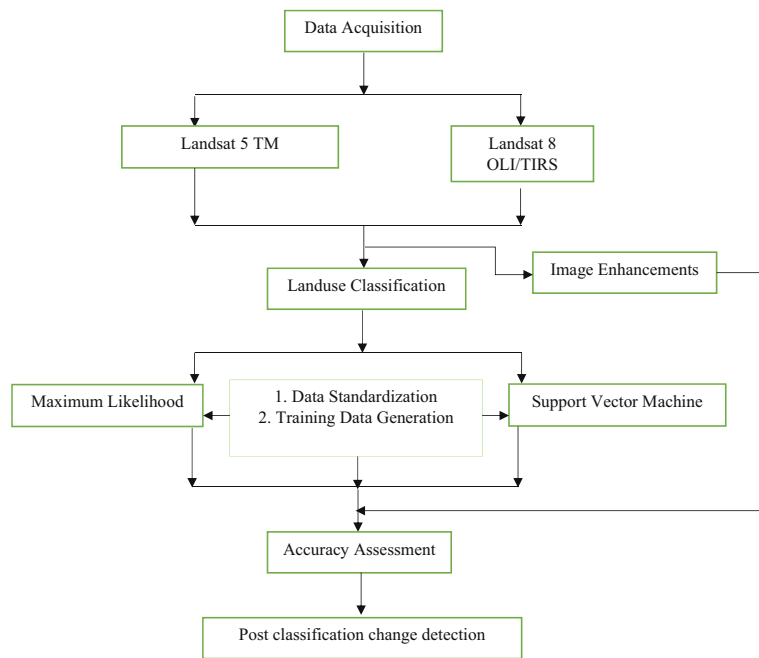
Materials and methods

Figure 1 depicts the overall methodology of the present study starting from data acquisition, followed by land use classification using maximum likelihood (ML) and SVM-based methods. After that, accuracy assessment was carried out to validate the classification results and to confirm which one of the two classification mechanisms was more accurate. After the validation part, change detection was performed using the results of classified images.

Study area

A part of the Jharia Coalfield, Dhanbad, India, has been selected for the present study to investigate the impact of coal mining on land use and land cover. The study area (Fig. 2) is situated at about 250 km west of Kolkata,

Fig. 1 Methodology flow chart of the present study



India, and lies between latitudes 23° 38' N and 23° 50' N and longitudes 86° 07' E and 86° 30' E. The Jharia Coalfield is regarded as the coal capital of India and is known as the exclusive storehouse of prime coking coal in the country. Due to extensive and unregulated coal mining over a century, the study area has witnessed unrestricted environmental degradation over the years (Prakash and Gupta 1998; Sarkar et al. 2007). According to the Central Pollution Control Board's (CPCB) Comprehensive Environmental Pollution Index (CEPI), Dhanbad is the 24th most critically polluted city of India (CPCB 2009). The continuous mining in the past has changed the landscape with remnants of old abandoned quarries, spoil dumps, subsided depressions and soil patches baked due to mine fires.

Data acquisition

In order to evaluate the change in land use/land cover over a period of 10 years, two different multispectral satellite images were procured. Landsat 5 Thematic Mapper (TM) surface reflectance data of 30 April 2006 and Landsat 8 Operational Land Imager (OLI)/Thermal Infrared Sensor (TIRS) surface reflectance data of 25 April 2016 were collected from USGS EarthExplorer (Table 1, Fig. 3a, b). Near anniversary images were collected to reduce the potential change detection errors due to variability arising from sun angle,

atmospheric condition and phenology. Both images were standardized and projected to the same projection system, Universal Transverse Mercator (UTM) Zone 45 N Datum WGS 1984 projection using ArcGIS 10.2.2. The specification of the satellite images is presented in Table 1.

Image enhancements

The primary objective of any image enhancement technique is to improve the visual interpretability of an image by increasing the probable distinction among the features in the scene (Lillesand et al. 2014). If two different features are of same colour, their isolation may become difficult, but if those features are sharply different in tone or brightness, their separation becomes easier. For the present study, some of the surface features had near similar reflectance on the mean signature plot (e.g. built-up and OB dump). Haze reduction and histogram equalization techniques were performed using ERDAS Imagine. These techniques helped in reducing the additive effect caused by atmospheric haze and thus improving the apparent distinction among the surface features. The enhanced images were used in visual analysis for the pseudo accuracy assessment study as no other reference data was available for the image of 2006.

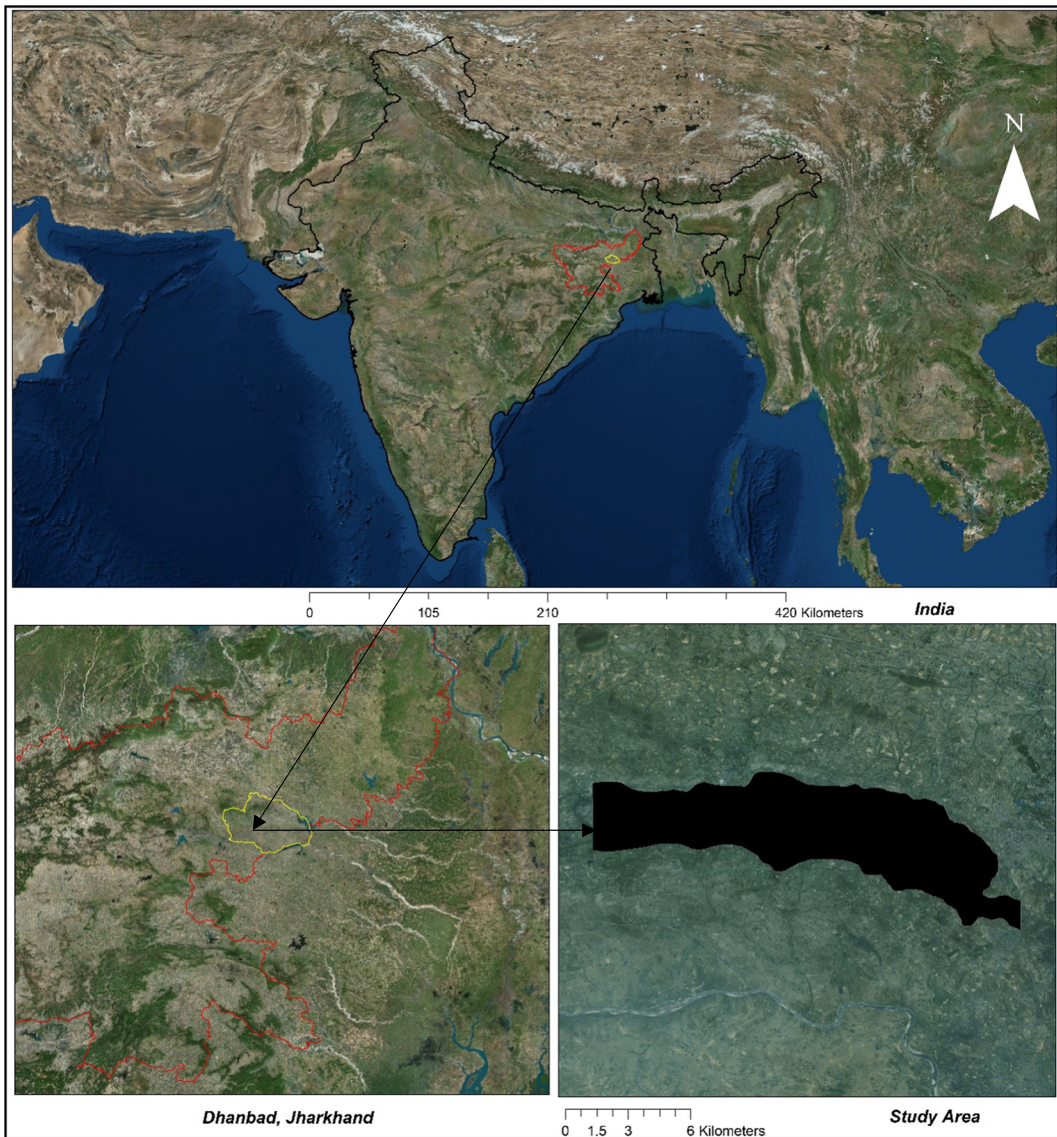
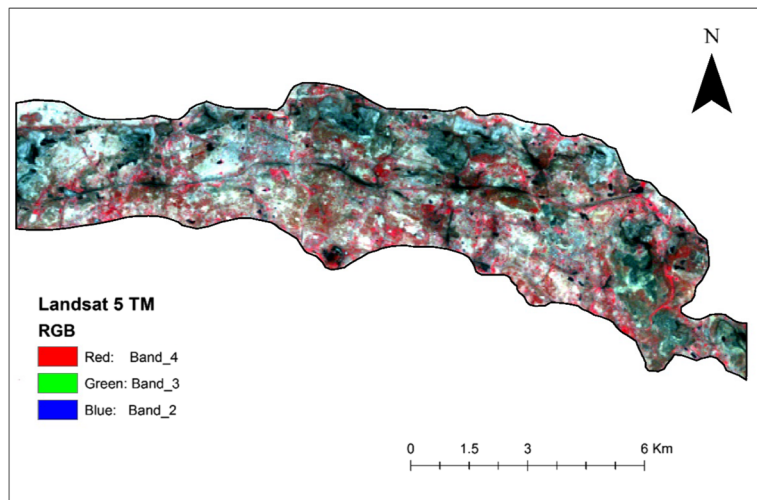


Fig. 2 Location of the study area

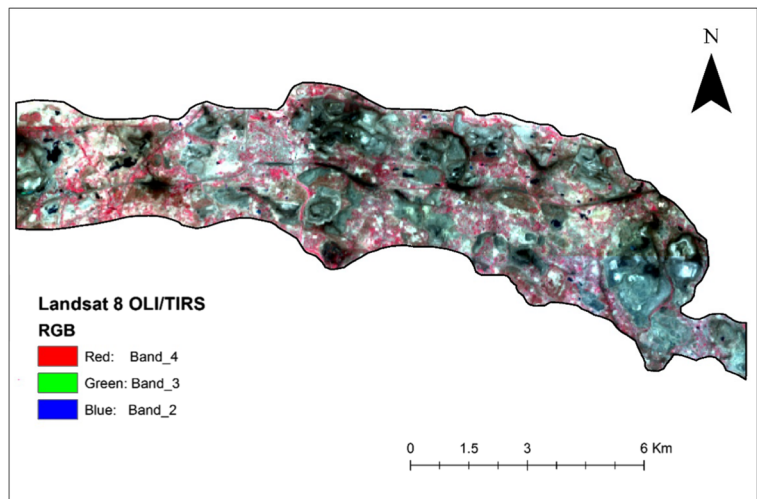
Table 1 Properties of satellite images

	Landsat 5 TM	Landsat 8 OLI/TIRS
Acquisition date	30 April 2006	25 April 2016
Metadata	Yes	Yes
Available number of bands	7	11
Spatial resolution (in m)	MULX 30	MULX 30
Spectral resolution (in μm)	Blue: (0.45–0.51) Green: (0.52–0.60) Red: (0.63–0.69) NIR: (0.76–0.90)	Blue: (0.45–0.51) Green: (0.53–0.59) Red: (0.64–0.67) NIR: (0.85–0.88)

Fig. 3 **a** Landsat 5 TM satellite image (year 2006). **b** Landsat 8 OLI/TIRS satellite image (year 2016)



a



b

Maximum likelihood classification

Maximum likelihood is one of the most commonly used classification algorithms and is dependent upon the probability distribution of the feature classes as per Bayes’ theorem (Liu and Mason 2013). In this algorithm, the normal probability distributions for each of the spectral classes are demarcated using a covariance matrix by selecting a sufficient number of pixels in each spectral class as training sample for the classification algorithm (Richards and Jia 2006). In MLC, the feature space distance between cluster ω_k and the image pixel Y_i are weighted by the covariance matrix Σ_k of ω_k with an offset relating to ratio N_k given in Eq. (2) (Liu and Mason 2013). The algorithm for maximum likelihood

classification is given in Eq. (1). N is the total number of pixels in the image Y . For all i , element Y_i ($Y_i \in Y$) to cluster ω_k if

$$\partial(Y_i, \omega_k) = \ln \left| \Sigma_k \right| + (Y_i - \mu_k)^T \Sigma_k^{-1} (Y_i - \mu_k) \tag{1}$$

$$-\ln \frac{N_k}{N} = \min \{ \partial(Y_i, \omega_r) \} \tag{2}$$

for $r = 1, 2, \dots, m$.

Both images were arranged in false colour composite (FCC) band order, being a general case of an RGB colour display, FCC images effectively highlight vegetation

distinctively in red (Liu and Mason 2013). ArcGIS 10.2.2 was used to perform MLC by taking seven broad land cover classes based on their representativeness in the image for the present study area. Training sets were generated manually by taking an appropriate number of training samples in each representative class and then merging them in training sample manager. A representative signature file was created for training and classifying the images.

Support vector machine-based classification

Support vector machine (SVM) technique is based on statistical learning theory presented by Cortes and Vapnik (1995). SVM classifier is a modern, powerful supervised classification method that can handle multiple-band imagery having high resolution and large segmented satellite data with ease when compared to other classification techniques, where attribute table management becomes difficult. Furthermore, SVM is a non-parametric learning technique; hence, no assumptions are made on the underlying data distributions (Fauvel et al. 2009). SVM works on the principle of generation of a hyperplane that epitomizes a precise separation of linearly separable classes in a hypersurface (Szuster et al. 2011). One major advantage of SVM over other classification algorithms is that it is less gullible to noise, correlated bands and an unbalanced number or size of training data within each class (Cawley and Talbot 2010). The purpose of SVM is to create a model (based on training data) that predicts the destination value of testing data from the training data attributes (Hsu et al. 2010). From the training set of instance-label pairs (X_i, Y_i) , $i = 1, \dots, l$ where $X_i \in R^n$ and $Y \in \{1, -1\}^l$, the SVM (Boser et al. 1992; Cortes and Vapnik 1995) require the solution for the optimization problem given in Eq. (3) and Eq. (4).

$$\min_{w,b,\epsilon} \frac{1}{2} W^T W + C \sum_{i=1}^l \epsilon_i \quad (3)$$

$$\begin{aligned} Y_i (W^T \phi(X_i) + b) &\geq 1 - \epsilon_i, \\ \text{Subject to,} & \\ \epsilon_i &\geq 0. \end{aligned} \quad (4)$$

In the above equation, the training vectors X_i are charted into a higher dimensional space by the function ϕ . The SVM locates a linear separating hyperplane

with the maximum margin in this higher dimensional space (Hsu et al. 2010). $C > 0$ is the error term penalty parameter.

ArcGIS 10.2.2 along with ArcPy tool (Wehmann 2013) using LIBSVM classification library (Chang and Lin 2011) with radial basis function (RBF) kernel given in Eq. (5) was used to perform SVM classification on both images. As SVM utilizes a maximum margin hyperplane for a decision boundary, only a part of the training data, known as support vectors are needed to describe its position. The RBF kernel non-linearly maps training samples to a higher dimensional space.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0. \quad (5)$$

The tool uses a training raster to learn how to produce the classified image. In the training raster, a small number of pixels are assigned to each training class. This is done by coding the land cover classes with numbers, and then changing the value of a number of pixels in the training raster which overlay that land cover type in the satellite image to the code of that class. All other remaining pixels are set to 0, in order to establish that the classifier should predict their classes based on the information that is extracted. Training and testing data should be a single band raster with known pixels having a value in the label set $\{1, \dots, n\}$ and unknown pixels are assigned a value of 0. These raster images should have the exact same extent and gridding as the input raster. This can be ensured by setting Extent, Snap Raster and Cell Size options appropriately in the environment settings during production. The sparse representation of training data along with non-linear mapping provides robust classification accuracy as compared to other classification techniques.

Accuracy assessment

After the land use classification process, it is important to assess the accuracy of the classified image, to peg and quantify mapping or classification errors. Several techniques are available for accuracy assessment study (Arnoff 1982; Piper 1983; Kalkhan et al. 1995; Rosenfield and Fitzpatrick-Lins 1986; Koukoulas and Blackburn 2001). The most commonly used technique is derived using a confusion or error matrix; for the present study, the same has been used. In this method, a simple cross tabulation of the mapped class label against that observed in the ground or reference data is

employed for a sample of cases at specified locations (Canters 1997). ArcGIS 10.2.2 was used for accuracy assessment study. Pseudo accuracy assessment was performed for the classified image of 2006 as no other reference data was available for that year. In pseudo accuracy assessment, the same data which was used in the classification process was enhanced and used as a reference for the classified images of 2006. Google Earth™, LISS IV high-resolution satellite image along with pan-sharpened Landsat 8 data of 2016 was used as a reference for the classified image of 2016. Random accuracy points were generated using the Create Accuracy Assessment Point tool available in ArcGIS. The tool computes the user’s and producer’s accuracy for each of the classes and calculates an overall kappa index to accommodate the effects of chance agreement. The locational accuracy for the reference data of 2006 was 100 % as it employed the same dataset for the accuracy

assessment study. The locational accuracy for the reference data of 2016 was limited to 5 m (spatial resolution of the LISS IV image). The error arising from the locational accuracy of the 2016 reference data was considered to be negligible as the reference data had higher resolution than the test dataset.

Change detection

Post classification change detection (CD) technique was used for assessing the change in land use pattern over the decade. This technique is often used as a benchmark for the subjective assessment of other CD techniques (Lunetta et al. 1999). During comparison, if the corresponding pixels lie in the same class label, the pixel has not been changed, or else the pixel has been changed. One major limitation of this technique is that it may also incorporate classification errors in the final CD map (Xu

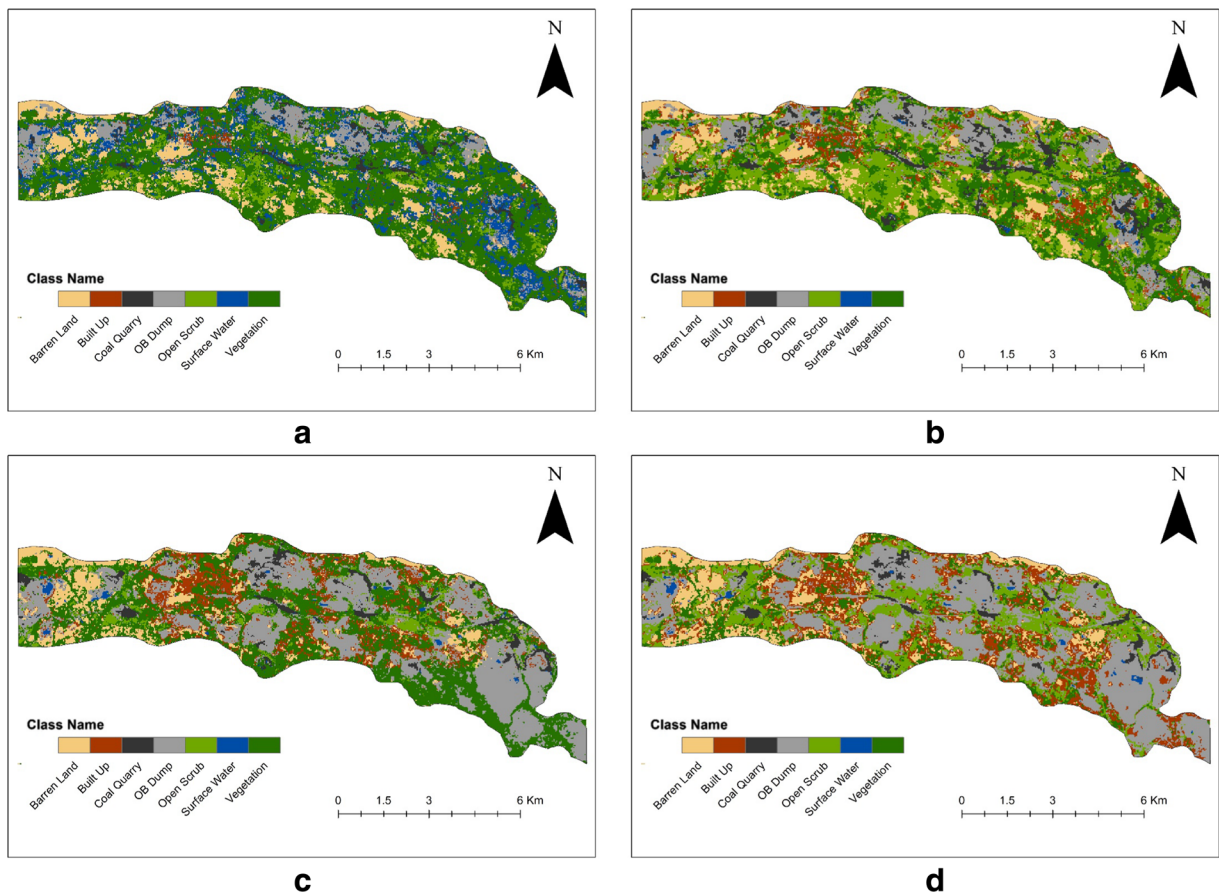


Fig. 4 **a** Maximum likelihood-based land use classification using the Landsat 5 TM image. **b** Support vector machine-based land use classification using the Landsat 5 TM image. **c** Maximum

likelihood-based land use classification using the Landsat 8 OLI/TIRS image. **d** Support vector machine-based land use classification using the Landsat 8 OLI/TIRS image

Table 2 Land use change detection from 2006 to 2016 based on the SVM and MLC classification techniques (total area 6541.23 ha)

Classes	Maximum likelihood (values in percent area)			Support vector machine (values in percent area)		
	Year		Relative change in percentage area	Year		Relative change in percentage area
	2006	2016		2006	2016	
Surface water	15.11	0.83	-94.50	0.85	0.69	-18.82
OB dump	11.88	34.10	187.03	17.66	37.61	112.96
Built-up	1.68	11.08	559.52	8.51	17.43	104.81
Coal quarry	3.16	3.77	19.30	6.18	3.33	-46.11
Vegetation	47.84	35.17	-26.48	25.94	10.70	-58.75
Open scrub	7.44	4.56	-38.70	28.81	18.32	-36.41
Barren land	12.89	10.49	-18.61	12.05	11.92	-1.07

et al. 2009; Hussain et al. 2013). ERDAS Imagine was used in the present study to perform CD analysis. A model was created using ERDAS's modeller to quantify the change. Pixel values of 0 represented no change; decrease and increase were represented by change in pixel values, either negative change or positive change. Also, a change matrix was generated to understand the change from one land cover category to another due to coal mining activities.

Results and discussions

Results of maximum likelihood classification

Seven different land use classes (i.e. surface water, OB dump, built-up, coal quarry, vegetation, open

scrub, and barren land) were successfully delineated using this technique (Fig. 4a, c). The results revealed that for both the years, Vegetation dominated the land cover type with 47.84 % in 2006 and 35.17 % in 2016, respectively. Coal quarry was observed as 3.16 % in 2006 and 3.77 % in 2016. OB dump was found to be 11.88 % in 2006 and 34.10 % in 2016. Likewise, the distribution of all land use classes is presented in Table 2. From visual interpretation, it became obvious that this technique was producing classification errors, in the form of overestimating the surface water for the year 2006. This error was attributed to the near similar spectral reflectance of surface water and coal quarry and the limitation of the MLC classifier. The distribution of built-up area and OB dump were also erroneous due to the similar spectral reflectance of these two features.

Table 3 Accuracy assessment using the confusion matrix generated from the test training samples for Landsat 5 TM image based on the MLC technique (overall classification accuracy 84.98 %)

Sl. no.	Cover	1	2	3	4	5	6	7	Sum	User's accuracy (%)
1	Surface water	18	2	7	17	0	0	0	44	40.91
2	OB dump	0	281	7	2	0	0	0	290	96.90
3	Built-up	0	6	17	0	0	0	0	23	73.91
4	Coal quarry	4	0	0	107	0	0	0	111	96.40
5	Vegetation	2	0	1	1	157	9	0	170	92.35
6	Open scrub	0	0	0	0	0	49	0	49	100.00
7	Barren land	0	6	0	0	0	0	213	219	97.26
	Sum	24	295	32	127	157	58	213		
	Producer's accuracy (%)	75	95.25	53.13	84.25	100	84.48	100		

Table 4 Accuracy assessment using the confusion matrix generated from the test training samples for Landsat 5 TM image based on the SVM classification technique (overall classification accuracy 91.20 %)

Sl. no.	Cover	1	2	3	4	5	6	7	Sum	User's accuracy (%)
1	Surface water	17	0	0	1	0	0	0	18	94.44
2	OB dump	1	287	10	1	0	0	0	299	95.99
3	Built-up	0	4	22	0	0	1	0	27	81.48
4	Coal quarry	4	0	0	125	0	0	0	129	96.90
5	Vegetation	1	0	0	0	153	5	0	159	96.23
6	Open scrub	1	0	0	0	4	52	0	57	91.23
7	Barren land	0	4	0	0	0	0	213	217	98.16
	Sum	24	295	32	127	157	58	213		
	Producer's accuracy (%)	70.83	97.29	68.75	98.43	97.45	89.66	100		

Results of support vector machine-based classification

The results of SVM classification revealed 112.96 % increase in the area coverage of OB dump (increased from 1155.42 ha in 2006 to 2460.8 ha in 2016). There was also a sharp increase in the area coverage of built-up area which increased from 8.51 % in 2006 to 17.43 % in 2016, showing an increase of 104 % (Table 2, Fig. 4b, d). Coal quarry was observed to be 6.18 % in 2006 and 3.33 % in 2016 respectively. Likewise, the distribution of other land use classes (*barren land, surface water, open scrub, and vegetation*) is presented in Table 2. Visual interpretation revealed higher classification accuracy with this method in comparison to MLC. The difference in classification accuracy may be explained by various factors like image quality, image resolutions, classification errors, software errors, and user errors. For the present study, SVM not only gave better results than MLC in classifying all the linearly separable classes but

it also produced better results while training the classes which were having near similar spectral reflectance(s). As SVM is based on the principle of separating hyperplanes, high-classification accuracy is attributed to SVM's ability to locate an optimal separating hyperplane.

Results of accuracy assessment

Initial attention was focussed on the accuracy of support vector machine-based classification over maximum likelihood classification technique. Confusion matrices revealed that SVM-based classifier gave better results than MLC technique by almost 6 % for the Landsat 5 TM image of 2006. Using MLC technique, the overall classification accuracy was found to be 84.98 % with kappa coefficient of 0.82 (Table 3) for Landsat 5 TM image of 2006, and using SVM, the overall classification accuracy was

Table 5 Accuracy assessment using the confusion matrix generated from the test training samples for Landsat 8 OLI/TIRS image based on the MLC technique (overall classification accuracy 90.7 %)

Sl. no.	Cover	1	2	3	4	5	6	7	Sum	User's accuracy (%)
1	Surface water	119	0	0	9	0	0	0	128	92.97
2	OB dump	7	300	1	1	0	1	0	310	96.77
3	Built-up	1	51	78	0	0	0	0	130	60.00
4	Coal quarry	0	0	0	55	0	0	0	55	100.00
5	Vegetation	11	0	0	0	97	6	0	114	85.09
6	Open scrub	0	0	0	0	1	105	0	106	99.06
7	Barren land	3	13	0	0	0	0	226	242	93.39
	Sum	141	364	79	65	98	112	226		
	Producer's accuracy (%)	84.40	82.42	98.73	84.62	98.98	93.75	100		

Table 6 Accuracy assessment using the confusion matrix generated from the test training samples for Landsat 8 OLI/TIRS image using the SVM classification technique (overall classification accuracy 93.37 %)

Sl. no.	Cover	1	2	3	4	5	6	7	Sum	User's accuracy (%)
1	Surface water	102	0	0	7	0	0	0	109	93.58
2	OB dump	38	350	12	0	0	0	0	400	87.50
3	Built-up	1	8	67	0	0	0	0	76	88.16
4	Coal quarry	0	0	0	58	0	0	0	58	100.00
5	Vegetation	0	0	0	0	97	0	0	97	100.00
6	Open scrub	0	0	0	0	1	112	0	113	99.12
7	Barren land	0	6	0	0	0	0	226	232	97.41
	Sum	141	364	79	65	98	112	226		
	Producer's accuracy (%)	72.34	96.15	84.81	89.23	98.98	100	100		

observed as 91.20 % with kappa coefficient of 0.89 (Table 4) for Landsat 5 TM image of 2006.

The built-up area does not appear to have any particular diagnostic spectral reflectance and its spectral signature resembles to that of any other highly reflective features in the study area such as overburden dump. Therefore, it is expected that algorithms would not be able to convincingly distinguish classes having analogous spectral reflectance(s). Similarly, surface water also resembles to coal quarry having similar peaks in the mean signature plot. For Landsat 5 TM image, the built-up area was delineated with an overall accuracy (including producer and user accuracy) of 63.52 % using MLC technique. SVM offered a better result by improving the overall accuracy to 75.1 %, an improvement of 11.5 % over MLC technique. For surface water, the

accuracy difference was surprisingly even more substantial, increasing almost 24.6 from 57.95 % (using MLC) to 82.63 % (using SVM). SVM-based classifiers have the ability to generalize the unseen data and it is also known for its higher accuracy on limited amount of training patterns. This is why the SVM classifier has exceeded MLC, giving higher classification accuracy for every individual class.

In case of Landsat 8 OLI/TIRS image of 2016, confusion matrices revealed an increase of almost 3 % in the overall classification accuracy from 90.7 % with kappa coefficient of 0.88 for MLC (Table 5) to 93.37 % with kappa coefficient of 0.91 for SVM (Table 6). An improvement of about 7 % was observed in classification of built-up area. Moreover, Landsat 8 OLI/TIRS has finer spectral resolution (Table 1) than Landsat 5 TM; this

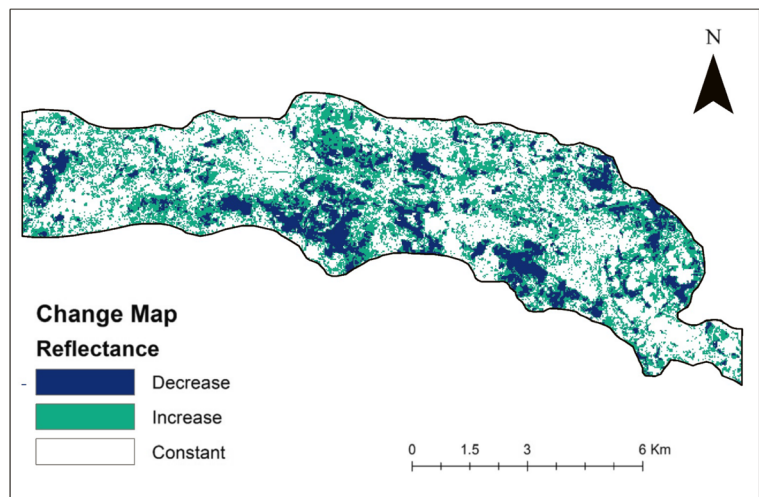
Fig. 5 Land use change map from 2006 to 2016 based on pixel reflectance values

Table 7 Land use change matrix from 2006 to 2016 (all figures are in ha)

	Area (ha)							2016 total
	Surface water	OB dump	Built-up	Coal quarry	Vegetation	Open scrub	Barren land	
Surface water	6.48	14.67	2.61	6.93	4.05	5.58	4.68	45.0
OB dump	34.11	765.72	159.48	228.24	358.56	726.39	188.28	2460.8
Built-up	3.42	101.34	234.27	7.11	306.45	331.47	156.42	1140.5
Coal Quarry	4.59	49.23	5.31	84.9	31.23	39.15	3.06	217.5
Vegetation	1.35	11.97	20.97	3.33	497.61	149.31	15.75	700.3
Open scrub	4.95	181.26	51.12	72.45	409.77	444.96	33.75	1198.3
Barren land	0.72	31.23	82.98	1.26	89.28	188.01	386.37	779.9
2006 total	55.62	1155.42	556.74	404.22	1696.95	1884.87	788.31	
Percent difference	-19.09	112.9	104.8	-46.2	-58.7	-36.4	1.06	

improved the apparent reflectance of different surface features. This is the reason why higher classification accuracies were observed using both techniques for Landsat 8 image. The improvement in classification accuracy was although close to 3 %, but considering the quality of the raw satellite image, this improvement may be termed substantial.

Results of change detection

Classification maps were generated for the years 2006 and 2016 using both SVM and MLC techniques (Fig. 4). SVM classified images reported higher accuracy for both 2006 and 2016, hence for post classification change detection, SVM classified images were used. A change detection (CD) map was created using ERDAS Imagine, the CD map was further sub-classified into three distinct classes (*decreased, increased, constant*) (Fig. 5). It was observed that due to continuous open-cast mining operations and high stripping ratio for coking coal, OB dump increased almost 112.96 %. Due to rapid coal extraction, large open scrub areas were converted into active mining land; as a result, open scrub area decreased nearly 36.41 % from 1884.87 ha in 2006 to 1198.3 ha in 2016 (Table 7). Surface water and vegetation both decreased by 18.82 and 58.75 %, respectively. Large tracts of Vegetation were converted into OB dump due to coal mining activities.

Statistical validation was not performed to evaluate the classification algorithms as individual performance of the algorithms cannot be established significantly (Dietterich 1998).

Conclusions

The main aim of this paper was to compare the accuracy of the two pixel-based classification techniques (support vector machine and maximum likelihood classification) in classifying satellite images of the Jharia Coalfield. Other objective was to study the change in land use pattern over a period of 10 years due to extensive opencast coal mining using Landsat 5 TM data of 2006 and Landsat 8 OLI/TIRS data of 2016. The ability of SVM to generate an optimal separating hyperplane resulted in a better performance of SVM over MLC in classifying high-resolution satellite imagery. Confusion matrix-based accuracy assessment revealed that the features with overlapping spectral reflectance values on the mean signature plot were classified more accurately using SVM with marginal errors, which was far better than MLC. SVM was superior to MLC in overall classification accuracy as well as individual classification accuracy for many classes. Post classification change detection was used to study the change in land use pattern over a period of 10 years from 2006 to 2016. A change detection map was prepared to analyse the change in land use pattern. It was observed during the study period that nearly 47 % of the total land area changed into other classes. The main drivers for this land use change for the present study area were coal mining activities. Large parts of land were converted into mining areas. The results also indicated that the rate of change of land cover is very high and the danger of severe land degradation is challenging the ecological resilience. This study will help in quantifying the land use change; in serving as a basis for further decision

support system, studies aiding variety of purposes such as environmental management, landscape planning, and monitoring reclamation success. Modern classification techniques such as SVM can be highly instrumental in monitoring land use and land cover change.

Acknowledgments The authors acknowledge the support provided by the Department of Environmental Science and Engineering, Indian School of Mines, Dhanbad, India, for carrying out the research work.

References

- Andermann, C., & Gloaguen, R. (2009). Estimation of erosion in tectonically active orogenies. Example from the Bhotekoshi catchment, Himalaya (Nepal). *International Journal of Remote Sensing*, 30, 3075–3096.
- Arnoff, S. (1982). Classification accuracy: a user approach. *Photogrammetric Engineering and Remote Sensing*, 48, 1299–1307.
- Boser, B. E., Guyon, I., & Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM Press, 144–152.
- Canter, F. (1997). Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification. *Photogrammetric Engineering and Remote Sensing*, 63, 403–414.
- Cao, X., Chen, J., Matsushita, B., Imura, H., & Wang, L. (2009a). An automatic method for burn scar mapping using support vector machines. *International Journal of Remote Sensing*, 30, 577–594.
- Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11, 2079–2107.
- Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(27), 1–27. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273–297.
- CPCB, (2009). Comprehensive Environmental Assessment of Industrial Clusters. Central Pollution Control Board, Ministry of Environment and Forests. December 2009. Retrieved from: www.cpcb.nic.in/divisionsofheadoffice/ess/NewItem_152_Final-Book_2.pdf
- Demirel, N., Emil, M. K., & Duzgun, H. S. (2011). Surface coal mining area monitoring using multi-temporal high resolution satellite imagery. *International Journal of Coal Geology*, 86, 3–11.
- Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7), 1895–1923.
- Fauvel, M., Chanussot, J., & Benediktsson, J. A., (2009). Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP Journal on Advances in Signal Processing*, Article ID: 783194
- Foody, G. M., Mathur, A., Hernandez, C. S., & Boyd, D. S. (2006). Training set size requirements for the classification of a specific class. *Remote Sensing of Environment*, 104, 1–14.
- Hsu, C. W., Chang, C. C., & Lin, C. J. L., (2010). Practical guide to support vector classification. Available at: <http://www.csie.ntu.edu.tw/~cjlin>. Accessed 26 May 2015).
- Hernandez, C. S., Boyd, D. S., & Foody, G. M. (2007). Mapping specific habitats from remotely sensed imagery: support vector machine and support vector data description based classification of coastal saltmarsh habitats. *Ecological Informatics*, 2, 88–88.
- Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: from pixel based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80, 91–106.
- Kalkhan, M. A., Reich, R. M., & Czaplowski, R. L. (1995). Statistical properties of five indices in assessing the accuracy of remotely sensed data using simple random sampling. *Proceedings ACSM/ASPRS Annual Convention and Exposition*, 2, 246–257.
- Karan, S. K., & Samadder, S. R. (2016). Reduction of spatial distribution of risk factors for transportation of contaminants released by coal mining activities. *Journal of Environmental Management*, 180, 280–290.
- Knorn, J., Rabe, A., Radeloff, V. C., Kuemmerle, T., Kozak, J., & Hostert, P. (2009). Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sensing of Environment*, 113, 957–964.
- Knudby, A., LeDrew, E., & Brenning, A., (2010). Predictive mapping of reef fish species richness, diversity and biomass in Zanzibar using IKONOS imagery and machine-learning techniques. *Remote Sensing of Environment*, 114, 1230 – 1241.
- Koukoulas, S., & Blackburn, G. A. (2001). Introducing new indices for accuracy evaluation of classified images representing semi-natural woodland environments. *Photogrammetric Engineering and Remote Sensing*, 67, 499–510.
- Lillesand, T., Kiefer, R. W., & Chipman, J., (2014). *Remote sensing and image interpretation*. John Wiley & Sons.
- Liu, J. G., & Mason, P., (2013). *Essential image processing and GIS for remote sensing*. John Wiley & Sons.
- Lu, D., & Weng, Q., (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28.
- Lunetta, R. S. (1999). Applications, project formulation, and analytical approach. In R. S. Lunetta & C. D. Elvidge (Eds.), *Remote sensing change detection: environmental monitoring methods and applications* (pp. 1–9). London: Taylor & Francis.
- Mantero, P., Moser, G., & Serpico, S. B. (2005). Partially supervised classification of remote sensing images through SVM-based probability density estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 43, 559–570.
- Melgani, F. (2006). Contextual reconstruction of cloud-contaminated multitemporal multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 442–455.
- Otukei, J. R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, 12S, S27–S31.
- Piper, S. E., (1983). The evaluation of the spatial accuracy of computer classification. In Proceedings of the 1983 Machine Processing of Remotely Sensed Data Symposium (303–310). West Lafayette: Purdue University.

- Prakash, A., & Gupta, R. P. (1998). Land-use mapping and change detection in a coal mining area—a case study in the Jharia Coalfield, India. *International Journal of Remote Sensing*, *19*, 391–410.
- Richards, J. A., & Jia, X., (2006). *Remote sensing digital image analysis—hardback*. Springer.
- Rosenfield, G. H., & Fitzpatrick-Lins, K. (1986). A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, *52*, 223–227.
- Sarkar, B. C., Mahanta, B. N., Saikia, K., Paul, P. R., & Singh, G. (2007). Geo-environmental quality assessment in Jharia Coalfield, India, using multivariate statistics and geographic information system. *Environmental Geology*, *51*(7), 1177–1196.
- Smith, A. H. V. (1997). Provenance of coals from Roman sites in England and Wales. *Britannia*, *28*, 297–324.
- Szuster, B. W., Chen, Q., & Borger, M. (2011). A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography*, *31*, 525–532.
- Tang, S., Chen, C., Zhan, H., & Zhang, T. (2008). Determination of ocean primary productivity using support vector machines. *International Journal of Remote Sensing*, *29*, 6227–6236.
- Van der Meer, F. D., Van der Werff, H. M., van Ruitenbeek, F. J., Hecker, C. A., Bakker, W. H., Noomen, M. F., et al. (2012). Multi-and hyperspectral geologic remote sensing: a review. *International Journal of Applied Earth Observation and Geoinformation*, *14*(1), 112–128.
- WCA, (2014). World Coal Association. Coal and the environment. www.worldcoal.org/coal-the-environment/coal-mining-the-environment. Accessed 24 June 2014.
- Wehmann, A., (2013). SVM tools for ArcGIS 10.1+. Columbus, OH, Software available at: <http://www.adamwehmann.com/>
- Xu, L., Zhang, S., He, Z., & Guo, Y., (2009). The comparative study of three methods of remote sensing image change detection. In *Geoinformatics, 2009 17th International Conference on* (pp. 1–4). IEEE.