FOLIA OECOLOGICA – vol. 49, no. 2 (2022), doi: 10.2478/foecol-2022-0010

Simulation of over-bark tree bole diameters, through the RFr (Random Forest Regression) algorithm

Maria J. Diamantopoulou*

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

Abstract

Diamantopoulou, M.J., 2022. Simulation of over-bark tree bole diameters, through the RFr (Random Forest Regression) algorithm. *Folia Oecologica*, 49 (2): 93–101.

The difficulty of locating and measuring the over-bark tree bole diameters at heights that are far from the ground, is a serious problem in ground-truth data measurements in the field. This problem could be addressed through the application of intelligent systems methods. The paper explores the possibility of applying the Random Forest regression method (RFr) in order to assess, as accurately as possible, the size of the tree bole diameters at any height above the ground, considering data that can be easily measured in the field. For this purpose, diameter measurements of pine trees (*Pinus brutia* Ten.) from the Seich–Sou urban forest of Thessaloniki, Greece, were used. The effectiveness of the Random Forest regression technique is compared with the results of non-linear regression models that fitted to the available data and evaluated. This research has shown that the RFr method can be a reliable alternative methodology in order to receive accurate information provided by the model, saving time and effort in field.

Keywords

artificial intelligence, non-linear regression, over bark tree bole diameters, *Pinus brutia*

Introduction

The measurements of trees are fundamental in the practice of forestry and forest science. The measurements are used both to understand the forest dynamics and to ensure that it is properly managed. The size of the tree bole diameters, at various heights from the ground, can outline the overall growth of the tree trunk. The knowledge of the diameter dimensions across the tree bole provides the necessary information needed for the accurate volume computation of the standing tree boles, using existing theoretical cross-sectional formulas, such as Smalian's, Huber's, Newton's methods, etc. (Avery and Burkhart, 2002; Van Laar and Akça, 2007; West, 2009). However, the difficulty of locating and measuring over-bark tree bole diameters at many heights that are far from the ground is a well-known subject. For this reason, the problem of finding appropriate estimation models for difficult-to-measure biological variables, is a field of intensive research in forest science.

The knowledge of the size of the diameters of a standing tree bole is essential, contributing directly and substantially to the accurate estimation of the tree bole volume, which depends on the number of known diameters at various heights. Moreover, this knowledge indirectly contributes to the description of the structure of the forest sections (Matis, 2004; West, 2009), while, at the same time, provides an essential information for the proper management of the forest ecosystem.

The most widespread methodology of producing estimation models is through the application of the regression analysis technique (Draper and Smith, 1998; Koulelis and Ioannidis, 2021). It is known that regression models give estimation accuracy with relatively small estimation errors. However, the prerequisite conditions for the application of the theory of regression modeling should be approached by the data in hand, with sufficient accuracy (Draper and Smith, 1998; Arkes, 2019). That is, problems that arise should be identified and addressed, such as: a) the unstable estimation of regression coefficients, b) incorrect decisions of parameters hypothesis tests, c) incorrect signs of regression coefficients, d) biased selection of variables of the model, e) the lack of observation independencies, etc. (Ratkowsky, 1990; Draper and Smith, 1998).

*Corresponding author:
e-mail: mdiamant@for.auth.gr

Due to its importance, the study and the construction of taper functions is a very interesting topic in forest research. Till today, the proposed taper models included polynomial, sigmoid, principal component analysis (PCA), and linear mixed functions, while contemporary machine learning (ML) approaches being explored (Salekin et al., 2021). According to the use of primary data from pine trees, the Max and Burkhart segmented taper equation under the nonlinear mixed-effects modeling technique was extensively used (Tricando and Burkhart, 2006; Özçelik and Alkan, 2020). Nicoletti et al. (2020) evaluated the accuracy of bivariate and generalized linear mixed modeling in the representation of the *Pinus taeda* L. trunk taper, while many researchers tested parametric and semi-parametric models for constructing reliable taper models (Özçelik et al., 2016; Algera et al, 2019; Marchi et al., 2020). As a step further, the scientific community that conducted research in this area, tested nonparametric modeling approaches, such as machine learning techniques and reported that it worth considering these methods as reliable alternatives to traditional taper-based equations (Leite et al., 2011; Nunes and Görgens, 2016; Özçelik et al., 2019).

When biological data are analyzed and modeled, such as data from tree measurement in the forest environment, the violation of the regression modeling prerequisites is a frequent phenomenon constituting a serious obstacle in finding a statistically reliable, and at the same time accurate, estimation model. In addition, the need of the specification of the appropriate form of the regression model that can accurately describe the data in hand, is a difficult and time-consuming requirement that has to be successfully addressed. For these reasons, the forest scientific research has focused on the application of new modeling methods, such as that of artificial intelligence systems (Artificial Intelligence, AI) and their comparative evaluation with the more classic modeling methods that were widely used and are still used today, such as the theory of regression analysis, (Diamantopoulou, 2005; Diamantopoulou et al., 2009; Diamantopoulou et al., 2018; Özçelik et al., 2019; Bayat et al., 2020; Boroughani et al., 2022; Güner et al, 2022), in order to determine their usefulness in solving problems of forest research.

As a part of AI, the algorithm of Random Forest regression (RFr) technique (Breiman, 2001; Seagal, 2003; Prasad et al,.2006; Cutler et al., 2012; Hastie et al., 2017), can be used to solve the estimation values of continuous variables. This algorithm constitutes an ensemble supervised machine learning model, which makes use of the decision trees learning process to optimize the learning procedure. It is hoped that the application of this modeling approach according to the simulation of ground-truth data of pine trees (*Pinus brutia* Ten.) from the suburban forest of Thessaloniki, will offer a general alternative reliable solution to the problem of accurate estimation of diameters at any height of the bole of standing trees. As a valuable landscape conifer species, pine has received considerable attention by many researchers (Tricando and Burkhart, 2006; Nicoletti et al., 2020; Özçelik and Alkan, 2020; Barna et al., 2020)

The aim of this work is, on the one hand, to develop a reliable estimation model by using the random forest regression technique and, on the other hand, to compare its performance with the nonlinear regression models adaptation to the data, which were developed in order to estimate the diameters of the pine (*Pinus brutia* Ten.) trunks at any height from the ground. Finally, the advantages and disadvantages offered by each modeling method are discussed.

Materials and methods

A total sample of 94 pine trees *(Pinus brutia)* from the Seich–Sou suburban forest of Thessaloniki, Greece, were measured. This forest is an almost pure planted pine forest. Its central location of the study area can be found using the geographic coordinates 40°37’33.0”N and 23°00’45.0”E. Systematic sampling was used to ensure that all different site classes would be included. In a 1:10,000 map of the forest, parallel lines were marked at equal distances, using the WS-NE direction and 42° exposure. A total of 94 points were systematically selected on the map lines. Then the relative trees were located on the ground using a GPS instrument. Tree measurements included stump diameter (0.3 m height from ground, $d_{0.3}$), and diameter at breast height (1.3 m height from ground, $d_{1.3}$), both measured by Finnish caliper, all diameters at one-meter height interval above breast height ($d_{2.3}$, $d_{3.3}$, $d_{4.3}$, $d_{5.3}$, $d_{6.3}$, $d_{7.3}$, $d_{8.3}$ and $d_{9.3}$) measured by Speigel Relaskop and total height (htotal) of the sampled trees, measured by the Blume-Leiss hypsometer (Philip, 1994; Avery and Barkhart, 2002). After the measurements were completed, a sample of size n = 445 rows of data was obtained.

To investigate and to construct the most appropriate, for our data, estimation model, in the case of modeling using the random forest regression algorithm, the sample of n = 445 lines of data was divided using random numbers into two distinct parts: a) the sample of the fitting data which constitutes the 90% of the total lines of data (n_1 = 401) and (b) the testing sample of the of the constructed model, which consist of the remaining 10% data lines (n_2 = 44). (Fig. 1). This is a reliable procedure that has been followed by many researchers (Moore et al., 1996), in order to ensure that the predictive ability of the constructed models is adequate.

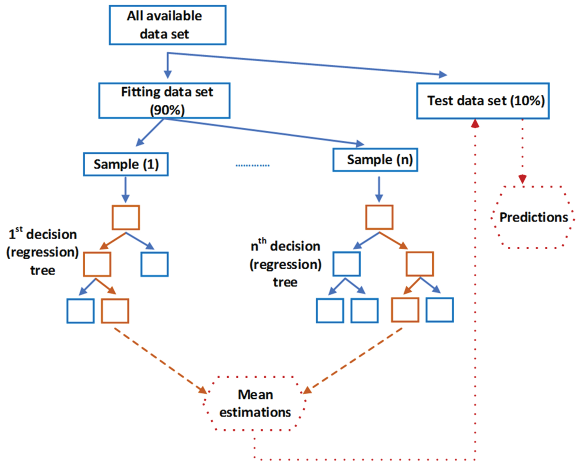

Fig. 1. Random Forest regression (RFr) structure.

In the case of the nonlinear regression models development, the above separation was not applied because it does not affect the training evaluation of the regression model in the way applied to intelligent learning models, while this handle could lead to a loss of additional information (Hirsch, 1991).

In order to establish a regression model that will be able to estimate the over-bark diameter at any height of the standing tree bole, a large number of multiple and nonlinear regression equations were adjusted to the data, using the sta-

tistical package IBM SPSS (2019), with the dependent variable ($d_i/d_{1.3}$) and independent variables the stump diameter ($d_{0.3}$), the total height (h_{total}) and the height (h_{di}) of the diameter in question This diameter value (d_i) is expected to be estimated by the constructed model. Both the Lilliefors statistical test of normality (Lilliefors, 1967), and the normal probability graphics were used to examine the normality of the dependent variable. Among the nonlinear regression models investigated, the logistic (1) model and the Gompertz (2) model (Ratkowsky, 1990) that showed the best fitting performance to the available data, were:

$$\frac{d_i}{d_{1.3}} = \frac{1}{1+e^{[-(\theta_1 + \theta_2 \cdot d_{0.3} + \theta_3 \cdot h_{total} + \theta_4 \cdot h_{di})]}} \quad (1)$$

$$\frac{d_i}{d_{1.3}} = \theta_1 \cdot e^{[-\theta_2 \cdot e^{(\theta_3 \cdot d_{0.3} + \theta_4 \cdot h_{total} + \theta_5 \cdot h_{di})}]} \quad (2)$$

where θ_i are the asymptotic regression coefficients.

Due to the heterogeneity of the variance error detected for the dependent variable, only weighted nonlinear least-squares regression was used. The investigation of the proper value of the weighted factor was applied by the use of the maximum likelihood estimation method (MLE) (IBM SPSS, 2019) in the range of [-3,3] values by 0.1 (Bi and Hamilton, 1998), Finally, the Levenberg-Marquardt algorithm (Ratkowsky, 1990) was used for optimization.

Because of the bias detected on the least squares' assumptions and the relatively high error values of estimates of the non-linear regression models, the Random Forest (Breiman, 2001), which is a non-parametric predictive methodology, was used as an alternative procedure to regression analysis. The Random Forest regression algorithm (RFr) is a supervised machine learning algorithm, that it is based on the ensemble optimization procedure from decision trees following a random way. That is, RFr model combines the estimations from multiple uncorrelated models in order to produce more accurate estimations. In our case, each decision tree was optimized using random sample both from the fitting data set and the available predictors (Fig. 1). Within the RFr, the individual decision trees had not interacted each other, but they were optimized separately. Finally, the random forest estimation was calculated as the average of the estimations produced by the optimization of the decision trees included in the analysis (Fig. 1). In order to construct the RFr model, the technique of bootstrapping has been followed. Bootstrapping (Hastie et al., 2017) is a sampling technique in which subsets of observations, under replacement, were created from the original dataset. This technique is also referred to as Bagging.

Each individual regression tree consisted of a connected flowchart, where there was a unique starting node from which two edges (branches) start and end at "children" nodes that come from the parental nodes. A condition of satisfaction has been settled, for each node. If this goal was not achieved, the process moved on to a new hub and new "children". By working in that manner, the RFr algorithm presents an important advantage: it avoids overfitting of the final model. However, the main disadvantage of the method is the lack of extrapolation. That is, it cannot produce reliable predictions for the dependent variable beyond the range of the predictors' values the model used in its training phase. The learning of the RFr model was achieved by using libraries of scikit-learn (Pedregosa et. al., 2011) and the Python programming language (Vanrossum and Drake, 2011, Python Software Foundation, 2022).

For the evaluation of the non-linear regression models and the RFr model, the following goodness of fit measures were used: 1) the correlation coefficient (R) between the actual values and the corresponding values of the model, 2) the maximum absolute error (MaxAE) between the actual values and the corresponding values of the model, 3) the percentage relative error (RE%) of the estimates, as a measure of accuracy of the model, 4) the root of the average square error (RMSE) between the actual values and the estimated values from the model and 5) the Furnival's (FI) index, in order to derive the precision obtained by the weighted non-linear models in terms of the unweighted dependent variable. The smaller the FI index value, the better adaptation of the model to the data received (Furnival, 1961). Furthermore, the exploration of the normality of the residuals derived by the models, was made using the test statistic of the Lilliefors test. It was calculated by the formula:

$$T_i = sup|(F^*(x) - S(x)|, \quad (3)$$

where $F^*(x)$ is the standard normal distribution function and the $S(x)$ is the empirical distribution function of the transformed (x_i) values under the transformation $z_i = (x_i - \underline{x})/s$, where $\underline{x}$ is the sample mean and s is the standard deviation of the x_i values.

Results

The descriptive statistics for the stump diameter ($d_{0.3}$), the breast height diameter ($d_{1.3}$) and the total height (h_{total}) of the trees for the total data set, are given in Table 1.

Because of the non-homogeneity of the error variance of the dependent variable (MENG and TSAI, 1986), weighted nonlinear least-squares regression was used. From exploratory graphing of $d_{1.3}$ against the dependent variable and the investigation of the proper value of the weighted factor using the maximum likelihood estimation method (MLE) (IBM SPSS, 2019) in the range of [-3,3] values by 0.1 resulted to variance proportional to the 0.2 power of the $d_{1,3}$. Thus, the weighting factor used was $w_i = \frac{1}{(d_{1.3})^{0.2}}$.

Table 1. Descriptive statistics for the raw data set

Variable	Arithmetic mean	Mean standard error	Maximum value	Minimum value	Variance
$d_{0.3}$ (cm)	18.93	0.3045	39.0	9.0	41.25
$d_{1.3}$ (cm)	15.02	0.3191	38.5	6.0	45.31
h_{total} (m)	6.95	0.0848	12.0	3.3	3.20

According to the weighted dependent variable, the exploratory data analysis conducted showed accepted deviation from normality. As can be seen in the relative Q-Q plots of Fig. 2, the values of the weighted variable are close enough to the straight line that represents the normality (Fig. 2-(2)), while the unweighted values of the same variable show a S shape pattern (Fig. 2-(1)), indicating deviation from the required normality. The box-plots of Fig. 2-(3) showed the correction of the heterogeneity of the variance of the weighted dependent variable, as well.

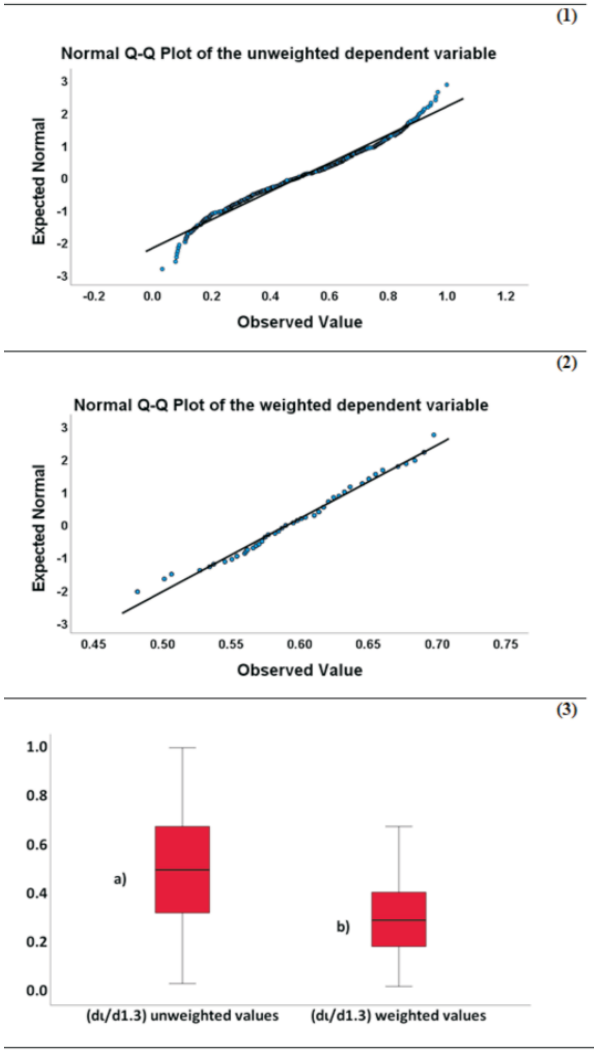

Fig. 2. Normal Q-Q plots (1 and 2) and box-plots of the unweighted (3a) values and the weighted (3b) values of the dependent variable, respectively.

Many multiple and nonlinear regression models tested to fit the available data set, The models that were best adapted to the data were the weighted logistic model and the weighted Gompertz model (Table 2). The significance tests of the regression coefficients for the two equations are given in Table 2.

According to the learning process of the random forest regression model, this was achieved using 100 regression trees (Fig. 3) whose number was selected after testing different numbers of trees ranged from 2 to 150. It was found that after the use of the 100 trees, there was no significant improvement in the model's average estimation error.

Bootstrap Aggregation

The size of subsets created for bagging may be less than the original set. In our case, in order not to miss any of the available information, the size of the subsets used is the same as that of the original set. In this technique a generalized result is obtained by combining the results of various predictive models.

The bootstrap method applied for each decision (regression) tree training data selection, so the procedure to be completely randomized. Finally, the number of the 10 branches was selected as the depth of the tree (Fig. 3), in order to avoid the learning over-parameterization of each decision tree.

The evaluation criteria for the non-linear models and the Random Forest regression model for the total sample of the 445 lines, are given in Table 3.

According to the weighted nonlinear models, Furnival's index was used to transform the error of the weighted equations (transformed error values). In this way, the weighted error would be comparable to the corresponding error of the RFr model.

As shown in Table 3, all statistical evaluation metrics are better for the RFr model as compared to the weighted nonlinear models, while both the non-linear equations gave similar results. In particular, the RMSE value for the RFr model, is 8.18% less than the corresponding error of the logistic model, while the RE% value of the estimates that used as a measure of accuracy of the models, shows that the RFr model's estimates are 7.89% more accurate than the estimates derived from the logistic model.

Table 2. Regression coefficient values, and 95% confidence intervals (a = 0.05), for the weighted equations: (1) Weighted logistic model $\frac{d_i}{d_{1.3}} = \frac{1}{1+e^{[-(\theta_1+\theta_2\cdot d_{0.3}+\theta_3\cdot h_{total}+\theta_4\cdot h_{ib})]}}\cdot w_i$, and (2) Weighted Gompertz model $\frac{d_i}{d_{1.3}} = \theta_1\cdot e^{[-\theta_3\cdot e^{(\theta_5\cdot d_{0.3}+\theta_5\cdot h_{total}+\theta_5\cdot h_{ib})}]}\cdot w_i$

Regression coefficient	Non-linear regression weighted models					
	Weighted logistic model			Weighted Gompertz model		
	Value (SE)	Lower limit	Upper limit	Value (SE)	Lower limit	Upper limit
θ_1	1.184 (0.103)	0.982	1.387	1.108 (0.081)	0.949	1.268
θ_2	–0.024 (0.005)	–0.035	–0.014	0.410 (0.066)	0.279	0.540
θ_3	0.302 (0.021)	0.262	0.343	0.015 (0.004)	0.008	0.022
θ_4	–0.638 (0.020)	–0.676	–0.599	–0.192 (0.026)	–0.244	–0.141
θ_5	---	---	---	0.381 (0.040)	0.301	0.460

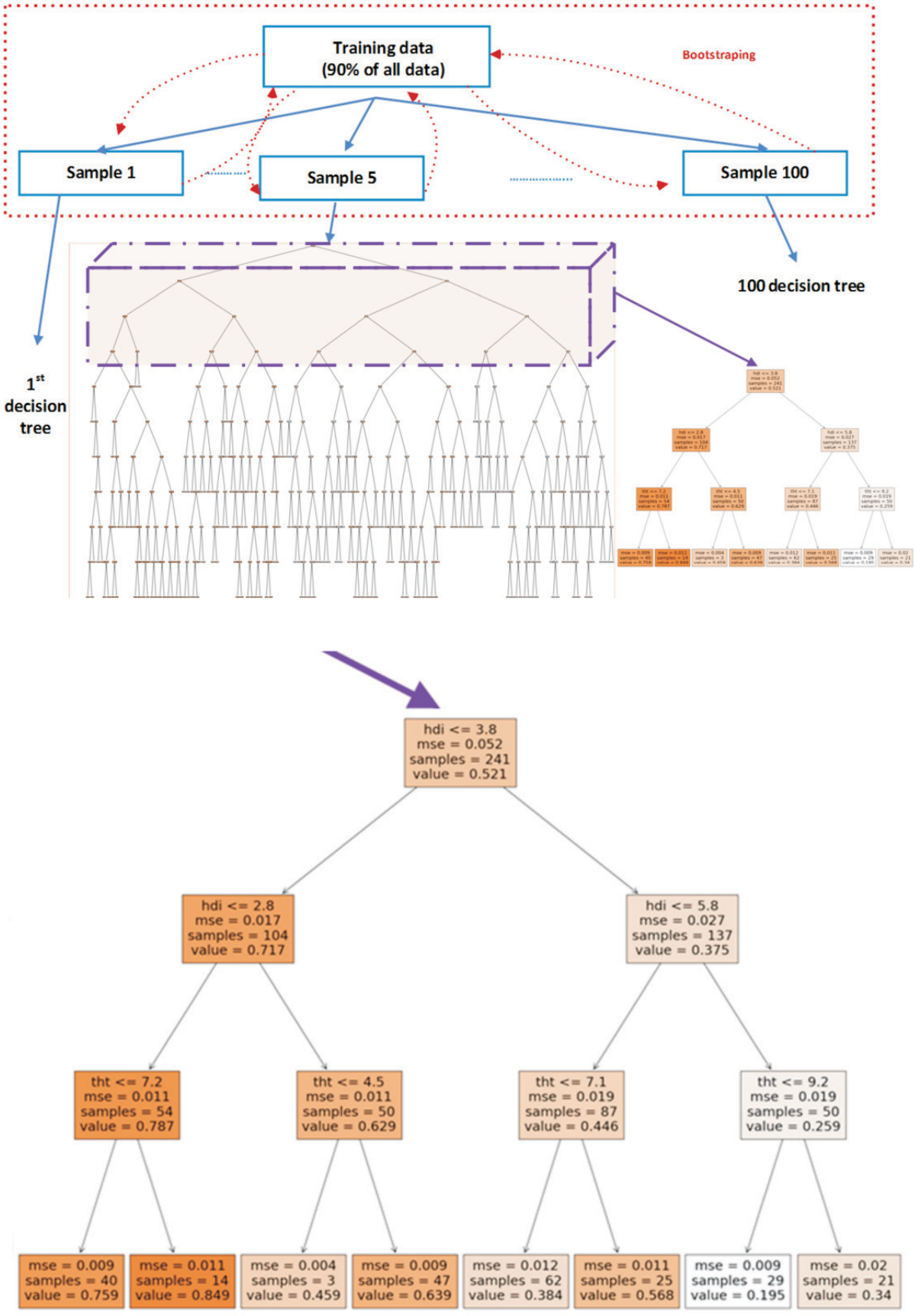

Fig. 3. Random Forest regression architecture.

Table 3. Evaluation statistics for the three models, the logistic (1), the Gompertz (2) and the Random forest regression (RFr) mode

Model	Dependent var.	R	MaxAE	RE%	RMSE	Factor[a]	FI
Logistic (1)	$d_i/d_{1.3}\cdot w_i$	0.8928	0.3404	16.16	0.0611	1.6921	0.1034
Gompertz (2)	$d_i/d_{1.3}\cdot w_i$	0.8897	0.3421	16.47	0.0697	1.6921	0.1049
RFr	$d_i/d_{1.3}$	0.9696	0.2790	8.27	0.0561	1	0.0561

[a]Anti-log of the reverse derivative of the model's dependent variable, in terms of the untransformed variable values. It is a necessary quantity for the calculation of the FI index.

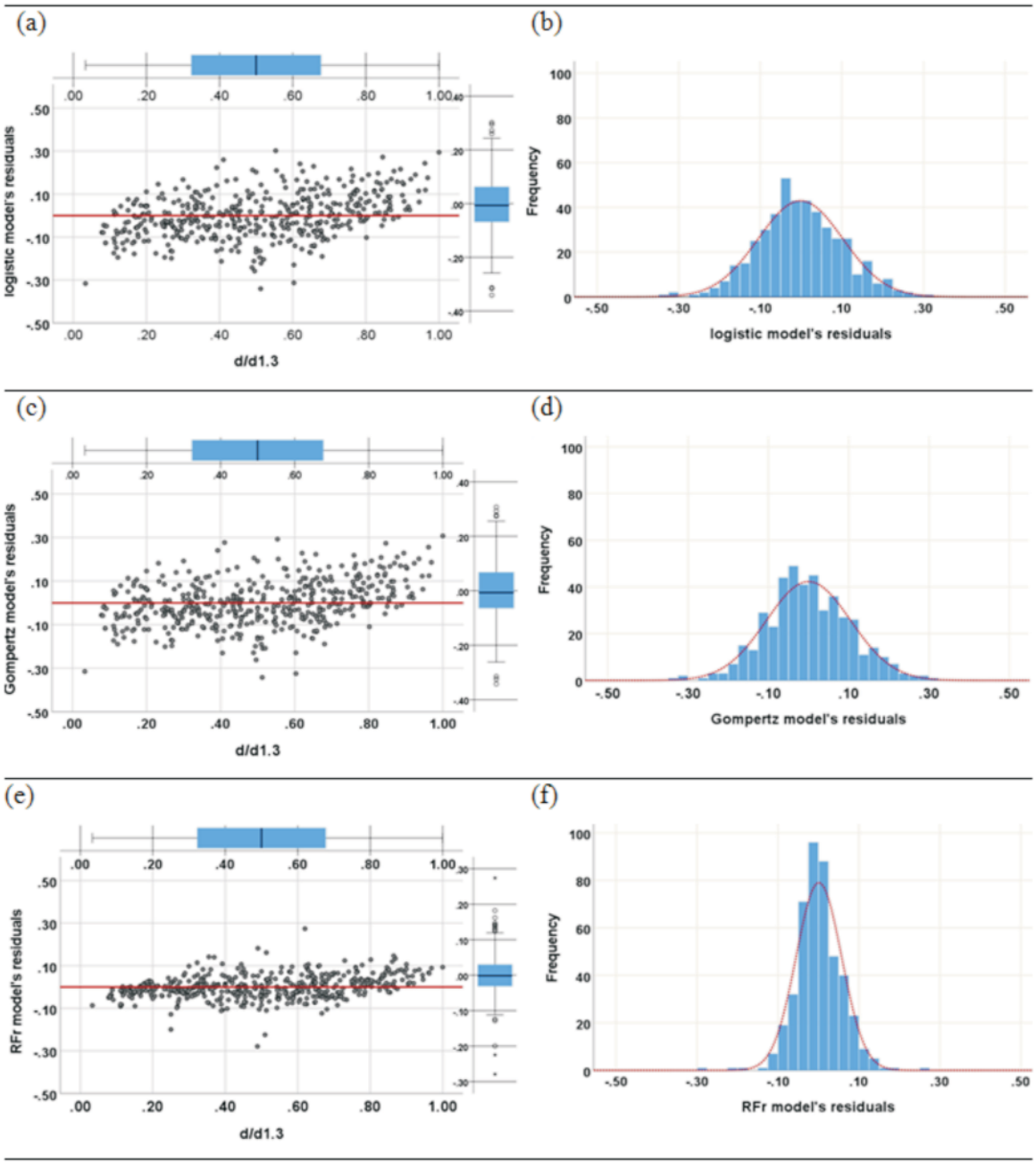

Fig. 4. Errors dot diagrams (a and c) and errors histograms (b and d) for the non-linear regression models and for the Random Forest regression model (e and f), respectively.

The residual analysis of the three models revealed normally distributed errors, with a small and homogeneous variance, which supports the statistically correct construction of the models (Fig. 4).

The same result was derived from the Lilliefors test for normality. The SPSS package was used for conducting the analysis. The null hypothesis of normality for the residuals for the three models, was not rejected in the significant level of a = 0.05, with significance of the test for the residuals, equals to 0.200 (Sig. = 0.200) of all tested models.

Discussion

Measuring the diameter values of a standing tree bole, at any height from the ground, is a non-trivial task that requires substantial field effort. The knowledge of the accurate values of these diameters enables, not only the reliable tree volume computation, but also the knowledge of the structure of the forest Matis, 2004; West, 2009, resulting in their sustainable management (Nanos and Montero, 2002). In this paper, the random forest (Random Forest regression, RFr) method was applied, and a comparative evaluation with the more classic modeling approach such as the nonlinear regression methodology, followed. The necessity of applying an intelligent system method arises from the fact that it has the ability to overcome problems in forest data, such as nonlinear relationships, non-Gaussian distributions, noise in the data, and it appears promising as an alternative to the traditional regression models in many forest modeling applications. This fact, combined with the requirement of the knowledge of the appropriate form of the regression equation, which will be able to reliably describe the primary data, makes the application of the regression methodology, difficult and time-consuming. Furthermore, nonlinear regression models' convergence relies on the good initial values for the parameters of the nonlinear models, which needs further effort by the modeler. On the other hand, intelligent systems, such as the random forest regression methodology, have the ability to use the primary data without any preconditions and they do not require by the modeler to know the form of the model which can reliably describe the data. The system itself creates the model. However, the appropriate values of the intelligent model parameters, have to appropriate be selected. For this, iterative process of learning is used, with different parameter values and different combinations of them. At the end, the final choice of the most appropriate parameter values leads to the most accurate learning. In particular, for the case of the RFr model, it was required to select the appropriate number of decision trees and to determine the complexity of each of these trees. Random Forest regression technique appeared more flexible for the stem diameter prediction than the logistic and the Gompertz non-linear regression models. The results of this research were supported by the findings of many researchers that evaluated the performance of many machine learning methods, such as artificial neural metworks (Soares et al., 2011) or random forest regression (Nunes and Görgens, 2016) in modelling stem taper. Furthermore, the Random Forest technique has not prerequisites, requires little configuration and at the same time can produce models without overtraining (Prassad et al., 2006). According to the tree taper modeling, until now, there are limited research results from the Random Forest technique applications in forest research. Nevertheless, the insights derived from the conducted research in this area, lead to the superiority of this machine learning method, as compared to other conventional methods.

Conclusions

The ability of the random forest regression model to assess more accurately as compared to nonlinear regression models, the diameters at any height of the tree bole, showed that this intelligent process of modeling learning can be successfully applied to forest data.

It also showed that the model is generic and can be safely used to estimate the tree bole diameters at any height from the ground. This concluded using primary data from pine trees measured in the Seich–Sou suburban forest of Thessaloniki, Greece. It is worth mentioning that the constructed random forest regression model can be safely used within the range of the model's construction primary data.

Despite the fact that the random forest regression model does not have the form of a conventional model, its application to the forest primary data of Pine trees from the suburban forest of Thessaloniki offered an alternative reliable solution to the problem of accurate estimation of diameters at any height of the trunk of the standing trees.

References

Algera, M., Kätsch, C., Chirwa, P.W., 2019. Developing a taper model for the Pinus elliottii x P. caribaea var. hondurensis hybrid in South Africa. Southern Forests: *a Journal of Forest Science*, 81: 141–150. https://doi.org/10.2989/20702620.2018.1520027

Arkes, J., 2019. *Regression analysis*. *A practical introduction*. London: Routledge. 362 p.

Avery, T.E., Burkhart, H.E., 2002. *Forest measurements*. New York, USA: Mc Graw Hill. 456 p.

Barna, M., Ferezliev, A., Tsakov, H., Mihál, I., 2020. Investigations of mature Scots pine stands in windthrow areas in Norway spruce forests in Western Rhodopes. *Folia Oecologica*, 47: 1–9. doi: 10.2478/foecol-2020-0001

Bayat, M., Bettinger, P., Heidari, S., Henareh Khalyani, A., Jourgholami, M., Hamidi, S.K, 2020. Estimation of tree heights in an uneven-aged, mixed forest in northern Iran using artificial intelligence and empirical models. *Forests*, 11 (3): 324. https://doi.org/10.3390/f11030324

Bi, H., Hamilton, F., 1998. Stem volume equations for native tree species in southern New South Wales and Victoria. *Australian Forestry*, 61 (4): 275–286. https://doi.org/10.1080/00049158.1998.10674752

Boroughani, M., Soltani, S., Ghezelseflu, N., Pazhouhan, I., 2022. A comparative assessment between artificial neural network, neuro-fuzzy, and support vector machine models in splash erosion modelling under simulation circumstances. *Folia Oecologica*, 49 (1): 23–34. https://doi.org/10.2478/foecol-2022-0003

Breiman, L., 2001. Random forests. *Machine Learning*, 45: 5–32. http://dx.doi.org/10.1023/A:1010933404324

Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random forests. In Zhang, C., Ma, Y.Q. (eds). *Ensemble*

machine learning. New York: Springer, 2015, p. 157–175. http://dx.doi.org/10.1007/978-1-4419-9326-7_5

Diamantopoulou, M.J., 2005. Artificial neural networks as an alternative tool in pine bark volume estimation. *Computers and Electronics in Agriculture*, 48: 235–244. https://doi.org/10.1016/j.compag.2005.04.002

Diamantopoulou, M.J., Milios, E., Doganos, D., Bistinas, I., 2009. Artificial neural network modeling for reforestation design through the dominant trees bole-volume estimation. *Natural Resource Modeling*, 22: 511–543. https://doi.org/10.1111/j.1939-7445.2009.00051.x

Diamantopoulou, M.J., Özçelik, R., Yavuz, H., 2018. Tree-bark volume prediction via machine learning: a case study based on black alder's tree-bark production. *Computers and Electronics in Agriculture*, 151: 431–440. https://doi.org/10.1016/j.compag.2018.06.039

Draper, N.R., Smith, H., 1998. *Applied regression analysis*. New York: Wiley. 706 p.

Furnival, G., 1961. An index for comparing equations used in constructing volume tables. *Forest Science*, 7: 337–341.

Güner, S.T., Diamantopoulou, M.J., Poudel, K.P., Çömez, A., Özçelik, R., 2022. Employing artificial neural network for effective biomass prediction: an alternative approach. *Computers and Electronics in Agriculture*, 192: 106596. https://doi.org/10.1016/j.compag.2021.106596

Hastie, T., Tibshirani, R., Friedman, J., 2017. *The elements of statistical learning data mining, inference, and prediction*. 2nd ed. New York: Springer. 764 p.

Hoaglin, D.C., Mosteller, F., Tukey, J.W., 2006. *Exploring data tables, trends and shapes*. New York: John Wiley and Sons Inc. 538 p.

Hirsch, R.P., 1991. Validation samples. *Biometrics*, 47: 1193–1194. [cit. 2022-01-09]. http://www.jstor.org/stable/2532672

IBM SPSS Corp., released 2019. *IBM SPSS Statistics for Windows*, Version 26.0. Armonk, NY: IBM Corp.

Leite, H.G., Da Silva, M.L.M., Binoti, D.H.B., Fardin, L., Takizawa, F.H., 2011. Estimation of inside-bark diameter and heartwood diameter for Tectona grandis Linn. trees using artificial neural networks. *European Journal of Forest Research*, 130: 263–269. https://doi.org/10.1007/s10342-010-0427-7

Lilliefors, H.W., 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association* 62: 399–402. https://doi.org/10.2307/2283970

Koulelis, P.P., Ioannidis, K., 2021. Constructing single-entry stem volume models for four economically important tree species of Greece. *Folia Oecologica*, 48 (2): 136–146. https://doi.org/10.2478/foecol-2021-0014

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: 405. https://doi.org/10.3390/f11040405

Matis, K.G., 2004. *Forest biometry: II. Dendrometry*. 2nd ed. Thessaloniki, Greece: Pegasus. 674 p. (In Greek).

Meng, C.H., Tsai, W.Y., 1986. Selection of weights for a weighted regression of tree volume. *Canadian Journal of Forest Research*, 16 (3): 671–673. https://doi.org/10.1139/x86-118

Moore, J.A., Zhang, L., Stuck, D., 1996. Height-diameter equations for ten tree species in the Inland Northwest. *Western Journal of Applied Forestry*, 11: 132–137. https://doi.org/10.1093/wjaf/11.4.132

Nanos, N., Montero, G., 2002. Spatial prediction of diameter distribution models. *Forest Ecology and Management*, 161: 147–158. https://doi.org/10.1016/S0378-1127(01)00498-4

Nicoletti, M.F., Carvalho, S.D.P.C., Do Amaral Machado, S., Costa, V.J., Silva, C.A., Topanotti, L.R., 2020. Bivariate and generalized models for taper stem representation and assortments production of loblolly pine (Pinus taeda L.). *Journal of Environmental Management*, 270:110865.https://doi.org/10.1016/j.jenvman.2020.110865

Nunes, M.H., Görgens, E.B., 2016. Artificial intelligence procedures for tree taper estimation within a complex vegetation mosaic in Brazil. *PLoS ONE*, 11: e0154738. https://doi.org/10.1371/journal.pone.0154738

Özçelik, R., Karatepe, Y., Gürlevik, N., Cañellas, I., Crecente-Campo, F., 2016. Development of ecoregion-based merchantable volume systems for Pinus brutia Ten. and Pinus nigra Arnold. in southern Turkey. *Journal of Forestry Research*, 27: 101–117. https://doi.org/10.1007/s11676-015-0147-4

Özçelik, R., Diamantopoulou, M.J., Trincado, G., 2019. Evaluation of potential modeling approaches for Scots pine stem diameter prediction in north-eastern Turkey. *Computers and Electronics in Agriculture*, 162: 773–782. https://doi.org/10.1016/j.compag.2019.05.033

Özçelik, R., Alkan, O., 2020. Fitting and calibrating a mixed-effects segmented taper model. *Cerne*, 26: 464–473. https://doi.org/10.1590/01047760202026032737

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: 2825–2830.

Philip, M.S., 1994. *Measuring trees and forests*. 2nd ed. Wallingford, UK: CAB International. 310 p.

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression techniques: bagging and random forests for ecological prediction. *Ecosystems*, 9: 181–199. https://doi.org/10.1007/s10021-005-0054-1

Python Software Foundation, 2022. *Python 3.9*. https://docs.python.org/3.9/index.html

Ratkowsky, D.A., 1990. *Handbook of nonlinear regression models*. Statistics: Textbooks and Monographs, vol. 107. New York: Marcel Dekker Inc. 241 p.

Salekin, S., Catalan, C.H., Boczniewicz, D., Phiri, D., Morgenroth, J., Meason, D.F., Mason, E.G., 2021. Global tree taper modelling: a review of applications, methods, functions, and their parameters. *Forests*, 12: 913. https://doi.org/10.3390/f12070913

Segal, M.R., 2003. Machine learning benchmarks and random forest regression. San Francisco: Center for Bioinformatics and Molecular Biostatistics, University of California. [cit.2022-01-09]. https://escholarship.org/uc/item /35x3v9t4

Soares, F.A.A.M.N., Flôres, E.L., Cabacinha, C.D., Carrijo, G.A., Veiga, A.C.P., 2011. Recursive diameter prediction and volume calculation of Eucalyptus trees using multilayer perceptron networks. *Computers and Electronics in Agriculture*, 78: 19–27. https://doi.org/10.1016/j.compag.2011.05.008

Trincado, G., Burchart, H.E., 2006. A generalized approach for modeling and localizing stem profile curves. *Forest Science*, 52: 670–682.

Van Laar, A., Akça, A., 2007. *Forest mensuration.* Dordrecht, The Netherlands: Springer. 383 p.

Van Rossum, G., Drake, F.L., 2011. *The Python language reference manual*. Network Theory Ltd. 150 p.

West, P.W., 2009. *Tree and forest measurement*. 2nd ed. Berlin: Springer-Verlag. 191 p.

Received January 12, 2022
Accepted May 20, 2022