Forest Ecology and Management 419–420 (2018) 240–248

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Predicting tree height from tree diameter and dominant height using mixed-effects and quantile regression models for two species in Turkey

Ramazan Özçelik[a,*], Quang V. Cao[b], Guillermo Trincado[c], Nilsun Göçer[d]

[a] Faculty of Forestry, Süleyman Demirel University, East Campus, 32260 Isparta, Turkey
[b] School of Renewable Resources, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA
[c] Instituto de Bosques y Sociedad, Facultad de Ciencias Forestales y Recursos Naturales, Universidad Autral de Chile, Box 567, Valdivia, Chile
[d] Graduate School of Natural and Applied Sciences, Süleyman Demirel University, East Campus, 32260 Isparta, Turkey

ARTICLE INFO

Keywords:
Chapman-Richards
Calibration
Brutian pine
Taurus cedar
Sampling effort

ABSTRACT

Height-diameter models were developed for Brutian pine (*Pinus brutia* Ten.) and Taurus cedar (*Cedrus libani* A. Rich.) in Turkey. A modified Chapman-Richards model that includes dominant height was used to predict tree height from diameter. Using the twofold evaluation scheme, five alternative modeling approaches were evaluated: (1) fixed-effects model, (2) calibrated fixed-effects model, (3) calibrated mixed-effects model, (4) three-quantile regression method, and (5) five-quantile regression method. Parameters of fixed-effects, mixed-effects and quantile regression models were calibrated by use of a subset of height measurements, ranging from 1 to 10 sample trees per plot. Evaluation statistics show that both quantile regression models produced similar results, and that the mixed-effects model approach yielded the best results in predicting tree heights. Model performance improved with increasing sample size; but gains in performance generally increased at a decreasing rate. A sample size of four trees per plot appears to be a good compromise between sampling cost and predictive accuracy and precision.

1. Introduction

Brutian pine (*Pinus brutia* Ten.) and Taurus cedar (*Cedrus libani* A. Rich.) are two of the economically and ecologically most important tree species in Turkey and presently found primarily in the Mediterranean Region of Turkey. The forests of brutian pine and Taurus cedar are predominantly composed of mature stands and provide important global and national benefits related to carbon storage and biodiversity, along with the other ecosystem services in Turkey. Brutian pine forms extensive forests, especially in regions where the Mediterranean climate prevails (Boydak, 2004). This tree species grows in pure stands and is commonly found in fire-related ecosystems. Brutian pine grows on many soil types, but primarily on rendzina soils on soft limestones and marl deposits (Atalay et al., 1998). Taurus cedar forests are located primarily in the Taurus Mountains in the Mediterranean Region of Turkey at elevations between 800 m and 2100 m on calcareous formations. However, this species can also be found at lower and higher elevation as small populations.

In recent years, Turkey has adopted the approach of multipurpose and ecologically based forest management. The General Directorate of Forests (GDF) therefore needs to develop and evaluate growth and yield prediction models for sustainable management of forest resources. Many growth and yield models require total tree height (h) and breast height diameter (d) as basic input variables (Temesgen et al., 2007), therefore, the height-diameter (h–d) models are considered one of the most important components of growth and yield models. *H-d* models are very useful for yield estimation (Curtis, 1967; Parresol, 1992), site index and dominant height estimation (Curtis, 1967; Calama and Montero, 2004), stand structural analysis (Spies and Cohen, 1992; Morrison et al., 2012; Von Gadow et al., 2001), damage appraisal and stand stability (Parresol, 1992; Vospernik et al., 2010), stand growth dynamics (Curtis, 1967; Burkhart and Strub, 1974; Wykoff et al., 1982), individual tree and stand volume prediction (Peng, 2001a; Gómez-García et al., 2014), and product recovery and carbon budgeting models (Newton and Amponsah, 2007). Furthermore, h–d models are needed to better understand the nature of various relationships that characterize, differentiate, and influence the development of forest ecosystems (Peng, 2001b). However, so far, the available information about *h-d* relationships concerning these tree species is very limited.

The cost of measuring tree height in forest inventory is greater than that of measuring tree diameter, leading to the need for equations to predict tree height from measured diameter at breast height. Most h–d

* Corresponding author.
E-mail address: ramazanozcelik@sdu.edu.tr (R. Özçelik).

https://doi.org/10.1016/j.foreco.2018.03.051
Received 22 February 2018; Received in revised form 14 March 2018; Accepted 26 March 2018
0378-1127/

models have been applied to pure even-aged stands or plantations (e.g. Soares and Tomé, 2002; Sánchez et al., 2003; Diéguez-Aranda et al., 2005). Despite the homogeneous characteristics of these types of forest, a single *h–d* model is not usually adequate for all possible situations, since the *h–d* relationship varies from stand to stand, and is not even constant within the same stand over time (Curtis, 1967). Some *h-d* models also have been developed to mixed, uneven-aged stands in which different species, age, structure, and levels of competition (Temesgen and Gadow, 2004; Corral-Rivas et al., 2014; Crecente-Campo et al., 2014).

The most widely used method for minimizing this level of variance is to fit a local *h–d* equation for each plot and measurement occasion. The main problem with this approach is that it requires a high sampling effort. Recently, several studies have shown that the use of plot-level information in addition to dbh could provide more accurate prediction of height than using only dbh. Stand variables that have been added to the base model include stand-density measures (Temesgen and Gadow, 2004), stand density, dominant height, and diameter distribution percentiles (Calama and Montero, 2004; Dorado et al., 2006; Temesgen et al., 2008), stand density and relative position (Temesgen et al., 2007), stand density, species composition, and top height (Huang et al., 2009), relative spacing index and quadratic mean diameter (Ducey, 2009; Saud et al., 2016). The inclusion of these variables improved the precision of height estimates. In contrast, Sharma and Zhang (2004) and Huang et al. (2009) found that site index and species composition did not improve the precision of height prediction in their studies, respectively. Huang et al. (2009) found that top height is the most significant contributor among different stand-level variables. Moreover, dominant height is usually measured for use in site index equations or other submodels of growth simulators (Gómez-García et al., 2014).

H-d relationships are often modelled by use of ordinary least squares (*OLS*) regression technique. However, this method is not reliable because the assumption of random and independent observations is often violated and the presence of autocorrelation is not accounted for. These problems can be addressed with nonlinear mixed-effects models (NMEM), which allow for both *population-averaged* and *subject-specific* models. The first considers only fixed-parameters, common to the population, while the second considers both fixed- and random-effects parameters, common to each subject. The inclusion of random parameters, specific for each plot, allows for modeling the variability of the *h-d* relationship among different locations, after defining a common fixed functional structure (Lindstrom and Bates, 1990). If prediction for a new stand is required and prior information from a small sample of trees measured for *h* and *d* is available, the *h-d* curve can be calibrated for that particular stand. Many studies have used mixed-models to describe *h-d* relationship (e.g., Lappi, 1997; Calama and Montero, 2004; Trincado et al., 2007; Sharma and Parton, 2007; Huang et al., 2009; Crecente-Campo et al., 2010; Özçelik et al., 2013; VanderSchaaf, 2014; Gómez-García et al., 2015; Saud et al., 2016; Zang et al., 2016). Temesgen et al. (2008) used a correction factor to calibrate a nonlinear fixed-effects model to local conditions, and found that generally, although the calibrated mixed-effects model performed better than the calibrated fixed-effects model, the differences were greatly dampened when additional tree- and stand-level factors were included.

Some authors have also tested the performance of NMEM for predicting tree height when a subsample of tree heights from a new stand is available. However, there is no unified agreement on what should be the sample size to localize *h-d* curves. Calama and Montero (2004) recommended using four tree heights for calibration and Trincado et al. (2007) observed that increases in sample sizes from one tree to two and three trees provided successively smaller gains. Similar findings have been reported by Temesgen et al. (2008) and Huang et al. (2009) for sample sizes ranging from 1 to 15 trees and from 1 to 9 trees, respectively.

Quantile regression, introduced by Koenker and Bassett (1978), has been gaining popularity in forestry research in recent years. Quantile regression is a method of estimating the complete conditional distribution of dependent variables and assessing the effects of predictors at different quantiles, whereas the mean regression estimator addresses only the conditional mean or the central effects of the covariates. As a result, quantile regression is a flexible method to depict varying patterns of the relationship between *y* and *x*. Quantile regression has been used in research dealing with error assessment in forest inventory (Mäkinen et al., 2008), self-thinning boundary lines (Zhang et al., 2005; Ducey and Knapp, 2010), spreading rates of insects (Evans and Gregoire, 2007) and diseases (Evans and Finkral, 2010), diameter percentiles (Mehtätalo et al., 2008), diameter growth (Bohora and Cao, 2014), and tree taper (Cao and Wang, 2015).

Zang et al. (2016) used quantile regression to describe various patterns of the *h-d* relationship, but did not show how to use curves from different quantiles to predict tree height from diameter. Calibration of quantile regression models has so far been based on only one observation, either dbh measurement at a certain age (Bohora and Cao, 2014) or bole diameter measurement at a relative tree height (Cao and Wang, 2015). New procedures were therefore necessary to deal with calibration data consisting of multiple observations, i.e., more than one sampled height per plot.

The goals of this study were to (a) evaluate the predictive capability of *h-d* models developed using fixed-effects, mixed-effects, and quantile regression, (b) evaluate the impact of including stand-level variables into the model, (c) compare different approaches for calibrating *h-d* curves to local conditions, and (d) determine the required sample size of trees for calibration.

2. Material and methods

2.1. Data

The Mediterranean region starts from Gelibolu peninsula, follows the Aegean and Mediterranean coasts and ends up in the Amanos Mountains. While the taxa of Brutian pine, Black pine, Taurus cedar and Cilician fir constitute the coniferous forests in this region; juniper and oaks species also accompany them. Data for this study were obtained from natural, even-aged, and pure stands (more than 90% of trees) throughout the area of distribution of these species in Mediterranean Region of Turkey. A total of 58 temporary sample plots for Brutian pine and 88 temporary sample plots for Taurus cedar were measured. The distribution of sample plots represent the existing range of ages, stand densities and sites with altitude ranging from 50 to 1025 m for Brutian pine and 1210 to 1750 m for Taurus cedar. Plot size ranged from 120 to 3000 m^2, depending on stand density, in order to achieve a minimum of 30 trees per plot. For each tree, two perpendicular diameters (outside-bark 1.3 m above ground level) were measured to the nearest 0.1 cm and were then averaged to obtain diameter at breast height (*d*, cm). In each plot, approximately one-third of the trees were selected to ensure a representative distribution by diameter and height classes. Total heights of these trees were measured to the nearest 0.5 m with a Blume-Leiss hypsometer. The dominant height was defined as the average height of the 100 largest-diameter trees per hectare, depending on plot size.

The data were randomly divided into two groups; each contained the same number of plots (Table 1). We used the two-fold evaluation scheme (Bohora and Cao, 2014), in which parameters of the *h-d* model fitted to one group was applied to predict for the other group. The predictions from both groups were then used to compute evaluation statistics for different methods.

2.2. Nonlinear fixed-effects model

A large number of nonlinear model forms were evaluated for both tree species; including those reported by Huang et al. (2000) and Peng (2001b). The nonlinear models were fitted by use of *OLS* to the whole