ASSESSMENT OF CLASSIFICATION ACCURACIES OF SENTINEL-2 AND LANDSAT-8 DATA FOR LAND COVER / USE MAPPING

Raziye Hale TOPALOĞLU[a], Elif SERTEL[a], Nebiye MUSAOĞLU[a] - (topaloglur, sertele, musaoglune) @itu.edu.tr

[a] ITU, Civil Engineering Faculty, Geomatics Engineering Department, 34469, Istanbul, Turkey

Commission VIII, WG VIII/8

KEY WORDS: SENTINEL-2, LANDSAT-8, Remote Sensing, Supervised classification, Land Cover/ Use Mapping

ABSTRACT:

This study aims to compare classification accuracies of land cover/use maps created from Sentinel-2 and Landsat-8 data. Istanbul metropolitan city of Turkey, with a population of around 14 million, having different landscape characteristics was selected as study area. Water, forest, agricultural areas, grasslands, transport network, urban, airport- industrial units and barren land- mine land cover/use classes adapted from CORINE nomenclature were used as main land cover/use classes to identify. To fulfil the aims of this research, recently acquired dated 08/02/2016 Sentinel-2 and dated 22/02/2016 Landsat-8 images of Istanbul were obtained and image pre-processing steps like atmospheric and geometric correction were employed. Both Sentinel-2 and Landsat-8 images were resampled to 30m pixel size after geometric correction and similar spectral bands for both satellites were selected to create a similar base for these multi-sensor data. Maximum Likelihood (MLC) and Support Vector Machine (SVM) supervised classification methods were applied to both data sets to accurately identify eight different land cover/ use classes. Error matrix was created using same reference points for Sentinel-2 and Landsat-8 classifications. After the classification accuracy, results were compared to find out the best approach to create current land cover/use map of the region. The results of MLC and SVM classification methods were compared for both images.

1. INTRODUCTION

Land cover and land use mapping and analysis are vital for different environmental and mapping applications. Mapping and monitoring of land cover have been widely recognized as an important scientific goal since created information could be used to support environmental and atmospheric models, decision making procedures etc. (Wulder et al., 2008; Zhu and Woodcock, 2014). Since it is possible to obtain rapid, periodic and accurate data from remote sensing system, satellite images are important source of creating land cover/use information. Land cover/ use studies are multidisciplinary in nature resulted in varying participants from decision makers to private companies, and academia to the public.

Basically, land cover/ use maps are created using image classification approaches. Land cover/ use classification is an important and challenging research field in remote sensing and multi-temporal land cover/use maps emphasizing the land changes have become a basic information source to analyse several environmental issues. Image classification has a long history in the remote sensing and is the fundamental for many applications, such as carbon modeling, land cover/use change, forest monitoring and management, and crop yield estimation (Woodcock et al., 2001; Zhu et al., 2012). Many classifiers have been developed and tested for land cover classification, such as Maximum Likelihood, neural networks, decision trees, support vector machines, and random forest (Zhu et al., 2012).

There are two international land cover classification systems, the Food and Agriculture Organization of the United Nations (FAO)/ United Nations Environment Programme (UNEP)/ Global Land Cover Network (GLCN) so-called Land Cover Classification System (LCCS) and Coordination of Information on the Environment CORINE Land Cover (CLC) of the European Union (Weber; 2009). CORINE program is the biggest European land Cover/ use monitoring project targeting to create periodic and standard geographical information on land cover/ use for its member countries in order to support environmental policies in various areas such as agriculture. CORINE Land Cover Nomenclature is a 3-level hierarchical classification system, composed of 5 classes in Level l, 15 classes in level 2 and 44 classes at the 3rd and most detailed level (Alp et al., 2015). CORINE land cover nomenclature provides a conceptual framework for understanding the different types of land cover.

Accuracy assessment is an established component of the process of creating and distributing thematic maps (Stehman and Wickham, 2011). Validating land cover and land-cover change products provides critical data quality information to users and producers of these maps. A common practice is to conduct an accuracy assessment of the classified map based on a spatially explicit comparison of the map to a higher-quality reference data. The goal of an accuracy assessment is to estimate a population error matrix and various descriptive measures of accuracy derived from this error matrix to quantitatively provide accuracy information of each land cover/ use categories (Stehman, 2013).

The main purpose of this study is to evaluate and compare the classification accuracies of new generation Sentinel-2 and Landsat-8 optical satellite images in Istanbul metropolitan area. Two different classification approaches namely Maximum Likelihood (MLC) and Support Vector Machine (SVM) were applied to multi-sensor data to classify eight different land categories using same training sites. Classification accuracies of two different methods applied to Sentinel-2 and Landsat-8 were compared to evaluate the impact and performance of two algorithms and different remotely sensed data.

2. STUDY AREA AND DATA

2.1 Study Area

The study area is located in Istanbul in Turkey (Figure 1) and covers a total area of 309.02 km². Istanbul is the fifth most crowded city in the world and the population of the city is 14,160,467 which is 18.5% of the total population of Turkey (URL-1). Istanbul is one of the most important cities in the world due to its historical, cultural, and industrial importance along with its natural characteristics and location where the two continents; Europe and Asia meet.
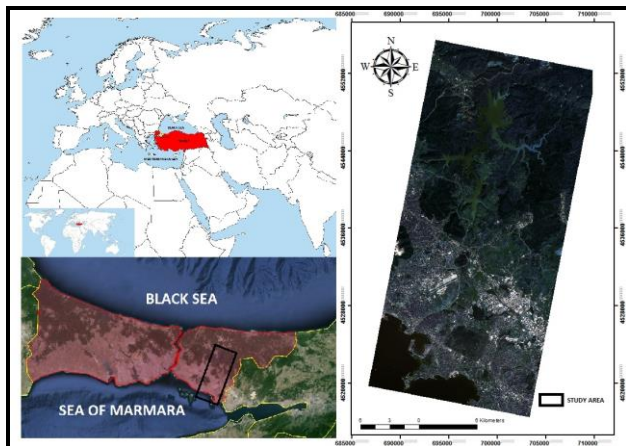

Figure 1. Study Area

The study region has different landscape characteristics and contains the diversity of the land cover/ use categories such as, forest, agricultural areas, urban, water etc. in the Anatolian side of Istanbul.

2.2 Data

Landsat-8 and Sentinel-2 satellite images were used in the study. Over the 40 years, Landsat data provide one of the most valuable datasets for mapping and monitoring the Earth surface (Kennedy et al., 2014; Zhu et al., 2015). Landsat-8, a NASA and USGS collaboration, acquires global moderate-resolution measurements of the Earth's surface in the visible, near-infrared, short wave, and thermal infrared.

Landsat-8 has increased capabilities such as new spectral bands in the blue part and cirrus cloud-detection portion of the spectrum, two new thermal bands, improved sensor signal-to-noise performance and several developments in radiometric resolution and duty cycle that allows a significant increase in collection of number of images per day (Roy et al., 2014). Landsat 8 satellite, launched on 11.02.2013, has totally 11 bands and spatial resolution ranging from 30 m to 100 m. Landsat-8 provides 30 m spatial resolution optical imagery on 8 spectral bands via the Operational Land Imager (OLI) sensor with 16 days temporal resolution.

Sentinel-2 mission, launched on 23.06.2015, is a land monitoring constellation of two satellites (Sentinel- 2a and Sentinel- 2b) providing global optical imagery with 13 spectral bands using MSI (Multispectral Imager) instrument. Temporal resolution of Sentinel-2 is 10 days with one satellite and will be 5 days with 2 satellites that will create huge amount of Earth Observation data for several research and application projects with a spatial resolution ranging from 10 m to 60 m. Sentinel-2

mission provides continuity for the current SPOT and LANDSAT missions with an expected lifetime of 7.25 years, over a 20-year period. Since both Sentinel-2 and Landsat-8 cover spectral range between 440 and 2300 nm (Table 1), these two data sets could complement each other.

In this study, Landsat 8 and Sentinel 2 images acquired respectively on 22.02.2016 and 08.02.2016 were used 6 multispectral bands having similar spectral resolution and covered by both satellites and 6 multispectral bands which the same spectral range of this two images were used (Table 1) (2, 3, 4, 5, 6, 7 for Landsat8; 2, 3, 4, 8a, 11, 12 for Sentinel-2).

Table 1: Landsat OLI/TIRS, and Sentinel 2 spectral bands.

OLI/TIRS bands (µm)	Resolution (m)	Sentinel 2 bands (µm)	Resolution (m)
Band 1 (0.43–0.45)	30	Band 1 (0.433–0.453)	60
Band 2 (0.45–0.51)	30	**Band 2 (0.458–0.523)**	10
Band 3 (0.53–0.59)	30	**Band 3 (0.543–0.578)**	10
Band 4 (0.64–0.67)	30	**Band 4 (0.650–0.680)**	10
Band 5 (0.85–0.88)	30	**Band 5 (0.698–0.713)**	20
Band 6 (1.57–1.65)	30	**Band 6 (0.733–0.748)**	20
Band 7 (2.11–2.29)	30	**Band 7 (0.765–0.785)**	20
Band 8 (0.50–0.68)	15	Band 8 (0.785–0.900)	10
Band 9 (1.36–1.38)	30	Band 8a (0.855–0.875)	20
Band 10 (10.60–11.19)	100	Band 9 (0.930–0.950)	60
Band 11 (11.50–12.51)	100	Band 10 (1.365–1.385)	60
		Band 11 (1.565–1.655)	20
		Band 12 (2.100–2.280)	20

3. METHODOLOGY

After obtaining Landsat-8 and Sentinel-2 data, different processes were applied to create land cover maps from these data using two different classification approaches. Figure 2 summarizes the whole flow chart of this research.
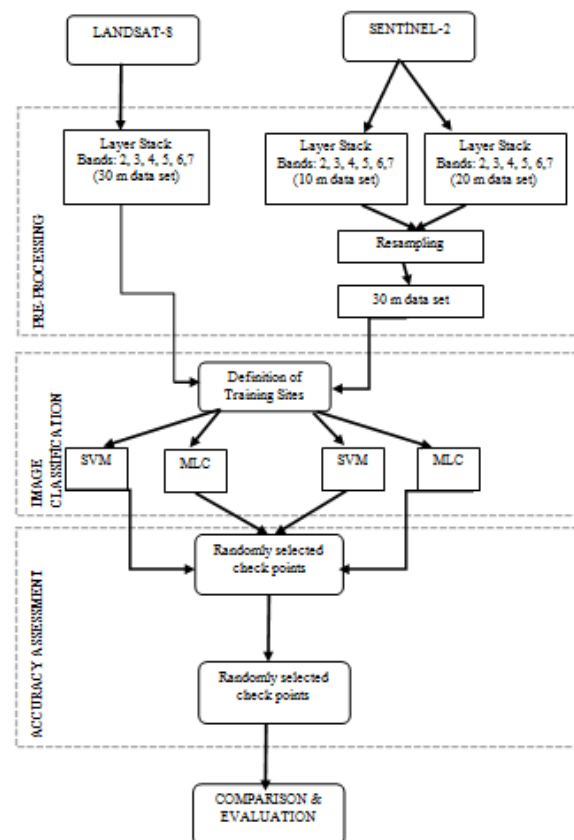

Figure 2. The whole flow chart of this research

Firstly, similar spectral bands of Landsat-8 and Sentinel-2 were obtained and layers stacked. For Landsat-8; band 2, band 3, band 4, band 5, band 6 and band 7 having 30 m spectral resolution, for Sentinel-2; band 2, band 3 and band 4 which having 10 m spectral resolution and band 8a, band 11, band 12 with 20 m spectral resolution were stacked together. Afterwards, 10 m and 20 m spectral resolution bands of Sentinel-2 image were resampled to 30 m and 6 multispectral band image with 30 m spatial resolution was obtained using layer stacking. Common spatial resolution data set of similar spectral band range has been created to compare the classification accuracy of results obtained from Landsat-8 and Sentinel-2 datasets. (Figure 3).
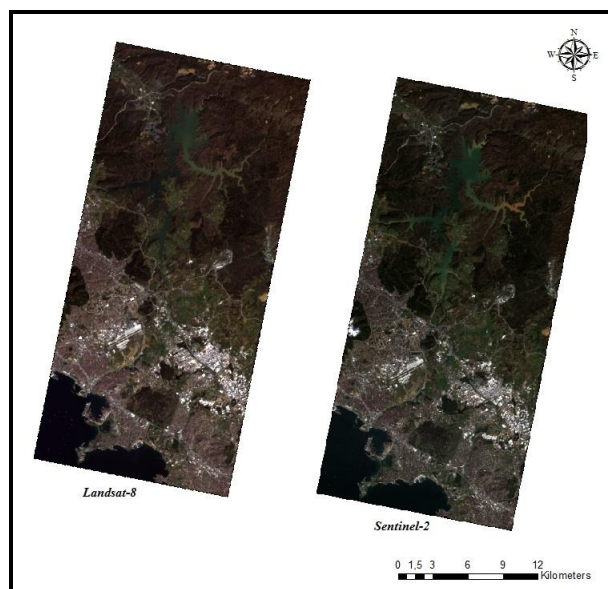

Figure 3. Landsat-8 and Sentinel-2 with 6 multispectral bands.

Both satellite images were converted to Top-of-Atmosphere reflectance values. Atmospheric correction was not necessary because these images were quite clear within the study area. Landsat-8 and Sentinel-2 images were acquired as geometrically corrected and both images are in Universal Transform Mercator (UTM) coordinate system and World Geodetic System (WGS-84, Zone 35) and geometrically matching each other with sub-pixel accuracy.

8 land cover classes namely water, forest, agricultural, grassland, transport network, urban, industrial unit- airport and barren land-mine CORINE nomenclature were used in this study. Training sites were created for each class using the reference data obtained from Google Earth. Satellite images are available in Google Earth with high spatial and temporal resolution which were fit to aim of this research to be used as reference. For each land cover/use class, at least 30 samples were collected and totally, 4300 samples were used for the classification both images in ENVI©. Same training sites were used for both data sets and two different classification algorithms.

Maximum Likelihood (MLC) and Support Vector Machine (SVM) methods were applied to 30 m resampled 6 band Landsat-8 and Sentinel-2 images to create comparable land cover maps of the study region.

Support vector machines

The support vector machine (SVM) approach is one of the successful machine learning algorithms for classifying high-dimensional remotely sensed data sets. Training sites are selected to create hyperplanes to separate dataset into predefined number of classes represented by training areas (Vapnik, 1979; Mountrakis et al., 2011). The penalty parameter in SVM is important to control certain level of misclassification which can not be separated by training sets. A soft margin could be created using penalty parameter that allows some misclassifications. The principal advantage of SVM is that it could successfully work with small number of training samples (Taskın et al., 2011).

Maximum Likelihood classification

Maximum Likelihood classification is a well-known supervised algorithm. According to Erdas (1999), the weighted distance or likelihood **D** of a unknown measurement vector **X** belonging to one of the known classes **Mc** is computed based on the Bayesian equation:

$$D = \ln (ac) - [0.5 \ln (|Covc|)] - [0.5(X - Mc)T(Covc - 1)(X - Mc)]$$

MLC method considers the variance–covariance (Cov) within the class distributions and assumed data as normally distributed. Better results could be obtained with MLC compared to other known parametric classifies such as Minimum Distance, Mahalonobis Distance if data set is normally distributed (Erdas, 1999). However, for data with a non-normal distribution, MLC might not produce high accuracy values (Otukei and Blaschke).

At the end of the land cover classification step, the classification accuracy was evaluated using the error matrix. A separate but same data sets was used for accuracy assessment using ERDAS©. Reference points were selected from Google Earth through visual interpretation.

4. RESULTS

4.1 Classification results

Both Landsat-8 and Sentinel-2 images were classified using SVM and MLC to create land cover maps of the study region. In total, four classifications were performed and eight land cover classes were created (Figure 4). Training sites were generated for water, forest, grassland, agricultural, transport network, urban, airport and industrial areas and mine and barren land classes and same training sites were used for all classifications. Due to mixed pixels in 30 m resolution images, especially for Maximum Likelihood methods; it is not possible to clearly identify transport network, industrial areas, airport, barren land and mine classes separately. Therefore, classes having similar spectral response and landscape characteristics were merged into one class such as airport and industrial areas as one class and mine sites and barren land as one class. Although, creation of this integrated classes improved the accuracy of the classification, there have been mixing problems between different classes. Mostly, SVM produces better results compared to MLC for all related land cover classes.
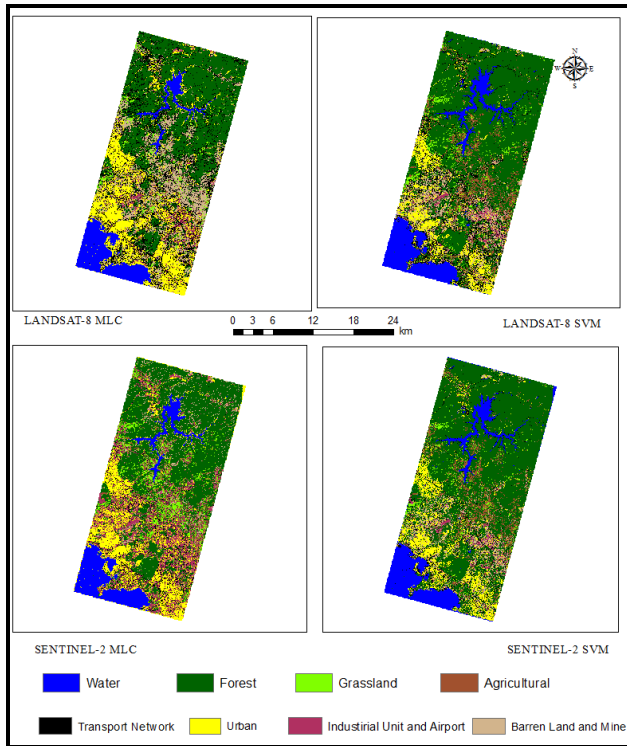
Figure 4. Land Classification using MLC and SVM

4.2 Accuracy Assessment

To determine the accuracy of each classification and class, thematic accuracy assessment was performed. For this purpose, firstly a reference data set including a total of 240 points was created. These points were selected over different locations representing different land cover/use classes. Same reference data was used for the accuracy assessment of Landsat-8 and Sentinel-2 images and different classification algorithms.

Error matrix of each classification was created and total, user's and producer's accuracy values of each class were analysed to evaluate classification accuracies.

Results are as shown in the following tables (Table 2, 3, 4, 5).

Overall accuracy of MLC method for Landsat-8 data was found as 70.60 % with a kappa value of 0.6580; whilst, the overall accuracy was 76.40 % with a kappa value of 0.7212 for Sentinel-2 data for MLC. Accuracy of MLC results obtained from Sentinel-2 data are higher than MLC results of Landsat-8 data.

In addition, overall accuracy of SVM method for Landsat-8 data was found as 81.67 % with a kappa value of 0.7896; however, the overall accuracy was 84.17 % with a kappa value of 0.8190 for Sentinel-2 data by using the same approach.

Although user' s and producer' s accuracy values of each class are higher than 70 % with SVM approach for both Sentinel-2 and Landsat-8 data, user' s and producer' s accuracy values with MLC approach range from 50 % to 100 %. User's and producer's accuracy values in SVM were similar for both Landsat-8 and Sentinel-2 data; however, user's and producer's accuracy values in MLC have huge differences such as for

grassland, agriculture, transport and urban classes (Table 1 and Table 3).

Classification results illustrated that water class could be successfully identified using both MLC and SVM for both satellite images. Although producer's accuracy values are good for urban class with MLC, user's accuracy values are comparatively lower for both Landsat-8 and Sentinel-2 data. The best accuracy results for urban areas are obtained from Sentinel- 2 data with SVM method.

Among the four different classifications, the highest accuracy was obtained from SVM results obtained from Sentinel-2 data.

Table 1: Error matrix for Landsat-8 classification using MLC.

Classified Data	Water	Forest	Grassland	Agricultural	Transport N.	Urban	Airport and Indust.	Mine and Barren L.	Classified Totals	Producer's Accuracy
			Reference Data							
Water	25	0	0	0	0	0	0	0	25	86.21%
Forest	0	17	3	0	1	0	0	0	21	51.52%
Grassland	0	1	22	1	0	0	0	0	24	59.46%
Agricultural	0	1	1	10	0	0	0	0	12	45.45%
Transport N.	1	7	4	0	19	2	4	0	37	70.37%
Urban	3	2	3	0	6	44	6	1	55	91.67%
Airport and Indust.	0	0	0	1	1	0	22	3	27	66.67%
Mine and Barren L.	0	5	4	10	0	2	1	17	39	80.95%
Reference Totals	29	33	37	22	27	48	33	21	Overall Classification Accuracy	70.60%
User's Accuracy	100.00%	80.95%	91.67%	83.33%	51.35%	67.69%	81.48%	43.59%	Overall Kappa Statistics	0.6580

Table 2: Error matrix for Landsat-8 classification using SVM.

Classified Data	Water	Forest	Grassland	Agricultural	Transport N.	Urban	Airport and Indust.	Mine and Barren L.	Classified Totals	Producer's Accuracy
			Reference Data							
Water	27	0	0	0	0	0	0	0	27	100.00%
Forest	0	35	1	2	0	0	1	0	39	74.47%
Grassland	0	2	22	1	0	0	0	0	25	88.00%
Agricultural	0	2	2	24	0	0	1	0	29	77.42%
Transport N.	0	3	0	2	24	4	2	0	35	82.76%
Urban	0	1	0	1	1	19	1	0	23	73.08%
Airport and Indust.	0	0	0	0	2	1	19	2	24	70.37%
Mine and Barren L.	0	4	0	1	2	2	3	26	38	92.86%
Reference Totals	27	47	25	31	29	26	27	28	Overall Classification Accuracy	81.67%
User's Accuracy	100.00%	89.74%	88.00%	82.76%	68.57%	82.61%	79.17%	68.42%	Overall Kappa Statistics	0.7896

Table 3: Error matrix for Sentinel-2 classification using MLC.

Classified Data	Water	Forest	Grassland	Agricultural	Transport N.	Urban	Airport and Indust.	Mine and Barren L.	Classified Totals	Producer's Accuracy
			Reference Data							
Water	25	0	0	0	0	0	0	0	25	100.00%
Forest	0	27	1	0	0	0	0	0	28	58.70%
Grassland	0	3	35	6	0	0	0	0	44	71.43%
Agricultural	0	0	1	9	0	0	0	1	11	50.00%
Transport N.	0	3	2	1	8	0	2	0	16	66.67%
Urban	0	8	5	1	1	36	4	1	56	100.00%
Airport and Indust.	0	4	4	1	3	0	43	2	57	82.69%
Mine and Barren L.	0	1	1	0	0	0	3	8	13	66.67%
Reference Totals	25	46	49	18	12	36	52	12	Overall Classification Accuracy	76.40%
User's Accuracy	100.00%	96.43%	79.55%	81.82%	50.00%	64.29%	75.44%	61.54%	Overall Kappa Statistics	0.7212

Table 4: Error Matrix for Sentinel-2 classification using SVM.

Classified Data	Reference Data								Classified Totals	Producer's Accuracy
	Water	Forest	Grassland	Agricultural	Transport N.	Urban	Airport and Indust.	Mine and Barren L.		
Water	30	0	0	0	0	0	0	0	30	100.00%
Forest	0	28	1	1	0	0	0	0	30	7.78%
Grassland	0	2	26	2	0	0	0	0	30	81.25%
Agricultural	0	1	5	24	0	0	0	0	30	75.00%
Transport N.	0	2	0	1	25	0	2	0	30	83.33%
Urban	0	1	0	2	0	26	1	2	30	92.86%
Airport and Indust.	0	0	0	0	3	2	22	3	30	78.57%
Mine and Barren L.	0	2	0	2	2	0	3	21	30	87.50%
Reference Totals	30	36	32	32	30	28	28	26	Overall Classification Accuracy	84.17%
User's Accuracy	100.00%	93.33%	86.67%	80.00%	83.33%	86.67%	73.33%	70.00%	Overall Kappa Statistics	0.8190

5. DISCUSSION AND CONCLUSIONS

As a result of the study; for Landsat-8 image, 70.60% classification accuracy with MLC method and 81.67% classification accuracy with SVM method have been acquired. On the other hand, for Sentinel-2 image, 76.40% classification accuracy with MLC method and 84.17% classification accuracy with SVM method have been acquired. For both data sets, SVM produced better results compared to MLC. The best overall accuracy and kappa values were obtained from SVM application to Sentinel-2 data. Considering that similar spectral bands of Sentinel-2 and Landsat-8 were used in classification, it is assumed that spatial resolution is also contributed to better accuracy obtained from Sentinel-2 classification taking into account that resampled 30m Sentinel data inherited the original spatial signal.

According to overall classification accuracies; it can be seen that, for both MLC and SVM classification methods, image classification accuracies of Sentinel-2 dataset is better than Landsat-8 dataset. Although, Sentinel-2 dataset resampled to 30 m using Landsat-8 as a base, overall classification accuracies of Sentinel-2 are much higher than Landsat-8 data.

REFERENCES

Alp, G., Algan, I. Y., & Sertel, E. (2015, July). Determination of agricultural land changes in Mugla, Turkey using remotely sensed data and Corine methodology. In *Agro-Geoinformatics (Agro-geoinformatics), 2015 Fourth International Conference on* (pp. 7-10). IEEE.

Erdas Inc. *Erdas Field Guide*. Erdas Inc., Atlanta, Georgia, 1999.

Weber, L. J. (2009, April). Land Cover Classification for Land Cover Accounting. In *14th Meeting of the London Group on Environmental Accounting Canberra* (pp. 27-30).

Kennedy, R.E., Andréfouët, S., Cohen, W.B., Gómez, C., Griffiths, P., Hais, M., et al. (2014). Bringing an ecological view of change to Landsat-based remote sensing. Frontiers in Ecology and the Environment, 12(6), 339–346.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247-259.

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*, *12*, S27-S31.

Roy, D. P., Wulder, M. A., Loveland, T. R., Woodcock, C. E., Allen, R. G., Anderson, M. C., ... & Scambos, T. A. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, *145*, 154-172.

Stehman, S. V., & Wickham, J. D. (2011). Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment. *Remote Sensing of Environment*, *115*(12), 3044-3055.

Stehman, S. V. (2013). Estimating area from an accuracy assessment error matrix. Remote Sensing of Environment, 132, 202-211.

Taskin Kaya, G., Musaoglu, N., & Ersoy, O. K. (2011). Damage assessment of 2010 Haiti earthquake with post-earthquake satellite image by support vector selection and adaptation. *Photogrammetric Engineering & Remote Sensing*, *77*(10), 1025-1035.

URL-1: <https://www.tuik.gov.tr>, [Accessed Date 28.03.2016].

Vapnik, V., 1979. Estimation of Dependences Based on Empirical Data. Nauka, Moscow, pp. 5165–5184, 27 (in Russian) (English translation: Springer Verlag, New York, 1982).

Woodcock, C. E., Macomber, S. A., Pax-Lenney, M., & Cohen, W. B. (2001). Large area monitoring of temperate forest change using Landsat data: Generalization across sensors, time and space. Remote Sensing of Environment, 78(1–2), 194–203.

Wulder, M.A., White, J. C., Goward, S. N., Masek, J. G., Irons, J. R., Herold, M., et al. (2008). Landsat continuity: Issues and opportunities for land cover monitoring. Remote Sensing of Environment, 112(3), 955–969.

Zhu, Z., Woodcock, C. E., Rogan, J., & Kellndorfer, J. (2012). Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. *Remote Sensing of Environment*, *117*, 72-82.

Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote sensing of Environment*, *144*, 152-171.

Zhu, Z., Wang, S., & Woodcock, C. E. (2015). Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment*, *159*, 269-277.