RS & GIS for Natural Resources
(Vol. 11/ Issue 2) Summer 2020
Indexed by ISC, SID, Magiran, Noormags, Civilica, Google Scholar
journal homepage : www.girs.iaubushehr.ac.ir

Using the bootstrap approach for comparing statistical modeling methods to estimate remotely-sensed aboveground biomass in Zagros forests

Amir Safari, Hormoz Sohrabi

Received: 28 April 2019/ Accepted: 19 May 2020
Available online 5 July 2020

Abstract

Background and Objective Considering the increasing importance of forest ecosystems in climate change mitigation projects, reliable and cost-effective methods are required to estimate the aboveground biomass (AGB). Common methods used to estimate the aboveground biomass (AGB) include in-situ measurement, the biomass calculation using aalometric equations and using remote sensing techniques. Remote sensing has been widely used to estimate the biomass of forests in recent decades.The used statistical modeling method is one of the most important factors to use remotely-sensed data for estimation of the aboveground biomass. A large number of researches have been carried out about using the modeling methods. However, these studies face the following different challenges: 1) no modeling method has been recommended as the best method 2) the performence of these modeling methods is affected by forest type, the forest structure, and the present disturbance intensity 3) the performance evaluation and the comparion of the results of these methods were done by using goodness-of-fit test and cross-validation methods. The purpose of this study is to considering the role of choosing statistical modeling methods to estimate remotely-sensed aboveground biomass, the current study was conducted to investigate nine statistical modeling methods including linear regression (LR), generalized additive model (GAM), random forest (RF), support vector machine (SVM), boosted regression tree (BRT), k-nearest neighbor (kNN), cubist regression (CR), Gaussian process model (GPR), multivariate adaptive regression spline (MARS) using bootstrap process and 1000-repeated 10-fold cross-validation approach to estimate the aboveground biomass of Zagros forests using Landsat 8 images.

Materials and Methods The cuurent study was conducted in Kermanshah forests which is mostly dominated by oak species trees (*Quercus* spp.) and is located in western Iran on the Zagros Mountains. Zagros forests are generally sparse and open and comprise approximately 20% of Iran's area and 40% forest regions of Iran. In order to conduct this study, two forest regions with different levels of human disturbances were chosen; SarfiruzAbad region with highly degraded (HD) forests, and Gahvareh forest region with minor degradation (MD).
Geographical coordinates of SarfiruzAbad and Gahvareh regions are 33º57′-34º04′N / 47º03′-47º17′E & 34º21′- 34º24′N / 46º16′-46º23′ E respectively. The Leaf area index (LAI) map derived from the Landsat images based on a global model was used to collect field-based sample plots in both regions of the study. Both regions were divided into three low, moderate and high Leaf area index (LAI) strata, and the locations of the sample plots were located by using a systematic inventory at the intersections of a 200m×200 m grid in each stratum.

A. Safari [1], H. Sohrabi(✉) [2]

1. PhD of Forestry, Faculty of Natural Resources and Marine Sciences, University of Tarbiat Modares, Thehran, Iran
2. Associate Professor, Department of Forestry, Faculty of Natural Resources and Marine Sciences, University of Tarbiat Modares, Iran
e-mail: hsohrabi@modares.ac.ir

124 georeferenced square plots of field-based sample plots (63 plots in Gahvareh region and 61 plots in SarfiruzAbad region) with 30m×30m dimensions the same size as a Landsat 8 image's pixel were collected. Allometric equation developed for oak tree in Zagros forests was used to calculate the amount of the aboveground biomass of each individual tree or sprout-clump. The allometric equation used in this study uses two vertical tree crown diameters to estimate the amount of the biomass of each individual tree or sprout-clump. The sum of the amount of the biomass of each individual tree in sample plot was used to calculate the amount of the biomass plot in sample plot level at a ton per hectare. Our study regions were located in a frame of Landsat 8 images (path/row:167/36). A cloud-free Landsat image relating to 19th Mordad 1394 (10th August 2015) relating to the time when the tree canopies are completely closed and near to the date of land inventory was downloaded from earthexplorer.usgs.gov site. Based on the previous studies, the pre-processing of the used image comprising the radiometric and topographic corrections was done.using C method.

To estimate the aboveground biomass in the study areas by using remote sensing, 38 spectral variables including band values, simple band ratios, vegetation indices and common linear transformations like tasseled cap and principle component analysis were extracted from the used Landsat 8 image. Generally, the efficiency of nine different statistical modeling methods including parametric methods (Linear Regression, LR), semiparametric (Generalized Additive Model, GAM), and nonparametric Random Forest (RF), Support Vector Machine (SVM), K-nearest neighbor (KNN), Boosted regression trees (BRT), multivariate additive regression splines, cubist regression (CR), and Gaussian processes regression/model) were compared in order to estimate aboveground biomass. To assess the models, two common quality statistics: (i) determination coefficient and (2) root mean square error via 10 fold cross validation repeated 1000 times approach were calculated. This number of repeats helps to ensure an acceptable assessment of robustness of the results.

Results and Discussion The measured statistical characteristics of the field sample plots showed that the mean aboveground biomass of SarfiruzAbad and Gahvareh regions were 12.6 ton/ha and 20.5 ton/ha respectively. ANOVA indicated significant differences between modelling methods (treatment effect: $p< 0.001$) for both R^2 and RMSPE calculated in 1000-time repeats using 10-fold cross- validation.

The Cubist modeling method with the mean determination coefficient of 0.61 outperformed other methods in SarfiruzAbad region. These results for Gahvareh region showed better efficiency of linear regression (LR), generalized additive model (GAM), and k-nearest neighbor (KNN) with the mean determination coeffieient of 0.87.

The multiple comparisons of different models by using Tukey test concerning RMSE showed that in SarfiruzAbad region, cubist method with the mean of RMSE 3.3 ton/ha and KNN and RF methods with the mean of RMSE 5.8 ton/ha had a significant difference in comparison to the other methods. Totally, the results of the research revealed the suitable efficiency of Landsat 8 image for AGB estimation in Zagros forests. The acceptable results are due to the low AGB in our study regions that did not reached the saturation point as one of challenges of using optical images like Landsat.

The other results of this research is the assessment of the effiecieny of modeling method in order to increase the accuracy of the estimation of remotely-sensed aboveground biomass.Unlike the results of the previous studies, linear regression yielded better results compared to nonparametric methods that can be due to the presence of the linear relationship between aboveground biomass and spectral variables derived from Landsat images. Among the used various spectral variables, red, near infrared, and shortwave infrared 1 and 2 band ratios were selected as the final variable in most modeling methods.

Conclusion In this study, we evaluated the effieincy of different statistical modeling methods to estimate AGB in Zagros forests by using Landsat images. The biomass estimations were compared by using nine parametric, semi-parametric, and non-parametric methods and using 1000-repeated 10-fold cross-validation. The results illustrated the acceptable potentiality of Landsat images for cost-efficient AGB estimating in Zagros oak forests. The accuracy of AGB estimation in Gahvareh region with low-degraded forest stands was higher than SarfiruzAbad region with highly degraded stands.

Keywords Aboveground biomass, Bootstrap, Statistical modeling, Zagros forest, Landsat 8