

Comparative analysis of taper models for *Pinus nigra* Arn. using terrestrial laser scanner acquired data

Issam Boukhris⁽¹⁻²⁾,
Nicola Puletti⁽³⁾,
Christian Vonderach⁽⁴⁻⁵⁾,
Matteo Guasti⁽³⁾,
Said Lahssini⁽⁶⁾,
Monia Santini⁽²⁾,
Riccardo Valentini⁽¹⁻²⁾

Taper equations are indispensable tools for characterizing the stem profile of trees, providing valuable insights for forest management, timber inventory, and optimal assortments allocation. The recent progress in Terrestrial Laser Scanning (TLS) has revolutionized forest inventory practices by enabling non-destructive data collection. In this study, four taper models from three different model categories were established based on point cloud data of 219 *Pinus nigra* trees. The taper equations fitted with TLS data were used to predict the diameter at specific stem heights and the total stem volume. The results show that among fitted models, the Max and Burkhart segmented model calibrated by the means of a mixed-effects approach provided the best estimate of the diameter at different heights and the total stem volume evaluated for different diameter at breast height (DBH) classes. In numerical terms, this model estimated the diameter and the volume with a respective overall error of 0.781 cm and 0.021 m³. The predicted profile also shows that above a relative height of 0.7, the diameter error tends to increase due to the low reliability of data collected beyond the base of the crown primarily caused by interference from branches and leaves. Nevertheless, this study shows that TLS technology presents a compelling opportunity and a promising non-destructive alternative for generating taper profiles and estimating tree volume.

Keywords: Taper Equations, Volume Equations, Forest Mensuration, Forest Assessment, Environmental Management, Max and Burkhart, B-Splines, Random Forest, TLS

Introduction

Precision forestry is becoming increasingly important in the face of the urgent challenges posed by climate change and the rising call for sustainable forest management (SFM) practices (Kováčsová & Antalová 2010). This emerging trend depends on leveraging precise data obtained using advanced technologies to guide well-informed decision-making (Fardusi et al. 2017). In the realm of SFM, a noteworthy potential exists for value creation through implementing improved practices. Besides the ecological advantages related to en-

hanced productivity, which contribute to carbon sequestration and alleviate pressure on forests, the adoption of this paradigm shift is fundamentally linked with substantial economic and social value (Choudhry & O’Kelly 2018).

One concrete and increasingly prevalent instance of the potential being realized is precision harvesting (PH), which aims to maximize the efficiency of the harvesting operations while minimizing the impact on the environment and overcoming several limitations inherent to conventional logging practices (e.g., lack of precision and in-

efficient resource utilization due to lack of tree-level data – Olivera Farias & Visser 2016). Within the scope of PH, the optimal allocation of assortments is a crucial component of the wood products supply chain and the carbon stock projections since different wood products hold different economical values and carbon storage potential (Brunet-Navarro et al. 2017, Marchi et al. 2020). Stem taper equations, which predict the change in stem form from ground to tip can accurately disaggregate trees into specific products based on certain specifications like log length and diameter and can help reach a better optimization of wood products (Calders et al. 2015, 2022, Puletti et al. 2019).

In the academic literature, a panoply of taper model formulations and methods of parameter estimation is currently in use and the selection of a proper model form is more important than the actual fitting method (Weiskittel et al. 2011). Following the classification of McTague & Weiskittel (2020), taper models could be grouped into three main categories: (i) parametric models (e.g., simple-taper, variable-exponent, and segmented equations) which include most taper equations, are calibrated by the means of parametric approaches (e.g., nonlinear least squares – NLS – and nonlinear mixed effects – NLME – methods) and can ensure biologically consistent behavior (Mäkelä & Valentine 2020). The

□ (1) Department for Innovation in Biological, Agri-Food and Forest Systems (DIBAF), University of Tuscia, 01100 Viterbo (Italy); (2) CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy, 01100 Viterbo (Italy); (3) CREA, Research Centre for Forestry and Wood, v.le Santa Margherita 80, IT-52100 Arezzo (Italy); (4) Forest Research Institute Baden-Württemberg, 79100, Freiburg (Germany); (5) Chair of Forest Growth and Dendroecology, University of Freiburg, Tennenbacherstraße 4, 79106 Freiburg (Germany); (6) Department of Forest Development, National School of Forest Engineers, 11000 Salé (Morocco)

@ Issam Boukhris (issam.boukhris@unitus.it)

Received: Nov 17, 2023 - Accepted: Jun 25, 2024

Citation: Boukhris I, Puletti N, Vonderach C, Guasti M, Lahssini S, Santini M, Valentini R (2024). Comparative analysis of taper models for *Pinus nigra* Arn. using terrestrial laser scanner acquired data. *iForest* 17: 203-212. - doi: [10.3832/ifor4525-017](https://doi.org/10.3832/ifor4525-017) [online 2024-07-22]

Communicated by: Angelo Nolè

Max and Burkhart model stands as an instance of this class, particularly within the segmented equations category, and has been demonstrated to exhibit high accuracy in predicting stem taper (Gordon 1983, Kozak & Smith 1993, Max & Burkhart 1976). (ii) Semiparametric models (e.g., B-splines and P-splines) which offer greater flexibility in the fit without requiring an extensive addition of parameters, may provide a better representation for complex form species (Kuzelka & Marušák 2014). The B-spline regression model is a particular instance of semi-parametric model that was, first introduced by Kublin et al. (2013), and has been successfully used to predict stem taper (Erber et al. 2022, Hansen et al. 2023). (iii) Nonparametric models (e.g., Random Forest – RF – and Artificial Neural Network – ANN) can offer a strong predictive performance without requiring the testing of statistical hypotheses (Baker et al. 2018). In particular, RF, an ensemble learning method in machine learning, has been gaining popularity in modeling taper by providing precise predictions and demonstrating generalizable performance (Nunes & Görrens 2016, Salekin et al. 2021).

Developing taper equations requires multiple observations on upper-stem diameters and other covariates collected on standing or felled trees (Gómez-García et al. 2013). Precision forestry in general offers substantial opportunities that may be beneficial for the forest sector including the development of decision tools such as taper equations. The advancements in Terrestrial Laser Scanning (TLS) technology present an unprecedented opportunity for acquiring essential data with high precision and non-destructive methods, thereby significantly reducing the time required compared to conventional approaches (Calders et al. 2015, Pitkänen et al. 2019). Data acquisition for TLS-based forest missions can be performed following single or multi-scan approaches. The single-scan approach has the simplest data acquisition setting and the fastest speed, however, a major problem with that scan method is that only parts of the trees are covered due to occlusion effects by other objects. On the other hand, the multi-scan approach appears the most accurate technique and has the potential to fully cover trees within the sampling area, since scans are performed from multiple directions to overcome the occlusion issue (Liang et al. 2016). Many studies using the multi-scan mode reported a stem detection rate between 62.1% and 100% depending on the forest structure and the scanning setup (Liang & Hyyppä 2013, Maas et al. 2008).

Pinus nigra Arn. is a fast-growing conifer with a wide though fragmented distribution across Europe and Asia Minor, predominantly in mountainous areas. Due to its ecological flexibility, it is one of the most widely used tree species for reforestation worldwide, and it is considered a potential substitute for native coniferous species in Central Europe under future climate change (Thiel et al. 2012). Economically, it is one of the most important conifers in Southern Europe, as its wood is highly suitable for general construction, indoor flooring, furniture industry, fuelwood, and paper pulp (Enescu et al. 2016). In Italy, *Pinus nigra* Arn. stands cover an area of 444,785 hectares, representing 4.25% of the total national forest area, with age ranging from 50 to 95 years (Marchi et al. 2020). The previously developed yield equations for *Pinus nigra* Arn. were based on data collected from young trees and have been shown to underestimate the volume of mature trees in the current conditions (Bernetti et al. 1969). Therefore,

the primary aim of this study was to compare the performance of four taper models from tree class categories in predicting diameter over bark (DOB) and total stem volume of *Pinus nigra* Arn. specifically using data collected from multi-scan approach with TLS. The developed tool will provide forest managers with more accurate estimates of volume and enable better optimization of assortments for industrial purposes.

Materials and methods

Study area and sampling protocol

The study was conducted in the Vallombrosa forest (43° 44' N, 11° 34' E), a biogenetic reserve located about 50 km east-southeast of Florence, Tuscany, Italy. The native vegetation of this forest is mainly represented by beech (*Fagus sylvatica*) at higher elevations, oak-hornbeam stands (*Quercus* spp. mixed with *Carpinus betulus* L. and *Ostrya carpinifolia* Scop.), and chestnut (*Castanea sativa* Mill.) at lower altitudes (Dálya et al. 2019). As part of the main reforestation program that occurred after the World War II, the European black pine (*Pinus nigra* Arn.) was introduced in the extreme western side of this forest (Bottalico et al. 2012). Covering 11% of the total forest area, *Pinus nigra* Arn. stands were specifically studied here.

The *Pinus nigra* Arn. area was stratified as part of the sampling scheme based on two criteria: stand density and total tree height. Each criterion was divided into three levels to ensure optimal representation of forest stand variability, resulting in a total of nine strata. An airborne lidar flight was utilized to achieve this stratification, providing high-resolution data on forest structure and height. The ALS survey was carried out under leaf-on condition using a Eurocopter AS350 B3 equipped with a LiDAR RIEGL LMS-Q680i sensor. The flight height was 1000 m above ground level. Full-waveform LiDAR data were registered and discretized to a point density of 10 points m². Standard procedures for pre-processing ALS data (e.g., outliers and noise removal, classification of ground/non-ground, and computation of height on the ground) were carried out with the LAsTools software (Isenburg 2017) to obtain ALS normalized point clouds. Subsequently, one scan center was randomly placed in each stratum; then, a multi-scan TLS mission was performed on each scan center to collect detailed information on forest characteristics (Fig. 1).

TLS data collection and processing

TLS data were acquired using a FARO Focus 3D × 130 (FARO Technologies Inc., Lake Mary, FL, USA). The instrument uses a phase-shift-based technology with a maximum range of 120 m and acquires data with an azimuth scan angle of 360°. It collects the x, y, and z coordinates, and the intensity of the laser returns with a scanning noise of ± 1 mm.

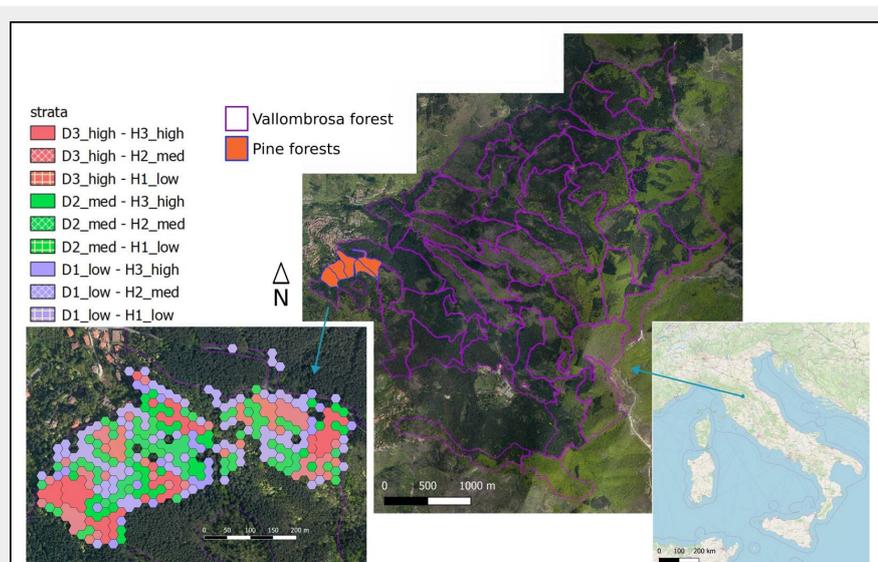
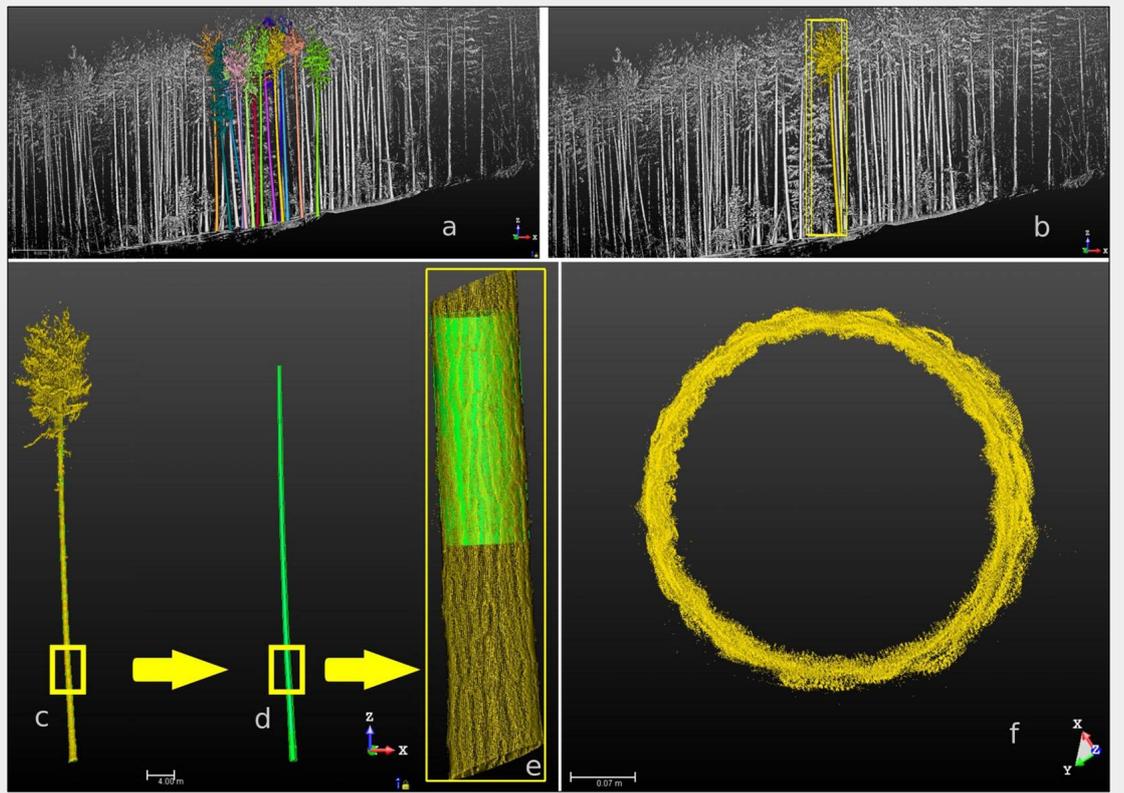


Fig. 1 - Map of the geographical situation of the Vallombrosa forest and the location of *Pinus nigra* Arn. stands in this forest (Orange color). Stratification of those stands by density (D1-D3) and canopy height (H1-H3; down left sub-figure).

Fig. 2 - Schematic illustration of the point cloud data exploration. (a): The resulting point cloud after the pre-processing steps from a small region within one stratification unit; (b): identification of a representative tree within the point cloud; (c): zoomed-in view of the stem of the identified tree; (d): removal of branches for stem cleaning; (e): fitting a cylinder to a specific section of the tree; (f): cross-sectional view of the point cloud for the section depicted in panel (e).



Using the TLS device, the data acquisition was set to a reasonable scan configuration providing a good trade-off between a sufficient density and the required time for a single scan (resolution of 7.6 mm at 10 m and 1/5 4× overall quality), for a total of about 28 million pulses per scan. With the aim of reducing occlusion due to other obstacles or vegetation and ensuring sufficient coverage, scanning was performed in multiple positions, at least 8 scans per plot subjectively distributed based on the density and structure. Each of the trees in the stand was at least scanned from 3 positions and up to 12 white polystyrene registration spheres (14 cm diameter) were placed in each plot to aid in the digital registration of individual scans.

Individual scans were merged on the plot level using the automatic registration algorithm implemented in Trimble Real Works® (TRW) software. The software automatically joins overlapping redundant points to create one seamless 3D point cloud suited for the analysis. Details of the operation (settings, criteria, and thresholds) performed by the software are not declared nor accessible. The process attained a very low plot-level registration error, and TRW achieved a high precision scan placement (mean = 2.33 mm, sd = 3.18 mm, max = 8.21 mm). Using the TRW tool, each pine tree stem was manually separated from the rest (i.e., from the ground, the crown, and other non-pine points). Diameter measurements were acquired at approximately 1-meter intervals through the fitting of cylinders to point cloud data (Fig. 2). To determine the volume, Huber's formula was em-

ployed, which entails using the diameter measured at the midpoint of a log segment (eqn. 1):

$$V = h \cdot s_{0.5} \quad (1)$$

where V is the log volume, H is the log length, and $s_{0.5}$ is the mid-section sectional area.

Detailed descriptive statistics are provided for the trees inventoried in this study (Tab. 1). Statistics are derived from stem attributes estimated from point cloud data collected by the TLS device. The diameters and total height of the stems were estimated using the method described earlier, while the stem volume was determined by applying the Huber formula to each of the 1m-length cylinders. In total, 219 trees were extracted from the inventoried area and were systematically allocated to 5 folds. Four folds were exclusively designated for model training, while the remaining fold

was reserved for rigorous testing during each iteration of the cross-validation procedure.

(1) Used models

The data acquired by TLS was used to fit 4 taper models from 3 different model categories, namely parametric, semi-parametric, and non-parametric. Thereafter, the fitted models were evaluated in terms of their ability to predict diameters at different heights and the total stem volume as well. In the following, the different used taper models, as well as the evaluation metrics, are explicitly introduced.

Max and Burkhardt (1976)

The segmented stem profile model developed by Max & Burkhardt (1976) was used to represent the parametric model's category. Numerous studies have shown that this model is accurate in estimating upper stem diameters of many species (Jiang &

Tab. 1 - Mean, standard deviation (SD), and range for tree characteristics from all the samples. (*): Disk refers to a given cross-section of the stem, at a given height level, for which the diameter is measured.

Parameter	Mean	SD	min	max
DBH (cm)	45.05	11.22	21.0	82.0
Total height (m)	30.7	4.0	20.5	40.5
Disk* dob (cm)	29.3	12.5	0	82.5
Disk height (m)	14.5	8.7	0.9	40.5
Crown base height (m)	21.0	4.0	11.7	31.3
Volume (m ³)	2.27	1.32	0.38	8.04

Liu 2011, Ozçelik & Brooks 2012, Pang et al. 2016). The model consists of three quadratic functions representing the stem's lower, middle, and upper parts grafted together at two points (a_1 and a_2) with continuous polynomials and continuous first derivatives at each point. In the present study the taper equation was constrained to pass through diameter at breast height (dbh) while it is by construction constrained for the total height (ht).

Max & Burkhart (1976) taper equation can be expressed as follows (eqn. 2):

$$y_{ij}(z_{ij}) = b_1 x_{ij} + b_2 z_{ij}^2 + b_3 (z_{ij} - a_1)^2 I_1 + b_4 (z_{ij} - a_2)^2 I_2 + \varepsilon_{ij} \quad (2)$$

where $y_{ij} = (d_{ij} / dbh_i)^2$, dbh_i is the dbh of tree i , with $i = [1, 2, \dots, N]$, N is the number of trees in the sample, d_{ij} is the i -th upper-stem diameter over bark (dob) at height h_{ij} on tree i , with $j = [1, 2, \dots, n_i]$, n_i is the number of diameter measurements for tree i , $z_{ij} = 1 - (h_{ij}/ht)$ is the complement of the relative height, a_1 and a_2 are the join points to be estimated from the data, $I_k = 1$ if $z > a_k$, and 0 otherwise with $k = [1, 2]$, b_p 's are regression coefficients with $p = [1, 2, 3, 4]$ and ε_{ij} is the residual error term associated to the prediction of y_{ij} .

To constrain the taper equation to pass through dbh at the breast height, only one regression coefficient is needed to be changed. In this case, the b_1 parameter was modified as suggested in Cao (2009) and replaced by the modified b_1^* which can be expressed as follows (eqn. 3):

$$b_1^* = \frac{1 - y_i(z_{bh}) + b_1 z_{bh}}{z_{bh}} \quad (3)$$

where $z_{bh} = 1 - (1.3/ht)$. To estimate the parameters of eqn. 2, we used the Levenberg-Marquardt algorithm, which is implemented within the "nlS" function in the R stats package (R Core Team 2022).

Mixed-effects model

The stem taper models may show a high degree of multicollinearity, autocorrelation, and heteroscedasticity, all of which can be addressed with appropriate statistical methods. Many studies reported a reduction in the within-tree autocorrelation when including a tree-level random effect (Garber & Maguire 2003, Cao & Wang 2011). Using a mixed model for example, the autocorrelation within the stem can be accounted for at each height measurement level.

In the mixed effect formulation, all parameters of eqn. 2 are expressed as fixed effect (population level coefficients) while some parameters may contain additional random effect (for each tree within the population). In the case of the Max & Burkhart (1976) taper equation, different combinations of the parameters ($b_1, b_2, b_3, b_4, a_1, a_2$) could be a potential candidate for random parameters.

Here is an expression for the stem profile

including the fixed and random effects (eqn. 4):

$$y_{ij}(z_{ij}) = b_{1i} z_{ij} + b_{2i} z_{ij}^2 + b_{3i} (z_{ij} - a_{1i})^2 I_1 + b_{4i} (z_{ij} - a_{2i})^2 I_2 + \varepsilon_{ij} \quad (4)$$

where a_{1i} and a_{2i} are join points to be estimated from the data for each tree i , $I_k = 1$ if $z > a_{ki}$, and 0 otherwise with $k = [1, 2]$, and b_{pi} 's are the regression coefficients to be estimated for each tree i , with $p = [1, 2, 3, 4]$.

To define the statistical properties of the terms in eqn. 4, its vectorial formulation presented in the following offers a more simplified notation and will be considered (eqn. 5):

$$Y_i = (A_i B + B_i U_i) X_i + \varepsilon_i \quad (5)$$

where Y_i is a vector of predictions, X_i is a matrix of predictors, A_i is a matrix for the fixed effects, B_i is a matrix for random effects, B is a vector of fixed parameters, U_i is the vector for random effects and ε_i is a vector of random errors, with the assumptions that $\varepsilon_i = N(0, R)$ and $U_i = N(0, D)$ where R is the diagonal variance-covariance matrix of ε_i and D is a diagonal variance-covariance matrix of U_i .

Using the "saemix" package in R (Comets et al. 2017), the stochastic approximation expectation maximization algorithm was used to estimate the parameters of eqn. 4.

The volume equation derived from the integration of the Max and Burkhart taper equation can be expressed as follows (eqn. 6):

$$V = K \cdot dbh^2 \cdot ht \left\{ \frac{b_1}{2} (z_y^2 - z_x^2) + \frac{b_2}{3} (z_y^3 - z_x^3) + \frac{b_3}{3} [(z_y - a_1)^3 I_1 - (z_x - a_1)^3 J_1] + \frac{b_4}{3} [(z_y - a_2)^3 I_2 - (z_x - a_2)^3 J_2] \right\} \quad (6)$$

where $z_y = 1 - (h_y/ht)$ and $z_x = 1 - (h_x/ht)$, h_x and h_y being respectively the lower height of interest (m) and the upper height of interest, K is a conversion factor ($= \pi / 40,000$), $I_k = 1$ if $z_y > a_k$, and 0 otherwise, $J_k = 1$ if $z_x > a_k$, and 0 otherwise, with $k = [1, 2]$.

B-splines

Semiparametric models are a class of statistical models that provide a flexible compromise between the rigidity of fully parametric models and the complexity of fully non-parametric models. By avoiding rigid assumptions about functional form, semiparametric models offer the potential to capture the underlying relationship between variables more accurately, particularly for unknown or complex functional forms, while parametric approaches can be used for estimating model coefficients. Basis-splines (B-splines) are a type of semiparametric model that possess favorable numerical properties. To construct a B-spline model, a fixed number of splines, also called polynomials, are connected at specific points known as knots while ensur-

ing the condition of continuity of their second derivatives across the knots.

To formalize, here we consider that the population mean diameter for a given relative height (h_r) may be approximated by a B-spline function of degree p and can be expressed as follows (eqn. 7):

$$d_i = f(hr) + \varepsilon = \sum_{l=0}^{d_1} \beta_l B_{l,p}^{(1)}(hr) + \varepsilon \quad (7)$$

where $f(hr)$ is the response of the model which renders the diameter value d_i at a given h_r , β_l 's are the parameters of the model, d_1 is the number of parameters in the B-spline function with the condition that $d_1 \leq k_1 + 1 + p$; k_1 corresponds to the number of internal knots, $B_{l,p}^{(1)}(hr)$ is the B-spline basis function computed based on the "de Boor recurrence relation".

Due to the nonlinear nature of stem tapering and the hierarchy of data acquired from sampled trees, semiparametric models can be calibrated more efficiently using mixed approaches, which can also consider the random deviation of an individual tree from the population average. Linear mixed effect models can be used for the calibration if knot values are fixed *a priori*, which has the advantage of being less computationally expensive and more numerically stable than nonlinear mixed models.

The B-spline based taper equation considering both fixed and random effects can be written as (eqn. 8):

$$d_{ij} = f(hr_{ij}) + g(hr_{ij}) + \varepsilon_{ij} = \sum_{l=0}^{d_1} \beta_l B_{l,p}^{(1)}(hr_{ij}) + \sum_{l=0}^{d_2} \theta_{l,i} B_{l,p}^{(2)}(hr_{ij}) + \varepsilon_{ij} \quad (8)$$

where $f(hr_{ij})$ corresponds to the population mean response, $g(hr_{ij})$ represents the tree-specific deviation from $f(hr_{ij})$, ε_{ij} is the tree-specific residual error, d_{ij} corresponds to the j -th diameter from the i -th sample tree, hr_{ij} corresponds to the j -th relative height from the i -th sample tree, β_l 's are the fixed regression coefficient of the model, θ_{li} 's are the random effect parameters to be estimated for each tree, d_2 is the number of random parameters in the g function with the condition that $d_2 \leq k_2 + 1 + p$; k_2 corresponds to the number of internal knots in the random effect part of the model, $B_{l,p}^{(2)}(hr_{ij})$ is the B-spline basis function for the random effect part of the model.

The vectorial formulation of eqn. 8 can be expressed as follows (eqn. 9):

$$Y_i = X_i \beta + Z_i \theta_i + \varepsilon_i \quad (9)$$

where Y_i is the vector of diameter predictions at the level of the i -th tree, X_i is the matrix of B-spline basis values for the fixed effect part of the model, Z_i is the matrix of B-spline basis values for the random effect part of the model, β is a vector of fixed parameters, θ_i is the vector of random effect parameters, ε_i is the vector of residual error with the assumption that $\varepsilon_i = N(0, R)$ and $\theta_i = N(0, G)$ where R is the diagonal variance-covariance matrix of ε_i and G is a

positive non-diagonal matrix of θ_i . For more details on this method, see Kublin et al. (2013).

According to the suggestions by Kublin et al. (2013). The parameters of eqn. 8 were estimated considering knots [0.0, 0.1, 0.75, 1.0] corresponding to the B-spline basis function $B^{(1)}(hr)$ for the population mean function and [0.0, 0.1, 1.0] for $B^{(2)}(hr)$ representing the deviation from the population average. Hence, the regression model depends only on relative height. Measurements of dbh and further diameter measurements of a specific tree are merely used to calibrate the tree specific deviation from the average taper curve by estimating the corresponding random effects. To ensure diameter estimates of zero at tree top, the last spline in both $B^{(1)}(hr)$ and $B^{(2)}(hr)$ are omitted (see Fig. S1 in Supplementary material for a graphical visualization and Kublin et al. 2013).

The R package “Taper” was used to calibrate the model by the Restricted Maximum Likelihood (REML) method (default in the internally used R package “nlme” – Pinheiro et al. 2023) providing the dbh values for each sampling tree of training data. To estimate stem volumes, the implemented numerical integration method included in the “Taper” package was used (Kublin et al. 2023).

Random Forest

Nonparametric methods have been recently used for taper predictions providing reasonable performances (Nunes & Görrens 2016, Yang & Burkhart 2020). The present study used Random Forest to estimate the diameter at different heights of the stem. This algorithm works by training many decision trees on random subsets of the features, then averaging out their predictions. Regarding the structure of the model, the diameter was set as the dependent variable, which were predicted by four independent predictors namely *dbh*, *h*, *ht* and *h/ht* (abbreviations as above). For fine-tuning the model, the grid search method was used to evaluate all possible combinations of hyperparameters (viz., the number of decision trees; the number of features to consider when looking for the best split; the minimum number of samples required to split an internal node) within the search space as well as the cross-validation method to prevent overfitting. The Python library “scikit-learn” (ver. 0.24.2) which includes an implementation of the Random Forest algorithm was used here (Pedregosa et al. 2011).

Since nonparametric methods are not algebraically integrable, including the random forest algorithm, the Monte Carlo integration, which is a numerical integration approach, was self-implemented and used to predict the stem volume for each sample tree (Fig. 3). This technique relies on a random choice of points at which the integrand is evaluated. Monte Carlo integration could be expressed as follows (eqn.

$$V = K \cdot (h_b - h_a) \cdot \frac{1}{N} \sum_1^N [RF(h_i)]^2 \approx K \cdot \int_{h_a}^{h_b} [RF(h)]^2 dh \tag{10}$$

where h_a and h_b being respectively the lower and the upper heights of interest (m), K is the conversion factor ($= \pi/40,000$), h_i corresponds to the i -th height on the tree generated randomly from a uniform distribution $U(h_a, h_b)$, $i = [1, 2, \dots, N]$, and N is the number of samples generated with the distribution U .

Evaluation of the methods

Taper models should yield unbiased estimates with minimal variance in both diameter outside bark (dob) and stem volume. A single overall measure of bias or residual error of estimation calculated for dob and stem volume does not provide an exclusive indicator of the goodness in evaluating several taper equations for a set of trees of a given species. This holds true when calculating bias metrics at different parts of stems and for different trees can result in a zero or close-to-zero average given that bias can be positive or negative. Alternatively, the standard error which can report the variability of the biases is considered of better reliability. Moreover, when dealing with small sample size, the residual variance can be influenced by the degrees of freedom, making the correlation index, also known as the fitness index, a useful metric to address this concern. All those metrics were used to rank the fitted models.

The most reliable measure of mean bias

(B), overall residual variance (SSE) and the fitness index (FI) can be expressed as follows (eqn. 11 – eqn. 13):

$$B = \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)}{n} \tag{11}$$

$$SEE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n - k} \tag{12}$$

$$FI = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \tag{13}$$

where Y_i is the actual observation, \hat{Y}_i is the predicted value of the actual observation, \bar{Y} is the mean of the actual observations, n is the number of observation and k corresponds to the number of estimated parameters.

As part of our evaluation process, we employed a grouping methodology to assess the performance of the taper models in conjunction with the overall assessment presented earlier. For dob, biases and standard errors were calculated for each 10% step of the height along the stem, while for the total stem volume (from ground to top), the biases and standard errors were calculated by DBH classes. This allowed us to gain a comprehensive understanding of the accuracy of the taper equations across various tree dimensions as well as across different parts within the trees.

Results

Taper prediction

The model by Max & Burkhart (1976) was fit using both fixed and mixed-effects approaches. Different mixed-effects combinations of parameters were tested (Tab. S1 in

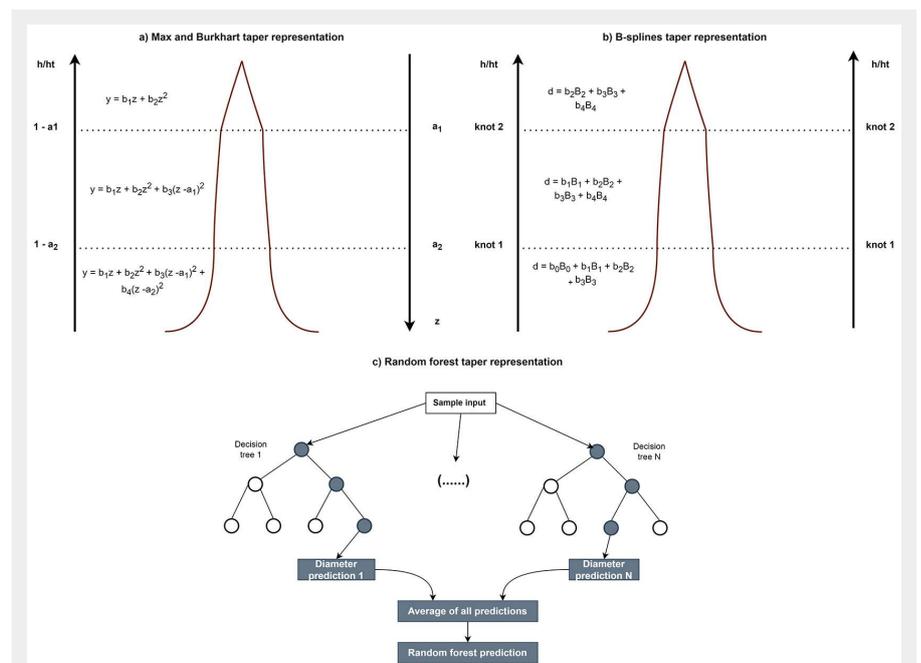


Fig. 3 - Schematic tree form representation for the three applied models. (a) the Max and Burkhart parametric model; (b) B-splines semi-parametric model; (c) Random Forest non-parametric model.

Tab. 2 - Estimates of parameters (± standard errors) for fixed-effects and mixed-effects taper models, based on the validation data set.

Parameters	Fixed-effect model	Mixed-effects model
a_1	0.6384 ± 0.0242	0.71 ± 0.007
a_2	0.8756 ± 0.0043	0.89 ± 0.001
b_1	0.9577 ± 0.0085	0.96 ± 0.014
b_2	-0.2546 ± 0.0166	-0.25 ± 0.017
b_3	1.1693 ± 0.2123	2.06 ± 0.083
b_4	25.5313 ± 1.9310	38.67 ± 1.428
var(a_1)	-	0.006 ± 0.00074
var(a_2)	-	0.0001 ± 0.00002
var(b_1)	-	0.042 ± 0.0042
var(b_2)	-	0.063 ± 0.0065
var(b_4)	-	52.250 ± 20.10

Tab. 3 - Estimates of parameters (± standard errors) for B-splines model, based on the validation data set.

Parameters	B-splines model
β_1	59.10 ± 0.89
β_2	41.58 ± 0.52
β_3	31.76 ± 0.22
β_4	29.27 ± 0.44
β_5	13.88 ± 0.05
var(θ_1)	172.95 ± 13.15
var(θ_2)	114.94 ± 10.72
var(θ_3)	38.71 ± 6.22
var(θ_4)	136.16 ± 11.66

Supplementary material), and the combination involving a_1 , a_2 , b_1 , b_2 and b_4 produced the lowest values of Akaike's information criterion (AIC) and Bayesian information criterion (BIC), resulting in the following parameter values (Tab. 2) and model form (eqn. 14):

$$y_{ij}(z_{ij}) = b_1 z_{ij} + b_2 z_{ij}^2 + b_3 (z_{ij} - a_{1i})^2 I_1 + b_4 (z_{ij} - a_{2i})^2 I_2 + \epsilon_{ij} \quad (14)$$

The mixed B-splines model was fit using the linear mixed-effects approach implemented in the TapeR package and reached convergence. In order to constrain the model predictions to equal zero at the tree

top, d , which is the number of fixed parameters was set to 5, while d_2 which corresponds to the number of random parameters was set to 4, omitting the topmost spline, respectively (see Fig. S1 in Supplementary material). The results of B-splines model fitting are represented in Tab. 3.

Regarding the RF model, after rigorous fine-tuning, the model reached a state of convergence as the number of decision trees stabilized at an optimal value of 267. In addition, the hyperparameter tuning process led to the following optimized configuration: number of features to best split = 3; minimum samples to split internal nodes = 4.

For each taper equation, overall statistics of fit (Bias, SEE and FI) for the entire stem were calculated and are represented in Tab. 4 for dob by model. Among the four models under investigation, discernible trends in bias emerged. The Max and Burkhart models exhibited a noteworthy positive bias, suggestive of their propensity to overestimate diameters. In contrast, the B-spline and Random Forest models displayed a discernible negative bias, indicating a tendency to underestimate diameters. Notably, the Max and Burkhart fixed effect model demonstrated a particularly pronounced inclination towards overestimation. The results suggest that mixed-effect models outperformed models considering only fixed effects. In particular, the Max and Burkhart and B-splines mixed-effects models showed relatively comparable SEE values of 0.78 cm and 0.96 cm, respectively. In contrast, the Max and Burkhart fixed-effect and random forest models had comparable SEE values of 1.90 cm and 2.26 cm, respectively. This comparison holds true for FI as well, although to a lesser extent than for SEE.

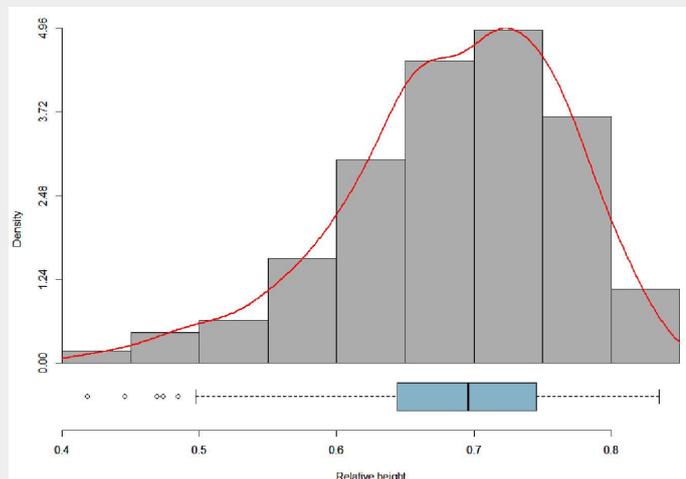
To assess the performance of taper models at different positions throughout the stem, the statistics of fit, derived from the 5-fold cross-validation results, were analyzed by relative height classes. These statistics are documented in Tab. S2 (Supplementary material) and graphically depicted in Fig. 1. To comprehensively assess the taper of all models, the average bias, SEE, and FI were calculated for each model by relative height class along the stem. The results show that models including mixed effects outperformed those considering only fixed effect roughly along all the section heights. Noticeably, a significant increase of SEE was observed for all the models from a relative height class of (70-80%) and was shown to coincide with the crown base height (CBH) for most of the trees (Fig. 4).

Bias measurements indicate that the Max and Burkhart fixed-effect model tends to consistently overestimate upper stem diameters throughout the entire stem. On the other hand, the random forest model tends to underestimate upper stem diameters. B-spline-based predictions tend to overestimate upper stem diameters in the

Tab. 4 - Total stem fit statistics for the four taper models from cross-validation results. The underlined values refer to the best statistics. M&B is used here as the acronym of Max and Burkhart.

Model	Bias ± SD (cm)	SEE ± SD (cm)	FI ± SD
M&B fixed effect	0.274 ± 0.292	1.90 ± 0.213	0.97 ± 0.030
M&B mixed effects	0.027 ± 0.073	0.78 ± 0.094	0.99 ± 0.021
B-splines	-0.015 ± 0.082	0.96 ± 0.102	0.99 ± 0.006
Random forest	-0.010 ± 0.357	2.26 ± 0.286	0.96 ± 0.046

Fig. 4 - Distribution of the relative crown base height (in gray color). The distribution fit curve (in red color). The boxplot of the relative crown base height (in blue color).



lower and upper portions while underestimating in the middle portion. In contrast, predictions from the Max and Burkhart mixed-effects models exhibit an opposing bias pattern in the lower and middle sections but demonstrate similar behavior in the upper portion of the stem.

Volume prediction

The predictions of volume (over bark) by four taper models were compared with each other. Overall statistics of fit (Bias, SEE, and FI) for the total stem volume are represented in Tab. 5. The results show that mixed-effects models outperformed models considering only fixed effect. Specifically, Max and Burkhart and B-splines models had a comparable SEE values of respectively 0.021 and 0.034 m³, while the Max and Burkhart fixed effect and the Random Forest models had comparable SEE values of respectively 0.221 and 0.273 m³. In addition, the overall results show that all the four models tend to generally overestimate the volume as the bias resulted in positive values for all the candidate models. Notably, the Max and Burkhart models (both fixed and mixed effects) exhibited the smallest overprediction, with an absolute bias value of less than 0.01 m³.

To assess the performance of stem volume prediction for different tree dimensions, the statistics of fit (SEE, bias, and FI) were analyzed by DBH class categories. These statistics are documented in Tab. S3 (Supplementary material) and graphically depicted in Fig. 6. In general, results show that prediction errors are larger in large diameter trees for both Max and Burkhart fixed effect and Random Forest models, while the prediction error tend to be constant over all the DBH classes for Max and Burkhart mixed effects and B-splines models. The results also demonstrate a strong correlation between volume predictions and volume observations (FI > ~ 0.5) for all four models across all DBH classes, except for the DBH class category (55-65 cm) and (> 65 cm), where both the M&B fixed effect and Random Forest models exhibit limitations (FI ~ 0.42 and 0.21, SEE ~ 0.39 m³ and 0.71 m³, respectively).

Model ranking

A taper-estimating system should provide an unbiased estimation with minimum variance of both diameter outside bark (dob) and total stem volume outside bark. Using the results from Tab. 4 and Tab. 5, it is possible to rank the four estimating systems. The rank sums were created by ranking the performance by dob and stem volume which were equally weighted here in this study. The rank sums reported in Tab. 6 were generated in such a way that each estimating system was assigned a rank separately for every relative height for dob and for every DBH class for the volume. These ranks were also summed for standard errors and biases. As a result, Max and Burkhart mixed-effects system ranked first

Tab. 5 - Total stem volume fit statistics for the four taper models from cross-validation results. The underlined values refer to the best statistics. M&B is used here as the acronym of Max and Burkhart.

Model	Bias ± SD (m ³)	SEE ± SD (m ³)	FI ± SD
M&B fixed effect	0.0085 ± 0.065	0.221 ± 0.044	0.97 ± 0.207
M&B mixed effects	0.0080 ± 0.001	0.021 ± 0.002	0.99 ± 0.001
B-splines	0.0240 ± 0.006	0.034 ± 0.005	0.99 ± 0.005
Random forest	0.0560 ± 0.066	0.273 ± 0.066	0.91 ± 0.169

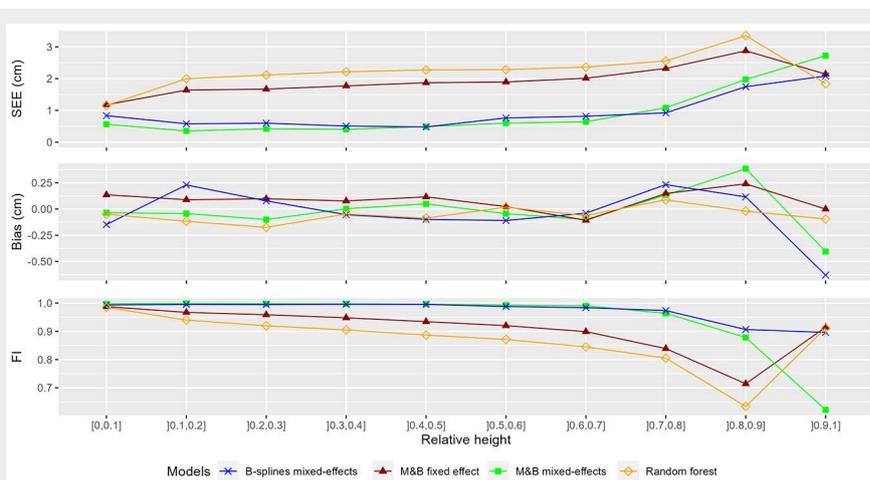


Fig. 5 - Bias, SEE and FI of the estimate of diameter over bark (dob) by relative class for the four models.

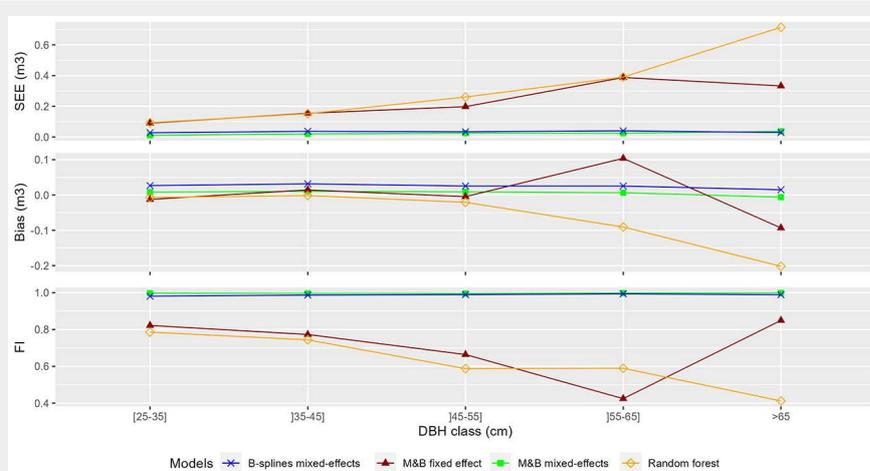


Fig. 6 - Bias, SEE and FI of the estimate of the stem volume by DBH class for the four models.

Tab. 6 - Rank sum analysis for the four estimating systems.

Model Parameters	Max and Burkhart		B-splines	Random forest
	Fixed effect	Mixed effects		
Overall SEE for diameter and volume	6 (3)	2 (1)	4 (2)	8 (4)
Overall bias for diameter and volume	6 (3)	4 (1)	5 (2)	5 (2)
SEE for diameter by relative height	31 (3)	17 (1)	16 (2)	36 (4)
Bias for diameter by relative height	29 (3)	21 (2)	30 (4)	20 (1)
SEE for volume by DBH classes	15 (3)	6 (1)	8 (2)	18 (4)
Bias for volume by DBH classes	16 (4)	9 (1)	13 (3)	12 (2)
Total	103 (4)	59 (1)	76 (2)	99 (3)

in most of all the ranking categories. B-splines based system ranked second in estimating diameter and volume by standard error of estimates while random forest system ranked third and Max and Burkhart fixed effect ranked fourth with regards to the same previously discussed estimates.

Discussion

Taper equations are an indispensable tool in the field of forest management and industry. However, the development of an accurate taper model present inherent challenges, mainly due to the need for comprehensive customization to account for the characteristics exhibited by each species and site-specific environmental conditions, and the availability of high-quality data required for the model development. In the present study, four taper models from three different class categories were compared in terms of their capacity to predict diameter over bark (dob) and total stem volume of *Pinus nigra* Arn. Across models, the Max and Burkhart and B-splines mixed-effects models demonstrated superior accuracy and reliability in predicting both dob and total stem volume, with a small advance for the Max and Burkhart model.

The Terrestrial Laser Scanner (TLS) has previously been employed as an effective tool for measuring trees in forest environments and generating taper equations (Saarinen et al. 2019, Sun et al. 2016). The case study conducted on black pine forests, which is presented here, demonstrated the feasibility of using TLS for surveys and highlighted the high quality of the acquired data. In the Italian context, where tree measurements for quantitative purposes have significantly decreased in recent decades, technological tools like TLS represent a pivotal advancement for rejuvenating our understanding of the Italian forest sector.

In the present study, biases and standard errors of estimate were evaluated for diameter outside bark (dob) and total stem volume. These biases and standard errors were also evaluated for different heights within the trees and across different tree sizes. The overall goodness-of-fit statistics (Bias, SEE, and FI) were calculated and reported in Tab. 4, Tab. 5 and Fig. 5, Fig. 6 for the four tested models. The results indicate that the Max and Burkhart and B-splines mixed-effects models explained more than 99% of the total variation in predicting the upper-stem diameter and total stem volume. The Max and Burkhart fixed-effect model explained more than 99% of this total variation, while the Random Forest model explained more than 96% for the diameter variation and 96% for the total volume variation.

The superiority of models that include random effects over those that only consider fixed effects has been demonstrated in various previous studies, and our findings are in agreement with this. For exam-

ple, a study conducted by Leites & Robinson (2004) showed that the inclusion of random effects considerably improved the fit of the Max and Burkhart taper equation for *Pinus taeda*. In a different study by Bronisz & Zasada (2020), the Kozak's taper equation fitted using a mixed-effects approach provided better results for the upper diameter and total stem volume for *Pinus sylvestris*. In our study, the Max & Burkhart (1976) model was fitted using both fixed and mixed effects approaches. Notably, the error was reduced by 52% for the upper-stem diameter and by 89% for the total stem volume when random effects were additionally considered.

It is also noteworthy that the evaluation of the proposed models exhibited a higher Standard Error of Estimation (SEE) in predicting the upper-stem diameter at a relative height of [0.7, 0.8] (Tab. S2 – Supplementary material). The analysis of the distribution of the tree's relative crown base height reveals that 50% of the observations are located between a value of 0.7 and 0.8, and the maximum relative crown base height was found to be approximately equal to 0.8 (Fig. 4). This implies that all observations above a relative height of 0.8 were necessarily measured inside the crown. This last result may partly explain the notable increase in error in predicting the upper-stem diameter at a relative height of 0.7. In fact, many studies have reported the inability of Terrestrial Laser Scanners to collect reliable data beyond the base of the crown due to interference from branches and leaves as well as the mutual occlusion between canopies. Practically, when approaching higher parts of the stem, occlusion starts to deteriorate the quality of cylinder fitting (Liu et al. 2018, Pitkänen et al. 2019). Additionally, as the distance from the scanner increases, spatial resolution decreases, which is a well-known effect in scanning that particularly impact the upper parts of trees (Liang et al. 2016). The interference from the crown, compounded with the increased distance, may explain the models' inability to detect a well-defined taper pattern above the crown base.

In this study, Monte Carlo numerical integration was employed for estimating the volume of tree stems using the non-differentiable and complex Random Forest model. This decision was justified by the model's inherent characteristics mentioned earlier, which render traditional integration methods less suitable. It is imperative to underscore the significance of comparing our chosen method with alternative numerical integration techniques to gain insights into result robustness and accuracy, potentially informing future research. Notably, the RF model, unlike the parametric and semi-parametric models used in this study, typically demands a substantial amount of training data. In our analysis, the RF model obtained the lowest score in terms of SEE (dob and volume). This high-

lights the potential impact of our relatively small dataset (219 trees) on RF performance. In addition, the use of various integration methods (Huber, Monte Carlo, and analytical integration) in our study may introduce minor biases affecting our results. Further research is needed to assess and potentially mitigate these biases, as their combined impact on the findings remains uncertain.

Considering the results of this study, the Max & Burkhart (1976) taper model, calibrated by means of a mixed-effects approach, and its compatible volume integration could be considered for operational use in estimating diameter outside bark at different heights, as well as total stem volume, respectively, with an overall precision of 0.781 cm and 0.021 m³ in *Pinus nigra* Arn.

Conclusion

This study explored the application of precise and non-destructive techniques, such as Terrestrial Laser Scanning, for collecting data to develop taper equations specific to *Pinus nigra* Arn., in line with the principles of precision forestry. The collected data were used to fit four taper models from three different class categories, evaluated for their ability to predict diameter at various heights and total stem volume. The Max & Burkhart (1976) taper model, calibrated with a mixed-effects approach, proved superior in estimating diameter and volume compared to other models. However, a common limitation among all models was their inability to accurately predict diameter beyond a relative height of 0.7 due to crown interference. The selected model, within the scope of this study, can be operationally utilized by forest managers to enhance predictions of diameter and stem volume, aiding in optimizing wood allocation to different harvested products. Future research perspectives may focus on refining algorithms and methods for extracting diameter at the upper parts of the stem.

Data, code availability and setup

The whole dataset is made available at <https://doi.org/10.5281/zenodo.8414408>. The study was performed on a laptop with Win10 22H2 operating system, an Intel Core i5 7200U CPU with maximal capacity of 2.70 GHz, and 8 Gb of RAM.

Authors contribution

NP and CV contributed equally to the manuscript.

References

- Baker ER, Peña J-M, Jayamohan J, Jé A (2018). Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters* 14 (5): 20170660. - doi: [10.1098/rsbl.2017.0660](https://doi.org/10.1098/rsbl.2017.0660)
- Bernetti G, Cantiani M, Hrelrig B (1969). Ricerche alometriche e dendrometriche sulle pinete di pino nero e laricio della Toscana [Alometric and dendrometric research on the black pine

- and larch pine forests of Tuscany]. *L'Italia Forestale e Montana* 26 (1): 10-40. [in Italian]
- Bottalico F, Chirici G, Travaglini D (2012). The forest management of Vallombrosa from 1876 to 2006: Analysis of forest maps. *L'Italia Forestale e Montana* 67 (6): Art. 6. - doi: [10.4129/ifm.2012.6.01](https://doi.org/10.4129/ifm.2012.6.01)
- Bronisz K, Zasada M (2020). Correction: Bronisz K. and Zasada M., Comparison of fixed- and mixed-effects approaches to taper modeling for Scots pine in West Poland. *Forests* 2019, 10, 975. *Forests* 11 (4): 437. - doi: [10.3390/f11040437](https://doi.org/10.3390/f11040437)
- Brunet-Navarro P, Jochheim H, Muys B (2017). The effect of increasing lifespan and recycling rate on carbon storage in wood products from theoretical model to application for the European wood sector. *Mitigation and Adaptation Strategies for Global Change* 22 (8): 1193-1205. - doi: [10.1007/s11027-016-9722-z](https://doi.org/10.1007/s11027-016-9722-z)
- Calders K, Newnham G, Burt A, Murphy S, Raunonen P, Herold M, Culvenor D, Avitabile V, Disney M, Armston J, Kaasalainen M (2015). Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecology and Evolution* 6 (2): 198-208. - doi: [10.1111/2041-210X.12301](https://doi.org/10.1111/2041-210X.12301)
- Calders K, Verbeeck H, Burt A, Origo N, Nightingale J, Malhi Y, Wilkes P, Raunonen P, Bunce RG, Disney M (2022). Laser scanning reveals potential underestimation of biomass carbon in temperate forest. *Ecological Solutions and Evidence* 3 (4): e12197. - doi: [10.1002/2688-8319.12197](https://doi.org/10.1002/2688-8319.12197)
- Cao QV (2009). Calibrating a segmented taper equation with two diameter measurements. *Southern Journal of Applied Forestry* 33 (2): 58-61. - doi: [10.1093/sjaf/33.2.58](https://doi.org/10.1093/sjaf/33.2.58)
- Cao QV, Wang J (2011). Calibrating fixed- and mixed-effects taper equations. *Forest Ecology and Management* 262 (4): 671-673. - doi: [10.1016/j.foreco.2011.04.039](https://doi.org/10.1016/j.foreco.2011.04.039)
- Choudhry H, O'Kelly G (2018). Precision forestry: a revolution in the woods. McKinsey & Co., Singapore, pp. 1-11. [online] URL: <http://www.mckinsey.com/industries/packaging-and-paper/our-insights/precision-forestry-a-revolution-in-the-woods>
- Comets E, Lavenue A, Lavielle M (2017). Parameter estimation in nonlinear mixed effect models using saemix, an R Implementation of the SAEM algorithm. *Journal of Statistical Software* 80 (3): 1-41. - doi: [10.18637/jss.v080.i03](https://doi.org/10.18637/jss.v080.i03)
- Dálya LB, Capretti P, Ghelardini L, Jankovsky L (2019). Assessment of presence and distribution of *Armillaria* and *Heterobasidion* root rot fungi in the forest of Vallombrosa (Apennines Mountains, Italy) after severe windstorm damage. *iForest - Biogeosciences and Forestry*, 12 (1): 118-124. - doi: [10.3832/ifor2929-012](https://doi.org/10.3832/ifor2929-012)
- Enescu C, De Rigo D, Caudullo G, Durrant T (2016). *Pinus nigra* in Europe: distribution, habitat, usage and threats. In: "European Atlas of Forest Tree Species" (San-Miguel-Ayanz J, De Rigo D, Caudullo G, Houston Durrant T eds). EU Publication Office, Luxembourg, pp. 126-127.
- Erber G, Gollob C, Krassnitzer R, Nothdurft A, Stampfer K (2022). Stem-level bucking pattern optimization in chainsaw bucking based on terrestrial laser scanning data. *Croatian Journal of Forest Engineering* 43 (2): 287-301. - doi: [10.5552/crojfe.2022.1596](https://doi.org/10.5552/crojfe.2022.1596)
- Fardusi MJ, Chianucci F, Barbati A (2017). Concept to practice of geospatial-information tools to assist forest management and planning under precision forestry framework: a review. *Annals of Silvicultural Research* 41 (1): 3-14. - doi: [10.12899/asr-1354](https://doi.org/10.12899/asr-1354)
- Garber SM, Maguire DA (2003). Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. *Forest Ecology and Management* 179 (1): 507-522. - doi: [10.1016/S0378-1127\(02\)00528-5](https://doi.org/10.1016/S0378-1127(02)00528-5)
- Gordon A (1983). Comparison of compatible polynomial taper equations. *New Zealand Journal of Forestry Science* 13 (2): 146-155. - [online] URL: <http://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=73892f2a14285dd3cd937f9eddda6c4dabbc5b86>
- Gómez-García E, Crecente-Campo F, Diéguez-Aranda U (2013). Selection of mixed-effects parameters in a variable-exponent taper equation for birch trees in northwestern Spain. *Annals of Forest Science* 70 (7): 707-715. - doi: [10.1007/s13595-013-0313-9](https://doi.org/10.1007/s13595-013-0313-9)
- Hansen E, Rahlf J, Astrup R, Gobakken T (2023). Taper, volume, and bark thickness models for spruce, pine, and birch in Norway. *Scandinavian Journal of Forest Research* 38 (6): 413-428. - doi: [10.1080/02827581.2023.2243821](https://doi.org/10.1080/02827581.2023.2243821)
- Ilsenburg M (2017). LAStools - Efficient LiDAR processing software. Version 141017 academic, web site. [online] URL: <http://rapidlasso.com/LAStools>
- Jiang L, Liu R (2011). Segmented taper equations with crown ratio and stand density for Dahurian larch (*Larix gmelinii*) in Northeastern China. *Journal of Forestry Research* 22 (3): 347-352. - doi: [10.1007/s11676-011-0178-4](https://doi.org/10.1007/s11676-011-0178-4)
- Kováčová P, Antalová M (2010). Precision forestry-definition and technologies. *Sumarski List* 34 (11-12): 603-610.
- Kozak A, Smith JHG (1993). Standards for evaluating taper estimating systems. *The Forestry Chronicle* 69 (4): 438-444. - doi: [10.5558/tfc69438-4](https://doi.org/10.5558/tfc69438-4)
- Kublin E, Breidenbach J, Kändler G (2013). A flexible stem taper and volume prediction method based on mixed-effects B-spline regression. *European Journal of Forest Research* 132 (5): 983-997. - doi: [10.1007/s10342-013-0715-0](https://doi.org/10.1007/s10342-013-0715-0)
- Kublin E, Breidenbach J, Vonderach C (2023). TapeR: flexible tree taper curves based on semi-parametric mixed models (0.5.2). Computer software, web site. [online] URL: <http://cran.r-project.org/web/packages/TapeR/index.html>
- Kuzelka K, Marušák R (2014). Comparison of selected splines for stem form modeling: a case study in Norway spruce. *Annals of Forest Research* 57 (1): 137-148 - doi: [10.15287/afr.2014.177](https://doi.org/10.15287/afr.2014.177)
- Leites LP, Robinson AP (2004). Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework. *Forest Science* 50 (2): 204-212. - doi: [10.1093/forestscience/50.2.204](https://doi.org/10.1093/forestscience/50.2.204)
- Liang X, Hyypä J (2013). Automatic stem mapping by merging several terrestrial laser scans at the feature and decision levels. *Sensors* 13 (2): 1614-1634. - doi: [10.3390/s130201614](https://doi.org/10.3390/s130201614)
- Liang X, Kankare V, Hyypä J, Wang Y, Kukko A, Haggrén H, Yu X, Kaartinen H, Jaakkola A, Guan F, Holopainen M, Vastaranta M (2016). Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing* 115: 63-77. - doi: [10.1016/j.isprsjprs.2016.01.006](https://doi.org/10.1016/j.isprsjprs.2016.01.006)
- Liu G, Wang J, Dong P, Chen Y, Liu Z (2018). Estimating individual tree height and diameter at breast height (DBH) from Terrestrial Laser Scanning (TLS) data at plot level. *Forests* 9 (7): 398. - doi: [10.3390/f9070398](https://doi.org/10.3390/f9070398)
- Maas H-G, Bienert A, Scheller S, Keane E (2008). Automatic forest inventory parameter determination from terrestrial laser scanner data. *International Journal of Remote Sensing* 29 (5): 1579-1593. - doi: [10.1080/01431160701736406](https://doi.org/10.1080/01431160701736406)
- Marchi M, Scotti R, Rinaldini G, Cantiani P (2020). Taper function for *Pinus nigra* in Central Italy: is a more complex computational system required? *Forests* 11 (4): 405. - doi: [10.3390/f11040405](https://doi.org/10.3390/f11040405)
- Max T, Burkhart H (1976). Segmented polynomial regression applied to taper equations. *Forest Science* 22: 283-289.
- McTague JP, Weiskittel A (2020). Evolution, history, and use of stem taper equations: a review of their development, application, and implementation. *Canadian Journal of Forest Research* 51 (2): 210-235. - doi: [10.1139/cjfr-2020-0326](https://doi.org/10.1139/cjfr-2020-0326)
- Mäkelä A, Valentine HT (2020). Applications and future outlook. In: "Models of Tree and Stand Dynamics: Theory, Formulation and Application" (Mäkelä A, Valentine HT eds). Springer International Publishing, Cham, Switzerland, pp. 245-266. - doi: [10.1007/978-3-030-35761-0_10](https://doi.org/10.1007/978-3-030-35761-0_10)
- Nunes MH, Görgens EB (2016). Artificial intelligence procedures for tree taper estimation within a complex vegetation mosaic in Brazil. *PLoS One* 11 (5): e0154738. - doi: [10.1371/journal.pone.0154738](https://doi.org/10.1371/journal.pone.0154738)
- Olivera Farias A, Visser R (2016). Using the harvester on-board computer capability to move towards precision forestry. *New Zealand Journal of Forestry* 60: 3-7. - doi: [10.1186/s40490-016-0059-x](https://doi.org/10.1186/s40490-016-0059-x)
- Ozçelik R, Brooks JR (2012). Compatible volume and taper models for economically important tree species of Turkey. *Annals of Forest Science* 69 (1): 105-118. - doi: [10.1007/s13595-011-0137-4](https://doi.org/10.1007/s13595-011-0137-4)
- Pang L, Ma Y, Sharma RP, Rice S, Song X, Fu L (2016). Developing an improved parameter estimation method for the segmented taper equation through combination of constrained two-dimensional optimum seeking and least square regression. *Forests* 7 (9): 194. - doi: [10.3390/f7090194](https://doi.org/10.3390/f7090194)
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12 (85): 2825-2830.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, Heisterkamp S, Willigen BV, Ranke J, R Core Team (2023). nlme: linear and nonlinear mixed effects models. Version 3:1-162, web site. [online] URL: <http://cran.r-project.org/web/packages/nlme/index.html>
- Pitkänen TP, Raunonen P, Kangas A (2019). Measuring stem diameters with TLS in boreal forests by complementary fitting procedure. IS-

PRS Journal of Photogrammetry and Remote Sensing 147: 294-306. - doi: [10.1016/j.isprsjprs.2018.11.027](https://doi.org/10.1016/j.isprsjprs.2018.11.027)

Puletti N, Grotti M, Scotti R (2019). Evaluating the eccentricities of poplar stem profiles with terrestrial laser scanning. *Forests* 10 (3): 239. - doi: [10.3390/f10030239](https://doi.org/10.3390/f10030239)

R Core Team (2022). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [online] URL: <http://www.r-project.org/>

Saarinén N, Kankare V, Pyörälä J, Yrttimaa T, Liang X, Wulder MA, Holopainen M, Hyyppä J, Vastaranta M (2019). Assessing the effects of sample size on parametrizing a taper curve equation and the resultant stem-volume estimates. *Forests* 10 (10): 848. - doi: [10.3390/f10100848](https://doi.org/10.3390/f10100848)

Salekin S, Catalán CH, Boczniewicz D, Phiri D, Morgenroth J, Meason DF, Mason EG (2021). Global tree taper modelling: a review of appli-

cations, methods, functions, and their parameters. *Forests* 12 (7): 913. - doi: [10.3390/f12070913](https://doi.org/10.3390/f12070913)

Sun Y, Liang X, Liang Z, Welham C, Li W (2016). Deriving merchantable volume in poplar through a localized tapering function from non-destructive terrestrial laser scanning. *Forests* 7 (4): 87. - doi: [10.3390/f7040087](https://doi.org/10.3390/f7040087)

Thiel D, Nagy L, Beierkuhnlein C, Huber G, Jentsch A, Konner M, Kreyling J (2012). Uniform drought and warming responses in *Pinus nigra* provenances despite specific overall performances. *Forest Ecology and Management* 270: 200-208. - doi: [10.1016/j.foreco.2012.01.034](https://doi.org/10.1016/j.foreco.2012.01.034)

Weiskittel AR, Kershaw JA, Vanclay JK (2011). Individual-tree static equations. In: "Forest Growth and Yield Modeling". John Wiley and Sons, Ltd., Hoboken, NJ, USA, pp. 115-137. - doi: [10.1002/9781119998518.ch7](https://doi.org/10.1002/9781119998518.ch7)

Yang S-I, Burkhart HE (2020). Robustness of parametric and nonparametric fitting procedures of tree-stem taper with alternative definitions

for validation data. *Journal of Forestry* 118 (6): 576-583. - doi: [10.1093/jofore/fvaa036](https://doi.org/10.1093/jofore/fvaa036)

Supplementary Material

Tab. S1 - Model selection statistics for evaluating the inclusion of random parameters in the Max and Burkhart taper model.

Fig. S1 - Population mean-fixed effects B-spline basis.

Tab. S2 - Bias (cm), standard error (cm), and fitness index of the estimate of dob by relative height for four models.

Tab. S3 - Bias (m³), standard error (m³), and fitness index of the estimate of stem volume by DBH class for four models.

Link: Boukhris_4525@suppl001.pdf