Estimation and validation of stem volume equations for *Pinus sibirica* in Russia

V V Zavarzin, A V Lebedev* and A V Gemonov

Russian State Agrarian University - Moscow Timiryazev Agricultural Academy, 49, Timiryazevskaya st., Moscow, 127550, Russia Federation

*E-mail: alebedev@rgau-msha.ru

**Abstract.** Stem volume models using diameter at breast height (DBH) and height were developed for *Pinus sibirica*. The data were obtained from Irkutsk Oblast (Eastern Siberia), Tyumen Oblast (Western Siberia) and Sverdlovsk Oblast (Ural). Logarithmic scale models that are widely used in forestry, which are often used to model stem volume and tree biomass, were slightly inferior to models using the response at the original scale. The obtained models can be used for forest inventory, and the proposed methodological approaches can be used for modeling the stem volume of other tree species.

1. Introduction

Siberian cedar pine (*Pinus sibirica*) is one of the main forest-forming species in Siberia. Its wood is a valuable raw material in the timber industry. For forest resource management, forest stock is basically represented by stem volume [1], and volume equations are widely applied using diameter at breast height (DBH) and height to calculate stem volume [2]. Many scientists have studied various volume equations [3, 4, 5]. In the USSR, a large number of forest inventory standards for determining the volume of cedar stems were compiled by the method of graphical data alignment [6]. The volume of cedar stems in the USSR was studied Nikolay Tretyakov, Pavel Gorsky [7], Nikolay Anuchin [8], Ivan Semechkin [9] and others.

The use of modern methods for the analysis of experimental data makes it possible to obtain more accurate predictive models in forestry. The aims of this study were to provide new parameter estimates of total stem volumes, to compare the models with previous studies for validation, and to verify the best stem volume equation for *Pinus sibirica* in Russia.

2. Methods and materials

The data on the morphometric characteristics of Siberian cedar pine stems used in the work were obtained as a result of forest inventory work in the Irkutsk region (Zhigalovsky and Cheremkhovsky districts) - Eastern Siberia, Tyumen region (Nizhnevartovsky, Surgutsky and Yarkovsky districts) - Western Siberia. In addition, the work uses collected on the territory of the Sverdlovsk region (Verkhotursky district) - Middle Urals.

The sample for modeling includes information about 689 model trees. The sample is characterized by the following statistical indicators. The average value of the DBH is 20.0 cm (standard deviation is 11.5 cm), the average value of the height is 14.5 m (standard deviation is 6.4 m). The average age of the model trees is 151 years (the standard deviation is 67 years), the average volume is 0.3881 m$^3$ (the
standard deviation is 0.4481 m$^3$). The largest share of the experimental material falls on stems with DBH of up to 30 cm, while the distribution of the experimental material by height classes is close to uniform. The distribution of the experimental material by classes of stem volumes is characterized by a pronounced right-sided asymmetry, more than half of the model trees have a stem volume of up to 0.5 m$^3$.

Based on the study of the literature data [10, 11, 12], 16 mathematical models of stem volume ($v$) were selected using diameter at breast height ($d$) and stem height ($h$) as predictor variables. The structure of mathematical models is presented in Table 1.

### Table 1. Stem volume equations using DBH and height

| Model No. | Equation | No. of coefficient |
|-----------|----------|-------------------|
| 1         | $v = \beta_0 + \beta_1 d + \beta_2 h$ | 3 |
| 2         | $v = \beta_0 + \beta_1 d + \beta_2 d^2 h + \beta_3 h$ | 4 |
| 3         | $v = \beta_0 + \beta_1 h d^2$ | 2 |
| 4         | $v = \beta_0 + \beta_1 d^2$ | 2 |
| 5         | $v = \beta_0 + \beta_1 d + \beta_2 d^2$ | 3 |
| 6         | $\ln v = \beta_0 + \beta_1 \ln(h d^2)$ | 2 |
| 7         | $\ln v = \beta_0 + \beta_1 \ln d + \beta_2 \ln h$ | 3 |
| 8         | $\ln v = \beta_0 + \beta_1 \ln d + \beta_2 \ln^2 d + \beta_3 \ln h$ | 4 |
| 9         | $\ln v = \beta_0 + \beta_1 \ln d + \beta_2 \ln^2 d + \beta_3 \ln h + \beta_4 \ln^2 h$ | 6 |
| 10        | $\ln v = \beta_0 + \beta_1 \ln d + \beta_2 \ln^2 d + \beta_3 \ln(h - 1.3) + \beta_4 \ln^2(h - 1.3)$ | 5 |
| 11        | $\ln v = \beta_0 + \beta_1 \ln d + \beta_2 \ln(h - 1.3)$ | 3 |
| 12        | $v = \frac{\pi}{40000} dh(\beta_0 + \beta_1 d)$ | 2 |
| 13        | $v = \beta_0 hd^2 + \beta_1 dh + \beta_2 d^3 + \beta_3 dh^2$ | 4 |
| 14        | $v = \frac{\pi}{4}(\beta_0 d^2 h + \beta_1 d^2 h \ln^2 d + \beta_2 d^2)$ | 3 |
| 15        | $v = \beta_0 h + \beta_1 h \left(\frac{d}{200}\right)^2 + \beta_2 \left(\frac{d}{200}\right)^2$ | 3 |
| 16        | $\ln v = \beta_0 + \beta_1 \ln d + \beta_2 \ln(d + 20) + \beta_3 \ln h + \beta_4 \ln(h - 1.3)$ | 5 |

The parameters of the models were estimated using the least squares method. For each model, the coefficient of determination ($R^2$), the adjusted coefficient of determination (adj. $R^2$), the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the mean square error (MSE) were calculated at the baseline scale [13]. For nonlinear models, along with the coefficient of determination, the index of determination was calculated. The overall significance of the regression model was assessed using the F-test ($\alpha = 0.05$).

To confirm the adequacy of the constructed models, the analysis of regression residuals was carried out. An important step was the detection of heteroscedasticity, which leads to ineffectiveness of the estimates obtained by the least-squares method, since the classical estimate of the covariance matrix of the least squares estimates of parameters is biased and inconsistent. Heteroscedasticity was detected using the Brousch-Pagan test ($\alpha = 0.05$). If the fact of heteroscedasticity was revealed, then to obtain more accurate and correct statistical conclusions, standard errors in the White form were used.

The statistical significance of the regression parameters in the case of homoscedasticity of errors was assessed using the t-test ($\alpha = 0.05$), and in the case of heteroscedasticity with the introduction of White's correction using the z-test ($\alpha = 0.05$).

### 3. Results and discussion

The characteristics of the stem volume models are shown in Table 2. For all models, except for model No. 1 ($R^2 = 0.858$), which does not consider the nonlinearity of the relationship between the DBH with
stem volume and stem height with volume, the coefficient of determination ($R^2$) ranges from 0.964 to 0.997, and the adjusted coefficient of determination (adj. $R^2$) is 0.963 to 0.997. Models No. 4 and No. 5, in which only DBH is used as a predictor variable, are characterized by relatively low values of the coefficient of determination, equal to 0.964 and 0.968, respectively.

Table 2. Coefficients and fit statistics of stem volume equations

| No  | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $R^2$ | Adj. $R^2$ | MSE  | AIC  | BIC  |
|-----|----------|----------|----------|----------|----------|----------|-------|------------|------|-------|------|
| 1   | -2.7095E-01 | 4.2446E-02 | -1.3209E-02 | -         | -         | -         | 0.858 | 0.857      | 0.0285 | -490  | -476 |
| 2   | -1.3501E-02 | 2.3280E-03 | 3.4000E-05  | -1.2600E-04 | -         | -         | 0.989 | 0.989      | 0.0023 | -2226 | -2208|
| 3   | 1.0969E-02  | 3.6000E-05 | -         | -         | -         | -         | 0.988 | 0.988      | 0.0024 | -2199 | -2190|
| 4   | -6.0302E-02 | 8.4000E-04 | -         | -         | -         | -         | 0.964 | 0.963      | 0.0073 | -1430 | -1421|
| 5   | 2.2256E-02  | -9.4220E-03| 1.0380E-03 | -         | -         | -         | 0.968 | 0.968      | 0.0064 | -1522 | -1508|
| 6   | -9.7494E+00 | 9.5078E-01 | -         | -         | -         | -         | 0.996 | 0.996      | 0.0028 | -772  | -763 |
| 7   | -9.7768E+00 | 1.8733E+00 | 9.2111E-01| -         | -         | -         | 0.996 | 0.996      | 0.0029 | -772  | -758 |
| 8   | -9.5768E+00 | 1.7097E+00 | 5.2041E-02 | 9.2063E-01| -         | -         | 0.997 | 0.997      | 0.0029 | -959  | -941 |
| 9   | -9.1233E+00 | 1.9086E+00 | 5.6605E-02 | -9.2710E-03| 2.3575E-01| 1.3814E-01| 0.997 | 0.997      | 0.0027 | -986  | -959 |
| 10  | -8.8770E+00 | 2.0470E+00 | -9.2040E-03| 1.1044E-01| 1.5164E-01| -         | 0.997 | 0.997      | 0.0028 | -980  | -958 |
| 11  | -9.0550E+00 | 2.1124E+00 | 4.7843E-01 | -         | -         | -         | 0.992 | 0.992      | 0.0048 | -269  | -255 |
| 12  | 5.0901E-01  | 4.5744E-01 | -         | -         | -         | -         | 0.964 | 0.964      | 0.0024 | 1986  | 1996 |
| 13  | 4.7000E-05  | 1.5900E-04 | -4.0000E-06| -1.4000E-05| -         | -         | 0.994 | 0.994      | 0.0023 | -2233 | -2215|
| 14  | 5.1007E-05  | -5.3213E-07| 5.5832E-05 | -         | -         | -         | 0.994 | 0.994      | 0.0023 | -1905 | -1891|
| 15  | 6.3700E-04  | 1.3174E+00 | 2.7405E+00 | -         | -         | -         | 0.993 | 0.993      | 0.0023 | -2222 | -2208|
| 16  | -1.0479E+01 | 1.9507E+00 | 1.2110E-01 | 1.4806E+00 | -5.0027E-01| -         | 0.997 | 0.997      | 0.0026 | -994  | -971 |

Among the models presented in Table 1, according to the indicators $R^2$, MSE, AIC, BIC, model No. 2 can be selected as the best ($R^2 = 0.989$; MSE = 0.0023; AIC = -2226; BIC = -2208), model No. 3 ($R^2 = 0.988$; MSE = 0.0024; AIC = -2199; BIC = -2190), model No. 13 ($R^2 = 0.994$; MSE = 0.0023; AIC = -2233; BIC = -2215) and model No. 15 ($R^2 = 0.993$; MSE = 0.0023; AIC = -2222; BIC = -2208).

Models on a logarithmic scale of traits and responses have found wide application in forestry, for example, in modeling tree volumes and in modeling tree biomass. At the same time, the inverse transformation from the logarithmic scale is rarely used when assessing the quality of models. Including a very small number of studies both in our country and abroad, in which, to compare the quality of models, the value of the standard deviation is calculated in the original scale. Special attention should be paid to the coefficient of determination, which is often used for comparing models, which is known to have limited application in the case of nonlinear models. R. Anderson-Sprecher [14] notes that the widespread use of $R^2$ inevitably leads to its erroneous use in practice.

In our case, none of the models in which the parameters are estimated on a logarithmic scale were included in the list of the best models, despite the fact that models No. 8-10 are characterized by a fairly high coefficient of determination ($R^2 = 0.997$). MSE values at the scale of the original features are in the range from 0.0027 to 0.0029, which is slightly higher than the models selected as the best. Thus, as a result of the logarithmic transformation, there was a shift in the estimates of the coefficient of determination. The coefficient of determination (the index of determination) in the original scale acquires the following values: model No. 8 - 0.986; model number 9 - 0.987; model No. 10 - 0.986. At the same time, it should be noted that information criteria turned out to be a more reliable tool for testing the quality of models, since they allow considering the accuracy of the approximation and the complexity of the model.
Despite the fact that the models with the logarithmic response turned out to be slightly worse, they have an important advantage over the models selected as the best. In models of this type, the decrease in the accuracy of the approximation is compensated by minimizing the relative deviations of the volume values for small-sized wells.

For small-sized shafts, incorrect prediction of volumes using model No. 2 was noted. When the stem DBH are less than 3 cm and the stem height is less than 3 m, the calculated values of the volumes stem out to be less than 0, which contradicts the physical meaning of the value under consideration.

For a stem with a DBH of 20 cm and a height of 15 m, the volume value calculated according to model No. 2 is 0.3576 m$^3$, according to model No. 3 - 0.2270 m$^3$, according to model No. 13 - 0.2347 m$^3$, according to model No. 15 - 0.2346 m$^3$. The difference between the minimum and maximum predicted values is 0.1306 m$^3$. The average value of the volume over the model trees with the given taxation characteristics is 0.2485 m$^3$. Thus, we see that in this case model No. 2 gives a strong overestimation of the true value (44 %). The volumes that are closest to the true value were obtained using models No. 13 and No. 15, which give the least underestimation (~5.5 %).

4. Conclusion

The models obtained as a result of the study of the dependence of the volume of Siberian cedar pine stems on the DBH and on the height of the stem can be used to compile new forest inventory standards, for example, volumetric tables. The methodological techniques described in the work can be used in the development of models of the volumes of stems of other tree species.

References

[1] Dubenok N N, Kuzmichev V V, Lebedev A V, Gemonov A V 2020 Growth and productivity of larch stands in conditions of urbanized environment, in European Russia Baltic Forestry 1(26) 248
[2] Avery T E and Burkhart H 2001 Forest Measurements (5th ed.) New York, NY: McGraw-Hill 456
[3] Schumacher F X and Hall F D S 1933 Logarithmic expression of timber-tree volume J Agric Res 47 719-34
[4] Honer T G 1965 A new total cubic foot volume function Forest Chron. 41 476-93
[5] Baskerville G L 1972 Use of logarithmic regression in the estimation of plant biomass Can J For Res 2 49-53
[6] Zavarzin V V and Lebedev A V 2017 Concerning the simulation method of trunks volume by the example of the siberian cedar pine (Pinus sibirica) Prirodooobstrojstvo 3 96-103
[7] Tretyakov N V, Gorskiy P V and Schultz V E 1940 Taxator’s reference book Leningrad, Gosleskhizdat 393
[8] Anuchin N P 1949 Assortment and market tables for larch, cedar and fir Moscow, Leningrad, Goslesbumizdat 84
[9] Semechkin I V 2002 Structure and dynamics of Siberian stone pine forests Novosibirsk, Publishing house of SB RAS 253
[10] Vibrans A C, Moser P, Oliveira L Z and Maçaneiro J P 2015 Generic and specific stem volume models for three subtropical forest types in southern Brazil Annals of Forest Science 6(72) 865-74
[11] Zianis D, Muukkonen P, Mäkipää R and Mencuccini M 2015 Biomass and stem volume equations for tree species in Europe Silva Fennica Monographs 63
[12] Brandel G 1990 Volume Functions for Individual Trees, Scots Pine (Pinus sylvestris), Norway Spruce (Picea abies) and Birch (Betula pendula & Betula pubescens): in Rapport - Institutionen för Skogsproduktion, Sveriges Lantbruksuniversitet Swedish University of Agricultural Sciences: Garpenberg, Sweden
[13] Lebedev A and Kuzmichev V 2020 Verification of two- and three-parameter simple height-diameter models for the birch in the European part of Russia J. For. Sci 66 375-82
[14] Anderson-Sprecher R 1994 Model comparisons and R$^2$ The American Statistician 2(48) 113-7