Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

ISSN: 0972-6268	Vol. 15	No. 1	pp. 243-248	2016

Original Research Paper

Comparison to Supervised Classification Modelling in Land Use Cover Using Landsat 8 OLI Data: An Example in Miyun County of North China

Wenfeng Gong*, Li Yuan, Wenyi Fan***†, Xiaofeng Wang* and Philip Stott******
*College of Hydraulic and Electrical Engineering, Heilongjiang University, Harbin 150086, China
**College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China
***College of Forest, Northeast Forestry University, Harbin 150040, China
****School of Animal and Veterinary Sciences, The University of Adelaide, Roseworthy 5371, Australia
†Corresponding author: Wenyi Fan

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 10-01-2015
Accepted: 15-03-2015

Key Words:
Land use cover
Image classification
Support vector machines
Maximum likelihood classifier

ABSTRACT

Land use cover (LUC) classification is one of the most important applications of optical remotely sensed data, while LUC mapping outcomes are used for global, local mapping, ecosystem assessment and environmental process monitoring. Hence, in this study, in order to evaluate the advantages and drawbacks of supervised classification schemes, the paper chose the optical image data of Landsat 8 OLI in Miyun county to test supervised classification and introduced Parallelepiped Method (PM), Minimum Distance (MD), Maximum Likelihood Classifier (MLC) and Support Vector Machines (SVMs) to improve classification accuracy of LUC mapping and to obtain the reliable LUC distribution. The four classified images reveal that the study area is dominated by considerable areas of forest land, with the overall accuracy found to be 87.89% (kappa = 0.8524) using SVMs, 85.26% (kappa = 0.8205) using MLC, 82.11% (kappa = 0.7813) using MD, and 74.74% (kappa = 0.6920) using PM. Based on the overall accuracy and kappa statistics, SVMs might be the first option in terms of classification accuracy without taking into account of the time costly and standard PC and laptops. MLC was the second accurate model classifiers from the classified image, which was always used to obtain LUC map information for economic potential in time and cost; and PM has shown the lowest overall classification accuracy with greater omission errors and commission errors.

INTRODUCTION

Land cover can be defined as changes in land resources related to biophysical factors such as biodiversity, soil and ecological processes (Weng 2001), whereas land use refers to the alteration of land for human needs, such as transportation, recreation and housing (Yemane 2003, Pillay et al. 2014). Furthermore, land use influences the environment, mainly by land cover (Wang et al. 2010), and thus land use cover (LUC) needs to be assessed simultaneously with an integrated evaluation of the results, as there are multifarious relationships between biophysical and social factors (Robson 2005, Sim et al. 2006). Remotely-sensed data (i.e., satellite or aerial imagery) has been one of the most important data sources for studies of LUC spatial and temporal changes (Fichera et al. 2012), and also can often be used to define land-use through observations of the LUC (Brown et al. 2000, Karl & Maurer 2010).

In fact, gathering information about LUC from RS data is a fundamental tool for RS interpretation and information extraction. Satellite data classification is a method by which labels or class identifiers are attached to individual pixels on the basis of their spectral characteristics in various bands (Sharma et al. 2013), and then becomes a vital role in image interpretation because classification results are basis for interpretation, analysis and modelling for various environmental and socio-economic applications (Jensen & Cowen 1999). However, due to the characteristics of RS data with large amount, higher-dimension and difference between spectrum and spatial distribution of objects, classification of remotely sensed satellite data into a thematic map remains a challenge (Sharma et al. 2013). Further, the traditional classification algorithms, which are commonly used in remote sensing studies and currently available in today's software systems, including algorithms of MLC, MD, PM (Abdul-Qadir 2014). While comparing the studies of different classification algorithms, SVMs, as a non-parametric supervised learning classifier, is a very effective and useful tool for the remote sensing image classification on LUC, and has been increased significantly in the past decade (Mountrakis et al. 2010).

Miyun county, which is situated in the northeastern of Beijing, the capital of China, is very important surface source of drinking water and natural ecological zone. While in recent years, the natural factors and irrational human activities caused great changes in LUC of the study region, and then made a certain impact on eco-environment quality. As for the actual eco-environment situation, it is not only directly influenced the local economic development and residential living, but also made an impact on the safety of water supply for Beijing. Therefore, in order to improve the water quantity and quality of Miyun reservoir, it is very necessary to master the LUC current situation and implement the protective measures for regional eco-environment. In this paper, with optical image data of Landsat 8 OLI, the four supervised classification schemes in discrimination of geological materials were evaluated to improve classification accuracy of LUC mapping and to obtain the reliable LUC distribution. The objectives of this paper are as follows: (1) to identify and classify dominant LUC classes of study region using GIS; (2) to evaluate and supervise the classification result performance in SVMs, PM, MD and MLC by using standard measures of accuracy assessment; (3) to find the best result in supervised classification among the different classifiers.

MATERIALS AND METHODS

Study area: Miyun county (Fig. 1), is the greatest county in Being between latitude 40°14'-40°48'N and longitude 116°41'-117°30' E, which covers an area of 2229.5 km^2. The gently mountain region, is mainly situated in northwest of the study area, which accounts for 79.47% of the total area, while the mainly hilly and partially plain areas are located in the southeast region, and the altitude varies from 150 to 1800 m above sea-level with an average elevation of 962 m. The climate of the region is warm and semi-humid continental monsoon with a range of annual rainfall between 407 mm in the northwest zones to 797 mm in the southeast (Li et al. 2013). About 80% of the rainfall occurs between June and September. The distribution of precipitation is generally decreasing from southeast to northwest. Annual average temperatures in the upper and lower watershed are 9°C and 25°C respectively (Chen et al. 2011).

Data processing: Landsat 8 OIL (Path/Row:148/45) dated October 23rd, 2013 was used for this study. Prior to the interpretation for obtaining accurate data, data processing is very important for satellite images with standard procedures including georeferencing, radiometric correction and image enhancement (Qin et al. 2006, Masoud & Koike 2006, Wang et al., 2010). Firstly, the standard false colour composites were generated displaying bands 6, 5, 4 for this image as red, green and blue respectively, then based on the 1:50,000 scale topographic maps, supported by the ENVI 5.1 remote sensing software, the image was geo-rectified by Transverse Mercator (TM) projection, while the resampling method was the linear interpolation, the root mean square errors (RMSE) was under 1 pixel. With the support of the FLAASH (Fast Line of Sight Atmospheric Analysis of the Spectral Hypercubes) module, the image was enhanced by linear contrast stretching and histogram equalization methods (Liu et al. 2007), and then the image of the study area was masked by using the boundary with raster format.

Image classification: Prior to image classification, the regional land use classification system was set up with reference to the current Land Use Condition Classification (Chinese National Standard GB/T 2000-2007), then some high resolution images were used as referenced data, such as World View-2 image (0.5m in the panchromatic band and 2.0m in multispectral bands) and Googlearth Map. A congruous number of training samples have been selected as reference data and ancillary information by using a handled GPS. Meanwhile, according to the actual regional land use state, a hierarchical classification system of 6 land cover classes was extracted from the image data, which were croplands, forest lands, grasslands, water bodies, built-up areas and unused

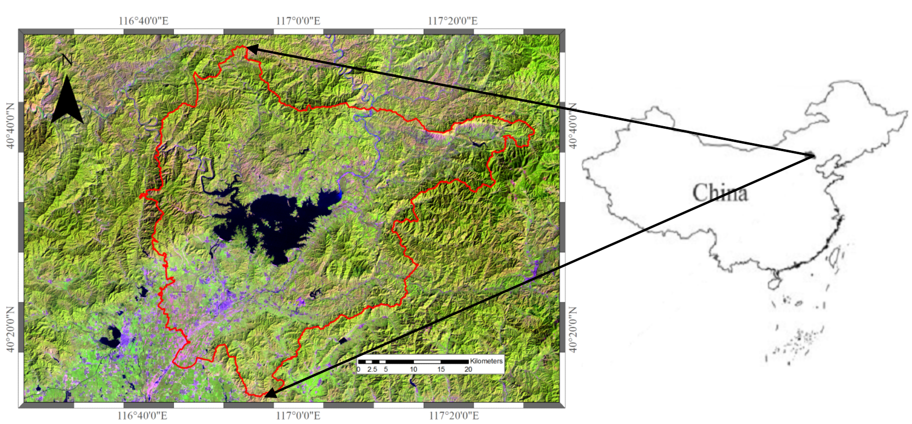

Fig. 1: Geographic location of the study area.

lands. While, croplands included paddy and dry farming land; forestlands included forest, shrub and others, grasslands included natural grasslands and pastures; water bodies included rivers and reservoirs; built-up land mainly included urban areas, rural settlements and others; unused lands included wetland, bare rock and so on. In training stage, the ROIs (Regions of Interest) were selected as training areas (polygons) on the screen for image classification, the number of sample ROIs (from visual interpretation, filed data and some high resolution images) for croplands, forestlands, grasslands, water bodies, built-up lands and unused lands was 58, 136, 75, 37, 89 and 64 respectively, then the total of 459 training sites were used for classification algorithms. Finally, 3×3 majority filter has been applied to the classified LUC data to reduce the "salt & pepper" effect (Lillesand & Kiefer 1999).

RESULTS AND DISCUSSION

Classification accuracy analysis: In order to quantitatively analyse the classification performance, the accuracy assessment of classification and the LUC mapping are performed in this paper, which can be described as the method for the analyst to distinguish, to what extent of reliability and confidence can be drawn from the data analysis (Pillay et al. 2014), meanwhile, the accuracy statistics provides objective information about the quality of the LUC classification (Wondie 2011). Therefore, the 190 random points of accuracy assessment were chosen using the random stratified method to represent different land cover classes, 164 points from field data by using GPS and 26 points from the referenced data. Meanwhile, error and accuracy assessment of classification result has been calculated. On the basis of confusion matrix by which several characteristics about classification performance can be expressed, including the omission (exclusion) and commission (inclusion) (Abdul-Qadir 2014), the overall accuracy of the classification, producers and users accuracy and Kappa coefficient were generated from this error matrix, which are reported in Tables 1, 2, 3 and 4. According to the specialized literature (Congalton & Green 1999), the Kappa statistics values can be divided into three groups as follows: (1) higher than 0.80 representing a strong agreement, (2) values in the range 0.40-0.80 expressing a moderated agreement, (3) and values below 0.40 representing a weak agreement (Vorovencii 2014). We can see that the classification results using the SVMs obtained the best overall accuracy of 87.89% and Kappa as 0.8524, followed by MLC kept the overall accuracy of 85.26% and Kappa of 0.8205, while the classification method of the PM obtained the least overall accuracy of 74.74% and Kappa as 0.6920. It is obvious that the Kappa accuracy values of 0.8524 for SVMs and of 0.8205 for MLC show a strong agreement between map classification and ground reference data, however, the Kappa accuracy values of 0.7813 and 0.6920 for two classification methods of MD and PM keep a moder-

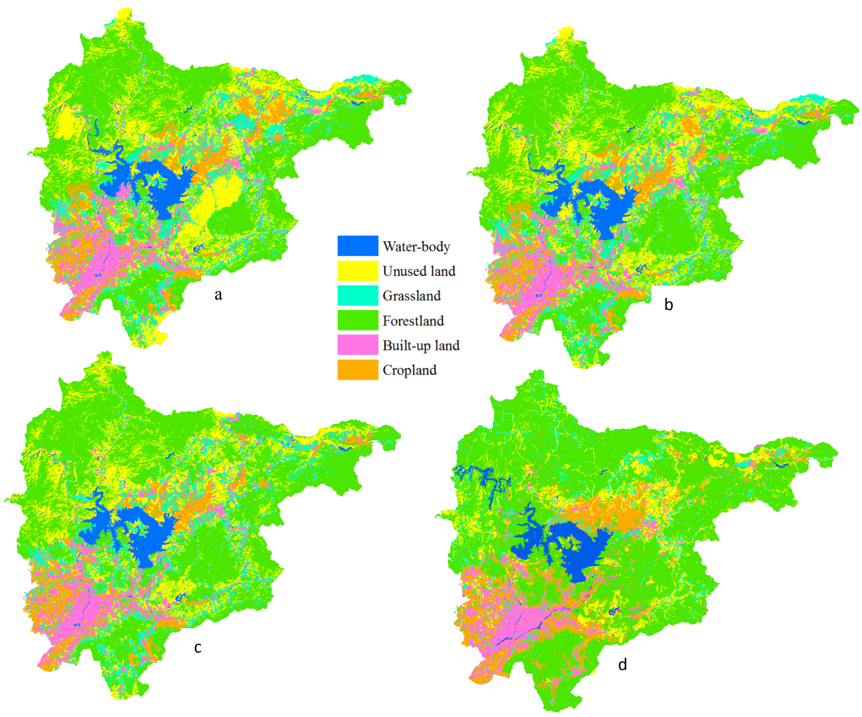

Fig. 2: LUC map using different supervised classifiers (a. PM; b. MD; c. MLC; d. SVMs).

Table 1: The error matrix for classification method of PM.

Class name	Cropland	Forestland	Grassland	Water-body	Built-up land	Unused land	Total	Users accuracy
Cropland	21	1	3	1	2	2	30	0.7000
Forestland	3	36	2	1	0	4	46	0.7826
Grassland	4	2	29	0	0	3	38	0.7632
Water-body	2	0	0	16	1	1	20	0.8000
Built-up land	1	1	0	3	18	2	25	0.7200
Unused land	2	2	2	2	1	22	31	0.7097
Total	33	42	36	23	22	34	190	
Producers accuracy	0.6364	0.8571	0.8056	0.6957	0.8182	0.6471		

Overall classification accuracy = 74.74%; Kappa statistics = 0.6920

Table 2: The error matrix for classification method of MD.

Class name	Cropland	Forestland	Grassland	Water-body	Built-up land	Unused land	Total	Users accuracy
Cropland	24	1	2	1	2	0	30	0.8000
Forestland	1	41	2	0	0	2	46	0.8913
Grassland	3	2	31	0	0	2	38	0.8158
Water-body	0	1	0	17	1	1	20	0.8500
Built-up land	2	1	1	0	19	2	25	0.7600
Unused land	2	2	1	1	1	24	31	0.7742
Total	32	48	37	19	23	31	190	
Producers accuracy	0.7500	0.8542	0.8378	0.8947	0.8261	0.7742		

Overall classification accuracy = 82.11 %; Kappa statistics = 0.7813

Table 3: The error matrix for classification method of MLC.

Class name	Cropland	Forestland	Grassland	Water-body	Built-up land	Unused land	Total	Users accuracy
Cropland	25	1	0	1	2	1	30	0.8333
Forestland	0	42	2	0	0	2	46	0.9130
Grassland	3	2	31	0	0	2	38	0.8158
Water-body	0	0	0	19	0	1	20	0.9500
Built-up land	1	1	0	0	20	3	25	0.8000
Unused land	1	2	2	0	1	25	31	0.8065
Total	30	48	35	20	23	34	190	
Producers accuracy	0.8333	0.8750	0.8857	0.9500	0.8696	0.7353		

Overall classification accuracy = 85.26 %; Kappa statistics = 0.8205

Table 4: The error matrix for classification method of SVMs.

Class name	Cropland	Forestland	Grassland	Water-body	Built-up land	Unused land	Total	Users accuracy
Cropland	26	1	0	1	2	0	30	0.8667
Forestland	0	43	2	0	0	1	46	0.9348
Grassland	2	1	33	0	0	2	38	0.8684
Water-body	0	0	0	18	1	1	20	0.9000
Built-up land	1	1	0	0	21	2	25	0.8400
Unused land	0	2	2	0	1	26	31	0.8387
Total	29	48	37	19	25	32	190	
Producers accuracy	0.8966	0.8958	0.8919	0.9474	0.8400	0.8125		

Overall classification accuracy = 87.89 %; Kappa statistics = 0.8524

Table 5: Classification results of different land use catalogues.

Class	PM		MD		MLC		SVMs	
	km^2	%	km^2	%	km^2	%	km^2	%
Cropland	295.40	13.27	228.84	10.28	250.43	11.25	272.59	12.25
Forestland	1119.05	50.27	1220.56	54.83	1294.47	58.15	1317.37	59.18
Grassland	162.95	7.32	140.69	6.32	114.87	5.16	120.19	5.40
Water-body	115.98	5.21	115.09	5.17	103.29	4.64	109.99	4.94
Built-up land	184.10	8.27	181.65	8.16	160.50	7.21	158.55	7.12
Unused land	348.60	15.66	339.25	15.24	302.52	13.59	247.39	11.11

ated agreement respectively. Meanwhile, the users and producers accuracy of individual classes for the same image were 63-95%, the MLC exhibited the highest users and producers accuracy for water body of 95% due to the integration of the visual interpretation with the classified image, while PM has shown the worst users and producers accuracy for cropland of 70.00% and 63.64% respectively. That the lowest accuracy was cropland could be explained by the fact that the shape of cropland is very similar to grassland, which led to the confusion with the grassland.

In a word, the SVMs exhibited the highest classification accuracy, accurately discriminated the ground objects, especially for the easily misclassified feature, and classified cropland, forestland, grassland and unused land more accurate than other classifiers. However, application of SVM is time costly when using standard PC and laptops during the image interpretation. Inferior to the SVMs, MLC shows the potential to discriminate the configuration and lithologic contents when accurate spectral behaviours for training pixels are considered (Abdul-Qadir 2014), and classifying the built-land and water-body more accurate than other classifiers. Moreover, an advantage of MLC is that it costs a short time during the processing of image interpretation. Hence, comparing to other classification methods, in practice, we are still using the MLC as a main processing method in dealing with digital remote sensing techniques when the strong training dataset was used. In comparison, the results from PM performed the least overall classification accuracy with greater errors of omission and commission, the main reason is that the ground surface features is easily disturbed by the similar spectral information.

Classification results analysis: Classification results of four different LUC maps along with Landsat 8 OLI data are presented in Fig. 2, originally storing in raster format (30m pixel resolution), and then have been converted into the shape file (*.shp) vector format using ArcGIS 9.3. Meanwhile, description of these land cover classes derived using classification schemes are presented in Table 5, and area calculation of LUC is shown in Fig. 3. We can see that the land use cover classes of cropland (295.40km^2), grassland (162.95km^2), water-body (115.98 km^2), built-up land (184.10 km^2) and unused land (348.60 km^2) for PM have been overestimated, and shown the maximum classes against other three classification methods. Moreover, comparing to other three classifiers, PM was shown the minimum coverage of forestland (1119.05km^2), while the SVMs was shown the maximum forestland (1317.37km^2) among the four classification algorithms. On the whole, the forestland area took up greatest proportion, with the ratios from 50.27% to 59.18% among all the classifiers. The four classified images reveal that the study area is dominated by considerable areas of forestland, and controlled the overall land use structure, functional and dynamic processes. Moreover, taking the actual situation into consideration, the Miyun reservoir needs more forestland to reduce the soil erosion and protect the environment and improve the water quality. On the contrary, the water-body accounted for the least ratio against the other land classes. In comparison, built-up land (anthropogenic pattern) and unused land (semi-natural land use pattern) mapped the minimum land classes by SVMs, which accounts for 158.55 km^2 and 247.39 km^2 respectively, and classified the built-up land less accurate than MLC. Obviously, the current overall land use pattern of the study area is dominated by forestland scattered with water-body, grassland and built-up land. In addition to the outside factors and natural conditions, the present land use pattern is influenced by its location, important sources of drinking water in Beijing, population, level of economic development and regional industrial structure of the study area.

CONCLUSION

In order to evaluate the advantages and drawbacks of supervised classification, this paper chose the Landsat TM OLI data and ancillary data of Miyun county in North China to test four supervised classification schemes in discrimination of geological materials and introduced PM, MD, MLC and SVMs with the objective to obtain an accurate classification of LUC mapping. The major findings are as follows: The four classified images reveal that the study area is dominated by considerable areas of forestland. Based on the overall

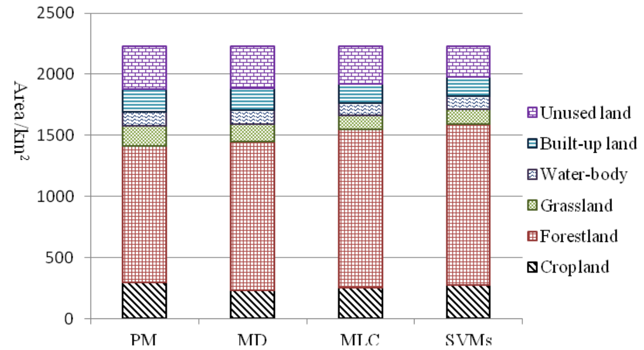

Fig. 3: Area distribution in different classification schemes.

accuracy and kappa statistics, SVMs might be the first option in terms of classification accuracy, without taking into account of the time cost and standard PC and laptops, and classified cropland, forestland, grassland and unused land more accurate than other classifiers; MLC was the second accurate model classifier from the classified image, which classified the built-up land and water-body more accurate than SVMs, and is always used to obtain LUC maps information for economic potential in time and cost. PM shown the lowest overall classification accuracy with greater omission errors and commission errors.

ACKNOWLEDGMENTS

This research was supported by National High Technology Research and Development Program 863 (Grant No. 2012 AA102001), Natural Scientific Foundation of Heilongjiang Province (Grant No.D201410), Natural Science Foundation of Heilongjiang Province Department Education Commission (Grant No.12531513). We appreciate the staff of the Ecological Research Central of Songhua River in 3S laboratory of Heilongjiang University. We also want to express our respect and thanks to the anonymous reviewers and the editors for their helpful comments in improving the quality of this paper.

REFERENCES

Abdul-Qadir, A. M. 2014. Supervised classification for lithologic discrimination in ShaikhIbrahim Area, NW Iraq using landsat images. Arab. J. Sci. Eng., 39: 437-451.

Brown, D. G., Pijanowski, B. C. and Duh, J. D. 2000. Modeling The relationships between land use and land cover on private lands in the upper midwest, USA. Journal of Environmental Management, 59: 247-263.

Chen, T., Niu, R. Q. and Wang, Y. et al. 2011. Assessment of spatial distribution of soil loss over the upper basin of Miyun reservoir in China based on RS and GIS techniques. Environ. Monit. Assess., 179: 605-617.

Congalton, R.G. and Green, K. 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Boca Raton, Florida: Lewis Publishers.

Fichera, C.R., Modica, G. and Pollino, M. 2012. Land cover classification and change-detection analysis using multi-temporal remote sensed imagery and landscape metrics. European Journal of Remote Sensing, 45: 1-18.

Jensen, J. R. and Cowen, D.J. 1999. Remote sensing of urban suburban infrastructure and socio-economic attributes. Photogram. Eng. Remote Sens., 65: 611-622.

Karl, J.W. and Maurer, B.A. 2010. Multivariate correlations between imagery and field measurements across scales: comparing pixel aggregation and image segmentation. Landscape Ecology, 25: 591-605.

Lillesand, T.M. and Kiefer, R.W. 1999. Remote Sensing and Image Interpretation. New York: John Wiley and Sons.

Liu, L. C., Dong, X. F. and Wang, J. H. 2007. Dynamic analysis of eco-environmental changes based on remote sensing and geographic information system: an example in Longdong region of the Chinese Loess Plateau. Environ. Geol., 53: 589-598.

Li, X.S., Wu, B.F. and Zhang, L. 2013. Dynamic monitoring of soil erosion for upper stream of Miyun reservoir in the last 30 Years. J. Mt. Sci., 10(5): 801-811.

Masoud, A.A. and Koike, K. 2006. Arid land salinization detected by remotely-sensed land cover changes: a case study in the Siwa region, NW Egypt. J. Arid Environ. 66(1): 151-167.

Mountrakis, G., Im, J. and Ogole, C. 2010. Support vector machines in remote sensing: A review. ISPRS. Journal of Photogrammetry and Remote Sensing, 66: 247-259.

Pillay, K., Agjee, N.H. and Pillay, S. 2014. Modelling changes in land cover patterns in Mtunzini, South Africa using satellite imagery. J. Indian Soc. Remote Sens., 2(1): 51-60.

Qin, Z., Li, W., Burgheimer, J. and Karnieli, A. 2006. Quantitative estimation of land cover structure in an arid region across the Israel-Egypt border using remote sensing data. J. Arid Environ., 66(2): 336-352.

Robson, T. F. 2005. An Assessment of Land Cover Changes Using GIS and Remote Sensing: A Case Study of the UMhlathuze Municipality, KwaZulu-Natal, South Africa. Diss. University of KwaZulu-Natal, Pietermaritzburg

Sharma, R., Ghosh, A. and Joshi, P. K. 2013. Decision tree approach for classification of remotely sensed satellite data using open source support. J. Earth Syst. Sci., 122(5): 1237-1247.

Sim, V., Strauss, A., Wand, D. and Forbes, T. et al. 2006. Umhlanga Node Study, Final Report, for Thekwini Municipality, Development Planning and Management Unit, Durban, South Africa.

Wang, S.Y., Liu, J.S. and Ma, T.B. 2010. Dynamics and changes in spatial patterns of land use in Yellow River Basin, China. Land Use Policy, 27: 313-323.

Weng, Q. 2001. A remote sensing-GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, China. International Journal of Remote Sensing, 22(10): 1999-2014.

Wondie, M., Schneider, W. and Melesse, A. M. et al. 2011. Spatial and temporal land cover changes in the Simen Mountains National Park, a world heritage site in north western Ethiopia. Remote Sens., 3: 752-766.

Yemane, M. M. 2003. An Assessment of Changes in Land Use/Cover Patterns in the Albert Falls Area KwaZulu-Natal. Unpublished Masters dissertation, Pietermartizburg: University of KwaZulu-Natal.

Vorovencii, I. 2014. Assessment of some remote sensing techniques used to detect land use/land cover changes in South-East Transilvania, Romania. Environ. Monit. Assess., 186: 2685-2699.