Allometric models for predicting above ground biomass of dominant shrub and tree species grown in semi-arid Bundelkhand region of India

*1Dibyendu Deb, 2Arunava Ghosh, 1J.P. Singh, 2Parthendu Poddar and 1R.S. Chaurasia

1Indian Grassland and Fodder Research Institute, Jhansi (UP)-284003, India
2UttarBangaKrishiViswavidyalaya, CoochBehar(W.B.)-736165, India

ABSTRACT

T*he aim of the present study was to find out suitable predictor variables for total tree biomass relationship and to fit 'allometric model for the total biomass of 5 dominant tree species of semi-arid region of India. The selection of sample sites were based on NDVI generated through IRSP6LISS-III imagery of Central Highland physiographic zone of Bundelkhan region of India, which lies between 21º 17´ N to 26º 52´ N latitudes and 74º 08´ E to 82º 49´ E longitudes. The dataset contained 302 trees with DBH ranging from 3.20 to 45.50 cm. For developing the models dbh, height and dbh² * height were used as predictor variables. All 5 species were taken for developing allometric models and after examining model residuals and site specific relationships it was found that using dbh² * height alone as the predictor variable produced the most stable model. Thus it makes regional estimation of above ground biomass production easier with precision as accurate as site specific allometry. The fitted model was then validated using model diagnostics and statistical validation techniques and tested with an independent dataset to see the accuracy of prediction.*

Keywords: Bundelkhand region, allometry, normalized difference vegetation index, residual diagnostics

INTRODUCTION

For estimation of above ground biomass and hence terrestrial carbon stock, different tree physical parameters are generally used. Tree-height measurements are expensive and time- consuming and often they are performed only on a small portion of sampled trees. The development of a relationship between total tree height and diameter at breast height (DBH) is considered crucial in forest inventories as well as in stand projection systems since it provides height predictions using an easily measured predictor variable such as DBH (Nanos *et al*., 2004). The construction of height/diameter (H/D) models is sometimes performed using mixed effect models, assuming that parameters of the model can vary randomly across stands (Nanos *et al*., 2004). The localisation of the H/D curve has been made with the use of other stand variables (Lappi, 1991, 1997) or according to the site or region where the curve is going to be used (Fulton, 1999).

Volume prediction of a particular tree using the parameters

If a particular tree has its dbh and height are estimated as D and H respectively then its volume (V) will be certainly:

$$V = a + b * D^2 H$$

But this equation involves a white noise $\in$, as the two trees having the same D and H cannot always have the same volume, this is called intrinsic variability of a particular tree equal to σ^2. When attempting to predict volumes, this intrinsic variability is supplemented by the variability due to the imprecision of the α and β parameter estimations. Thus, for a linear regression, the semi-amplitude of the confidence interval at the threshold α (typically 5%) of V is equal to (Saporta, 1990):

$$t_{n-2}\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(D^2H - \overline{D^2H_e})^2}{\sum_{i=1}^{n}(D_i^{\,2}H_i - \overline{D^2H_e})^2}}$$

where t_{n-2} is the quantile $1- \alpha/2$ of a Student's distribution with $n - 2$ degrees of freedom, D^2H_e is the empirical mean of the D^2H values measured in the sample:

$$D^2H_e = \frac{1}{n}\sum_{i=1}^{n}(D_i^{\,2}H_i)$$

and $\hat{\sigma}$ is an estimate of the standard deviation of the residuals:

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}[V_i - (a + bD_i^{\,2}H_i)]^2$$

E-mail: debiasri@gmail.com

The lowest value of this semi-amplitude (when $n \to \infty$) is $1{,}96\sigma$. So to set the precision target for the estimation as a deviation of $E\%$ from this incompressible minimum, i.e. in an approximate manner one should look for sample size n (Picard *et al*., 2012) such that:

$$1 + E \approx \sqrt{1 + \frac{1}{n} + \frac{(D^2H - \overline{D^2H_e})^2}{\sum_{i=1}^{n}(D_i^2H_i - \overline{D^2H_e})^2}}$$

These approaches to H/D modelling are based on the assumption that site conditions and previous silvicultural practices influence the relationship between total tree height and DBH. In the models that use stand variables for prediction, there does not seem to be a single or universal variable to explain variation in stand-specific parameters, since these depend on the investigator's intuition, model specification, site productivity (Fulton, 1999), and thinning regime (Zhang *et al.*, 1997). Fang *et al.* (1998), for example, used percentiles of the diameter distribution; Zakrzewski and Bella (1988) found that the quadratic mean diameter and height are the best explanatory variables, while Parresol (1992) used the basal area per hectare.

In this paper the variations in different biomass equations of 5 different dominant tree species of the study area have been examined for a common allometric equation and its validation, which can be used to estimate the aboveground biomass of the area and thereby the carbon stock.

MATERIALS AND METHODS

Data source

In the present study ground data was used as primary data which was collected on different biometric / physiological characters of sample tree and shrub species, grown at different sites representing the Central Highland physiographic zone of Madhya Pradesh and Uttar Pradesh. The sampling of tree data was carried out from total 42 sites (each of 250 x 250m size) spread over 6 districts of UP and MP and finally 4 sample plots were selected from each sites cumulating 168 permanent sample plots. The database contains information on altitude, soil type, crown density, stand height, CBH, Dia, tree-height, ht-1st-forking, DBH, radius, and basal area. Based on these biometric values total wood biomass of particular tree species can be calculated using the volume equations given in the State of Forest Report 2005, forest survey of India, Ministry of Environment & Forests, Dehradun.

Study area

The dataset used in the present study contains different biometric characters of tree species located in Bundelkhand region of India. The study area is a typical semi-arid region of the country. In the present investigation an estimation of tree biomass is done for this region and total 42 sites were selected for the collection of data, using NDVI (Normalised Difference Vegetation Index) map based on IRS imagery The study areas were located on Central Highland physiographic zone of Bundelkhan region of India, lies between 21º 17´ N to 26º 52´ N latitudes and 74º 08´ E to 82º 49´ E longitudes (Fig. 1).

Field data collection and data processing

The sampling design and identification of sample sites are broadly based on (i) forest density as per Forest Survey of India (FSI) (ISFR 2011) and (ii) NDVI based on AWiFS (IRS Resourcesat 2) data. Sample plots were laid down in different homogeneous strata on the basis of NDVI values. Four plots, each of 0.1 ha in size were selected for collection of tree samples. One pixel of AWiFS image covered nearly four plots (the plot size was 31.61 x 31.61m) marked under one selected site of approximately 0.1ha size. Nested two stage sampling approach was adopted for sample of trees (0.1 ha), shrubs (25 m^2) and herbs (1m^2). In total 302 tree samples were collected representing existing diameter range and forest types. The sample trees were randomly selected from each 5 cm class intervals of the existing dbh range so that each class has a nearly even tree distribution. These samples belonged to the 5 most abundant botanical genera. The species selected, no of trees of each species and their dbh ranges are given in table 1. Individual tree volume was calculated (from the collected field data) using site specific tree volume equations available in the literature of Forest Survey of India (ISFR 2003). Few trees species those volume equations were not readily available, geometric relationships were used to approximate the volume of standing tree bole. Biomass of each tree was then worked out by multiplying with specific gravity and biomass expansion factor (Priyanka *et al.* 2013). The mean BEF value of 1.5 was used for this study as prescribed by Brown and Luge (1992).
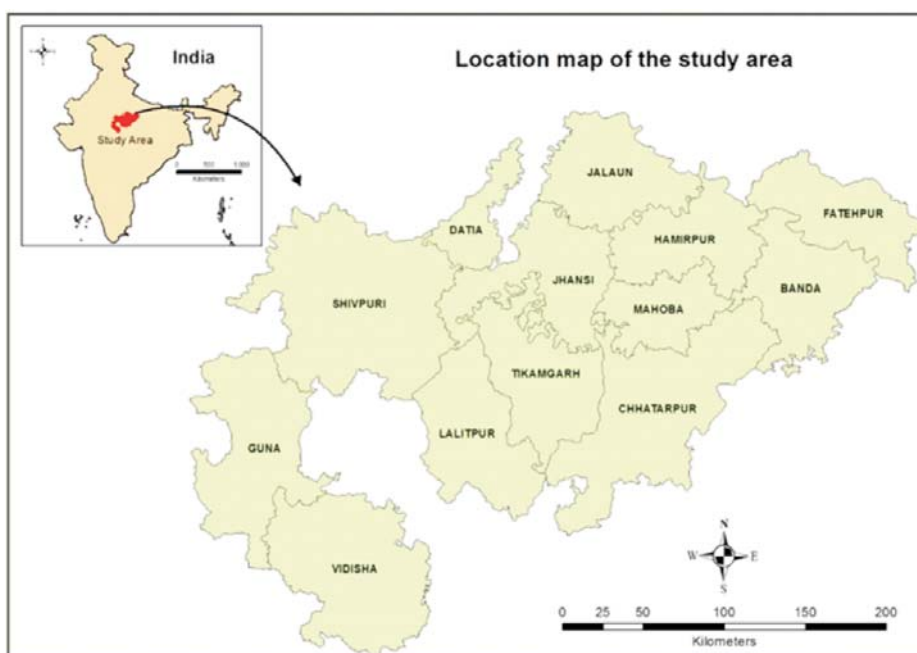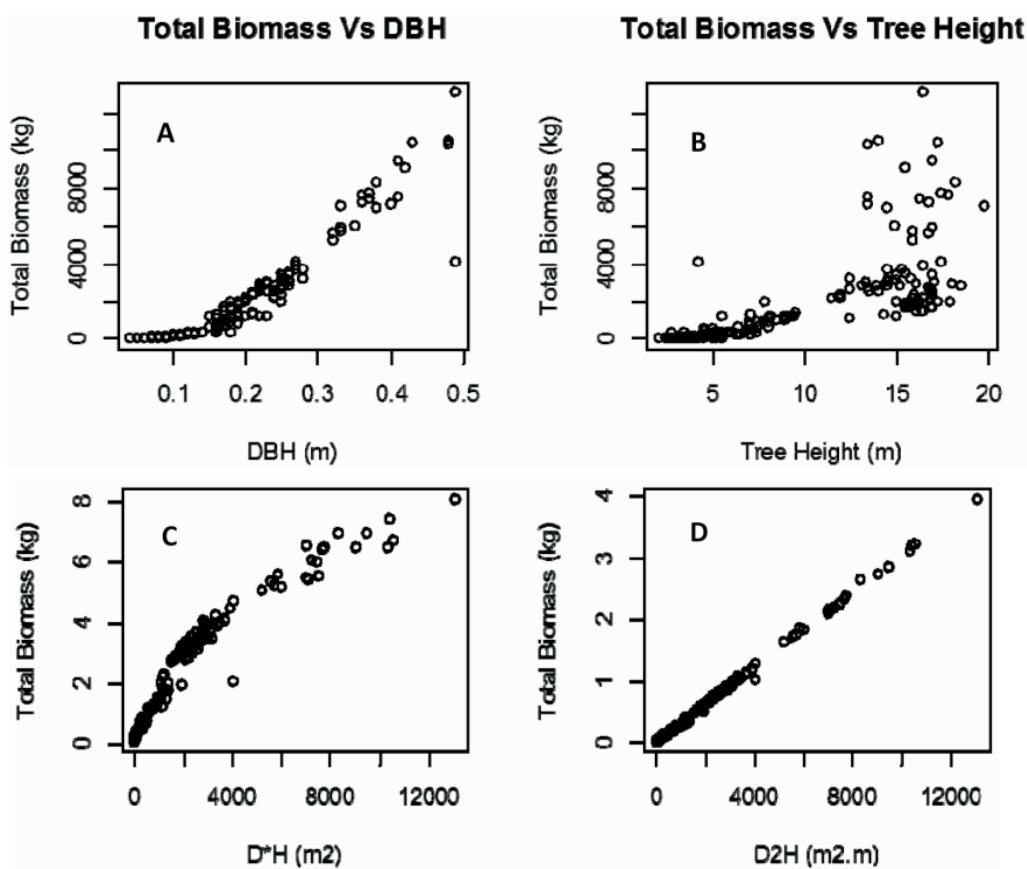
Fig.1: Study area

Fig. 2: **Scatter plot of Total biomass of individual tree against (A)DBH, (B) Tree height, (C) DBH*Height and (D)(DBH)2*Height respectively**
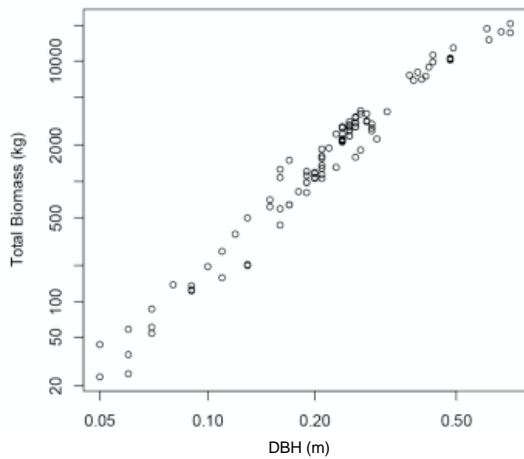
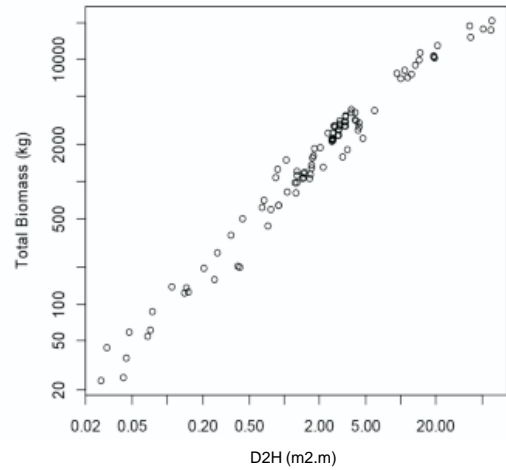**Fig. 3 :** *DBH* Vs Total Biomass of after log transformation.

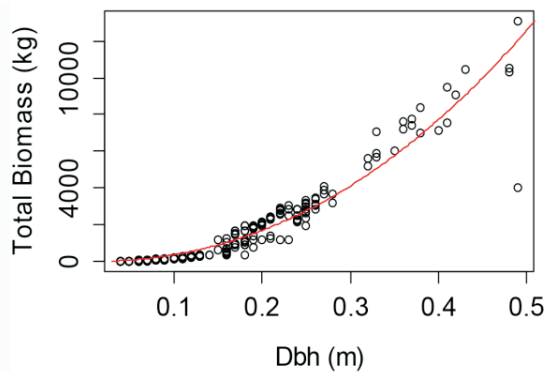**Fig. 4:** D^2H Vs Total Biomass after log

Fig. 5 : Fit plot for power model:
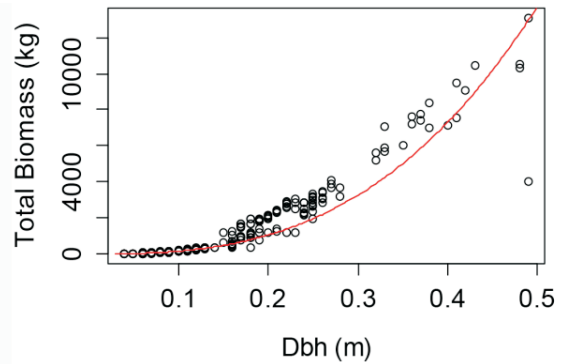*Total Biomass = 4375.00 * DBH$^{2.035}$*

Fig. 6: Fit plot for power model with combined variable:
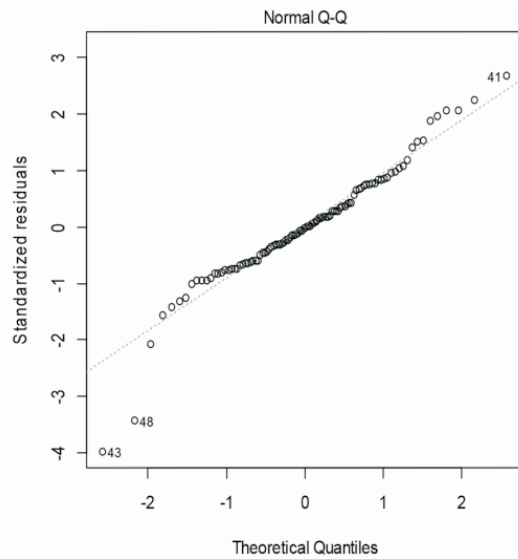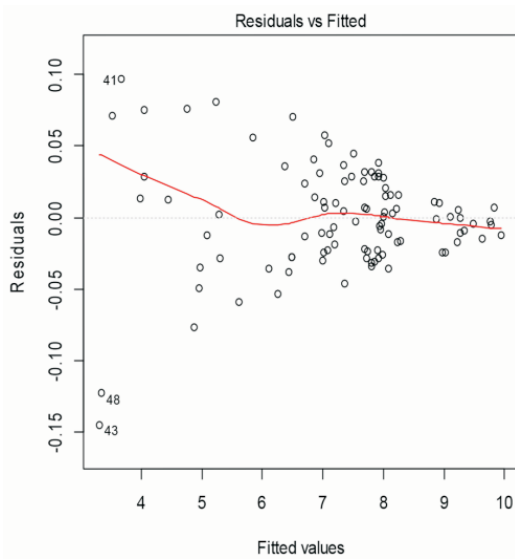*Total Biomass = 4512.00 * DBH$^{1.003}$H$^{0.887}$*

Fig. 7 : Residuals plotted against fitted values (left) and quantile–quantile plot (right) for residuals of the multiple regression of ln(B) against ln(D) and ln(H)

Table 1: Basic statistics of the samples collected

Sl. No.	Tree/Shrub species	DBH Range(cm)	CBH Range(cm)	Basal Area Range(cm2)	No. of trees
1	Azadirachtaindica	4.90 - 45.50	15.7 - 143.0	190 - 1627	70
2	Acacia catechu	3.30 - 34.50	10.6 - 108.5	80 - 9360	52
3	Dalbergiasissoo	15.40 - 38.60	48.5 - 121.5	1870 - 1174	60
4	Butea monosperma	3.20 - 40.90	10.2 - 128.6	80 - 1316	60
5	Tectonagrandis	22.90 - 38.70	–	4140 - 1176	60

Graphical exploration of the data

To achieve the objective, first of all graphical exploration was performed as the first step in analysis. It deals with visually assessing the existing relationship between the variables in order to gain an idea about what kind of model were to be fit. Here the points are plotted in a graph as cluster, where the dependent/response variable was plotted on the Y axis and the independent /effect variable (different tree parameters) was plotted on the X axis. We have here total biomass as the response variable and several other tree parameters like DBH, height, CBH, Basal diameter and combined variables viz. *DBH * Height and DBH² * H* etc. as effect variables. To visualize the relation between biomass against DBH, Height and combined variables, they are plotted first (Fig 2). From the plot it is very clear that the relation between them is not linear and the variance in biomass observations increases with the increase in DBH values. The other variables like tree height and CBH were also plotted in the same manner as DBH. Tree height appeared as the worst predictor to be considered for biomass model.

As there are more than one effect variables here and when the no. of effect variables is larger than 3 then it becomes quite difficult to analyze the relationship between the response variable and each of the effect variables specially if the effect variables are not independent and it also requires huge data set to allow a valid graphical explorations of the relations between variables. So an effort was made to combine them to build a single combined effect variable. As different references suggested trunk volume of any tree depends on the DBH and height (Louppe *et al.*1994), so here the effect variables, DBH and *DBH² * H* were selected to analyze their relation with total biomass and then model that accordingly. From the scatter plots (Fig 2(A) and 2(C) respectively) we can see that the relation between biomass and both these two effect variables is nonlinear and variance of biomass increases with *DBH * H* and *DBH² * H* and also from figure 3 and 4 it can be observed that log transformation of these two explanatory variables made the scatter plots linear, which suggests that the relationship is of exponential type between them.

RESULTS

First of all linear model is tried. The exploratory analysis showed that the relation between ln(total biomass) and *ln(DBH)* is a linear one (Figure 2), with a variance of l*n(total biomass)* that was approximately constant. The linear model calculated taking DBH as independent variable has the expression *ln(total biomass)* $= -0.71 + 2.61 * \ln(DBH) + \in$ with Residual standard error : 876.1, $R^2 = 0.8587$ and the model is highly significant. Whereas considering *DBH² * H* as independent variable the linear model calculated as *ln(Total Biomass)* $= 7.843 + 1.084 * ln(DBH^2 * H)$ with less residual standard error 0.3192 and a higher value of multiple R^2 : 0.9247.

As it was clearly evident from figure 2 that the variability of biomass increases with the increase in tree DBH, so the variance of error also varies/increases with DBH which is a contradiction to the homoscedasticity hypothesis required for regression. Though the log transformation stabilizes the error variance and the error variance becomes constant but unfortunately it does not remain any more in liner form as the dependent variable Total Biomass becomes non linearly dependent on the coefficients a_0, a_1 and a_2. So in this occasion nonlinear modelling technique was to be followed, which could be written as:

$$Y = f(X_1, X_2, \ldots, X_n; \theta) + \in$$

where Y is the response variable, $X_1, ..., X_n$ are the effect variables, θ is the vector of all the model coefficients, ε is the residual error, and f is a function. Here a power function of the form:

$$Y = a * X_b$$

was tried as it may prove a good fit according to the scatter plots prepared earlier during graphical exploration and this form also the most common form used in biomass studies (Ter-Mikaelian and Korzukhin 1997, Pilli *et al*., 2006). The power model or allometric model resulted even better fit in case of both the independent variables as in the table below:

Model equation	Residual standard error	R^2
Total Biomass = $4375.00 * DBH^{2.035}$	1182	0.9485
Total Biomass = $4512.00 * (DBH^2)^{1.003} H^{0.887}$	134.9	0.9656

The fit plots of both the models are given in figure 5 and 6 respectively.

To ensure that the model also satisfies the assumptions of regression and for validation, it was essential to perform residual diagnostics to confirm that of regression the assumptions are satisfied (Fig 7). Even though there is a hint of slight structure in the residual vs fitted values, it may be considered that the errors of the fitted model are normal with constant variance.

DISCUSSION

Though it was found in the present study that for predicting standing tree biomass or estimation of aboveground biomass the allometric/power model proved to be the best one with $DBH^2 * H$ as effect variable, use of this model for extrapolation beyond the data range may cause biasness in case of estimation of above ground biomass of large trees due to high accuracy in measurement is required, which can be checked only by harvesting the trees and comparing the estimated and observed biomass. Theallometric modeldeveloped by includingdbh and height both as combined variable proved to be the best one, the second best estimate can also be achieved by using only dbh as effect variable, so, in case tree height measurement seems to be difficult, this model can be a better choice by sacrificing little precision in biomass estimation. However care should be taken in case of use of these models in other sites of unknown size structures.

REFERENCES

Brown, S. and A. E. Luge, 1992, Aboveground biomass estimates for tropical moist forests of the Brazilian Amazon, *Jaterciercia*, **17(1)**: 8-18

Fulton, M., 1999, Patterns of height-diameter relationships for selected tree species and sites in eastern Texas,Can. *J. For. Res*, **29**:1445-144

Lappi, J., 1991, Calibration of height and volume equations with random parameters. *For. Sci.* **37**: 781-801.

Lappi, J., 1997, A longitudinal analysis of height/diameter curves,*For. Sci*. **43:** 555-70

Louppe, D., Koua, M. and Coulibaly, A, 1994,Tarifs de cubage pour Afzelia Africana Smith enforet de Badenou (nord Cote d'Ivoire). Tech. rep., Institut des Forets (IDEFOR), departementforesterie, Cote d'Ivoire

Nanos, Nikos., Calamab, Rafael., Monterob, Gregorio and Gil Luis, 2004, Geostatistical prediction of height/diameter models, *For. Ecol.Manage.* **195**: 221-35

Pilli, R., Anfodillo, T. and M. carrer, 2006, Towards a functional and simplified allometry for estimating forest biomass, *For. Ecol. Manage*. **237**: 583-93

Ter-Mikaelian, M.T., and Korzukhin, M. D. 1997, Biomass equations for sixty-five North American tree species, *For. Ecol.Manage*, **97**: 1-24

Parresol, B.R., 1992,Baldcypress height-diameter equations and their prediction confidence intervals, *Canadian J. Forest Res.* **22**:1429-34

Picard N., Saint-André L., Henry M, 2012, Manual for building tree volume and biomass allometric equations: from field measurement to prediction. Food and Agricultural Organization of the United Nations, Rome, and Centre de CoopérationInternationaleenRechercheAgronomique pour le Développement, Montpellier, 215 pp.

Priyanka, B., Chaubey, O.P. and Singhal, P.K., 2013, Biomass Accumulation and Carbon Sequestration in Tectonagrandis Linn. f. and GmelinaarboreaRoxb., International Journal of Bio-Science and Bio-Technology, 5(3), 153-173

Saporta, G., 1990,Probabilités, analyse des donnéesetstatistique. Paris, Technip. 493, 35, 38, 41, 46, 173, 174, 176, 177, 179, 182

Zhang, S., Burkhart, H.E., Amateis, R.L., Zhang, S.A., 1997,The influence of thinning on tree height and diameter relationships in loblolly pine plantations. South. J. Appl. For., 21, 199-205

Fang, Z., Bailey, R.L., Fang, Z.X., 1998, Height-diameter models for tropical forests on Hainan Island in southern China,For. Ecol. Manage., 110, 315-327

Zakrzewski, W.T., Bella, I.E., 1988,Two new height models for volume estimation of lodgepole pine stands,Can. J. For. Res, 18, 195-201