Annals of the University of Oradea, Fascicle: Environmental Protection Vol. XXIX, 2017
Analele Universității din Oradea, Fascicula: Protecția Mediului Vol. XXIX, 2017

DETERMINING THE SIZE OF THE SAMPLE TO COLLECT THE DENDROMETRIC ELEMENTS REQUIRED FOR GENERATION THE SHAPE OF SPRUCE SPINDLE

Muscă Ciprian Dumitru*

*University of Suceava, Faculty of Forestry, 13 Universității St., 720229, Suceava, Romania, e-mail: ciprianmusca@yahoo.com

Abstract

Using taper equations has some major advantages in terms of predicting dendrometrical elements (diameter, height, volume). Considering that to generate the coefficients for the taper equations or for the volume calculation we need a large number of data (diameters measured along the spindle every meter on trees felled) and the collection of this data is most often expensive, the sample size becomes very important.

In this sense it is very useful to calculate the number of trees sampled and the number of diameters measured on each tree. So far, researchers have developed mathematical models based on samples consisting of 400 trees up to more than 10,000 trees.

This paper aims to analyze the sample size using correlation and regression analysis technique. When between two variables (dependent and independent) there are correlative connection, through an equation we can express the link between the two variables.

Another method is the sequential analysis which involves checking the null hypothesis after the extraction and verification of each element in the sample. The measurement data is stopped when the tested sample is accepted. The examples are built on analyzing a data set consisting of 863 diameters measured on 23 trees, spruce species. The data come from a natural arboretum, located at an altitude of 950 m, SV exposure, in Suceava county, near Vatra Dornei.

The analysis concludes that the statistical methods applied are effective. Regarding the sample size in terms of costs, especially as time spent for collecting data, the conclusion would be that it is more effective to measure a diameter plus on the trunk of the tree, at an accessible height, than to expand the sample by including more trees.

Key words: form, taper, equation, volume, sample, regression

INTRODUCTION

An important role in forest inventory is to establish the volume of trees. This allows us to understand the relations that exist in the forest ecosystem and to see how the competition, natural or anthropogenic interventions have an impact on the forest. These influences are seen studying the tree form (the taper). Starting from the idea that wood accumulates in the taper form we can also see the influences on the calculation of the volume of trees.

Setting the volume and the taper has a huge economical role for both forest fund administrator and the sector of forestry exploitation and industrialization of wood, too.

To describe the taper and to calculate the volume of trees, it's applying mathematical techniques using regression equations. The great advantage of these regression equations is that they can estimate the diameter at any height of taper, as well as the volume on sections of different length (height).

By using regression equations can be calculated local regression coefficients specific to certain stationary conditions, being much closer to the specific area. Thus, by this calibration of the regression equations, using regression coefficients calculated on local particularities, we get much more accurate information both in terms of form but especially in terms of volume of trees, information that farther are useful both in studies related to the growth process of forest in question and in economical relations of the production process and wood exploitation.

To succeed on calculating regression coefficients, an important factor is determining the sample base to be measured so the data thus obtained should be extended to a local population.

In this sense, we try to identify a method for calculating the size of the sample diameters, heights, number of trees that are required to be measured, starting from the relation diameter height, relation used also in volume calculation or in generating taper of trees and as well we want to identify the relation that exists between diameters measured at different heights, successive on taper. It is known that the tree form (taper) is given by the way of decreasing diameter with the increase in height, Philip, 1994.

MATERIAL AND METHOD

Although more than 100 years there are many studies to determine regression equations to calculate volume or taper, still, there is not a single model that can be used with good results for all species (Clutter et al., 1983). Also, mathematical models that are used to describe the taper and the volume calculation must be constantly improved to keep up with market needs.

Given the importance of knowing its characteristics, the curve shape of the taper was and will be one of the main objectives of the dendrometrical research.

In practice it is necessary to know the dendrometrical characteristics of stands, on large areas. A maximum accuracy of these determinations could be obtained by a complete inventory of stands, Giurgiu 1968. A total inventory takes time and high costs, but these should not do this activity unprofitable.

In Romania researchers used samples of different sizes to calculate, or to verify volume equations. For the study, related to taper form and volume calculation, Giurgiu has measured, in Romania, 10,000 trees, including 1008 Spruce trees and 1429 Fir trees.

To estimate the taper form and to calculate the volume, he uses Huber's formula. For the study, Toma uses data from 1,240 trees, from parquet covered with cuttings in Călimani Mountains. In 1968, the issue of taper of the trees form the spruce stands in Suceava is discussed also by Ichim, who made measurements in parquet traveled with cuttings measuring a number of 4,600 trees.

In 1961 Iacovlev is conducting a study on the accuracy of the yield tables, for pin from the Eastern Carpathians, using a sample of 818 trees. Goulding and Murray (1975) have conducted measurements on a sample of 1,267 trees, Max and Burkhart (1976) used a sample of 652 trees, Perez et al. (1990) used a sample of 405 trees, and Kozak and Smith (1993) used a sample of 603 trees, Kitikidou, Chatzilazarou (2008).

The data presented in this researches did not reveal how the sample was established, the number of trees required to be measured, or the number of variables (diameter, height) required.

One of the methods used and accepted for determining the sample size is the one proposed by Demaerschalk and Kozak (1974), based on the simple linear regression method. Thus, they are assuming that if there is an indication that between two variables there is a linear correlation connection strong enough, we can apply regression analysis method for the sample. In this way, will be calculated the arithmetic average and the variance of the independent variable, who depend on the acceptable error, for the calculation of dependent variables. From these calculations it will be used the highest value as the minimum required sample size.

To perform this study there were measured diameters and lengths on the trees of the spruce species. The area from which data were collected is located geographically in the Eastern Carpathians, Romania, town of Vatra Dornei, Suceava County, at an altitude of 950 m, SV exposure.

Thus it was established a set of measurements made on a total of 23 trees, on which were measured a number of 863 diameters, and also on a model consisting on the average of the diameters measured along the taper. The data set was formed by actual measurement in field, on the felled trees, by sections of 1m diameter. In the first part of the tree section, diameters were taken down at length: 0.3; 0.5 1; 1.3; 1.5; 2 ..... n.

The method used

Method Demaerschalk and Kozak (1974), Kitikidou, Chatzilazarou (2008). To apply this method, we will use the linear correlation relation that exists between variables: the diameter at breast height (DBH) and height, as well as the correlative link that exists between diameters measured along the length of the taper.

To establish the link correlation between the two variables we will use simple linear regression, trying to find a strong association between the two elements that lead us to increase the accuracy of prediction of a variable expense of the other. Thus, in a first phase we will determine the regression equation, and in the second phase we will use it in order to achieve the prediction.

The study will focus on setting correlative relations between the diameter at breast height (DBH) and the diameter measured at a height of 0.5 m above the ground. In this case the independent variable is established as the diameter at breast height and the dependent variable as the diameter of the h0,5. In this case will be processed elements Dbh and Dh0,5, of 23 trees.

Also, given that in form equations of trees, diameters are measured at different we try to calculate the sample starting with relative height as the independent variable and diameter as the dependent variable. In this way 863 samples will be processed, meaning diameter measurements made on 23 trees.

To achieve this we apply simple linear regression formula:

$$y = b_0 + b_1x,$$

where y = dependent variable, x = independent variable, b_0 – intercept, b_1 – slope.

Confidence interval when $x = x_0$

$$\hat{y} \pm t_{\alpha/2,\,DF}\, s\sqrt{\frac{1}{n} + \frac{(x_P - \overline{x})^2}{\sum (x_i - \overline{x})^2}} \quad C_i = 2t_{n-2,\alpha/2,}\, s\sqrt{\frac{1}{n} + \frac{(x_P - \overline{x})^2}{\sum (x_i - \overline{x})^2}}$$

$t_{n-2,\,\alpha/2,}$ - value of t distribution for (n–2) degrees of freedom and significance level α,

s – standard error of the estimate,

n – final sample size,

x_0 – mean of the independent variable distribution.

$\sum (x_i - \overline{x})^2$ = VarX - – variance of the independent variable distribution.

The main usefulness of linear model is predicting the dependent variable values. Forecast value is obviously a statistic, because it is produced by the estimated model (from experimental data). Then we can speak of the random distribution of the predicted value, distribution that u underlies the determination of confidence intervals for the predicted values.

If simple regression (the regression line), last confidence interval has the form

$$\hat{y} - t_{\alpha/2,DF} s \sqrt{\frac{1}{n} + \frac{(x_P - \overline{x})^2}{\sum (x_i - \overline{x})^2}} \leq y \leq \hat{y} + t_{\alpha/2,DF} s \sqrt{\frac{1}{n} + \frac{(x_P - \overline{x})^2}{\sum (x_i - \overline{x})^2}}$$

from where we obtain the conclusion that forecast values have intervals of confidence, on the same confidence level, higher as the value x_0 is furthest from the average x.

From here it appears recommendation that a linear model should not be used for forecasting if the independent variables have values far from the center of the data considered in estimating the model.

Thus, our goal is to find value $t_{n-2,\alpha/2,}$, but for that we need to know the value of the sample size. The proposed method of Demaerschalk and Kozak (1974) is based on the determination of sample size based on solving the confidence interval for simple linear regression equation.

The regression equation can be used to make predictions (estimations) only if the two variables are perfectly linear correlated. For this we use significance tests for regression coefficients. Among the elements used in significance testing are standard error of prediction test t – for the slope and intercept respectively for analysis of variance (ANOVA).

RESULTS AND DISCUSSION

A. Starting from the basic relationship between the diameter and the diameter measured at 0.5 m from the ground we calculate the sample size necessary to validate a mathematical model based on the relationship between the diameter measured at different heights on the tree trunk.
For this calculation we use as independent variables – $d_{1,3}$ and as the dependent variable – $d_{h0,5}$.

To achieve this we apply simple linear regression formula:

$$y = b_0 + b_1x,$$

By calculating the above equation we obtain the coefficients: -0,0358 intercept or slope – 1,2184. y = 1,2184x - 0,0358

The standard error calculated is very small, only 0,022623 (s) and the coefficient of determination (R^2) is 0,979003. Which means that 97,9% of the variation of the diameter at breast height is determined by the variation of the diameter measured at a distance of 0.5 m from the ground and only 2.1% of the variation of the diameter at breast height is determined by other variables.

Based on the above relation, for a 10% accepted error Ci = 0,10X = 0,10X0,405 = 0.0405, the value X is 0.11223, and the size of the range is 51 trees.

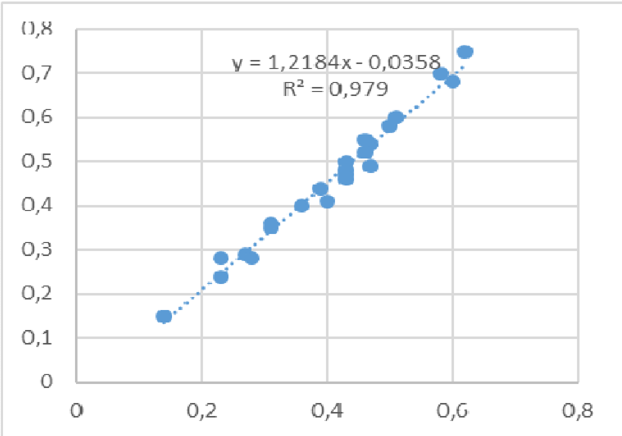

Fig. 1. The link between $d_{1,3}$ and $d_{h0,5}$

Testing the significance of the regression coefficients using test t and analysing the variance (ANOVA) we observe that the value of ratio F (ratio of variance expressed by the equation of regression and the variance of residuals) corresponds to $P < 0.01$.

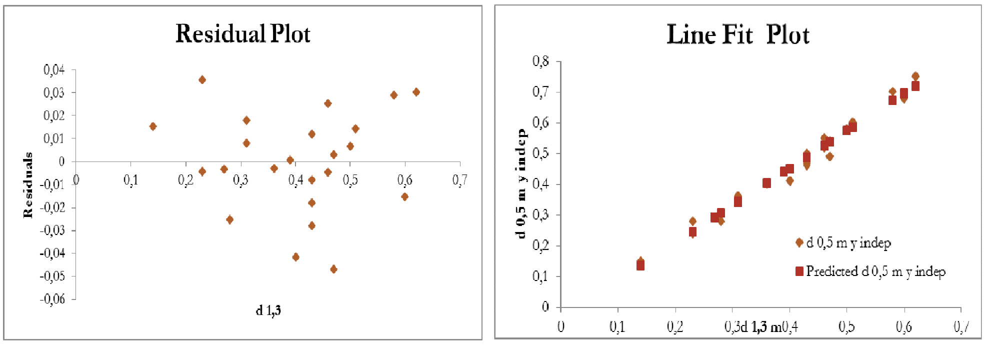

Fig. 2. Statistical presentation

B. Starting from the basic relationship between the diameter and height we calculate the sample size necessary to validate a mathematical model based on the relation between diameter and height.

For this calculation we use as independent variables - $h_r = h_i/H$ and as the dependent variable – $dr = D_{hi}/D$.

To achieve this, we apply simple linear regression formula:

$$y = b_0 + b_1x,$$

By calculating the above equation we obtain the coefficients: 1,0954 intercept or slope –
-0,9602.
$y = 1,0954 - 0,9602x$

The standard error calculated is 0,1004 (s) and the coefficient of determination (R^2) is 0,90162. Which means that 90,1% of the variation of the diameter at breast height is determined by the variation of the height and only 9,9% of the variation of the diameter at breast height is determined by other variables.

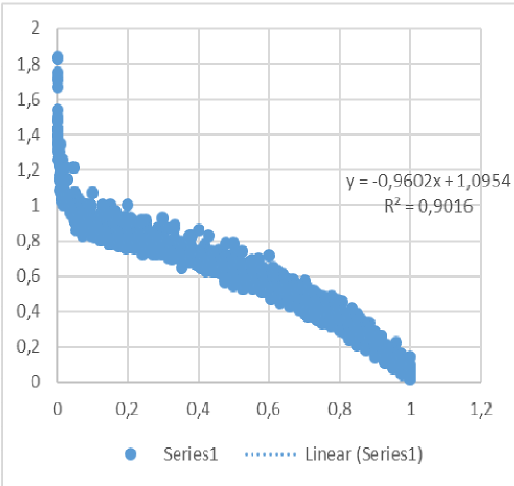

Fig. 3. The link between $d_{1,3}$ and $d_{h0,5}$

Based on the above relation, for a 10% accepted error Ci = 0,10X = 0,10X0,450 = 0.0450, the value X is 0.6223, and the size of the range is 768 trees.

Testing the significance of the regression coefficients using test t and analysing the variance (ANOVA) we observe that the value of ratio F (ratio of variance expressed by the equation of regression and the variance of residuals) corresponds to P <0.01.

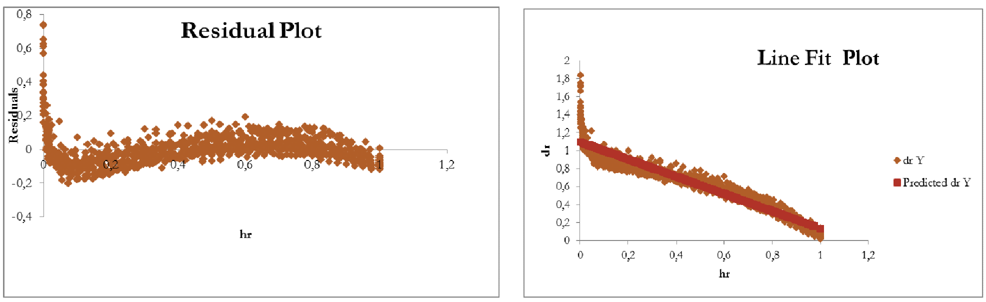

Fig. 4. Statistical presentation

CONCLUSIONS

The analysis concludes that the statistical methods applied are effective. Regarding the sample size in terms of costs, especially as time spent for collecting data, the conclusion would be that it is more effective to measure a diameter plus on the trunk of the tree, at an accessible height, than to expand the sample by including more trees.

Regarding the number of measured diameters on tree, we can see that the method of data collection from m to m is satisfying, for the chosen degree of accuracy is not required to be measured so many diameters.

And this study confirms the close correlative relations that exists between height and diameter as well as between different diameters measured successively on the trunk.

Concerning the sample size, it is dependent on the intensity of the link between diameter and height and also on the accuracy required.

REFERENCES

1. Behre C.E., 1923, Preliminary Notes on Studies of Tree Form. Journal of Forestry, 21, pp.507- 511
2. Burkhart H., Tome M., 2012, Modeling Forest Trees and Stands. Springer
3. Cruz de León G., Uranga-Valencia L.P., 2013, Theoretical evaluation of Huber and Smalian methods applied to tree stem classical geometries. Bosque (Valdivia), 34(3), pp.311-317
4. Forslundr R., 1982, A geometrical tree volume model based on the location of the centre of gravity of the bole. Can. J. For. Res., 12, pp.215-221
5. Giurgiu V., 1975, Expresii matematice ale tabelelor dendrometrice romanesti. Analele ICAS, Volumul 33 (1), pp.75-90
6. Giurgiu V., 1979, Dendrometrie şi auxologie forestieră. Ed. Ceres, Bucureşti
7. Giurgiu V., Drăghiciu D., 2005, Modele matematico-auxologice şi tabele de producţie pentru arborete. Ed. Ceres, Bucureşti
8. Kitikidou K. Chatzilazarou G., 2008, Estimating the sample size for fitting taper equations. Journal of Forest Science
9. Ormerod D.W., 1973, A simple bole model. Forestry Chronicle, v.49, pp.136-138