



Diameter distributions in *Pinus sylvestris* L. stands: evaluating modelling approaches including a machine learning technique

Şükrü Teoman Güner¹ · Maria J. Diamantopoulou² · Ramazan Özçelik³

Received: 5 September 2022 / Accepted: 31 January 2023 / Published online: 2 June 2023
© The Author(s) 2023

Abstract The diameter distribution of trees in a stand provides the basis for determining the stand's ecological and economic value, its structure and stability and appropriate management practices. Scots pine (*Pinus sylvestris* L.) is one of the most common and important conifers in Turkey, so a well-planned management schedule is critical. Diameter distribution models to accurately describe the stand structure help improve management strategies, but developing reliable models requires a deep understanding of the growth, output and constraints of the forests. The most important information derived by diameter distribution models is primary data on horizontal stand structure for each diameter class of trees: basal area and volume per unit area. These predictions are required to estimate the range of products and predicted volume and yield from a forest stand. Here, to construct an accurate, reliable diameter distribution model for natural Scots pine stands in the Türkmen Mountain region, we used Johnson's S_B distribution to represent the empirical diameter distributions of the stands using ground-based measurements from 55 sample plots that included 1219 trees in natural distribution zones of the forests. As an

alternative, nonparametric approach, which does not require any predefined function, an artificial intelligence model was constructed based on support vector machine methodology. An error index was calculated to evaluate the results. Overall, both Johnson's S_B probability density function with a three-parameter recovery approach and the support vector regression methodology provided reliable estimates of the diameter distribution of these stands.

Keywords Diameter distribution · Johnson's S_B · Support vector regression · Scots pine · Türkmen mountains

Introduction

Scots pine (*Pinus sylvestris* L.) is one of the main forest tree species in high demand in Turkey and covers 1.4 million ha (6.2% of the total forest area) in Turkey (GDF 2015). Ecologically, Scots pine forests provide important ecological services such as protecting soil and water resources, mitigating, and adapting to the negative effects of climate change, and protecting biological diversity (Fischer et al. 2008; Lebedev 2022). Moreover, due to its valuable, versatile wood, it is in high demand and has high economic value. Last but not least, Scots pine forests that dominate the study region and Anatolian black pine (*Pinus nigra* Arnold. subsp. *pallasiana* Lamb.), trembling poplar (*Populus tremula* L.) and oriental beech (*Fagus orientalis* Lipsky) support a huge variety of wildlife worthy of consideration in a management strategy for the forest. To create planning and management strategies to protect these forests, reliable and accurate information on the status, growth and development characteristics of these forests is needed.

Tools capable of considering the structures and special attributes of Scots pine stands are required for managing

The online version is available at <http://www.springerlink.com>.

Corresponding editor: Tao Xu.

✉ Maria J. Diamantopoulou
mdiamant@for.auth.gr

¹ Department of Forestry, Ulus Vocational School, Bartın University, 74600 Ulus, Bartın, Turkey

² Faculty of Agriculture, Forestry and Natural Environment, School of Forestry and Natural Environment, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

³ Faculty of Forestry, Isparta University of Applied Science, East Campus, 32260 Isparta, Turkey

these natural resources in a multifunctional approach. Detailed and accurate predictions of stand development under different management strategies often needed. Diameter classes help to obtain information such as stand structure, age distribution, and stand establishment that is needed to plan important silvicultural interventions. Knowledge on tree diameters, a principal factor in the production phase, is required because the type of machinery need for production and the transportation of the wood must be suitable for the diameters in the stands (Bailey and Dell 1973). Information needed for forest management studies such as estimates of the energy that can be obtained from biomass, carbon, or wood for each diameter class can be easily obtained with the help of diameter distribution models (Borders et al. 2008). Bankston et al. (2021) stated that the diameter distribution of a stand also gives information about the ecological or economic value of this stand. While the diameter distributions of the stands can provide information about stand structures that benefit different wildlife characteristics, estimating future product quantity from the relevant stand and its distribution for different commercial standards (paper production, timber, etc.) is economically important and will inform management design. Actual diameter distributions of the stand are needed to accurately estimate of growth and yield, and design management plans suitable for that forest. Management decisions about the management of forests are shaped by the current and future conditions of the resources. Since the volume and other properties of the tree depend on the diameter of the tree, the diameter distribution can characterize the properties of a stand quite well (Bailey and Dell 1973).

Forest resource managers benefit from knowing how much stand volume is distributed into different size classes at any given time. Diameter distribution models, which employ a probability density function (PDF) to distribute a stand feature over size classes such as diameter at breast height (DBH) or tree height, are of great importance for forest management decisions. Although the PDF is a continuous function defined by a parameter vector that does not have any direct biological significance on its own, it is used in many forestry studies to model the distribution of trees in stands in different diameter or height classes. Since the studies analyzing the diameter distribution numerically in the early 1900s until today, many PDFs have been used in modeling diameters such as lognormal, Weibull, Gamma, normal and Johnson's S_B (Fonseca et al. 2009; Mirzaei et al. 2016; Sun et al. 2019; Ogana et al. 2020; Araújo et al. 2021). Because of the flexibility of Johnson's S_B distribution for the possible shape that a tree diameter distribution can produce, we selected it as the most appropriate for our study. In addition, the ease of integration of the cumulative density function (CDF) (Zhang et al.

2003) is a considerable advantage. Finally, the relative forest literature supports the Johnson's S_B distribution use for tree: (1) diameter description (Hafley and Schreuder 1977; Lopes 2001; Kiviste et al. 2003; Parresol 2003; Fonseca 2004; Özçelik et al. 2016; Ogana et al. 2017; Gorgoso-Varela et al. 2020; Vega et al. 2022), (2) height (Hafley and Schreuder 1977), and (3) volume ratio (Newberry and Burk 1985). However, it is a difficult procedure to select the appropriate theoretical distribution that can model the source of randomness appropriately in the study variable. For this reason, we used probability theory to express the likelihood of chance as a source for the difference in end outcome. On the other hand, a distribution's parameter recovery method has to be carefully selected, so as to produce reliable results. Furthermore, if there are no trees in some diameter classes, then both the sampling error and the sampling distribution of these classes may lead to a violation of regression assumptions. Free of such assumptions and due to its computational efficiency, in order to overcome the above difficulties, the support vector machine (SVM) learning methodology (Vapnik et al. 1997; Basak et al. 2007; García-Nieto et al. 2012; Gu et al. 2016) with the support vector regression (SVR) procedure was used to simulate as accurately as possible the empirical diameter distribution of the ground truth data to predict the structure of a stand. Previous studies highlighted the effectiveness of the support vector machine methodology for solving forest modeling problems. Guo et al. (2005) applied SVMs to predict the distribution of sudden oak death in California, Monnet et al. (2011) showed that the support vector regression models adequately predicted dominant tree height, basal area, mean diameter, and stem density, and Jiao et al. (2012) used SVR methodology to estimate timber volume. SVM approaches were also used by Diamantopoulou et al. (2018) to predict tree bark volume, by Nguyen et al. (2019) to classify tree species, and by Iizuka et al. (2022) to estimate DBH of Japanese cypress.

DBH values for trees are related to basal area, stand density and tree volume determination, and can be used for downscaling from stand-level information to tree level (Newton et al. 2005). However, the available information concerning the diameter distribution of Scots pine species is highly limited, and no such models have been developed for the Türkmen Mountain region. For this purpose, Johnson's S_B distribution was used due to its flexible structure in representing or explaining biological variables. The three-parameter recovery approach, which is based on the percentile moment method suggested by Fonseca et al. (2009), was used to estimate the model's parameters. Furthermore, the support vector machine for regression methodology was used for the simulation based on the empirical diameter distribution modeling. The use of support vector methodology is a fairly new approach for

diameter distribution modeling and has not been used for Scots pine stands.

Materials and methods

Study area

The study region is between 39°16' and 39°38' N and 30°06' and 30°36' E on Turkmen Mountain (Fig. 1). Rhyolite and dacite are the most common geological parent minerals in the study area; others are also present, e.g., basalt, claystone, and limestone. Grey brown and podsollic grey brown forest soils are the most common soil types (Güner 2006). The mean annual temperature is from 10.6 °C to 11.1 °C, and annual precipitation is between 374 and 562 mm, according to data from meteorological stations in Eskişehir, Kütahya, and Afyonkarahisar. In the Thornthwaite water-balancing system, the climate type of the research region ranges between semihumid and humid (Güner 2006). Scots pine dominates the research area. The other main plant species in the study area are Anatolian black pine, trembling poplar, and oriental beech.

In natural stands, sample plots were carefully selected to reflect all the existing variability such as site, stand age and stand density. Plot size varied between 200 and 300 m², depending on the density of the stands. All trees in the sample plots (total 1219) were numbered, then the

diameters of all trees at 1.30 m height were measured. The height of some trees was measured with an accuracy of 0.1 m using Blume-Leiss ALTimeter (Carl Leiss Berlin, Berlin, Germany) to estimate stand height. In addition, for each plot, variables such as average diameter (\bar{d}), quadratic mean diameter (d_g), number of trees per hectare (N), basal area per hectare (G), minimum diameter (d_{min}), maximum diameter (d_{max}), and median diameter ($d_{0,50}$) were calculated. Descriptive statistics (mean, maximum,

Table 1 Descriptive statistics for stand variables used to develop diameter distribution models (1219 trees from 55 plots)

Variable	Mean	Min	Max	SD
\bar{d} (cm)	23	17	31	3.66
d_g (cm)	23	18	32	3.69
$d_{0,50}$ (cm)	23	17	34	3.85
d_{min} (cm)	15	9	26	3.40
d_{max} (cm)	32	23	46	4.80
n (no. of trees/plot)	22	15	40	5.85
G (m ² ha ⁻¹)	45	24	67	8.89
N (no. of trees/hectare)	1092	500	2000	312.90
Size (m ²)	205	200	300	22.92

\bar{d} mean diameter, d_g , quadratic mean diameter, $d_{0,50}$ median diameter, d_{min} minimum diameter, d_{max} , maximum diameter, n number of trees per plot, G basal area, N number of trees per hectare, SD standard deviation

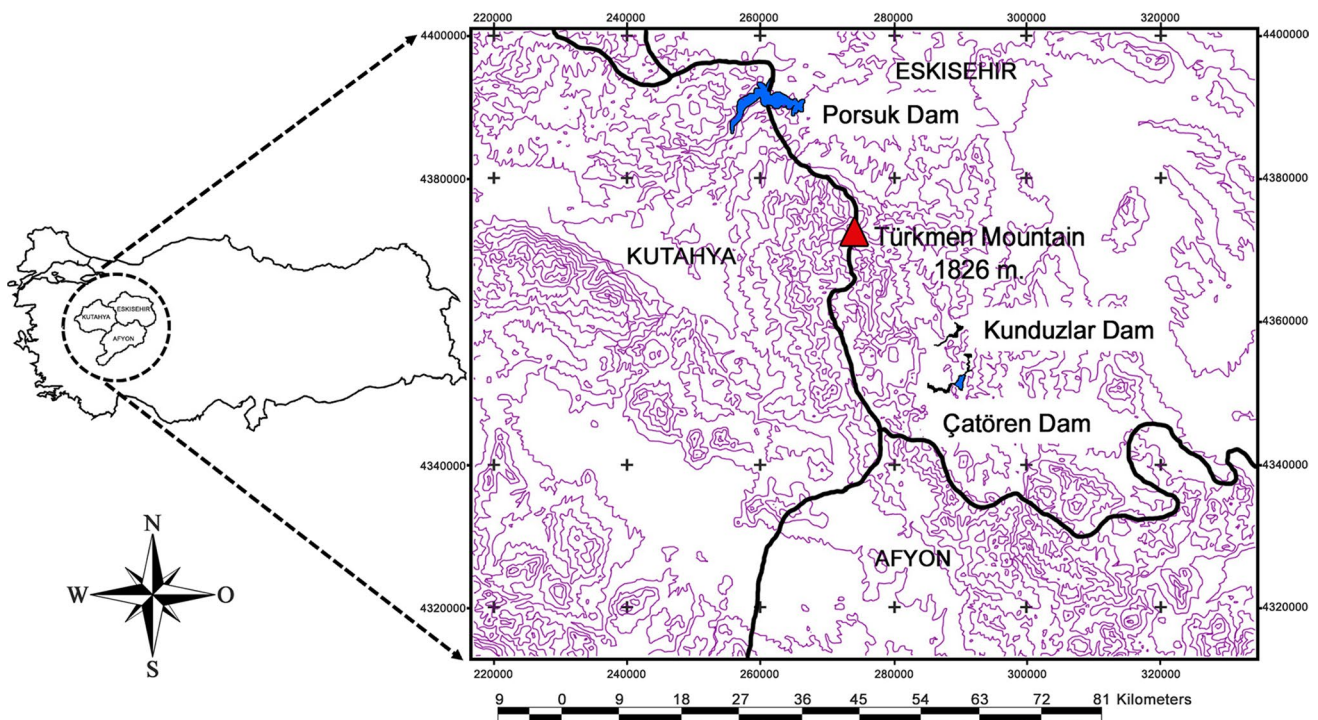


Fig. 1 Location of study area in Turkey

and minimum values and SD) of the main stand variables are given in Table 1.

The SB distribution and localization of the empirical distributions

Due to the moment statistics ability to provide sufficient information to construct a frequency distribution function, the third standardized moment (skewness, β_1) and the fourth standardized moment (kurtosis, β_2) were used to estimate and evaluate the asymmetry (left or right skewed distribution) and kurtosis (heavy- or light-tailed distribution) of our ground-truth data, respectively. To obtain the estimator ($\sqrt{b_1}$) of the coefficient of skewness, the centered third-order moment was divided by the sample standard deviation raised to the third power. The estimated skewness ($\sqrt{b_1}$) values ranged from -1.47 to 1.20 . The estimator (b_2) of kurtosis was obtained by dividing the centered fourth-order moment by the sample standard deviation raised to the fourth power. The estimated kurtosis values ranged from -1.50 to 4.36 . Figure 2 shows these estimates in the (β_1, β_2) space with two reference lines. β_1 represents the square of the standardized measure of skewness and β_2 is the standardized measure of kurtosis. Some combinations of β_1 and β_2 are mathematically impossible and occur above the $\beta_1 - \beta_2 - 1 = 0$ line. In Fig. 2, 54 empirical distributions are in the S_B region (some of which are quite close to the S_L region) and the remaining one is in the S_U region. Figure 2 shows that the diameter distribution for approximately 1.8% of the sample areas used in the study can be better represented by another distribution instead of Johnson’s S_B distribution.

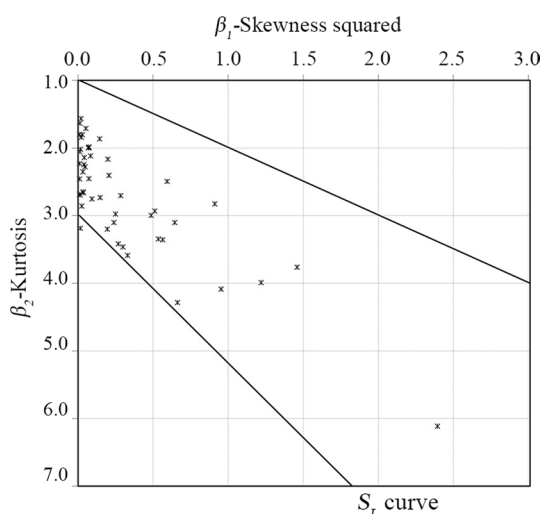


Fig. 2 Representation of the observations in the (β_1, β_2) space of skewness squared and kurtosis

Johnson’s S_B probability density function is one of the components of Johnson’s distribution function family, which was first introduced by Johnson (1949). This distribution system consists of S_U, S_L and S_B distributions which are for unbounded variates, variates bounded at one end, and bounded from below to above, respectively. The PDF for a variable X which follows an S_B PDF can be expressed as:

$$f(x) = \frac{\delta \lambda}{\sqrt{2\pi}(x - \xi)(\xi + \lambda - x)} \exp\left(-\frac{1}{2}\left[\gamma + \delta \ln\left(\frac{x - \xi}{\xi + \lambda - x}\right)\right]^2\right), \quad (1)$$

$$x \in (\xi, \xi + \lambda),$$

where, $f(x)$ is the probability density associated with diameter x ; $\lambda, \delta > 0, -\infty < \xi < \infty, -\infty < \gamma < \infty$. The parameter λ gives the range parameter, ξ is the location parameter and represents lower bond, δ and γ are shape parameters, and $\gamma = 0$ indicates symmetry.

The S_B PDF can represent variables that have natural or physical constraints on their range thanks to the lower (ξ) and upper ($\xi + \lambda$) bounds. Moreover, a remarkable amount of flexibility to fit a wide range of distribution can be achieved by two parameters (δ and γ) which control the shape. These attributes of the Johnson’s S_B PDF make this distribution system suitable for representing the biological variables (Fonseca et al. 2009).

The approaches commonly used to estimate Johnson’s S_B distribution parameters are the linear and nonlinear regression, maximum likelihood, percentile and the moment methods. However, for most of these techniques, knowledge of the distribution boundaries is essential. The use of Johnson’s S_B distribution for diameter distribution in growth and yield models and the use of parameter prediction or parameter recovery methods for parameter estimation have been tested, albeit in limited numbers (Özçelik et al. 2016). However, the parameter prediction method has some important disadvantages such as not being able to provide a match between the estimated stand value and the stand value obtained from the distribution, along with the possibility that a very small part of the variation in the parameters will be explained by the stand variables. For example, the shape parameter of the function shows a very weak relationship with age. Better results will be obtained with the parameter recovery-based approach (Scolforo et al. 2003; Fonseca 2004).

Three-parameter recovery method (3-PRM)

Extreme diameter values in the sample data set were used to estimate the location and range parameters of the Johnson’s S_B PDF distribution, which has been converted from a four-parameter to a three-parameter distribution. The

remaining parameters of the distribution model can be estimated using a percentile or moment method. Scolforo et al. (2003) defined a moment method to estimate the shape parameters. Parresol (2003) developed a percent-moment method for simultaneous solution of range and shape parameters. Alternatively, it can be used to directly estimate average stand properties using parameter recovery models and then to estimate the base diameter distribution. The parameter recovery method provides compatibility between the stand characteristics estimated from the regression and created from the distribution function. For applying the parameter recovery method, the equation system should include certain tree characteristics, and S_B parameters should be developed (Fonseca et al. 2009).

Parresol (2003) introduced a new estimation approach for the recovery of three parameters. In the method, a parameter recovery method has been developed for the range and shape parameters by using the median, first and second noncentral moments of the diameter distribution in general. In this approach, the location parameter is also estimated. Parresol (2003) estimated the location parameter ξ using a regression technique to extrapolate the random variable breast height diameter (d) to the lower bound. Using this estimated location parameter and the transformation given by Johnson and Kotz (1970), the two shape and range parameters can be solved. Although there is no closed form expression for the S_B probability density function (PDF), if the random variable is $X \sim S_B(\delta, \lambda, \gamma, \delta)$, where X is the diameter (Parresol 2003; Fonseca et al. 2009), then

$$z = \gamma + \delta \ln \left[\frac{(x - \xi)}{(\xi + \lambda - x)} \right] \sim N(0, 1) \tag{2}$$

Given a new variable,

$$y = \frac{(x - \xi)}{\lambda}, \tag{3}$$

It follows from Eq. 2 that

$$z = \gamma + \delta \ln \left[\frac{y}{(1 - y)} \right] \sim N(0, 1). \tag{4}$$

The new random variable Y will follow a distribution with the same shape parameters as X (Johnson and Kotz 1970). Using the Y random variable, the S_B PDF Eq. (1) becomes (Parresol 2003; Fonseca et al. 2009):

$$f(y) = \frac{\delta}{y(1 - y)\sqrt{2\pi}} \exp \left(-\frac{1}{2} \left[\gamma + \delta \ln \left(\frac{y}{1 - y} \right) \right]^2 \right), 0 < y < 1. \tag{5}$$

Setting z in Eq. (4) equal to 0 and rearranging in terms of parameter γ gives

$$\gamma = \delta \ln \left(\frac{1}{y_{0.50}} - 1 \right), \tag{6}$$

where $y_{0.50}$ is the median of Y .

Following the expected value of X^p in terms of the Y variable, then

$$\bar{d} = \xi + \lambda \mu'_1(Y) \tag{7}$$

$$dg^2 = E(\xi + \lambda Y)^2 = \xi^2 + 2\xi \lambda \mu'_1(Y) + \lambda^2 \mu'_2(Y) \tag{8}$$

where $\mu'_1(Y)$ and $\mu'_2(Y)$ are first and second noncentral moment of distribution of Y , respectively. \bar{d} is the average tree diameter. Equation 7 denotes average tree diameter (\bar{d}) as a function the first noncentral moment of Y . It is worth noticing that Eq. 8 is a function of the first two noncentral moments of Y . The quadratic mean diameter (dg) is functionally related to the number of trees per unit area (N) and basal area per unit area (G) : $G = kNdg^2$, where k is the conversion factor for the basal area per square meter ($k = \frac{\pi}{40000}$). Hence, Eq. 8 is equivalent to

$$G = kN[\xi^2 + 2\xi \lambda \mu'_1(Y) + \lambda^2 \mu'_2(Y)]. \tag{9}$$

Because z is a unit normal variance, the r th noncentral moment of Y is

$$\mu'_r(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left[1 + e^{-\frac{z-\gamma}{\delta}} \right]^{-r} e^{-z^2/2} dz \tag{10}$$

As indicated by Fonseca et al. (2009), the relationship in Eq. 6 is first used to eliminate γ in Eqs. 7 and 9 by substitution in Eq. 10. The solution system of the two equations and the two unknown parameters is nonlinear and must be solved by numerical procedures (see Parresol 2003 for details). Given the estimates of G , N , \bar{d} , median tree diameter ($d_{0.50}$), and location parameter, and Eqs. 7 and 9, the system of equations for δ and λ must be solved by iteration. The parameter γ is then found from Eq. 6.

As a result, ξ is predetermined, λ and δ are solved by iteration using Eqs. 7 and 9, and then the range parameter γ is solved with the help of Eq. 6. More details on technical solution for the three-parameter recovery approach can be found in Parresol (2003) and Fonseca et al. (2009). Parameters of Johnson's S_B distribution were estimated in SAS 9.2 (SAS Institute, Cary, NC, USA) with the parameter recovery method based on the percent-moment method. This program is implemented with the nonlinear Levenberg–Marquardt (NLPLM) method with the help of the interactive matrix language CAPABILITY subroutine (SAS Institute 2010). Detailed information can be found in Parresol et al. (2010).

Support vector regression models (SVRs)

The support vector regression modeling approach belongs to the type of supervised machine learning algorithms and can be considered as a generalization of the support vector machine (SVM) methodology for regression-type problems (Vapnik 1995, 1998, 1999, 2000; Vapnik et al. 1997). The SVR methodology represents a promising nonparametric learning algorithm with surprising characteristics such as the recognition of all patterns in the available data set. The methodology used in this paper is the SVR with the ϵ -insensitive loss function based on the determination of the support vectors (SVs) that produced an ϵ -insensitive tube, supported by the nonlinear kernel functions (Smola and Schölkopf 2004). That is, in order for the learning error to be disclosed for the nonlinear SVR, the input vector x (training sample) with a dimension mapped onto a higher dimensional space (let's say n -dimensional space) via the fixed nonlinear mapping function $\varphi(x) \in R^n$, a linear combination is constructed:

$$f(x) = [w, \varphi(x)] + b, \tag{11}$$

with $(x) \in R$, that due to $\varphi(x)$ leads to a nonlinear function. By ignoring the errors that are within the ϵ -insensitive tube, the function is minimized:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + \xi'_i \tag{12}$$

$$\text{with constraints : } \begin{cases} y_i - f(x) \leq \epsilon + \xi'_i \\ f(x) - y_i \leq \epsilon + \xi_i \\ \xi_i, \xi'_i \geq 0, i = 1, \dots, n \end{cases} \tag{13}$$

where y_i is the output value, w denotes the weight parameters, C is the parameter that represents the smoothness of the model, and ξ_i and ξ'_i are slack variables (Fig. 3), that show the deviation of points outside the ϵ -insensitive zone.

For the input vector x to be mapped onto a higher dimensional feature space, the Gaussian radial basis function (RBF) kernels was used:

$$k(x, x') = \exp \left(-\gamma_{SVR} \|x_i - x_j\|^2 \right), \gamma_{SVR} > 0 \tag{14}$$

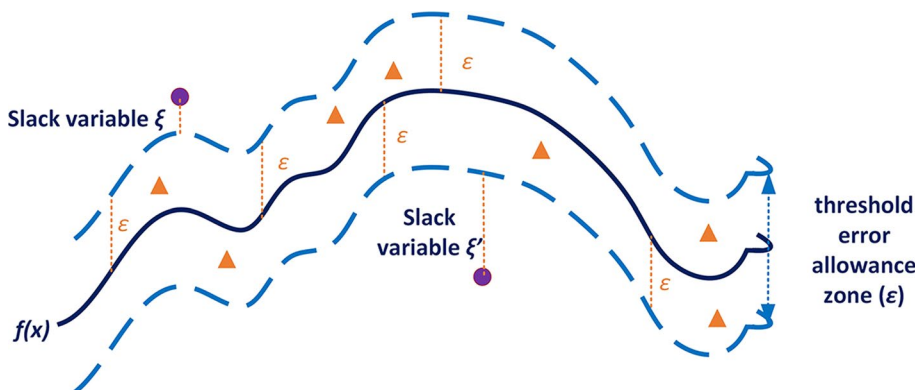
where $\gamma_{SVR} = (1/2\sigma^2)$ is the free parameter of the RBF kernels and $\|x_i - x_j\|$ is the Euclidean distance between the support vectors (SV).

From the above, it is clear that the robustness and the effectiveness of an ϵ -SVR model depends on three meta-parameters: ϵ , that determines the width of the ϵ -insensitive zone; γ_{SVR} , the parameter that depends on the variance of the training samples (σ^2) and thus sets the spread of the kernel; C , the cost parameter, (which can balance the resulting inaccuracy against the desired simplicity of the constructed ϵ -SVR model).

For the construction of the ϵ -SVR model, the programming language Python 3.9 (Van Rossum and Drake 2011; Python Software Foundation 2022) and libraries of scikit-learn (Pedregosa et al. 2011) were used. The state-of-the-practice approach utilized for a machine learning model construction (Olson and Delen, 2008), includes the 90% and 10% percentages division for the fitting and testing data sets, respectively. For this purpose, the available data set of the trees of the 55 sample plots was randomly divided into fitting (90% of the total sample plots) and testing (the remaining 10% of the total sample plots) samples to test the predictive ability of the constructed ϵ -SVR model for new, never-seen data in its construction phase. Furthermore, to prevent overfitting for the best generalization of the constructed ϵ -SVR model, the k =tenfold cross validation procedure was applied to the fitting data set.

Finally, for a quantitative approach of the concentration of the empirical data around the ϵ -SVR model fit, the root mean square error (RMSE) function from scikit-learn metrics library was used (Pedregosa et al. 2011) to produce a risk metric that corresponded to the expected values of the root squared (quadratic) error or loss by the ϵ -SVR model fit.

Fig. 3 Representation of ϵ -SVR mapping



Evaluation of the simulated distributions using different approaches

For testing the simulated diameter distributions using different approaches, the diameter class width of 5 cm was selected as the most accepted one. The error index (EI) introduced by Reynolds et al. (1988) was calculated related to the basal area of the diameter class as a function of weight as can be calculated as an exact value. Moreover, using the basal area as the weight factor ensures that the different dimensions of the tree bole would be considered in the economic evaluation of the trees (Fonseca et al. 2009; Özçelik et al. 2016). The formula for the error index (EI) is given as:

$$EI = \sum_{j=1}^M |G_j - \hat{G}_j| = \sum_{j=1}^M \left| \int_{C_j} \frac{\pi x^2}{4} [\hat{f}(x) - f_{emp}(x)] dx \right|, \quad (15)$$

where M is the number of diameter classes, G_i is the observed basal area of the j th diameter class, \hat{G}_j is the predicted basal area of the j th diameter class, C_j denotes diameter class number j , and \hat{f} is the function of the different approaches used for the diameter distribution simulation.

Results

For presenting the values of the parameter estimates obtained from the 55 sample plots using the three-parameter recovery method based on the percent-moment method, the parameter estimations obtained by the simulated Johnson’s S_B distribution for the first 10 sample plots are given in Table 2. Similar results were obtained for the rest of the plots. According to the three-parameter solution method introduced by Parresol (2003), two parameters (range and lambda) and both shape parameters (gamma and delta) of the four parameter Johnson’s S_B distribution that were estimated

for all sample plots, the L1-norm (output variable, which should have a value close to zero) values (Parresol et al. 2010), for almost all sample plots were quite small similar to the results given in Table 2. A delta value less than 0.7 generally indicates a bimodal distribution. More than half of the δ values for the 55 sample areas were less than 0.7. As seen in Fig. 2, Johnson’s S_B distribution cover a broad spectrum of shapes, fitting both positively and negatively skewed data. As indicated by Parresol (2003), because the Johnson’s S_B distribution is obtained by transformation of a standard normal variate, integration over specific classes can be accomplished by application of the well-tabulated standard normal. Further, the distribution can easily be extended to multivariate forms.

As for the simulated diameter distribution by the constructed ϵ -support vector regression (ϵ -SVR) model using the radial basis function (RBF) kernel, the best combination of the set of ϵ , the γ_{SVR} parameter, and the cost parameter (C) was explored (Fig. 3) by the grid search method (Pedregosa et al. 2011). The grid-search method was used to test and evaluate all possible combinations of the parameters’ values that comprised the grid. Specifically, the tested ϵ values ranged from 0.00 to 0.45 in steps of 0.01, γ_{SVR} from 0.00 to 1.00 in steps of 0.01, while the tested C values ranged from 4 to 30 in steps of 1.

The root mean square error (RMSE; as the square root of the mean of the squares of the deviations between the observed and the predicted by the ϵ -SVR model diameter values) was used as the statistical evaluation criterion that represented the adaptation of the model to the fitting data set along with its predictive ability to the testing data set, respectively. This measure was used for a quantitative approach of the empirical data concentration around the ϵ -SVR model fit. Finally, the combination that gave the smallest root mean square errors for both the fitting and the testing data set was $\epsilon = 0.01$, $\gamma_{SVR} = 0.01$ and $C = 10$, with RMSE

Table 2 Parameter estimates for the first 10 sample plots obtained with Johnson’s S_B PDF

Plot	XI*	λ	γ	δ	L1_Norm	Converged	RC*
1	19	11.86663	1.20472	0.75483	0.036319	Yes	3
2	14	13.65007	0.18093	0.74471	2E-10	Yes	3
3	11	12.47177	-0.28727	0.69875	1.55E-07	Yes	3
4	9	16.42203	0.02198	0.42758	3.3E-07	Yes	3
5	10	12.73792	-0.09352	0.26050	6.39E-06	Yes	3
6	10	18.91966	-0.09684	0.84708	5.92E-06	Yes	3
7	11	13.56259	-0.03630	0.56251	3.68E-06	Yes	3
8	14	75.59411	3.69980	1.39746	0.022806	Yes	3
9	15	61.26548	2.61986	1.17991	0.183885	Yes	3
10	16	36.06415	0.97485	1.01762	1.52E-10	Yes	3

*XI: Location parameter; RC: SAS return code from IML routine NLPLM indicating which converge criteria was met or reason for non convergence. Positive value indicates successful termination, while a negative value indicates unsuccessful termination

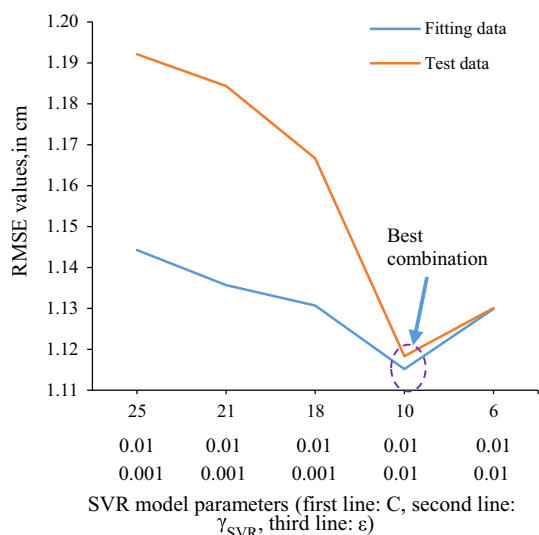


Fig. 4 Best combination of ϵ -SVR parameters as produced by the grid search method

equal to 1.1152 and 1.1183 for the fitting (training plus validation data sets) and the test data sets, respectively (Fig. 4).

Although the behavior of the simulated distributions by both approaches was evaluated for the total available data set of the 1219 trees, it was considered as significant information, and the specific simulation ability of the different approaches according to the small number of trees of each plot separately was also evaluated. The diameter distribution simulation of both approaches for plots 1, 2, and 5–10 plots in Table 2 is shown in Fig. 5. As can be seen, the simulation of both models is more or less acceptable for all plots, with the Johnson’s S_B distribution outperforming for plots 1, 2 and 5.

The experimental distribution for the sample plot 10 showed a typical right-tie flat distribution while the simulated ϵ -SVR distribution was a multi-modal distribution. As can be seen, the ϵ -SVR model showed the ability to capture the diameter distribution pattern of plot 10, except for the case of the 25 cm diameter class (Fig. 5).

For examining the reliability of the simulated distribution using both Johnson’s S_B distribution and the constructed SVR model, plots of the experimental distributions and the observed distributions for sample plots 4, 15, and 35, are given in Fig. 6. The motivation behind the usage of these plots was to explore the adaptation of the simulated distributions to the actual slightly skewed distribution. The experimental Johnson’s S_B distribution was constructed with δ values less than 0.7 and, according to the 35-sample plot, γ value equal to 0.02. On the other hand, delta, and gamma values for the sample plot 35 showed a left-skewed bimodal distribution with $\delta=0.5$ and $\gamma=-0.59$ (Fig. 6). As can be seen (Fig. 6) the ϵ -SVR constructed model gave well-shaped

curves that followed the original curve shape. As for the 35-plot, the constructed ϵ -SVR distribution showed a significant peak of diameter in the 27.5–32.5 cm diameter class. Despite this fact, the distribution simulation by the SVR model can be considered as a successful diameter distribution.

The shape of the 18- and 26-plot distributions are shown in Fig. 7. As can be seen (Fig. 7), the simulated distributions for the 18-plot were sufficient, producing unimodal and uniform distribution patterns for both approaches, while for plot 26, Johnson’s S_B simulation produced a bimodal distribution. According to plot 26, for the estimated distribution obtained with the help of Johnson’s S_B distribution, the peak values are at 15 cm and 30 cm, while in the observed distribution, the peak value is at 25 cm. As for the SVR approach, the simulated peak was followed the actual distribution’s peak (plot 26). Although for both approaches, partial differences between the observed distribution and the estimated distribution were observed, the diameter distribution was sufficient.

However, in some sample areas with quite a small number of trees, the estimates with both approaches (Johnson’s S_B distribution and SVR model) seem to be quite unsuccessful as for plot 16 (Fig. 8). Quite large differences were observed between the values for estimated distributions and the observed distribution for the sample plot area 16, which includes measured diameters for 15 trees. On the contrary, according to those sample plots that include a relatively large number of trees, for example, plot 3, with measured diameters of 33 trees, the simulated distributions for both approaches were considered as adequate (Fig. 8).

For evaluating the simulation behavior of the methods applied, the summary statistics of the error index (EI) values for both approaches are given in Table 3. The EI values for 55 sample plots ranged from 2.67 to 28.57 and 4.92 to 13.76 with mean error index values equal to 13.61 and 9.25 and median values equal to 13.88 and 9.27 for the simulation using Johnson’s S_B distribution and the ϵ -SVR model, respectively. Furthermore, the proportion of plots that produced lower values of error index was found equal to 72.73% for the ϵ -SVR approach, meaning that the diameter distribution simulation by the constructed ϵ -SVR model was more reliable than the simulation derived by Johnson’s S_B distribution for 45 plots, while according to the remaining 15 plots, the Johnson’s S_B distribution produced simulation with better accuracy.

For both methods, the EI values for each sample plot are given in Fig. 9a and the frequency distribution of the EI values in Fig. 9b; the error distribution values for the sample plots can thus be seen to vary. Johnson’s S_B distribution gave larger errors for most of the plots than those derived from the simulated distribution using the constructed ϵ -SVR model (Fig. 9a). In general, as seen from Fig. 9b,

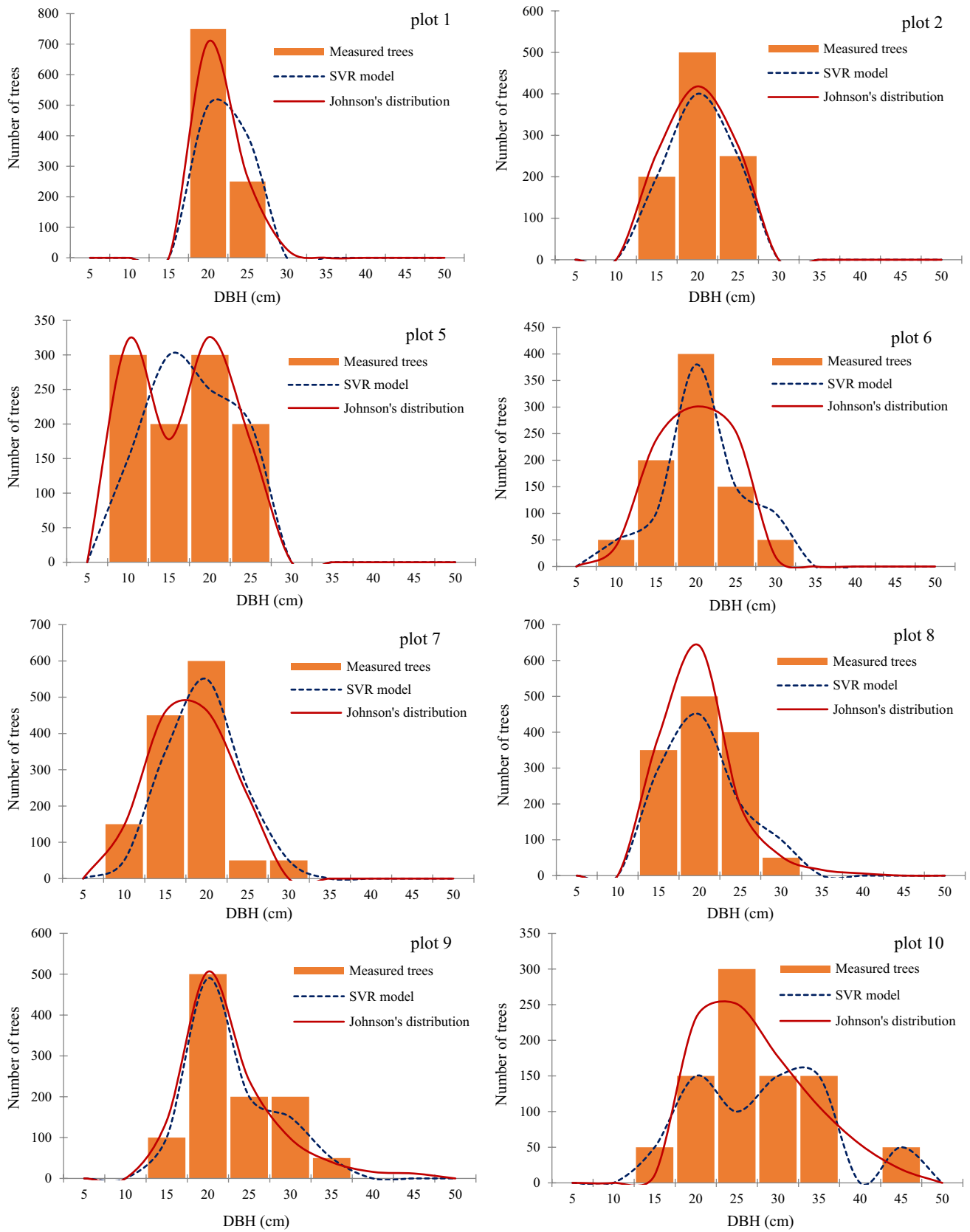


Fig. 5 Observed and simulated diameter distributions by both methods for sample plots 1, 2 and 5–10

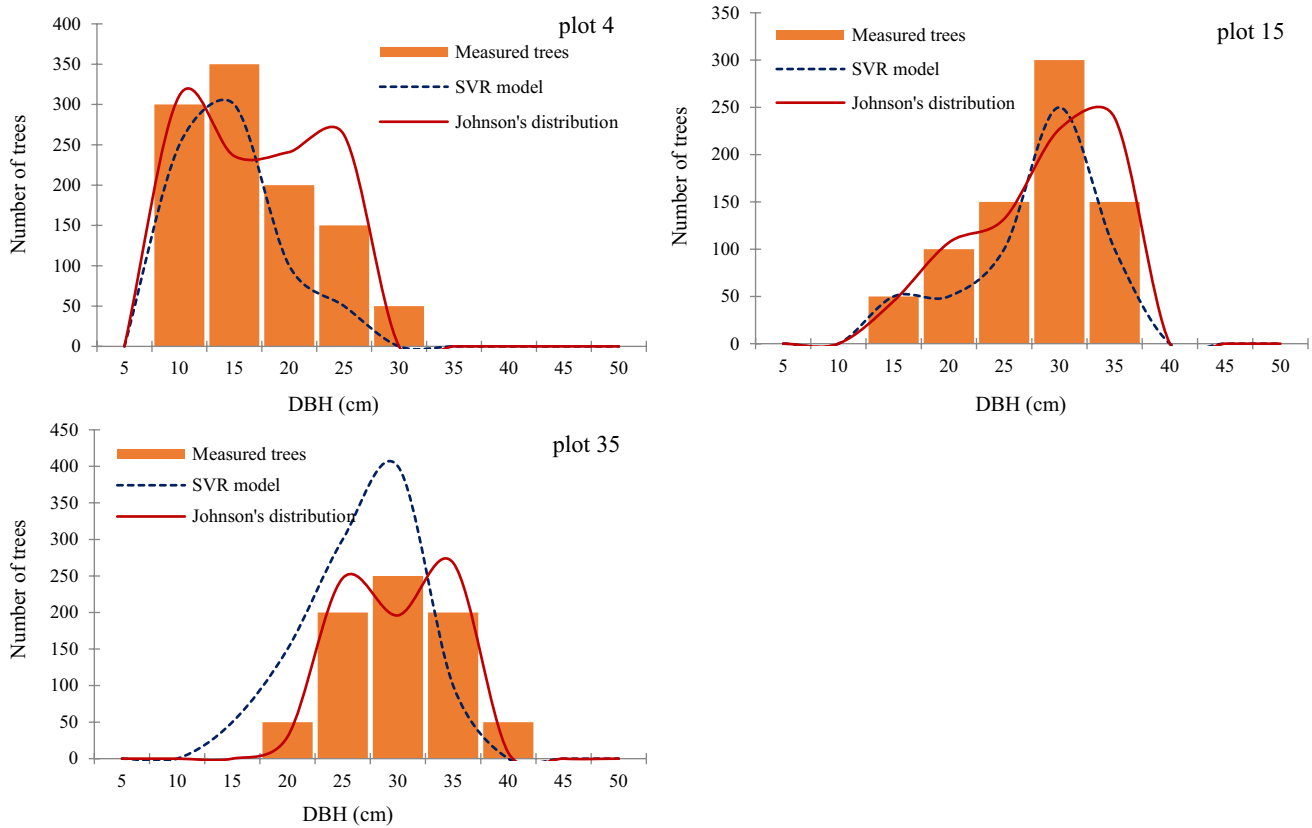


Fig. 6 Observed and simulated diameter distributions by the SVR model and Johnson's S_B distribution for sample plots 4, 15, and 35

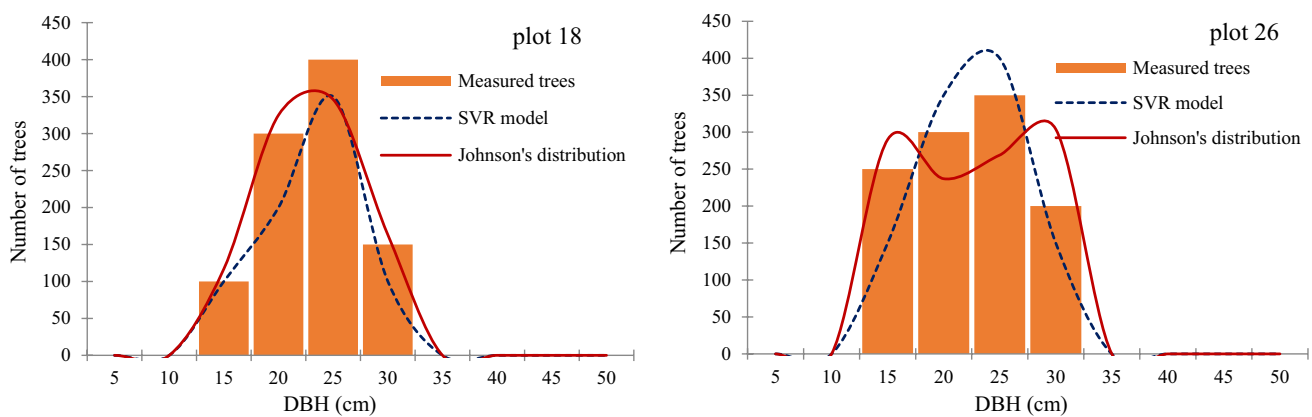


Fig. 7 Observed and simulated diameter distributions by both methods for sample plots 18 and 26

a significant part of the EI values for the sample plots are between 4 and 26 m^2/ha for Johnson's S_B distribution and between 4 and 14 m^2/ha for the ε -SVR model, meaning that the ε -SVR model adequately simulated the diameter distribution for most of the sample plots.

Discussion

To explore the use of a machine learning model, such as a support vector regression model, to reliably describe the diameter distribution of a natural Scots pine stand, as an alternative procedure to the known, accepted theoretical distribution method, such as the Johnson's S_B distribution, we used a three-parameter recovery method based on the

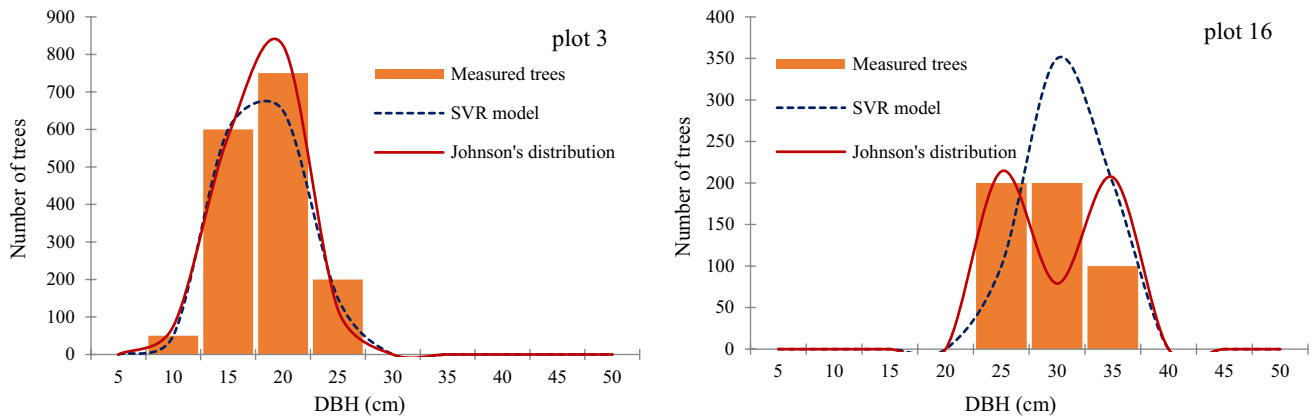


Fig. 8 Observed and simulated diameter distributions by both methods for sample plots 3 and 16

Table 3 Summary statistics of the error index (EI) values for the different methods

Method	Mean	Minimum	Maximum	SD	Median
Johnson's S_B distribution (3-PRM)	13.61	2.67	28.57	6.62	13.88
ϵ -SVR	9.25	4.92	13.76	1.91	9.27

percentile-moment method to estimate parameters of Johnson's S_B PDF. This distribution was preferred because it allows modeling of different distribution patterns due to its two shape parameters and enables better representation of biological variables (Fonseca et al. 2009; Özçelik et al. 2016). Many studies have shown that Johnson's S_B distribution produces remarkably successful results in describing the diameter distributions (Kiviste et al. 2003; Parresol 2003; Lei 2008; Fonseca et al. 2009; Mateus and Tomé 2011; Özçelik et al. 2016).

Further, due to its nonparametric nature and ability to reveal unknown relationships among real world data, the machine learning method support vector regression (SVR) was used (Wang et al. 2009; Alonso et al. 2013; Diamantopoulou et al. 2018). Specifically, the non-linear ϵ -SVR algorithm, including the radial basis function (RBF) kernel was used to simulate the distribution of diameter classes of pine trees. Because of the discontinuous nature of the number of trees in each diameter class, the continuous numbers produced by the ϵ -SVR model were considered equal to the nearest integer.

Data required in this study were obtained from various Scots pine stands in the Türkmen Mountain region. The adaptation of the simulated distributions developed with the Johnson's S_B PDF and the ϵ -SVR model in this study was evaluated using error index values. For this purpose, diameter classes of 5-cm intervals were created. The basal area is used as a weight function for the error index calculation, since, on one hand, it can be calculated as an exact value

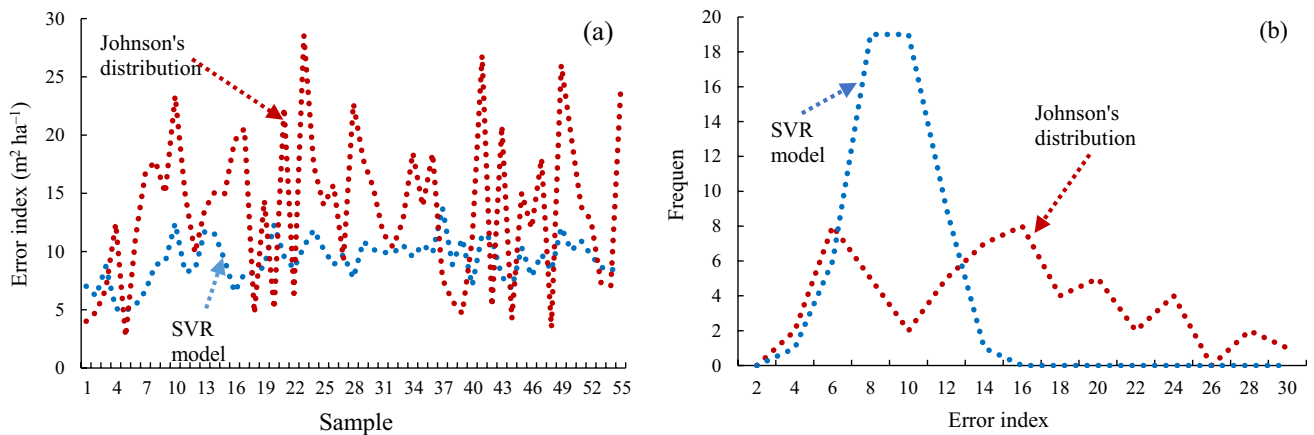


Fig. 9 Error index values for each plot (a) and their frequency distribution (b) using the SVR model and Johnson's S_B method

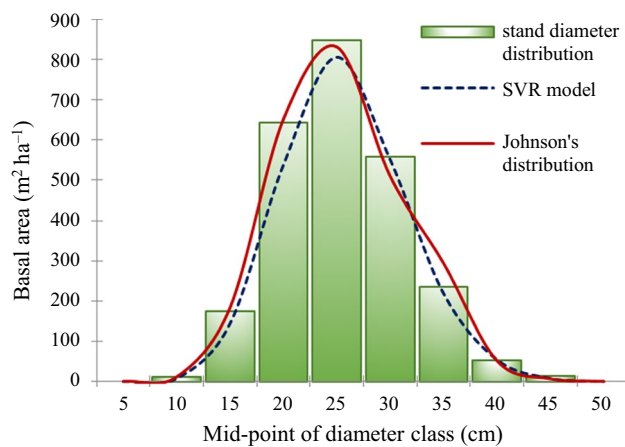


Fig. 10 Simulated distributions of basal areas for the whole tree stand using the SVR model and Johnson's S_B method

and, on the other hand, it considers dimensional differences between trees (Fonseca et al. 2009; Özçelik et al. 2016).

The statistical evaluations revealed that the diameter distribution of the natural Scots pine stands in the Türkmen Mountain region can be modeled successfully with both modeling approaches tested. According to the simulated distributions for the whole pine stand (Fig. 10) examined in this study, it is obvious that both simulated distributions can reliably estimate the basal area of the forest per diameter class and per hectare.

The diameter distribution of some stands was bimodal, others were unimodal, and some were right or left skewed. Due to their features, both the simulated distributions have been found to be successful approaches in describing the ground truth behavior of the tree diameter distribution. Specifically, the most important advantage of estimating the parameters of Johnson's S_B distribution with the parameter recovery method is that it allows estimating the future diameter distributions in growth and yield models, by producing significant information about forecasting the number of trees in a stand by diameter classes. However, the maximum likelihood method cannot be used directly for this purpose. According to the support vector regression approach, it showed great ability for describing the tree diameter distribution, providing a significant alternative to theoretical distribution.

The selection of the appropriate methodology that enables unbiased and accurate diameter distribution simulations requires a multifaceted design that takes into account the nature of the model and its applicability both in the field and in the office. In addition, this decision also requires setting the appropriate priorities for prediction accuracy and convenience. In this context, the two approaches compared in this study make use of the same sample size and data and therefore involve the same field effort. Given the obtained

results, the ϵ -SVR approach enables the adequate capture and simulation of the complex, nonlinear structure of the diameter distributions without the need for first specifying the model form. This is not the case for other nonlinear modeling techniques. As far as office work is concerned, both modeling approaches require extensive knowledge, programming skills, and the corresponding effort to apply the constructed approach. Specifically, given the SVR modeling approach, it is worth noting that the ϵ -SVR model was constructed using the freely available programming language, (Python). Using Python's export capabilities, the ϵ -SVR models can be exported into file(s) and thus very easily loaded and used by a third-party user. Of course, requirements such as proper equipment or skills that are needed for the application of the ϵ -support vector regression models are not always met in practical forestry. Despite this fact, potential users can gain the experience, knowledge, and accuracy needed for predictions and using advanced models. When convenience is the limiting factor in a survey, loss of prediction accuracy should be seriously taken into account.

Conclusions

For the natural Scots pine forests in the Türkmen Mountain region, which have quite different stand structures due to ongoing destruction over many years, both parametric and nonparametric modeling approaches (Johnson's S_B function and the support vector regression (SVR) procedure, respectively) were tested as accurate and reliable diameter distribution models that can easily fit an empirical diameter frequency distribution. Johnson's S_B probability density function based on the percentile moment approach, due to its flexibility according to the possible shape that a tree diameter distribution, produced a reliable tool to describe the empirical diameter distribution of the pine trees. However, a predetermined function is a requirement for this approach. On the other hand, the data-driven support vector regression modeling, a nonparametric supervised learning algorithm with the ϵ -insensitive loss function supported by nonlinear kernel functions, provided the desired characteristics (e.g., recognition of all patterns in the available data set) and insignificant ones.

Users can gain the experience, knowledge, and prediction accuracy needed to use advanced models such as the ϵ -SVR model. In other words, the ϵ -SVR algorithm has potential to accurately simulate the diameter distribution dynamics and thus can be safely used as a data-driven alternative for the efficient management of a forest ecosystem.

Acknowledgements The collection of the data for this research was supported by Turkish General Directorate of Forestry.

Author contributions ŞTG: aided fieldwork, data curation, manuscript review and editing; MJD and RÖ: analyzed primary data, designed methodology, did modeling work, evaluated results, did the manuscript writing, review, and editing.

Funding Open access funding provided by HEAL-Link Greece.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alonso J, Castañón AR, Bahamonde A (2013) Support vector regression to predict carcass weight in beef cattle in advance of the slaughter. *Com Elec Agric* 91:116–120. <https://doi.org/10.1016/j.compag.2012.08.009>
- Araújo LA, Oliveira RM, Dobner M, Jarochinski e Silva CS, Gomide LR, (2021) Appropriate search techniques to estimate weibull function parameters in a *Pinus* spp. *Plantation J Forestry Res* 32:2423–2435. <https://doi.org/10.1007/s11676-020-01246-z>
- Bailey RL, Dell TL (1973) Quantifying diameter distributions with the weibull function. *For Sci* 19(2):97–104. <https://doi.org/10.1093/forestscience/19.2.97>
- Bankston JB, Sabatia CO, Poudel KP (2021) Effects of sample plot size and prediction models on diameter distribution recovery. *For Sci* 67(3):245–255. <https://doi.org/10.1093/forsci/xfaa055>
- Basak D, Pal S, Patranabis DC (2007) Support vector regression. *Neural information processing—Letters and reviews*. 10(11): 203–224.
- Borders BE, Wang ML, Zhao DH (2008) Problems of scaling plantation plot diameter distributions to stand level. *For Sci* 54(3):349–355. <https://doi.org/10.1093/forestscience/54.3.349>
- Diamantopoulou MJ, Özçelik R, Yavuz H (2018) Tree-bark volume prediction via machine learning: a case study based on black alder's tree-bark production. *Comput Electron Agric* 151:431–440. <https://doi.org/10.1016/j.compag.2018.06.039>
- Fischer R, Lorenz M, Köhl M, Becher G, Granke O, Christou A (2008) The conditions of forests in Europe: 2008 executive report. United Nations Economic Commission for Europe, Convention on Long-range Trans boundary Air Pollution, International Co-operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests (ICP Forests). 23 p.
- Fonseca TF, Marques CP, Parresol BR (2009) Describing maritime pine diameter distributions with Johnson's S_B distribution using a new all-parameter recovery approach. *For Sci* 55(4):367–373. <https://doi.org/10.1093/forestscience/55.4.367>
- Fonseca TF (2004) Modelação do crescimento, mortalidade e distribuição, do pinhal bravo no Vale do Tamega. Ph.D. dissertation Univ. of Trás-os-Montese e Alto Douro, Vila Real, Portugal, 248 p.
- García-Nieto PJ, Martínez-Torres J, Araújo-Fernández M, Ordóñez-Galán C (2012) Support vector machines and neural networks used to evaluate paper manufactured using *Eucalyptus globulus*. *Appl Math Model* 36(12):6137–6145. <https://doi.org/10.1016/j.apm.2012.02.016>
- GDF (2015) Forest Resources of Turkey. The General Directorate of Forests, Ankara. (in Turkish)
- Gorgoso-Varela JJ, Ogana FN, Ige PO (2020) A comparison between derivative and numerical optimization methods used for diameter distribution estimation. *Scand J for Res* 35:156–164
- Gu YH, Yoo SJ, Park CJ, Kim YH, Park SK, Kim JS, Lim JH (2016) BLITE-SVR: new forecasting model for late blight on potato using support-vector regression. *Comput Electron Agric* 130:169–176. <https://doi.org/10.1016/j.compag.2016.10.005>
- Güner ŞT (2006) Determination of Growth Nutrition Relationships as Related to Elevation in Scots Pine (*Pinus sylvestris* ssp. *hamata*) Forests on Turkmen Mountain (Eskişehir, Kütahya). PhD Dissertation, Anadolu University, Graduate School of Sciences, Eskişehir. (in Turkish)
- Guo QH, Kell M, Graham CH (2005) Support vector machines for predicting distribution of sudden oak death in California. *Ecol Model* 182:75–90. <https://doi.org/10.1016/j.ecolmodel.2004.07.012>
- Hafley W, Schreuder H (1977) Statistical distributions for fitting diameter and height data in even-aged stands. *Can J Forestry Res* 7(3):481–487. <https://doi.org/10.1139/x77-062>
- Iizuka K, Kosugi Y, Noguchi S, Iwagami S (2022) Toward a comprehensive model for estimating diameter at breast height of Japanese cypress (*Chamaecyparis obtusa*) using crown size derived from unmanned aerial systems. *Comput Electron Agric* 192:106579. <https://doi.org/10.1016/j.compag.2021.106579>
- SAS Institute (2010) SAS/OR(R) 9.2 User's Guide: Mathematical Programming. <<http://support.sas.com/documentation/cdl/en/orm-pug/59679/HTML/default/viewer.htm#optmodel.htm>, accessed on July 2022>.
- Jiao YQ, Zhao LX, Deng O, Xu WH, Feng ZK (2012) Calculation of live tree timber volume based on particle swarm optimization and support vector regression. *Trans Chin Soc Agric Eng* 29(20):160–167 ((in Chinese))
- Johnson NL (1949) Systems of frequency curves generated by methods of translation. *Biometrika* 36(1–2):149–176. <https://doi.org/10.2307/2332539>
- Johnson NL, Kotz S (1970) Continuous Univariate Distributions, vol 1. John Wiley & Sons, New York, NY
- Kiviste A, Nilson A, Hordo M, Merenäk M (2003) Diameter distribution models and height-diameter equations for Estonian forests. *Modelling forest systems*. CAB International Publishing. pp 169–179.
- Lebedev AV (2022) Changes in the growth of scots pine (*Pinus sylvestris* L.) stands in an urban environment in European Russia since 1862. *J Forestry Res*. <https://doi.org/10.1007/s11676-022-01569-z>
- Lei Y (2008) Evaluation of three methods for estimating the Weibull distribution parameters of Chinese pine (*Pinus tabulaeformis*). *J for Sci* 54(12):566–571
- Lopes S (2001) Modelação matemática da distribuição de diâmetros em povoamentos de pinheiro bravo. PhD dissertation, Faculdade de Ciências e Tecnologia da Universidade de Coimbra. (in Portuguese)
- Mateus A, Tomé M (2011) Modelling the diameter distribution of eucalyptus plantations with Johnson's S_B probability density function: parameters recovery from a compatible system of equations to predict stand variables. *Ann for Sci* 68(2):325–335. <https://doi.org/10.1007/s13595-011-0037-7>
- Mirzaei M, Aziz J, Mahdavi A, Rad AM (2016) Modeling frequency distributions of tree height, diameter and crown area by six probability functions for open forests of *Quercus persica* in Iran. *J Forestry Res* 27:901–906. <https://doi.org/10.1007/s11676-015-0194-x>
- Monnet J, Chanussot J, Berger F (2011) Support vector regression for the estimation of forest stand parameters using airborne laser

- scanning. *IEEE Geosci Remote Sens Lett* 8(3):580–584. <https://doi.org/10.1109/LGRS.2010.2094179>
- Newberry JD, Burk TE (1985) S_B distribution-based models for individual tree merchantable volume-total volume ratios. *For Sci* 31(2):389–398. <https://doi.org/10.1093/forestscience/31.2.389>
- Newton PF, Lei Y, Zhang SY (2005) Stand-level diameter distribution yield model for black spruce plantations. *Forest Ecol Manag* 209(3):181–192. <https://doi.org/10.1016/j.foreco.2005.01.020>
- Ogana FN, Itam ES, Osho JSA (2017) Modeling diameter distribution of *Gmelina arborea* plantation in Omo Forest Reserve, Nigeria with Johnson's S_B . *J Sustain Forest* 36:121–133
- Ogana FN, Gorgoso-Varela JJ, Osho JSA (2020) Modelling joint distribution of tree diameter and height using Frank and Plackett copulas. *J Forestry Res* 31:1681–1690. <https://doi.org/10.1007/s11676-018-0869-1>
- Özçelik R, Fonseca TF, Parresol BR, Eler Ü (2016) Modeling the diameter distributions of brutian pine stands using Johnson's S_B distribution. *For Sci* 62(6):587–593. <https://doi.org/10.5849/forsci.15-089>
- Parresol BR (2003) Recovering parameters of Johnson's S_B distribution. *US For Serv Res Paper SRS-31*:9p
- Parresol BR, Fonseca TF, Marques CP (2010) Numerical details and SAS programs for parameter recovery of the S_B distribution. *US For Serv Gen Tech Rep SRS-122*. USDA, USA, 27pp. <https://doi.org/10.2737/SRS-GTR-122>
- 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. *J Mach Learn Res* 12:2825–2830
- Python Software Foundation (2022) Python 3.9. <https://docs.python.org/3.9/index.html>
- Reynolds MR, Burk TE, Huang WC (1988) Goodness-of-fit tests and model selection procedures for diameter distribution models. *For Sci* 34(2):373–399. <https://doi.org/10.1093/forestscience/34.2.373>
- Van Rossum G, Drake FL (2011) *The Python Language Reference Manual*. Network Theory Ltd. 150 p.
- Scolforo JRS, Tabai FCV, de Macedo RLSG, Acerbi FW, de Assis AL (2003) S_B distribution's accuracy to represent the diameter distribution of *Pinus taeda*, through five fitting methods. *For Ecol Manag* 175(1):489–496. [https://doi.org/10.1016/S0378-1127\(02\)00183-4](https://doi.org/10.1016/S0378-1127(02)00183-4)
- Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14(3):199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- Sun S, Cao QV, Cao T (2019) Evaluation of distance-independent competition indices in predicting tree survival and diameter growth. *Can J For Res* 49(5):440–446. <https://doi.org/10.1139/cjfr-2018-0344>
- Vapnik V (1995) *The Nature of Statistical Learning Theory*. Springer, New York
- Vapnik V (1998) *Statistical Learning Theory*. Wiley-Interscience, New York
- Vapnik V (2000) *The Nature of Statistical Learning Theory*, 2nd edn. Springer-Verlag, New York, NY, USA
- Vapnik V, Golowich S, Smola A (1997) Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing, in M. Mozer, M. Jordan, and T. Petsche (eds.), *Neural Information Processing Systems*, Vol. 9. MIT Press, Cambridge, MA.
- Vapnik V (1999) Three Remarks on The Support Vector Method of Function Estimation. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods—Support Vector Learning*, pages 25–42, Cambridge, MA, 1999. MIT Press.
- Vega AA, Corral-Rivas S, Corralrivas JJ, Dieguez-Aranda U (2022) Modeling diameter distribution of natural forest in PuebloNuevo DurangoSate. *Rev Mex de Cienc For* 13:73
- Wang DC, Wang MH, Qiao XJ (2009) Support vector machines regression and modeling of greenhouse environment. *Comput Electron Agric* 66(1):46–52. <https://doi.org/10.1016/j.compag.2008.12.004>
- Zhang LJ, Packard KC, Liu CM (2003) A comparison of estimation methods for fitting Weibull and Johnson's S_B distributions to mixed spruce fir stands in northeastern North America. *Can J Forest Res* 33(7):1340–1347. <https://doi.org/10.1139/x03-054>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.