Original Research

Landsat-5 Time Series Analysis for Land Use/Land Cover Change Detection Using NDVI and Semi-Supervised Classification Techniques

Syeda Maria Zaidi[1]*, Abolghasem Akbari[1], Azizan Abu Samah[2], Ngien Su Kong[1], Jacqueline Isabella Aanak Gisen[1, 3]

[1]Faculty of Civil Engineering & Earth Resources, University Malaysia Pahang (UMP), Malaysia
[2]Institute of Ocean and Earth Sciences (IOES) University of Malaya (UM), Malaysia
[3]Centre for Earth Resources and Research Management (CERRM), UMP, Malaysia

Abstract

Rapid urbanization and the risk of climatic variations, including a rise in temperature and increased rainfall, have urged research in the development of methods and techniques to monitor the modification of land use/land cover (LULC). This study employed the normalized differencing vegetative index (NDVI) and semi-supervised image classification (SSIC) integrated with high-resolution Google Earth images of the Kuantan River Basin (KRB) in Malaysia. The Landsat-5 (TM) images for the years 1993, 1999, and 2010 were selected. The results from both classifications provided a consistent accuracy of assessment with a reasonable level of agreement. However, SSIC was found to be more precise than NDVI. Overall accuracy was 82% for 1993 and 1999, and 80% for 2010, with the kappa values ranging from 0.789 to 0.761. Meanwhile, NDVI accuracy was attained at 64% with kappa value at 0.527 for 1999. In addition, 70% and 72% accuracy were obtained for 1993 and 2010 with estimated kappa values of 0.651 and 0.672, respectively. The study is anticipated to assist decision makers for better emergency response and sustainable land development action plans, thus mitigating the challenges of rapid urban growth.

Keywords: LULC, Landsat-5, NDVI, SSIC, Kuantan

Introduction

The spatial and temporal changes in land use/land cover (LULC) have proceeded rapidly as a result of increased urban populations and anthropogenic activities.

The modification of LULC and the interaction of humans and the environment have caused variability of dynamic changes to the environment and climate [1]. Several flood plains and river deltaic regions are highly vulnerable to flooding due to rapid urbanization and the threat of changed climatic events such as tempe-rature rise, wind storms, and heavy precipitation [2]. Therefore, it is required to monitor the modification of LULC under

*e-mail: mariazaidi26@gmail.com

diverse demographic and environmental conditions for sustainable land mana-gement [3]. Various studies have shown that the appli-cation of LULC change detection plays an important role in solving problems related to different domains such as changes in environmental services [4], urban development [5], and watershed characteristics [6-7].

Currently the use of satellite remote sensing data is considered the most powerful source of information to detect the changes in LULC because of its wide range of temporal and spatial coverage [8-9]. [10] defined change detection techniques as a method of recognizing changes in LULC with the time variation. Among several methods of image classification, principle component analysis (PCA), simple band ratio (SBR), and normalized differencing vegetative index (NDVI) are more popular [9, 11].

Several studies have also employed the visual interpretation for LULC classification, and some others have used supervised and unsupervised classification [12-13]. From the technical point of view, the supervised classification technique provides more accurate and reliable results compared to the unsupervised method. But the supervised method takes more time and costs too much. It also needs special equipment for collecting samples that are not available in many cases. For that reason, the time-series satellite data has been used to monitor changes in LULC efficiently. In this study, LULC changes were investigated for almost three decades. The determination of this research is to enhance the performance of semi-supervised image classification (SSIC) compared to the standard normalized differencing vegetation index (NDVI) technique.

Fig. 1. Study area: Kuantan River Basin (KRB).

Material and Methods

Study Area

Kuantan is the capital of Malaysia's third-largest state, Pahang. It is located at coordinates 3.82° N, 103.33° E. Kuantan is considered the social, economic, and commercial hub for the east coast of Peninsular Malaysia due to its locality. Geographically, Kuantan covers approximately an area of 1,707 km². Kuantan River Basin (KRB) was chosen as the study area, which passes through the main Kuantan city. KRB comprises numbers of important tributaries that supply water to the urban, agricultural, and industrial sectors of Kuantan (Fig. 1).

Data Source

The study utilized Landsat-5 (TM) images of 4 April 1993, 8 March 1999, and 7 February 2010, which are available at *glovis.usgs.gov*. To minimize the effect of atmospheric noise on image classification, cloud cover was limited to 15% (although some researchers believe that cloud cover should be less than 10% [5]). Since the study area was near the equatorial region and most of the time covered with clouds, it was hard to get <10% or cloud-free images. The collected images contained cloud cover of 15% (1993), 5% (1999), and 12% (2010).

Image processing

The Kertau_RSO_Malaya_Meters map projection was used for the image as it is commonly used in Malaysia. Then according to [14], radiance calibration was performed by converting the digital number (DN) value to surface spectral reflectance and removal of atmospheric effect (atmospheric correction) with the darkest pixel subtraction techniques using ENVI version 5.1.

Radiometric Calibration Process

The atmospheric effects of dispersed aerosol particles and atmospheric gases that traveled all through from the earth's surface to the sensors have been reported by several researchers [15-16]. It is believed that even though the atmospheric and radiometric effects have no significant influence over the classified image for LULC changes detection, it is necessarily required to eliminate their effects for investigating unforeseen variations in green cover. Radiometric calibration includes the corrections associated with sensor sensitivity, topography, and sun angle. In this technique, the digital number (DN) values are converted to the radiance unit for Landsat (TM). The spectral radiance at the sensor's aperture can be calculated using equations 1, 2, and 3 as stated by [17].

$$L_\lambda = \left(\frac{LMAX_\lambda - LMIN_\lambda}{Q_{calmax} - Q_{calmin}} \right) \left(Q_{Calmax} - Q_{Calmin} \right) + LMIN_\lambda$$

(1)

Or

$$L_\lambda = \left(\frac{LMAX_\lambda - LMIN_\lambda}{255}\right) 255 + LMIN_\lambda \qquad (2)$$

Or

$$L\lambda = G_{rescale} \times Q_{cal} + B_{rescale} \qquad (3)$$

While, $\quad G_{rescale} = \dfrac{LMAX_{\lambda-}LMIN_\lambda}{255} \qquad (3a)$

$$B_{rescale} = LMIN_\lambda \qquad (3b)$$

Whereas Lλ is spectral radiance at the sensor's aperture [$W/(m^2.sr.\mu m)$].

Q_{cal}: Quantized pixel value [DN]

Q_{calmax}: Maximum quantized calibrated pixel value [DN = 255) corresponding to $LMAX_\lambda$

Q_{calmin}: Minimum quantized calibrated pixel value [DN = 0] corresponding to $LMIN_\lambda$

$LMAX_\lambda$: Spectral radiance scaled to Q_{calmax} [$W/(m^2.sr.\mu m)$]

$LMIN_\lambda$: Spectral radiance scaled to Q_{calmin} [$W/(m^2.sr.\mu m)$]

$G_{rescale}$: Band-spectral rescaling gain factor [$W/(m^2.sr.\mu m)$]

$B_{rescale}$: Band-spectral rescaling bias factor [$W/(m^2.sr.\mu m)$]

According to [18], the acquired radiance value was then converted to the top of atmosphere (ToA) reflectance in order to correct the effects of uncertain solar zenith angle to the sensors at different dates of acquisition using Eq. 4.

$$\rho = \frac{\pi L d^2}{ESun_{\lambda cos\theta}} \qquad (4)$$

…where λ is ToA reflectance for the band, d is the distance from earth to sun (astronomical units), ESunλ is exoatmospheric solar irradiance, and θ is solar zenith angle in degrees. The ESunλ values for TM images were used from the available source [19]. Landsat images and atmospherically corrected images of the KRB are presented in Figs 2 and 3, respectively.

Semi-Supervised Image Classification (SSIC)

A natural color composite from the Landsat-5 bands b7, b4, and b2 were associated with red, green, and blue (RGB), respectively. The sample polygons were created based on visual interpretation on the image to recognize the land cover feature classes. Spectral analogous sub-areas were then demarcated with specified class name using a training signature. Then the images were classified by using the maximum likelihood classifier (MLC), which is the most common statistical technique for evaluating the standard LULC classification. According to [14], the vector feature of classes could be used to compute the likelihood of unknown pixels in each class. Hence the maximum probability of pixels is considered to
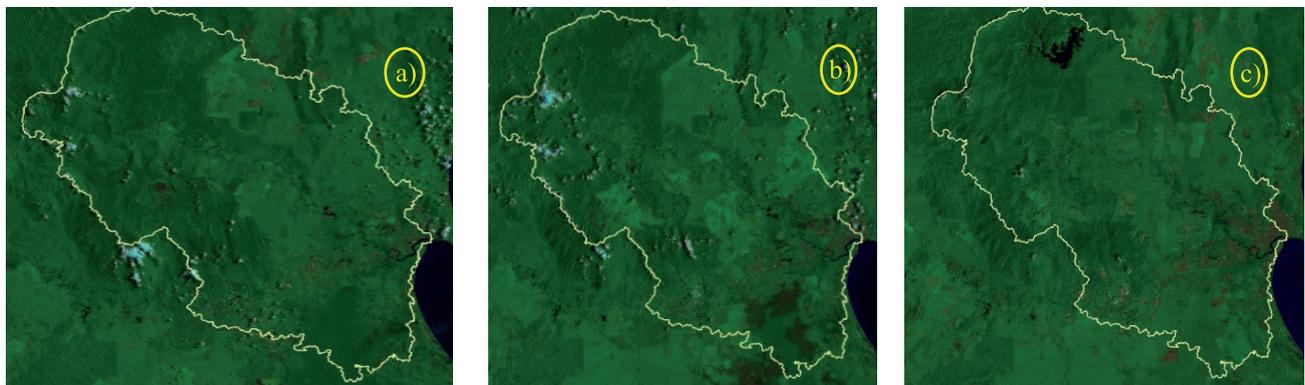

Fig. 2. Raw natural Landsat-5 color composite of the study area: a) 4 April 1993, b) 8 March 1999, and c) 7 February 2010.
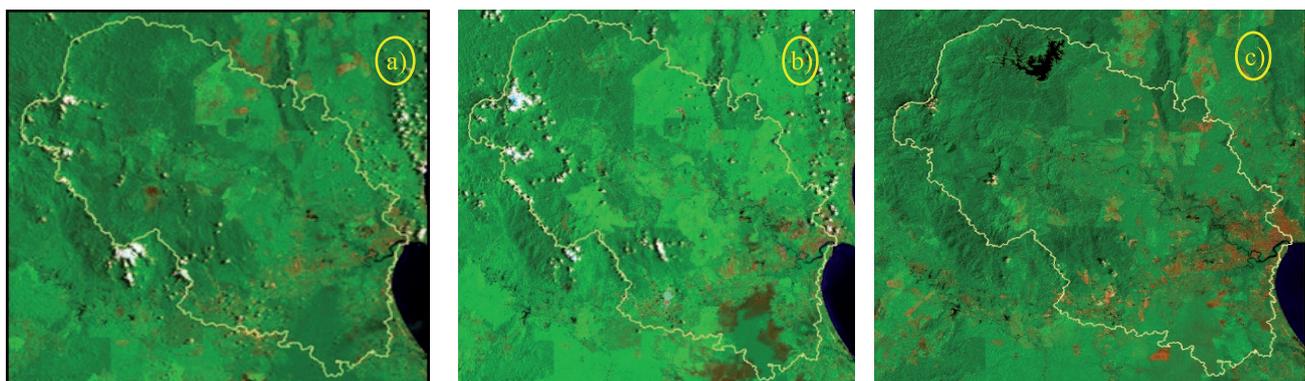

Fig. 3. Radiometric-corrected Landsat-5 images: a) 4 April 1993, b) 8 March 1999, and c) 7 February 2010.

belong to the assigned corresponding feature class. Five major LULC classes were identified in the study, namely water body, built-up land, agricultural land, forest, and bare soil. The accuracy assessment for each classified image was performed using error matrix and kappa value [20].

The generated random points depend on criteria and area situation. In this study, 10 reference points were selected in each feature class from Google Earth as the base map. Google Earth offers high resolution satellite imagery in different dates and times for many places. Overall accuracy was carried out using error matrix and kappa value indexes. The error matrix describes the level of accuracy between the training sample and classified values of the same image. This method is the most commonly used and is considered fundamental for assessing the overall classification using user accuracy, producer accuracy, and kappa coefficient [21].

NDVI Classification

The NDVI is a well-known technique for LULC classification. It was identified by [22] and is used to generate the normalized band ratio to enhance the presence or absence of vegetation cover. The equation is defined as NDVI = (NIR-IR)/(IR+NIR). In the case of Landsat-5, near infrared (NIR) refers to band 4 and infrared (IR) refers to band 3. According to this equation, NDVI values range from +1 to -1. The negative values represent the non-vegetative surface, whereas positive values indicate the green cover and as higher the positive values, the denser the green surface. In the study, NDVI values were assigned with reference of some previous studies such as [23-25] and the pixel by pixel value cross over. Negative NDVI readings indicate the water body, built-up land, and bare soil surface that range from 0.1 to 0.4 while agriculture and forest land cover describe the normal healthy vegetative index that falls between 0.4 to 0.76. Meanwhile, in April 1993 the values lie up to 0.98 because of dense vegetation index netted during the acquisition date.

Results and Discussions

The SSIC and NDVI classified maps for the years 1993, 1999, and 2010 are presented in Figs 4 and 5. To compare the statistical accuracy of proposed classifications, the error matrix and kappa coefficient methods were applied. The implied classifications showed good agreement. The overall accuracy of LULC of SSIC from 1993 to 2010$\geq$ 80% with kappa values from 0.781 to 0.761 showed good agreement compared with NDVI accuracy of 64% (1999), 70% (1993), and 72% (2010), with kappa value ranging from 0.52 to 0.671. The reason could be that SSIC relied on creating homogenous training areas based on visual interpretation and information about an area followed by the method of maximum likelihood that contributed to the best result. But SSIC also showed the

values mixed-up between bare soils and built-up land because these classes shared more similar reflectance values that caused misclassification. NDVI has been
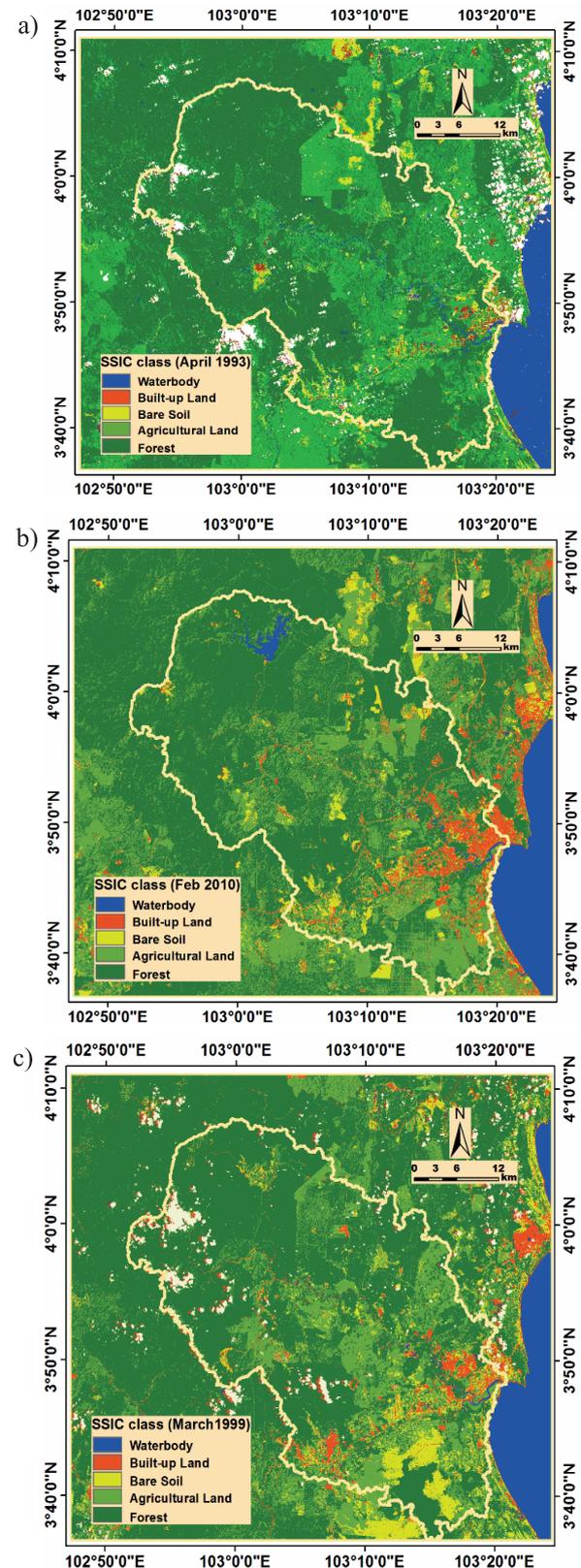

Fig. 4. LULC semi-supervised image classification (SSIC) for: a) 4 April 1993, b) 8 March 1999, and c) 7 February 2010.

used as a vital tool characterizing different classes that predominantly have great influence on vegetation cover. In NDVI classification, the pixel values were computed for

categorizing a natural feature class. The negative NDVI values indicated a water body whereas built-up land and bare soil surface ranges lie close to 0.1 to 0.4. Although both classifications have different levels of accuracy, they share almost the same pattern of LULC distribution.

In graphical analysis (Fig. 6), NDVI noticed the rising trend in built-up land from 1993 (4.24%) to 2010 (4.87%), as KRB has been experiencing urbanization since the majority of the population started migrating from rural to urban areas for economic and social reasons. The same trend was shown in SSIC in images from 1993 with 6.80% and 2010 with 5.90%, as values dramatically fell in 1999 with the area covered at 3.20%. On the contrary, bare soil class increased during the period from 1993 to 2010 with a significant difference of 1.75% to 4.49%. In NDVI, bare soil decreased during the corresponding years by 7.50%. There is a possibility that built-up land and bare soil were comprised of quite similar varieties. For this reason, both classes suffered from a confused classification percentage.

In addition, the forest area (the denser green cover land) in comparison with SSIC and NDVI classifications showed the highest ratio of 68.74% to 62.75% for 1999. Low forest area was observed during the period of 1993 (48.34%) in SSIC and (56.11%) in NDVI. On the other hand, the lowest ratio of agricultural land was observed
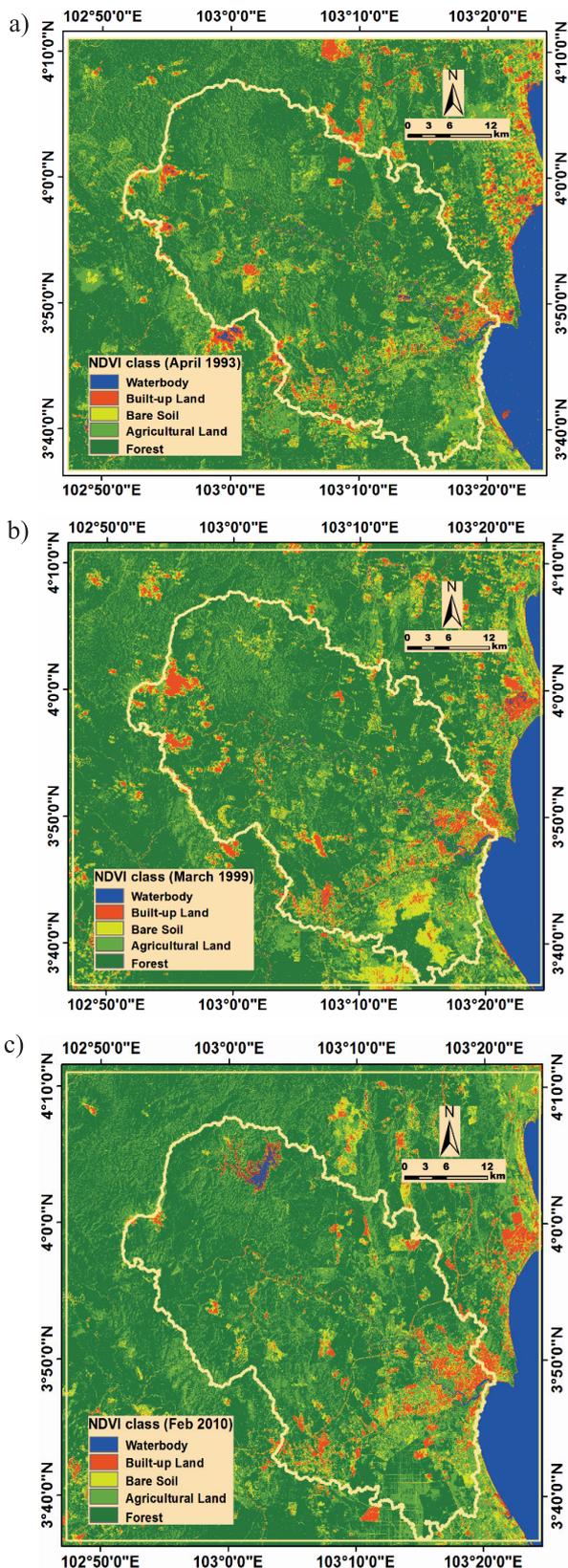

Fig. 5. LULC NDVI classification for: a) 4 April 1993, b) 8 March 1999, and c) 7 February 2010.
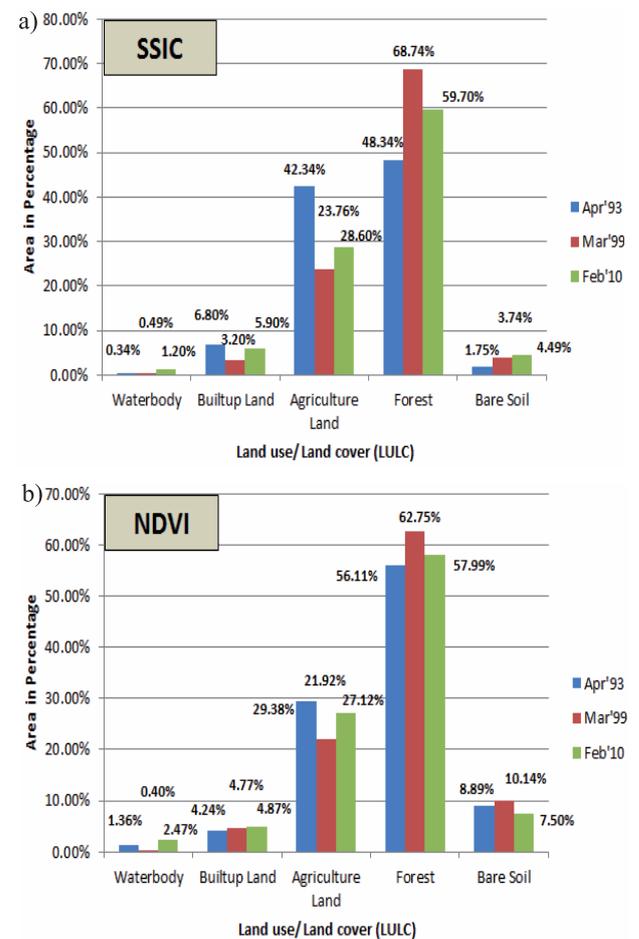

Fig. 6. LULC changes in the study area for the investigated period: a) SSIC and b) NDVI.

Table 1. SSIC and LULC error matrix for the 4 April 1993 image.

Actual Data	Predicted Data (PD)							
	WB	BL	AL	FO	BS	Total	PA (%)	UA (%)
WB	10	0	0	0	0	10	100	100
BL	0	8	0	1	1	10	88.88	80
AL	0	0	8	2	0	10	80	80
FO	0	0	1	9	0	10	64.28	90
BS	0	1	1	2	6	10	85.71	60
Total	10	9	10	14	7	50		

Table 2. SSIC and LULC error matrix for the 8 March 1999 image.

Actual Data	Predicted Data (PD)							
	WB	BL	AL	FO	BS	Total	PA (%)	UA (%)
WB	10	0	0	0	0	10	100	100
BL	0	8	1	1	0	10	80	80
AL	0	0	7	3	0	10	77.77	70
FO	0	0	0	10	0	10	66.66	100
BS	0	2	1	1	6	10	100	60
Total	10	10	9	15	6	50		

Table 3. SSIC and LULC error matrix for the 7 February 2010 image.

Actual Data	Predicted Data (PD)							
	WB	BL	AL	FO	BS	Total	PA (%)	UA (%)
WB	10	0	0	0	0	10	100	100
BL	0	7	0	0	3	10	100	70
AL	0	0	5	4	1	10	83.33	50
FO	0	0	0	10	0	10	66.66	100
BS	0	0	1	1	8	10	66.66	80
Total	10	7	6	15	12	50		

Note: WB is water body, BL is built-up land, AL is agricultural land, FO is forest, BS is bare soil, UA is user accuracy, and PA is producer accuracy.

Table 4. NDVI LULC error matrix for the 4 April 1993 image.

Actual Data	Predicted Data (PD)							
	WB	BL	AL	FO	BS	Total	PA (%)	UA (%)
WB	10	0	0	0	0	10	100	100
BL	0	9	0	1	0	10	81.81	90
AL	0	0	5	5	0	10	62.5	50
FO	0	0	3	6	1	10	40	60
BS	0	2	0	3	5	10	100	50
Total	10	11	8	15	6	50		

Table 5. NDVI LULC error matrix for the 8 March1999 image.

Actual Data	Predicted Data (PD)							
	WB	BL	AL	FO	BS	Total	PA (%)	UA (%)
WB	9	1	0	0	0	10	100	90
BL	0	7	1	2	0	10	63.63	70
AL	0	0	5	5	0	10	50	50
FO	0	0	3	7	0	10	43.75	70
BS	0	3	1	2	4	10	100	40
Total	9	11	10	16	4	50		

Table 6. NDVI LULC error matrix for the 7 February 2010 image.

Actual Data	Predicted Data (PD)							
	WB	BL	AL	FO	BS	Total	PA (%)	UA (%)
WB	9	1	0	0	0	10	100	90
BL	0	8	0	0	2	10	61.53	80
AL	0	1	6	3	0	10	60	60
FO	0	0	2	8	0	10	72.72	80
BS	0	3	2	0	5	10	71.42	50
Total	9	13	10	11	7	50		

Note: WB is water body, BL is built-up land, AL is agricultural land, FO is forest, BS is bare soil, UA is user accuracy, and PA is producer accuracy

in 1999 (23.76% and 21.91%) and the highest ratio was computed for 1993 in both NDVI (21.92%) and SSIC (42.34%). The reason for the denser cover class in 1999 was due to the period for monsoon rainfall during that date of acquisition. For the water body, the greater values were observed in 2010, which was about 1.20% (SSIC) and 2.47% (NDVI). The main reason for this change in water area is the construction of a DAM reservoir at the upper stream of KRB. Detailed statistical analysis of SSIC and NDVI classifications for the LULC change detection of KRB is represented in Tables 1-3 and 4-6, respectively. Overall accuracy can be calculated by Eq. 5 and Kappa coefficient (kc) is determined by Eq. 6, shown in Table 7.

Overall Accuracy =
$$\frac{Sum\ of\ true\ random\ points}{Cumulative\ sum\ of\ all\ random\ points} \qquad (5)$$

$$k_c = \frac{Observed - Expected}{1 - Expected} \qquad (6)$$

Conclusion

In this work, the proposed change detection approach based on the semi-supervised method was performed. The accuracy of both classification methods (NDVI and SSIC) were validated with Google Earth images. The obtained results showed that the level of accuracy increased from 80 to 82% (SSIC) and 64 to 72% (NDVI), which represent the significant improvement in image classification. Thus the implementation of the generated homogenous sample point helps as a useful and reliable approach for data verification. This technique is useful in case there is

Table 7. Accuracy assessment of Landsat-5 in LULC classification.

	SSIC		NDVI	
Landsat 5	Overall accuracy (%)	Kappa coefficient (k_c)	Overall accuracy (%)	Kappa coefficient (k_c)
April 1993	82	0.789	70	0.651
March 1999	82	0.789	64	0.527
February 2010	80	0.761	72	0.672

not enough time or budget for a field survey or any other constraint and limitation to perform standard classification techniques. According to this study, SSIC with Google Earth imageries seem to be reliably cost-effective and a time-saving technique. It also enhances the online sources as an alternative for verification of image classification when there is the lack of financial assistance to arrange field survey and/or sparse availability of referenced maps.

Acknowledgements

References

1. YANG, Z.-L., Modeling land surface processes in short-term weather and climate studies. **3**, **2004**, World Scientific Series on Meteorology of East Asia.
2. ARKEMA K.K., et al., Coastal habitats shield people and property from sea-level rise and storms. Nature Climate Change, **3** (10), 913, **2013**.
3. LIU D., et al., Effects of land use classification on landscape metrics based on remote sensing and GIS. Environmental earth sciences, **68** (8), 2229, **2013**.
4. GILANI H., et al., Decadal land cover change dynamics in Bhutan. Journal of environmental management, **148**, 91, **2015**.
5. TAN K.C., et al., Landsat data to evaluate urban expansion and determine land use/land cover changes in Penang Island, Malaysia. Environmental Earth Sciences, **60** (7), 1509, **2010**.
6. MCCOLL C., AGGETT G. Land-use forecasting and hydrologic model integration for improved land-use decision support. Journal of environmental management, **84** (4), 494, **2007**.
7. AKBARI A., et al., The application of the Water Erosion Prediction Project (WEPP) model for the estimation of runoff and sediment on a monthly time resolution. Environmental Earth Sciences, **74** (7), 5827, **2015**.
8. KAWY W.A.M.A., EL-MAGD I.H.A. Use of satellite data and GIS for assessing the agricultural potentiality of the soils South Farafra Oasis, Western Desert, Egypt. Arabian Journal of Geosciences, **6** (7), 2299, **2013**.
9. YEGANEH H., et al., Monitoring rangeland ground cover vegetation using multitemporal MODIS data. Arabian Journal of Geosciences, **7** (1), 287, **2014**.
10. SINGH A. Review article digital change detection techniques using remotely-sensed data. International journal of remote sensing, **10** (6), 989, **1989**.
11. SUN J., QIN X. Precipitation and temperature regulate the seasonal changes of NDVI across the Tibetan Plateau. Environmental Earth Sciences, **75** (4), 1, **2016**.
12. NIRAULA R.R., et al., Measuring impacts of community forestry program through repeat photography and satellite remote sensing in the Dolakha district of Nepal. Journal of environmental management, **26**, 20, **2013**.
13. VOGELMANN J.E., et al., Monitoring gradual ecosystem change using Landsat time series analyses: Case studies in selected forest and rangeland ecosystems. Remote Sensing of Environment, **122**, 92, **2012**.
14. JIA K., et al., Land cover classification of finer resolution remote sensing data integrating temporal features from time series coarser resolution data. ISPRS Journal of Photogrammetry and Remote Sensing, **93**, 49, **2014**.
15. TOKOLA T., LÖFMAN S., ERKKILÄ A. Relative calibration of multitemporal Landsat data for forest cover change detection. Remote Sensing of Environment, **68** (1), 1, **1999**.
16. JANZEN D.T., FREDEEN A.L., WHEATE R.D. Radiometric correction techniques and accuracy assessment for Landsat TM data in remote forested regions. Canadian Journal of Remote Sensing, **32** (5), 330, **2006**.
17. CHANDER G., MARKHAM B.L., BARSI J.A. Revised Landsat-5 thematic mapper radiometric calibration. IEEE Geoscience and remote sensing letters, **4** (3), 490, **2007**.
18. CHANDER G., MARKHAM B.L., HELDER D.L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote sensing of environment, **113** (5), 893, **2009**.
19. CHANDER G., MARKHAM B. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. IEEE Transactions on geoscience and remote sensing, **41** (11), 2674, **2003**.
20. FOODY G.M. Status of land cover classification accuracy assessment. Remote sensing of environment, **80** (1), 185, **2002**.
21. FOODY G.M. Classification accuracy comparison: hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. Remote Sensing of Environment, **113** (8), 1658, **2009**.
22. ROUSE JR J.W., et al., Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication, **351**, 309, **1974**.
23. PANTANAHIRAN W. Land use change on sloping areas in Phuket Province, Thailand. in Agro-geoinformatics (Agro-geoinformatics 2014), Third International Conference on. **2014**. IEEE.
24. GHOBADI Y., et al., Assessment of spatial relationship between land surface temperature and landuse/cover retrieval from multi-temporal remote sensing data in South Karkheh Sub-basin, Iran. Arabian Journal of Geosciences, **8** (1), 525, **2015**.
25. PRAVALIE R., SÎRODOEV I., PEPTENATU D., Detecting climate change effects on forest ecosystems in Southwestern Romania using Landsat TM NDVI data. Journal of Geographical Sciences, **24** (5), 815, **2014**.