BEECH TREE DENSITY ESTIMATION USING SENTINEL-2 DATA (CASE STUDY: KHYROUD FOREST)

Gh. Ronoud¹, A.A. Darvish Sefat^{*2}, P. Fatehi³

¹ Dept. of Forestry and Forest Economics, Faculty of Natural Resources, University of Tehran, Iran - ronoud.q@ut.ac.ir ² Dept. of Forestry and Forest Economics, Faculty of Natural Resources, University of Tehran - adarvish@ut.ac.ir

³ Dept. of Forestry and Forest Economics, Faculty of Natural Resources, University of Tehran - parviz.fatehi@ut.ac.ir

KEY WORDS: Tree Density, Sentinel-2 Data, Machine Learning Algorithms, Fagus Orientalis Stands

ABSTRACT:

Obtaining information about forest attributes is essential for planning, monitoring, and management of forests. Due to the time and cost consuming of Tree Density (TD) using field measurements especially in the vast and remote areas, remote sensing techniques have gained more attention in scientific community. Khyroud forest, a part of Hyrcanian forest of Iran, with a high species biodiversity and growing volume stock plays an important role in carbon storage. The aim of this study was to assess the capability of Sentinel-2 data for estimating the tree density in the Khyroud forest. 65 square sample plots with an area of $2025m^2$ were measured. In each sample plot, trees with diameter at the breast height (DBH) higher than 7.5-cm were recorded. The quality of Sentinel-2 data in terms of geometric correction and cloud effect were investigated. Different processing approaches such as vegetation indices and Tasseled Cap transformation on spectral bands in combination with an empirical approach were implemented. Also, some of biophysical variables were computed. To assess the model performance, the data were randomly divided into parts, 70% of sample plots were used for modelling and 30% for validation. The results showed that the SVR algorithm (linear kernel) with a relative RMSE of 23.09% and a R² of 0.526 gained the highest performance for tree density estimation.

1. INTRODUCTION

Forest's ecosystem is one of the most important carbon sinks of the terrestrial ecosystem. Therefore, quantifying them in the different aspects such as volume, biomass, carbon stock, tree density, and basal area is virtual. Field data measurement is a conventional method for estimating the forest structural attributes. Although, this method produces the most accurate results, it is labor intensive, expensive and time consuming. Moreover, it still lacks providing the spatially explicit of forest attributes in large area. In contrast, remote sensing data have been shown to provide a solution for the above-mentioned challenges (Vashum and ayakumar, 2012). They have advantages such as broad coverage, lower cost, and providing continues information and have been used frequently in applications such as geology, agriculture, and forestry. To the best of our knowledge, a few studies have been conducted to estimate tree density using remote sensing data over Hyrcanian and Zagros forests of Iran (Pir Bavaghar, 2011; Kalbi et al., 2013; Noorian et al., 2014).

For instance, Pir Bavaghar (2011) has investigated the capability of SPOT5-HRG to estimate tree stem density in the northern Zagros forests. Kalbi *et al.*, (2013) used linear regression model to estimate tree stem density using of ASTER data. Noorian *et al.*, (2014) shows that with implementing the Classification and Regression Trees (CART) algorithm and using ASTER image data, the tree stem density can be estimated. Recognition of available resources and accessibility to an update database are important for forest planning and management (Ahmadi Sani, 2008) in this study, we aim to investigate the relationship between the Sentinel-2 derived features and the tree density, hereafter TD, for Beech (*Fagus Orientalis*) stands in Khyroud forests.

2. MATERIALS AND METHOD

2.1. Study area and reference data

The study area with coordinates of 51° 32' to 51° 44' and 36° 27' to 36° 40' covers an area of about 500-ha. It is located in the Khyroud forest, North of Iran. The plot positions at Gorazbon and Namkhaneh district in Khyruod research forest were shown at Fig. 1.The elevation above sea level (a.s.l) ranges from 540 to 1350-m. *Fagus Orientalis* (beach) is the dominant species of this forest. In-situ data for developing tree density models and validating the models were provided through field measurements. We conducted a selective design sampling inventory and 65 sample plots were collected in August, 2014. The dimensions of plots were choose $45m\times45$ -m, with roughly corresponding to sentinel2 spatial resolution and all trees with a diameter at breast height (dbh) greater than 7.5 cm were considered for measuring TD.

^{*} Corresponding author



Figure 1. Location of the plots in Khyroud forest

2.2. Remote sensing data and processing

We used Sentinel-2 level-1C (L1C) Multispectral data acquired on 26 August 2016. The Sentinel-2 data was downloaded free of Open from Copernicus Access Hub charge (https://scihub.copernicus.eu/dhus/#/home). Sentinel's Application Platform (SNAP) (version. 6) (http://step.esa.int/main/toolboxes/snap/) (Sentinel-2_Team, 2015) were used for Sentinel-2 data processing. The image was a relatively cloud-free scene. First as a preprocessing step, 1:25000 topographic maps were used to check the geometric quality of Sentinel-2 dataset.

The quality of data in aspects of radiometry was evaluated. For this reason, we used band-9 of Sentinel-2 for assessing cloud cover. Sentinel-2 data were transferred from level-2A into level-1C using atmospheric processor of SEN2COR (http://step.esa.int/main/third-party-plugins-2/sen2cor/). In addition to main spectral bands, we used from other enhancement processing for extracting related features to forest biophysical parameters. Second as processing step, vegetation indices, biophysical variables such as Leaf Area Index (LAI), Leaf Chlorophyll Content (Cab), Canopy Water Content (CWC) and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) (Liu et al., 2019; Castillo et al., 2017; Chen et al., 2018; Frampton et al., 2013) and Tasseled Cap transformation (Nedkov, 2017) were derived as independent variables (Table 1). These mentioned biophysical indices are computed using radiation PROSAIL transformation model (For detailed information, please refer to Weiss and Baret, 2016 and Jacquemoud et al., 2009). All of desired features were resampled to 5-m resolution. The flowchart of research procedure have presented at Fig. 2.

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-4/W18, 2019 GeoSpatial Conference 2019 – Joint Conferences of SMPR and GI Research, 12–14 October 2019, Karaj, Iran

Predictor variable	Relevant	Description/ Resolution		
	band/index			
Multispectral	B2 Blue (B)	458–523nm/10m		
Bands	B3 Green (G)	543–578nm/10m		
	B4 Red (R)	650–680nm/10m		
	B5 Red edge 1 (RE1)	698–713nm/20m		
	B6 Red edge 2 (RE2)	733–748nm/20m		
	B7 Red edge 3 (RE3)	773–793nm/20m		
	B8 Near infrared (NIR)	785–900nm/10m		
	B8a Near infrared narrow (NIRn)	855–875nm/20m		
	B11 Shortwave infrared 1 (SWIR1)	1565–1655nm/20m		
	B12 Shortwave infrared 2 (SWIR2)	2100–2280nm/20m		
Biophysical	LAI	Leaf Area Index		
variables	CAB	Chlorophyll content in the leaf		
	CWC	Canopy Water Content		
	FAPAR	Fraction of Absorbed Photosynthetically		
		Active Radiation		
	FCOVER	Fraction of Vegetation Cover		
Vegetation indices	NDI45	(B5-B4)/(B5+B4)		
	PSSRA	(B7/B4)		
	S2REP	705 + 35*((((B7 + B4)/2)- B5)/(B6-B5))		
Tasseled Cap	Greeness	-		

Tabla 1	Continal 2 haged	indiana and	amininal ha	nd wood oo	mundiators f	on the domait	· actimation
rable r.	. Senunei-2 Daseo	i muices and	onginai da	nu useu as	breatciors re	or tree density	v esumation
						· · · · · · · · ·	



Figure 2. The Flowchart of research procedure for TD estimation

2.3. TD modeling and model validation

We randomly divided our in-situ data into two groups i.e. modeling (70%) and validation (30%) dataset. Two main approaches i.e. parametric and non-parametric approaches are commonly used to predict forest structural attributes from optical data. Parametric approach has a dependency to the model structure and needs a normal distribution assumption, which is particularly challenging. Therefore, the non-parametric approaches i.e. Machine learning algorithms including K-Nearest Neighbors (k-NN), Support Vector Regression (SVR), and Random Forest (RF) were used for estimating TD. The accuracy assessment of estimations was based on validation grope of reference data. The coefficient of determination R², absolute and relative root mean square error (RMSE and RMSE %) and Akaike Information Criterion (AIC) were calculated based on estimated TD from different machine learning algorithms (Equations 1-4).

$$R^{2} = 1 - \frac{\sum_{i}(yi - \hat{y}_{i})^{2}}{\sum_{i}(yi - \bar{y})^{2}}$$
(1)

$$RMSE = \sqrt{\frac{\sum_{i}(yi - \hat{y})^{2}}{n}}$$
(2)

$$RMSE\% = \frac{RMSE}{\bar{y}} \times 100$$
(3)

$$AIC = 2K - 2\ln(L)$$
(4)

Where is y_i observed value, \overline{y} average of observed values, $\hat{y_l}$ predicted value, K the number of variables used in the model, L the maximum value of likelihood function for the estimated model, and n is the number of observations.

3. RESULTS

After overlaying reference roads layer on the image of study area, it was observed that the reference roads completely corresponded to the roads observed in the image and there is no need for image rectification. Minimum, maximum and average value of measured TD were 600, 3080 and 1704 tree.ha⁻¹ respectively.

The results of k-NN have showed in Table 2 for Sentinel-2 dataset. Chebychev metric had better performance rather than Metrics and produced more accurate results with RMSE of 27.26% and R^2 of 0.121.

Table 2. The results of k-NN algorithm with different height metrics for estimating TD

Metrics	RMSE	RMSE%	(R^2)
	(tree.ha ⁻¹)		
Euclidean	444.4	27.3	0.1077
Euclidean Squared	453.2	27.84	0.073
Chebychev	443.73	27.26	0.121
Manhatan	495.34	30.42	0.037

The results of Random Forest have showed in Table 3 for Sentinel-2 dataset. As it can be seen, 7 predictor variables with 350 trees produced the best results with RMSE of 27.99% and R^2 of 0.085.

Table 3. The results of RF algorithm with different number of K predictors for estimating TD

Optimal	Number of	RMSE	RMSE%	(R^2)
number	predictor (k)	(tree.ha ⁻¹)		
of trees				
350	3	460.9	28.31	0.061
350	4	465.85	28.61	0.0414
350	5	459.36	28.21	0.07
350	6	462.66	28.42	0.056
350	7	455.73	27.99	0.085

SVR results were reported at Table 4. Linear Kernel had better performance rather than other Kernels. So, linear kernel with RMSE and R^2 of 23.09% and 0.526 showed the better results.

Table 4. The results of SVR algorithm with different kernels for estimating TD

Kernel	RMSE	RMSE	(R^2)
	(tree.ha ⁻¹)	%	
Linear	375.96	23.09	0.526
Polynomial	578.6	35.54	0.021
Radial Basis	419	25.73	0.247
Function kernel (RBF)			
Sigmoid	623.67	38.31	0.001

In general, we can see that Sentinel-2 dataset with Linear Kernel algorithm produced best results for TD estimation at our study area. The scatter plot of predicted values versus observed values was reported at Fig. 3.



Fig. 3. The predicted values versus measured values for Linear Kernel (SVR algorithm)

4. CONCLUSION

Applied machine learning approaches strengthened the hypothesis of estimating TD of Fagus Orientalis stands using Sentinel-2 images, but a need for further investigation is remain. The results showed that the SVR algorithm with linear kernel has better performance than other machine learning algorithms $(rRMSE = 23.09\% \text{ and } R^2 = 0.526)$. The same results have obtained by López-Serrano et al., (2016), Mountrakis et al (2011) and Camps-Valls (2009). One of the reasons for this result is that SVR can resolve the problems associated to low number of plots. Also, it can analyse the non-linear relationships between dependent and independent variables. There are several sources of errors and uncertainties in the estimation of TD. Understanding and quantification of these sources of errors is important in order to improve the accuracy of estimations. These sources of errors and uncertainties can be such as temporal decorrelation between remote sensing measurements and ground data and GPS positional errors. The performance of the selected method is still insufficient for accurate TD estimation and cannot be generalized to other forest areas. The error of this model can be accepted over the large scale such as watersheds. The use of lidar data and radar data is recommended for future researches.

REFERENCES

Ahmadi Sani, N. Darvishsefat, A. A. Zobeiri, M. and Farzaneh, A., 2008. Potentiality of ASTER images for forest density mapping in Zagros (Case study: Marivan forests). Journal of the Iranian Natural Res., 61(3): 603- 614.

Camps-Valls, G., Muñoz-Marí, J., Gómez-Chova, L., Richter, K. and Calpe-Maravilla, J., 2009. Biophysical parameter estimation with a semisupervised support vector machine. IEEE Geoscience and Remote Sensing Letters, 6(2): 248-252.

Castillo, J.A.A., Apan, A.A., Maraseni, T.N. and Salmo, S.G., 2017. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the

Philippines using Sentinel imagery. ISPRS Journal Photogrammetry Remote Sensing, 134: 70–85.

Chen, L., Ren, C., Zhang, B., Wang, Z. and Xi, Y., 2018. Estimation of forest above-ground biomass by geographically weighted regression and machine learning with sentinel imagery. Forests, 9(10): 582.

Frampton, W.J., Dash, J., Watmough, G. and Milton, J.M., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS Journal Photogrammetry. Remote Sensing. 82: 83–92.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C. and Ustin, S.L., 2009. PROSPECT + SAIL models: A review of use for vegetation characterization. Remote Sensing Environment, 113: S56–S66.

Kalbi, S., Fallah, A., and Shataee, Sh., 2014. Estimation of forest biophysical properties using SPOT HRG data (Case Study: Darabkola Experimental Forest), J. of Wood &Forest Science and Technology, 20(4): 117-132.

Liu, Y., Gong, W., Xing, Y., Hu, X. and Gong, J., 2019. Estimation of the forest stand mean height and aboveground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery. ISPRS journal of photogrammetry and remote sensing, 151:277-289.

López-Serrano, P. M., López-Sánchez, C. A., Álvarez-González, J. G. and García-Gutiérrez, J., 2016. A comparison of machine learning techniques applied to landsat-5 TM spectral data for

biomass estimation. Canadian Journal of Remote Sensing, 42(6): 690-705.

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

Nedkov, R., 2017. Orthogonal transformation of segmented images from the satellite Sentinel-2. Comptes rendus de l'Acad'emie bulgare des Sciences, 70(5): 687-692.

Noorian, N., Shataee, Sh., Mohammadi, J., and Yazdani, S., 2014. Estimating forest structural attributes by means of ASTER imagery and CART algorithm (Case study: Shastkolateh forest, Gorgan). Iranian Journal of Forest and Poplar Research, 22(3): 434-446.

Sentinel-2_Team, 2015. Sentinel-2 User Handbook, European Space Agency.

Pir Bavaghar, M., 2011. Evaluation of capability of IRS-P6 satellite data for predicting quantitative attributes of forests (case study: Northern Zagros forests). Iranian Journal of Forest, 3(4): 277-289.

Vashum, K. T., & Jayakumar, S. (2012). Methods to estimate above-ground biomass and carbon stock in natural forests-a review. *J. Ecosyst. Ecogr*, 2(4): 1-7.

Weiss, M. and Baret, F., 2016. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER, Version 1.1. In ESA Contract n° 4000110612/14/I-BG (p. 52). INRA Avignon, France.