Computers and Electronics in Agriculture 151 (2018) 431–440

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Tree-bark volume prediction via machine learning: A case study based on black alder's tree-bark production

Maria J. Diamantopoulou[a,*], Ramazan Özçelik[b], Hakkı Yavuz[c]

[a] Faculty of Forestry and Natural Environment, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece
[b] Faculty of Forestry, Süleyman Demirel University, East Campus, 32260 Isparta, Turkey
[c] Faculty of Forestry, Karadeniz Technical University, 61080 Trabzon, Turkey

ARTICLE INFO

Keywords:
Support vector regression
Non-linear regression
Forest resources management

ABSTRACT

Tree-bark volume estimation is a multi-faceted problem and at the same time of vital importance in the area of forest resources management. This importance relies on the fact that it constitutes a key variable for accurately assessing timber quantities, while at the same time its use has been spread as a soil-covering product or as a soil fertilizer or as a substitutional medicinal product. Consequently, due to its substantial economic impact, the accurate prediction of the tree-bark volume is of utmost importance. In this study, we propose three bark volume prediction models for black alder trees (*Alnus glutinosa* (L.) Gaertn subsp. *barbata* (C.A. Mey.) Yalt.) each targeting a different creation source of the black alder forest. Hence, we used data from naturally regenerated, plantation and coppice stand types. 1334 stem analysis data were collected for three different stand types. Two different modeling techniques were used, the weighted nonlinear regression and the ε-support vector regression techniques. These two modeling approaches were selected due to the fact that the need to handle regression analysis problems (noise in the data, high variability and/or non-normal distributions) is essential. The state-of-the-art approach suggests the usage of machine learning techniques in an effort to build reliable and robust models able to deal with complex environmental problems. An overall illustration of the precision obtained by the constructed models was conducted by statistical criteria such as the root mean square error, the correlation coefficient, the Furnival's index of fit and the Akaike's information criterion. Although the estimation and prediction errors of the two different modeling techniques seem to be close in pure numbers, the ε-support vector regression models gave the most accurate results for all stand types as compared to the nonlinear regression. Based on the results obtained from this study, the constructed ε-support vector regression models for modeling tree-bark volume showed a great ability to generalize, and thus worth considering as an alternative to regression modeling that enables increasing our ability for successful forest management.

1. Introduction

Black alder (*Alnus glutinosa* (L.) Gaertn subsp. *barbata* (C.A. Mey.) Yalt.) forests are naturally widespread across all Europe from mid-Scandinavia to the Mediterranean countries including northern Turkey. Despite the fact that most of the growing sites have been converted to agriculture, these forests are typically a subdomain of the mixed broadleaved forests (Turok et al., 1996). These forests cover an area of about 146.730 ha and approximately 1% of forest cover in Turkey (GDF, 2015). Black alder trees exhibit a widespread use in several domains such as energy and paper production which originates from the aesthetic and mechanical properties of their wood. Furthermore, they appear to have a high potential for timber production, a prerequisite of which is their growing in suitable ecological conditions. Under these conditions, the species grow in a similar way to ash, maple or cherry (Claessens et al., 2010). In this context, black alder forests play a key role by providing important benefits along with environmental services to riverine ecosystems in the Northern Turkey. In addition, they contribute to biodiversity by providing habitats for a specific flora and fauna both on the tree itself and in the flooded root system (Dussart, 1999). At the same time, these forests assist with water filtration and purification in waterlogged soils. Their root system also helps to control floods and stabilize riverbanks (Piegay et al., 2003). Finally, black alder is important in alluvial and marshy ecosystems for nature conservation and watershed management (Claessens et al., 2010).

Recently, Turkey has adopted the approach of a multipurpose and ecologically-based forest management which primarily aims at maintaining biological diversity, productivity, regenerative capacity, and

* Corresponding author.
E-mail address: mardi.diamant@gmail.com (M.J. Diamantopoulou).

https://doi.org/10.1016/j.compag.2018.06.039
Received 23 September 2017; Received in revised form 10 June 2018; Accepted 13 June 2018
0168-1699/

vitality and along with fulfilling relevant ecological, economical, and social functions (Zengin et al., 2013). Therefore, the General Directorate of Forests (GDF, 2015) is in need for evaluating forest growth and yield both qualitatively and quantitatively through the development of prediction models be relevant to forests raised from specific creative procedures. As a result, the need to provide more accurate and versatile information regarding important characteristics of forest restand types arises. However, the already available research outcomes regarding the prediction of growth and yield in black alder trees are very limited.

Tree volume estimation is an integral part of forest management and planning (de-Miguel et al., 2012; Rodriguez et al., 2015), due to several reasons. At first, it enables the projection of forest products' industry (Fang et al., 2000; de-Miguel et al., 2012), and facilitates the monitoring of forest health and productivity, while at the same time can be used for the estimation of biomass and carbon stocks (Castedo-Dorado et al., 2012; Gomez-Garcia et al., 2015).

In forest management, the estimation of tree volume is usually based on different volume functions including bark. Tree bark can comprise a large amount of the total tree volume. According to FAO (2009) forest trees provide an industrial output of about 363 million m^3 with the volume of bark accounting for an estimate of 10% of total tree volume. According to Health et al. (2009), the amount of tree bark varies by species and ranges from 12% to 20% of the total tree volume. The substantial economic impact of bark volume which is the most important non-woody product, not only for itself but also for the timber production, constitutes accurate bark volume prediction of utmost importance. As indicated by Marshall et al. (2006), the use of an inaccurate bark volume estimation can cost the landowner approximately 11% of his timber value.

Bark volume varies not only by species but also by rate of growth, environmental conditions (Laasasenaho et al., 2005) and the genetic constitution of trees. In many countries, those differences that originate from the different genetic, site and growth conditions are not taken into consideration in the models applied towards estimating bark thickness and volume (Stangle et al., 2015). Until now, timber volume of trees in Turkey has been calculated including bark, despite the fact that the bark may represent large amount of total tree volume. As a result, the lack of bark volume models makes it impossible to accurately estimate/predict bark volume and biomass, which negatively affects forest industry sector.

Although several studies in the field of forest management estimate bark thickness, the research efforts in the direction of estimating bark volume is limited. At first, pine-bark volume estimation has been conducted by Diamantopoulou (2005) through non-linear regression models and cascade-correlation artificial neural network models. Another approach by Laasasenaho et al. (2005) involves fitting taper curves with cubic spline functions, while Malone and Liang (2009), and Pompa-Garcia et al. (2012) employ linear regression models. On the other hand, regression models using mixed-effects modeling approach were presented by Li and Weiskittel (2011), while linear and nonlinear regression models were used by Wehenkel et al. (2012). A different approach to bark volume prediction that differs from the aforementioned paradigms is the one proposed by Valipour et al. (2009) which involves the evaluation of the interrelationship of bark thickness and geographic variables, while Sönmez et al. (2007) explores the effects of aspect, tree age, and tree diameter on bark thickness for oriental spruce (*Picea orientalis*).

Most of the research efforts in the area of tree-bark volume prediction propose the use of regression models. However, while dealing with real data, which is a typical paradigm in environmental modeling, the need to handle problems such as noise in the data, high variability and/or non-normal distributions is essential. These problems originate from the known and necessary assumptions made by the regression analysis (Draper and Smith, 1998; Archontoulis and Miguez, 2013). Fundamental assumption of least squares regression is that errors are independent and normally distributed with a mean of zero and constant variance. However, In many forest modeling situations, there is a common pattern whereby the error variation increases as the values of the dependent variable increase (Parresol, 1993).

In an effort to find alternatives, the state-of-the-art approach suggests the usage of machine learning (ML) techniques in an effort to build not only reliable, but also robust models able to deal with complex environmental problems. ML algorithms iteratively learn from data and are able to discover and model hidden data patterns without the need for explicit programming effort (Swingler, 2001). On the contrary, regression analysis is based on coefficients estimation of an equation that has to be specified first and thus requires the modeler to fully understand the role of each variable that attempts to describe the specific natural problem in hand.

An investigated and reliable branch of machine learning in forest modeling is the Artificial Neural Network (ANN) methodology. Since nowadays, ANNs' applicability in several domains such as software engineering (Schmidhuber, 2015; Papamichail et al., 2016; Dimaridou et al., 2017) or forest science (Diamantopoulou, 2005; Diamantopoulou and Milios, 2010; Özçelik et al., 2010; Tiruveedhula et al., 2010; Özçelik et al., 2013; Vahedi, 2016) has been explored by many researchers. The main concluding remark of the aforemantion research is that ANN methodology can be a useful tool in practice. While the use of a single hidden layer ANNs has been proved to be a powerful tool in forest modeling (Özçelik et al., 2010; Leite et al., 2011; Satir et al., 2016), they are still confined with the need for properly define the architecture and calculate/evaluate parameters such as the initial weights or the stopping criteria. Non-optimal settings can lead to convergence on local minima or overfitting by considering the noise as part of the estimation pattern (Navarro and Bennun, 2014).

Another promising alternative that worths exploration in forest modeling is the Support Vector Machine (SVM) methodology that seems to represent an intuitive and at the same time powerful technique for outlier detection, classification and regression type problems. Reliable predictions based on complex data sets, typical in real world problems, require models able to recognize all patterns including the subtle ones. This predictive ability seems to be supported by the SVM learning theory, as this theory has originally been enriched with nonlinear kernel functions (Boser et al., 1992) for solving mainly classification problems. A branch of SVM s is the Support Vector Regression (SVR) methodology (Vapnik et al., 1997; Basak et al., 2007; García-Nieto et al., 2012). Vapnik et al. (1997) introduced the ε-Support Vector Regression (ε-SVR) algorithm theory in which the model depends on these training data points that lie within the margin determined by the Support Vectors (SV). Relying solely on kernel functions; ε-SVR is considered as a nonparametric algorithm in contrast with traditional regression. The ε-SVR theory has been extensively described in Vapnik (1995, 1998, 1999, 2000), Smola and Schölkopf (1998), Basak et al. (2007) and García-Nieto et al. (2012). This algorithm deals with solving regression type problems and has been found able to overcome both overfitting and local minima problems by minimizing the bound of generalization error (instead of the training error). Due to these advantages ε-SVR has been considered as a reliable alternative algorithm against the widely-accepted non-linear regression approach for building our tree-bark volume prediction models.

During the last decade, research efforts focused on examining the effectiveness of the SVR models in the field of environmental modeling (Wang et al., 2009, 2017; Alonso et al., 2013; Liu et al., 2013; Gu et al., 2016; Thomas et al., 2017). Specifically, according to the forest modeling, Monnet et al. (2011) demonstrated the SVR models' ability to adequately address the prediction of forest stand parameters (dominant height, basal area, mean diameter, and stem density). Youquan et al. (2012) calculated the live tree timber volume using support vector regression, while Binoti et al. (2016) used the support vector machines methodology for the Eucalyptus tress volume prediction. Except for the above, while SVR models' ability has been explored in diameter at breast height and total tree volume prediction (Wu et al., 2015) as well,