TESTING THE POTENTIAL OF VEGETATION INDICES FOR LAND USE/COVER CLASSIFICATION USING HIGH RESOLUTION DATA

A. Karakacan Kuzucu [a,b,]*, F. Bektas Balcik [b]

[a] Buyukcekmece Municipality, 34535 Buyukcekmece Istanbul, Turkey - aysin.kuzucu@bcekmece.bel.tr
[b] ITU, Civil Engineering Faculty, Geomatics Engineering Department, 34469 Maslak Istanbul, Turkey - bektasfi@itu.edu.tr

KEY WORDS: Support Vector Machine, Maximum Likelihood Classification, SPOT 7, Land use/Land cover, Vegetation Indices

ABSTRACT:

Accurate and reliable land use/land cover (LULC) information obtained by remote sensing technology is necessary in many applications such as environmental monitoring, agricultural management, urban planning, hydrological applications, soil management, vegetation condition study and suitability analysis. But this information still remains a challenge especially in heterogeneous landscapes covering urban and rural areas due to spectrally similar LULC features. In parallel with technological developments, supplementary data such as satellite-derived spectral indices have begun to be used as additional bands in classification to produce data with high accuracy. The aim of this research is to test the potential of spectral vegetation indices combination with supervised classification methods and to extract reliable LULC information from SPOT 7 multispectral imagery. The Normalized Difference Vegetation Index (NDVI), the Ratio Vegetation Index (RATIO), the Soil Adjusted Vegetation Index (SAVI) were the three vegetation indices used in this study. The classical maximum likelihood classifier (MLC) and support vector machine (SVM) algorithm were applied to classify SPOT 7 image. Catalca is selected region located in the north west of the Istanbul in Turkey, which has complex landscape covering artificial surface, forest and natural area, agricultural field, quarry/mining area, pasture/scrubland and water body. Accuracy assessment of all classified images was performed through overall accuracy and kappa coefficient. The results indicated that the incorporation of these three different vegetation indices decrease the classification accuracy for the MLC and SVM classification. In addition, the maximum likelihood classification slightly outperformed the support vector machine classification approach in both overall accuracy and kappa statistics.

1. INTRODUCTION

Land use/land cover classes are very essential to characterize information of natural environment and human activities on the Earth's surface, to monitor spatial-temporal pattern (Song et al., 2012; Ikiel et al., 2012). In recent decades, remotely sensed data have been widely used to provide the land use and land cover information for analyzing many socio-ecological issues (Hasmadi et al., 2009; Gálvez et al., 2013).

LULC classification with high accuracy and reliable is necessary. However, land use/cover classification remains a difficult task and it is especially challenging in heterogeneous landscapes. The heterogeneous landscape is regarded as a complex combination of different landforms such as residential, commercial and industrial areas, transportation lands, agricultural fields, pasture/scrublands, bare areas, quarry/mining areas, forest and natural areas, water surfaces, and so on (Weng, 2012; Lu et al., 2014). One of the main issues when generating land use/cover maps from complex areas is the confusion of spectral responses from different features. The spectrally similar LU/LC classes are quite common in such regions. The classification accuracy depends on the spatial and spectral resolution using image, classification methods, training data, the seasonal variability in vegetation cover and crop types, soil moisture conditions, etc. (Poursanidis et al. 2015; Ustuner and Sanli, 2015). All these factors limit the accuracy of image classification.

Various researchers have made efforts to improve the accuracy for these areas. The most common approach is to incorporate the

ancillary data before, during and after the classification (Kongwongjan et al., 2012; Mwakapuja et al., 2013; Thakkar et al., 2014; Luo et al., 2015; Mare and Mihai, 2016; Thakkar et al., 2017). One of these ancillary data is derived remote sensing indices using satellite image.

These indices assist to enhance the spectral information and to increase the separability of the classes of interest so it affects the quality of the LULC mapping produced (Ustuner and Sanli, 2015). Vegetation indices are used to separate each land use and land cover from each other, to reduce intra-class variation, to enhance the inter-class variability. In addition, the different combinations of vegetation indices enhance spectral characteristics of some crops while suppressing others.

Vegetation indices are mathematical combinations of different spectral bands that are designed to numerically separate or stretch the pixel value of different features in the image (Vinã et al., 2011; Mwakapuja et al., 2013).

Many indices have been developed that implement various band combinations of different remote sensing data. In this study, three vegetation indices were used such as the Normalized Difference Vegetation Index (NDVI), the Ratio Vegetation Index (RATIO), and the Soil Adjusted Vegetation Index (SAVI).

In this study, the sensitivity and applicability of three different vegetation indices of SPOT 7 multispectral imagery on land use/cover classification were investigated using MLC and SVM

---

* Corresponding author

classification. The image of the study area was classified with both classification methods and thematic maps were produced.

2. STUDY AREA & DATA

For this project, Catalca district has been selected as the study area. Catalca is located in Kocaeli-Catalca part of Marmara region, Istanbul (Figure 1). It is an urban center where established on western boundary of Istanbul's European side. The district consists thirty nine neighborhoods. It is the largest district of Istanbul with a surface area of 1115.50 square kilometers and the land cover is mostly composed forest areas and rural (agricultural) fields. The other land cover/use categories include artificial surfaces, water surfaces, pasture/scrublands, quarry/mining areas. The population of the region is 68.935 according to the population census of 2016 (http://www.tuik.gov.tr). Agriculture, the source of livelihood for the Catalca people, is the most important labor force in the district. It has a great potential in terms of its geographical structure, its natural resources, its ecological conditions and crop quantity/variety in agricultural production. 90% of the Catalca surface area is in the ISKI protection basins. Most of Istanbul's drinking water is supplied from Durusu Lake and Buyukcekmece Dam Lake, which is located near the district borders. Catalca climate has Mediterranean climate characteristics. Besides this, it is seen that the Marmara transition-type climate is one of the end points in the north. In addition, the northern part of the district is exposed to the Black Sea climate zone and is affected by this climate type.
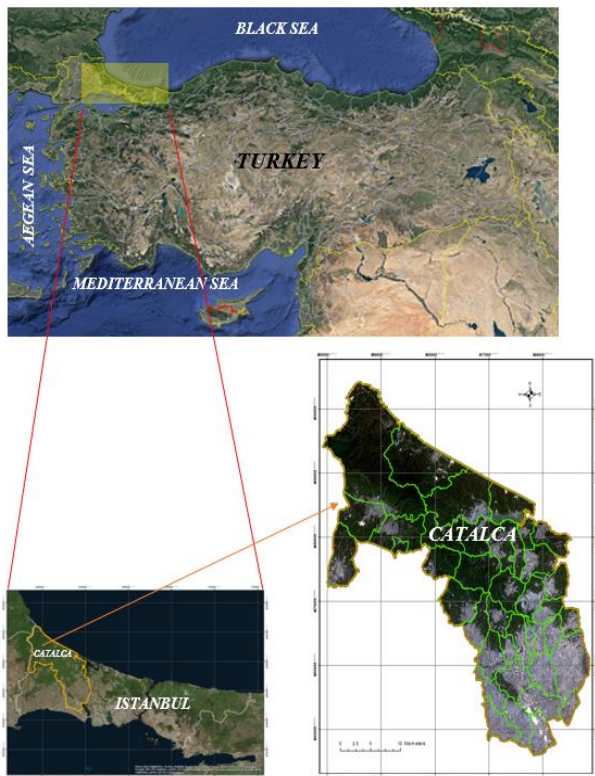

Figure 1. Study Area

SPOT 7 is a high-resolution imaging spacecraft built and operated by AIRBUS Defence & Space was launched on June 30, 2014. The instrument covers five spectral bands – the panchromatic band of 450 to 745 nm and four multispectral bands including blue (455 to 525 nm), green (530 to 590 nm), red (625 to 695 nm) and near infrared (760 to 890 nm). Images of the

panchromatic band can reach 1.5 m resolution and images of multispectral bands obtain 6 m resolution. SPOT 7 satellite has daily revisits everywhere, wide coverage capacity and its imaging swath is 60 km at nadir. Dynamic range at acquisition is 12 bits per pixel (http://www.intelligence-airbusds.com).

In the study, SPOT 7 multispectral satellite image was used to determine the land use/cover area. Properties of the SPOT 7 image were given as below (Table 1).

Product Attribute	Description
Multispectral Imagery (4 bands)	Blue (455 nm - 525 nm)
	Green (530 nm - 590 nm)
	Red (625 nm - 695 nm)
	Near Infrared (760 nm - 890 nm)
Resolution (GSD)	6.0 m (B, G, R, NIR)
Processing Level	ORTHO
Coordinate System	WGS_1984_UTM_Zone_35N
Acquisition Date	16.08.2015

Table 1. Properties of the SPOT 7 image

Detection of sample areas is considered as one of the most important process steps for the purpose of the study. Because, the spatial reference data identified are a basis to characterize the spectral characteristics of consisting of different land use/cover and agricultural product types, to improve the visual interpretation of satellite images, to select sample areas in the image classification stage as well as in the evaluation of the post-classification accuracy. Field survey has been carried out to collect different land cover/use categories location information. For this purpose, partially field studies were made with hands-GPS measuring; partially sample areas were selected through 1: 5000 scaled orthophotos and google earth images. In field trips, coordinates of sample areas were obtained with 5 m accuracy using Garmin-iQue 3600 model hand-GPS. Sample areas were grouped according to their properties and the photos of each sample were stored. The sample site selection was conducted with the acquisition date of satellite image simultaneously. Considering the phenological development process of agricultural products, field studies were carried out between June and August in 2015.

3. METHODOLOGY

The methodology of the study includes four steps which are pre-processing satellite image (radiometric, atmospheric and geometric correction); extraction vegetation indices and producing new data set; classification four original SPOT 7's spectral bands and combination between four original SPOT 7's spectral bands and each vegetation index including NDVI, RATIO and SAVI using maximum likelihood method and support vector machine algorithm; evaluating and comparing classification accuracies.
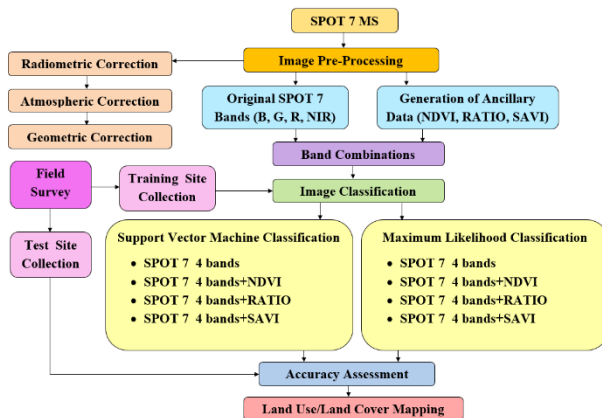
Figure 2. Flowchart of the study

3.1 Image Pre-Processing

In the image pre-processing stage radiometric, atmospheric and geometric correction were applied. Digital number values for the SPOT 7 satellite image were converted to radiance values and FLAASH Atmospheric Correction Model was used for atmospheric correction. Scene center location, sensor type, sensor altitude, ground elevation, pixel size, flight date, flight time GMT, atmospheric model, water column multiplier, initial visibility were used as input parameters in the model. SPOT 7 image was geometrically corrected by image-to-map registration using 1:5000 scaled topographic maps and orthophotos. It took into consideration to be homogeneously distributed on the image of the selected control points. Connection between the image coordinate system and the ground coordinate system was accomplished by polynomial transformation. As a result of the conversion, it was obtained under the root mean square error of 0.5 pixels. Cubic convolution method was used as resampling method.

3.2 Extraction Vegetation Indices

The Normalized Difference Vegetation Index (NDVI), the Ratio Vegetation Index (RATIO) and the Soil Adjusted Vegetation Index (SAVI) were selected as the three different spectral indices in this study. All spectral indices uses spectral bands in the red and the near-infrared (NIR) intervals of the electromagnetic spectrum. Derived vegetation indices and formulations were shown in Table 2 below.

Three different indices of the SPOT 7 image were produced by using the "Band math" function. Then, with "layer stacking", each index image was subjected to classification process by combining with SPOT 7 multispectral bands.

Vegetation Index	Formulation	Referances
NDVI	$NDVI = (NIR - RED) / (NIR + RED)$	Rouse et al. (1974)
RATIO	$RATIO = NIR / RED$	Rouse et al. (1974)
SAVI	$SAVI = [(NIR - RED)(1 + l)] / (NIR + RED + l)$	Huete (1988)

Table 2. Vegetation Indices

For the SPOT 7 image, the band numbers of 3,4 refer to red (625 – 695 nm) and near-infrared (760 – 890 nm), respectively.

where l = soil adjustment factor in SAVI equation

For high vegetation cover the value of l is 0.0 (or 0.25), and for low vegetation cover -1.0. For intermediate vegetation cover l= 0.5, and this value is used most widely. In this study, soil adjustment factor was taken as 0.5.

3.3 Classification Methods

3.3.1 Maximum Likelihood Classification

Maximum likelihood classification (MLC), which is the most common supervised classification method in remote sensing, was used to derive land use/land cover categories of selected study area. MLC classification is based on Bayesian probability theory. The maximum likelihood classifier basically develops a probability function based on inputs from a dataset collected for training. After that the method considers each individual pixel in an image, compares it with known pixels and assigns unknown pixels to a class based on similarity and highest probability to belong to one of the already known classes (Jensen, 2005; Kwesi, 2012). Maximum likelihood classification method involves the estimation of class mean variance and covariance matrices using training patterns chosen from known pixels of each particular class.

3.3.2 Support Vector Machine Classification

The Support Vector Machine (SVM) is a new generation of supervised machine learning method based on the principle of statistical learning theory. The main working principle of SVM is to predict the most appropriate decision function that can distinguish two classes. In other words, it is based on the definition of the hyperplane, which can distinguish the two classes most appropriately (Vapnik, 1995; Vapnik, 2000).

Support vector machine classifier has proven to be more powerful and effective than other statistically based classifiers especially. And also, SVM can work with large dimensional data sets and integrate with auxiliary data set. In addition, it is possible to get accurate classifications with small training sets.

Related parameters vary according to used kernel function such as Linear kernel, Polynomial kernel, Radial Basis Function kernel, Sigmoid kernel (Yang, 2011; Shi and Yang, 2012). In this study, radial basis function kernel was used and optimal parameter values for the selected kernel function determined based on literature review. The parameters for the RBF kernel were set to 0,250 and 100 for γ (gamma= 1/number of the band) and C (error penalty), respectively for SPOT 7's 4 original bands classification, set to 0,200 and 100 for γ (gamma= 1/number of the band) and C (error penalty), respectively for new composite images classification. The pyramid parameter was set to a value of 0 to process the satellite data at full resolution (6 m).

4. RESULTS & CONCLUSIONS

In this research, the applicability and potential of three different vegetation indices of SPOT 7 imagery of Catalca district, Istanbul, Turkey for land use/cover mapping has been examined using support vector machine algorithm and maximum likelihood classification. In the classification, at the first stage, training and test data were prepared to classify satellite image that is covering the study area. Both training and test data were selected at the same number. The same training and test data were used for all classifications problems in this study.

As a result of the classification methods, eight land-use categories were distinguished in the selected region: artificial

surfaces, forest and natural areas, water body, quarry/mining areas, pasture/scrubland, sunflower, other agricultural fields and cloud+shadow (Figure 3).
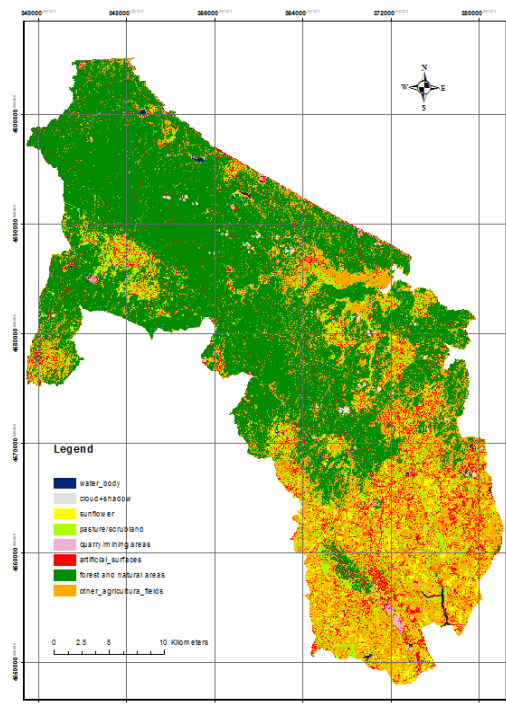
(a)
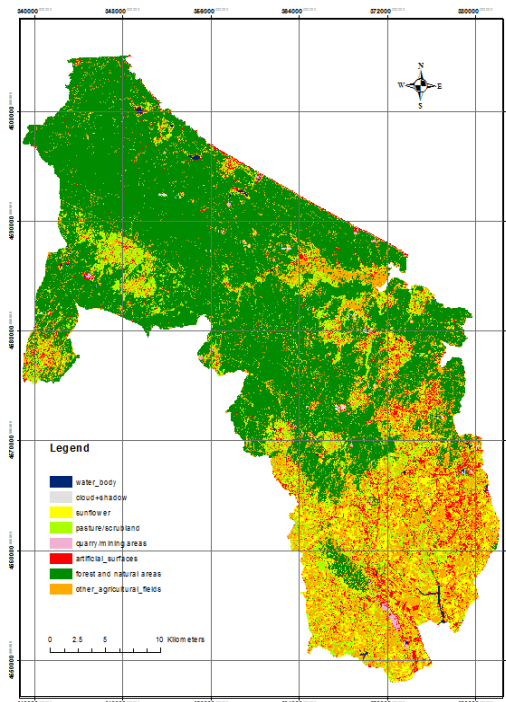
(b)

Figure 3. Land Use/Land Cover Classification with Highest Accuracy for: (a) SVM, and (b) MLC

An accuracy assessment was evaluated for all classification using a standart error matrix. Overall accuracy and kappa statistics were used to compared to determine the performance of the classification results for the selected heterogeneous region (Table 3).

Analysis	O.A	K.C.
MLC (B+G+R+NIR)	86.3451%	0.8462
MLC (B+G+R+NIR+NDVI)	84.8490%	0.8299
MLC (B+G+R+NIR+RATIO)	86.2907%	0.8454
MLC (B+G+R+NIR+SAVI)	84.6677%	0.8280
SVM (B+G+R+NIR)	83.5887%	0.8131
SVM(B+G+R+NIR+NDVI)	83.2351%	0.8090
SVM(B+G+R+NIR+RATIO)	83.1535%	0.8081
SVM(B+G+R+NIR+SAVI)	83.5162%	0.8123

Table 3. Classification Results

The evaluation results demonstrated that the MLC algorithm with an overall accuracy and a kappa coefficient has a higher accuracy in comparison with the SVM algorithm in land use/land cover mapping.

The original SPOT 7's 4 band combination has the highest overall accuracy and kappa coefficient for both classification results such as MLC and SVM.

Vegetation indices derived from original spectral bands of SPOT 7 imagery have different sensitivity on classification methods for MLC and SVM. While original SPOT 7's 4 band combination with the addition of RATIO has the highest classification accuracy for MLC between three indices, original SPOT 7's 4 band combination with the addition of SAVI has the highest classification accuracy for SVM between three indices.

In future work, we plan to evaluate the potential of any other vegetation indices for the same purpose incorporating NIR bands from SPOT 7 imagery using different MLC/SVM options and different classification techniques.

ACKNOWLEDGEMENTS

REFERENCES

Gálvez, J. P., Mas, J. F., Moré, G., Cristóbal, J., Martínez, M. O., Luz, A. C., Guèze, M., Macía, M. J., García, V. R., 2013. Enhanced land use/cover classification of heterogeneous tropical landscapes using support vector machines and textural homogeneity. *International Journal of Applied Earth Observation and Geoinformation*, Vol.23, pp. 372-383.

Hasmadi, M. I., Pakhriazad, H. Z, Shahrin, M. F., 2009. Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Malaysian Journal of Society and Space*, Vol.5, Issue 1, pp. 1-10.

http://www.intelligence-airbusds.com (15 Sep. 2017).

http://www.tuik.gov.tr (15 Sep. 2017).

Huete, A. R., 1988. A soil adjusted vegetation index (SAVI). *Remote Sensing of Environment*, Vol. 25, Issue 3, pp.295-309.

Ikiel, C., Dutucu, A. A., Ustaoglu, B., Kilic, D. E., 2012. Land Use and Land Cover (LULC) Classification Using SPOT-5 Image in the Adapazari Plain and Its Surroundings, Turkey. *The Online Journal of Science and Technology*, Vol.2, Issue 2, pp.37-42.

Jensen, R. J., 2005. Introductory Digital Image Processing. 3rd edition, New York: Prentice Hall.

Kongwongjan, J., Suwanprasit, C., Thongchumnum, P., 2012. Comparison of vegetation indices for mangrove mapping using THEOS data. *Proceedings of the Asia-Pacific Advanced Network*, Vol.33, pp.56-64.

Kwesi, N. I., 2012. Oil Palm Mapping Using Support Vector Machine with Landsat ETM+Data. Enschede, The Netherlands.

Lu, D., Li, G., Kuang, W., Moran, E., 2014. Methods to extract impervious surface areas from satellite images. *International Journal of Digital Earth*, Vol. 7, No. 2, pp. 93-112.

Luo, X., Zhao, W., Wei, S., Fu, Q., 2015. Study on Urban Remote Sensing Classification Based on Improved RBF Network and Normalized Difference Indexes. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 8, No.10, pp. 257-270.

Mare, M. R. R., Mihai, B. A., 2016. Mapping Land Cover Using Remote Sensing Data and GIS Techniques: A Case Study of Prahova Subcarpathians. *Procedia Environmental Sciences*, Vol. 32, pp.244-255.

Mwakapuja, F., Liwa, E., Kashaigili, J., 2013. Usage of Indices for Extraction of Built-up Areas and Vegetation Features from Landsat TM Image: A Case of Dar Es Salaam and Kisarawe Peri-Urban Areas, Tanzania. *International Journal of Agriculture and Forestry*, Vol.3, Issue 7, pp.273-283.

Poursanidis, D., Chrysoulakis, N., Mitraka, Z., 2015. Landsat 8 vs. Landsat 5: A Comparison based on urban and peri-urban land cover mapping. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 35, pp.259-269.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS-1 Symposium, pp.309-317, Washington, DC: NASA.

Shi, D., Yang, X., 2012. Support Vector Machines for Landscape Mapping from Remote Sensor Imagery. Proceedings - AutoCarto. Columbus, Ohio, USA.

Song, X., Duan, Z., Jiang, X., 2012. Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image. *International Journal of Remote Sensing*, Vol. 33, No.10, pp. 3301-3320.

Thakkar, A., Desai, V., Patel, A., Potdar, M., 2014. Land Use/Land Cover Classification of Remote Sensing Data and Their Derived Products in a Heterogeneous Landscape of a Khan-Kali Watershed, Gujarat. *Asian Journal of Geoinformatics*, Vol.14, No.4.

Thakkar, A. K., Desai, V. R., Patel, A., Potdar, M. B., 2017. Post-classification corrections in improving the classification of Land Use/Land Cover of arid region using RS and GIS: The case of Arjuni watershed, Gujarat, India. *The Egyptian Journal of Remote Sensing and Space Sciences*, Vol. 20, pp.79–89.

Ustuner, M., Sanli, F. B., 2015. Testing the Sensitivity of Vegetation Indices for Crop Type Classification using RapidEye Imagery. FIG Working Week 2015, Sofia, Bulgaria, 17-21 May 2015.

Vapnik, V. N., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Vapnik, V. N., 2000. The Nature of Statistical Learning Theory. Vol.2, Springer-Verlag, New York.

Vinã, A., Gitelson, A. A., Robertson, A. L. N., Peng, Y., 2011. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, Vol.115, pp. 3468-3478.

Weng, Q., 2012. Remote Sensing of impervious surface in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, Vol. 117, pp. 34-49.

Yang, X., 2011. Parameterizing Support Vector Machines for Land Cover Classification. *Photogrammetric Engineering and Remote Sensing*, Vol.77, Issue 1, pp. 27-37.