Evaluating some artificial neural networks and multiple linear regression model for predicting carbon of pure oriental beech stand in Göldağ forests

Alkan Günlü[1,*], İlker Ercanlı[1]

[1] Çankırı Karatekin University, Faculty of Forestry, Department of Forest Engineering, Çankırı, Turkey
[*] Corresponding author: alkangunlu18@gmail.com

Abstract: Forest ecosystems provide timber and non-timber products, recreation; water and soil protection, biological diversity conservation as well as they decreased a huge amount of CO_2 in the atmosphere. As stated in some international agreements and acts like "Kyoto Protocol", the simplest, easiest and cheapest way for decreasing climate change is protection of forest ecosystems, reforestation and afforestation. Recently, many scientific studies have been conducted for determining carbon quantities of forest trees in the World's forest ecosystems. On the other hand, remote sensing methods have been used to estimate aboveground carbon. Many prediction models based on linear regression analyze for stand attributes and stand carbon have been developed in forest literature. However, this analyze based on some statistical assumptions, normally distributed residuals and homoscedastic trends in predictions, and if these assumptions are violated, these predictions can be biased and erroneously obtained in forest applications. As remedy for this problem in predictions, Artificial Neural Network Analysis (ANN) has been successfully introduced in forest literature for modeling different individual and stand attributes. In this study, stand carbon storage were firstly calculated from BEF coefficients of Tolunay (2013) for 70 sample plots in Göldağ Forests. Vegetation indices values were obtained from Landsat 7 ETM. To model the relationships between stand carbon and vegetation indices, the multiple linear regression analysis and some artificial neural networks such as the feed-forward backprop, Cascade Correlation, Elman backprop, Layer Recurrent, NARX network and Radial basis with training function of Levenberg-Marquardt and transfer function of Log-sigmoid transfer function were used in this study. These artificial neural networks (ANNs) structures such as the feed-forward backprop, Cascade Correlation, Elman backprop, Layer Recurrent, NARX network and Radial basis network with multiple linear regression model were compared based on evaluations of the magnitudes and distributions of models' residual and six goodness-of-fit statistics: sum of squared errors (SSE), Akaike's information criterion (AIC), Bayesian information criterion (BIC), Root Mean Square Error (RMSE), Mean Squared Error (MSE) and Adjusted Coefficient of Determination (R^2_{adj}). The ANN based on the Cascade backprop produced better predictive ability with SSE (28353.36), AIC (424.280), RMSE (20.126), R^2 (0.510) than the regression model and the other studied ANNs. These fit statistic results in the ability of ANN to predict stand carbon values by using vegetation indices from satellite image and producing more precise predictions than multiple linear regression models.
Keywords: Stand carbon, Remote sensing data, Artificial neural network, Prediction

1. Introduction

Forest biomass is cursor of carbon sequestration. Nearly 50% of forest dry biomass is carbon (Brown, 1997). The greater part of biomass evaluations is performed for the above ground biomass (AGB) of trees. Traditional biomass and carbon evaluation technique depend on ground measurements are the most accurate technique; however, but it is an extremely time overwhelming and destructive method, generally limited to minor areas and minor tree sample sizes (Attarchi and Gloaguen, 2014). Recently, remote sensing technique have been used to gather data regarding the AGB and carbon on large forest areas (Maynard et al., 2007). Remote sensing technique has been applied to the AGB evaluate in many revisions (Maynard et al., 2007; Wannasiri et al., 2013). Investigators have used different modelling technique such as linear regression models, multiple regressions models, artificial neutral networks and so on. The aim of this study is to evaluate the relationships between band reflectance values and vegetation indices obtained from Landsat 7 ETM satellite image and the stand carbon storage from field measurements by using multiple regression analysis and some artificial neutral networks for pure beech forests in north of Turkey.

2. Material

2.1. Study area

The research area, a part of Göldağ Forest Planning Unit is located in Kastamonu Regional Forest Directorate with a total area of 600 ha. It is bounded by 647000-650000 on the east longitudes and 4629000-4632000 on the North latitudes (ED 1950, UTM Zone 36N). Average altitude, precipitation and temperature of research area are 775 m, 677.3 mm and 17.6 C°, respectively (Günlü et al., 2008). The study area is covered by trees that include unmanaged, even-aged, pure stands of oriental beech (*Fagus orientalis* Lipsky.). In this study, required data were obtained from 70 temporary sample plots ranging crown closure. In the field survey, the size of circular plots ranged from 400 to 800 m2, depending on crown closure. The diameter at breast height (dbh) of each tree and spatial coordinates of sample plots were measured. Dbh was considered for every trees greater than 8.0 cm. Stand volume was calculated for each plot and converted to values in hectare. In addition to, The Landsat 7 ETM satellite image, which was consisted of six spectral bands (ETM1, ETM2, ETM3, ETM4, ETM5 and ETM7) with 30 m spatial resolution, was acquired on May 3, 2000 used as the remote sensing data.

3. Method

3.1. Carbon estimation through standing volume

Prediction of forest biomass are required for monitoring changes in Carbon stocks. Investigators use different methods such as allometric biomass equations or Biomass Expansion Factors (BEFs) to forest inventory in biomass estimations. A common method is the use of BEFs. In fact, BEFs is simple method of converting from forest tree stem volume to total forest biomass (Brown et al. 1999; Hu and Wang 2008; Keleş et al. 2012; Kadıoğulları and Karahalil 2013; Karahalil 2013). In this study, above ground biomass of each sample plot was calculated. The aboveground biomass for each sample plot was calculated using BEFs by developed Tolunay (2013).
The amounts of above ground carbon of each sample areas were calculated using the above ground biomass values calculated for each sample area.

3.2. Estimation vegetation indices values

Some vegetation indices (NDVI, SR, DVI, TVI, NLI, SAVI, ND53, ND54, ND57, ND32 and ND73) were calculated using bands of Landsat 7 ETM satellite imagery.

3.3. Analyzed models

To model the relationships between stand carbon and vegetation indices, the multiple linear regression analysis was used in this study. This multiple linear regression models based on stepwise variable selection method was developed though Ordinary Least Squares (OLS) Technique using vegetation indices as independent variable, which dependent variables in models were stand carbon values. The multiple stepwise regressions were performed using SPSS version 12.0 (SPSS Institute, 2004). The stepwise regression technique was used to select the best vegetation indices variables that are significant ($p<0.05$) with the highest value of coefficient of determination adjusted by number of parameters (R^2_{adj}), also called adjusted the coefficient of determination. In this study, the following linear relationship was assumed:

$$SC = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + ... + \beta_n \cdot X_n + \varepsilon \qquad (1)$$

where SC is the stand carbon values, X_1 …… X_n are variable vectors corresponding to vegetation indices values, β_1 …… β_n represent model coefficients and ε is the additive error term (Corona et al., 1998) as other technique to model the relationships between stand carbon and vegetation indices, the neural network model building and multiple linear regression model were used in this study. both training, verification and testing data sets that randomly partitioned into training (75% of all data), verification (15% of all data) and test (the remaining 10% of all data) data sets were used for modeling the relationships between input variables (vegetation indices, independent variables in regression analysis) and target variable (stand carbon values dependent variables in regression analysis). In ANN training process, input variables were some vegetation indices that gave the best predictive results in regression analysis. Target variable is stand carbon values calculated as sum of individual tree carbon in each sample plots at hectare. The ANNs selected for predicting stand carbon are multiple layer network structures such as the feed-forward backprop, Cascade Correlation, Elman backprop, Layer Recurrent, NARX network and Radial basis with training function of Levenberg-Marquardt and transfer function of Log-sigmoid transfer function. In ANNs training process, the number of neurons is used as 10 with number of two layers including hidden and output layers, since this structure are the most frequently chosen values in ANNs. All these applications for ANN was carried out using MATLAB-nntool module. These artificial neural networks (ANNs) structures such as the feed-forward backprop, Cascade Correlation, Elman backprop, Layer Recurrent, NARX network and Radial basis network with multiple linear regression model were compared based on evaluations of the magnitudes and distributions of models' residual and six goodness-of-fit statistics: sum of squared errors (SSE), Akaike's information criterion (AIC), Bayesian information criterion (BIC), Root Mean Square Error (RMSE), Mean Squared Error (MSE) and Adjusted Coefficient of Determination (R^2adj).

4. Results and discussions

The values of goodness-of-fit statistics, such as SSE, AIC, BIC, RMSE, MSE and R^2_{adj}, for these ANNs, including the feed-forward backprop, Cascade Correlation, Elman backprop, Layer Recurrent, NARX network and Radial basis network and multiple regression model were given in table 1. In this regression model for stand carbon values, the F statistics and coefficients were significant at a probability level of 95 percent ($p<0.05$). The stand carbon model based on the vegetation indices were developed by NDVI and SAVI as independent variables with SSE (36207.78), AIC (441.397), RMSE (22.743), R2 (0.374). Table 1 showed that ANN based on the Cascade backprop produced better predictive ability with SSE (28353.36),

AIC (424.280), RMSE (20.126), R2 (0.510) than the regression model and the other studied ANNs. Using the Cascacade backprop neural network model building, the explanatory in stand carbon predictions determined as Adjusted Coefficient of Determination (R^2adj) increased by % 13.6 and prediction's error determined as RMSE decreased by % 11.51.

Table 1. The goodness-of-fit statistics of number of trees predictions for the ANNs types and regression model.

Prediction Methods including the ANNs and regression model	SSE	R^2_{adj}	MSE	RMSE	AIC	BIC
ANN based on feed-forward backprop	31879.25	0.449	455.418	21.341	432.485	436.982
ANN based on Cascade backprop	28353.36	0.510	405.048	20.126	424.280	428.777
ANN based on Elman Recurrent	31350.12	0.458	447.859	21.163	431.313	435.810
ANN based on Layer Recurrent	31790.74	0.451	454.153	21.311	432.290	436.787
ANN based on Radial Basis	29615.53	0.488	423.079	20.569	427.329	431.826
Linear Regression	36207.78	0.374	517.254	22.743	441.397	445.894

These fit statistic results in the ability of ANN to predict stand carbon values by using vegetation indices from satellite image and producing more precise predictions than multiple linear regression models. This empirical relationship between vegetation indices from satellite image and stand carbon helped to develop successfully for mapping stand carbon. Based on the results obtained from the research, forest managers could use satellite images data for estimating stand carbon and this information would also be beneficial for generating maps and developing forest management plans. It is probable that this study will present appreciated contribution to the literature indicated that ANN scan be utilized in various stand attributes such as stand biomass, carbon, volume others. However, these fitting results should be further examined and assessed at different forest areas meanwhile this modeling consequences of this carbon values from satellite images are specific to this type of forest structure and areas that was studied.

5. Acknowledgment

This work was supported by the Karadeniz Technical University Project Department, Project No: 2005.113.001.3.

References

Attarchi, S., Gloaguen, R., 2014. Improving of above groud biomass using dual polarimetric palsar and ETM+ data in the Hyrcanian fore tainnuomts (Iran). Remote Sensing, 6, 3693–3715.

Brown, S., 1997. Estimating biomass and biomass change of tropical forests; FAO forest resources assessment publication, Roma, Italy. p. 55.

Brown, S., Schroeder, P., Kern, J., 1999. Spatial distribution of biomass in forests of the eastern USA. Forest Ecology and Management, 123; 81-90.

Corona P., Scotti R., Tarchiani N., 1998. Relation-ship between environmental factors and site in-dex in Douglas-fi r plantations in central Italy. Forest cology and Management 101: 195-207.DOI: 10.1016/S0378-1127(98)00281-3.

Tolunay, D., 2013. Türkiye'de ağaç servetinden bitkisel kütle ve karbon miktarlarının hesaplamasında kullanılabilecek katsayılar. "Ormancılıkta Sektörel Planlamanın 50.Yılı Uluslararası Sempozyumu" Bildiriler Kitabı s. 240-251 Ankara.

Günlü, A., Başkent, E.Z., Kadıoğulları, A.İ., Ercanlı, İ., 2008. Classifying oriental beech (Fagus orientalis Lipsky.) forest sites using direct, indirect and remote sensing methods: a case study from Turkey. Sensors, 8(4), 2526-2540.

Hu, H., Wang, G.G., 2008. Changes in forest biomass carbon storage in the South Caroline Piedmont between 1936 and 2005. Forest Ecology and Management 255; 1400-1408.

Kadıoğulları, A.İ., Karahalil, U., 2013. Spatiotemporal change of carbon storage in forest biomass : a case study in Köprülü Canyon National Park. Kastamonu University Journal of Forestry Faculty 13(1); 1-14.

Karahalil, U., Köse, S., Çil, B., Kılıç, B., 2013. Analyzing carbon storage changes considering spatiotemporal dynamics in Tepedüzü between 1971 and 2008. International Caucasian Forestry Symposium (ICFS), 24-26 October, Artvin, Turkey.

Keleş, S., Kadıoğulları, A. İ., Başkent, E.Z., 2012. The Effects of land use and land cover changes on carbon storage in forest timber biomass: a case study in Torul, Turkey. Journal of Land Use Science 7(2); 125-133.

Maynard, C.L., Lawrence, R.L., Nielsen, G.A., Decker, G., 2007. Modeling vegetation amount using bandwise regression and ecological site descriptions as an alternative to vegetation indices. GISci. Remote Sens., 44, 68–81.

Wannasiri, W., Masahiko, N., Kiyoshi, H., Santitamnont, P., Miphokasap, P., 2013. Extraction of mangrove biophysical parameters using Airborne Lidar. Remote Sens., 5, 1787–1808.

SPSS Institute Inc., 2004. SPSS Base 12.0 user's guide. Chicago, IL: SPSS.