IJEG

International Journal of Engineering and Geosciences

International Journal of Engineering and Geosciences (IJEG),
Vol; 4; , Issue; 1, pp. 1-7, Month, 2019, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.404426

OUTLIER DETECTION OF LAND SURFACE TEMPERATURE: KÜÇÜKÇEKMECE REGION

Kuşak, L., [1]* Küçükali, U. F. ,[2]

[1]Istanbul Aydin University, Architecture and Design Faculty, Department of Interior Architecture, Istanbul, Turkey (lutfiyekusak@aydin.edu.tr); **ORCID ID 0000-0002-7265-245X**

[2]Istanbul Aydin University, Architecture and Design Faculty, Department of Architecture, Istanbul, Turkey (ufkucukali@aydin.edu.tr); **ORCID ID 0000 – 0002 – 2715 – 7046**

*Corresponding Author, Received: 12/03/2018, Accepted: 25/05/2018

ABSTRACT: Unplanned and rapid urbanization is one of the reasons for the rising surface temperature in urban areas. There is a large amount of literature demonstrating the association of urbanization with surface temperatures. Küçükçekmece Lake, an important lake that has been meeting the utility water needs of Istanbul, and the unplanned and rapid urbanization around it, has been affected by this inevitable change for years. Although surface temperatures generally correlate strongly with each other, very high and very low temperature values should not be disregarded and need to be investigated. The current study was conducted with the assumption that these values could be outliers; thus, they were analyzed using the box plot method for the selected region. Correlations between land surface temperature (LST) values obtained for Küçükçekmece and its vicinity were examined using Landsat Operational Land Imager (OLI) images from June 20, 2016 and June 23, 2017, and LST outliers and regions with common outliers on both days were determined. In the study, 310 LST outliers were identified for June 20, 2016 and 34 LST outliers for June 23, 2017; in both images, 33 outliers were found to be common and these clustered in two different buildings. The reasons for the outliers outside the standard surface temperature values as well as the recommended solutions were discussed.

Keywords: *LST, Outlier, Box Plot, Urbanization, Land use/cover*

1. INTRODUCTION

According to the data for January 2018 from the Worldometers website (URL 1), 53.25% of the world's population, about 7.5 billion people, lives in cities. Rapid population growth in cities, which is caused by the fact that the human population abandons rural areas and migrates to urban areas, results in changes in land use as well as unplanned or poorly planned urbanization (Deng et al., 2009), leading to many ecological (Zhao et al., 2006; Wang et al., 2008), climatic (Liu et al., 2017), sociological (Vlahov and Galea, 2002), and environmental problems (Liao et al., 2017). Changes in land use/cover and the resultant problems have been analyzed using ecological (Nacef et al., 2016), economic, and sociological data, and remote sensing images (Shen et al., 2016).

One of the biggest problems brought about by rapid urbanization and consequent changes in land use/cover are changes in surface temperatures in intensely urbanized areas (Weng, 2001; Jenerette et al., 2007; Cui and Shi, 2012; Tayyebi et al., 2018). Numerous studies have been conducted to observe these changes and determine whether there is a relationship between land use/cover and the change in land surface temperatures (Chen et al., 2006; Hasanlou and Mostofi, 2015; Alhawitti and Mitsova, 2016; Li et al., 2017).

The developments in remote sensing technology allow land surface temperatures (LST) to be obtained using various satellite and airborne sensors that support the acquisition of thermal infrared information (Kaya et al., 2012; Chen et al., 2013; Gunawardena et al., 2017; Ranagalage et al., 2017).

Atmospheric conditions, geographical factors, and urbanization affect surface temperature values. However, as analyzed in this study, the LST values are affected by unusual local changes and phenomena, such as very low or very high values, which may be called outliers.

Outliers, which are often used in the field of statistics, are widely used in the field of machine learning owing to developments in software and hardware technology that allow storage and processing of much more data. Outliers, also called abnormalities, discordants, deviants, or anomalies, are used in a wide range of activities, including data security, medical diagnosis, various fields of physics, network security, credit card and insurance fraud detection, critical systems, defense activities, and earth science (Iglewicz and Hoaglin, 1993; Bramer, 2007; Kumar and Mathur 2014, URL 2, URL 3). Statistical or model-based models, proximity-based models, linear regression models (Principal Component Analysis (PCA), Least Mean Squares (LMS)), information theory models, high-dimensional outlier-detection methods, and many other methods were developed for the detection of outliers, with data type and data size (univariate/multivariate) being very important in the selection of the method to be used (Seo, 2006). Statistical or model-based models, including standard deviation (SD), Z-scores, and box plots, are preferred for the detection of outliers for univariate data (Olson and Delen, 2008; Han et al., 2011).

A box plot is a graphical method for displaying variation in a group of numerical data. In the literature, box plot is widely used in environmental studies (Xie et al., 1999; Zhang et al., 2009), urban studies (Cipolla and Maglionico, 2014), and other scientific studies concerned with outlier detection.

The outlier obtained in some studies may contain highly valuable information (Rousseeuw and Hubert, 2017). As non-spatial data such as temperature, pressure, and image pixel color intensity are obtained on the basis of spatial locations, unusual local changes in these data can be seen as outliers (Aggarwal, 2015).

In this study, the LST values obtained with Landsat 8 Operational Land Imager (OLI) satellite images were prepared given that unusual local changes might affect them. As the obtained LST values are univariate, the box plot method (Tukey, 1997) is used in this study. In addition, box plot was preferred here because it is less sensitive to extreme values (Mendenhall and Sincich 2016).

The interquartile range (IQR) calculations used to construct the box plot were created, and the outlier values obtained from the analysis were discussed and evaluated at the end of the study by taking into account the spatial characteristics (anthropogenic effect, slope, aspect, and water cooling island (WCI)) (Du et al., 2016).

2. DATA AND METHODS

The study comprises three main stages, including acquisition of Normalized Difference Vegetation Index (NDVI), LST images and values of indices; determination of LST values causing outliers by the box plot method; and discussion of potential reasons that might be responsible for LST values described as outliers.

2.1 Study Area

An area within a distance of six km from the Küçükçekmece Lake and its coastline was chosen as the study area. The Küçükçekmece District is situated on the shores of the Marmara Sea on the European side of Istanbul (Fig. 1). The study area is located between 40º5730 N–41º0630 N and 28º390 E–28º5030 E coordinates.

As stated in the 1/100,000 environmental plan of Istanbul, in accordance with east-west development, the study area includes Küçükçekmece Lake, residential areas, industrial areas, agricultural areas, sports facilities, educational facilities, and airport functions. As is the case in Istanbul, the study area has faced highly intensive urbanization for the last 30 years (Kucukali and Kuşak, 2017).

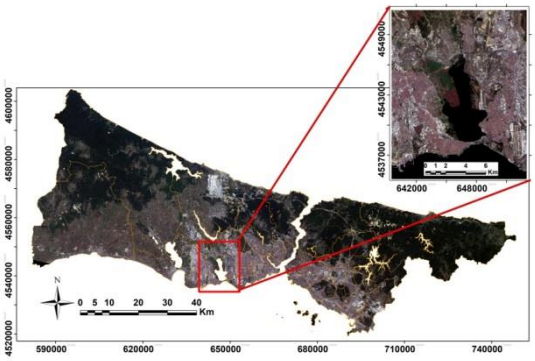


Figure. 1 Study area

2.2 Data

In the study, June 20, 2016 and June 23, 2017 Landsat 8 OLI images were used to obtain LST values and determine outlier values. Using Landsat 8 OLI images, the June 20, 2016 and June 23, 2017 images were prepared by the EROS Science Processing Architecture (ESPA) on-demand interface that provides Landsat higher-level science data products, including Climate Data Records (top of atmosphere (TOA) reflectance, brightness temperature, cloud masks, and surface reflectance) and spectral indices (e.g., NDVI, Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI), and Normalized Burn Ratio (NBR)). Processed bands, surface reflectance, and NDVI were used in this study. The Landsat 8 satellite payload consists of two science instruments: the OLI and the Thermal Infrared Sensor (TIRS).

Additionally, the slope and aspect properties obtained from the ASTER DEM image, Istanbul environmental plan of 1/100,000 scale and report, and current Google Earth images were employed to interpret the relationship between outlier values and land use.

2.3 Methods

Digital number values for OLI and TIRS bands should first be converted into spectral radiance values for TOA to obtain LST images and spectral indices. The LST values are calculated using brightness temperature (BT) values. When BT values are being calculated, TOA data and thermal conversation constant values are utilized. As the file prepared by ESPA contains TOA reflectance, BT, and surface reflectance NDVI, these steps were omitted. LST was calculated using the data prepared by ESPA.

LST based on satellite brightness temperature (BT) was computed using the following equation Eq. (1). Therefore, emissivity values are also defined in the study.

$$LST = \frac{BT}{1+(\lambda+\frac{BT}{\alpha})\ln\varepsilon} \quad (1)$$

BT is the effective at-satellite temperature in Kelvin, λ is the wavelength of the emitted radiance in meters, α=1.438 x 10^{-2}mK, and ε is the surface emissivity.

There is a relationship between land cover and emissivity. The use of NDVI has often been the preferred method for demonstrating this relationship (Van de Griend and Owe 1993; Valor and Caselles, 1996). The NDVI-based threshold method was then developed for NDVI-based studies (Sobrino et al., 2001; Sobrino et al., 2004). In this method, three main threshold ranges consisting of soil, vegetation, and mixed areas were determined. Certain emissivity values are used for soil and vegetation, but they are created with the emissivity formula calculated using NDVI values for mixed areas. Currently, the three main threshold ranges defined by Sobrino (2004) have increased and different emissivity values and calculation methods for these ranges are being developed (Stathopoulou et al., 2007; Stathopoulou and Cartalis, 2007; Xie et al., 2012; Tang et al., 2015).

In this study, the method adopted by Shen, which is used in the LST calculation, is applied. Here, emissivity values of 0.9923, 0.923, and 0.986 for water areas, urban impervious and bare soil areas, and vegetation areas, respectively, are accepted. The formula in Eq. (2) for mixed areas is also used (Shen et al., 2016).

$$\varepsilon = 1.0094 + 0.047\ln(NDVI) \quad (2)$$

Obtaining the pixel values for the LST images is another step of the study to detect outliers. The tools in ARCGIS 10.2 were exploited to obtain pixel values. With the help of spatial analysis tools used concurrently, multipoint values were obtained from the clipped images. LST, NDVI, Normalized Difference Built Index (NDBI), slope (URL 4), aspect (URL 5) pixel values, and the distance of their spatial location to the lake were obtained using a spatial analysis tool for both dates. During interpretation of the values obtained, correlations between data were investigated first by forming a correlation matrix.

A box plot was used to determine the pixel values of the obtained values, which show temperature values contrary to the data set, and to discuss the situations that could be caused by these outliers. MS Excel and RStudio software are preferred for calculating outliers. IQR is used in the box plot method developed by Tukey. The interquartile range is the range between the first and third quartiles (the edges of the box). Tukey took into account any data point that fell outside either 1.5 times the IQR below the first or 1.5 times the IQR above the third quartile to be “outside” or “far out” (Tukey, 1977).

3. RESULTS

NDVI is an index often used in the detection of live vegetation in the area. NDVI values of the study area from two different days were evaluated. It is clear that the NDVI values were close to 1 in and around the northern parts of the lake in both images (Fig. 2). NDVI is also used intensively to calculate LST values.

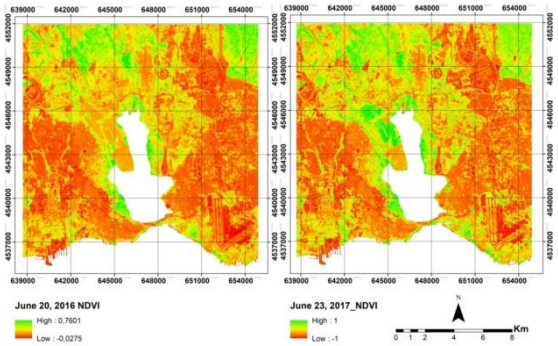


Figure. 2 NDVI images

LST images generated using Eq. (1) and Eq. (2) as well as other emissivity values (Fig. 3) showed that the surface temperature in close vicinity to the lake was lower than those of the inner parts of the study area, as expected. Similarly, the surface temperature of the northern areas, where the NDVI value was close to 1, was also low. On the other hand, the surface temperature was noted to be high in the eastern and western parts of the lake.

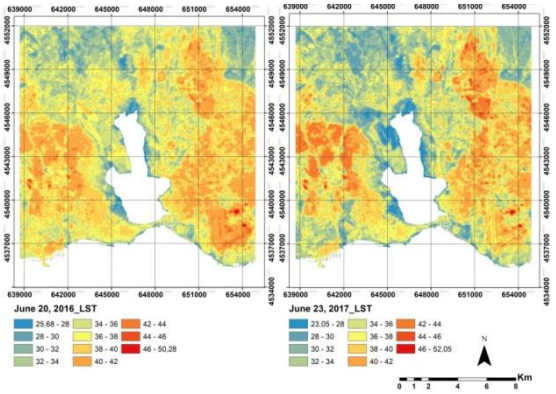


Figure. 3 LST images

Following visual interpretation, 256592 values were obtained by the ARCGIS spatial analyst tool to determine the numeric counterparts of surface temperatures and LST average, SD, Min., Max., and R2 values of the two different LST images obtained (Table 1). Table 1 demonstrates that there is an average drop of 0.82 °C in the selected region in the LST image for June 23, 2017. The LST correlation for June 20, 2016 was 0.965, and that for June 23, 2017 was 0.988.

Table 1 LST values and correlations

Year	Average	SD	Min	Max	R^2
6/20/16	36.73	3.27	25.68	50.28	0.965
6/23/17	35.91	3.93	23.24	52.05	0.988

The correlations between LST, NDVI, NDBI, slope, aspect pixel values, and distance of their spatial location to the lake were investigated. As expected, the highest negative correlation was found between NDVI and LST. In addition, there was no significant relationship between the other parameters. Moreover, the correlation between LST values of the two different years was 0.889.

Despite this high correlation, very low and high LST values were found when the LST values were listed in increasing order. A box plot analysis (Fig. 4) was performed to determine whether these values were outlier values.

According to the results, the upper level LST value was 46.72 °C, while the lower level LST value was 27.12 °C in 2016. For the year 2017, the upper level LST value was found to be 48.42 °C, while the lower level LST value was found to be 23.66 °C.

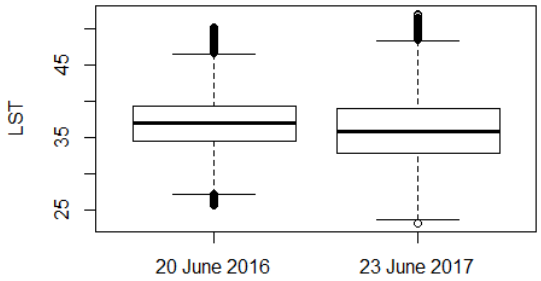


Figure. 4 The results of the Box Plot method

In 2016, 234 items of data were below the lower level, while 84 were above the upper level. A total of 318 outlier data were found.

One datum of 2017 was below the lower level, whereas 33 were above the upper level. A total of 34 data were determined to be outliers.

After the outlier points determined for the year 2016 were excluded, the correlation value increased to 0.9675, while the correlation value for the year 2017 increased to 0.9883. LST correlations for both days were unchanged and remained at 0.889, compared to the correlations before exclusion of the outlier data.

However, when outlier data are the focus (Figs. 5 and 6), this may lead to results that are too important to neglect, especially in micro-scale studies, although there is not any noticeable change.

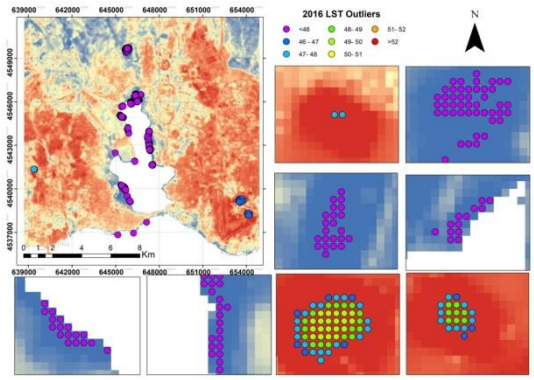


Figure. 5 2016 LST outliers images

When 318 outlier points determined by the box plot analysis of the LST image from June 20, 2016 were examined, 234 points at the lower level were general distributed along the lake shoreline, seaside, and wetlands, whereas 84 outlier points were in the Atatürk Airport area, located in the southeastern part of the image.

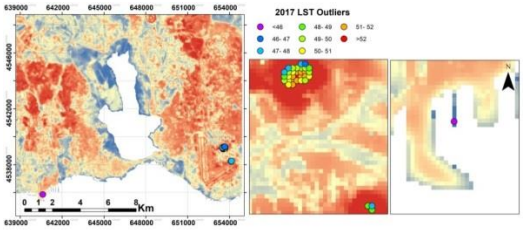


Figure. 6 2017 LST outliers images

When 34 outlier points detected by the box plot analysis of the LST image from June 23, 2017 were examined, one point at the lower level was at the seaside, whereas 33 outlier points were densely distributed in the Atatürk Airport area, located in the southwestern part of the image. When the points where the overlapping outlier points determined by the box plot analysis of the LST images from June 20, 2016 and June 23, 2017 were examined, it is noteworthy that the points at the lower level did not actually overlap, whereas 33 points at the upper level did overlap. When position checks of these points were made, they corresponded to the CNR EXPO (first building) and Atatürk Airport Warehouse (second building); the reasons for this were then investigated (Fig. 7).

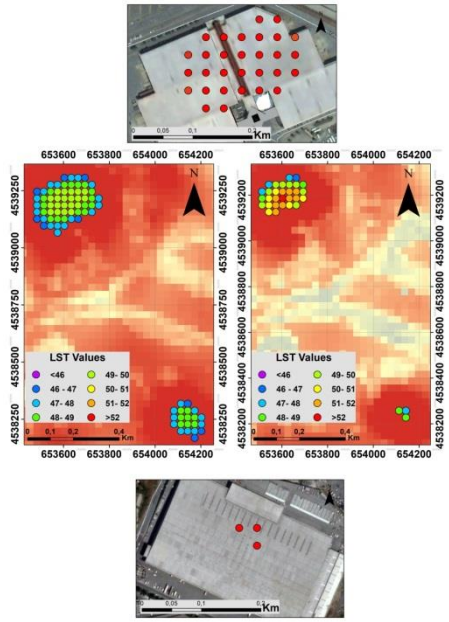


Figure. 7 CNR EXPO (First Building), Atatürk Airport Depo (Second Building) outlier images

4. DISCUSSION

In this study, the box plot method developed by Tukey was used because the data were univariate, and took into account that unusual local changes might have affected the LST values obtained from Landsat 8 OLI satellite images prepared to include the area within a distance of six km from Küçükçekmece Lake and its vicinity (Tukey 1997). The IQR calculations used to construct the box plot were undertaken and the outlier values obtained from the analysis were discussed and evaluated at the end of the study by taking into account the spatial characteristics (anthropogenic effect, slope, aspect, and WCI).

According to the results obtained by the box plot method of LST data from 2016, outlier temperature values lower than 27.12 °C and higher than 46.72 °C were ascertained. Visual assessment of these outlier data revealed that the outlier values lower than 27.12 °C in particular were located on the lake shoreline and seaside. This result was ascribed to similar atmospheric conditions, similar geomorphologic properties of the locations of these outliers, and the cooling effects of the water surface. On the other hand, the outliers higher than 46.72 °C were densely clustered in the southeastern region of the study area (Fig. 7). These outliers were located in the CNR EXPO (first building) and Atatürk Airport Warehouse (second building), which are in the region of the Atatürk Airport.

Additionally, on the basis of the obtained by the box plot method of LST data in 2017, outlier temperature values higher than 48.42 °C were determined. They completely matched the locations of the 2016 LST outliers, which were the CNR EXPO (first building) and the Atatürk Airport Depot (second building) buildings. This result suggests that the massive size of these buildings, determined according to the results of both analyses, the technical properties of their roofing material, and the positioning of the building's air-conditioning elements have an impact on the formation of outliers.

5. CONCLUSION

In this type of study, temperature, pressure, wind, and pixel values which can be more significant in the interpretation of the results, obtained from satellite images should not be evaluated independently of spatial location characteristics. Thus, the correlations between slopes, aspect, lake distance, NDVI, NDBI, and LST values of the study area were identified in the preliminary phase of the study. However, except for the high correlation between generally accepted NDVI, NDBI, and LST indices for the selected study area, there was no relationship between slope, aspect, distance to the lake, and LST; therefore, these were excluded from the evaluation, but were included as secondary data in the interpretation of the results.

The structures identified as a result of the examination of the outlier locations determined by box plot analysis of the data obtained from the LST images were found to have a very large mass in terms of their architectural design. In such large flat-roofed structures, the reflection and solar orientation design of the roof, planting of an appropriate selection of species in the vicinity of the building, selection of proper roofing material and color, green roof applications, correct positioning of air-conditioning elements, and restoration and improvement of structures may help to reduce the LST value.

It is recommended that further research studies include continuous observation and control of locations with outliers on a seasonal and annual basis, and taking of precautions, where necessary, by expanding the methodology, scale, and spatial and temporal characteristics of the current study.

REFERENCES

Aggarwal, C. C. (2015). Outlier analysis. In Data mining (pp. 237-263). Springer, Cham.

Alhawitti, R. H., and Mitsova, D. (2016). Using Landsat-8 data to explore the correlation between urban heart island and urban land uses. IJRET: International Journal of Research in Engineering and Technology, 5(3), 457-466.

Bramer, M. (2007). Principles of data mining (Vol. 180). London: Springer.

Chen, L., Li, M., Huang, F., and Xu, S. (2013, December). Relationships of LST to NDBI and NDVI in Wuhan City based on Landsat ETM+ image. In Image and Signal Processing (CISP), 2013 6th International Congress on (Vol. 2, pp. 840-845). IEEE.

Chen, X. L., Zhao, H. M., Li, P. X., and Yin, Z. Y. (2006). Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. Remote sensing of environment, 104(2), 133-146.

Cipolla, S. S., and Maglionico, M. (2014). Heat recovery from urban wastewater: Analysis of the variability of flow rate and temperature. Energy and Buildings, 69, 122-130.

Cui, L., and Shi, J. (2012). Urbanization and its environmental effects in Shanghai, China. Urban Climate, 2, 1-15.

Deng, J. S., Wang, K., Hong, Y., and Qi, J. G. (2009). Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization. Landscape and urban planning, 92(3-4), 187-198.

Du, H., Song, X., Jiang, H., Kan, Z., Wang, Z., and Cai, Y. (2016). Research on the cooling island effects of water body: A case study of Shanghai, China. Ecological indicators, 67, 31-38.

Gunawardena, K. R., Wells, M. J., and Kershaw, T. (2017). Utilising green and bluespace to mitigate urban heat island intensity. Science of the Total Environment, 584, 1040-1055.

Han, J., Pei, J., and Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

Hasanlou, M., and Mostofi, N. (2015). Investigating urban heat island estimation and relation between various land cover indices in Tehran city using Landsat 8 imagery. In Proceeding.

Iglewicz, Boris and David Hoaglin (1993), How to Detect and Handle Outliers. American Society for Quality Control, Milwaukee WI.

Jenerette, G. D., Harlan, S. L., Brazel, A., Jones, N., Larsen, L., and Stefanov, W. L. (2007). Regional relationships between surface temperature, vegetation, and human settlement in a rapidly urbanizing ecosystem. Landscape ecology, 22(3), 353-365.

Kaya, S., Basar, U. G., Karaca, M., and Seker, D. Z. (2012). Assessment of urban heat islands using remotely sensed data. Ecology, 21(84), 107-113.

Kucukali, U. F., and Kuşak, L. (2017). Environmental, Social, and Economic Indicators of Urban Land Use Conflicts. Urbanization and Its Impact on Socio-Economic Growth in Developing Regions, pp. 285, IGI Global,USA.

Kumar, M. and Mathur, R., (2014). April. Outlier detection based fault-detection algorithm for cloud computing. In Convergence of Technology (I2CT), 2014 International Conference for (pp. 1-4). IEEE.

Li, W., Cao, Q., Lang, K., and Wu, J. (2017). Linking potential heat source and sink to urban heat island: Heterogeneous effects of landscape pattern on land surface temperature. Science of the Total Environment, 586, 457-465.

Liao, J., Jia, Y., Tang, L., Huang, Q., Wang, Y., Huang, N., and Hua, L. (2017). Assessment of urbanization-induced ecological risks in an area with significant ecosystem services based on land use/cover change scenarios. International Journal of Sustainable Development and World Ecology, 1-10.

Liu, Y., Chen, Z. M., Xiao, H., Yang, W., Liu, D., and Chen, B. (2017). Driving factors of carbon dioxide emissions in China: an empirical study using 2006-2010 provincial data. Frontiers of Earth Science, 11(1), 156-161.

Mendenhall, W. M., and Sincich, T. L. (2016). Statistics for Engineering and the Sciences. Chapman and Hall/CRC.

Nacef, L., Bachari, N. E. I., Bouda, A., & Boubnia, R. (2016). "Variability and decadal evolution of temperature and salinity in the mediterranean sea surface". International Journal of Engineering and Geosciences, 1(1), 20-29.

Olson, D. L., and Delen, D. (2008). Advanced data mining techniques. Springer Science and Business Media.

Ranagalage, M., Estoque, R. C., and Murayama, Y. (2017). An urban heat island study of the Colombo metropolitan area, Sri Lanka, based on Landsat data (1997–2017). ISPRS International Journal of Geo-Information, 6(7), 189.

Rousseeuw, P. J., and Hubert, M. (2017). Anomaly detection by robust statistics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.

Seo, S. (2006). A review and comparison of methods for detecting outliers in univariate data sets (Doctoral dissertation, University of Pittsburgh).

Shen, H., Huang, L., Zhang, L., Wu, P., and Zeng, C. (2016). Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: A 26-year case study of the city of Wuhan in China. Remote Sensing of Environment, 172, 109-125.

Sobrino, J. A., Jiménez-Muñoz, J. C., and Paolini, L. (2004). Land surface temperature retrieval from LANDSAT TM 5. Remote Sensing of environment, 90(4), 434-440.

Sobrino, J. A., Raissouni, N., and Li, Z. L. (2001). A comparative study of land surface emissivity retrieval from NOAA data. Remote Sensing of Environment, 75(2), 256-266.

Stathopoulou, M., and Cartalis, C. (2007). Daytime urban heat islands from Landsat ETM+ and Corine land cover data: An application to major cities in Greece. Solar Energy, 81(3), 358-368.

Stathopoulou, M., Cartalis, C., and Petrakis, M. (2007). Integrating Corine Land Cover data and Landsat TM for surface emissivity definition: application to the urban area of Athens, Greece. International Journal of Remote Sensing, 28(15), 3291-3304.

Tang, B. H., Shao, K., Li, Z. L., Wu, H., and Tang, R. (2015). An improved NDVI-based threshold method for estimating land surface emissivity using MODIS satellite data. International Journal of Remote Sensing, 36(19-20), 4864-4878.

Tayyebi, A., Shafizadeh-Moghadam, H., and Tayyebi, A. H. (2018). Analyzing long-term spatio-temporal patterns of land surface temperature in response to rapid urbanization in the mega-city of Tehran. Land Use Policy, 71, 459-469.

Tukey, J. W. (1977). Box-and-whisker plots. Exploratory data analysis, 39-43.

Valor, E., and Caselles, V. (1996). Mapping land surface emissivity from NDVI: Application to European, African, and South American areas. Remote sensing of Environment, 57(3), 167-184.

Van de Griend, A. A., and Owe, M. (1993). On the relationship between thermal emissivity and the normalized difference vegetation index for natural surfaces. International Journal of remote sensing, 14(6), 1119-1131.

Vlahov, D. and Galea, S. (2002). Urbanization, urbanicity, and health. Journal of Urban Health, 79(1), pp.S1-S12.

Wang, J., Da, L., Song, K., and Li, B. L. (2008). Temporal variations of surface water quality in urban, suburban and rural areas during rapid urbanization in Shanghai, China. Environmental Pollution, 152(2), 387-393.

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.

Xie, S., Dearing, J. A., Bloemendal, J., & Boyle, J. F. (1999). Association between the organic matter content and magnetic properties in street dust, Liverpool, UK. Science of the Total Environment, 241(1-3), 205-214.

Xie, Q., Zhou, Z., Teng, M., and Wang, P. (2012). A multi-temporal Landsat TM data analysis of the impact of land use and land cover changes on the urban heat island effect. J. Food Agric. Environ, 10(2), 803-809.

Zhang, C., Tang, Y., Luo, L., & Xu, W. (2009). Outlier identification and visualization for Pb concentrations in urban soils and its implications for identification of potential contaminated land. Environmental Pollution, 157(11), 3083-3090.

Zhao, S., Da, L., Tang, Z., Fang, H., Song, K., and Fang, J. (2006). Ecological consequences of rapid urban expansion: Shanghai, China. Frontiers in Ecology and the Environment, 4(7), 341-346.

URL 1. Web Map Tile Service, http://www.worldometers.info [Accessed 1 Jan 2018]

URL 2. Web Map Tile Service, http://www.kdd.org [Accessed 1 Jan 2018]

URL 3. Web Map Tile Service, https://towardsdatascience.com [Accessed 12 Dec 2017]

URL 4. Web Map Tile Service, http://esdac.jrc.ec.europa.eu [Accessed 30 Jun 2017]

URL 5. Web Map Tile Service, http://eusoils.jrc.ec.europa.eu [Accessed 30 Jun 2017]

International Journal of Engineering and Geosciences (IJEG),
Vol; 4, Issue; 1, pp. 008-015, February, 2019, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.416336

EDUCATION FOR REAL ESTATE VALUATION IN TURKEY

Mehmet Ertaş[1]

[1]Selçuk University, Technical Sciences Vocational Higher School, City and Region Planning Department, Geomatic & Cadastre Program, Konya, Turkey (mertas@selcuk.edu.tr); **ORCID ID: 0000-0002-9458-3395**

*Corresponding Author, Received: 18/04/2018, Accepted: 05/07/2018

ABSTRACT: In scientific terms, the real estate valuation is first started in Spain (to recover Castilian treasure from the bad condition) on 10th October 1749 in the world. Later, the "value" concept came to agenda to cadastral legislation of all European countries. For this reason the "value" issue was added cadastral legislation to France in 1808, Germany in 1872 and Holland in 1810.

The first study based on the cadastral valuation in the Ottoman Empire mentioned in the Cadastre Law dated on 5th February 1912, but has dealt with the technical and legal aspects of the cadastre without taking of valuation into account. After the Republic of Turkey founded, although some of the Cadastre Laws (10th April 1924, 22nd April 1925, 15th December 1934, 15th March 1950, 21st June1987 and 22nd February 2005) include the term ”valuation” again only the technical and legal aspects were used since they are not exactly evaluative.

The first scientific studies on the valuation in Turkey was started to at Yıldız Technical University on graduate training and undergraduate training with the course called Land Valuation in 1978 and Real Estate Valuation in 1988. The studies remained at the academic level.

Within the Customs Union agreement in 1st January 1996 between Turkey and European Union, a great deal of foreign capital is started to enter into our country. Since there is no reliable, real time and correct records of the real estate values, some of the foreign capital owners were hesitated to enter our country. As a result, the need for the “honest people who can make valuations according to the scientific criteria” became obvious. Because of the applications of the foreign capital owners to the Turkish Government, the authorization of making licensing exams was given to the Capital Markets Board in order to find out the required valuers for the country. The need is partly satisfied by the licensing exams. However the necessity and importance of the valuation training in Turkey is started with these licensing exams. Today, valuation training is given at many universities (in associate, undergraduate and postgraduate degrees) in our country. The subject of this paper is on the quality and content of the training given in our country.

Keywords: *Valuation, Valuation Education, Real Estate Valuation Expert*

1. INTRODUCTION

As is known, Market Value is defined as " an estimated amount of a property on the valuation date after the appropriate marketing between independent willing seller and willing buyer without any enforcement and under the circumstances that parties will not affected form any relationship, and behave in the frame of a treaty that is informed, foresighted and rightminded manner" (SPK, 2005).

Again according to IVS (International Valuation Standards) Professional Property Valuer is a person who has the quality, ability and experience on real estate that is to given as an assurance to loan or mortgage including transferring the property rights, property as a subject to trial or tax payment problem to be solved, and many kinds of operations that need generally property valuation like the properties dealed as immovable assets in financial reports (URL10). In addition, Professional Property Valuers must have the sepecial expertise needed for other valuation categories on personal property, financial rights and benefits.

Real Estate Valuation Experts are *special occupational group* that carry out the valuation and making reports from the results as a economic activity. Professional valuation experts must have quite difficult educationing, education and pass from qualification tests and show their skills. In addition, these people must obey working principles (ethical principles and competencies) and professional activity standards and protect them and conform to Generally Accepted Valuation Principles.

Valuation Expert is a person who has the necessary qualification, skill and experience to make valuation (SPK, 2005). Working as a Valuation Expert in some countries requires a license.

Valuation Expert is a person who:

a) Has received an education degree from an institution or equivalent recognized center of education with appropriate academic qualifications,
b) Has gained experience and skills to do valuation in the market and asset categories,
c) Knows, understands and uses accepted methods and techniques properly for reliable valuation,
d) Member of a recognized national professional valuation organizations,
e) Follows a professional educationing program during his career,
f) Conducts the requirements of the Behaving Rules,
g) Respected (SPK, 2005).

With all these explanations, the Association of Turkish Valuation Experts (ATVE) searches some minimum conditions for those who want to carry out the valuation profession and sets out rules for it. Making and recording the real estate valuation according to scientific principles are important as;

a) The real estate tax, which is an important revenue source for municipalities,
b) To fight against criminals,
c) The income tax for the economy,
d) Confidence for the market,
e) Pursuing rant for a consistent land policy.

However real estate valuation in Turkey is scattered in legal and institutional structure aspect (AYOP, 2012). It is determined that the problems related to valuation are education level, the museum, the academic structure, the banking system, the market behavior. To produce reliable and quality information, education should be structured accordingly (Erdem, 2018). Therefore, only valuation education is considered in this study.

1.1 The Method of the Study

Related data of the study were obtained from; web site of the National Thesis Center of Higher Education Council (URL2), website of Turkey Valuators Association website (URL8, URL9), website of the Capital Markets Board (URL6), Magazine Park from academic journals that are followed by Dergi Park (Journal Park), "real estate valuation , valuation training, problems in valuation "keywords results in Google Scholar search engine. 70 theses for this subject have been made in our country. These were grouped in themselves as "doctorate" and "master" after they were divided into "science" and "philosophy".

2. VALUATION EDUCATIONING IN TURKEY

Real estate valuation in the world in scientific sense was born in Europe in the early 1900 where large rural migration took place. Valuation was suspended due to the World Wars I and II. Mathematical modeling was started to be done especially in Germany in 1946. After 1960, real estate management and valuation has been performed by the people educated on this field in developed countries especially in Germany and the UK.

In our country, real estate management, valuation and the economy has begun to gain importance due to urbanization, industrialization and economic development since 1990 (Ertaş, 2000, Poon, 2011). Now the necessity of making real estate valuation by using scientific method was felt instead of trading, managing and valuating the real estate and their additions in the usual sense in our country. The private sector and some civil organizations connected to the private sector noticed the case has started some education and re-establishment activities since 1998. As it is known, since the real estate valuation and management wasn't carried out with scientific methods in our country until 2000s, especially EU countries demanded a very important task from the ministry that manages economy and the ministry gave the duty to CMB (Capital Market Board). The task was scientific valuation and management of real estate. Because;

a) The firms of EU countries wishing to investment,
b) Multi-national corporations,
c) Insurance companies,
d) Social Security Organizations,
e) EU citizens who want to live in Turkey after they retired

were cheated previously, the EU countries wanted real estate and their additions valuation to be made according to the *valuation* based on scientific and mathematical valuation instead of *value estimation* as real estate agents are doing today (which is open to all kinds of speculation) (Ertaş, 2011). On the other hand, since our country's public and private sector have suffered from this issue, they began to work immediately to make arrangements for the legal and educational regulations.

But because of recent studies show multi-headiness, they will be explained in groups.

2.1 Postgraduate Education

The first post-graduate educationing on this issue was started with the course called *Land Valuation* by Ahmet AÇLAR at Yıldız Technical University in 1978. Similar courses were started at Science Institutes (SI) of some of the universities which have Geomatic Engineering Department Since 1991 and some Civil Engineering and Agricultural Engineering departments followed them (Ertaş, 2000). Many universities like Yıldız Technical University (YTU), Istanbul Technical University (ITU), Karadeniz Technical University (KTU), Selcuk University (SU), Istanbul University(IU) and Ankara University (AU) can be given example.

Table 1: The number of the thesis completed and continuing carried in universities of Turkey on immovable (real estate) valuation (end of 2017)

Universities	Postgraduate				Doctorate			
	Science Institutes		Social Science Institutes		Science Institutes		Social Science Institutes	
	Completed	Continuing	Completed	Continuing	Completed	Continuing	Completed	Continuing
Afyon Kocatepe	2							
Aksaray	1	4						
Anadolu							2	
Ankara	1	1					1	
Bahçeşehir	4							
Balıkesir	1							
Dokuz Eylül	1						1	
Ege	3							
Erciyes			1		1			
Gazi			8	1				
Gebze YTE	2							
Gümüşhane		1						
Hacettepe			1					
İstanbul Aydın				1				
İstanbul Kültür	1							
İstanbul Teknik	13	1			1			
Karadeniz Teknik	1				2	1		
Ondokuz Mayıs	1		1					
Selçuk	4				2		1	
Süleyman Demirel				1				
Marmara	1						1	
Mersin			1					
Orta Doğu Teknik	1							
Yıldız Teknik	5				1			
Zonguldak BE	1							
Çukurova	1							
Total	**38**	**4**	**18**	**3**	**3**	**1**	**5**	**0**

In addition, Ankara University, Institute of Science established Real Estate Development Department independently. This institute has 35 instructers and 75 courses (URL1). The related courses on valuation have been given in Social Sciences Institutes since 2000.

25 universities of our country have given post-graduate educationing on the relevant subject. 7 of the universities have given doctoral education. The number of thesis completed and continuing in the inquiry system of HEC (Higher Education Council) under the keyword *immovable* (*real estate*) *valuation* is given in (Table 1) and its distribution to the departments are given in (Table 2) (URL2).

According to Table 1 the sum of the completed thesis in science institutes is 39, the most dissertation managing university is ITU with 13 dissertations. Currently, there are 3 theses continuing in our 4 universities. The total number of doctoral studies completed in this institute is 7 and the universities are Istanbul Technical, Karadeniz Technical and Selcuk University, Erciyes University and 1 thesis study is still continuing.

Total postgraduate work done in the institutes of social sciences is 19 and Gazi University has 8 theses and Marmara University has 6 theses. 2 postgraduate works are also underway. Again, the number of completed doctoral study in this institute is 5, and there is no continuing doctoral study.

The first postgraduate study completed in 1992 and the first doctoral study completed in 2000 in science institute. Similarly, the first postgraduate study completed in 2001 and the first doctoral study completed in 2009 in the institute of social sciences.

Table 2: The number of the theses completed on immovable (real estate) valuation according to their main science departments in Turkey (end of 2017)

Main Science Department	Postgraduate				Doctorate			
	Science Institutes		Social Science Institutes		Science Institutes		Social Science Institutes	
	Completed	Continuing	Completed	Continuing	Completed	Continuing	Completed	Continuing
Geomatic Eng.	26	4			6	1		
Civil Engineering	3	1						
Urban Planning	8	1				1		
Architecth	2					1		
Agriculture	2							
Environment	2							
Others	5				1			
Administration			9	2			2	1
Economy			4				2	
Banking			2				1	
Public Management			1					
Law			1					
Finance			1	1			1	
Total	**48**	**6**	**18**	**3**	**7**	**3**	**6**	**1**

When Table 2 is studied we can clearly see that Geomatic Engineering is coming first with 26 completed dissertation and 4 continuing postgraduate dissertations in science institutes. There is no doctoral dissertation completed in other science departments except Civil and Urban Planning Departments, there are 2 study going on in that department.

Administration department is leading with 9 completed and 2 continuing postgraduate theses among social sciences. There is no continuing doctoral study in the social sciences.

Geomatic Engineering is coming first about completed and continuing postgraduate studies. I think there are two main reasons for this;

a) The interest of the profession itself is already real estates,

b) The first study of this subject in Turkey was made as a doctoral study in 1975 by Ahmet AÇLAR who is a lecturer in Yıldız Technical University, Geomatic Engineering. So the first study in this area was done at Geomatic Engineering.

2.2 Undergraduate Education

Undergraduate education in the universities in our country is given by have Kadir Has (URL3) and Istanbul Arel Universities (URL4). In these universities, Real Estate and Property Valuation department take part in Applied Departments Faculty (College). Courses taught here are; 26% *common courses*, 47% *geomatic + construction + architecture*, 27% *accountancy + economy + law*. Teaching staff have engineering and economy backgrounds. Total quota of these two departments is (40 + 34) 74 including students won scholarship.

Vocational courses in the department of Kadir Has University were carried out by 4 instructors (1 instructor, 1 lecturer, and 2 guest instructors). All of the instructors have economics origin. Vocational courses in the department of Istanbul Arel University were carried out by 5 instructors (2 instructors, 1 lecturer, and 2 guest instructors). 3 of the instructors have Geomatic and 2 instructors have economics origin. Each of these departments has not got the expected demand by students.

2.3 Associate Education

To respond the need of intermediate staff in this area Real Estate and Property Management Programs were opened in Istanbul University in 2000 (URL5). However, when the program definition and working fields of the graduates were studied, it is obvious that the issue is perceived as *real estate agent*. The task definition is given as "a person who gives consulting service on real estate sale and hire like houses, lands and shops, and establishing the house construction cooperatives based on the transitory and indefinitely contract prepared in a workplace". Their work areas are explained as;

a) Mediator of all the work of property owned by sale, mortgage and inheritance in the registry offices,

b) Provides consulting services about property buying and selling,

c) Helps in making the contract of sale at notary,

d) Provides consulting services, gives choices and presents comparative examples on making the construction contracts in exchange for flat,

e) Fills the declaration on behalf of clients,

f) Makes recommendations on construction projects to facilitative marketing,

g) Conducts market research and helps establishing cooperatives,

h) Finds lands and members for cooperatives,

i) Informs the owner or his representative about the economic and social status of the renter candidates,

j) Prepares rent agreement.

However, these tasks have already fulfilled by Estate Brokers in the free market. Real Estate Brokers are opening their offices according to the Article 8; paragraph B of "Regulation on the Business Counted Industry, Trade, Agriculture and Forestry Affairs" (Resmi Gazete, 2008). So there is no condition that Real Estate Brokers must graduate from Real Estate and Property Management Department. Graduates of this program have to be registered to the Chamber of Real Estate Agents and they have to succeed the courses organized by the chamber.

In addition, Compulsory Standard Communiqué (Communiqué No: OSG-2003/59) has been published based on the Law No. 3143 published in 08 July 2003 Official Journal No. 25162. In the Communiqué, all matters relating to the profession of Real Estate Brokerage are described in detail. There is no condition of graduating *Real Estate Management* department. Then the Institute of Turkish Standards stated the standards on the issue in the enacted TS 11816 standard dated 13.04.2010 (TS 11816, 2002). Again there is no condition of graduating *Real Estate Management* department as a standard. Ministry of Industry constituted a speciality committee including members of the Chamber of Real Estate Agents from various cities and Survey and Cadastre Engineers Chamber members while they were making studies on TS 11816 standard. This speciality committee took the view of Survey and Cadastre Engineers Chamber.

There are 32 Real Estate and Real Estate Management program in 18 state universities (26 vocational college) and 6 private universities. Among these universities Istanbul University Social Sciences Vocational College is clearly leading the others in terms of students' interest. The quota of the department is 50 (URL5), vocational courses are conducted by a total of 3 instructor who are economic origin. In addition, the program was opened at the Vocational Colleges of private universities like İstanbul Beykent, Doğuş, Arel, Aydın, Kadir Has. This program also took place in Anadolu University Eskişehir Vocational College and Open University Vocational College. Many Courses in other universities in Real Estate and Real Estate Management Program has been modified on the basis of this university. Courses taught here are; 35% *common courses*, 25% *geomatic + construction*, 40% *accountancy+ economy*. The courses are carried out by the teaching staff that have architecture, economy and law backgrounds.

2.4 Licensing Examinations

After our country accepted the International Valuation Standards in 2000, *Valuation Certificate* of which validity was accepted by Europe was needed. However, at that time, there was no organization giving

education on this issue, CMP took licensing duty because valuation is related in terms of international capital for Europe. Foreign companies, banks, mediator organizations which want to operate in our country need values which valuated by independent organizations and its responsibility is taken. The task was given to Capital Markets Board of Turkey (CMB) on that day.

CMB, started to make licensing exams in 2002 by means of ASPC (Assessment Selection and Placement Center) in the leadership of Proffessor Dr. Ahmet AÇLAR one of the lecturers of Yıldız University Geomatic Engineering Department (URL6).

Exams on Valuation

a) Real Estate Valuation Expertise,
b) Residential Valuation Expertise,

and they are conducted two times a year. The only prerequisite for entry into these exams is being a graduate of any 4-year undergraduate department. Whereas Real Estate Valuation Expertise exams were held out of 5 modules in early years, but today they are carried out of 4 modules (basic lessons);

a) Narrow extent capital market legislation and professional rules,
b) Real Estate Valuation principles,
c) Construction and Real Estate accounting,
d) Real Estate Legislation

Examination is made on the topics of 4 modules. Residential Valuation Expertise exam is made from 2 modules;

a) Narrow extent capital market legislation and professional rules,
b) Real Estate Valuation principles,

The number of licensed real estate valuation experts is 7 759 and the number of licensed home valuation experts is 4891 according to the Capital Markets Licensing Registration and Training Organization (CML) data in 31.12.2017, The number of licensed members from 2010 is shown in Table 3 (URL8).

Table 3: Range of valuation experts by years

Yıllar	TLDU	TDUBÜ	TLDU Rate%	TDUBÜ Rate	Membership %
2010	1482	1012	0,00	0,00	68,3
2011	1768	1259	19,3	24,4	71,2
2012	2080	1476	40,4	45,8	71,0
2013	2424	1654	63,6	63,4	68,2
2014	3288	2128	121,9	110,3	64,7
2015	4929	3318	232,6	227,9	67,3
2016	5941	3953	300,9	290,6	66,5
2017	**7759**	**4891**	**423,5**	**383,3**	**63,0**

(TLDU: Total licensed valuation experts,
TDUBU: Member of, the Association of Turkish Valuation Experts)

As can be seen from Table, by 2013, the annual increase in valuation experts is about 20%, but it has increased to 30% since 2014. The rate of organization of valuation experts in the TDUB is declining compared to years. 26% of these valuation experts reside in Istanbul, 14% in Ankara and 9% in İzmir.

Today, the number of valuation organizations in Turkey is 132 (URL8). 70 of these institutions are in Istanbul, 33 are in Ankara and 10 are in Izmir. The remaining 19 companies are scattered to be a few companies with a total of 11.

When we look at these data, we can see that the valuation experts, are clustered in the big cities especially Istanbul, and in other cities there are no valuation experts or they do not actively work.

Table 4: Number of students taking the exam and success rate by year

Exam Date	Candidate number	Candidates Getting certificate	Success rate (%)
2002	108	6	6
2003	309	61	20
2004	445	56	13
2005	928	67	7
2006	3822	190	5
2007	7023	305	4
2008	7371	418	6
2009	5968	327	5
2010	1470	139	9
2011	4488	466	10
2012	6273	701	11
2013	10812	865	8
2014	8730	873	10
2015	17217	1033	6
2016	5978	343	6
Total	**80912**	**5941**	**7**

When table 4 is examined we can see that the average success is 7%. In fact, this ratio shows us that the exam is not easy. The distribution of 6% succeeds according to the profession are;

a) Engineers (*Geomatic + Construction + architecture + urban planners*) 67%
b) The Economists (*Economist + Administration+ Banking + Finance + Lawyer*) 33%.

3. CONCLUSION: HOW SHOULD VALUATION EDUCATION BE?

Due to the globalization of the economy, real estate trade has become a global sector. On the contrary, the education of valuation training is varied (except in European countries). For that reason international and interdisciplinary approaches are needed. Institutional investors are not only in operative in their countries but in countries those protect themselves economically. Investors firstly examine the real estate legislation of other countries and make comparison to their countries (Shulte, 2005).

Turkey needs competent valuators, who can valuate according to standards accepted and valid internationally, for corporatizations, property sales to foreigners and to attract global capital. Besides, there is a need of valuators to satisfy the needs of

a) Community
b) Courts
c) Property Owners

d) Municipalities
e) Expropriators
f) Field and land arrangers

in domestic market. Therefore this topic should be handled in two phase as education in global world and education in Turkey. Nonetheless these educations should not contradict with each other while satisfying the need of its field.

After valuation education given in aforementioned institutions, an environment where famous real estate agents, researchers and valuators can share their experiences with students like in U.S.A. and England (Kohnstam, 1994). In some areas especially like land management and hypothec, an extensive education may not be needed. Procedures in these topics should be taught thoroughly. However, economists and investors must be included in the process of education of commercial and agricultural immovable valuation by giving importance to the interdisciplinary study.

In educational studies which are about to be given, CBS (GIS) which allows spatial interrogation and aspects which will form the basis of real estate tax studies, should be taken into the consideration (Yomralioglu, 2006). In these educations authentic determining the market value which is sub-level of financial crimes, illicit money and multi-purposed cadaster system, should be taken as a main principal and should include information about generating value map (Thief, 2011).

Furthermore, Cadaster Declaration Principals in EU promulgated by EU member countries which are gathered in Spain in 2002. Deed and Cadaster organizations are envisaged to be rearranged to meet the needs of valuation and land management requirements of 5th,6th and 7th articles of this principals. Valuators should be educated according to this in order to satisfy the needs of valuations in global aspect (EU, 2002).

3.1 Global Valuation Education

For the first time, Georgia University started its global valuation education as a graduate level is first satarted in Georgia University in 1998 and offered job guarantee to its graduates. Later on, these trainings spread to Australia and then to Europe. The first comprehensive study on this subject was made by the Swedish Royal Institute of Technology in 2000 by accepting students from Europe, Africa, Asia, Former Russia (Shulte, 2005). The Institute invited academicians as trainers from other countries. Education issues were the basic subjects like; finance, real estate marketing and consumer rights. Naturally, common academic initiatives were also targeted. In some countries, educational institutions accredited to RICS and TEGOVA were born. The following questions should be asked here;

a) Content depth of courses,
b) information, skills and competence for professionals need,
c) The skills and competence of the graduates,
d) Experts' depth of technological knowledge,
e) Which courses should be given,
f) Consumer demands,
g) Ethical principles

There are some organizations working globally in valuation field and accrediting. RICS and TEGOVA (The European Group of Valuers Association) can be examples of them. These kind of organizations accredit coworkers operating globally in immovable and construction fields. To do this they protect their customers and consumers with firm ethical rules by educating and setting vocational standards.

A research shows that there is a momentous difference between the expectations of valuation organizations and courses accredited by RICS (Royal Institution of Chartered Surveyors) (Shulte, 2005). Both employees and graduates after starting to work, say that there is lack of application and knowledge in education process. Employees also express that immovable sector is a part of global economy thus, also commercial consciousness should be thought to educationees. Graduates actually are not aware of they are equipped with these kind of knowledge and they do not know their deficiency is lack of application either.

Today corporative investors not only invest on their countries but also the countries which provide the opportunities of cheap energy, cheap and qualified labor and available market. Hence banks and credit institutions are densely in the land market.

A global valuation is not an individual study, but it can be done corporately and by including all shareholders (Poon, 2011). Organization should have academic support in the process of valuation. All individuals who take part in valuation process should

a) Be post-graduate
b) Be fluent in a foreign language
c) Think analytically
d) Know the International and related country's tax regulations
e) Analyze the risks with conjunctive view
f) Be highly experienced
g) Be able to use technology
h) Have the ability of ascendant communication
i) Have social and vocational ethics
j) Be sensitive against environment and to the community.

In addition, It should contain the training information required by the valuation items 5, 6, 7, which should be carried by the cadastral systems of the Member States of the European Union and summarized in Item 12 (EU, 2002).

3.2 Valuation Education in Turkey

Every stage of the valuation education in Turkey is formed by the initiative of the people and cyclical effect. Thus continuity, levelling and power sharing cannot be seen much in education. So, undergraduate educationing is not the continuation of associate educationing; postgraduate educationing is not the continuation of undergraduate educationing, because each stage has emerged in a different conjectural effect. This continues sometimes in a positive and sometimes in a negative manner.

Each step of the profession educationing in our country completes each other under normal conditions and shares authorization and responsibility in parallel. For example, authorization, task and responsibility in an institution or a project of each of the *Geomatic Technician* and *Geomatic Engineer* were defined separately according to the regulations. The valuation authorization was only given to *valuation experts* who

were succeeded in *licensing examination,* and graduates of

a) Associate,
b) Undergraduate,
c) Postgraduate

educationing don't have any authorization on preparing valuation reports. Therefore, each stage of the education should be revised and a curriculum and continuation should be formed for each stage.

As can be understood from the Licensing Examinations section, the licensing examinations conducted by the CMB are far from satisfying the following expectations. The only condition for entering these exams is having an undergraduate degree. Many under graduates who are from the out of the field like chemistry, biology, veterinary, psychology are able to enter and succeed in this exam. Because the questions about the licensing exams are now repeating themselves. In addition, the subjects of this test are also insufficient to respond to everyday problems. Today, there are plenty of undergraduate and graduate programs in this field. Therefore, the licensing exams made by the CMB were no longer required.

Courses for valuation training should be accredited to RICS and TEGOVA and must be member of the Bologna Process (D'Arcy.2009). After valuation training was given in the institutions, an environment should be created in which the experienced real estate agents, researchers and valuators in the field can trensfer their experiences to the sudents as in America and England (Kohnstamm 1994). Comprehensive training may not be necessary, especially on issues like management and mortgages. Procedures should be taught in detail in these matters. Especially interdisciplinary work on commercial and agricultural immovables should be emphasized and academicians, economists and investors should be included in the process of valuation training.

GIS which allows spatial inquiry and issues that will base on real estate tax studies should be taken into consideration in educational studies to be given (Yomralıoğlu 2006). In these trainings, market value which is determination of subgrade of financial crime, black money and multi-purpose cadastre system accurately should be the basic principle and it should include the information to produce value map. (Thief 2011).

As real estate valuation profession that hosts various disciplines is a new content is known in our country. For our country, this profession is very important for our country (anti-money laundering, preventing speculation, capital shift trade, ...) that it cannot be fit in it is licensing examinations. So the new members of the profession must take enough education of;

a) Engineering,
b) Law,
c) Economics,

to respond the needs of the field.

Educationing will be given should respond the needs of;

a) Banks,
b) Public Institutions,
c) Municipalities,
d) Courts,
e) Citizens,
f) Shareholders of real estate investments.

The issue is multicultural, so the educationing can be accredited by the institutions like RICS and TEGOVA (Köktürk, 2011). The educationing can include the courses like (Ertaş, 2011);

a) Architectural design and aesthetics,
b) Construction knowledge,
c) Construction accounting,
d) Economics,
e) Financial mathematics,
f) Basic Law,
g) Real Estate Law,
h) Real Estate Financing,
i) Real Estate Development,
j) Real Estate Marketing,
k) Real Estate Management,
l) Real Estate Economy,
m) Real Estate valuation,
n) Urban and Rural Areas Development,
o) Geographic Information System,
p) Cadastral Information,
q) Map Drawing,
r) Value Map Generation,
s) Urban Science and Planning,
t) Error Information and Statistics
u) Business ethics,

Course time should be adjusted according to the importance of the issue.

The last word to be said is real estate valuation is a very important profession that cannot be allotted into one discipline.

REFERENCES

AYOP (2012). II. Arazi Yönetim Çalıştayı *"Türkiye'de Taşınmaz Değerlemesi"* 21, 22 Mayıs 2012, İTÜ, İstanbul

D'Arcy E., and et all, (2009). *Real Estate Education İn Europe: Some Perspectives on a Decade of Rapid Change* Journal of European real Estate Research, Vol. 2 Issue: 1, pp 69-78

Erdem N. (2018). *Türkiye'de Taşınmaz Değerleme Sisteminin Yeniden Yapılandırılmasına Yönelik Bilimsel Çalışma ve Öneriler* ÖHÜ Müh. Bilim Derg. / OHU J Eng Sci ISSN: 2564-6605, Niğde

Ertaş M. (2000). *Kentsel Alanlarda Taşınmaz Değerlemesi* doktora tezi, İTÜ Fen Bilimleri Enstitüsü, İstanbul,

Ertaş M. (2011). Frances Plimmer'e rapor, *The Occupation of Valuation in Turkey,* Konya,

Ertaş M. (2015). Real Estate Valuation Training İn Turkey, The World Cadatre Summit-Congress and Exhibition, İstanbul,

EU, (2002). Declaration of Cadaster, The Permanent Committee on Cadastre

Kohnstamm P. (1994) *Trends in European Investment, Performance and real Estate Education* journal of property Valuation and Investment, Vol. 13 Iss 2pp 51-58,

Köktürk E. (2011). Taşınmaz Değerlemesi, ISBN: 978-605-873-346-0-9, s. 1187-1192,

Poon J., et al. (2011) Real Estate Education: An Investigation of Multiple Stakeholders, Property Management, Vol. 30 ISS 5 pp. 416-434,

Resmi Gazete, (2008). Sanati, Ticaret, Tarım ve Orman İşlerinden Sayılan İşlere İlişkin Yönetmelik,

Shulte K. W., et al. (2005). *İnternalizasyon of Real Estate Education*, bildiri, FIG Working Week 2005, Cairo,

SPK, (2005). Sermaye Piyasasında Uluslararası Değerleme Standartları Hakkında Tebliğ, (seri: VIII, No:45),

Thief F. (2011). *Education, Property Valuation, and Taxation Capacity in Developing Countries: The Example of Cambodia,* FIG Working Week 2011, Marrakech, Morocco,

TS 11816, (2002). Türk Standartları Enstitüsü, Emlak Komisyonculuğu Hizmetleri,

Yomralıoğlu T. ve ark, (2006). *Dünya'da ve Türkiye'de Taşınmaz Değerlemesi*, Türkiye'de Sürdürülebilir Arazi Yönetimi Çalıştayı, 20-27 Mayıs 2011, Okan Üniversitesi, İstanbul,

URL1, http://fenbilimleri.ankara.edu.tr/tr/anabilim-dallari/abd/50171/tasinmaz-gelistirme,

URL2, https://tez.yok.gov.tr/UlusalTezMerkezi/tez Sorgu Sonuc Yeni.jsp

URL3, http://ubyo.khas.edu.tr/index.php?page=1005

URL4, https://www.arel.edu.tr/tr/akademikbolum/81/gayrimenkul-ve-varlik-degerleme

URL5, http://sosyalbilimlermyo.istanbul.edu.tr/?p=6373

URL6, http://www.spk.gov.tr/indexcont.aspx?action=showpage&showmenu=yes&menuid=1&pid=1&su

URL7, http://www.eurocadastre.org/pdf/draft1.pdf

URL8, https://www.spl.com.tr/spl/eep/Specific/OEP/Upload/SPL%202016%20OCAK-EYL%C3%9CL%20D%C3%96NEM%C4%B0%20%C3%96ZET%20VER%C4%B0LER.pdf

URL9, http://www.tdub.org.tr/degerleme-uzmanlari.html

International Journal of Engineering and Geosciences (IJEG),
Vol; 4, Issue; 1, pp. 028-035, February, 2019, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.422570

COMPARISON OF DIFFERENT INTERPOLATION TECHNIQUES IN DETERMINING OF AGRICULTURAL SOIL INDEX ON LAND CONSOLIDATION PROJECTS

Mevlut Uyan[1]

[1]Selcuk University, Vocational School of Technical Sciences, Konya, Turkey (muyan@selcuk.edu.tr);
ORCID 0000-0002-3415-6893

*Corresponding Author, Received: 10/05/2018, Accepted: 13/06/2018

ABSTRACT: Land consolidation (LC) is a tool to improve the processing efficiency of agricultural area and the promotion of rural development same time an indispensable application for the promotion of sustainable agriculture. In order to achieve the reallocation process after LC, determining the correct of soil index (SI) for each of the agricultural parcels is very important for the success of LC projects. Nowadays, interpolation methods are extensively applied in the mapping processes to estimate the SI at unsampled sites.
The objective of this study was to evaluate and compare the performance of three interpolation methods for the agricultural SI values maps with GIS technology for LC projects. The SI data were determined from 132 observation points. Three spatial interpolation methods Ordinary Kriging (OK), Inverse Distance Weighted (IDW), and Radial Basis Functions (RBFs) were utilized for modeling the agricultural SI values. The results indicated that all methods provided a high prediction accuracy of the mean concentration of SI.In this study, although the best performed interpolation method was the OK, the results showed that the performance differed slightly among three methods. Results show that all the methods present a good performance in the estimation with RMSE (root mean square error) and ME (mean error) close to 0%.

Keywords: *Land consolidation (LC), Soil index (SI), Ordinary Kriging (OK), Inverse Distance weighting (IDW), Radial Basis Functions (RBFs)*

1. INTRODUCTION

Agriculture is the most important and indispensable part of Turkey's economy. The agriculture sector in Turkey still maintains the distinction of being social and economic sector with the effect of nutrition and labor, contribution to GDP and raw materials provided to the industrial sector. The majority of the industrial plant in Turkey use as agricultural raw materials. This situation is of great importance in the development of the industry. Approximately one-third of the population earns one's keep with agricultural activities. Agricultural production has exceeded $ 60 billion for 2015 in Turkey.

Agricultural land fragmentation, in which a single farm uses several parcels of land, is a common phenomenon in many countries (Orea et al., 2015; Latruffe and Piet, 2014). Land fragmentation is the biggest problem for sustainable agriculture (Uyan, 2016). It causes an increase in traveling time between fields which induces both lower labor productivity and higher transport costs for inputs and outputs and reduces the efficiency of machines compared to that obtainable in large, rectangular fields (Orea et al., 2015). In Turkey, the main cause of land fragmentation over the past has been population pressure on agricultural land.

Land consolidation (LC) is one of the most important tools for rural development and avoiding the negative impact of land fragmentation on agricultural productivity (Muchová et al., 2016; Guo et. al., 2015). LC may be described as a planned readjustment and rearrangement of land parcels and their ownership. With LC, land quality and agricultural infrastructures such as irrigation systems and roads are improved, and land fragmentation is reduced (Wang et al., 2015).

In Turkey, LC projects began in 1961. Only 450,000 ha of fragmented agricultural land were consolidated from 1961 to 2002. 5 million ha of fragmented agricultural land were consolidated between 2002 and 2013. The aim is to consolidate 1 million ha of fragmented agricultural land for each year and complete the land consolidation in Turkey until 2023 (Esen et. al., 2017).

Land reallocation stage is the most important step of consolidation. The purpose of this step is to ensure equivalent that the new parcels will be given to the landowners after LC with their previous parcels (Uyan, 2016). Land degrees (valuation) maps are created for this process by valuation experts (Derlich, 2002). Errors in degrees maps cause uneven distributions among landowners. Although soil quality defined as soil index (SI) is not the only factor in valuation, it is the most important factor. The value of a parcel can be affected all factors that have a substantial impact on the use of the agricultural land.

SI based on solely the soil characteristics and is obtained by evaluating factors such as soil depth, structure of the surface layer, subsoil characteristics, drainage, salinity, alkalinity, pH, erosions and relief. Soil index is rated productivity capabilities, potential benefit opportunities according to the land of the soil characteristics. SI is marked as 100 points.

Soil index is calculated according to the following formula:

$$SI = A * B * C * X \quad (1)$$

where A is the topsoil profile group (the kind of main material, shape and accumulation formation, age of soil material, variation, erosion resistance), B is the topsoil texture (sand, silt and clay rates according to various size groups in topsoil), C is the land slope and X is the soil profile group (drainage, salinity, alkalinity, acidity, toxicant and erosions (Uyan, 2016).

If SI values cannot be determined reliably for a study area, LC projects is affected negatively. Different interpolation techniques can be used in order to estimate the unmeasured points in accordance with the data values measured in the SI values. Rare measured SI data contain considerable uncertainty. Therefore, mapping the spatial distribution of SI requires spatial interpolation methods. Geographic Information System (GIS) techniques play an important role in managing complex relationships such as the storage, processing, and analysis of a wide variety of spatial data (Kuşçu Şimşek et al., 2018). Surface interpolation methods in a GIS are very powerful tools for predicting surface values. Geostatistics provides very useful techniques for handling spatially distributed data (Mirzaei and Sakizadeh, 2016). Geostatistical interpolation techniques utilize the statistical properties of the measured data to produce the raster maps (Kamali et al., 2015).

The kriging methods are widely applied to geographical sciences as the best linear unbiased prediction technique (Zhu et al., 2016; Peng et al., 2014). Likewise, the inverse distance weighting (IDW) interpolation algorithm is also one of the most commonly used spatial interpolation methods in geographical sciences (Mei and Tian, 2016). The most frequent applications for establishing spatial representations rely on the principle of ordinary kriging (OK) or IDW (de Amorim et al., 2016). Radial Basis Functions (RBFs) is known as a powerful tool for scattered data interpolation problem (Su et al., 2015) and widely applied to approximate scattered data (Zhang and Li, 2016).

The purpose of this study was to evaluate and compare the performance of three interpolation methods (Ordinary Kriging (OK), Inverse Distance Weighted (IDW), and Radial Basis Functions (RBF)) for the agricultural SI values maps with GIS technology for LC project in the village of Ortaoba in the Karaman Province, Turkey.

2. MATERIAL AND METHODS

2.1 Study Area

Ortaoba Village covers an area of approximately 2700 ha in the Karaman Province, Turkey. It is located 42 km away in the northwest of Karaman (Fig. 1). The topographical structure is generally flat and close to flat. The cultivated products are mostly wheat, barley and chickpea. The SI data were determined from 132 observation points.

Documents of cadastral parcels, SI data and other information on the area were obtained from the Konya Provincial Directorate of Agriculture.

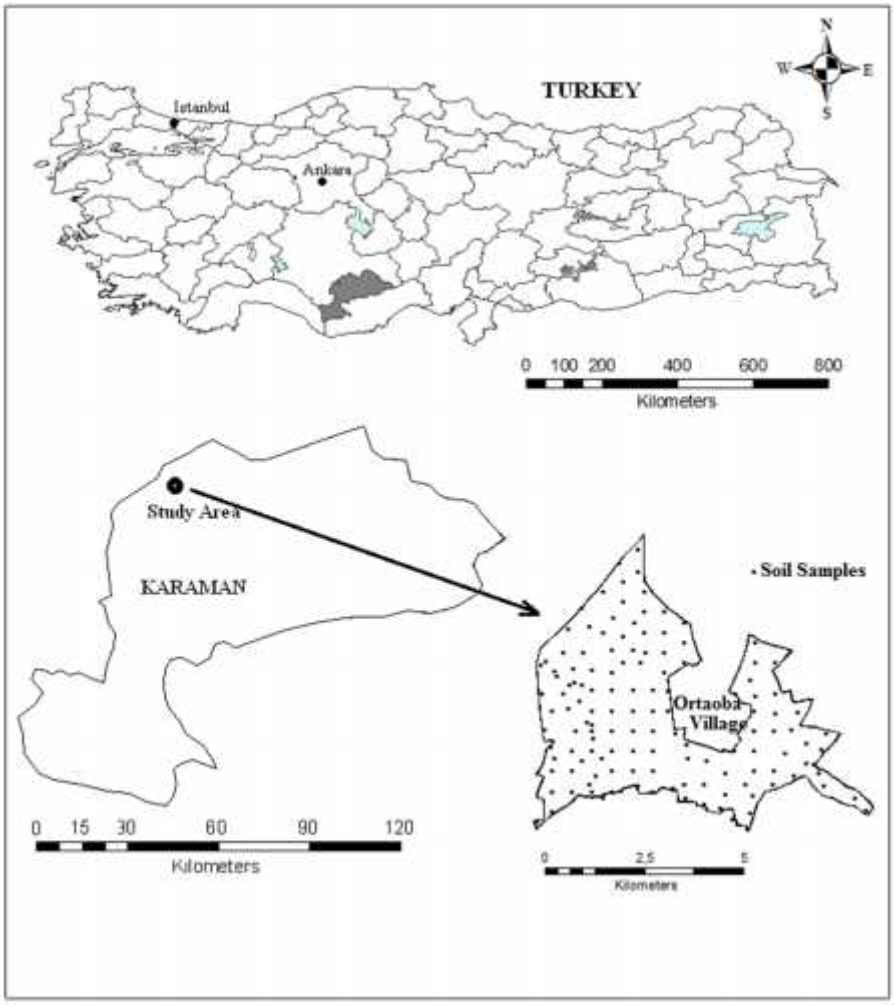


Figure.1 Study area and spatial distribution of sample points

2.2 Interpolation methods

2.2.1 Ordinary kriging (OK)

The OK method is widely applied to geographical sciences as the best linear unbiased prediction technique because it minimizes the variance of estimation error based on the statistical properties of the random field. The accuracy of OK is highly dependent on the effectiveness of the stochastic model of the random field (Zhong et al., 2016; Zhu et al., 2016). OK uses a linear combination of weights at known points to estimate the value at an unknown point (Uyan, 2016). OK assumes a constant, but unknown, local mean of a random variable, which is not plausible when spatial data exhibit a strong trend (Cafarelli et al., 2015). The OK method was detailed at Uyan (2016).

The general equation of the kriging method is as follows (Uyan, 2016):

$$Z^*(X_p) = \sum_{i=1}^{n} \lambda Z(X_i) \quad (2)$$

In order to achieve equitable estimations in ordinary kriging the following set of equations should be solved simultaneously.

$$\sum_{i=1}^{n} a_i \gamma(x_i, x_j) - \mu = \gamma(x_i, x_j) \quad (3)$$

$$\sum_{i=1}^{n} a_i = 1 \quad (4)$$

Here, $Z^*(x_p)$ is the kriged value at location x_p, $Z(x_i)$ is the known value at location x_i, λ_i is the weight associated with the data, μ is the Lagrange multiplier, and $\gamma(x_i, x_j)$ is the value of variogram corresponding to a vector with origin in x_i and extremity in x_j.

2.2.2 Inverse distance weighted (IDW)

A type of deterministic method widely applied in spatial modelling (de Amorim et al., 2016). The IDW calculates the interpolated values of unknown points by

weighting average of the values of known points. The name given to this type of methods was motivated by the weighted average applied since it resorts to the inverse of the distance to each known point when calculating the weights (Mei and Tian, 2016). The IDW interpolation method is mathematically expressed as (Osmanli et al., 2017):

$$Z_0 = \frac{\sum_{i=1}^{N} z_i . d_i^{-n}}{\sum_{i=1}^{N} d_i^{-n}} \quad (5)$$

where:

Z_0=The estimation value of variable z in point I.

Z_i= The sample value in point I.

d_i= The distance of sample point to estimated point.

N= The coefficient that determines weigh based on a distance.

n= The total number of predictions for each validation case.

2.2.3 Radial basis functions (RBFs)

RBFs method is one of the primary tools for interpolating multidimensional scattered data and used a basic equation dependent on the distance between the interpolated point and the sampling points for calculating smooth surfaces from a large number of data points (Xie et al., 2011). RBFs have become a popular numerical technology for scattered data approximation in a wide variety of fields for solving partial differential equations (Boyd, 2015). The prediction value by RBFs can be expressed as the sum of two components (Xie et al., 2011):

$$Z(x) = \sum_{i=1}^{m} a_i f_i(x) + \sum_{j=1}^{n} b_j \psi(d_j) \quad (6)$$

where $\psi(d_j)$ shows the radial basis functions and d_j the distance from sample site to prediction point x, $f_i(x)$ is a trend function, a member of a basis for the space of polynomials of degree <m. The coefficients a_i and b_j are calculated by means of the resolution of the following system of n+m linear equations; n is the total number of known points used in the interpolation as below:

$$Z(x_k) = \sum_{i=1}^{m} a_i f_i(x_k) + \sum_{j=1}^{n} b_j \psi(d_{jk}) \quad for\ k = 1,2,\dots n \quad (7)$$

$$\sum_{j=1}^{n} b_j f_k(x_j) = 0 \quad for\ k = 1,2,3,\dots m \quad (8)$$

RBFs methods have five different basis functions (thin plate splines, spline with tension, completely regularized spline, multiquadratic function and inverse multiquadratic function). We used thin plate splines (TPS) function in this study because it gave the lowest estimation error. The basic equation for TPS is below (Xie et al., 2011):

$$\psi(d) = c^2 d^2 \ln cd \quad (9)$$

where d is the distance from sample to prediction location, c is a smoothing factor.

2.3 Comparison of interpolation methods

Cross validation and validation with an independent data set are the commonly used methods for comparing the interpolation methods (Xie et al., 2011). Cross validation is an estimator widely used to estimate generalization error for its practicability and flexibility (Jiang and Wang, 2017). Cross-validation was obtained by taking the value of a sample point out and estimating it from the remaining values. The errors produced (observed-predicted) were used to assess the accuracy of each interpolation method. In order to assess the performance of different interpolation methods, the mean error (ME) and the root mean square error (RMSE) were calculated according to the formulas (Pereira et al., 2015):

$$ME = \frac{1}{N}\sum_{i=1}^{n}\{z(x_i) - \hat{z}(x_i)\} \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\{z(x_i) - \hat{z}(x_i)\}^2} \quad (11)$$

Where (x_i) is the observed value, $\hat{z}(x_i)$ predicted value and N the number of data points. Lower ME and RMSE values indicate fewer errors. If these values are small enough, interpolation becomes so accurate. Cross validation only confirms the prediction accuracy at sampling sites and cannot reflect the accuracy at unsampled sites (Cui et al., 2016).

In this study, all of the digitization, conversion and analysis processes of the maps were performed using ArcGIS 9.3 software.

3. RESULTS AND DISCUSSION

The descriptive statistics of SI parameters have been given in Table 1. According to this table, the calculated SI ranged from 10.53 to 100 with a mean value of 46.14.

The interpolation methods give the best results if the distribution of samples to be normal distributed. Otherwise, a suitable log transformation should be applied to the sample data. According to Table 1 results, histogram of SI values was plotted with a normal distribution curve.

Considering OK method, the nugget-sill rate can be used to classify spatial dependency. If the result is <25, 25–75, and >75 % suggest intense, medium, and weak spatial autocorrelations, respectively (Mirzaei and Sakizadeh, 2016; Uyan, 2016). The nugget-sill ratio of SI

in the area was 0.23 (23 %), which showed intense spatial autocorrelations according to Table 2.

Table 1 Descriptive statistics of the SI data

Statistics	SI data
Count	132
Minimum	10.53
Maximum	100
Mean	46.14
Std. Dev.	19.94
Skewness	0.48
Kurtosis	2.84
1st Quartile	29.92
Median	44.46
3rd Quartile	57.06

Geostatistical analysis uses the semivariogram to quantify the spatial variation of a regionalized variable and derives important parameters used for kriging spatial interpolation (Reza et al., 2016). Different semivariogram functions (Circular, Spherical, Tetraspherical, Pentaspherical, Exponential, Gaussian) were evaluated to select the best fit with the data. OK validity of semivariogram was assessed using cross-validation parameter Root Mean Square Standardized (RMSS). The estimations for a good-fitting semivariogram are an RMSS near 1. If RMSS < 1, there is a tendency toward overestimation of the variance; if > 1, there is a tendency toward underestimation (Uyan, 2016). Table 2 shows cross-validation indicators of SI obtained from OK method. According to the cross validation parameters, circular model was fitted to the empirical semivariograms. RMSS of circular model is close to 1, more than others. The model fitted to experimental variogram was circular as it results from cross-validation.

Table 2 Cross-validation performance and ranking of OK interpolation method

	ME	RMSE	RMSS	Nugget/sill	Regression Function
Circular	0.07	11.53	0.89	0.23	0.68x+14.59
Spherical	0.12	11.89	0.84	0.25	0.68x+14.33
Tetraspherical	0.13	11.89	0.83	0.26	0.68x+14.21
Pentaspherical	0.13	11.89	0.80	0.26	0.68x+14.14
Exponential	0.38	11.74	0.82	0.02	0.69x+13.48
Gaussian	0.09	12.33	0.89	0.49	0.68x+14.94

According to IDW method, ME and RMSE are 1.49, 12.38, respectively.

Five different basis functions (thin plate splines, spline with tension, completely regularized spline, multiquadratic function and inverse multiquadratic function) were evaluated for the RBFs. The accuracy performances of these functions are given in Table 3. Best results were obtained with thin plate splines function.

Table 3 Performance values of RBFs basis functions

	ME	RMSE	Regression Function
Completely regularized spline	0.49	11.62	0.69x+13.6
Spline with tension	0.45	11.64	0.69x+13.7
Multiquadratic function	0.47	11.83	0.75x+10.6
Inverse multiquadratic functio	0.63	11.70	0.68x+13.9
Thin plate splines	0.45	11.62	0.77x+8.99

All of the interpolation methods have similar prediction accuracy. Table 4 shows cross-validation indicators of SI obtained from OK, IDW, and RBF methods. The best performance was obtained by OK

(RMSE=11.53). RBF (RMSE=11.62) and IDW (RMSE=12.38) methods performs nearly as well as OK.. The results of three different interpolation methods had a quite similar SI distribution tendency according to Fig. 2.

Table 4 Cross-validation performance and ranking of different interpolation methods for SI

Methods	ME	RMSE	Regression Function
OK (circular)	0.07	11.53	0.68x+14.59
IDW	1.49	12.38	0.64x+14.91
RBF (thin plate splines)	0.45	11.62	0.77x+8.99

The ME shows that bias is very small for three methods. However, the OK method is much better than the others with a value of 0.07. The best fitted regression line between measured and estimated SI values and 1:1 line is showed for OK, IDW and RBF in Fig. 3. The deviation from the 1:1 line is greater for the IDW and OK methods. It shows that within interpolation methods used, the RBF method is the one that best estimated the measurements results of the SI.

If we take the results as a whole, the OK and RBF methods gives better results than the IDW method.

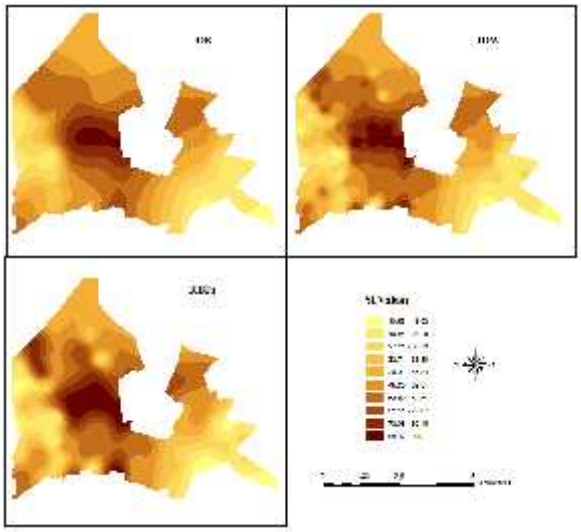


Figure.2 Interpolation effect of SI in study area through the three interpolation models

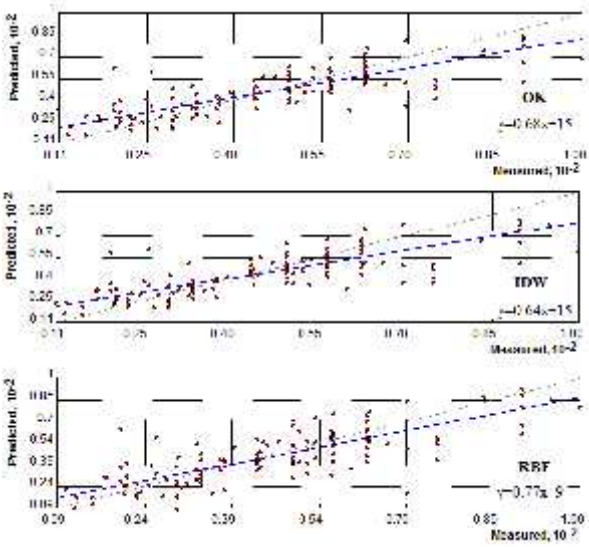


Figure.3 Least-squares regression line and the 1:1 line between the measured and the estimated SI values

4. CONCLUSION

Spatially continuous data is often required for environmental sciences and management. However, information for environmental variables is usually collected by point sampling. Thus, methods that generate such spatially continuous data by using point samples become essential tools. Spatial interpolation methods are, however, often data-specific or even variable-specific. Many factors affect the predictive performance of the methods. Hence it is difficult to select an appropriate method for a given dataset (Li and Heap, 2014). Even with the same type of interpolation method, the results varied with the parameters of the method. The target of the interpolations was to estimate the spatial means as accurately as possible (Xie et al., 2011).

Effectiveness of land consolidation depends on accurate and efficient mapping of SI. Usually, a larger number of soil samples will produce a more accurate SI map. However, due to the cost of sample collection and analysis, the use of large numbers of soil samples is usually impractical.

In this study, according to the RMSE for cross validation, OK and RBF are more accurate than IDW method. IDW have the biggest estimated error. According to Fig.2, SI estimated by OK and RBF gave quite similar results.

Kriging is an example of a group of geostatistical techniques used to interpolate the value of a random field. Kriging is based on statistical models involving autocorrelation. Autocorrelation refers to the statistical relationships between measured points. Not only do geostatistical methods have the capability of producing a prediction surface, but they can also provide some measures of the certainty and accuracy of the predictions (Ly et al., 2013). Kriging gives the best linear unbiased prediction of the intermediate values (Uyan, 2016). In IDW, each interpolated value is a weighted combination of every examination data point. In the classical formulation (IDW), the weights are inversely proportional to power of the distances to the data points (Smith et al., 2017). RBF methods are a series of exact interpolation techniques; that is, the surface must pass through each measured sample value (Liu et al., 2016).

SI values of land are one of the important criteria for LC studies. Therefore, it must be determined very precisely (Uyan, 2016). This study is conducted by making spatial analyses of 132 observation points using interpolation methods in order to determine the agricultural SI values in the LC project area. Tested three interpolation methods have a high prediction accuracy of the mean content for SI. The best performance was obtained by OK (RMSE=11.53). RBF (RMSE=11.62) and IDW (RMSE=12.38) methods performs nearly as well as OK. The results of three different interpolation methods had a quite similar SI distribution tendency according to Fig. 2. The best fitted regression line between measured and estimated SI values and 1:1 line was determined by RBF (y=0.77x+8.99). The modeling results show that the interpolated SI values similarly matched the observed SI values.

REFERENCES

Boyd, J.P. (2015). Convergence and error theorems for Hermite function pseudo-RBFs: Interpolation on a finite interval by Gaussian-localized polynomials, Applied Numerical Mathematics 87, 125-144.

Cafarelli, B., Castrignan, A., De Benedetto, D., Palumbo, A.D., Buttafuoco, G. (2015). A linear mixed effect (LME) model for soil water content estimation based on geophysical sensing: a comparison of an LME model and kriging with external drift, Environmental Earth Sciences 73(5), 1951-1960.

Cui, Y. Q., Yoneda, M., Shimada, Y., Matsui, Y. (2016). Cost-Effective Strategy for the Investigation and Remediation of Polluted Soil Using Geostatistics and a Genetic Algorithm Approach, Journal of Environmental Protection 7(01), 99-115.

de Amorim Borges, P., Franke, J., da Anunciação, Y. M. T., Weiss, H., Bernhofer, C. (2016). Comparison of spatial interpolation methods for the estimation of

precipitation distribution in Distrito Federal, Brazil, Theoretical and applied climatology 123(1-2), 335-348.

Derlich, F. (2002). Land Consolidation: A Key for Sustainable Development – French Experience, FIG XXII International Congress, Washington, USA.

Esen, Ö., Çay, T., Toklu, N. (2017). Evaluation of Land Reform Policies in Turkey, International Journal of Engineering and Geosciences 2(2), 61-67.

Guo, B., Jin, X., Yang, X., Guan, X., Lin, Y., Zhou, Y. (2015). Determining the effects of land consolidation on the multifunctionlity of the cropland production system in China using a SPA-fuzzy assessment model, European Journal of Agronomy 63, 12–26.

Jiang, G., Wang, W. (2017). Markov cross-validation for time series model evaluations, Information Sciences 375, 219-233.

Kamali, M. I., Nazari, R., Faridhosseini, A., Ansari, H., Eslamian, S. (2015). The Determination of Reference Evapotranspiration for Spatial Distribution Mapping Using Geostatistics, Water Resources Management 29(11), 3929-3940.

Kuşçu Şimşek, Ç., Türk, T., Ödül, H., Çelik, M.N. (2018). Detection of Paragliding Fields by GIS, International Journal of Engineering and Geosciences 3(3), 119-125.

Latruffe, L., Piet, L. (2014). Does land fragmentation affect farm performance? A case study from Brittany, France, Agricultural Systems 129, 68–80.

Li, J., Heap, A.D. (2014). Spatial interpolation methods applied in the environmental sciences: A review, Environmental Modelling & Software 53, 173-189.

Liu, Y., Hu, S., Sheng, D., Chang, L., Jia, M. (2016). Study of precipitation interpolation at Xiangjiang River Basin based on Geostatistical Analyst, In Geoinformatics, 2016 24th International Conference on (pp. 1-5). IEEE.

Ly, S., Charles, C., Degré, A. (2013). Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review, Biotechnologie, Agronomie, Société et Environnement 17(2), 392.

Mei, G., Tian, H. (2016). Impact of data layouts on the efficiency of GPU-accelerated IDW interpolation, SpringerPlus 5(1), 104.

Mirzaei, R., Sakizadeh, M. (2016). Comparison of interpolation methods for the estimation of groundwater contamination in Andimeshk-Shush Plain, Southwest of Iran, Environmental Science and Pollution Research 23(3), 2758-2769.

Muchová, Z., Leitmanová, M., Petrovič, F. (2016). Possibilities of optimal land use as a consequence of lessons learned from land consolidation projects (Slovakia), Ecological Engineering 90, 294-306.

Orea, L., Perez, J. A., Roibas, D. (2015). Evaluating the double effect of land fragmentation on technology choice and dairy farm productivity: A latent class model approach, Land Use Policy 45, 189-198.

Osmanli, N., Sert, E., Uyan, M. (2017). Social texture map and Social Information Center design and integration for social welfare in the city of Konya, Turkey, International Social Work 60(2), 521-533.

Peng, X., Wang, K., Li, Q. (2014). A new power mapping method based on ordinary kriging and determination of optimal detector location strategy, Annals of Nuclear Energy 68, 118-123.

Pereira, P., Oliva, M., Misiune, I. (2015). Spatial interpolation of precipitation indexes in Sierra Nevada (Spain): comparing the performance of some interpolation methods, Theoretical and Applied Climatology 126(3), 1-16.

Reza, S. K., Baruah, U., Sarkar, D., Singh, S. K. (2016). Spatial variability of soil properties using geostatistical method: a case study of lower Brahmaputra plains, India. Arabian Journal of Geosciences, 9(6), 1-8.

Smith, T.B., Smith, N., Weleber, R.G. (2017). Comparison of nonparametric methods for static visual field interpolation, Medical & biological engineering & computing 55(1), 117-126.

Su, L.D., Jiang, Z.W., Jiang, T.S. (2015). Numerical Method for One-Dimensional Convection-diffusion EquationUsing Radical Basis Functions. Journal of Physical Mathematics 6(2), 2015.

Uyan, M. (2016). Determination of agricultural soil index using geostatistical analysis and GIS on land consolidation projects: A case study in Konya/Turkey, Computers and Electronics in Agriculture 123, 402-409.

Wang, J., Yan, S., Guo, Y., Li, J., Sun, G. (2015). The effects of land consolidation on the ecological connectivity based on ecosystem service value: A case study of Da'an land consolidation project in Jilin province, Journal of Geographical Sciences 25(5), 603-616.

Xie, Y., Chen, T.B., Lei, M., Yang, J., Guo, Q.J., Song, B., Zhou, X.Y. (2011). Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: accuracy and uncertainty analysis. Chemosphere, 82(3), 468-476.

Zhang, Y.F., Li, C.J. (2016). A Gaussian RBFs method with regularization for the numerical solution of inverse heat conduction problems. Inverse Problems in Science and Engineering 24(9), 1606-1646.

Zhong, X., Kealy, A., Duckham, M. (2016). Stream Kriging: Incremental and recursive ordinary Kriging over spatiotemporal data streams, Computers & Geosciences 90, 134-143.

Zhu, Z., Chen, Z., Chen, X., He, P. (2016). Approach for evaluating inundation risks in urban drainage systems, Science of the Total Environment 553, 1-12.

International Journal of Engineering and Geosciences (IJEG),
Vol; 4, Issue;1, pp. 036-044, February, 2019, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.427531

ANN APPROACH FOR ESTIMATION OF COW WEIGHT DEPENDING ON PHOTOGRAMMETRIC BODY DIMENSIONS

Sakir Tasdemir[1] * and Ilker Ali Ozkan[1]

[1]Selçuk University, Faculty of Technology, Department of Computer Engineering, Konya, Turkey
(stasdemir/ilkerozkan@selcuk.edu.tr); **ORCID 0000-0002-2433-246X; ORCID 0000-0002-5715-1040**

*Corresponding Author, Received: 27/05/2018, Accepted: 08/07/2018

ABSTRACT: Computer technology and software are widely used in every multi-discipline field. Geomatics engineering can be seen as a pioneer of these disciplines especially in photogrammetry and image processing. Photogrammetry is a method where geometric parameters of objects on digitally captured images are determined and make measurements on them. Capturing the digital images and photogrammetric processing include several fully defined stages, which allows to generate three-dimension or two-dimension digital models of the body as an end product. The aim of this study is to predict Holstein cows' live weight via artificial neural network whose body dimensions were determined with photogrammetry method. The body dimensions to be used in this study are obtained metric from analysis of cows' images captured by synchronized three-dimension camera environment from different aspects. Wither height, hip height, body length, hip width of cows determined with photogrammetry. Artificial neural network prediction model was developed by using these body measurements. Dataset is divided into two after preprocessing as training and testing dataset. Different structured artificial neural network models are generated and the artificial neural network model which has the best performance is determined. Then with this artificial neural network model live weight of animals is estimated by using measurements obtained from images. After comparison of estimated live weights and weights obtained from scale, correlation coefficient is found (R=0.995). The statistical analysis shows that both groups are meaningful and artificial neural network can be used in live weight prediction safely.

Keywords: *Photogrammetry, Image Analysis, Artificial Neural Network, Live weight.*

1. INTRODUCTION

There have been new developments in image analysis systems and applications recently. One of the opportunities in these visual applications is that measurements can be made from the obtained images. Contactless analyses of objects are preferred more common than other methods because distortion and changes can happen on the object that will be measured in contact. Computer-aided digital image processing and analysis methods have a great number of application areas in addition to having some advantages such as being time saving, accurate and economic.

Area and length measurements with image analysis can be computed as pixels. These calculations can be made extremely accurate and clear according to the metric system. In order to conduct these analyses according to the metric system, there are a few preprocessing requirements. First process is to find the best calibration parameters explaining camera system. For this process, a particular number of control points in the test are (the reference points whose metric system equals are known) should be created by providing a laboratory environment with a camera system and photographs should be taken. Using these points, transformation process will be performed and the transition from 2-dimensions (2D) to 3-dimensions (3D) or vice versa can be performed. By analyzing these images, length measurements and geometric performances are determined in this way. Thus, metric measurements can be performed as contactless over images with photogrammetry technique (Goktepe, 2010; Tasdemir et. al., 2009; Tudes, 1996; Doğan et.al. 2018).

Image analysis with computer and various prediction applications are commonly used in different disciplines. one of which is husbandry. The number of dairies and their capacity is increasing so the importance of computer-aided studies is gradually increasing and usage of them are being widespread. In dairies, relations among live weight (LW), milk efficiency and forage consume can be accepted as signals for care and nourishment level in business and current animal conditions. It is crucial and necessary to watch these periods. When they are out of the boundaries, this affects the cows' immune system and economic efficiency significantly. Negative changes that can happen in especially live weight may indicate health problems, inappropriate environment conditions and nourishment errors. Therefore, live weight (LW) is certainly important for dairy cows (Tasdemir, 2010).

Artificial Neural Network (ANN) is a rather complex mathematical algorithm that can solve all the problems that are not subject to static mathematical laws. They simulate the way of human brain works in recognizing sounds, words and images. Neural networks can deal with complex issues in a smart way to reach an acceptable degree of expected solutions. The neural network creates interconnected neurons with different weights and each neuron is responsible for one input. For example, we have a problem consisting of ten inputs and one output. When we configure the grid, we actually have created a network of 10 neurons and each neuron responsible for one input. In training, the network continuously adjusts the weights associated with each neuron so that the output is closest to reality. The training process continues on all inputs available and the weights associated with each neuron are adjusted, provided that the resulting value is as close as possible to the true output value. Therefore, the training process and the abundance of information is very important to make the network forecast as close as possible to reality. One of the most important usage area of ANN is used is prediction. ANN can reveal the relations that are unknown among data and hard to notice. Many studies indicate that ANN is used commonly as traditional methods in prediction and it produces better results. For this reason, by using an artificial neural network a model was designed that enables input and output data pattern to be learned by the network.

When the literature reviews are examined, LW predicted studies using regression equations are mostly seen. Artificial intelligence techniques especially LW prediction using ANNs are available for different animal types and for different purposes. Some literature review for LW prediction using ANN; Prediction of bulls' slaughter value from growth data using artificial neural network (Adamczyk et. al., 2005), artificial intelligence technologies in dairy science (Akilli et. al., 2014), comparison of artificial neural network and multiple linear regression for prediction of live weight in hair goats (Akkol et. al., 2017), comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period (Ali et. al., 2015), weight prediction of broiler chickens using 3D computer vision (Mortensen et. al., 2016), prediction of body weight of goats (Raja et. al., 2012), artificial neural network to predict body weights of rabbits (Salawu et. al. 2014), prediction of carcass meat percentage in young pigs using linear regression models and artificial neural networks (Szyndler-Nędza et. al., 2016), walk-through weighing of pigs using machine vision and an artificial neural network (Wang et. al. 2008), extracting the three-dimensional shape of live pigs using stereo photogrammetry (Wu et. al., 2004), digital image analysis to estimate the live weights of pigs (Wongsriworaphon at. Al., 2015), estimation of mature live weight using some body measurements in Karya sheep (Yilmaz et. al., 2013).

The aim of this study is to predict LW of cows with artificial neural networks whose body dimension (BD) were determined with photogrammetric method. Thus, images of the cows captured from various angles and BDs were found in 3D via photogrammetric method. Wither height (WH), hip height (HH), body length (BL) and hip width (HW) of cows were detected with photogrammetry. Besides LW of cows were also measured with a digital scale and saved to computer. ANN prediction model was developed by using these parameters (input parameters are WH, HH, BL, HW and output parameter is LW). At the conclusion predicted LW with ANN and saved LW from scale are compared statistically.

2. PHOTOGRAMMETRY AND CAMERA CALIBRATION

Photogrammetry is the technology to obtain quantitative and qualitative information about a region's natural and industrial features by means of photographic or non-photographic images of the area. The pictorial area differs from the ground area in that the ground area deals with nature directly, and the imaging area is to obtain information and measurement of images without direct contact with nature in most stages of work. Photogrammetry uses ways from many department, included projective geometry and optics. Capturing of

digital image and photogrammetric processing include several fully defined stages, which permit to generate 3D or 2D digital models of the body as an end product. Figure 1 shows the type of information that can enter and come out of photogrammetric methods.

The 3-D coordinates (or 2-D) determine the locations of body points in the 3-D space. The image coordinates determine the locations of the body points' images on the film. The outer orientation of a camera determines its location in space and its vision direction. The inner orientation determines the geometrical parameters for the imaging process. This is the focal length of the lens but can also include the lens distortions. Photogrammetry algorithms, generally tries to reduce the sum of the squares of errors on the coordinates. This reduction called (bundle adjustment) is known as minimal modification of the beam.

Technical definition of photogrammetry is precise 3D measurement technique over images of objects and metric interpretation of image data. The first step in photogrammetry is to capture images in an intended way. In order to make stereological evaluation, gather 3D data and reach high accuracy, it is essential to capture synchronized images of object in at least two different stations (Figure 1). One of the other conditions to use an image as a measurement image in photogrammetry, is marked points whose 3D coordinates are known on the image are needed. Later, these coordinates are measured from 3D model generated by photogrammetry software. Measurements are done without contact. Photogrammetry that prevents deformation of object and provides safety for measurer, is more efficient and faster compared to classical methods. Images include both metric and semantic information in high density. Apart from the solid and stable objects, dynamic and deformable objects can also be registered and measured (Aguilar et. al., 2005; Atkinson, 1996; Cooper et. al., 1996; Tasdemir et. al., 2008; Tudes, 1996).

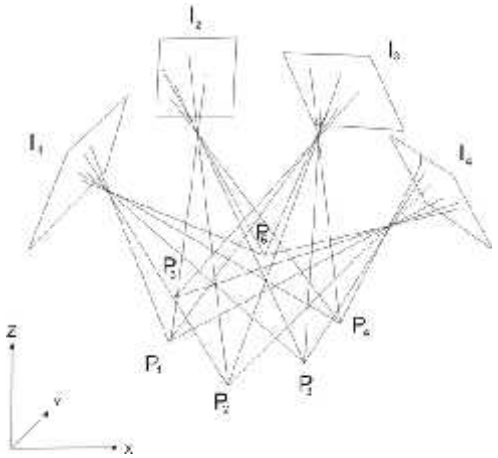


Figure 1. Coordinate of object points (Pi) by triangulating bundles of observation rays from different image planes (Ii)

Before image capturing process studio is prepared, cameras are set in stations and necessary regulations are done. In studio, space coordinates of reference object, whose 3D coordinates are known, are marked and measured. With the help of these measurements, camera calibration and rotation processes can be executed. Later on camera calibration process is executed on prepared experimental test setting (Akçay et. Al., 2017; Tudes 1996; Yilmaz et. al., 2008).

Camera calibration can be defined as finding parameters that represent camera system best. It is compulsory to know geometric performances of the cameras in order to obtain accurate measurements. Geometric accuracy can only be found via calibration of related camera and determining some parameters. These parameters are image capturing distance, directions and angles of image coordinate system axis and distortion parameters. Calibration of a camera can be defined as reverse process of photogrammetric point marking process. In photogrammetric inner orientation elements are known and coordinates of object spots are asked. In calibration on the other hand, coordinates of object spots are known and inner orientation elements are looked for (Atkinson 1996; Goktepe 2010; Karslı ve ark. 2003).

There are many camera calibration methods in literature such as linear, nonlinear and multi-step techniques. Linear method is faster than the others. The most favorite methods in this category is Direct Linear Transformation (DLT), Bundle Adjustment and Tsai's (Roger Tsai's perspective projection camera model) (Tasdemir ve ark. 2009).

To have a transformation between coordination systems, the systems should be described and their status according to each other should be determined. In photogrammetry, measured image coordinates should be transformed to object coordinate system. Image coordinates can be transformed directly into object coordinates via DLT method developed by Abdel-Aziz and Karara. This method defines transformation between image plane and space coordinate system. In the application, relationship between 2D image plane and 3D object space coordinate system, is modelled and 11 camera calibration parameters (transformation parameters), which defines cameras, are calculated. Basic DLT equation is shown in Eq. (1) 1 in matrix form.

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1X_1 & -x_1Y_1 & -x_1Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1X_1 & -y_1Y_1 & -y_1Z_1 \\ & & & \vdots & & & & & & \vdots & \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -x_nX_n & -x_nY_n & -x_nZ_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -y_nX_n & -y_nY_n & -y_nZ_n \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_{10} \\ L_{11} \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ \vdots \\ x_n \\ y_n \end{bmatrix} \quad (1)$$

In order to have mathematical conclusion to the equation, at least 6 point (from two images) whose image coordinates and object coordinates are known, is required. Before calibration object coordinate system of to be measured point and coordinate system of image should be defined. Points object coordinates are named as X, Y, Z, image coordinates are named as x, y and camera calibration parameters are names as L_1, L_2, L_3,…,L_{11} (Abdel-Aziz et. al. 1971; Fang-Jenq,. 1997; Goktepe, 2010; Tasdemir et. Al. 2009).

3. IMAGE CAPTURING AND DATASET

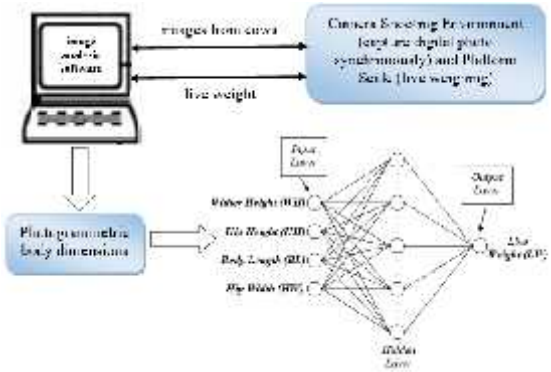


Figure 2. Block diagram of the image capture unit and structure of ANN for LW

A digital photograph shooting studio (platform) designed according to block diagram in Figure 1, is created where lighting system suitable for stereoscopic shooting technique is installed with the help of hardware components. The images of animals captured from both bottom and over with Canon EOS 400D cameras synchronously and at the same time LW of animals measured via digital scale and all data are registered in the computer.

As metric system will be used in measurements, before the computational image capturing process, reference points, whose coordinate numbers are known, were installed the way where animals would pass and be scaled (these points were given coordinates via geodesic methods-spatial coordinates). During installation of the mechanism, since size calculation will be made from analysis of computational images, calibration of appropriately located cameras are done. In this application relationship between 2D image plane and 3D earth coordination system is modelled and transformation process is executed. Suitable to this relation method, DLT method and camera calibration test area whose 3D coordinates were known, are used to calculate calibration parameters that identifies cameras. On captured images, these reference points whose coordinates were known, and marks were used to calculate animals' body dimensions (wither height (WH), hip height (HH), body length (BL), hip width (HW)) with photogrammetric method via Image Analysis Software (Tasdemir, 2010). LW of cows were also measured with a digital scale and saved to computer. Data set is used in the designed ANN.

Table 1 gives the descriptive statistical results that summarize the numerical values of the data set and convert them to descriptive indexes.

Table 1. Descriptive statistics of data set using ANN

Parameters*	Unit	Minimum	Maximum	Sum	Mean	Std. Deviation
Wither Height (*WH*)	cm	125	149	15673	136.3	4.18
Hip Height (*HH*)	cm	130	149	16131	140.3	4.72
Body Length (*BL*)	cm	148	181	18889	164.3	6.34
Hip Width (*HW*)	cm	45	65	6232	54.2	3.95
Live Weight (*LW*)	kg	409.5	771	65574	570.2	67.14

*N (number of data) for each parameter is 115

Dispersion of BDs obtained with photogrammetric technique against live weights measured by digital scale, is given in Figure 3. Data set given in table and figure is used to develop Artificial Neural Network model.

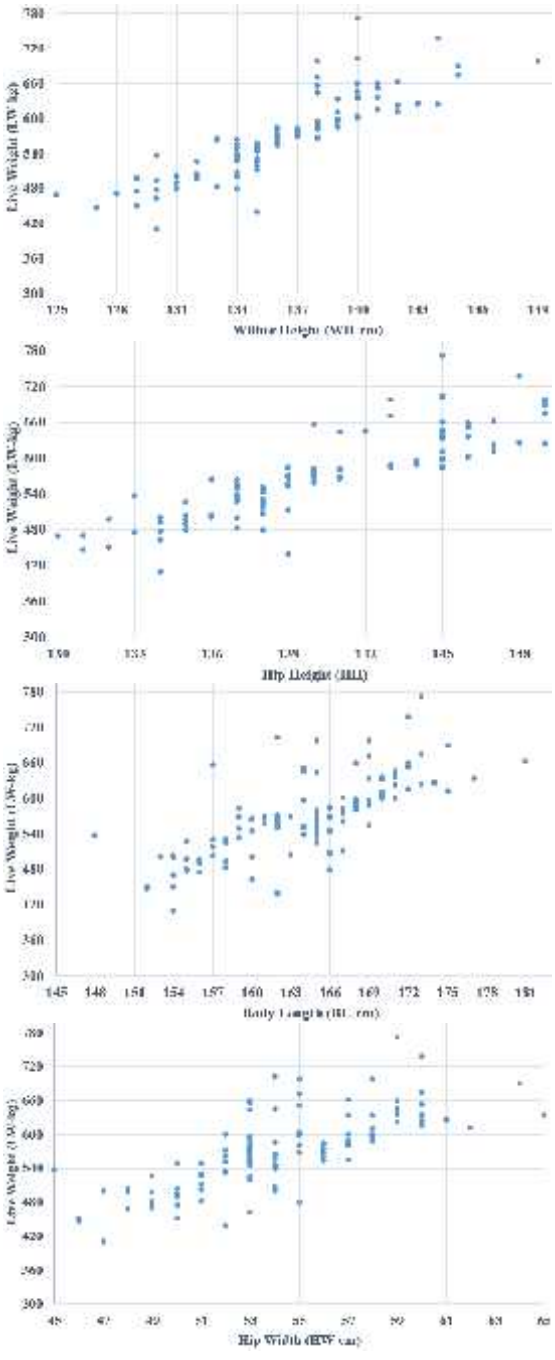


Figure 3. Data set graphics showing dispersion of BDs against live weights

4. ARTIFICIAL NEURAL NETWORK (ANN) PREDICTION MODEL AND RESULTS

Artificial Neural Network (ANN) is an artificial intelligence method fulfilling learning function which is the most basic feature of human brain. ANN has many important features such as learning over examples, producing solutions to non-linear problems, making generalizations, working with missing information and being able to work with infinite number of variance. ANN enables linear and non-linear modelling between input and output variances without requiring any prior knowledge. Thus, ANNs make generalizations from the examples given to themselves during training and they produce information about new samples with these generalizations (Gurney, 2014; Hagan et al., 1996; Maind and Wankar, 2014; Jain et al., 1996; Saritas, 2010)

ANN prediction model is a structure formed through artificial neurons connecting to each other. ANN consists of three main layers as Input Layer, one or more Hidden Layers and Output Layer. Although artificial neurons exist as much as input number of the model, generally inputs are transmitted to interlayers without undergoing any processes. Information going out from the input layer comes to the hidden layer. The number of neurons in hidden layers is independent from the number of input and output. Although the increase of the number of hidden layers and neurons in these layers increase calculation complexity and time, it enables artificial neural network to be used in solutions to more complex problems (Gurney, 2014). The number of hidden layers and neurons in hidden layers are determined with trial and error method. Data are gathered by multiplying their connections weight values in hidden layer and output layer and they are transferred to transfer function. As for output layer, it is the layer producing network outputs by processing the information coming from hidden layers. In ANN with feedback, output produced from this layer is used and new weight values are calculated. (Idris et al., 2017; Maind and Wankar, 2014) ANN prediction model used in this study is shown in Figure 2.

One of the most important factors in an artificial neural network is the connections providing data transfer from one artificial neurons to another. A connection transmitting information from one neuron to another has also a weight value. Add function calculates need input which comes to a neuron. Input is formed for add function by multiplying input variances and weight coefficients. (Altan et al., 2009). A neuron structure is given in Figure 4.

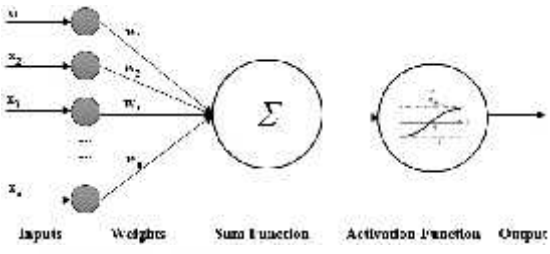


Figure 4. A basic artificial neural network cell

$x_1, x_2, x_3 \ldots . x_n$ show input values coming to neurons, $w_1, w_2, w_3 \ldots . w_n$ show weights of each inputs, b shows bias, y shows output value, in Eq. (2) and (3) where mathematical expression of an artificial neuron is given (Hasni et al., 2012)

$$net = \sum_{i=1}^{n} w_{ij} x_j + b \quad (2)$$

$$y = F(net) = F\left(\sum_{i=1}^{n} w_{ij} x_j + b\right) \quad (3)$$

The alteration in weight values provides ANN's learning process. F is activation function and is the function that processes the input and generates output.

In order to predict LW with ANN, four inputs which include BDs, were used. These inputs are wither height (WH), hip height (HH), body length (BL), hip width (HW) in order. Data set inputs and outputs normalized between 0-1 by using Eq. (4) in order to regulate dispersion of data set values.

$$xn = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (4)$$

x_n= Normalized x_i (real value) value
x_{min}= Min value of data to be normalized
x_{max}= Max value of data to be normalized

Data set consists of 115x5=575 measurements which include a cow's (BD). Dataset is divided into two parts as 75% train and 25% test dataset. As performance function of network Mean Squared Error (MSE) (Tumer et. al., 2015) in Eq. (5) was used.

$$MSE = \frac{1}{n}\sum_{k=1}^{n}(y_k - t_k)^2 \quad (5)$$

In this equation y_k is target or real value, t_k is the network output or estimated value and n is the output data number.

Determining neuron quantity in hidden layer of ANN structure, is important since it states network's generalization ability. In order to determine the neuron quantity in hidden layer MATLAB software is used and various ANN models with different neuron quantities are generated. Figure 5 is MSE variance graph of various ANN models which were generated as a result of changing neuron quantity in hidden layer from 2 to 50. [4-10-1] structured ANN model whose hidden layer has 10 neurons, has lowest MSE result and is chosen as the most appropriate model to be used in the study.

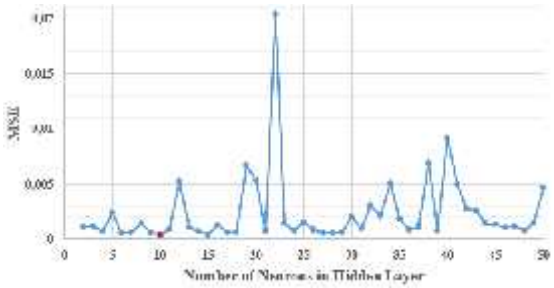


Figure 5. MSE variance for different ANN structures

In this study according to body dimensions (WH, HH, BL, and HW) of cows, live weight of cow is predicted. BDs used in the study are obtained via photogrammetric methods. In literature various methods were used to determine the live weight of cows. In this study an ANN prediction model, which can predict live weight of animal

according to its BDs, is generated since it can model linear relations along with non-linear relations in data set successfully and it has no need for any presupposition over data.

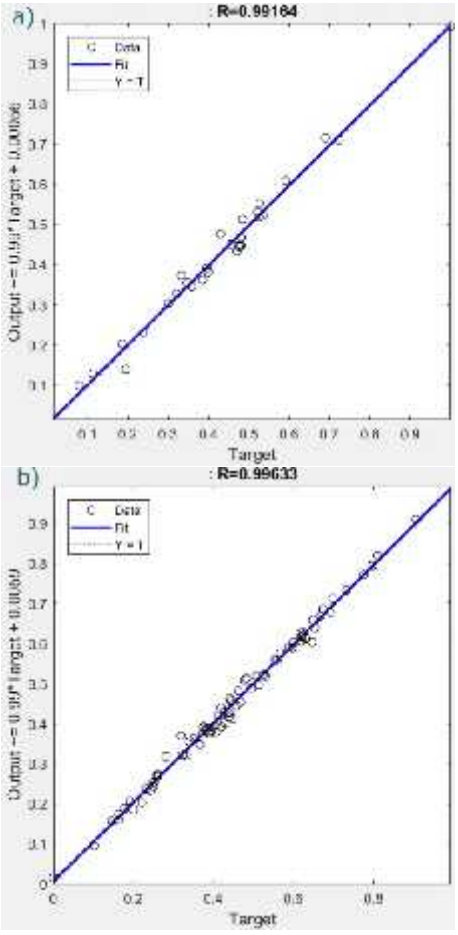
In Table 2'de parameters of chosen ANN structure are given.

Table 2. ANN Parameters

Hidden Layer Number	1
Activation Function, Hidden Layer	Tansig
Activation Function, Output Layer	Lineer
Learning Rate	0.01
Minimum performance gradient	1e-5
Performance goal	1e-3
Maximum number of epochs to train	5000

Correlation Coefficient (R) is used to evaluate the performances measured LW and predicted values. R value changes from 0 to 1, the stronger the relation between value calculated with model and value measured. In other words, The closer the R value to 1, the more successful the model's prediction ability (Cohen et al., 2013).

In the study [4-10-1] structured ANN model is used and 0.9916 (for testing), 0.9963 (for training) and 0.9949 (training and testing) correlation coefficient (R) values are obtained in test dataset, train dataset and all dataset in order. Linear regression graphics of network outputs according to weights are given in Figure 6.



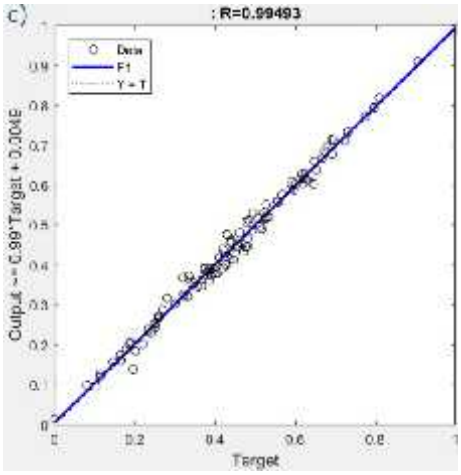


Figure 6. Linear regression graphics a) Testing dataset b) Training dataset c) Training and Testing dataset (all)

In Figure 7 for 115 data example comparison of LW scale values and ANN model prediction values are given graphically.

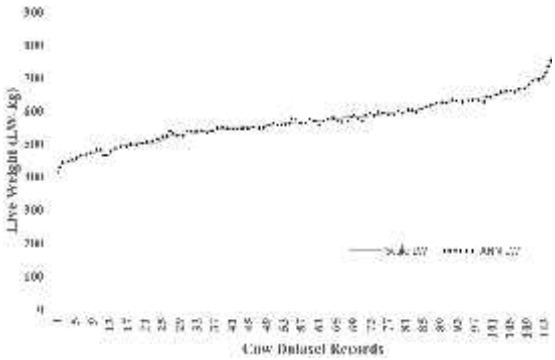


Figure 7. LW Scale and ANN prediction values variation

In this study Eq. (6) is used to predict LW according to BDs with ANN model.

$$\boldsymbol{ANN\ Prediction = denorm(w_2.tansig(w_1.[x(1);x(2);x(3);x(4)] + b_1) + b_2)} \quad (6)$$

In this equation x(1), x(2), x(3) ve x(4) variations show WH, HH, BL, HW values respectively, w_1 ve b_1 show weight and bias in hidden layer, w2 ve b2 show LW and bias in output layer. In Table 3 weight and bias values for every layers obtained from optimum ANN structure, are given. Besides, aforementioned tansig and denorm functions are given in Eq. (7) and (8).

$$tansig(n) = \frac{2}{(1+e^{-2n})} - 1 \quad (7)$$

$$denorm(n) = n361.5 + 409.5 \quad (8)$$

Table 3. ANN's weight and bias values for LW

w_1				$(w_2)^T$	b_1	b_2
			-	0.52	-	
[0.6	2.12	1.47	0.34	454	2.785	
1404	059	590	855	0.20	84	
3.15	0.36	0.96	1.90	763	-	
654	502	831	868	0.51	1.599	
		-		140	58	
0.65	1.03	1.83	1.32	0.82	-	
207	833	813	029	795	2.137	
-	-		-	0.19	91	
1.04	0.68	1.01	0.09	490	1.456	
395	947	252	592	1.09	80	
-		-	-	210	0.620	
0.24	0.86	1.54	1.66	-	41	
973	463	535	778	0.24	-	
-	-	-		687	0.842	
0.57	0.84	0.13	1.97	0.54	07	
314	760	942	171	190	0.403	
-	-	-		0.59	49	
1.51	3.00	1.21	2.22	424	1.696	
778	138	291	314	0.94	79	
	-	-	-	914	-	
2.19	0.94	1.18	1.77		2.650	
149	908	213	171		61	
-	-	-	-		1.891	
0.63	2.24	1.73	0.05		76	
997	131	209	444			
			-			
1.25	1.45	0.91	0.18			
102	673	481	310]			

0.24977

Moreover, according to result of t-test conducted in SPSS statistical analysis software in 95% reliability interval, there is no meaningful difference between ANN prediction values and LW values. As in table 4 the mean level is 0.994 which is high and reliable.

Table 4. t-Test for equality of means

t	df	Significant (two-tailed)	Mean difference	Standard error difference	95% Confidence interval of the difference	
					Lower	Upper
0.008	227.9	0.9	-0.07	8.83	-17.47	17.33

5. CONCLUSION

In this study, ANN study with four inputs and one output was performed to predict LW using body dimensions of cows. The use of ANN is a useful model in this study due to the nonlinear data. The model of live weight prediction is useful and necessary in this kind of animals, because scale is a difficult way to measure LW and animals feel stressed on scale. In this study LW values obtained from ANN and scale measurements were compared. It is seen that there is a consistency between two groups of data through comparison. ANN produced successful results, and can be used in such kind of system modelling and prediction (Table 4). It has been observed that there is a strong correlation between the estimated LW data by ANN and the measured data with scale.

Analysis and comparisons made based on the consistency between the values obtained this way.

Prediction accuracy increases when the correlation coefficient value is close to 1. In this study, obtained correlation coefficients are very close to 1 for all the dataset which indicates that there is a perfect match between ANN prediction values and live weight values. Correlation coefficient for ANN prediction value in the test data set was found as 0.9916 for testing and R has an acceptable value.

When input parameter numbers are increased in designed model, the success rate of the study is thought to increase. What is more, this developed method and modelling can be used to measure weights of different kinds of species. When another smart system, algorithm, mathematical and statistical method are used alone or with ANN instead of ANN used in the study, it can affect the rate of success. In addition, this model has the skill of predicting LW values which are not found in data set.

REFERENCES

Abdel-Aziz, Y.I. ve Karara, H.M. (1971). Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry, Proceedings of the Symposium on Close-Range Photogrammetry, Urbana, Illinois, 1-8.

Adamczyk, K., Molenda, K., Szarek, J., & Skrzyński, G. (2005). Prediction of Bulls'slaughter Value From Growth Data Using Artificial Neural Network. Journal of Central European Agriculture, 6(2), 133-142.

Aguilar, M. A., Aguilar, F. J., Agüera, F., & Carvajal, F. (2005). The evaluation of close-range photogrammetry for the modelling of mouldboard plough surfaces. Biosystems Engineering, 90(4), 397-407.

Akçay, Ö., Erenoğlu, R , Avşar, E . (2017). The Effect of Jpeg Compression in Close Range Photogrammetry. International Journal of Engineering and Geosciences, 2 (1), 35-40. DOI: 10.26833/ijeg.287308

Akilli, A., & Atil, H. (2014). Artificial Intelligence Technologies in Dairy Science: Fuzzy Logic and Artificial Neural Network. Hayvansal Üretim, 55(1), 39-45.

Akkol, S., Akilli, A., & Cemal, İ. (2017). Comparison of Artificial Neural Network and Multiple Linear Regression for Prediction of Live Weight in Hair Goats.YYU Yüzüncü Yıl Üniversitesi Journal of Agricultural Sciences, 27(1), 21-29.

Ali, M., Eyduran, E., Tariq, M. M., Tirink, C., Abbas, F., Bajwa, M. A., ... & Shah, S. H. (2015). Comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in Harnai sheep. Pakistan J. Zool., vol. 47(6), pp. 1579-1585.

Atkinson, K.B. (1996). Close Range Photogrammetry and Machine Vision, Whittles Publishing, Scotland, 1996.

Altan, Ö., Gölcü, M. (2009). Daily means ambient temperature prediction using artificial neural network method: A case study of Turkey. Renewable Energy 34, 1158-1161.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S. (2013). Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.

Cooper, M. A. R. and Robson, (1996). S.Theory of Close Range Photogrammetry, Close Range Photogrammetry and Machine Vision, 9-51.

Doğan, Y., Yakar, M. (2018). GIS and three-dimensional modeling for cultural heritages. International Journal of Engineering and Geosciences, 3 (2), 50-55.

Fang-Jenq, C. (1997). Application of Least-Squares Adjustment Technique to Geometric Camera Calibration and Photogrammetric Flow Visualization, ISA 43rd International Instrumentation Symposium, Orlando, Florida.

Goktepe, A. (2010). A study on the calibration of stereo photogrammetric systems used in motion analysis. Scientific Research and Essays, 5(17), 2491-2496.

Gurney, K. (2014). An introduction to neural networks. CRC press.

Hagan, M.T., Demuth, H.B., Beale, M.H. (1996). Neural network design. Pws Pub. Boston.

Hasni, A., Sehli, A., Draoui, B., Bassou, A., Amieur, B. (2012). Estimating Global Solar Radiation Using Artificial Neural Network and Climate Data in the South-western Region of Algeria. Energy Procedia 18, 531-537.

Idris, N.H., Salim, N.A., Othman, M.M., Yasin, Z.M. (2017). Prediction of Cascading Collapse Occurrence due to the Effect of Hidden Failure of a Protection System using Artificial Neural Network. Journal of Electrical Systems 13.

Jain, A.K., Jianchang, M., Mohiuddin, K.M. (1996). Artificial neural networks: a tutorial. Computer 29, 31-44.

Karsli, E., Ayhan, E., Tunc, E. (2003). Genelleştirilmiş DLT metodu ile Dijital Kamera Geometrik Kalibrasyonu, TMMOB Harita ve Kadastro Mühendisleri Odası 9. Türkiye Harita Bilimsel ve Teknik Kurultayı, Ankara.

Maind, S.B., Wankar, P. (2014). Research paper on basic of artificial neural network. International Journal on Recent and Innovation Trends in Computing and Communication 2, 96-100.

Mortensen, A. K., Lisouski, P., & Ahrendt, P. (2016). Weight prediction of broiler chickens using 3D computer vision. Computers and Electronics in Agriculture, 123, 319-326.

Raja, T. V., Ruhil, A. P., & Gandhi, R. S. (2012). Comparison of connectionist and multiple regression approaches for prediction of body weight of goats. Neural Computing and Applications, 21(1), 119-124.

Salawu, E. O., Abdulraheem, M., Shoyombo, A., Adepeju, A., Davies, S., Akinsola, O., & Nwagu, B. 2014. Using artificial neural network to predict body weights of rabbits. Open Journal of Animal Sciences, 4(04), 182.

Saritas, I., Ozkan, I. A., & Sert, I. U. (2010). Prognosis of prostate cancer by artificial neural networks. Expert Systems with Applications, 37(9), 6646-6650.

Szyndler-Nędza, M., Eckert, R., Blicharski, T., Tyra, M., & Prokowski, A. (2016). Prediction of carcass meat percentage in young pigs using linear regression models and artificial neural networks. Annals of Animal Science, 16(1), 275-286.

Tasdemir, S., Yakar, M., Urkmez, A., & Inal, S. (2008, June). Determination of body measurements of a cow by image analysis. In Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing (p. 70). ACM.

Tasdemir, S., Urkmez, A., Yakar, M. ve Inal, S, 13-15 May 2009. Determination of camera calibration parameters at digital image analysis. IATS'09.

Tasdemir, S. (2010). Determination of Body Measurements On the Holstein Cows by Digital Image Analysis Method and Estimation of Their Live Weight. Ph. D. thesis, Selcuk University, Konya, Turkey.

Tudes, T., Yer Fotogrametrisi, (1986). KTÜ Basımevi, Mühendislik Fakültesi yayını, No: 105, Trabzon, 1996.

Tumer, A. E., & Edebali, S. (2015). An Artificial Neural Network Model for Wastewater Treatment Plant of Konya. International Journal of Intelligent Systems and Applications in Engineering, 3(4), 131-135.

Wang, Y., Yang, W., Winter, P., & Walker, L. (2008). Walk-through weighing of pigs using machine vision and an artificial neural network. Biosystems Engineering, 100(1), 117-125.

Wu, J., Tillett, R., McFarlane, N., Ju, X., Siebert, J. P., & Schofield, P. (2004). Extracting the three-dimensional shape of live pigs using stereo photogrammetry. Computers and Electronics in Agriculture, 44(3), 203-222.

Wongsriworaphon, A., Arnonkijpanich, B., & Pathumnakul, S. (2015). An approach based on digital image analysis to estimate the live weights of pigs in farm environments. Computers and Electronics in Agriculture, 115, 26-33.

Yilmaz, H. M., Yakar, M., Yildiz, F. 2008. Digital Photogrammetry in Obtaining of 3D Model Data of Irregular Small Surfaces Objects. ISPRS Congress, Beijing, 125-130.

Yilmaz, O., Cemal, I., & Karaca, O. (2013). Estimation of mature live weight using some body measurements in Karya sheep. Tropica.

International Journal of Engineering and Geosciences (IJEG),
Vol;4, Issue;1, pp. 045-051, February, 2019, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.440828

AIRBORNE LIDAR DATA CLASSIFICATION IN COMPLEX URBAN AREA USING RANDOM FOREST: A CASE STUDY OF BERGAMA, TURKEY

Sibel Canaz Sevgen [1]*

[1]Ankara University, Faculty of Applied Sciences, Department of Real Estate Development and Management, Ankara, Turkey (ssevgen@ankara.edu.tr); **ORCID 0000-0001-5552-6067**

*Corresponding Author, Received: 05/07/2018, Accepted: 06/08/2018

ABSTRACT: Airborne Light Detection and Ranging (LiDAR) data have been increasingly used for classification of urban areas in the last decades. Classification of urban areas is especially crucial to separate the area into classes for urban planning, mapping, and change detection monitoring purposes. In this study, an airborne LiDAR data of a complex urban area from Bergama District, İzmir, Turkey were classified into four classes; buildings, trees, asphalt road, and ground. Random Forest (RF) supervised classification method is selected as classification algorithm and pixel-wise classification was performed. Ground truth of the area was generated by digitizing classes into features to select training data and to validate the results. The selected study area from Bergama district is complex in urban planning of buildings, road, and ground. The buildings are very close to each other, and trees are also very close to buildings and sometimes cover the rooftops of buildings. The most challenging part of this study is to generate ground truth in such a complex area. According to the obtained classification results, the overall accuracy of the results is found as 70,20%. The experimental results showed that the algorithm promises reliable results to classify airborne LiDAR data into classes in a complex urban area.

Keywords: *Random Forest, LiDAR, Classification, Complex Urban Area*

1. INTRODUCTION

Classification of objects in an urban area is a popular subject in a variety of research areas, such as computer vision, machine learning, pattern recognition, photogrammetry, remote sensing, and urban planning. In the literature, satellite and aerial images have been widely used in urban area classification. Especially, land cover changes studies over the years by classifying satellite data are abundantly presence in the literature. For instance, Yu et al. (2012) monitored land cover changes and urban sprawl dynamics 1989, 1999, and 2009 of Yantai China by classifying satellite images in five classes. Atlanta, Georgia's land cover changes 1973-1998 were categorized into six different classes (Yang and Lo, 2002). Canaz et al. (2017), classified Istanbul, Turkey, in four different classes to monitor land cover change between the years of 1986-2015. On the other hand, comparing with optical sensor data, a new technology to collect remotely sensed data is called as Light Detection and Ranging (LiDAR) have also been subjected as a popular data for classification studies. LiDAR technology is capable of collecting 3 Dimensional (3D) point cloud data in a short time day or night. Because of the direct 3D data acquisition, LiDAR data also have been increasingly used for classifying urban areas into classes.

Classical data-driven techniques have been developed for urban area classification (Rottensteiner and Briese, 2002, Charaniya et al. 2004), the recent trend is to use machine learning techniques to classify LiDAR data in urban area (Lodha et al., 2006). Supervised machine learning techniques are based on selected features and classifier algorithm. In the literature, a variety of supervised classification techniques, support vector machines, neural networks, exists (Richards, J.A., and Jia), in this study one of the supervised classification technique, called as Random Forest (RF) was selected and used because of its stability and robustness to the features.

RF classification for airborne LiDAR data has been studied using different features in order to label different classes. For instance, Niemeyer et al. (2012) classified three different area from Vaihingen, Germany LiDAR dataset named as 'ISPRS Test Project on Urban Classification and 3D Building Reconstruction'. The authors classified data into five categories; building, low vegetation, tree, terrain, and asphalt ground using Conditional Random Field (CRF) approach. However, they only showed and evaluated the result only for classes building and tree. Their correction result for classification for the 3 subset area of the data was found in average 73% and 92% for the tree and building classes, respectively. Guo et al. (2010) use a combination of optical multispectral and LiDAR data to classify LiDAR data in urban area in four classes using the Random Forest (RF) algorithm. Many other studies using RF algorithm to classify LiDAR data can be found in the literature (Immitzer, et al., 2012; Rodriguez-Galiano et al. 2012; Guan et al., 2013).

Lodha et al. (2006) employed another LiDAR data classification work. The authors used Support Vector Machines (SVM) for classifying LiDAR data into buildings, trees, roads, and grass using five features: height, height variation, normal variation, LiDAR return intensity, and image intensity. To evaluate result they compare ground truth and the classification result and observed 90% accuracy. Chen et al. (2013) classified LiDAR data to detect landslides in Three Gorges, China by using the mean aspect, Digital Terrain Model (DTM), and slope textures based on four texture directions; aspect, DTM, and slope textures based on aspect; and the moving average and standard deviation (stdev) filter of aspect, DTM, and slope and RF algorithm. By combine feature selection method with RF algorithm, they found a reliable result for classifying LiDAR data and detection of landslides. Ma et al. (2017) studied a comparison between SVM and RF algorithm to classify LiDAR data. The authors classified data in four categories: trees, buildings, farmland, and ground. According to their findings, the RF algorithm gave a better result than the SVM algorithm for the classification of the LiDAR data.

In this study, an area from the Bergama district of İzmir province, Turkey was chosen as study area. The study area is very complex in shape. The feature classes in interest are located very close to each other and some buildings and trees are embedded. Thus, the originality of the study is that the selected study area is very complex in shape. Therefore, digitization and generation of ground truth for the study area were carried out very carefully. After creating the ground truth and 12 features (which were generated from LiDAR data such as intensity, planarity, DSM etc.) were used to employ classification of LiDAR data.

2. STUDY AREA AND DATA

The study area was chosen from Bergama District of İzmir. İzmir is one of the biggest provinces in Turkey and located in western Turkey. Bergama is the biggest district of İzmir in the size of the area. The area of Bergama is 1573 km^2. The population of the district in 2017 is 102.961.

The study area is located in the center of Bergama district (Fig. 1). The boundary of İzmir province is shown with the blue line, and the boundary of Bergama district is shown in red line in Fig. 1. The true orthophoto of the study area is also shown in Fig. 1. Since the study area's land cover mainly consists of ground, roads, trees, and buildings, the study area divided four groups for classification: buildings, trees, ground, and asphalt road.

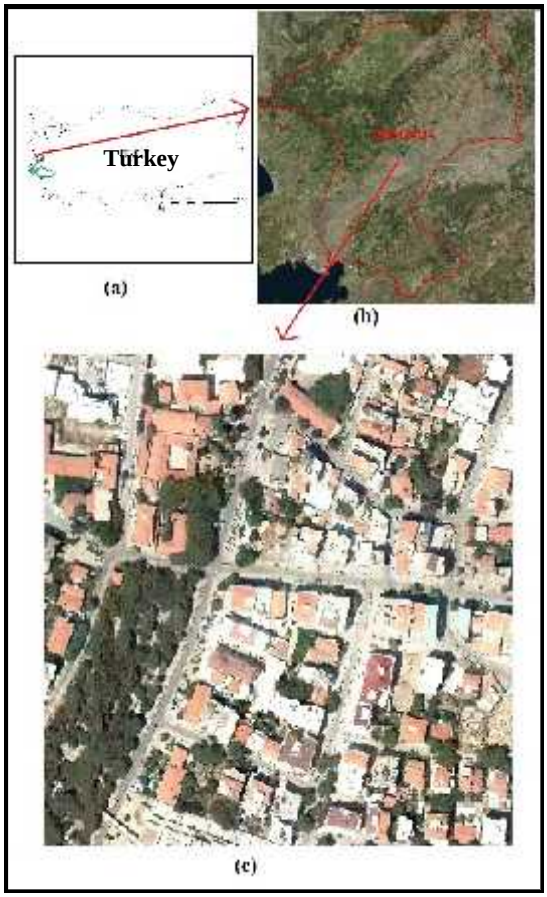


Figure 1. Location of study area, İzmir province boundaries (a), Bergama district boundaries (b) *(source: google maps)*, and the chosen study area (c)

True orthophotos of the study area were generated by Directory of Geographic Information Systems. The images were acquired in May 2016. The pixel size of the images is 10 cm. LiDAR data of the study area was collected by Optech Pegasus HA-500 technic by Turkish General Command of Mapping on 20-21 October 2014 (Kayı et al. 2015). Detailed information about the Optech Pegasus HA-500 is given in Table 1 (Optech, 2018).

Table 1. Technical information of Optech Pegasus HA-500 (Kayı et al. 2015)

Feature	Value
Height	150-5000 m
Effective laser repetition rate	100-500kHz
Scanning Angle	0-75º Adjustable
Accuracy (KOH)	≤ 5-20 cm.
Scanning Mechanism	Oscillating

3. METHODOLOGY

RF algorithm is often used in remote sensing applications to classify data such as multi and hyperspectral images, radar, LiDAR and thermal data sets. A literature review of these applications was presented in Belgiu and Dragut article (2016). This study is based on RF on one of the remote sensing data airborne LiDAR for a complex urban area. The flowchart of the methodology of this study is given in Figure 2.

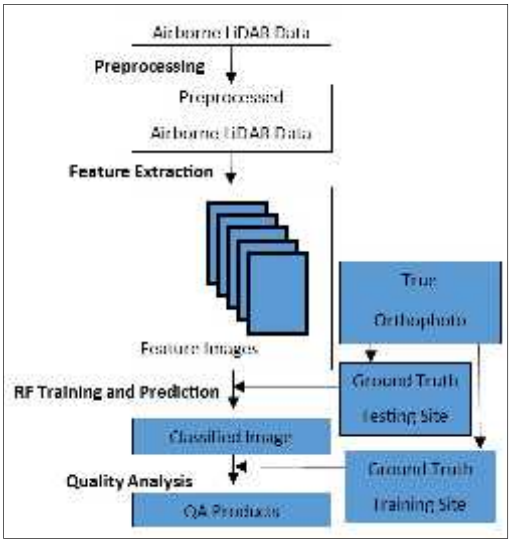


Figure 2. Flowchart of the methodology

The classification of the LiDAR data involves pixel-based classification; therefore, 12 features were generated and rasterized to 50 cm images. Before generating features images and classifying the study area, airborne LiDAR data was cleaned from noisy and duplicate points. After preprocessing, feature images were generated in four groups (Chehata et al., 2009; Dittrich et al., 2017) intensity, height, eigenvalue, and echo based. The intensity-based feature relies on the reflected energy of the objects in the LiDAR dataset. It helps to separate different characteristics objects such as asphalt road and ground classes. Intensity feature image is created using ArcGIS "Las to Raster" tool in 50 cm pixel size. Height based features, on the other hand, were generated from height values of the points and they play a really important role in separating ground and other non-ground classes, such as buildings and trees. The lidar data set was filtered to ground points, then from those points, a DTM in 0.5 m pixel size was generated. In addition to that, a 50 cm Digital Surface Model (DSM) was generated from all points in the LiDAR dataset. Normalized DSM (nDSM) was obtained by subtracting DSM from DTM. Besides, height features based on local neighborhood helps to determine objects, which are also different levels of the surface. *minh*, minimum height value in the neighborhood, and *Hd*, height difference from *minh* of the interested point, were generated for each point in the LiDAR dataset (Table 2). From those features, 0.5 m feature raster were generated using Python programming language (Python, version 2.7).

Table 2. Height based features

Feature	**Description**
nDSM	Normalized Digital Surface Model
minh	Minimum height in local neighborhood of a point
hd	The difference between minimum height in the local neighborhood of a point and that point height

Eigen-value based features were obtained from eigenvalues which were calculated from the local neighborhood covariance matrix. Eigen-values describe the shape of the object, thus they give valuable information about the object, whether it is a plane, line or sphere; therefore, those features are a good indicator of a tree or building roofs, depending on the feature (Table 3). Sphericity, *S*, planarity, *P*, linearity, *L*, anisotropy, *A*, the sum of eigenvalues, *Sum*, and change of curvature, *C*, were calculated and 0.5 m feature images for each feature were generated using Python Programming Language. Geometric features, sphericity, planarity, linearity, and anisotropy describe the shape of the object and give useful information about the object whether it is a line, plane or sphere. All geometric features were created in 3 m neighborhood points per point and then rasterized into 1 m range of mean values.

Table 3. Eigen-value based features

Feature	Description
Anisotropy	$\frac{\lambda 1 - \lambda 3}{\lambda 1}$
Planarity	$\frac{\lambda 2 - \lambda 1}{\lambda 1}$
Sphericity	$\frac{\lambda 3}{\lambda 1}$
Linearity	$\frac{\lambda 1 - \lambda 2}{\lambda 1}$
Change of curvature	$\frac{\lambda_3}{\sum_{i=1}^{3} \lambda_i}$
Sum of eigenvalues	$\sum_{i=1}^{3} \lambda_i$

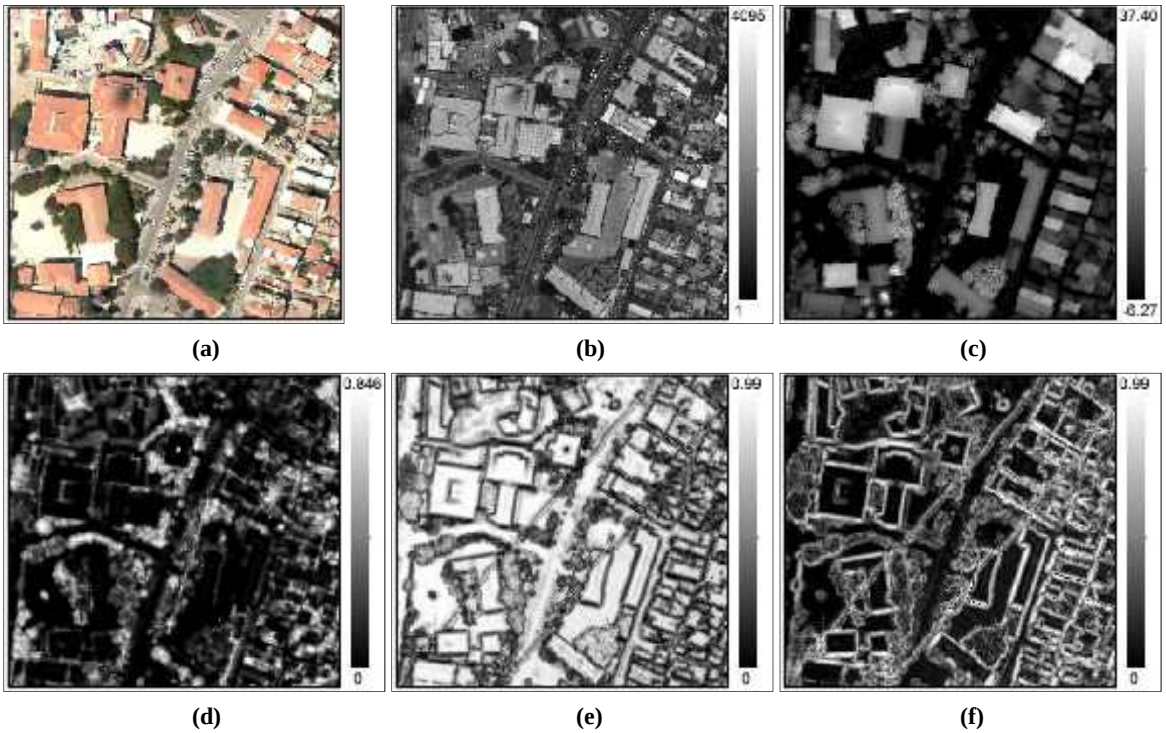
Last feature set, echo based features, helps to differentiate objects, which have multiple returns. Therefore, a total number of return, *n*, and the ratio of a number of return over a total number of returns, *t/n*, were calculated for each point and rasterized to 0.5 m images (Table 4).

Table 4. Echo based features

Feature	Description
n	Total number of returns
t /n	Number of returns over a total number of returns

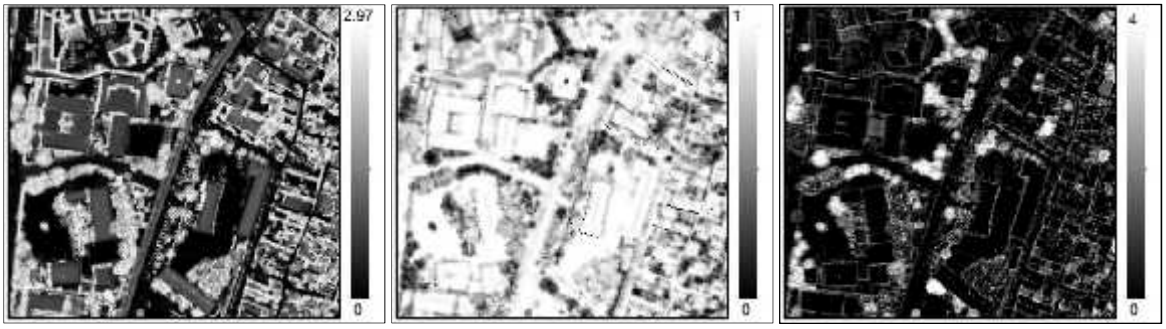
A total number of twelve features was selected and images were generated using Python programming Language and its machine learning and geospatial libraries, including scikit learn (Pedregosa et al., 2011) and GDAL (GDAL, 2018). Some of the features and orthophoto of a part of the study area are shown in Figure 3.

RF classification (Breimen, 2001) is an ensemble method of decision trees, which relies on randomly selecting a subset of features and creating multiple trees in training. and predicting new unlabeled data by voting each tree in the ensemble. Two parameters are required by the user, a number of trees, that define how much a tree can grow up and number of features, which determine how many new nodes can be split from parent node in the tree.



(a) **(b)** **(c)**

(d) **(e)** **(f)**



(g) **(h)** **(i)**

Figure 3. (a) True orthophoto and example of generated feature images; (b) intensity, (c) nDSM, (d) sphericity, (e) planarity, (f) linearity, (g) total number of returns, (h) anisotropy, (i) number of total returns over number of return images.

Ground truth of the study area (red boundary) and the training area (blue boundary) are shown in Figure 4. The study area and training areas were chosen from a different area. According to the similar studies in the literature, the size of the training area was chosen as no lower than the following size: *0,3 x size of the study area*. The study area was fully digitized to use it for quality control of the classification results. Pink, green, black, and yellow colored features represent buildings, trees, asphalt road, and ground, respectively.

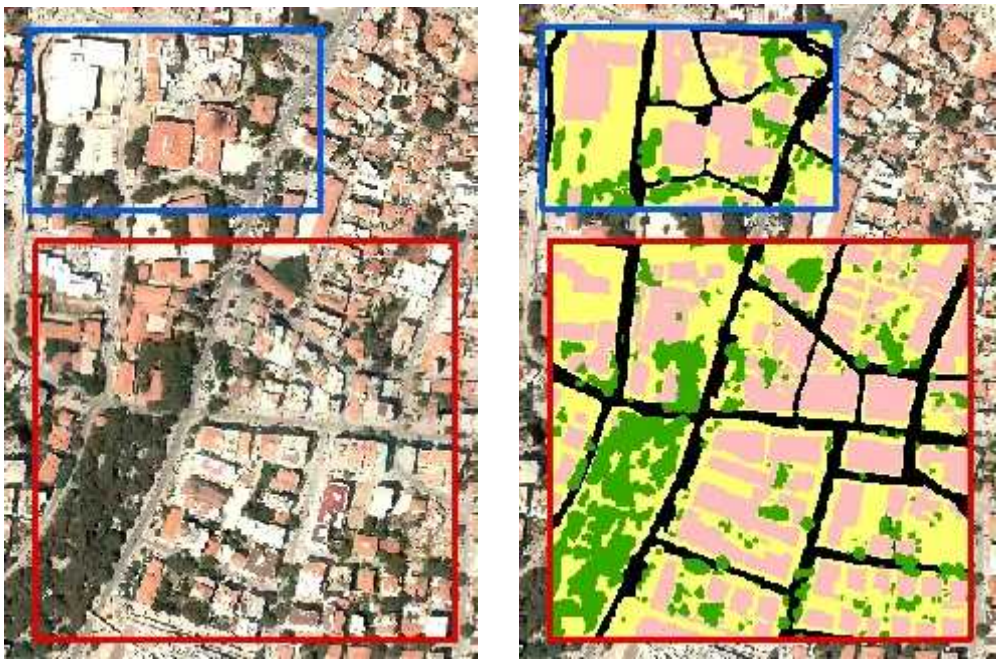


Figure 4. Study area (red), training area (blue) and their manually digitized features

Using the manually digitized training area (blue boundary Fig. 4) and the twelve features, the classification results were acquired by the RF algorithm. The results are described in the following section.

4. RESULTS

Ground truth of the area was created by digitizing the features from orthophoto of the area. Buildings, trees, asphalt road, and the ground were carefully digitized (Figure 5a). A part of the ground truth is used for classification as a training site, while the ground truth of the study area is used for the quality control of the results.

The results were classified into four groups is shown in Figure 5b. In the figure, red, green, gray and blue represent the buildings, trees, asphalt road, and ground classification results, respectively. As it can be seen in figure 5, the classes are extracted with high accuracy by comparing the proposed methodology classification results and the orthophoto of the study area. The qualitative analysis was employed by comparing the ground truth and the classification result. For this

purpose, the difference between the ground truth and result of the classes were created and illustrated in the Fig. 5c.

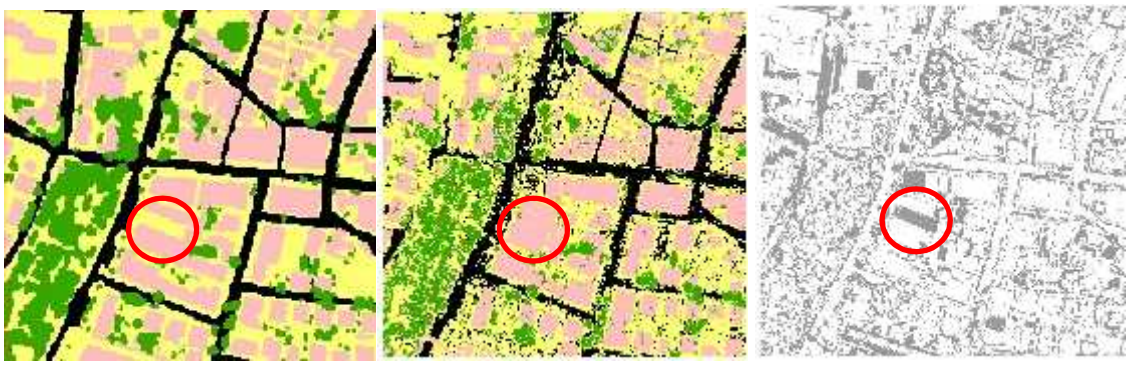


Figure 5. (a) Ground truth, (b) Classification result, (c) Difference

By using the ground truth and the classification results, the quality control was employed, and a confusion matrix was calculated. According to the results, the accuracy is found as, 77,90% , 58,37% , 72,90%, and 71,53% for the buildings, trees, asphalt road, and the ground, respectively. Overall accuracy for the results is 70,20%. Although the area is very complex, the classification results are reliable. Only trees class have lower results than results of the other classes. Some errors occurred since LiDAR data, and the orthophotos, which were used to create ground truth, were acquired in different years and seasons. Therefore, some of the trees might misclassified just because in LiDAR data acquisition time (October 2014), the trees might not have leaves on the trees. On the other hand, orthophotos were collected and created in May, when trees have leaves. Another reason that affects the results is that, in orthophotos, some of the buildings were demolished while they are present in the LiDAR data. For instance, one of the case for this kind of building is shown in Figure 5 with red circles. Finally, there were cars that were on the roads in the LiDAR data, while they are not presence in the orthophoto. This phoneme also mismatch the classification of asphalt roads

5. CONCLUSION

In conclusion, in this study, LiDAR data of a complex urban area from Bergama district, İzmir, Turkey was classified into four groups using the RF algorithm. The classes are as following, buildings, trees, asphalt road, and ground. The area is very complex in terms of city planning for instance buildings' shapes are irregular. The most challenging part for this study was a generation of the ground truth since the area is very complex in shape. Digitization of roads and buildings was very difficult and carried out very carefully. After digitization of the area, twelve features were created from LiDAR data, and using the features and ground truth together, the area is classified by RF algorithm. According to the results, the RF algorithm was classified the area reliably with 70,20% overall accuracy. However, some errors occur because the LiDAR data was acquired in October 2014 and the orthophoto used in this study was collected in May 2016. Because of the seasonal effect, some of the trees were not classified by the proposed methodology. Moreover, in some cases, some building and trees that are available in the orthophoto images, is not found in the LiDAR data, Finally, for the asphalt road, there are car on the roads, which may not be on the LiDAR data or vice versa. These affected the classification results. Even though, these limitations, the proposed methodology is able to classify the complex urban area with high accuracy.

ACKNOWLEDGEMENTS

The author is very thankful to Eray Sevgen, a Ph.D. student at the Hacettepe University for sharing Python scripts for the RF algorithm, helping the feature extraction and digitizing of ground truth data. The author is also thankful to the Turkish Directory of Geographic Information Systems for providing true orthophoto images of the study area and the Turkish General Command of Mapping for providing the LiDAR data of Bergama district.

REFERENCES

Belgiu, M. and Drăguț L. (2016). Random forests in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens., 114 , pp. 24-31.

Breiman, L. (2001). "Random Forests." Machine Learning 45: 5–32.

Canaz S., Aliefendioğlu Y. and Tanrıvermiş H. (2017). Change detection using Landsat images and an analysis of the linkages between the change and property tax values in the Istanbul Province of Turkey. Journal of Environmental Management. Vol. 200:446-45.

Chehata, N., Li, G. and Mallet, C. (2009). Airborne LIDAR feature selection for urban classification using random forests. Geomat. Inform. Sci. Wuhan Univ. 38, 207–212.

Dittrich, A., Weinmann, M. and Hinz, S. (2017). Analytical and numerical investigations on the accuracy and robustness of geometric features extracted from 3D point cloud data. ISPRS J. Photogramm. 126, 195–208.

Charaniya, A.P., Manduchi, R. and Lodha, S.K. (2004). Supervised parametric classification of aerial LiDAR data. In Proceedings of 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04), Washington, DC.

Chen, W., Li, X., Wang, Y., Chen, G. and Liu S. (2014). Forested landslide detection using LiDAR data and the random forest algorithm: a case study of the Three Gorges, China. Remote Sens. Environ. 152 (2014), pp. 291-301.

GDAL/OGR contributors (2018). GDAL/OGR Geospatial Data Abstraction software Library. The Open Source Geospatial Foundation. URL http://gdal.org

Guan, H., Li, J., Chapman, M., Deng, F., Ji, Z. and Yang, X. (2013). Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. International Journal of Remote Sensing, vol. 34, issue 14, pp. 5166-5186.

Guo, L., Chehata, N., Mallet, C. and Boukir, S. (2011). Relevance of airborne LiDAR and multispectral image data for urban scene classification using Random Forests. ISPRS Journal of Photogrammetry and Remote Sensing, 66 (1) , pp. 56-66/

Kayı A. Erdoğan M. and Eker O. (2015). Results of LİDAR test performed by OPTECH HA-500 and RIEGL LMS-Q1560. Harita Dergisi, Volume 153, pp 42-46.

Lodha, S.K., Kreps, E.J., Helmbold, D.P. and Fitzpatrick, D. (2006). Aerial LiDAR data classification using support vector machines (SVM). The Third International Symposium on 3D Data Processing, Visualization, and Transmission pp. 567-574.

Ma L., Zhou M. and Li C. (2017). Land Covers Classification Based On Random Forest Method Using Features From Full-Waveform Lidar Data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-2/W7, 2017 ISPRS Geospatial Week 2017, 18–22 September 2017, Wuhan, China

Niemeyer, J., Rottensteiner, F. and Soergel, U. (2012). Conditional random fields for LiDAR point cloud classification in complex urban areas. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3, pp. 263-26.

Optech, (2018), http://www.Optech.Com/Specification [Accessed on: 14 May 2018]

Richards, J.A. and Jia, X. (1999). Supervised Classification Techniques Remote Sensing Digital Image, Analysis, Springer-Verlag GmbH, Heidelberg (1999) pp. 193–247.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blonde,l M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12:2825-2830.

Python Software Foundation. Python Language Reference, version 2.7. available at http://www.python.org. Retrieved on 15.04.2018.

Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 67 (0), pp. 93-104.

Rottensteiner, F. and Briese C. (2002). A new method for building extraction in urban areas from high-resolution LiDAR data. Int Arch Photogramm Remote Sens Spat Inf Sci 34(3A):295–301

Yang, X. and Lo, C.P. (2002). Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area , International Journal of Remote Sensing, 23, pp. 1775—1798.

Yu, X., Zhang, A., Hou, X., Li, M., and Xia, Y. (2013). Multi-temporal remote sensing of land cover change and urban sprawl in the coastal city of Yantai, China. International Journal of Digital Earth. Vol. 6, Supplement 2, 137-154.