International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 03 – Issue 03 || March 2018 || PP. 114-120

Estimating Carbon Storage through Machine Learning Algorithms

Tuğrul Varol[1], *Birsen Durkaya[2], Eda Okan[3]

[1]*Bartın University, Faculty of Forestry,*
Department of Forest Engineering, 74100, Bartın-Turkey
[1]*Bartın University, Faculty of Forestry,*
Department of Forest Engineering, 74100, Bartın-Turkey
[1]*Bartın University, Faculty of Forestry,*
Department of Forest Engineering, 74100, Bartın-Turkey

Abstract: Forest ecosystems have an important place in carbon conversion by transforming the CO2 they receive from the atmosphere and storing it in large quantities on the earth. Standard models are established and carbon calculations are made for biomass estimations of forest trees. While regression equations are frequently used in the prediction of biomass, estimations made with machine learning algorithms using stand parameters are rarely tested. In this study, it is evaluated whether the parameters of the stand type can be used without using the standard models or equations in the biomass estimation procedure. Verification of biomass estimates via kNN (Kernel Nearest Neighbor), RF (Random Forest) and RPART (Recursive Partitioning and Regression Trees) from machine learning algorithms for coniferous, broad-leaved and mixed stands in the Amasra, Arıt and Kurucaşile Sub-district Directorates of Bartın Forestry Directorate have been carried out. Amasra, Arıt and Kurucaşile regions, where the work was carried out, have 121, 79 and 121 stand types respectively. Carbon calculations for five diametric classes were carried out using the data for the tables of stand identification. Total carbon stocks were found to be 111 tons/ha, 115 tons/ha and 179 tons/ha for Amasra, Arıt and Kurucaşile regions respectively. Carbon stock values calculated by regression equations; it can be estimated as 40%, 85%, 99% with the KNN algorithm, 42%, 57%, 85% with the RPART algorithm and 71%, 78%, 80% with the RF algorithm in the mixed, coniferous and broad-leaved stands, respectively.

Keywords: Carbon storage, Machine Learning, kNN, Regression tree, Random forest.

1. Introduction

Forest ecosystems play an important role in carbon conversion by converting CO2 from the atmosphere and store it in vast amount on earth [1, 2]. It is estimated by the scientists that the average surface temperature will rise from 1.4oC to 5.8oC to the end of the 21. Century [3] and is thought that the increasing amount of fossil fuels used causes this problem. While it is expected that this increase in the temperatures boosts the CO2 emission because of the organic mineralization in soil, an increase is seen in the biomass of plants [4]. Thus, forest inventory is the most important resource that could be used to evaluate the carbon change on earth [5]. Therefore, it was tried to form standard models in order to make aboveground biomass estimations in many areas. Carbon accounts are made using biomass that accumulates in forest ecosystems. There are two basic approaches to the calculation of forest biomass. (1) above-ground and below-ground carbon values are calculated from stem volume using biomass expansion factors (BEF). (2) Allometric biomass equations use biomass models developed for each tree type and region. In this method, independent variables such as diameter, height, and specific weight of the tree are used for biomass calculations [6, 7, 8, 9, 10]. Depending on the approach used, biomass quantities are converted to carbon values with different coefficients. This mainly stems from that forest areas involves a wide range of tree species and various growth conditions. However, the regression models in use are obtained from chopped trees in few quantities and they sometimes contain a very limited number of large-diameter trees. So, it is important question how much these trees represent the sampled area. It could be the answer to why two regression models make different estimations or why the results from regression models differ from the estimations from mostly used BEF coefficients. This difference at stand level is greater for large-diameter trees and may cause uncertainty in estimations [5, 11, 12]. Since there are a few research studies about regression equations obtained for the areas having mostly the same or similar features, it gets harder to evaluate the quality of these models.

While regression equations are frequently used in above-ground carbon estimation, estimates made with automatic learning algorithms using stand parameters have rarely been tested. In this study, biomass estimations can be made with the regression equations obtained associating with the diameter, length values, dry weights of chopped trees, also it is estimated how successful machine learning algorithms could be and how much quality they would be. If accurate estimations can be achieved with machine learning algorithms, it will not be required the tree cuts to represent the entire area and the calculation of their dry weights because the estimations will be able to be made with the number of trees and average diameters from sample areas in the preparation of management plans. In this study, the confirmations of carbon estimations were made through kNN (Kernel Nearest Neighbor), RF (Random Forest) and RPART (Recursive Partitioning and Regression Trees) as machine learning algorithms for broad-leaved, coniferous and mixed stands in the Amasra-Arıt-Kurucaşile Forest Sub-district Directorates affiliated to Bartın Forestry Operation Directorate.

2. Study Area

This study was carried out in the Arıt-Amasra-Kurucaşile sub-district Directorates. The three areas have a total of 30766.5 ha forest land, 83.2% of which is productive according to the 2011-2030 management plan and this region is located between 32017′55″ -320 46′37″ east longitudes and 410 33′ 90″ – 41051′ 01″ north latitudes (Figure 1). 81.7% of the forest lands at the average altitude, 853m, are productive. The average annual temperature in the region is 12.80C and the average annual rainfall is 1140 mm. The precipitation can be seen in any season and month ranging from 5-13%. The geological structure of the study area is constituted by fine grained and impermeable rock and brown podzolic soil spread in the area in the south part. While the south part of the area is of red sandstone, clay Stone and dolomite limestone, marn, andesite, tuff and agglomerate are seen in the north.

121 stand types in Amasra, 79 in Arıt and 121 in Kurucaşile exist. In each of these take place broad-leaved, coniferous and mixed forests and tree species are black pine, scotch pine, calabrian pine, stone pine, abies, fagus, oak, hornbeam, chestnut, lime. The number of trees and average diameters in each hectare of the forest lands in these three regions can be seen in Table 1.

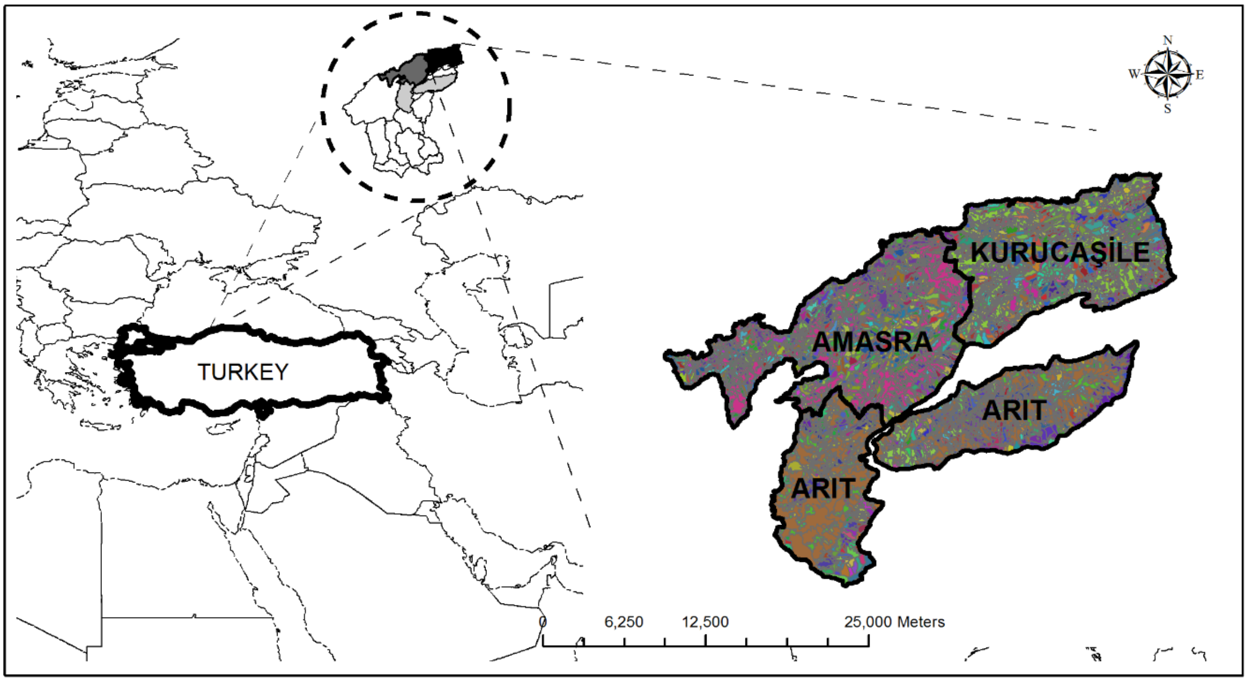

Figure 1: Stand Type Map according to 2011-2030 management plans belonging to Amasra-Arıt-Kurucaşile regions

Table 1: Data from Amasra-Arıt-Kurucaşile regions

Region	Stand Type	Area (ha)	Average diameter (cm)	Tree number (N/ha)
Amasra	Coniferous	710.8	15.8	244
	Broad-leaved	8649.1	22.8	756
	Mixed	313.3	27.0	648
	Degraded	1371.4	-	-
	Total	***11044.6***	**-**	***1648***
Arıt	Coniferous	833.5	18.0	703
	Broad-leaved	5650.5	21.5	598
	Mixed	643.3	21.1	597
	Degraded	1342.3	-	-
	Total	***8469.6***	**-**	***1898***
Kuruca□ile	Coniferous	613.2	19.8	520
	Broad-leaved	7327.4	21.6	839
	Mixed	860.7	20.4	647
	Degraded	951.5	-	-
	Total	***9752.8***	**-**	***2006***
TOTAL		**29267**	**-**	**5552**

3. Method

Biomass was calculated by taking tree species into account for coniferous, broad-leaved and mixed stands in each region. Biomass values are also multiplied by various coefficients to find the amount of carbon. Above-ground and below-ground biomass calculations were made with the regression equations by Durkaya et al. [13] for oak (Quercus spp.), Saraçoğlu [14] for fagus (FagusorientalisLipsky), İkinci [15] for chestnut (Castaneaspp.), Durkaya et al. [16] for black pine (Pinusnigra), Durkaya et al. [17] for scotch pine (Pinussylvestris), Durkaya et al. [18] for abies (Abiesnordmanniana subsp. bornmulleriana), Durkaya et al. [19] for cedar (Cedruslibani L.), Durkaya et al. [20] for calabrian pine (Pinusbrutia Ten.). These values are converted to carbon values by multiplying by 0.5. Carbon values determined with kNN, RPART and PRF algorithms by using the data belonging to the area, number of trees and five diameter classes (I. diameter class: 8.0-19.9 cm, II. diameter class: 20.0-35.9 cm, III. diameter class: 36.0-51.9, IV. diameter class: 52.0-59.9, V. diameter class: > 60.0 cm). It is thought that successful estimations with such types of data found as standard in management plans would make carbon estimations easier.

Learning methods with artificial intelligence have become more useful for modeling the complex relationships and interactions without limiting the assumptions of parametric statistics [21, 22]. Some of these methods are kNN [23], RPART [24] and RF [25].

kNN is one of the earliest and simplest methods used in model classifications among machine learning methods. kNN tags each of tagged samples according to the nearest neighbors in the data set. Therefore, it performance depends on distance calculations used in the calculation of nearest neighbors (Euclidean distance, Minkowski distance, Mahalanobis distance) [26].

RPART stands out as a powerful statistical tool in order to analyze the complex ecological data sets. One of the most important reason is that independent variables bear useful alternatives while modeling the non-linear data interacting with each other [24]. Regression trees have been used in numerous ecological practices such as the relationship between the severity and frequency of forest fires [27].

RF is a learning algorithm that produces multiple classifiers instead of single classifier and later groups the new data (x) with the estimations (h (x, Ø_k), k = 1, ...). R statistical language [28] was used in the quadratic error measurements. In R software, the following software packages were used: "Random Forest" for RF, "kknn" for kNN calculations and "rpart" for RPART calculations. For the estimation, the70% of the measured data is used and the remaining data (30%) in validation.

The accuracy of the estimations obtained via algorithms was tested by confusion matrices. P or PPV (accuracy), R or TPR (precision), V or F1 (weighterd mean of TPR and PPV) and Ac. (accuracy of classification) values were calculated through TN (True negative), FP (False positive), FN (False Negative) and TP (True Positive) which were determined by using real and estimated results in confusion matrices (Table 2).

Table 2: Confusion Matrix

		PREDICTED CLASS	
		POSITIVE	NEGATIVE
ACTUAL CLASS	POSITIVE	True positives (TP)	Falsenegatives (FN)
	NEGATIVE	Falsepositives (FP)	True negatives (TN)

4. Results

As a result of the study, the carbon values (Table 3) obtained from the calculations with the 2011-2030 management plan data belonging to Amasra-Arıt-Kurucaşile Forest Sub-district Directorate and carbon estimations made with machine learning algorithms for coniferous, broad-leaved, mixed and all stand types belonging to these regions are listed below (Table 4).

As could be seen in Table 3, the current amount of carbon storage in the management plan (2011-2030) for each of three regions (Amasra-Arıt-Kurucaşile) is respectively 1073553 tons, 819590 tons and 1573495 tons. The greatest carbon storage is in Kurucaşile and Amasra region is of the largest forest land (11045 ha) while Kurucaşile is of a forest land of 9753 ha and Arıt has 8470 ha.

Table 3: Carbon amounts calculated from biomass equations

Region	**Aboveground Carbon**	**Underground Carbon**	**Total (ton)**	**Carbon Sequestration in Hectare**
Amasra	862497	211056	1073553	97
Arıt	649683	169906	819590	97
Kurucaşile	1259349	314146	1573495	161
TOTAL	**2771529**	**695108**	**3466638**	**355**

As can be seen in Table 4, when taken into account all the stands, each of the three algorithms bear very close results (kNN and RPART 88.46% [V:84.21], RF 84.85% [V:82.30%]. The biggest difference among the algorithms occurs in mixed stands and the highest accuracy belongs to RF algorithm 78.95% (V:71.43%). While kNN algorithm gives the best results in broad-leaved and coniferous stands, RPART and RF algorithms have very close results. In coniferous stands, differences between algorithms are at almost similar rates (10%). RPART algorithm using the method of regression trees has similar or lower estimation values than RF algorithm using the multiple version of this method. It is thought that the reason why RF algorithm has more successful results especially in mixed stands stems from that multiple regression tree was formed with the area, number of trees and diameter variables and the neighboring relationship of coniferous and broad-leaved trees at various diameters was complex. This assumption is also supported by that kNN algorithm show better results in coniferous and broad-leaved stands and when considered all the stands and mixed stands, kNN algorithm results are of either lower estimation percentages or similar results.

Table 4: Carbon estimation results obtained via machine learning algorithms

Stand Type	kNN				RPART				RF			
	P(%)	R(%)	V(%)	Ac.(%)	P(%)	R(%)	V(%)	Ac.(%)	P(%)	R(%)	V(%)	Ac.(%)
Coniferous	100.00	75.00	85.71	90.00	50.00	66.67	57.14	70.00	84.62	73.33	78.57	80.65
Broad-leaved	98.75	100.00	99.37	98.77	89.28	80.65	84.75	88.89	80.49	80.49	80.49	86.89
Mixed	25.00	100.00	40.00	53.85	45.52	39.61	40.35-	41.67	71.43	71.43	71.43	78.95
All type	91.43	78.05	84.21	88.46	88.89	80.00	84.21	88.46	81.58	83.04	82.30	84.85
Average	96.73	84.35	89.77	92.41	76.06	75.77	75.37	82.45	82.23	78.95	80.45	84.13

5. Conclusion and Recommendations

It is an important issue to determine the carbon stored in forest ecosystems accurately and reliably in the framework of fight against global climate change. Forests absorb carbon dioxide from the atmosphere like oceans and store in their structures for a long period. With this aim, various methods have been developed in order to determine the stored carbon amount in stands. As the storage capacity could be determined via biomass equations for plant species in each region, the results can also be obtained from the several coefficients. In Turkish forestry system, coefficients are used by the Regulation on Ecosystem based Functional Forest Management Plans [29] and the practices in different disciplines also exist [30]. The calculation method with these coefficients produces more practical results in determining the carbon storage capacity in large areas.

The biomass equations bear more realistic results than EFFMP method [26, 27], the results belonging to these equations were used in the estimations with machine learning algorithms. In management plans are used the stand size, number of trees and average diameter values on calculating with machine learning algorithms. So, no additional workload is required to estimate the carbon amount in present and future. When Table 4 is analyzed, it is seen that the algorithms such as kNN, also showing successful results in other branches of forestry [33, 34] could be used in the studies which do not require precise results for large spaces. We believe that in the prospective studies it would be also useful to test different machine learning algorithms such as SVM (Support Vector Machine), artificial neural networks and Naive Bayes for determining the carbon storage amounts.

References

[1] Fuchs, H., Magdon, P., Kleinn, C.,Flessa, H. "Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory". Remote Sensing of Environment, 113(3), 518-531, 2009.

[2] Jung, J., Kim, S., Hong, S., Kim, K., Kim, E., Im, J.,Heo, J. "Effects of national forest inventory plot location error on forest carbon stock estimation using k-nearest neighbor algorithm". ISPRS journal of photogrammetry and remote sensing, 81, 82-92, 2013.

[3] Ramachandran, A., Jayakumar, S., Haroon, R. M., Bhaskaran, A.,Arockiasamy, D. I. "Carbon sequestration: estimation of carbon stock in natural forests using geospatial technology in the Eastern Ghats of Tamil Nadu, India". Current Science, 92(3), 323-331, 2007.

[4] White, A., Cannell, M. G. R., Friend, A. D. "The high-latitude terrestrial carbon sink: a model analysis". Global Change Biology, 6(2), 227-245, 2000.

[5] Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., ...,Lescure, J. P., "Tree allometry and improved estimation of carbon stocks and balance in tropical forests". Oecologia, 145(1), 87-99, 2005.

[6] Schroeder P., Brown S., Mo J., Birdsey R., Cieszewski C., "Biomass estimation for temperate broadleaf forests of the United States using inventory data", For. Sci. 43,424–434, 1997.

[7] VandeWalle I, Van Camp N, Perrin D, Lemeur R, Verheyen, K, Van Wesemael B, Laitat E. "Growing stock-based assesment of the carbon stock in the Belgian forest biomass". Ann.For.Sci. 62, 853-864, 2005.

[8] Durkaya, B.,Varol, T., Durkaya, A. "Determination of carbon stock changes: biomass models or biomass expansion factors" Fresenius Environmental Bulletin. Vol:23 (3), pp. 774-781, 2014.

[9] Durkaya, A., Durkaya, B., Makineci, E., Orhan, İ. "Turkish Pines' Aboveground Biomass and Carbon Storage Relationships" Fresenius Environmental Bulletin Vol:24 (11), pp. 3573-3583, 2015.

[10] Güner, Ş.T., Çömez, A, "Biomass Equations And Changes In Carbon Stock In Afforested Black Pine (PinusNigra Arnold. Subsp. Pallasiana (Lamb.) Holmboe) Stands In Turkey". Fresenius Environmental Bulletin. Vol:26 (3), pp 2368-2378, 2017.

[11] Clark, D. B., Clark, D. A. "Landscape-scale variation in forest structure and biomass in a tropical rain forest". Forest ecology and management, 137(1), 185-198, 2000.

[12] Houghton, R. A., Lawrence, K. T., Hackler, J. L., Brown, S. "The spatial distribution of forest biomass in the Brazilian Amazon: a comparison of estimates", Global Change Biology, 7(7), 731-746, 2001.

[13] Durkaya, B., Durkaya, A.,Saraçoğlu, N. "Construction of Biomass Tables of Quercus in Zonguldak Forest Region Administration". XXI. IUFRO World Congress, 7-12 August 2000, Kuala Lumpur, Poster Abstracts, Volume 3, pp. 149, 2000.

[14] Saraçoğlu, N. "Biomass tables of beech (Fagus orientalisLipsky)." Turkish Journal of Agriculture and Forestry, 22(1), 93-100, 1998.

[15] İkinci, O. ZonguldakOrmanBölgeMüdürlüğüKestaneMeşcereleriBiyokütleTablolarınınDüzenlenmesi. YüksekLisansTezi (yayımlanmış), ZKÜ Fen BilimleriEnstitüsü, OrmanMühendisliğiAnabilim Dalı, Zonguldak, 135 s, 2002.

[16] Durkaya, A., Durkaya, B.,Çakıl, E. “Predicting the above-ground biomass of crimean pine (Pinusnigra)stands in Turkey”, Journal of Environmental Biology, Vol 31, pp115-118, 2009.

[17] Durkaya, A., Durkaya, B.,Atmaca, S. “Predicting the Above-ground Biomass of Scots Pine (Pinussylvestris L.) Stands in Turkey”. Energy Sources, Part A, 32:485-493, DOI:10.1080/15567030802612473, 2010.

[18] Durkaya, B., Durkaya, A.,Makineci, E.,Karabürk, T. “Estımatıng Above-Ground Bıomass and Carbon Stock of Indıvıdual Trees in Uneven-Aged Uludag Fir Stands”. Fresenius Environmental Bulletin, Vol:22 (2), pp. 428-434, 2013.

[19] Durkaya, B., Durkaya, A., Makineci, E.,Ülküdür, M. ”Estimation of Above-Ground Biomass and sequestered Carbon of Taurus Cedar (Cedruslibani L.) in Antalya, Turkey”. iForest-Biogeosciences and Forestry. 6:278-284. DOI:10.3832/ifor0899-006, 2013.

[20] Durkaya, A., Durkaya, B.,Ünsal, A. ” Predicting the above-ground biomass of calabrian pine (Pinusbrutia Ten.) stands in Turkey”. African Journal of Biotechnology, Vol 8 (11), pp. 2483-2488 ISSN 1684-5315 © 2009 Academic Journals, 2009.

[21] Olden, J. D., Lawler, J. J.,Poff, N. L. “Machine learning methods without tears: a primer for ecologists”. The Quarterly review of biology, 83(2), 171-193, 2008.

[22] De'Ath, G. “Boosted trees for ecological modeling and prediction”. Ecology, 88(1), 243-251, 2007.

[23] Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., Pereira, J. M. “Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest”. Forest Ecology and Management, 275, 117-129, 2002.

[24] De'ath, G.,Fabricius, K. E. “Classification and regression trees: a powerful yet simple technique for ecological data analysis”. Ecology, 81(11), 3178-3192, 2000.

[25] Stojanova, D., Panov, P., Kobler, A., Džeroski, S.,Taškova, K. “Learning to predict forest fires with different data mining techniques”. In Conference on Data Mining and Data Warehouses (SiKDD 2006), Ljubljana, Slovenia , pp. 255-258, 2006.

[26] Weinberger, K. Q., Blitzer, J., Saul, L. “Distance metric learning for large margin nearest neighbor classification”. Advances in neural information processing systems, 18, 1473, 2006.

[27] Rollins, M. G., Keane, R. E., Parsons, R. A. “Mapping fuels and fire regimes using remote sensing, ecosystem simulation, and gradient modeling”. Ecological Applications, 14(1), 75-95, 2004.

[28] R Development Core Team. R, “A language and environment for statistical computing”. http://www.R-project.org, 2006. [Online]. Available: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing. [Accessed: March. 13, 2018].

[29] OGM, EkosistemTabanlıFonksiyonelOrmanAmenajmanıPlanlarınınDüzenlenmesineAitUsulveEsaslar. Ankara, 2014.

[30] Gençay, G., Birben, Ü., Durkaya, B. “BirAraziKullanımDeğişikliğiÖrneği 2/B Arazileri” XIII. Congress of Ecology and Environmental With International Participation, Edirne, 2017.

[31] Durkaya, B., Varol, T.,Okan, E. “ Carbon Stock; Kurucaşile Forest Sub-District Directorate”, International Journal of Recent Engineering Research and Development, 2(4), 59-66, 2017.

[32] Durkaya, B., Varol, T., Durakaya, A. “Determinatıon of Carbon Stock Changes: Bıomass Models Or Bıomass Expansıon Factors”, Fresenius Environmental Bulletin, 23 (3), 774-781, 2014.

[33] Özel, H.B., Varol, T.,Bilir, N., Kapukıran, İ. “Estimation of Number of Natural Juvenilities in Oriental Beech (Fagus orientalisLipsky.) andUludag Fir (Abiesnordmanniana subsp. bornmüllerianaMattf) Mixed Stands in the Kumluca Forest Range Directorate By Different Automatic Learning Algorithms”, XIII. Congress of Ecology and Environmental With International Participation, Edirne, 2017.

[34] Karayılmazlar, S., Özel, H.B., Varol, T.,Varol, H.N. “Determination of Potential Rehabilitation Areas Using by Dynamic Analytical Hierarchy Process (DAHP) Method and Geographic Information Systems (GIS) in the Bartın-Sökü Forest Range District in Turkey”, International Forest Symposium, Kastamonu, 2017.

Author Profile

Tuğrul Varol was born 04 June 1972. Tuğrul Varol finished Bachelor Sc. at 1992, Master of Science at 1996 and Ph.D at 2004 in the Karadeniz Technical University, Zonguldak Karaelmas University in Trabzon and Bartın city in Turkey. Tuğrul Varol has been special issues forest road, forest fire, GIS and Forest Mechanization. Varol was started working research assitant between 1994-2004 and he worked Assist. Prof. Dr. at 2004 still in the Bartın Faculty of Forestry. He published a lot of paper and manuscript various national and international journals about his special issues.

Birsen Durkaya was born 08 July 1973. Birsen Durkaya finished Bachelor Sc. at 1993, Master of Science at 1998 and Ph.D at 2004 in the Karadeniz Technical University, Zonguldak Karaelmas University in Trabzon and Bartın city in Turkey. Birsen Durkaya has been special issues forest biomass, management, forest yield. Durkaya was started working research assitant between 1994-2004 and she worked Assist. Prof. Dr. at 2004, she has worked in the Bartın Faculty of Forestry as Assoc. Prof. Dr since 2013.Shepublished a lot of paper and manuscript various national and international journals about his special issues.

Eda Okan was born 21September 1989. Eda Okan finished Bachelor Sc. at 2012. She has Master of Science at 2018 from the Bartin University