EVALUATION OF SENTINEL-2 MSI DATA FOR LAND USE / LAND COVER CLASSIFICATION USING DIFFERENT VEGETATION INDICES

[1] Filiz BEKTAŞ BALÇIK

[1]İstanbul Technical University, Civil Engineering Faculty, Department of Geomatic Engineering, 34469, İstanbul, Turkey
[1] bektasfi@itu.edu.tr

ABSTRACT: Accurate determination of Land Use/Land Cover (LULC) categories has very important role for environmental monitoring and management applications. Classification of remotely sensed data is one of the popular method to determine LULC information in different scale. Many methods have been developed and applied to classify satellite images. Freely available Sentinel-2 MSI data is new generation remotely sensed data which can be used efficiently to determine the land use and land cover categories for environmental monitoring applications. In this study, Sentinel-2A level 1C data acquired in July 2018 were downloaded from Earth Explorer web page. A test site from Çatalca District of İstanbul, Turkey was selected as the study area. Çatalca is very important district for İstanbul because of its valuable agricultural fields. Different land use/cover types have been defined in the selected study area such as; water surfaces, forest areas, agricultural fields (sunflowers), open mining area, settlements, and road. Sentinel-2 data four bands with 10 m spatial resolution was classified by maximum likelihood classification (MLC) method to investigate the potential of the data to determine the LULC types in selected region, as the first data set. Beside the original bands, different vegetation indices such as Normalized Difference Vegetation Index (NDVI), Green–red normalized difference vegetation index (GRNDVI), were calculated for Sentinel-2 data. These calculated indices and red-edge band were added to the original bands, and classified as the other data sets. The results of these 4 data sets of Sentinel-2 image were compared based on the field collected ground control data and error matrix. Sentinel-2 data had a satisfactory performance in land cover classification; (the overall classification accuracy using the MLC classifier applied data set 2 was higher than the other three data set).

Key Words: GRNDVI, LULC, Maximum likelihood classification, NDVI, Red-Edge band, Sentinel-2 MSI

Farklı Bitki Endeksleri Kullanarak Arazi Kullanımı/ Arazi Sınıflandırması İçin SENTINEL-2 MSI Verilerinin Değerlendirilmesi

ÖZ: Arazi Kullanımı / Arazi Örtüsü (AKAÖ) sınıflarının doğru belirlenmesi, çevresel izleme ve yönetim uygulamaları için çok önemli bir role sahiptir. Uzaktan algılama verilerinin sınıflandırılması, farklı ölçeklerde AKAÖ bilgilerinin belirlenmesi için kullanılan popüler yöntemlerden biridir. Uydu görüntülerini sınıflandırmak için birçok yöntem geliştirilmiş ve uygulanmıştır. Bu çalışmada, Temmuz, 2018 tarihli yeni nesil Sentinel-2 MSI uydu görüntüleri AKAÖ sınıflarını belirlemek amacıyla kullanılmıştır. Çalışma alanı olarak İstanbul Çatalca İlçesi'nden bir test bölgesi seçilmiştir. Çatalca, değerli tarım alanları nedeniyle İstanbul için çok önemli bir bölgedir. Seçilen arazi alanında farklı AKAÖ tipleri tanımlanmıştır; su yüzeyleri, orman alanları, farklı tarım alanları (Ayçiçeği), açık maden alanı, yerleşim yerleri ve yol. Bu çalışma için dört farklı veri seti oluşturulmuş ve seçilen bölgede AKAÖ kategorilerini belirlemek için maksimum benzerlik sınıflaması (MLC) yöntemiyle sınıflandırılmıştır. Birinci veri seti olarak Sentinel-2 uydu görüntüsünün 10 m mekansal çözünürlüğe sahip dört bandı

kullanılmıştır. Orijinal bantların yanı sıra, Sentinel-2 verileri için Normalleştirilmiş Fark Bitki İndeksi (NDVI), ve Yeşil-kırmızı normalleştirilmiş fark bitki indeksi (GRNDVI) gibi farklı indeksler hesaplanmıştır. Bu hesaplanmış indeksler ve kırmızı kenar bandı, orijinal bantlara eklenmiş ve diğer veri setleri olarak sınıflandırılmıştır. Dört veri setine ait sınıflandırma sonuçları, yer kontrol verileri ve hata matrisi kullanılarak karşılaştırılmıştır. Sentinel-2 verileri arazi örtüsü sınıflandırmasında genel doğruluk dört veri seti için tatmin edici bir performans sergilemiştir (% 85 ve üzeri). Veri seti 2 için hesaplanan genel doğruluk ve Kappa değerleri diğer üç veri setinden daha yüksek elde edilmiştir.

Anahtar Kelimeler: *GRNDVI, AKAÖ, Maksimum Benzerlik Sınıflandırması, NDVI, Kırmızı Kenar Bandı, Sentinel-2 MSI,*

INTRODUCTION

Reliable, accurate and up-to date determination of Land Use/Land Cover (LULC) categories has significant role for sustainable urban and rural planning, crop types mapping, agricultural activities, soil management, environmental monitoring and management, climate change modelling, natural hazard management, air quality monitoring (Bektaş Balçık, 2014; Kahya et al., 2017).

Remote sensing is an important tool for LULC applications due to its ability to acquire measurements of land surfaces cost effectively at various spatial and temporal scales (Esetlili et al., 2018). The classification is one of the widely used approaches to extract LULC information from remotely sensed data. Variety of classification methods have been developed and applied to get accurate and reliable LULC categories such as supervised and unsupervised classification, support vector machine classification, artificial neural network classification, and sub-pixel mapping (Lu and Weng, 2007; Esetlili et al., 2018). Therefore, different remote sensing indices (vegetation, built-up, water, soil), image transformation methods, and image fusion methods have been conducted to improve the accuracy of classification (Bektaş Balçık, 2014; Bektaş Balçık and Kuzucu, 2016).

Classification of remotely sensed data which have different spatial, spectral and temporal resolution is one of the common method to determine LULC information in different scale. However, land use/cover classification with high classification accuracy is a difficult image processing step. Especially, the classification of heterogeneous areas that have rural and urban LULC categories together is still faced with mixed pixel problems.

In this study, Maximum Likelihood supervised classification was applied to freely available, new generation Sentinel-2 MSI data. A heterogeneous test site in Çatalca, İstanbul was selected for the study. Sentinel-2 data four bands with 10 m spatial resolution was classified by maximum likelihood classification (MLC) method to investigate the potential of the data to determine the LULC types in selected region, as the first data set. Beside the original bands, different vegetation indices such as Normalized Difference Vegetation Index (NDVI), Green–red normalized difference vegetation index (GRNDVI), were calculated for Sentinel-2 data. These calculated indices and red-edge band were added to the original bands, and classified as the other data sets. The results of these 4 data sets of Sentinel-2 image were compared based on the field collected ground control data and error matrix.

STUDY AREA

Istanbul is one of the most important cities in the world due to its historical, cultural, and industrial importance (Figure 1). İstanbul lies between the continents of Europe and Asia and has an area of approximately 5313 km^2 with the population over 15 million.

For the past 65 years, Istanbul has been experiencing an accelerated urban expansion because of high population rate. Determination of LULC categories and monitoring its' changes have huge importance for sustainable city management.

Figure 1. Study Area (İstanbul/Çatalca Test Site)

In this study, one heterogeneous test region was selected from the European side of the Istanbul. The region includes one of the valuable agricultural areas of mega city İstanbul. Çatalca test site covers the land cover categories of water surfaces, green area, bare lands, agricultural fields, mining areas, road, and artificial surfaces. Figure 2 shows three field study photos that belong to agricultural field samples in the region (From July, 2018).

Figure 2. Agricultural fields A) ploughed B) harvested C) cultivated (sun flower)

MATERIAL

New generation high resolution Sentinel-2 MSI image was used in the study. Sentinel sensors, designed to perform continuous measurements, the sensors have global coverage and provide free data in both the optical (Sentinels-2 and 3) and microwave (Sentinel-1) sections of the electromagnetic spectrum (EM) (Topaloglu et al., 2016).

Sentinel-2A (S-2) satellite, launched in June 2015, has a swath width of 290 km and provides multi-spectral images in 13 spectral bands at different spatial resolutions. These are four visible and near-infrared (NIR) bands at 10 m resolution, six red-edge and shortwave infrared (SWIR) bands at 20 m resolution, and three atmospheric correction bands at 60 m resolution (Drusch et al. 2012; İrfanoğlu and Bektaş Balçık, 2018) (Table 1). Freely available Sentinel-2 multispectral space borne imagery has a 5-day revisit temporal resolution. Sentinel-2 also features those traditional bands, alongside additional bands in the red-edge region that is very sensitive to determine and monitor crop characteristics. Moreover, an acceptable compromise between the revisit time and spatial resolution, with increased spectral abilities for vegetation monitoring compared to previous public domain space-borne imagery. Sentinel-2 provides for compatibility with the historical Landsat data, while also improving measurement capabilities.

Table 1. Spatial and spectral resolution of Sentinel 2A

Band	Resolution	Central Wavelength	Description
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B9	60 m	940 nm	Short Wave Infrared (SWIR)
B10	60 m	1375 nm	Short Wave Infrared (SWIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)
B1	60 m	443 nm	Ultra blue (Coastal and Aerosol)
B5	20 m	705 nm	Visible and Near Infrared (VNIR)
B6	20 m	740 nm	Visible and Near Infrared (VNIR)
B7	20 m	783 nm	Visible and Near Infrared (VNIR)
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B8a	20 m	865 nm	Visible and Near Infrared (VNIR)

In this study, ground truth data were collected in the selected test site. Especially different crop fields were visited and coordinates collected for classification purposes. Field works conducted to determine location of different crop pattern and other land cover types. And also, spatial properties of objects that classified were observed during the fieldwork. Approximately 200 field points were visited and data gathered to use in MLC.

METHOD

Main aim of the study is to test the potential of Sentinel-2 MSI data for LULC classification in rural part of mega city İstanbul. A summary of the methodology used in the study is illustrated in Figure 3 and the image processing details are given in the following sections.
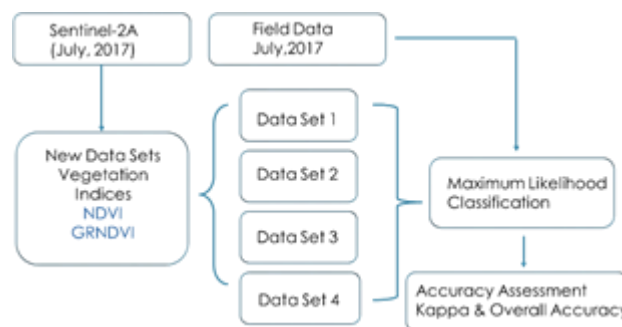
Figure 3. Flowchart of the study

In the first step, four different data set were created by using original bands with 10 m spatial resolution, two vegetation indices and red-edge band (Band 8A).

Data set 1: Sentinel-2 MSI with 4 bands (A)

Data set 2: Sentinel-2 MSI data (4 bands) + NDVI (B)

Data set 3: Sentinel-2 MSI data (4 bands) + GRNDVI (C)

Data set 4: Sentinel-2 MSI data (4 bands) + Red-Edge Band (D)

In this study, Band 2, Band 3, Band 4 and Band 8 were selected as four original bands to create different data set. Two different vegetation indices such as Normalized Difference Vegetation Index (NDVI) and Green Red Normalized Difference Vegetation Index (GRNDVI) were selected four different data set were classified using MLC classification method. Related equations and used bands were given in Table 2.

Table 2. Selected vegetation indices

Vegetation Indices	Equation
Normalized Difference Vegetation Index	$(Band8 - Band4)/(Band\ 8 + Band\ 4)$ Band 4= Red Band 8= Near Infrared
Green-Red Normalized Difference Vegetation Index	$(Band8 - (Band3 + Band4))/(Band\ 8 + (Band\ 3 + Band\ 4))$ Band 3=Green Band 4= Red Band 8= Near Infrared

Supervised classification is a technique that based on the statistics of training areas representing different earth surface objects selected subjectively by users on the basis of their own knowledge or experience (Liu and Mason, 2009). In this study, Maximum Likelihood Classification (MLC), which is the most common supervised classification method in remote sensing, was used to derive land use/cover categories. In this method, each pixel is assigned to the class-category for which the probability of pixel belonging is the highest (Gong, 2002).

For the classification same signature files were used for four data set. Field work synchronized with the remotely sensed data and 200 field data collected during field study using handhold GPS for classification process. Nine general LULC classes including Water Surfaces, Road, Artificial Surfaces, Green Areas, Mining Area, Crop 1, Crop 2, Agricultural Field 1, Agricultural Field 2 were utilized in this case study.

There are a number of equations that can show the level of error statistically, such as producer accuracy, user accuracy, overall accuracy, and Kappa, which can be calculated using the error matrix to determine the accuracy of thematic maps (Foody, 2002). An accuracy assessment was performed using field collected ground-truth data for classification using a standard error matrix. Kappa statistics and overall accuracy were used to determine the performance of the created data set for the selected heterogeneous region (Table 3).

RESULTS AND DISCUSSION

In order to determine the performance of Sentinel-2 MSI data for selected test site, maximum likelihood supervised classification method applied 4 different data set by using field collected ground truth data. Same reference field points were used in classification process for all of the data set. Figure 4 indicates the MLC classification results for 4 data set.
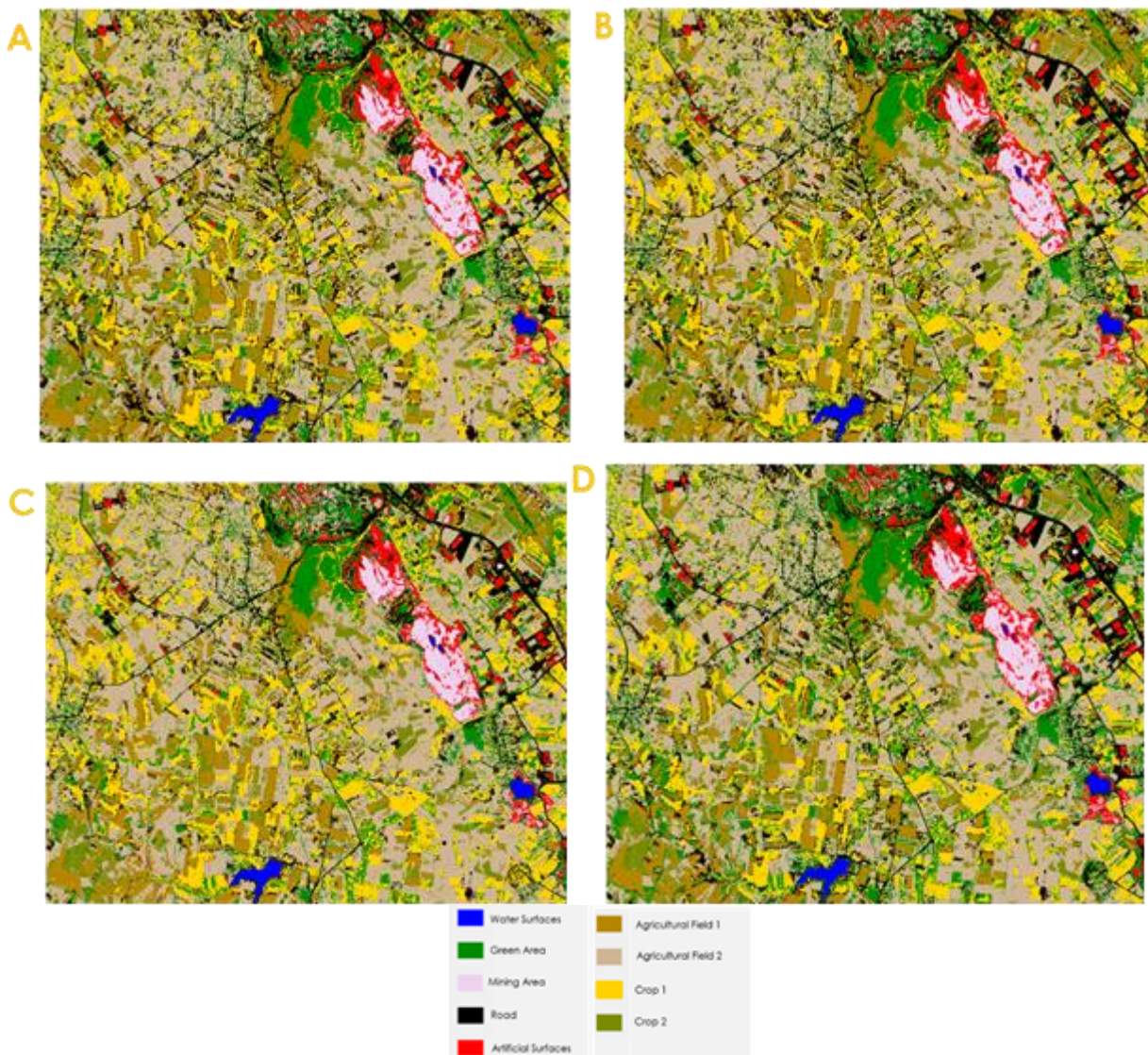
Figure 4. Maximum likelihood classification results A) data set 1 B) data set 2 C) data Set 3 D) data Set 4

LULC images were obtained from Sentinel-2 data sets using pixel based MLC method, and the results were evaluated using accuracy assessment by same test points. Overall accuracy and Kappa statistics were given for 4 different data set in Table 3.

Table 3. Accuracy Assessment Results

Data	Overall Accuracy	Kappa
Data Set 1	86.10 %	0.83
Data Set 2	89.24 %	0.87
Data Set 3	87.82 %	0.86
Data Set 4	86.83 %	0.84

The overall accuracy of the Data Set-2 data was slightly higher than for Data Set -3 data, 89.24% versus 87.82 %, respectively. The approach that yielded the highest value of Kappa coefficient was the Data Set 2 data with 0.87, followed by Data Set 3 with 0.86. From Table 3, it can be observed that the Data Set 1 data had low overall accuracy (86.10%) and kappa coefficient (0.83).

Data Set 2 and Data Set 3 data set created by using NDVI and GRNDVI vegetation indices, respectively. Index added data sets in this paper provided results with higher accuracies than the four original band of Sentinel 2A-MSI (Data Set 1). This result suggests that vegetation indices have potential as an alternative method (over original bands) to extract LULC information from freely available high resolution satellite imagery captured over agricultural lands in Çatalca, Turkey.

CONCLUSION

The aim of this study is to generate LULC images from Sentinel-2 MSI data sets that created by using original bands and vegetation indices using pixel-based MLC supervised classification method, and to reveal which LULC image presents better accuracy results. The results of this study indicate differences in the classification accuracy between four different data set that created using original four bands of Sentinel-2 A, selected vegetation indices and red-edge band for mapping LULC types using pixel based Maximum Likelihood Supervised classification method. The results highlighted that all created data set as highly useful (over 85 %) in mapping of LULC types in the study region. The success of created data set also underlines the potential of Sentinel-2A MSI data for mapping LULC in heterogeneous test site.

Future work would test the classification potential of using the different band composites of Sentinel-2 images for whole Çatalca Region in İstanbul. Different transformation methods and remote sensing indices and pan sharpening methods would test with advance classification methods such as random forest, SVM and object based classification.

REFERENCES

Bektas Balcik, F., Karakacan Kuzucu, A., 2016, "Determination of Land Cover/Land Use Using SPOT 7 Data With Supervised Classification Methods", *GeoAdvances Workshop*, Istanbul.

Bektas Balcik, F., 2014, "Determining the Impact of Urban Components on Land Surface Temperature of Istanbul by using Remote Sensing Indices", *Environmental Monitoring and Assessment*, Vol. 186, pp. 859-872.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012, "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services", *Remote Sensing of Environment*, Vol. 120, pp. 25–36.

Esetlili, T., Bektaş Balçık, F., Balı,k Şanlı, F., Üstüner, M., Kalkan, K., Göksel, Ç., Gazioğlu, C., Kurucu Y., 2018, "Comparison of Object and Pixel-Based Classifications for Mapping Crops Using Rapideye Imagery: A Case Study of Menemen Plain, Turkey", *International Journal of Environment and Geoinformatics (IJEGEO)*, Vol. 5(2), pp. 231- 243.

Foody, G.M., 2002, "Status of Land Cover Classification Accuracy Assessment", *Remote Sensing of Environment*, Vol. 80, pp. 185– 201.

Gong, P., 2002, Remote Sensing and Image Analysis Textbook, http://nature.berkeley.edu/~penggong/textbo ok/chapter7/html/sect73.htm.

İrfanoğlu, F., Bektaş Balçık, F., 2018, "Determination of LULC Categories using Object Based Classification and Sentinel-2 MSI data (In Turkish)", *UZALCBS*, Eskisehir, Turkiye.

Lu, D., Weng, Q., 2007, "A Survey of Image Classification Methods And Techniques For Improving Classification Performance", *International Journal of Remote Sensing*, Vol. 28 (5), pp. 823–870.

Liu, J. G., Mason, P. J., 2009, *Essential Image Processing and GIS for Remote Sensing*, John Wiley Sons, Inc.

Kahya, C., Bektaş Balçık, F., Öztaner Y. B., Ozcomak D., Seker, D.Z., 2017, "Spatio Temporal Analysis of PM 2.5 Over Marmara Region, Turkey", *Fresenius Environmental Bulletin*, Vol. 26,1, pp. 310-317.

Topaloğlu, H., Sertel, E., Musaoğlu, N., 2016, "Assessment of Classification Accuracies of Sentinel-2 and Landsat-8 Data For Land Cover / Use Mapping", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLI-B8, 2016 XXIII ISPRS Congress, Prague, Czech Republic.