DOI: 10.1007/s11676-007-0052-6

Comparing the model forms estimating generalised diameter-height relationships in *Tecomella undulata* plantations in hot arid region of India

Vindhya Prasad Tewari

Forest Management Unit, Silviculture Division, P.O. Krishi Mandi, New Pali Road, Arid Forest Research Institute, Jodhpur 342005, India

Abstract: Four generalised diameter-height equations were developed and compared for pure and even-aged stands of *Tecomella undulata* in hot arid region of Rajasthan State in India. The data used to fit the equations consisted of 1 540 diameter-height observations collected from the plots laid out in uniformly stocked stands of varying age and density. The performance of four equations was tested by non-linear least squares regression and evaluated using different statistical criteria. Finally, these equations, with the same values of coefficients obtained during the fitting phase, were validated by an independent data set consisting of 854 diameter-height observations. Overall, equation (4) (Hui and Gadow function) was found to perform best for both the fitting data set as well as validation data set.

Key words: Generalised height-diameter equations; *T. undulata*; Model evaluation & validation; Hot desert; India

Introduction

Tecomella undulata is a tree species found in the Thar Desert of northwest and western India. Distribution of *T. undulata* is restricted to the drier parts of the Arabia, southern Pakistan and northwest India up to an elevation of 1200 m. It is a medium sized tree that produces quality timber and is the main source of timber amongst the indigenous tree species of hot desert regions. Its wood is strong, tough and durable. The wood is excellent for firewood and charcoal (Anon 1994). *T. undulata* is an accepted tree species in agro-forestry and large population is found in agricultural lands. It acts as a soil-binding tree by spreading a network of lateral roots on the top surface of the soil and helps in stabilizing shifting sand dunes.

Growth models for many indigenous species in India are not yet developed and information about the growth of T. undulata trees is rarely available. T. undulata fetches a high price in the domestic market where it is extensively used as timber for making furniture etc. There is a high demand of T. undulata especially by sawmills and other timber and handicraft industries.

Measurements of individual tree diameters and heights are commonly applied in most forest inventory situations to obtain estimates of growing stock. Diameters can be measured easily at low cost but height measurements are time consuming, often inaccurate, very difficult to measure in dense plantations. Hence, the heights are derived indirectly from the diameters, using a known or estimated relationship between diameters and heights (Van Laar and Akça 1997) and modelling the development of this relationship over time. The relationship between diameters and heights may be described using a height regression or a bivariate diameter-height distribution. In practice, all the trees in the plots are measured for diameter and sub-sampling is done for height. The data from trees sampled for height are then used to develop a diameter-height regression, while in turn, is used to estimate height of other tree in the plot (Arabatizis and Burkhart 1992).

A height regression may be derived separately for each stand, on the basis of pairs of diameter-height measurements obtained during a stand inventory. As this approach is rather time consuming and costly, a practical alternative is to develop generalised height regressions, which embody certain basic characteristics inherent in all individual height regressions (Prodan 1965; Kramer and Akça 1987; Wenk *et al*. 1990; Gadow and Hui 1993; Prodan *et al*. 1997; Hradetzky 1999).

Generalised diameter-height equations are useful tools for forest inventory purposes. Commonly, they are also used as important element of many size class models for simulating the development of silvicultural alternatives over time (Páscoa 1987). Additionally, generalised diameter-height equations are sometimes used to generate individual tree height increment data (Hasenauer 1999; Kahn and Dursky 1999).

The aim of the present paper is to develop and compare some generalised diameter-height equations for pure stands of *T. undulata* in arid region of Rajasthan State in India.

Materials and methods

Site description and data

The geographic location of the study area ranged from 27°17′ to 28°31′ across north latitude to 71°18′–72°51′ east longitude. The area is characterized by large variation in the diurnal and seasonal temperatures. The mean monthly maximum temperature varies between 39.5°C and 42.5°C while mean monthly minimum temperature varies between 14°C and 16°C. The mean annual rainfall in the area varies from 120 mm to 300 mm. The majority of the rainfall is received during the southwest monsoon season (July-September). The number of rainy days varies from

Received: 2007-05-08; Accepted: 2007-07-08

Biography: Vindhya Prasad Tewari (1960-), in the Forest Management Unit, Silviculture Division, P.O. Krishi Mandi, New Pali Road, Arid Forest Research Institute, Jodhpur 342005, India
E-mail: vptewari@yahoo.com
Responsible editor: Hu Yanbo

8 to 17 days in the area. The mean monthly relative humidity in the area fluctuates largely during the year from 15% to 80%. Wind speeds as high as 130 km per hour have been experienced during the summer months. The terrain of the area is very undulating and is frequently subjected to moving sand dunes. Dust storms are also common in the region. The area consists of dry undulating planes of hard sand and gravely soil and rolling planes of loose sand. The soil is rich in potash but poor in nitrogen and organic matter with very low productivity. There is a semi-consolidated lime concretionary or gypsum strata underneath at many places. The soils are coarse textured with low water retention capacity.

The data used in the present study were collected from the *T. undulata* plantations established by the State Forest Department of Rajasthan in the Indira Gandhi Canal Project area which is under drought prone hot arid region. The plantations available for the species under study cover age groups ranging from 14 to 20 years and stand densities from 450 to 2 038 stems per hectare.

Two different data sets were used, one for model fitting (referred to as fitting data set) and another one for model validation (referred to as validation data set). The fitting data set consisted of 1 540 height-diameter measurements from 15 plots of 0.1 hm^2 area installed in pure, uniformly stocked stands covering available range of ages and stand densities. The validation data set had 854 diameter-height observations collected from 7 sample plots located in other plantations of *T. undulata* in the same area. The summary of some whole stand statistics for both the data sets are given in Table 1.

Table 1. Summary of some whole stand statistics for the fitting and validating data sets

Variable	Minimum	Maximum	Mean	Standard deviation
Fitting data set				
Basal area (m^2/hm^2)	1.94	14.21	6.43	3.92
Dominant height (m)	4.56	8.47	6.06	1.18
Stems per hectare	450	1916	1116	425
Quad. mean dia. (cm)	9.18	17.77	13.35	2.79
Validation data set				
Basal area (m^2/ hm^2)	3.01	10.63	7.24	2.81
Dominant height (m)	4.91	8.54	6.14	1.34
Stems per hectare	517	2038	1263	454
Quad. mean dia. (cm)	11.50	20.20	14.32	3.11

Fitted models

The generalised diameter-height equations are different with the ordinary diameter-height equations in the sense that they include quadratic mean diameter, stand basal area or stems per hectare as extra independent variable so that the equation can be applied on the plantations available on different sites with varying stocking. Few such equations are available in the literature, which are modifications of Richards or Schumacher functions. This study compares predictive ability of four generalised diameter-height equations.

Pienaar (1991) derived an equation from the Richards function (cp. Richards 1959) for stands of slash pine in the southeastern United States as the follows:

$$h_i = \alpha_1 H_0 \left(1 - \alpha_2 e^{-\beta d_i / D_g}\right)^{\alpha_3} \quad (1)$$

Mirkovich (1958) derived the following equation from the Schumacher function (cp. Schumacher 1939) for oak stands in central Europe (see also Michailoff 1943; Prodan *et al*. 1997):

$$h_i = 1.3 + \left(\alpha_1 + \alpha_2 H_0 - \alpha_3 D_g\right) e^{-\beta / d_i} \quad (2)$$

The stands with the same Dg may have different stand densities and hence model may be further improved by incorporating a variable, which accounts for stand density. Schröder and González (2001) modified equation 2 by incorporating stand basal area as an independent variable:

$$h_i = 1.3 + \left(\alpha_1 + \alpha_2 H_0 - \alpha_3 D_g + \alpha_4 G\right) e^{-\beta / \sqrt{d_i}} \quad (3)$$

On the basis of the allometric growth theory, Gadow and Hui (1993) developed a generalised height regression for the stands of *Cunninghamia lanceolata* in southern China:

$$h_i = 1.3 + \alpha_1 H_0^{\beta_1} d_i^{\alpha_2 H_0^{\beta_2}} \quad (4)$$

In the above equations, h_i is the height of tree i (m), d_i the breast height diameter over bark of tree i (cm), H_0 the dominant stand height (m), D_g the quadratic mean diameter of the stand (cm), G the stand basal area (m^2/hm^2), α_1, α_2, α_3, α_4, β_1, β_2 are the parameters to be estimated, and1.3 is a constant used to avoid the prediction of a height less than 1.3 m when d_i is very small

Each equation was applied to the fitting data set. As the equations are intrinsically nonlinear, iterative nonlinear fitting is required to estimate the parameters (cp. Draper and Smith 1981). In the present paper, the simplex algorithm provided in the non-linear estimation procedure of STATISTICA statistical software package (Statistica 1994) was used for parameter estimation.

Model evaluation and validation

The comparison of the four equations fitted was based on graphic and quantitative analysis of the residuals (e_i). Graphical analysis of residuals searching for discrepancies or patterns is an important step in evaluating the fitted models (Gadow and Hui 1999). Residuals were graphically examined to check for any trend. Linear regression of predicting on observed values was also done to see the performance of the fitted models. The ideal value for the intercept and slope of the linear regression is 0 and 1, respectively. Five statistical criteria were examined: bias ($\overline{E}$), which tests the systematic deviation of the model from the observations; root mean squared error (RMSE), which measures the accuracy of the estimates; the adjusted coefficient of determination (R^2_{adj}), which shows the proportion of the total variance that is explained by the model, adjusted for the number of model parameters and the number of observations; model precision (MPR), a standardised sum of squares criterion proposed by Freese (1960) for evaluating precision of fitted model; and Akaike's information criterion differences (AICd), which is an index to select the best model based on minimizing the Kullback-Liebler distance (Burnham and Andarson 1998). The expressions for these criteria are as below:

Bias:

$$\overline{E} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)}{n} \tag{5}$$

Root mean squared error:

$$\text{RMSE}=\sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n-p}} \tag{6}$$

Adjusted coefficient of determination:

$$R^2_{adj}=1-\frac{n-1}{n-p}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y}_i)^2} \tag{7}$$

Model precision (ideal value 0):

$$MPR = \frac{1}{\sigma^2}\sum(y_i - \hat{y}_i)^2 \tag{8}$$

Akaike's information criterion differences:

$$\text{AICd}= n \ln \hat{\sigma}^2 + 2l - \min\left(n \ln \hat{\sigma}^2 + 2l\right) \tag{9}$$

where, y_i, $\hat{y}_i$ and $\bar{y}_i$ are the measured, predicted and average values of the dependent variable, respectively, σ^2 is the variance of y, n the total number of observations used to fit the models, p the number of model parameters, $l = p+1$, and $\hat{\sigma}^2$ the estimator of the error variance of the model, which the value is obtained as follows:

$$\hat{\sigma}^2 = \sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{n} \tag{10}$$

As a thumb rule, models with AICd $\leq$ 2 have substantial support and should receive consideration in making inferences. Models having AICd of about 4 to 7 have considerably less support, while models with AICd>10 have either essentially no support and might be omitted from further consideration or at least those models fail to explain some substantial explainable variation in the data.

All the four equations, with the same value of the coefficients obtained during the fitting phase, were applied on an independent data set consisting of 854 diameter-height observations for validation purpose. The validation of each model was based on the analysis of the model efficiency (ME) calculated as the adjusted coefficient of determination, the bias, the root mean squared error of the estimates, model precision and Akaike's information criterion differences.

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

Results and discussion

The values of model coefficients obtained by applying the four equations on the fitting data set are given in Table 2. The standard errors for various regression coefficients are given in the parentheses, which showed that all the partial regression coefficients were highly significant (p=0.0001) except the coefficients related to D_g in equation 3 and G in Eq. (4).

Table 2. Estimated coefficient values of different equations for fitting data set

Equation	α_1	α_2	α_3	α_4	β	β_1	β_2
Eq. (1)	1.254	0.960	-1.964		2.073		
	(0.065)	(0.066)	(0.133)		(0.044)		
Eq. (2)	2.954	0.631	0.020		4.765		
	(0.122)	(0.039)	(0.017)		(0.068)		
Eq. (3)	6.492	1.310	0.132	0.018	3.661		
	(0.409)	(0.087)	(0.042)	(0.025)	(0.052)		
Eq. (4)	0.412	0.571				0.436	0.064
	(0.016)	(0.117)				(0.065)	(0.005)

Note: The values in brackets are standard errors for the coefficients

Table 3 compares the fit statistics for the equations used and presents an overview of the performance of the equations based on the statistical criteria used for evaluating predictive ability of the models on fitting data set. It can be seen that adj. R^2 values were generally high (ranging from 0.887 for Eq. (1) to 0.920 for Eq. (4)) and acceptable for all the equations. The values of Bias and RMSE were the minimum for Eq. (4) while maximum for Eq. 1. Also, the first rank in model precision was Eq. (4), followed by Eq. 3. The values of Akaike's criterion differences (AICd) for fitting data set suggest that only Eq. (4) has substantial support in the model selection while all other equations failed badly in explaining some substantial explainable variation in the data and hence must not be given any consideration during model selection. The final rank showed that Eq. (4) ranked first while Eq. (1) ranked last. Thus, Eq. (4) placed first in the overall performance.

Table 3. Statistical criteria for model evaluation on fitting data set

Eqs.	Adj. R^2	Bias	RMSE	MPR	AICd	Rank	Final Rank
Eq. (1)	0.887 (4)	0.02983 (4)	0.39839 (4)	0.11352 (4)	535.26530 (4)	20	4
Eq. (2)	0.888 (3)	0.01500 (3)	0.39672 (3)	0.11258 (3)	522.38340 (3)	15	3
Eq. (3)	0.913 (2)	0.00680 (2)	0.34920 (2)	0.08722 (2)	16.96423 (2)	10	2
Eq. (4)	0.920 (1)	-0.00011 (1)	0.33483 (1)	0.08019 (1)	0.00000 (1)	5	1

Values in the parentheses are the ranks

Fig. 1 presents the plot of residuals against the actual values obtained from the fitting data set. As illustrated from the figure, Eq. (4) had the least dispersion of the residuals. The residual values varied from -1.05 to 1.20, -0.67 to 2.01, -0.58 to 1.37 and -0.90 to 0.97 for Eqs. (1), (2), (3) and (4), respectively.

Fig. 2 shows the plot of observed versus predicted values, which indicates that the Eq. (4) performed best and produced strongest correlation (R=0.959), followed by Eq. (3) (R=0.956). The value of the intercept (α) and slope (β) of the linear regression was -0.03063 and 1.00656, respectively. Slope and intercept of the straight line obtained were compared with the theoretical values of 1 and 0 by means of the joint-confidence ellipse *F*-test to test the null hypothesis. A confidence ellipse is a 2-dimensional interval in which, with a certain probability, the true parameter (a 2-dimensional vector) lies. At the confidence level of 95%, no evident difference was found.

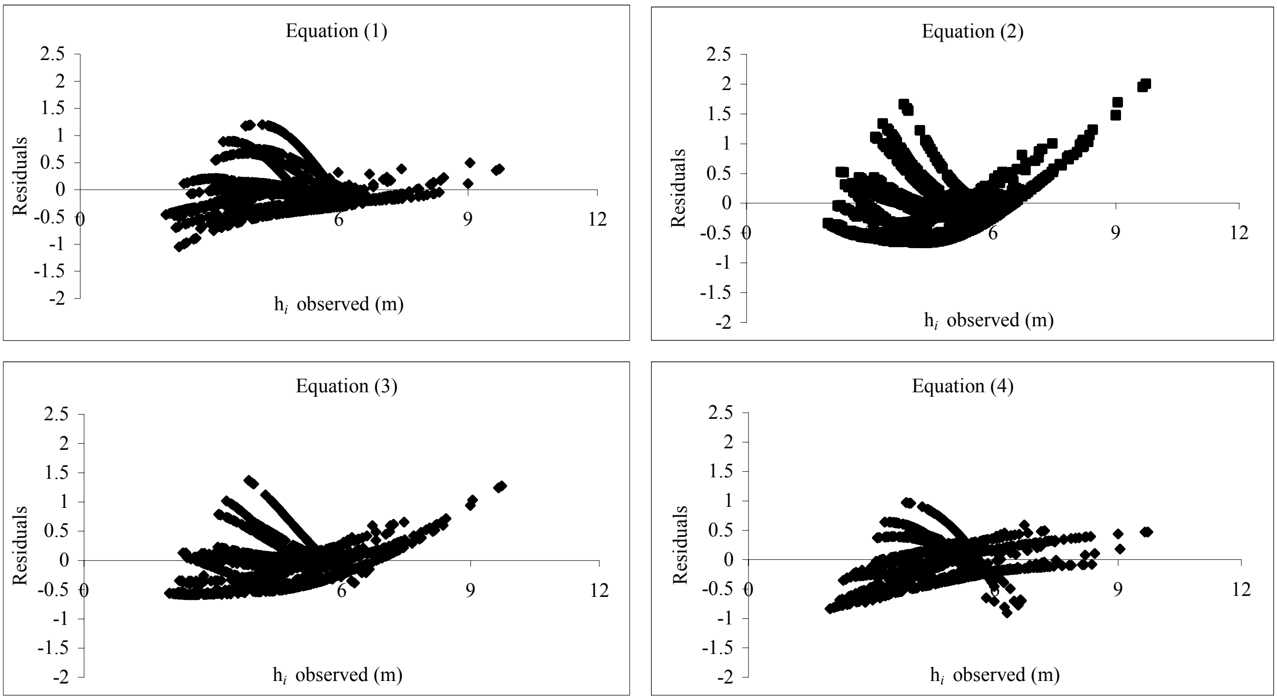

Fig. 1 Residuals for equations 1-4 plotted over observed values for fitting data set

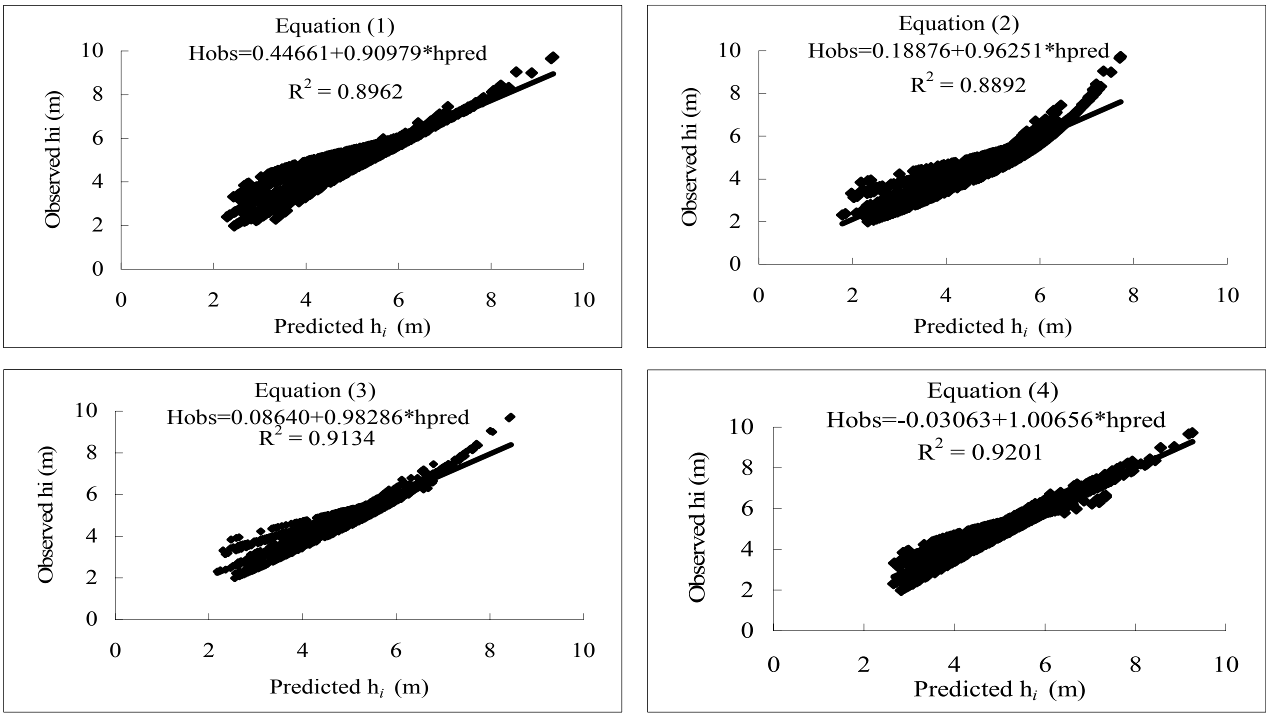

Fig. 2 Linear regression between observed and predicted values for equations 1-4 for fitting data set

For model validation, an independent data set (validation data set) was used to assess the predictive ability of different equations. The equations (with the parameter values obtained from the fitting data set) were applied to the validation data set. Table 4 compares the validation statistics for the four equations applied on the validation data set.

Table 4. Validation statistics for generalised diameter-height equations

Equation	Adj. R^2	Bias	RMSE	MPR	AICd	Rank	Final Rank
Eq. (1)	0.811 (3)	-0.09146 (1)	0.47815 (3)	0.18904 (3)	262.10420 (3)	13	3
Eq. (2)	0.810 (4)	-0.25237 (3)	0.47945 (4)	0.19007 (4)	266.74360 (4)	19	4
Eq. (3)	0.860 (2)	-0.23572 (2)	0.41116 (2)	0.13978 (2)	5.28459 (2)	10	2
Eq. (4)	0.861 (1)	-0.25851 (4)	0.41012 (1)	0.13908 (1)	0.00000 (1)	8	1

Values in the parentheses are the ranks

Equation 1 had minimum bias while Eq. (4) performed best in terms of model efficiency (adj. R^2), error (RMSE) and model precision (MPR). The values of Akaike's criterion differences (AICd) indicated that Eq. (4) had substantial support in the model selection while Eq. (3) had considerably less support. All other equations failed badly in explaining some substantial explainable variation in the data and hence must not be given any consideration during model selection. The final ranking showed that Eq. (4) ranked first while Eq. (2) ranked last. Thus, Eq. (4) placed first in the overall performance. It may also be seen that Eq. (1) performing poorest during fitting process, occupied the third place in model validation phase replacing Eq. (2).

Fig. 3 shows the plot of residuals against the actual values obtained from the validation data set. As illustrated from the figure, the Eq. (4) produced the best results and the least dispersion of the residuals while the dispersion was highest for Eq. (2). The residual values varied from -1.21 to 0.87, -0.84 to 2.20, -0.66 to 1.35 and -0.62 to 1.02 for Eqs. (1), (2), (3) and (4), respectively. This result is in conformity with the ranking shown in Table 4. The Figure also indicated that for equations 2 & 3 the residuals increased with the increase of the tree height and hence these models failed to predict height of larger trees with greater accuracy.

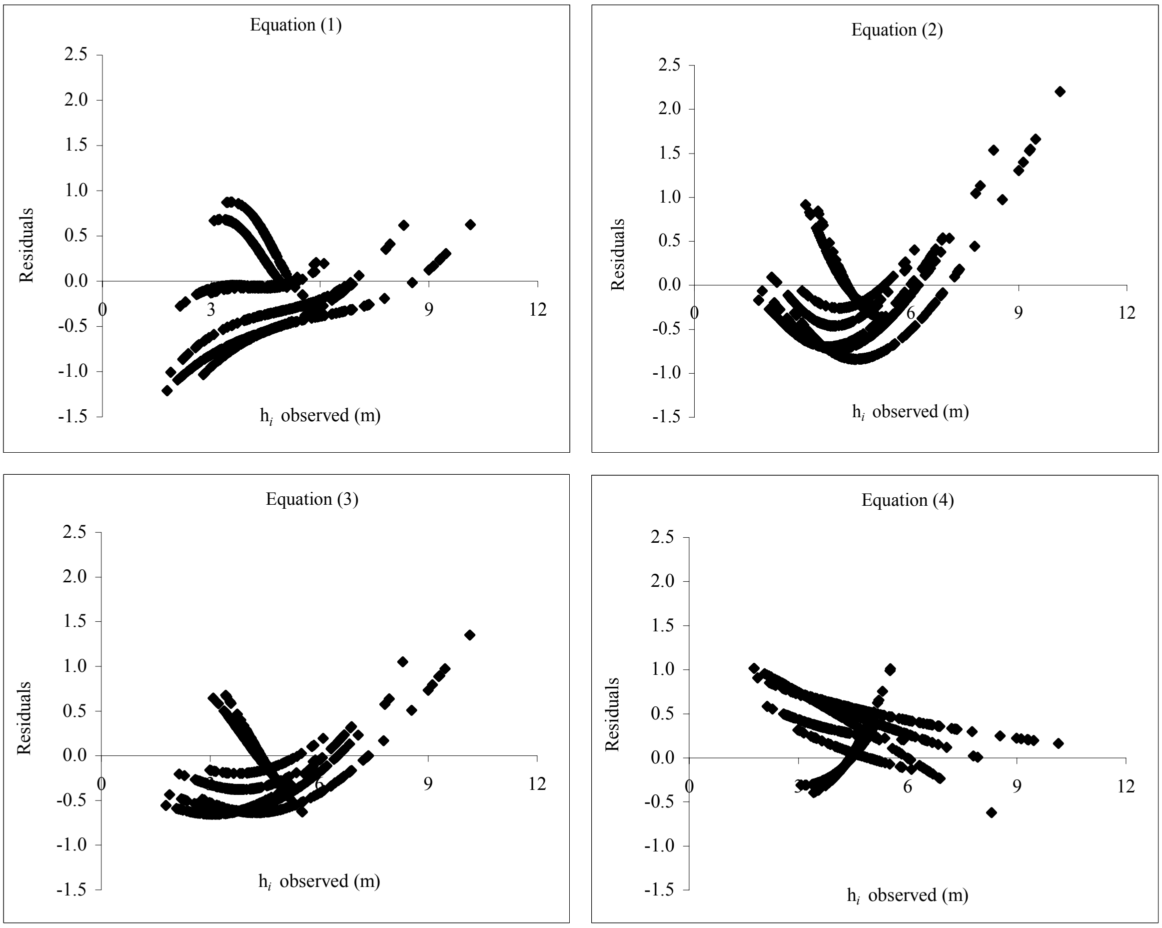

Fig. 3 Residuals for Eqs. (1)–(4) plotted over observed values for validation data set

In the present study we prior to recommend use of Eq. (4) based on the fit and validation statistics. The final equation based on pooling the fitting and validating data set is given below:

$$h_i = 1.3 + 0.70857\, H_0^{0.08300}\, d_i^{0.43503 H_0^{0.24053}}$$

Adj. R^2 = 0.913; RMSE = 0.34134

Conclusions

Four generalised diameter-height equations were tested for their performance and the evaluation criteria for fitting data set showed that performance of Eq. (4) was the best. Eq. (4) produced minimum value of mean residuals and root mean squared errors and model precision close to its ideal value. Finally, these equations were validated over an independent data set and again it was found that Eq. (4) performed best based on the statistical criteria. It was also noticed that Eq. (1) performed poorly during the fitting phase jumped to the third place in validation phase. Considering both the fitting as well as validating criteria, it was found that overall performance of Eq. (4) was better compared with other equations used. Moreover, Eq. (4) needed data only for dominant height and tree diameters (less independent variable) to estimate tree heights while other equations needed more independent variables for prediction of tree height. Thus, Eq. (4) is easy to use and may be preferred.

References

Anon. 1994. ROHIDA (*Tecomella undulata*). Dehradun, India: Indian Council of Forestry Research and Education.

Arabatzis, A.A., Burkhart, H.E. 1992. An evaluation of sampling methods and model forms estimating height-diameter relationships in loblolly pine plantations. *Forest Science*, **38**(1): 192–198.

Burnham, K.P., Anderson, D.R. 1998. Model Selection and Inference. A Practical Information-Theoretic Approach. New York, USA: Springer-Verlag, 353 p.

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

Draper, N.R. Smith, H. 1981. Applied Regression Analysis. New York, USA: John Wiley and Sons, 709 p.

Freese, F. 1960. Testing accuracy. *Forest Science*, **6**: 139–145.

Gadow, K.v., Hui, G.Y. 1993. Zur Entwicklung von Einheitshöhenkurven am Beispiel der Baumart Cunninghamia lanceolata. *Allgemeine Forst und Jagd Zeitung*, **110**(2): 41–48.

Gadow, K.v., Hui, G.Y. 1999. Modelling Forest Development. Dordrecht, The Netherlands: Kluwer Academic Publisher, 213 p.

Hasenauer, H. 1999. Statistische Probleme der Höhenzuwachsberechnung. *Centralblatt für des gesamte Forstwesen*, **116**(1/2): 91–104.

Hradetzky, J. 1999. Höhenermittlung bei Betriebsinventuren in Baden-Württemberg. *Centralblatt für des gesamte Forstwesen*, **116**(1/2): 119–128.

Kahn, M., Dursky, J. 1999. Höhenzuwachsfunktionen für Einzelbaummodelle auf der Grundlage quasirrealer Baumhöhenzuwächse. *Centralblatt für des gesamte Forstwesen*, **116**(1/2): 105–118.

Kramer, H., Akça, A. 1987. Leitfaden zur Waldmeßlehre. Frankfurt a.M, Germany: J.D. Sauerländer's Verlag, 266 p.

Michailoff, J.L. 1943. Zahlenmässiges Verfahren für die Ausführung der Bestandeshöhenkurven. *Forst-wissenschaftliches Centralbatt - Tharandter Forstliches Jahrbuch,* **6**: 273–279.

Mirkovich, D. 1958. Normale visinske krive za chrast kitnak i bukvu v NR Srbiji. Glasnik sumarskog faculteta 13, Zagreb (cited in: Wenk, G., Antanaitis, V. and Smelko, S. 1990. Waldertragslehre. Berlin, Germany: Deutscher Landwirtschaftsverlag, 448 p.).

Páscoa, F. 1987. Estructura, crecimento e produção em povomentos de pinheiro bravo. Um modelo de simulação. Ph.D. Thesis, Universidade Técnica de Lisboa, 241 p.

Pienaar, L.V. 1991. PMRC yield prediction system for slash pine plantations in the Atlantic coast flatwoods. PMRC Technical Report, Athens, USA (cited in: Van Laar, A. and Akça, A. 1997. Forest Mensuration. Göttingen, Germany: Cuvillier Verlag, 418 p.).

Prodan, M. 1965. Holzmeßlehre. Frankfurt a.M., Germany: J.D. Sauerländer's Verlag, 644 p.

Prodan, M., Peters, R., Cox, F., Real, P. 1997. Mensura Forestal. Instituto Interamericano de Cooperación para la Agricultura (IICA). San José, Costa Rica: Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH, 561 p.

Richards, F.J. 1959. A flexible growth function for empirical use. *Journal of Experimental Botany*, **10**: 290–300.

Schumacher, F.X. 1939. A new growth curve and its application to timber yield studies. *Journal of Forestry*, **37**: 819–820.

Schröder, J., Álvarez, J.G. 2001. Comparing the performance of generalized diameter-height equations for Maritime pine in northwestern Spain. *Forst-wissenschaftliches Centralblatt*, **120**: 18–23.

Statistica 1994. Statistica Version 4.5. StatSoft, Inc., Tulsa, Oklahoma, USA.

Van Laar, A. and Akça, A. 1997. Forest Mensuration. Göttingen, Germany: Cuvillier Verlag, 418 p.

Wenk, G., Antanaitis, V., Smelko, S. 1990. Waldertragslehre. Deutscher Landwirtschaftsverlag. Berlin., 448 p.