Modelling height-diameter relationship for Chir pine trees

R.P. Sharma[1]

Tree height-diameter relationship can be used as an important input component in growth and yield models, and description of stand dynamics. This study aims at establishing robust height-diameter models for Chir pine (*Pinus roxburghii*) trees using regression techniques. Among the twelve non-linear models fitted to height-diameter data from twenty-three Chir pine stands in Parbat and Shyangja districts, Hossfeld's model accounted for the largest proportion of height variations (R^2_{adj} = 86%), and appeared to be biologically most realistic. This model can be applied to similar stand conditions from where study data were procured.

Keywords: Chir pine, height-diameter models, model evaluation, stand attributes

The accurate information of tree height is required for both forest management and research. Diameter at breast height (dbh) and total height are the commonly measured variables in an inventory. Unlike dbh, total height is less frequently used for construction or application of forest models because measurement of dbh is more cost effective, easy and accurate than total height. An estimation of total height from height-diameter models might be a reliable option where such models are available. For height-diameter models, a representative sample of accurately measured total-height is used as the response variable and dbh as the predictor variable. Several of such models are available in the literature (e.g. Curtis, 1967; Wang and Hann 1988; Huang *et al.*, 1992, 2000; Moor *et al.*, 1996; Zhang; 1997; Fang and Bailey, 1998; Sharma and Portan, 2007; Trincado *et al.*, 2007; Newton and Amponsah, 2007; Wagle, 2007).

For a given species, height-diameter relationship differs from stand to stand due to different stand densities and site qualities, sometimes even within the same stand, variation might be high (Calama and Montero, 2004). Also, height-diameter relationship may change over time (Curtis, 1967). For more comprehensive and accurate height-diameter models, additional variables describing stand density (e.g. basal area or number of stems) and site quality (e.g. site index) should be included into the models (e.g. Sharma and Zhang, 2004; Tremesgen and Gadow, 2004; Sharma and Portan, 2007; Newton and Amponsah, 2007). However, getting information on such attributes demand a lot of resources, and therefore cannot be considered for general purpose models.

Chir pine (*Pinus roxburghii*) forest is located in a sub-tropical region with an altitude varying from 1000 m to 2000 m, and its standing volume is 6.3% of the total forest in the country (NFI/FINIDA, 1999). The economic contribution of Chir pine forest to national and local level development is valuable; and, therefore, its management is useful. For scientific management, species-specific individual tree or stand level models such as height-diameter models, site index models, growth models, and biomass and volume models need to be developed. Height-diameter models can be used as a sub-model (input) in the more comprehensive models such as biomass models, growth and yield models or their simulation systems. Modelling works for Chir pine forests in the country include Joshi, 1984; Joshi, 1985; Rauntiainen, 1992; and Sharma and Pukkala, 1990. But, none of these are height-diameter models. This study, therefore, aims at constructing height-diameter models using data from various Chir pine stands of two mid-hill districts.

Materials and methods

Data

Height-diameter data were obtained from Chir pine stands located in different localities such as Lunkhudeurali, Pakhapani, Kurgha, Falamkhani, Balakot, Bhorledanda, Ghantedeurali, Bayale, Karkineta, Khanigaun in Parbat and Syanja districts.

[1] Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences (UMB)
E-mail: ram.sharma@umb.no

Many of those stands, in that time, were managed in accordance with forest operation plans formulated by local users. The stands were visited during June-July in 2005. To cover wider variations of altitudes, aspects, slopes, stand origins (natural and plantation), stand densities, stand age and size classes and stand treatments, the selection of stands was subjective instead of using any complex sampling technique. Altogether twenty-three stands were selected for measurements.

Sample trees were chosen within each stand in such a way that the chosen trees would properly represent tree-population and micro-site of a subject stand. Measurements of deformed, top broken, suppressed, leaning and wolf trees were avoided. The number of sample trees varied from 5 to 31 within a stand, depending on variation of stand attributes. The diameter was measured with a diameter tape and total height was measured directly with a Vertex III Hypsometer. Due to higher accuracy of this instrument, the measured heights would be more precise, even on the slope. Because of not applying a plot inventory system, stand attributes such as stand density (basal area or number of stems), site index and stand ages were not recorded. The measurements of larger trees i.e., trees larger than 40 cm were mostly from old natural stands and those below 40cm were from young stands of plantation. Data summaries are presented in Table 1.

Models

Non-linear relationship between height and diameter was confirmed with a scattered plot diagram of height against dbh. Twelve different non-linear models (Table 2) were used to fit height-diameter relationship.

All those models were parsimonious (possessing few parameters), mathematically robust, and therefore have been commonly used for modelling various tree and stand attributes (e.g. Wang and Hann, 1988; Huang *et al.*, 1992, 2000; Fang and Bailey, 1998; Sharma, 2006; Sharma and Portan, 2007; Newton and Amponsah, 2007). The models in Table 2 are of the following form:

$$H_i = 1.3 + f(D_i, b) + \varepsilon_i \tag{1}$$

where H_i is the ith observation of the response variable- tree height (m), D_i is the ith observation of the predictor variable-dbh (cm), b is a vector of model parameters, and ε_i is the unexplained error, and it is assumed to be independent and normally distributed with a zero mean and a constant variance. A constant, 1.3 was added to the model to avoid prediction of a H_i less than 1.3 m when D_i approaches zero.

Analysis

Parameter estimation and model evaluation

The parameters were estimated using a non-linear least-squares procedure with full information maximum likelihood (FIML) methods in SAS/ETS PROC MODEL (SAS Institute, Inc. 2004). The fitted models were then evaluated using all of the following criteria:

1. *Significant parameter estimates*: Parameter estimates should be significantly different from zero ($p < 0.05$).

2. *Akaike information criterion (AIC):* This is one of the most reliable criteria to compare the models with different parameter numbers (Burnham and

Table 1 : Data summary (dbh class (cm): 0-10 = 0-10.99, 11-20 = 11-20.99, and so on)

Variables (dbh, cm; height, m)	Number of observations	Mean	Minimum	Maximum
dbh (0-10)		7.6	0.8	10.9
height	71	7.9	1.6	13.9
dbh (11-20)		16.1	11.0	20.9
height	227	14.6	7.0	24.3
dbh (21-30)		24.9	21.0	30.9
height	210	20.5	13.9	29.5
dbh (31-40)		36.4	31.0	40.5
height	96	25.6	18.0	31.3
dbh (41-50)		44.0	41.0	50.9
height	35	28.5	23.2	34.0
dbh (51-60)		56.2	53.6	60.1
height	5	29.4	26.5	33.0
dbh (overall)		22.9	0.8	60.1
height	644	18.3	1.6	34.0

Table 2 : Models considered

	Models *	References
M1	$H = 1.3 + aD^b$	Arabatzis and Burkhart (1992)
M2	$H = 1.3 + a \exp\left(\dfrac{b}{D}\right)$	Buford (1986)
M3	$H = 1.3 + \exp(a + bD^c)$	Wang and Hann (1988)
M4	$H = 1.3 + a + bD + cD^2$	Curtis (1967)
M5	$H = 1.3 + \dfrac{D^2}{(a + bD + cD^2)}$	Huang *et al.* (1992)
M6	$H = 1.3 + aD^{(b+cD)}$	This study
M7	$H = 1.3 + \dfrac{aD^2}{(D+b)^2}$	Hossfeld (1822)
M8	$H = 1.3 + \dfrac{D^a}{b + cD^a}$	Hossfeld (1822)
M9	$H = 1.3 + a[1 - \exp(-bD)]^c$	Richards (1959), Chapman (1961)
M10	$H = 1.3 + a[1 - \exp(-bD)]^3$	Bertalanffy (1949)
M11	$H = 1.3 + \dfrac{a}{1 + b \exp(-cD)}$	Logistic model, cited in Zeide (1993)
M12	$H = 1.3 + a[1 - \exp(-bD^c)]$	Weibull model, cited in Zeide (1993)

*H = total height (m); D = dbh (cm); and a, b, c = parameters; exp = exponent

Anderson, 2002). The smaller the AIC value, the better the model. It is defined as:

$$AIC = f(\beta) + 2p \tag{2}$$

where $f(\beta)$ is negative of the marginal log-likelihood function, β is a vector of parameter estimates, and p is the number of parameters in a model.

3. *Root mean squared error (RMSE):* It is defined as:

$$RMSE = \sqrt{\dfrac{\sum_{i=1}^{n}\left(h_i - \hat{h}_i\right)^2}{n - p}} \tag{3}$$

where h_i and h_i are the observed and predicted values for the dominant height of observation i, respectively; n is total non-missing observations used to fit the model; and p is the number of parameters in the model.

4. *Adjusted coefficient of determination* (R^2_{adj}): This shows a proportion of total variance explained by the model with the adjustment of the number of parameters, p and the number of non-missing observations, n. It is estimated as:

$$R^2_{adj} = 1 - \dfrac{(n-1)\sum_{i=1}^{n}\left(h_i - \hat{h}_i\right)^2}{(n-p)\sum_{i=1}^{n}\left(h_i - \bar{h}\right)^2} \tag{4}$$

where $\bar{h}$ is mean of the observed height, and other symbols are the same as above.

5. *Homogeneousness and independence of residuals*: The residuals with predicted height, $\hat{h}_i$ and observed height, h_i were plotted against the predicted heights and examined.

6. *Biological realism*: The curves generated with models were checked with respect to their biological realism, for example, height curves against dbh were assumed to demonstrate an approximately a sigmoid shape with clear inflection point (culmination of height growth) that occurred in an early stage and other height increment should be more than zero.

The validation of models with splitting data set was not considered in this study. Because, splitting a data set into two parts: one for calibration and the other for validation cannot be independent as they have

Table 3 : Model parameter estimates and fit statistics (n = 644)

Models	Parameter estimates			Fit statistics		
	a	b	c	AIC	RMSE (m)	R^2_{adj}
M1	1.8182	0.722		1273	2.68	0.8321
M2	37.6799	-16.2513		1292	2.72	0.8265
M3	4.6453	-6.4669	-0.4156	1236	2.60	0.8420
M4	-0.3450*	0.9876	-0.0083	1216	2.56	0.8457
M5	2.0498	0.7682	0.0183	1231	2.59	0.8430
M6	0.7793	1.073	-0.0031	1221	2.57	0.8453
M7	46.104	13.638		1236	2.60	0.8421
M8	1.2479	1.6736	0.022	1111	2.36	0.8559
M9	35.032	0.0374	1.2075	1226	2.58	0.8445
M10	26.5123	0.0976		1352	2.85	0.8095
M11	28.2951	6.5214	0.106	1231	2.59	0.8432
M12	33.5721	0.0205	1.157	1221	2.57	0.8447

*non-significant (p> 0.05)

identical statistical structures (Yang *et al.*, 2004). The validation with splitting data does not provide any additional information as compared to respective goodness of fit statistics obtained directly from models with entire data set (Kozak and Kozak, 2003). Validating models with independent data would be the best alternative. But, it was not possible to get those data due to resource limitations.

Results and discussion

Except one with model M4, all parameter estimates of the models were significant (p<0.05). Except M1, M2 and M10, other models described more than 84%

(R^2_{adj} > 0.84) of height variability (Table 3). The model M8 showed the best fits (smallest AIC and RMSE, and largest R^2_{adj}) followed by model M4. However, M4 was excluded from further analysis because one of its parameter estimates was non-significant (p>0.05). Also, M1, M2 and M10 were excluded because they demonstrated relatively poorer fit statistics.

Other remaining models showed almost identical fit statistics (Table 3). To identify the best model, graphical examination of fitted curves overlaid on observed data (Figure 1), height increment curves and residual plots (Figure 2) was performed.
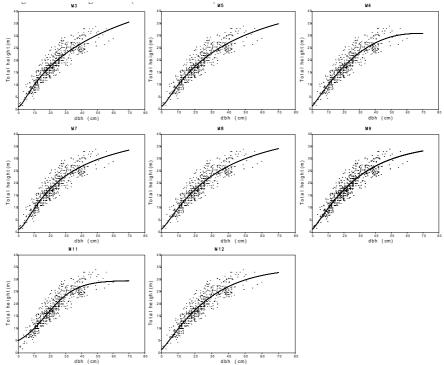

Fig. 1 : Fitted height curves (with selected models) overlaid on the observed data

The height curves generated with models M6 and M11 appeared to be biologically less relevant as they indicated almost leveling-off trends against dbh even within the observed data range where such trends were not visible. Also, M11 showed biologically illogical behaviour as its predicted heights at zero was more than 1.3 m. Therefore, both M6 and M11 were excluded from further examination. The biological logics rather than attractive fit statistics of the model should be important for biological such as height growth models (e.g. Vanclay and Skovsgaard, 1997; Ratkowsky, 1990; Schabenberger and Pierce; 2002) The height curves generated with models M5, M7, M8, M9 and M12 appeared to be identical up to 50 cm dbh. These models revealed biologically logical growth trends; for example, in early stage, height growth rate (change of height with respect to dbh) increased up to a certain dbh limit, but at the later stage it declined with increasing dbh (Figure 2, top). In the later stage, diameter growth should be faster than height growth because that a tree needs more strength to firmly withstand itself against external forces like wind blow by the thickening of its bole as tree grows to bigger and taller sizes (Khanna and Chaturvedi; 1994; Cato *et al.*, 2006).
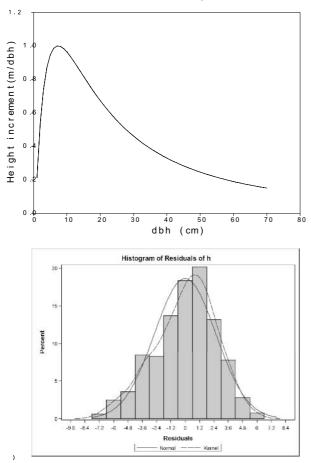

Most of the residuals were found within 95% confidence limits, and the residual histogram looked bell-shaped with symmetrical (normal) distributions (Figure 2, bottom). Thus, this hinted that there was no substantial heteroscedasticity problem with the promising models.

There was different predicting performance with five different models beyond observed data range (Figure 3). Above 50cm dbh, predicted height with model M5 is the largest, followed by models M8, M7, M9 and M12. Due to lack of independent test data, it was not possible to identify the best model that could be used for extrapolation. The most appropriate way to check a model's predictive performance beyond range of calibration data is to test it against independent data from different tree populations over the widest possible range of size, site and stand conditions (e.g. Vanclay, 1994; Vanclay and Skovsgaard, 1997; Kozak and Kozak, 2003; Yang *et al.*, 2004). Validation of these models with independent data has been left to future forest modelers.

The model M8, which described the data in the best way, can be used for the prediction of total heights only within observed data range (Table 1). The heights not described by the models may be due to the absence of input variables like stand density (basal area or stem numbers) and site quality (site index) into the models. It is because stand attributes substantially affect height-diameter relationship (Calama and Montero, 2004; Sharma and Zhang, 2004; Tremesgen and Gadow, 2000; Sharma and Portan, 2007; Newton and Amponsah, 2007). More accurate and comprehensive height-diameter models would be possible where stand attributes are incorporated into the models.
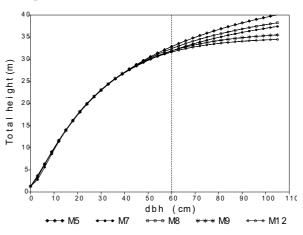

Fig. 2 : Height increment curve (top), and histogram of height residuals (bottom) with M8

Fig. 3 : Height curves beyond observed data range

Conclusion

Among the twelve models, Hossfeld's model (M8) demonstrated the best fit and accounted for the greatest proportion of total height variations (R^2_{adj} = 86%). The model was mathematically flexible and biologically robust. The model's update through re-calibration and validation against independent data from widest possible ranges of size, site and stand conditions including stand attributes across the country would be important tasks for the future.

References

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

Bertalanffy, L.V. 1949. Problems of organic growth. *Nature,* **163:** 156-158.

Buford, M.A. 1986. Height-diameter relationship at age 15 in Loblolly pine seed sources. *Forest Science,* **32:** 812-818.

Burnham, K.P., Anderson, D.R., 2002. **Model selection and inference. A practical information-theoretic approach,** Springer-Verlag, NY.

Calama, R. and Montero, G. 2004. Interregional non-linear height-diameter model with random coefficients for Stone pine in Spain. *Canadian Journal of Forest Research,* **34:** 150-163.

Cato, S. McMillan, L., Donaldson, L., Richardson, T., Echt, C. and Gardner, R. 2006. Wood formation from the base to the crown in *Pinus radiata*: Gradients of trachied wall thickness, wood density, radial growth rate and gene expression. *Plant Molecular Biology,* **60:** 565-58.

Chapman, D.G. 1961. Statistical problems in dynamics of exploited fisheries populations. In *Proceeding 4th Berkeley Symposium on Mathematical Statistics and Probability* (ed.) Neyman, J. **4:** 153-168.

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

DFRS. 1999. **Forest Resources of Nepal.** Department of Forest Research and Survey. Ministry of Forest and Soil Conservation, HMGN/ FINIDA, Report No.74, 48p.

Fang, Z. and Bailey, R.L. 1998. Height-diameter models for tropical forests on Hainan Island in Southern China. *Forest Ecology and Management,* **110:** 315-327.

Hossfeld, J.W. 1822. Mathematik für Forstmänner, Pkonomen und Cameralisten, pp. 310 [in German]

Huang, S., Price, D. and Titus, S.J. 2000. Development of ecoregion-based height-diameter models for White spruce in boreal forests. *Forest Ecology and Management,* **129:** 125-141.

Huang, S., Titus, S.J. and Wiens, D.P. 1992. Comparison of non-linear height-diameter functions for major Alberta tree species. *Canadian Journal of Forest Research,* **22:** 1297-1304.

Joshi, M.R. 1985. Prediction of biomass in a plantation stands of Chir pine (*Pinus roxburghii* Sarg.) in Nepal. MSc thesis, University of Oxford, UK.

Joshi, R.B. 1984. Total and merchantable volume equations for natural silver fir and Chir pine of Nepal. MSc thesis, University of Athens, Georgia, USA, 35p.

Khanna, L.S. and Chaturbedi, A.N., 1994. **Forest Mensuration.** International Books distributors, Dehradun, India.

Kozak, A. and Kozak, R. 2003. Does cross validation provide additional information in the evaluation of regression models? *Canadian Journal of Forest Research,* **33:** 976-987.

Moore, J.A., Zhang, L. and Stuck, D. 1996. Height–diameter equations for ten tree species in the Inland Northwest. *Western Journal Applied Forestry,* **11:** 132–137.

Newton, P.F. and Amponsah, I.G. 2007. Comparative evaluation of five height-diameter models developed for black spruce and jack pine stand-types in terms of goodness-of-fit, lack-of-fit and predictive ability. *Forest Ecology and Management,* **247:** 149-166.

Ratkowsky, D.A. 1990. **Hand Book of Non-Linear Regression Models.** Marcel Dekker, Inc. vol. 107, 241p.

Rauntiainen, O. 1992. Observations on the growth of planted *Pinus roxburghii* on high quality sites in Kathmandu Valley. *Banko Janakari,* **3 (3):** 37-42.

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

SAS Institute Inc. 2004. SAS/ETS1 9.1 **User's Guide.** SAS Institute Inc., Cary, NC.

Schabenberger, O. and Pierce, F.J. 2002. **Contemporary Statistical Models for the Plant and Soil Sciences.** CRC press, pp. 738.

Sharma, E.R. and Pukkala, T. 1990. **Volume and Biomass Prediction Equations of Forest Trees of Nepal.** Forest Survey and Statistical Division. Ministry of Forest and Soil Conservation, Kathmandu, Nepal

Sharma, M. and Portan, J. 2007. Height-diameter equations for boreal tree species in Ontario using a mixed-effects modelling apprach. *Forest Ecology and Management,* **249:** 187-198.

Sharma, M. and Zhang, S.Y. 2004. Height-diamater models using stand characteristics for *Pinus banksiana* and *Picea mariana. Scandinavian Journal of Forest Research,* **19:** 442-451.

Sharma, R.P. 2006. Modelling growing space requirement for *Alnus nepalensis* D. Don in Nepal. *Banko Janakari,* **16(2):** 30-36.

Temesgen, H. and Gadow, K.V. 2004. Generalized height–diameter models- an application for major tree species in complex stands of interior British Columbia. *European Journal of Forest Research,* **123:** 45-51.

Trincado, G., Curtis, L, Shaaf, V. and Burkhart, H.E. 2007. Regional mixed effects height-diameter models for loblolly pine (*Pinus taeda* L.) plantations. *European Journal of Forest Research,* **126:** 253-262.

Vanclay, J.K. 1994. **Modelling Forest Growth and Yield. Applications to Mixed Tropical Forests.** CAB International, Oxon, U.K., pp. 312.

Vanclay, J.K. and Skovsgaard, J.P., 1997. Evaluating forest growth model. *Ecological Modelling,* **98:** 1-12

Wagle, B. 2007. Growth of Bluepine (*Pinus wallichiana*) in Lete and Kunjo of Mustang District. MSc thesis submitted to TU/IOF, Pokhara, 55p.

Wang, C.H. and Hann, D.W., 1988. Height–diameter equations for sixteen tree species in the central western Willamette valley of Oregon. Oregon State University Forest Research Laboratory, Research Paper, 51p.

Yang, Y., Monserud, R.A., and Huang, S. 2004. An evaluation of diagnostic tests and their roles in validating forest biometric models. *Canadian Journal of Forest Research,* **34:** 619-629.

Zeide, B. 1993. Analysis of gorwth equation. *Forest Science,* **39:** 594-616.

Zhang, L. 1997. Cross-validation of non-linear growth functions for modelling tree height-diameter relationships. *Annals of Botany,* **79:** 251-257.