CONSTRUCTION AND VALIDATION OF TREE VOLUME FUNCTIONS FOR *DALBERGIA SISSOO* GROWN UNDER IRRIGATED CONDITIONS IN THE HOT DESERT OF INDIA

V. P. Tewari & V. S. Kishan Kumar

Arid Forest Research Institute, P.O. Krishi Mandi, New Pali Road, Jodhpur - 342 005, India

Received October 1999

TEWARI, V. P. & KISHAN KUMAR, V. S. 2001. Construction and validation of tree volume functions for *Dalbergia sissoo* grown under irrigated conditions in the hot desert of India. Six volume equations were compared on the basis of fit and validation statistics using data collected from *Dalbergia sissoo* stands at Indira Gandhi Nahar Pariyojana (IGNP) area of Rajasthan State in India. An equation fitting very well to a data set may not necessarily be the best when applied to another data set though collected from the same population. The contrasting results obtained sometimes between model fitting and validation emphasise the need for model validation as an important step in model construction process in order to get the best choice. The combined variable equation produced the best results and hence has been recommended for use in total wood volume prediction of *D. sissoo* in the study area.

Key words: *Dalbergia sissoo* - fit statistics - hot desert - irrigated plantations - validation - volume equations - India

TEWARI, V. P. & KISHAN KUMAR, V. S. 2001. Pembinaan dan pengesahan fungsi isipadu pokok bagi *Dalbergia sissoo* yang ditanam di bawah keadaan pengairan di gurun panas India. Enam persamaan isipadu dibanding berdasarkan statistik kesesuaian dan pengesahan yang menggunakan data daripada dirian *Dalbergia sissoo* di kawasan Indira Gandhi Nahar Pariyojana (IGNP) di Rajasthan, India. Satu persamaan yang sangat sesuai dengan satu set data tidak semestinya sesuai untuk set data lain yang diambil daripada populasi yang sama. Keputusan berbeza yang diperoleh antara kesesuaian dan pengesahan model menekankan betapa perlunya pengesahan model sebagai langkah yang penting dalam proses pembinaan model bagi mendapatkan pilihan yang terbaik. Penggabungan persamaan pembolehubah menghasilkan keputusan yang terbaik dan dengan itu disyorkan untuk meramalkan jumlah isipadu pokok *D. sissoo* di kawasan kajian.

Introduction

Estimation of tree volume with greater accuracy has always been a matter of interest for forest managers. Construction of volume equations for tree species is an important step in this direction. The importance of volume equations is well indicated by the existence of numerous such equations and the constant search for their improvement. In India, various volume equations and tables are constructed during forest inventories but these equations are not validated. The role of validation in examining the predictive ability of a model before its application has been stressed by various authors (Goulding 1979, Reynolds *et al.* 1981).

The aim of the present study was to develop and validate volume equations for *Dalbergia sissoo* stands grown as irrigated plantations of the Indira Gandhi Nahar Pariyojana (IGNP) (a canal project) area located in the arid parts of Rajasthan State in India.

Materials and methods

Data used in the present study were collected from IGNP area in Rajasthan. Plantations available for the species under study cover various age groups (3–30 y) and stand densities (600–2500 trees ha^{-1}). Trees of different diameter classes (5–40 cm, class interval 5 cm) were felled and their total heights, diameters at breast height (dbh) and volume data were recorded. Diameter class was considered to cover the trees of all sizes in the felled data but age and stand density had no direct role although they affect the diameter growth of the trees. The length of the felled tree was measured with a tape and stump height was added to get the total height. For the computation of total volume, stems and branches up to a minimum diameter of 5 cm were considered. The volume was then calculated by dividing the stems and branches into logs of 3-m length, measuring the mid-diameters and applying Huber's formula to estimate individual log volumes. A total of 71 trees was measured from the plantations.

The data was randomly divided into two sets. The models were fitted to the first set consisting of 70% of the data. The second set, consisting of 30% of the data, was used for validation purposes. These data sets will henceforth be referred to as fitting and validating data sets respectively. Table 1 summarises the statistics of the two data sets.

Table 1 Statistical summary of the fitting and validating data sets

Variable	Sample size	Min.	Max.	Mean	SD	Kurtosis	Skewness
				Fitting data set			
Dbh (cm)	50	5.8	35.9	15.8	7.7	0.4689	1.0317
Height (m)	50	8.2	19.8	12.5	2.9	- 0.0158	0.8708
Volume (m^3)	50	0.0041	0.8539	0.165	0.207	3.3738	1.9806
				Validating data set			
Dbh (cm)	21	6.0	39.8	18.6	8.5	0.5309	0.8712
Height (m)	21	6.6	20.2	13.4	3.4	0.1033	- 0.2070
Volume (m^3)	21	0.0079	1.0362	0.232	0.2637	3.6796	1.9626

SD = standard deviation.

With small data sets, there are chances that assignment of trees to the validation data set may be poor. Therefore iterative validation procedure (Williams 1997) was also adopted to avoid this problem. Here the regression equations were compared against one another for estimating volume from sample data by using cross-validated simulation study. The data were randomly partitioned into five different subsets. In turn, each of the five data sets containing 20% of the data was set aside for validation and the remaining 80% used to fit the model using each regression equation. The fitted models were then used to estimate the volume for each of the five validation subsets.

The data sets cover the small and middle ranges of tree sizes for *D. sissoo* as very large trees were not available.

Model construction

This study compared six volume equations (Table 2) selected from forestry literature based on their wide application (Spurr 1952, Loetsch *et al.* 1973, Clutter *et al.* 1983). Each model was applied to the fitting data set. The error structure in volume estimation was not homogeneous which implied that the observations were not measured with equal precision and ordinary least squares did not yield parameter estimates of the linear regression models with minimum variances. Hence weighted least square fitting technique was applied for fitting equations 1 to 4. It was not necessary for equations 5 and 6 as they were fitted with non-linear technique. The weight applied for equations 1 and 4 was $1/(D^2H)^k$ while for equations 2 and 3 it was $1/(D^2)^k$.

Table 2 Volume equations compared in the study

Equation type	Designation
$V = a + bD^2H$	1
$V = a + bD^2$	2
$V = a + bD + cD^2$	3
$V = a + bH + cD + dD^2 + eD^2H + fDH$	4
$V = aD^b$	5
$V = aD^bH^c$	6

The coefficient of determination (R^2) and the root mean square error (RMSE) were used to determine the quality of fit. For the non-linear regressions, a fit index (FI) analogous to R^2 in linear regression (Cornell & Berger 1987) was used which was computed as:

$$FI = 1 - [\sum_{i=1}^{n} (\hat{V}_i - V_i)^2 / \sum_{i=1}^{n} (V_i - \bar{V})^2]$$

where

V_i = observed volume for tree i
$\hat{V}_i$ = predicted volume for tree i
n = number of observations
$\bar{V}$ = mean observed volume of n trees

A rank was assigned to each equation based on each criterion (Cao *et al.* 1980). The smaller the rank the better the performance of the model. The ranks were then summed up to arrive at the final fit rank for each model that is indicative of its performance with respect to all the criteria considered.

Model validation

The standard error of estimate (SEE) and the average difference between predicted and observed values, the average bias (B), were used as evaluation criteria for model validation (Cao *et al.* 1980, Gordon 1983, Biging 1984, Fowler & Rennie 1988, Trincado *et al.* 1996). The SEE was given as

$$SEE = [\sum_{i=1}^{n}(\hat{V}_i - V_i)^2/(n - p - 1)]$$

and the B was calculated as

$$B = \sum_{i=1}^{n}(V_i - \hat{V}_i)/n$$

where

p = number of model parameter,
V_i, $\hat{V}_i$, and n are as given above.

In the cross validation study, the average prediction bias was given by

$$B = (\tfrac{1}{5})\sum_{i=1}^{n}B_i$$

Similarly, the SEE was also computed over the five validation subsets.

Results and discussion

Model fitting

The values of model coefficients obtained by applying various equations on the fitting data set are given in Table 3. The values of standard errors for various regression coefficients are given in parentheses, which showed that all the partial regression coefficients were significant except for equation 4 where only one coefficient relating to D^2H was found to be significant. The intercept values for equation 3 were also not significant. The values of the power 'k' estimated for the weights applied on equations 1 to 4 were 1.35, 2.80, 2.22 and 1.36 respectively.

Table 4 compares the fit statistics for the equations used. It can be seen that R^2 values were generally high and acceptable for all the equations. The final rankings showed that equations 1 and 4 ranked first, followed by equation 6 while equation 5 ranked last. However, equation 4 involved six variables and the standard errors for various regression coefficients in Table 3 showed that for this equation only one coefficient relating to D^2H was significant. Hence equation 1 was given preference over equation 4.

Table 3 Values of coefficients for different equations obtained for fitting data set

Equation	a	b	c	d	e	f
1	- 0.00303	3.69E-05				
	(0.00075)	(4.31E-07)				
2	- 0.01134	0.000505				
	(0.00113)	(1.66E-05)				
3	0.01182	- 0.00503	0.00075			
	(0.00635)	(0.00125)	(5.52E-05)			
4	- 0.11501	0.00059	0.00113	- 3.28E-05	3.84E-05	- 6.55E-05
	(0.02057)	(0.00228)	(0.00317)	(0.00013)	(8.63E-06)	(0.00028)
5	0.00016	2.38815				
	(0.00004)	(0.06722)				
6	0.00003	1.97492	1.04463			
	(4.16E-06)	(0.03186)	(0.05571)			

Values in parentheses are standard errors for the partial regression coefficients.

Table 4 Fit statistics for volume equations for *Dalbergia sissoo*

Equation	df	R^2	RMSE (m^3)	Rank	Final Rank (R_f)
1	48	0.993 (2)	0.00004 (2)	4	1
2	48	0.951 (6)	0.00001 (1)	7	4
3	47	0.963 (5)	0.00004 (2)	7	4
4	44	0.993 (2)	0.00004 (2)	4	1
5	48	0.986 (4)	0.03078 (6)	10	6
6	47	0.998 (1)	0.01049 (5)	6	3

Values in parentheses are the ranks.

Model validation

For model validation it is pertinent to use an independent data set (validating data set) to assess the predictive ability of the different equations. The volume equations obtained from the fitting data set were applied to the validating data set. The B and the SEE values were considered to assess the overall performance of each equation. The smaller these values the better the prediction. Table 5 compares the validation statistics for the six equations used over the validating data set (30% of original data set).

Equation 1 had minimum SEE while equation 6 produced the lowest bias. Equation 4 had maximum SEE and also produced bias nearly 20 times that of equations 1, 3, 5 and 6 and three times that of equation 2. Final ranking showed that equations 1 and 6 might be considered as the best predictor. Equation 4, which shared the first position during the fitting phase, occupied the last position for volume prediction in validation.

Table 5 Validation statistics for volume equations for *Dalbergia sissoo*

Equation	Bias (m^3)	SEE (m^3)	ΣRank	Final rank (R_v)
1	0.00455 (2)	0.01687 (1)	3	1
2	0.03371 (5)	0.09208 (5)	10	5
3	0.00460 (3)	0.04879 (4)	7	3
4	0.09970 (6)	0.12345 (6)	12	6
5	- 0.00509 (4)	0.04366 (3)	7	3
6	- 0.00415 (1)	0.01696 (2)	3	1

Values in parentheses are ranks.

Table 6 gives the average values of bias and SEE obtained for all the six equations during cross-validation procedure. It can be seen that equation 1 ranked first followed by equations 6 and 4. This again confirmed the superiority of the combined variable equations over the other models used.

Table 6 Statistics for volume equations for *Dalbergia sissoo* obtained through iterative validation procedure

Equation	Average bias (± m^3)	Average SEE (m^3)	ΣRank	Final rank (R_v)
1	0.00236 (1)	0.01009 (1)	2	1
2	0.02266 (6)	0.05410 (6)	12	6
3	0.00752 (4)	0.03097 (5)	9	4
4	0.00255 (3)	0.01279 (3)	6	3
5	0.00772 (5)	0.03020 (4)	9	4
6	0.00249 (2)	0.01113 (2)	4	2

Values in parentheses are ranks.

Figure 1 shows the plots of residuals (observed-predicted) obtained from the validating data set consisting of 30% of the original data against the actual volumes. The non-randomness of residuals for all the equations reflected the heteroscedasticity of the data. The plots also revealed that the least dispersion and minimum bias were for equations 1 and 6, while equation 4 produced maximum residuals which conformed with the rankings given in Table 5. All other equations produced high residuals in the higher range of volume, indicating that predictions from these equations were less accurate especially in this range. In the lower range of volume, all equations except equation 4 produced almost similar values of residuals.

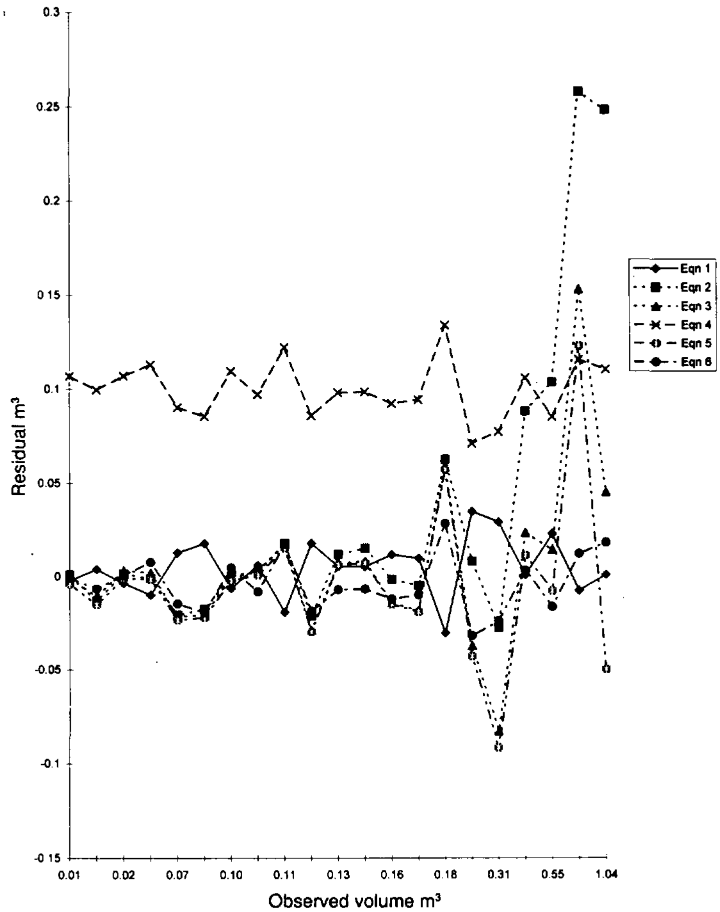

Figure 1 Plots of residuals (observed-predicted) against observed volume of *Dalbergia sissoo*

Tables 4 and 5 showed that equation 1 performed best in the validation as well as in the fitting phases. On the other hand, equation 5, which was placed lowest during fitting, jumped to the third place during validation. Equation 4, which was ranked first in the fitting phase along with equation 1, dropped to the last place during model validation. Almost similar results were also exhibited in the cross-validation process (Table 6). This analysis exemplifies the importance of model validation so that models can be used with greater confidence (Goulding 1979, Reynolds & Chung 1986).

Equation 1, the combined variable equation, has been well recognised in volume predictions of many tree species with R^2 usually above 95% (Avery & Burkhart 1994). From the present study, we also recommend the same equation (equation 1) on the basis of both fit and validation. The final equation based on the pooled fitting and validating data set (obtained through weighted least squares analysis, power k = 1.26) is given below:

$$V = -0.0023 + 0.0000364\ D^2H;\ df = 69;\ R^2 = 0.992;\ RMSE = 0.00006$$

Height, however, is often difficult to measure accurately and may not always be available. In such cases, volume-diameter equations may be the best alternative. In the present study, three such equations were analysed (equations 2, 3 and 5) and equation 3 was found best for such cases. The final equation based on the pooled fitting and validating data set (obtained through weighted least squares analysis, power k=2.18) is given below:

$$V = 0.01328 - 0.00538\ D + 0.000760\ D^2;\ df = 68;\ R^2 = 0.961;\ RMSE = 0.00005$$

Conclusion

It can be concluded from the study that the combined variable equation (model 1) performed well in both the fitting and validation phases. Therefore, it could be used to predict volume for *D. sissoo* in the study area. The contrasting results obtained in some cases between model fitting and validation emphasise the need for model validation as an important step in model construction process in order to get the best choice.

Acknowledgements

The authors wish to thank the Director, AFRI for providing the necessary facilities during this study. The guidance provided by R. C. Jain during the course of this study is gratefully acknowledged. The help rendered by B. Singh, S. L. Chouhan and J. P. Dadhich in data collection is highly appreciated. Help from A. Verma during data analysis is also highly appreciated.

References

Avery, T. E. & Burkhart, H. E. 1994. *Forest Measurements*. McGraw-Hill, New York. 408 pp.

Biging, G. S. 1984. Taper equations for second growth mixed conifers of northern California. *Forest Science* 30: 1103–1117.

Cao, Q. V., Burkhart, H. E. & Max, T. A. 1980. Evaluation of two methods for cubic volume prediction for loblolly pine to any merchantible limit. *Forest Science* 26: 71–80.

Clutter, J. L., Fortson, J. C., Pienaar, L. V., Brister, G. H. & Bailey, R. L. 1983. *Timber Management: A Quantitative Approach*. John Wiley and Sons, New York. 333 pp.

Cornell, J. A. & Berger, R. D. 1987. Factors that influence the value of the coefficient of determination in simple linear and non-linear regression models. *Phytopathology* 77: 63–70.

Fowler, J. H. & Rennie, J. C. 1988. Merchantible height in lieu of total height in stem profile equations. *Forest Science* 34: 505–511.

Gordon, A. 1983. Comparison of compatible polynomial taper equations. *New Zealand Journal of Forest Science* 13: 146–155.

GOULDING, C. J. 1979. Validation of growth models used in forest management. *New Zealand Journal of Forestry* 24: 108–124.

LOETSCH, F., ZOHRER, F. & HALLER, K. E. 1973. *Forest Inventory.* Volume II. BLV Verlagsgesellschaft, Munchen. 469 pp.

REYNOLDS, M. R., BURKHART, H. E. & DANIELS, R. F. 1981. Procedures for statistical validation of stochastic simulation models. *Forest Science* 27: 349–364.

REYNOLDS, M. R. & CHUNG, J. 1986. Regression methodology for estimating model prediction error. *Canadian Journal of Forestry Research* 16: 931–938.

SPURR, S. H. 1952. *Forest Inventory.* John Wiley and Sons, New York. 476 pp.

TRINCADO, G., GADOW, K. V. & TEWARI, V. P. 1996. Comparison of three stem profile equations for *Quercus robur* L. *South African Forestry Journal* 177: 23–29.

WILLIAMS, M. S. 1997. A regression technique accounting for heteroscedastic and asymmetric errors. *Journal of Agricultural, Biological and Environmental Statistics* 2: 108–129.