RESEARCH ARTICLE Agricultural Science Digest, Volume 43 Issue 2: 170-175 (April 2023)

Predictive Modelling of Height and Diameter Relationships of Himalayan Chir Pine

Afshan Tabassum[1], M. Iqbal Jeelani[1], Manish Sharma[1], Khalid Ul Islam Rather[1], Imran Rashid[1], Mansha Gul[1]

10.18805/ag.D-5555

ABSTRACT

Background: In this study 11 height and diameter prediction models were fitted and evaluated for Himalayan Chir pine (*Pinus roxburghii*) in Jammu region of UT J and K (India). The data were collected from 50 permanent sample plots in uneven aged stands of *Pinus roxburghii* and total of 500 individual tree height and diameter measurement were used for this study. At initial stage all the models fitted resulted in significant coefficients, besides various selection criteria's were also used to test the predictive performance of fitted models. The results of these criteria were generated from various libraries of R studio (version 3.5.1, 2018). The models were further cross validated and results revealed Manfred (MG) and Michaelis-Menten2 (MJ) models described the highest amount of height variation in terms of fit statistics and more crucially with lowest prediction error rate as compared to other models.

Methods: The study was carried out in Jammu region of UT J and K (India). Data used in this study were collected on 50 permanent sample plots of 0.25 ha in size. In order to achieve stipulated objectives, Height diameter data on 500 trees from Jammu forest division was utilized in this study.

Result: The summary statistics of height and diameter variable and the overall summary of the coefficients of various height and diameter models in Jammu forest division are presented. Almost all the coefficients of the statistical models were statistically significant which is an indication that fitted models are capturing the height diameter relationship an important aspect in context to biological realism.

Key words: Cross validation, Diameter, Model, Prediction error rate, R Studio.

INTRODUCTION

Pinus roxburghii (Sarg.) is named after William Roxburgh, also known as long leaf pine or chir pine and belongs to *Pinus* section *diploxylon*, characterized by hard timber and two xylem bundles, In India it occurs naturally in the outer ranges and valleys of the Himalayas; on the ridges of Shiwalik hills flanking the Himalayas, from altitude of 500-2500 m and covers approximately 6,77,813 ha area in the states *viz.* Himachal Pradesh, Jammu and Kashmir and Uttarakhand. Economically, pines are important source of wood, paper, resins, charcoal, food (particularly seeds) and ornamentals (dried cones *etc.*). In India major share of resin production comes from Himachal Pradesh, Uttarakhand and Jammu and Kashmir (Sharma, 2001). Chir pine provides a variety of wide ranged goods and services to the people both at local and national level. Hence its management is very important which requires the knowledge of height and diameter relationships. Various studies have revealed that these models adequately fit height diameter relationships, but sometimes they may involve various types of errors, while dealing with varying aged stands. In order to correct these situations the techniques of validation are adopted which is used when one wants to estimates how accurately statistical models will perform in practice. The idea of validation was originated in 1930s (Larson, 1931) and further generalised by (Mosteller and Turkey, 1968). In this study fitting and validation of models were carried out in R studio, which was developed in the year (2011) by J.J Allaire (R Development Core Team 2016). R has become the de-facto system for statistical analysis and graphics developed by Ross Ihaka and Robert Gentleman in the year 1995 andit handles large data sets much better than it used to, say even a decade earlier and has become a lot more versatile. Jeelani *et al.* (2015) compared linear regression model based on various sampling techniques like simple random sampling (SRS), systematic sampling (SYS) and rank set sampling (RSS) under validation technique (LOOCV). The result revealed that there was consistency in the measure of R^2, Adj R^2 and RMSE in case of RSS as compared to SRS and SYS. Results were also supported with an empirical study involving a real data set generated from *Pinus wallichiana*

[1]Division of Statistics and Computer Science, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Chatha-180 009, Jammu and Kashmir, India.

Corresponding Author: Afshan Tabassum, Division of Statistics and Computer Science, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Chatha-180 009, Jammu and Kashmir, India. Email: tabassumafshan85@gmail.com

How to cite this article: Tabassum, A., Jeelani, M.I., Sharma, M., Rather, K.U.I., Rashid, I. and Gul, M. (2023). Predictive Modelling of Height and Diameter Relationships of Himalayan Chir Pine. Agricultural Science Digest. 43(2): 170-175. doi: 10.18805/ag.D-5555.

Submitted: 17-01-2022 **Accepted:** 11-07-2022 **Online:** 23-07-2022

taken from Kashmir, India. Hasanzad *et al.* (2016) studied the relationship between height and diameter of velvet maple. A complete list of the selected height diameter models was used and nineteen candidate models were considered. Various criteria were chosen and applied to evaluate the predictive performance of the models. Jeelani *et al.* (2017) studied k-fold cross validation method for performance evaluation of different regression models. Regression analysis was used to study the relationship between fodder yield (dependent variable) and other parameters. The models were validated for its adequacy through different criteria. Jeelani *et al.* (2018), Kumar *et al.* (2021), Harsh *et al.* (2022) studied the relationship between fodder yield and other parameters of *Grewia optiva* in Jammu region. On the basis of set criteria, the models were ranked. This study aim at fitting, evaluating and validating different statistical models so that they can be used to predict height and diameter relationship in *pinus roxburghii* of Jammu region of UT J and K (India).

MATERIALS AND METHODS

The study was carried out in Jammu region of UT J and K (India). Data used in this study were collected on 50 permanent sample plots of 0.25 ha in size.In order to achieve stipulated objectives, Height diameter data on 500 trees from Jammu forest division was utilized in this study. The forest composition of Chir in terms of area in Jammu forest division is given in Table 1.

In view of the above various height and diameter statistical models were used to study the tree height and diameter relationships of chir trees in each division. The functional form of the models used in this study is as follows:

$$H_i = 1.3 + F(D_i, b) + \varepsilon_i$$

Where,

H_i = i^{th} observation of the response variable tree height (m).

D_i = i^{th} observation of the predictor variable diameter at breast height (cm).

b = A vector of model parameters.

ε_i = Unexplained error, that it is assumed to be independent and normally distributed with mean zero and a constant variance.

A constant, 1.3 (which is bh in models below) is added to the model to avoid prediction of a H_i less than 1.3 m when Di approaches zero.

Based on the above functional form eleven height diameter models were used in this study, whose description is given in Table 2.

Then adequacy of the fitted models was tested by using different selection criteria like AIC, BIC, RMSE, (Adj. R^2), *etc.* which is given below:

Adjusted R^2 (Adj. R^2)

$$R^2 \text{ adj} = 1 - (1 - R^2) \times \left(\frac{n - 1}{n - k - 1}\right)$$

Where,

R^2 = Coefficient of determination.

n = Number ofobservation and *k* is number of parameters.

Root mean square error (RMSE)

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

Where,

n = Total number of observations.

y_i = Actual observation.

$\hat{y}$ = Predicted value.

$\bar{y}$ = Mean of observed value.

Mean absolute error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Where,

n = Total number of observations.

y_i = Actual observation.

$\hat{y}_i$ = Predicted value

$\bar{y}$ = Mean of observed value, Raghav *et al.* (2022).

Akaike information criterion (AIC)

$$AIC = n \times \ln(RMSE) + 2k$$

Where,

n = Number of observations.

k = Number of parameters.

Bayesian information criterion (BIC)

$$BIC = -2\ln(l) + \ln(n) \times k$$

Where,

l = Maximum likelihood value of the model.

k = Number of degrees of freedom (or independently adjusted parameters) in the model.

n = Number of observations (total sample size).

Residual standard error (RSE)

$$RSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{df}}$$

y = The observed value.

$\hat{y}_i$ = The predicted value.

df = The degrees of freedom.

Further, the assumption of normality of errors of fitted models was accessed by Shapiro Wilk test of significance. Then to check the predictive performance of the fitted models cross validation method was used and data was randomly partitioned into training and testing sets, where 80 per cent

Table 1: Forest area by composition of Chir Pine in Jammu forest division.

Province	Circle	Forest division	(Area in Ha)
Jammu	East Circle	Jammu	31492
		Kathua	12002
		Billawar	20917
		Ramnagar	12104
		Udhampur	19815
		Total	96330

Table 2: Functional form of the height diameter models.

Model code	Functional form	Name
Model A (MA)	$h = bh + \frac{d^2}{(a + bd)^2}$	Manfred Naslund (1937)
Model B (MB)	$h = bh + a\left(\frac{d}{1 + d}\right)^b$	Curtis (1967)
Model C (MC)	$h = bh + ae^{-bd-1}$	Huang (1992)
Model D (MD)	$h = bh + a(1 - e^{-bd})$	Meyer (1940)
Model E (ME)	$h = bh + ad^b$	Zeide (1993)
Model F (MF)	$h = bh + \frac{d^2}{(a + e^b d)^2}$	Huang (1992)
Model G (MG)	$h = bh + \frac{d^2}{(ae^a + bd)^2}$	Huang (1992)
Model H (MH)	$h = bh + \frac{d^2}{(e^a + e^b d)^2}$	Huang (1992)
Model I (MI)	$h = bh + \frac{ad}{b + d}$	Michaelis-Menten (1913)
Model J (MJ)	$h = bh + \frac{d}{a + bd}$	Huang (1992)
Model K (MK)	$h = bh + \exp\left(a + \frac{b}{d + 1}\right)$	Wykoff (1982)

Where, *d* = A vector of tree diameters, usually in cm; *h* = A vector of tree heights, usually in meters; *a* and *b* = Parameters; *bh* is the applied height for the measurement of tree diameter (so called breast height).

of data was used for training and the remaining 20 per cent of the data for testing. Statistical analysis of above models was carried out in R studio (version 3.5.1, 2018) utilising its various libraries. Further the function geom_line and ggtitle of R studio was used for creating these GG-plots.

RESULTS AND DISCUSSION

The summary statistics of height and diameter variable and the overall summary of the coefficients of various height and diameter models in Jammu forest division are presented in Table 3 and 4. A glance of Table 4 disclose that almost all the coefficients of the statistical models were statistically significant which is an indication that fitted models are capturing the height diameter relationship an important aspect in context to biological realism. Also this table revealed the performance of height diameter models across the forest division utilizing various selection criteria like RMSE, MAE, BIAS, RSE, AIC, BIC, MAE, RMSE, R^2 and Adj. R^2. Among significant fitted models MG (Manfred N3) height diameter model was found out to be the best model with lower values of RMSE, MAE, AIC and BIC, also with non-significant Shapiro Wilk test for normality of errors.

In order to access the predictive performance of fitted models, cross validation method (80:20) was adopted, utilizing 80 per cent data as training set and remaining 20 per cent as testing set. Table 5 reveal the performance criteria of models under training and testing data set in case

Table 3: Descriptive summary.

	Jammu	
	Diameter (cm)	Height (m)
Mean	37.35	25.97
Median	36.12	24.47
Kurtosis	3.98	3.09
Min	2.30	1.22
Max	69.20	67.99

of Jammu forest division. A perusal of Table 5 revealed that MG (Manfred N3) and MJ (Michaelis-Menten2) performed better across both data sets in Jammu forest division as both the models resulted in lower values prediction error rates (PER).

Apart from this all the models were ranked based on prediction error rates (PER), a perusal of which is given in Table 6.

The evaluation of height diameter models with respect to training and testing data sets were also visualized graphically with the help of library (ggplot2) in R studio. A perusal of GGplots across Jammu forest division with respect to various performance criteria are given in Fig 1, 2 and 3.

At primary stage data on height and diameter variables of 500 Chir Pine trees were takenand accordingly eleven height diameter models were fitted on the data. Initial analysis revealed that almost all the parameters across the

Table 4: Estimates of parameter and performance criteria of fitted height and diameter models in Jammu forest division.

Models	A	B	RSE	AIC	BIC	MAE	RMSE	R^2	Adj. R^2	Shapiro wilk test
MA	5.02*	0.05*	5.52	571.75	578.54	3.53	5.46	0.77	0.72	0.90
MB	1.66	56.23	6.71	670.28	678.09	4.78	6.70	0.86	0.84	0.79*
MC	1.81**	54.94*	5.10	526.01	532.26	3.24	5.05	0.79	0.76	0.90
MD	1.02*	0.11**	6.37	658.19	666.00	3.63	6.30	0.75	0.72	0.99
ME	5.02*	2.94*	6.35	657.51	665.33	3.57	6.28	0.75	0.73	0.80
MF	0.16	-1.42	6.74	669.61	677.43	4.73	6.68	0.66	0.63	0.70*
MG	1.61*	0.85*	4.23	438.34	443.55	2.48	4.19	0.88	0.85	0.90
MH	1.61*	-2.94**	5.08	512.59	518.60	2.98	5.03	0.86	0.84	0.91
MI	7.45*	-18.48*	6.56	664.02	671.84	3.72	6.49	0.69	0.66	0.91
MJ	1.81**	-0.81*	5.03	496.97	502.76	2.83	4.98	0.86	0.84	0.91
MK	4.91**	-7.55*	6.63	666.37	674.18	3.76	6.57	0.66	0.64	0.81

*: Significant at 5% level of significance; **: Significant at 1% level of significance.

Table 5: Evaluation of height diameter models under training and testing data sets in Jammu forest division.

Models	Training data						Testing data					
	AIC	BIC	MAE	RMSE	Adj. R^2	PER	AIC	BIC	MAE	RMSE	Adj. R^2	PER
MA	496.97	502.76	3.67	5.67	0.68	0.37	110.24	113.84	2.72	4.43	0.79	0.33
MB	541.92	549.06	4.85	6.89	0.57	0.46	132.65	135.63	4.18	5.73	0.65	0.45
MC	463.59	469.80	3.39	5.25	0.70	0.34	102.34	107.58	2.70	4.26	0.81	0.32
MD	534.13	541.28	3.77	6.56	0.66	0.45	132.81	142.80	3.13	5.10	0.77	0.40
ME	533.11	540.27	3.70	6.52	0.67	0.43	125.93	137.41	3.11	4.96	0.78	0.38
MF	541.42	548.57	4.80	6.87	0.57	0.78	156.89	178.96	4.83	5.93	0.68	0.65
MG	355.42	360.18	2.59	4.35	0.79	0.29	85.28	87.27	2.07	3.40	0.90	0.25
MH	426.50	432.22	3.11	5.22	0.78	0.34	98.37	104.73	2.48	4.08	0.89	0.30
MI	538.79	545.94	3.85	6.76	0.60	0.45	137.69	158.19	3.27	5.32	0.71	0.48
MJ	415.75	421.24	2.99	5.12	0.78	0.35	97.54	99.84	2.09	3.81	0.89	0.28
MK	540.47	547.62	3.89	6.83	0.58	0.49	148.47	163.57	4.12	5.71	0.69	0.43

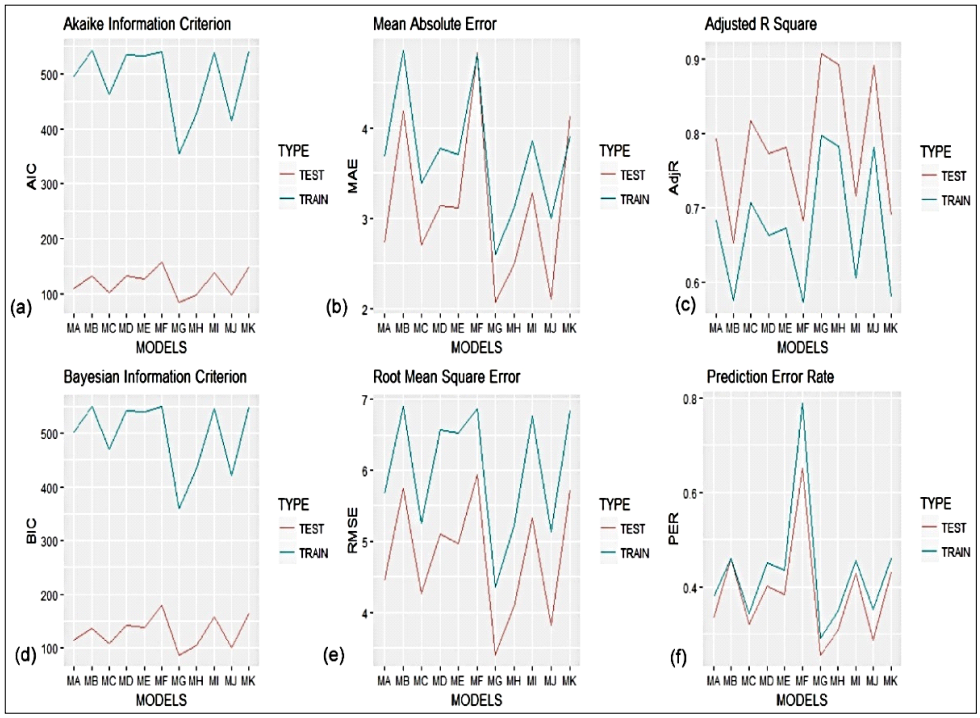


Fig 1: GG plots of performance criteria of models under Jammu forest division.

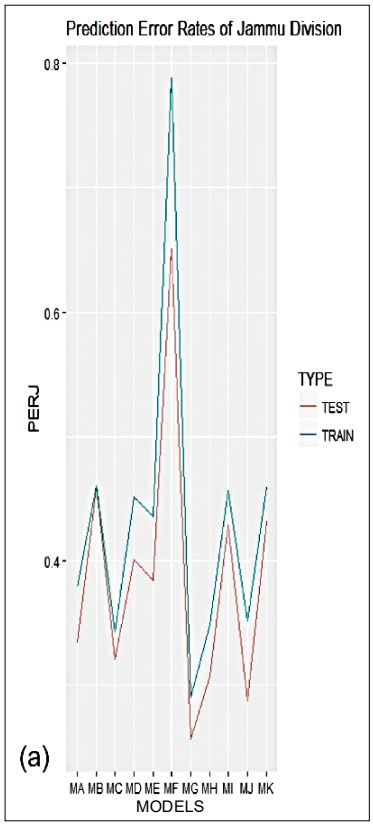
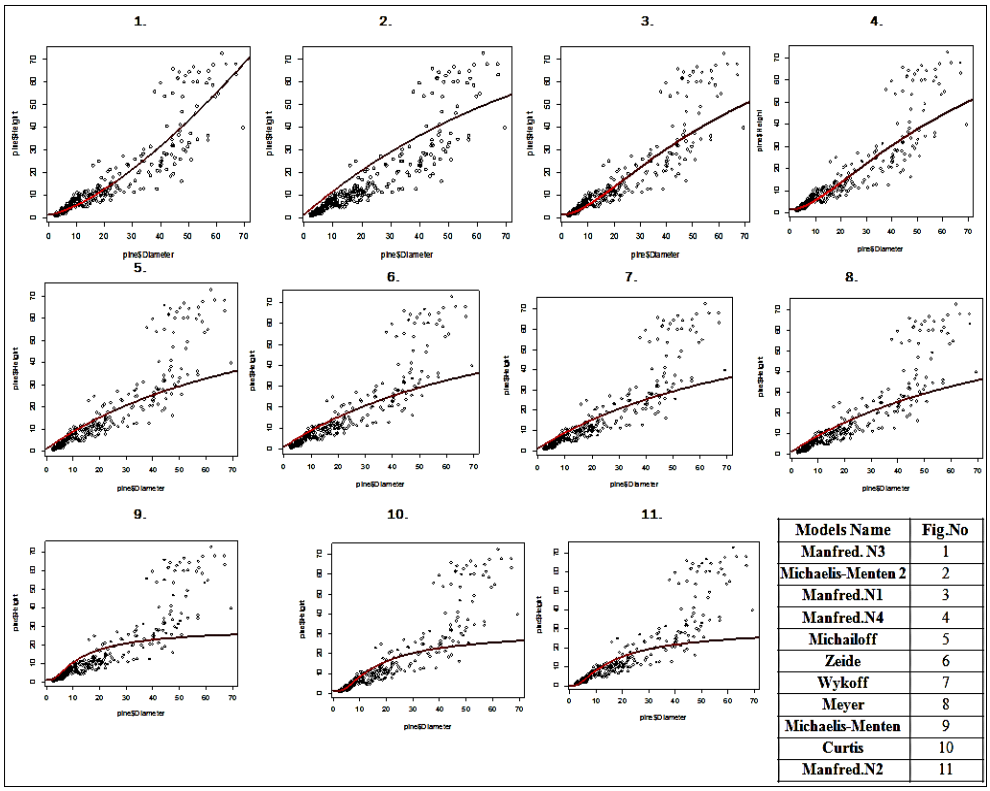


Fig 2: GG-plots of predication error rates of models under Jammu forest division.

Table 6: Summary of ranks of model performance for Chir Pine.

Rank	Jammu	
	Model	PER
1	MG	0.25
2	MJ	0.28
3	MH	0.30
4	MC	0.32
5	MA	0.33
6	ME	0.38
7	MD	0.40
8	MI	0.42
9	MK	0.43
10	MB	0.45
11	MF	0.65

models were found out to be significant. At second stage various selection criteria's like RMSE, MAE and BIAS *etc.* were used to study the predictive performance of fitted height diameter models. Various functions of R were utilized with the help of various libraries like minpack.lm, Metrics, caret, tidyverse and nlme of R to generate the results of various selection criteria's. It was revealed that RSE ranged from 4.23 to 6.74, while as AIC and BIC ranged from 438.34 to 670.28 and 443.55 to 678.09. Further MAE and RMSE ranged from 2.48 to 4.78 and 4.19 to 6.70, while as Adj.R^2 and R^2 from 0.66 to 0.85 and 0.66 to 0.88. Based on these results MG (Manfered N3) and MJ (Michaelis-Menten2) height diameter models performed well in Jammu forest



Models Name	Fig.No
Manfred. N3	1
Michaelis-Menten 2	2
Manfred.N1	3
Manfred.N4	4
Michailoff	5
Zeide	6
Wykoff	7
Meyer	8
Michaelis-Menten	9
Curtis	10
Manfred.N2	11

Fig 3: GG-plots of predicted height curves of the clubbed data.

division among all the models used in this study. Further the prediction error rate (PER) ranged from 0.29 to 0.77 and 0.25 to 0.65 in case of training and testing data sets, which revealed that results were in favour of MG (Manfered N3) and MJ (Michaelis-Menten2) model as both models resulted in lower values of PER, also have lowest values of RMSE, MAE, AIC, BIC and higher values of Adj.R^2. Further the function geom_line and ggtitle of R studio was used for creating these GG-plots. Predicted height on Jammu forest data sets based on the fitted models were also plotted using library (ggplot2).

CONCLUSION

The relationship between tree height and diameter is an important element of forest structure that can used to describe the strategies of tree species among individuals at a particular point, to estimate timber volume and site index and is also an important variable in growth and yield modelling. Hence development of the simple and most accurate height and diameter models enables forest managers to assemble the reliable prediction of tree height in forests. To achieve this, eleven height diameter models were used in this study. Results revealed that theselection of best and most accurate models should be based on the predication error rate (PER) which was used in this study. Among the eleven models, MG (Manfred N3) and MJ (Michaelis-Menten2) recognized as the best fit and accounted for the highest amount of total height variations and more crucially with lowest prediction error rate and both models appeared to be biologically more realistic. Hence it is proposed that height diameter relationship should be known in advance for any forest policy and planning as the information about such relationships is essential for better understanding of tree stand dynamics.

Conflict of interest: None.

REFERENCES

Curtis, R.O. (1967). Height-diameter and height-diameter-age equations for second growth Douglas fir. Forest Science. 13: 365-375.

Harsh, M., Pawan, K., Avinash C.R., Kaushal, R. and Anand, K.G. (2022). Evaluation of *Grewia optiva* clones for fodder yield under North Western Himalayas conditions. Journal of Sustainable Forestry. DOI: 10.1080/10549811.2022.2045502.

Hasanzad, N.I., Alavi, S.J., Ahmadi, M.K. and Radkarmi, M. (2016). Comparison of different non-linear models for prediction of the relationship between diameter and height of velvet maple trees in natural forests (Case study: Asalem Forests, Iran). Journal of Forest Science. 62: 65-71.

Huang, S., Titus, S.J. and Wiens, D.P. (1992). Comparison of nonlinear height-diameter functiond for major Alberta tree species. Can. J. For. Res. 22: 1297-1304. DOI: 10.1139/x92-172.

Jeelani, F., Sharma, M.K., Rizvi, S.E.H. and Jeelani, M.I. (2017). Predictive modelling and validation for estimating fodder yield of *Grewia optiva*. Malaysian Journal of Science. 36: 103-115.

Jeelani, F., Sharma, M.K., Rizvi, S.E.H. and Jeelani, M.I. (2018). A study on cross validation for model selection and estimation. International Journal of Agricultural Sciences. 14: 165-172.

Jeelani, M.I., Mir, S.A., Khan, I., Nazir, N. and Jeelani, F. (2015). Rank set sampling in improving the estimates of simple regression model. Pakistan Journal of Statistics and Operation Research. 11: 39-49.

Kumar, R.R., Chauhan, J. and Joshi, U. (2021). Social Economical and Medicinal Importanc of *Grewia optiva*. Agriculture and Food e-Newsletter. 3: 252-254.

Larson, S. (1931). The shrinkage of the coefficient of multiple correlations. Journal of Educational Psychology. 22: 45-55.

Menten, L. and Michaelis, M.I. (1913). Die kinetik der invertinwirkung. Biochem. Z. 49: 333-369.

Meyer, H.A. (1940). A mathematical expression for height curves. Journal of Forestry. 38: 415-420.

Míchal, I. (1992). Obnovaekologické stability lesù.Praha, Academia. 169. (in Czech).

Mosteller, F. and Turkey, J.W. (1968). Data Analysis, Including Statistics.In Handbook of Social Psychology. Addison-Wesley. pp. 601-720.

Naslund, M. (1937). Forest Research Institute's Thinning Experiments in Scots Pine Forests. 29: 7-9.

R Development Core Team, (2016). R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing. Vienna, Austria. R version 3.2.4 (2016-03-10). https://www.Rproject.org.

Raghav, Y.S., Mishra, P., Alakkari, K.M., Singh, M., Al Khatib, A.M.G. and Balloo, R. (2022). Modelling and forecasting of pulses production in South Asian countries and its role in nutritional security. Legume Research. 45: 454-461. DOI: 10.18805/LRF-645.

Sharma, K.R. (2001). Oleoresin tapping of pines in India. Forest Chemicals Review. 111: 10-17.

Wykoff, W.R., Crookston, N.L. and Stage, A.R. (1982). User's guide to the stand prognosis model. USDA For. Serv. Gen. Tech. Rep. 133.

Zeide, B. (1993). Analysis of growth equations [J]. Forest Science. 39: 594-616.