Regular Article
pISSN: 2288-9744, eISSN: 2288-9752
Journal of Forest and Environmental Science
Vol. 32, No. 3, pp. 280-290, August, 2016
http://dx.doi.org/10.7747/JFES.2016.32.3.280

JFES
Journal of Forest and Environmental Science

Modelling Stem Diameter Variability in *Pinus caribaea* (Morelet) Plantations in South West Nigeria

Peter Oluremi Adesoye*
Department of Forest Resources Management, University of Ibadan, Ibadan 200005, Nigeria

Abstract

Stem diameter variability is an essential inventory result that provides useful information in forest management decisions. Little has been done to explore the modelling potentials of standard deviation (SDD) and coefficient of variation (CVD) of diameter at breast height (dbh). This study, therefore, was aimed at developing and testing models for predicting SDD and CVD in stands of *Pinus caribaea* Morelet (pine) in south west Nigeria. Sixty temporary sample plots of size 20 m×20 m, ranging between 15 and 37 years were sampled, covering the entire range of pine in south west Nigeria. The dbh (cm), total and merchantable heights (m), number of stems and age of trees were measured within each plot. Basal area (m^2), site index (m), relative spacing and percentile positions of dbh at 24^{th}, 63^{rd}, 76^{th} and 93^{rd} (i.e. P_{24}, P_{63}, P_{76} and P_{93}) were computed from measured variables for each plot. Linear mixed model (LMM) was used to test the effects of locations (fixed) and plots (random). Six candidate models (3 for SDD and 3 for CVD), using three categories of explanatory variables (i.e. (i) only stand size measures, (ii) distribution measures, and (iii) combination of i and ii). The best model was chosen based on smaller relative standard error (RSE), prediction residual sum of squares (PRESS), corrected Akaike Information Criterion (AIC_c) and larger coefficient of determination (R^2). The results of the LMM indicated that location and plot effects were not significant. The CVD and SDD models having only measures of percentiles (i.e. P_{24} and P_{93}) as predictors produced better predictions than others. However, CVD model produced the overall best predictions, because of the lower RSE and stability in measuring variability across different stand developments. The results demonstrate the potentials of CVD in modelling stem diameter variability in relationship with percentiles variables.

Key Words: Stem diameter variability, standard deviation of diameter, coefficient of variation of diameter, Percentiles, *Pinus caribaea*

Introduction

Diameter variability, which is an index of diameter distribution, is a very important information for forest management planning. It gives an insight to the structure of the forest. According to Fries et al. (1997) and Mc Elhinny et al. (2005), diameter distribution is an important attribute for the management and conservation of biodiversity in forests. This is because larger diameter trees tend to host many species. Several theoretical distributions such as Weibull, beta and Johnson's SB functions have been used to describe stem diameter distribution (e.g. Bailey and Dell 1973; Hafley and Schreuder 1977; Kilkki and Paivinen 1986; Maltamo 1997; Tewari and Gadow 1997). Almost

Received: December 16, 2015. Revised: March 14, 2016. Accepted: March 14, 2016.

Corresponding author: Peter Oluremi Adesoye

Department of Forest Resources Management, University of Ibadan, Ibadan 200005, Nigeria
Tel: 2348063311799, E-mail: adesoyepet@yahoo.com

invariably, the parameters of these functions have been predicted as a function of stand structure variables or solved from a system of equations. While these functions have produced varying degrees of success, there is however room for improvement.

Currently, there are two main approaches for predicting distribution in tree size distribution modelling. These are, parameter prediction method (PPM) and parameter recovery method (PRM) (Hyink and Moser 1983). The PPM involves obtaining parameters of a distribution function (say Weibull) from the fit data via maximum likelihood estimation. They are then used as dependent variables in regression equation with stand variables like relative spacing, number of trees per unit area, site quality, etc. On the other hand, the PRM uses either the diameter moments or specific percentiles (or a hybrid of the two) which are predicted from stand attributes to recover the parameters of the distribution function (e.g. Lohrey and Bailey 1977; Parresol 2003; Knowe et al. 2005; Mehtatalo et al. 2007).

Several studies have been carried out on comparative assessment of these parameter estimation methods (e.g. Baldwin and Feduccia 1987; Cao 2004; Poudel and Cao 2013; Siipilehto and Mehtatalo 2013). The recurring conclusion from these studies is that the PRM seems to be a better approach. However, the focus of this study was not to compare parameter estimation methods for distribution function. The focus was rather, to investigate whether standard deviation of diameter (SDD) and coefficient of variation of diameter (CVD), which are chief characteristics of diameter variability are indeed independent from stand attributes like relative spacing, stand density, stand diameter, age, site quality, percentiles etc. Therefore, the main thrust of this study is to explore the relationship between SDD/CVD and stand size variables.

To date, little has been done to explore the modelling potentials of the two main characteristics of stem diameter distribution (i.e. SDD and CVD). Zeide and Zhang (2000), proposed a model for estimating SDD using stand variables such as average diameter, number of trees and age. Their proposed model, which focused mainly on SDD, explained 91% of the variation in SDD. However, standard deviation is often criticized to be difficult to interpret in magnitude. This is because standard deviation of smaller observations (say stem diameter of younger trees) tends to be small, while that of larger observations (e.g. diameter of larger trees) tends to be large. Furthermore, a lower standard deviation does not necessarily imply lesser variability. In this study, the coefficient of variation is proposed as a suitable alternative, because the standard deviation of observations must always be understood in the context of the mean of observations. The advantage of coefficient of variation is that it is unit-less. For comparison among different means, which is a recurring phenomenon when dealing with forest inventory data, coefficient of variation may be a better alternative. The purpose of this study is to investigate the modelling potentials of these two characteristics of diameter variability with stand size variables using *Pinus caribaea* stands datasets in south west Nigeria.

The study hypothesis is that measures of central tendency and distribution (which can be computed from forest inventory data) are sufficient to predict stem diameter variability in terms of SDD and CVD. The variability may differ from location to location because of differences in variables such as stand diameter, stand density, age, elevation and site quality. This study, therefore, propose a single model applicable to various stand conditions.

Materials and Methods

The Data

In this study, the data used for model fitting were collected in 2011 on 60 temporary sample plots ranging between 15 and 37 years (Oyebade 2014). The selected locations cover the entire range of *Pinus caribaea* in south west Nigeria. The dataset came from three locations, viz., Omo, Oluwa and Shasha Forest Reserves. The three locations covered three states in the south west Nigeria notable for industrial forest plantations (i.e. Ogun, Ondo and Osun states). The Reserves are contiguous and have a mean annual rainfall of about 2,050 mm and mean monthly temperature of about $27^{o}C$ (Adedeji and Adeofun 2014). The natural vegetation, which was previously lowland tropical rainforest (moist evergreen type), have been reduced to secondary forests. Substantial parts of the Reserves were converted to plantations with *Gmelina arborea* being the dominant species planted. The pine stands in the three locations had no record of thinning and pruning. The Reserves are made up of several soil types but all belong to the tertiary sediments (Ola-Adams 1999). The terrain is generally undulating with maximum elevation of 150 m above sea level.

Sixty temporary sample plots (TSPs) of size 20 m×20 m each and of different stand ages based on stocking density were sampled from Omo Forest Reserve (16 TSPs from stand ages 15 and 21 years); Oluwa Forest Reserve (36 TSPs from stand ages 18, 20, 35, 36 and 37 years) and Shasha Forest Reserve (8 TSPs from stand age 27 years) in Ogun, Ondo and Osun States respectively using stratified random sampling method. The tree size variables measured within each plot include, diameter at breast height (cm), total and merchantable heights in meters (the merchantable limit is taken as height to the minimum top diameter of 10 cm), number of stems and age.

Model Formulation

Diameter variability is a function of many processes that constitute stand dynamics. Attempt was made to deduce the existence of variability from available variables. The model formulation procedures include selecting response and explanatory variables, formulating appropriate models and evaluating the final models.

Response variables

In this study, two measures variability: standard deviation in cm (SDD) and coefficient of variation (CVD) of the most easily and accurately measured stand variable – the tree diameter at breast height (in cm) were used as the predictor (or response variable).

Explanatory variables

As explanatory variables, all available variables were considered. These are, location, arithmetic mean diameter (D_m in cm), quadratic mean diameter (D_q in cm), number of trees/ha (N), basal area/ha (BA in m^2), stand age (years), mean dominant height (H_d in m), site index (SI in m) and relative spacing (RS). In addition, the percentile positions at 24th (P_{24}), 63rd (P_{63}), 76th (P_{76}) and 93rd (P_{93}) of plot diameter distribution were also considered. A rectangular correlation matrix of SDD and CVD against the explanatory variables that are quantitative and continuous was carried out to give insights into the relationships between candidate response variables (i.e. SDD and CVD); and each explanatory variable. The relative spacing (RS) of the pine stands was computed using the formula:

$$RS = (10000/N)^{1/2}/H_d \quad \cdots\cdots (1)$$

Where, N is the number of trees per hectare and H_d is the average dominant total height (m).

For the purpose of site index estimation, four largest trees of pine were selected per plot (4 being representative of 100 largest trees per hectare in a 20 m×20 m plot). The four largest trees are dominant or co-dominant; have straight stems; are unsuppressed; showing no signs of disease or insect attack; and are healthy and vigorous. The average height of the four largest trees in each plot was referenced as the dominant height. The plantation age, according to the year of establishment of the species was used as stand age. The site index for each plot was obtained by using the site index equation developed by Oyebade (2014). The equation is presented as follows:

$$SI = e^{[\ln H_d - 23.495(A^{-1} - 0.04)]} \quad \cdots\cdots (2)$$

Where, SI denotes the site index (m), H_d is the average dominant total height (m), A is the stand age. An index age of 25 years was used.

Relationship between response and explanatory variables

The identification of which of the available explanatory variables and their interactions are significant is considered the first step in developing appropriate model for diameter variability. In this study, plots were randomly selected from each location. This study therefore, uses location as fixed factor and plots as random factor. As a result, diameter variability (DVAR), in terms of SDD and CVD can be expressed by the following linear model:

$$DVAR = X\beta + Z\gamma + \varepsilon \quad \cdots\cdots (3)$$

Where, DVAR denotes an N×1 vector of SDD or CVD observations (N represents the number of observations), β is a p×1 vector of unknown fixed effects parameters (p being the number of fixed-effects parameters), X is an N×P non-stochastic matrix of rank p of observations on the explanatory variables, γ is a k×1 vector of unknown random-effects parameters (k being the number of random-effects parameters), Z is an N×k known design matrix, containing either continuous or dummy variables, and ε is N×1 vector of random unobservable errors.

This is a mixed model because both fixed – effects parameters (β) and random – effects parameters (γ) are included. Both γ and ε are assumed to be normally distributed with expectation, E and variance Var:

$$E\begin{bmatrix}\gamma\\ \varepsilon\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}$$

$$Var\begin{bmatrix}\gamma\\ \varepsilon\end{bmatrix}=\begin{bmatrix}G & 0\\ 0 & R\end{bmatrix} \quad \cdots\cdots (4)$$

Where, G is the variance of γ that is equal to $I\sigma_\gamma^2$ (I is an N×N identity matrix and σ_γ^2 is the variance due to random-effects), and R equals $I\sigma_e^2$ (σ_e^2 is the residual variance).

The normality assumption is satisfied judging from the number of observations that is sufficiently large according to the central limit theorem. Locations (fixed) and plots (random) are class variables. The other set of explanatory variables and their interactions were tested. The test revealed that the effect of location is not significant. The covariance parameter of the plot variance was also found to be redundant. Analysis of variance (Eq.2) suggested the following predictors of DVAR (i.e. SDD and CVD): D_m, P_{63}, and P_{93}. As a further step to confirm the findings in linear mixed model, stepwise regression methods (using backward elimination and stepwise procedures) were used to identify suitable explanatory variables. The three explanatory variables were still found suitable. However, when their variance inflation factors (VIFs) were checked, there was the problem of multi-collinearity as the VIFs for D_m and P_{63} were found to be 31.1 and 29.3 respectively. These values exceed 10 and showed apparent multi-collinearity. The promising alternative was using P_{24} and P_{93} as explanatory variables.

Therefore, in terms of the explanatory variables considered, three categories of models were investigated. The three categories are:

(i) Models with explanatory variables that are mainly measures of average tree size, stand age, stand density and stand productivity (i.e. A, N/ha, D_m, D_q, BA/ha, RS and SI)

(ii) Models with explanatory variables that are mainly measures of distribution (i.e. P_{24}, P_{63}, P_{76} and P_{93})

(iii) Models with explanatory variables that are based on the combinations of (i) and (ii)

Multiple linear model was proposed and fitted to the data. The proposed models are generally of the form:

$$SDD=b_0+b_1X_1+b_2X_2+\cdots+b_nX_n \quad \cdots\cdots (5)$$

$$CVD=b_0+b_1X_1+b_2X_2+\cdots+b_nX_n \quad \cdots\cdots (6)$$

Where, X_1, X_2 ⋯ X_n are the explanatory variables (i.e. D_q, D_m, N, Age, P_{24}, P_{63}, P_{76} and P_{93}, Hd, SI, and RS). All the candidate explanatory variables, under each model categories were considered using backward elimination and stepwise approach. Furthermore, the suitable variables were checked for multicollinerity by observing their variance inflation factors (VIFs). For the purpose of addressing the issue of thinning and pruning in the future, it was essential to obtain future values for suitable explanatory variables. Hence prediction equations were also formulated for the explanatory variables of the best model for the purpose of obtaining the future values of the explanatory variables.

Goodness-of-fit Criteria

To evaluate models performance and to compare them, standard error of estimate (SEE), relative standard error (RSE), coefficient of determination (R^2), prediction residual sum of squares (PRESS) and Akaike's Information Criterion (AIC) were employed. Model comparison was essentially done under each category of response variable. Then, RSE was used to select the overall best model. The evaluation statistics are defined as follows:

$$SEE=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-p}} \quad \text{or} \quad \sqrt{\frac{SSE}{n-p}} \quad \cdots\cdots (7)$$

$$RSE\ (\%)=\frac{SEE}{Estimate}\times 100 \quad \cdots\cdots (8)$$

$$R^2=1-\left[\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\right] \quad \text{or} \quad 1-\frac{SSE}{SST} \quad \cdots\cdots (9)$$

$$PRESS = \sum_{i=1}^{n} (y_i - \hat{y} *_{ip})^2 \qquad (10)$$

$$AIC = n * ln(\frac{SSE}{n}) + 2k \qquad (11)$$

Where, y_i is the observed value for the i^{th} observation, $\hat{y_i}$ is predicted value for the i^{th} observation, $\bar{y}$ is mean of y observations, $\hat{y} *_{ip}$ is the predicted value of y for observation i as calculated from a regression equation derived through fitting the p parameter model to a data set obtained by deleting observation i from the original data set, n is the total number of observations used to fit the model, SSE is the residual or error sum of squares and k is the number of parameters in the model including the error term. Model with smaller values of PRESS, AIC, SEE, RSE; and higher value of R^2 was considered to have better fit.

When n is small compared to k (usually, the rule of the thumb is that if n/k <40, then n is small), it has been shown that AIC is too small and bias (e.g. Burnham and Anderson 2002; Royall 1997). The corrected AIC value (AIC_c) is more accurate and can be computed as follows:

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1} \qquad (12)$$

Results

Data Summary

The summary statistics of the datasets for model fitting are presented in Table 1. The data covered a relatively wide age range (i.e. 15-37 years). The two measures of variability (i.e. standard deviation－SD; and coefficient of variation－CV) differ in their summary of stand variables. The standard deviations of the stand variables are quite unreliable in the summary statistics (Table 1). The smallest SD was obtained for relative spacing (i.e. RS of 0.07), while the largest SD was obtained for number of stems per hectare (i.e. N of 403.49). It can be observed that stand variables having larger means tend to have larger SD values and those with smaller means tend to have smaller SD. Conversely, an examination of the coefficient of variation reveals a more reliable measure of variability. The smallest CV was obtained for mean dominant height (i.e. CV of 0.166), while the largest CV was obtained for number of stems per hectare. Larger means was not necessarily attributed to larger CV. Furthermore, the two measures of variability differ in their summary of percentile positions. Standard deviation increases with increasing percentile position, whereas, coefficient of variation decreases with increasing percentile

Table 1. Summary statistics of plot data used for model development

Variable	Minimum	Maximum	Mean	SD	CV
A (years)	15	37	26.72	8.28	0.310
D_m (cm)	15.74	46.52	25.78	7.46	0.289
N/ha	125	1,750	663.33	403.49	0.608
SDD (cm)	3.75	13.70	7.50	2.42	0.323
CVD	0.14	0.42	0.29	0.06	0.207
P_{24} (cm)	12.29	46.18	20.38	7.13	0.350
P_{63} (cm)	16.79	52.42	27.87	8.32	0.299
P_{76} (cm)	18.43	54.21	30.56	8.86	0.290
P_{93} (cm)	20.5	54.62	36.37	9.20	0.253
BA/ha (m^2)	9.00	70.25	31.07	10.94	0.352
D_q (cm)	17.45	48.00	27.36	7.97	0.291
H_d (cm)	10.80	20.60	16.61	2.76	0.166
SI (cm)	5.77	26.95	17.47	7.15	0.409
RS	0.17	0.49	0.27	0.07	0.245

Note: A is the plot age, D_m is the arithmetic mean diameter at breast height in each plot, N/ha is the number of stems per hectare, P_{24}, P_{63}, P_{76} and P_{93} are the 24^{th}, 63^{rd}, 76^{th} and 93^{rd} percentiles distribution of the stem diameter at breast height in each plot, BA/ha is the basal area per hectare, D_q is the quadratic mean diameter at breast height in each plot, H_d is the average dominant height in each plot, SI is the site index, RS is the relative spacing, SDD and CVD are as earlier defined.

position. It can be deduced that variability is indeed higher in lower percentile positions than higher percentile positions.

The result of the rectangular correlation analysis of SDD and CVD against stand variables is presented in Table 2. The SDD was positively related to all the stand variables except number of stems per hectare with negative correlation. Relative spacing and basal area per hectare were not correlated with SDD. The highest correlation with SDD was found for P_{93}. On the other hand CVD was negatively correlated with P_{24} and RS. CVD was not correlated with the other stand variables. In general, the stand variables considered in this study may be grouped into four broad categories. These are:

(i) average tree size or dimension (i.e. D_m, D_q and H_d)
(ii) stand density or stand productivity (i.e. N/ha, BA/ha, SI and RS)
(iii) tree size distribution (i.e. P_{24}, P_{63}, P_{76} and P_{93}) and
(iv) stand age (i.e. A)

Consistent positive values of correlation coefficients between SDD and each average tree size, stand age and tree size distribution variables suggested that SDD increased with increasing tree size, age and tree size distribution variables. Negative value of correlation coefficient between SDD and N/ha illustrated that SDD decreased with increasing stand density. The relationship between CVD and the various categories of stand variables can be depicted using similar approach.

Models for Predicting SDD and CVD

The specific candidate models fitted to the data with their corresponding parameter estimates, fit and prediction statistics are presented in Table 3. The candidate models under CVD and SDD were based on three categories of explanatory variables as earlier pointed out. All the parameter estimates of the candidate models are statistically significant at 5% probability.

On the basis of estimated R^2 values, 40 to 77% of the total variation in observed CVD values was explained by the three candidate CVD models. The fit statistics for the candidate CVD models showed that model 2 with explanatory variables that are mainly based on measures of distribution was the best. The prediction statistics (i.e. PRESS and RSE) obtained for the candidate CVD models provide some indication of the prediction accuracy of the three candidate CVD models. The PRESS and RSE values of model 2 were consistently smaller. Also, the AIC_c and SEE for model 2 were the smallest among the CVD candidate models. All the fit and prediction statistics were consistent in their judgements.

Similarly, on the basis of estimated R^2 values, 66 to 90%

Table 2. Rectangular matrix of correlation coefficients of SDD and CVD against stand variables

	D_m	N/ha	A	BA/ha	D_q	P_{24}	P_{63}	P_{76}	P_{93}	H_d	SI	RS
SDD	0.67*	−0.50*	0.80*	0.10	0.58*	0.44*	0.72*	0.75*	0.86*	0.63*	0.79*	0.24
CVD	−0.17	0.11	0.21	0.23	−0.16	−0.40*	−0.09	−0.05	0.15	0.21	0.23	−0.45*

Marked correlations are significant at $p<0.05$.

Table 3. Candidate models with parameter estimates and fit statistics for CVD and SDD

Model No	Model	R^2	AICc	SEE	RSE%	PRESS
	Candidate CVD Models					
1.	$CVD=0.292-0.0074D_m+0.0016BA+0.0083SI$	0.43	−357.06	0.0481	16.36	0.1497
2.	$CVD=0.230-0.0126P_{24}+0.0088P_{93}$	0.77	−414.67	0.030	10.20	0.0695
3.	$CVD=0.348+0.0035P_{93}-0.682RS$	0.40	−356.50	0.0489	16.63	0.1538
	Candidate SDD Models					
4.	$SDD=1.50+0.0395BA+0.274SI$	0.66	126.08	1.4340	19.11	128.6910
5.	$SDD=-1.716-0.226P_{24}+0.366P_{93}$	0.90	−22.77	0.7891	10.52	51.5931
6.	$SDD=1.67+0.0888P_{76}+0.178SI$	0.66	48.16	1.4244	18.99	130.7000

of the total variation in observed SDD values was explained by the three candidate SDD models. The fit statistics for the candidate SDD models again, showed that model *5* with explanatory variables that are mainly based on measures of distribution was the best. All the fit and prediction statistics were consistent in their judgements of the best model. The R^2 value was highest for model *5*. The AIC_c, SEE, PRESS and RSE values were the smallest for model *5*.

Although, it is not appropriate to compare two models with different response variables, using the coefficient of determination and standard error of estimate values, the use of relative standard error (RSE) can, however, be very useful for such comparison. Hence, comparison of the two best models (2 and *5*) on the basis of relative standard error indicate that model 2 is slightly superior to model *5*. Fig. 1 shows the residual distributions of models 2 and *5* across their respective range of fitted values. It can be seen that model 2 showed a more constant error variance compared to model *5* that showed increasing error variance across the range of fitted values. This implies that model *5* violates the constant error assumption. Fig. 2 shows the relationships between the fitted values and the observed values of these two models. Figs. 3 and 4 show the predicted surface of CVD and SDD respectively, for a range of observed P_{24} and P_{93}. The CVD increased with increasing P_{93}, but decreased with increasing P_{24} (Fig. 3). Similar trend was observed for SDD versus P_{24} and P_{93} (Fig. 4). On the basis of the above analysis, the CVD model 2, with explanatory variables that are mainly based on measures of distribution, displays sufficiently high predictive power to constitute a final model for predicting stem diameter variability of *Pinus caribaea* trees in the study area.

Prediction Equation for P_{24} and P_{93}

To obtain the future values of P_{24} and P_{93}, prediction equation was obtained for the two explanatory variables. The final equations obtained for these variables are presented as follows:

$$lnP_{24} = 2.0051 + 0.0438D_m - 0.0063A \quad \cdots\cdots\cdots\cdots\cdots (13)$$

$R^2 = 0.93$, $SEE = 0.0772$, $RSE = 2.60\%$, $PRESS = 0.3749$

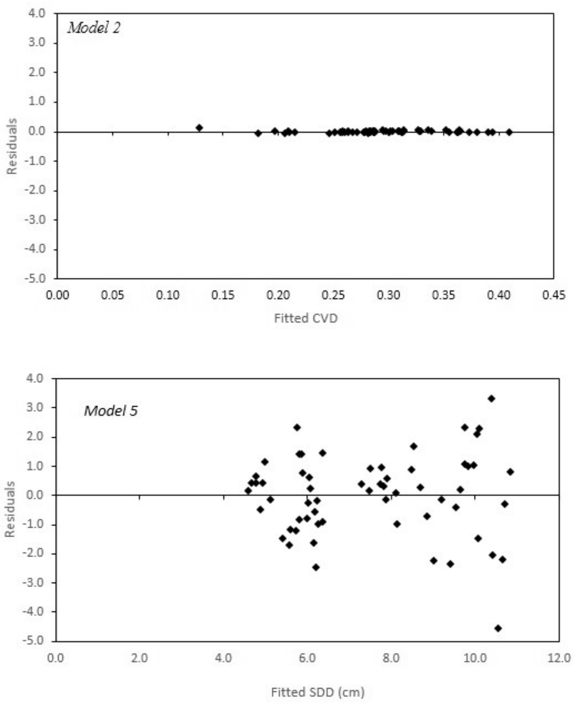

Fig. 1. Distributions of residuals of CVD with model 2 and SDD with model *5*.

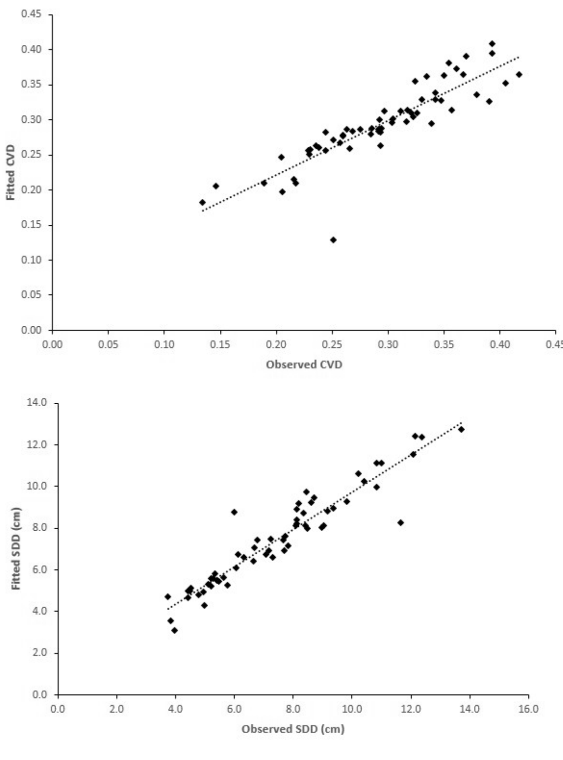

Fig. 2. Relationships between the fitted and the observed values of CVD and SDD of models 2 and *5*.

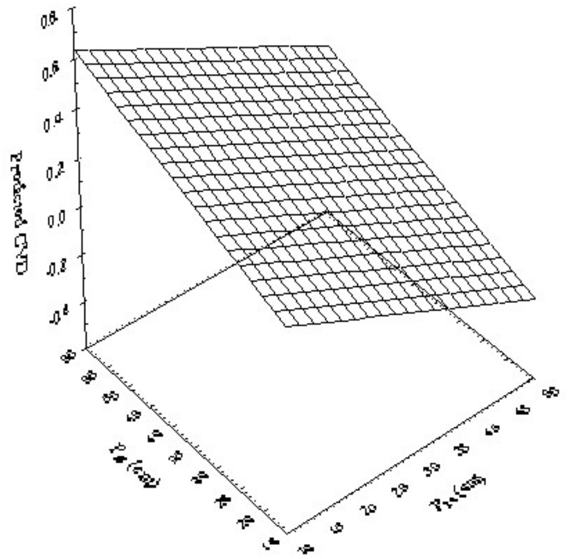

Fig. 3. Surface plot of predicted CVD versus P_{24} and P_{93} for *Pinus caribaea* using model 2 (n=60).

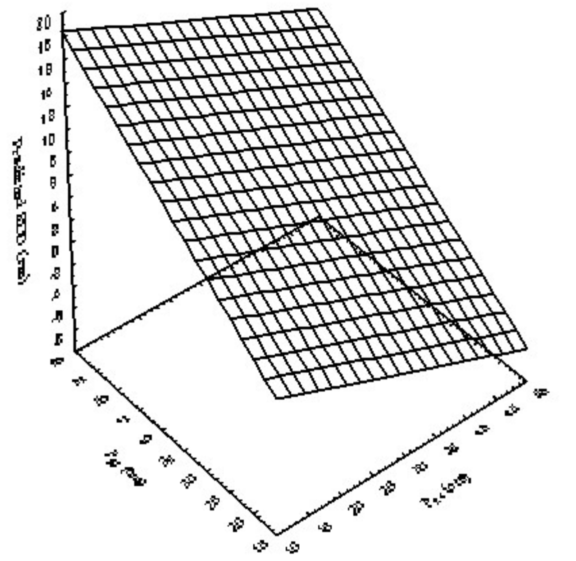

Fig. 4. Surface plot of predicted SDD versus P_{24} and P_{93} for *Pinus caribaea* using model 5 (n=60).

$$lnP_{93} = 2.7256 + 0.0191D_m + 0.0129A \quad \cdots\cdots\cdots\cdots\cdots (14)$$
$$R^2=0.89,\ SEE=0.0845,\ RSE=2.41\%,\ PRESS=0.4500$$

It is noteworthy that P_{24} is negatively related to age, and positively related to arithmetic average diameter at breast height. The implication is that P_{24} increases with increase in average diameter, but decrease with increase in age. Whereas, P_{93} is positively related to age and arithmetic average of diameter at breast height. The two prediction equations are found suitable for the purpose of obtaining the future values of P_{24} and P_{93}.

Discussion

In this study, models predicting diameter variability in terms of coefficient of variation and standard deviation of diameter at breast height were estimated using backward elimination and stepwise regression methods. Stand variables considered as possible explanatory variables were categorized into: (i) measures of tree size, stand density, age or stand productivity, (ii) measures of stem size distribution, and (iii) combination of (i) and (ii). Initially, fixed effects of location and random effect of plots were tested using linear mixed model. However, the test revealed that location was not statistically significant and the random effect of plots was redundant. This is probably, because the three locations representing the entire native range of pine in the south-western Nigeria are contiguous. The pine stands in these locations had no record of thinning and pruning.

Quite many studies have provided information on the capacity of stand variables to predict percentile based diameter distribution (e.g. Baldwin and Feduccia 1987; Borders et al. 1987; Maltamo et al. 2000; Cao 2004; Poudel and Cao 2013). However, few studies, to date, have pointed out the relationship between measures of variability (i.e. SDD and CVD) and stand variables (e.g. Zeide and Zhang 2000). The apparent significant correlations between standard deviation of diameter (SDD) and most of the stand variables obtained in this study is similar to the findings of Zeide and Zhang (2000). It was observed that the coefficient of variation of diameter was not significantly correlated with most of the stand variables. This could probably be due to the fact that coefficient of variation of diameter is a ratio. Cohen et al. (2003) indicated that the Pearson product-moment correlation coefficient may be spurious and misleading when used to measure relationship between a ratio and another ratio or variable. Hence, serious consideration was not given to the results of the correlation during model development.

The standard deviation of diameter was positively correlated with the 24th percentile (P_{24}) of diameter distribution. However, when the same variable (i.e. P_{24}) was used as explanatory variable together with P_{93} in predicting SDD, it was negatively related. This indicate that SDD could be very unstable and unreliable under different stand developments. This trend was not found for CVD in this study. The trend observed in the correlation (at least in terms of direction of relationship) was consistent with that observed in the regression. McDonald (2014) has pointed out that coefficient of variation may be a more useful measure of variability than standard deviation.

Among the six candidate models tested, models with measures of distribution as the explanatory variables ranked best. This confirms the findings of Koesmarno et al. (1995) and Koesmarno (1996) who found percentile intervals technique to be flexible and efficient in modelling diameter variability. The best two models showed similar trend in relationship with the explanatory variables (i.e. P_{24} and P_{93}). Both CVD and SDD were negatively related to P_{24} and positively related to P_{93}. This finding implies that CVD and SDD increase with P_{93}, but decrease with increase in P_{24}. Each of these variables is statistically significant. At the same time, this finding raise a question: why are CVD and SDD negatively related to P_{24}, but positively related to P_{93}? A possible reason for this could be as a result of the trend in the variability of the underlying data (i.e. Table 1). A closer observation of the data summary, especially, the measures of distribution (P_{24}, P_{63}, P_{76} and P_{93}), revealed that real dispersion (i.e. checking the values of CV) was highest for P_{24} and least for P_{93} across the range of stand conditions in the study area. Again, naturally, P_{24} values will be higher in older stands than younger stands, and relative variability will be smaller in older stands. Hence, there is a high tendency for negative relationship between P_{24} and CVD/SDD.

The CVD model having measures of distribution (i.e. P_{24} and P_{93}) was found to be the best in terms of goodness-of-fit criteria and prediction accuracy. The reason for this was also based on the fact that CVD as a measure of variability is relativized over a range of stand conditions and is unaffected by the form or size of inventory data, compared to SDD. The use of relative standard error to compare SDD and CVD models provides the basis for such comparison. It is also important to note that this study was conducted in unthinned and unpruned stands of *Pinus caribaea*. Essentially, during thinning, a stand is affected in terms of number of stems, average diameter at breast height and average dominant height. Prediction equations obtained for P_{24} and P_{93} in this study were found suitable for the purpose of obtaining the future values of the explanatory variables. This confirms the findings in previous studies that a good relationship exists between percentile positions of diameter at breast height and stand variables (e.g. Eerikainen and Maltamo 2003; Mehtatalo et al. 2008; Yatich 2009; Lumbres and Lee 2014).

Conclusion

The modelling potentials of two chief characteristics of stem diameter variability, standard deviation of diameter (SDD) and coefficient of variation of diameter (CVD) were investigated using *Pinus caribaea* stands in south west Nigeria. Six candidate models (3 for CVD and 3 for SDD) were investigated. This study shows that models with only measures of stem diameter distribution (i.e. P_{24} and P_{93}) as explanatory variables ranked the best. The CVD model 2 was found to be the overall best for the prediction of stem diameter variability of *Pinus caribaea* stands in the south west Nigeria. This is because of the lower relative standard error (RSE) and the stability of the CVD in measuring variability across the different stand developments unlike the standard deviation of diameter. The CVD increases with P_{93} and it decreases with increase in P_{24}. This trend was also observed with SDD, although, the correlation results showed opposing trend. This indicated that SDD could be unstable and unreliable. Each of the explanatory variables is statistically significant and, to some extent, ecologically meaningful. Furthermore, prediction equations were obtained for P_{24} and P_{93}, for the purpose of addressing thinning and pruning scenarios. Again, age was negatively related to P_{24}, but positively related to P_{93}. These results raise a question: why does P_{24} decrease with increase in age, SDD and CVD? It is expected that stand level growth models based on stem diameter distribution can be improved by using the CVD model 2.

Acknowledgements

Author is grateful to Dr. Bukola Oyebade of the

Department of Forestry and Wildlife, University of Port Harcourt, Nigeria for kindly sharing his valuable data set for this study.

References

Adedeji OH, Adeofun CO. 2014. Spatial Pattern of Land Cover Change using Remotely Sensed Imagery and GIS: A Case Study of Omo-Shasha-Oluwa Forest Reserve, SW Nigeria (1986-2002). Journal of Geographic Information System 6: 375-385.

Bailey RL, Dell TR. 1973. Quantifying Diameter Distributions with Weibull Function. Forest Science 19: 97-104.

Baldwin VC Jr, Feduccia DP. 1987. Loblolly Pine Growth and Yield Prediction for Managed West Gulf Plantations. USDA Forest Service, New Orleans, Research Paper SO-236, 27 pp.

Borders BE, Souter RA, Bailey RL, Ware KD. 1987. Percentile-Based Distributions Characterize Forest Stand Tables. Forest Science 33: 570-576.

Burnham KK, Anderson DR. 2002. Model selection and multi-model inference: a practical information-theoretic approach. 2nd Edition. Springer, New York, 515 pp.

Cao QV. 2004. Predicting Parameters of a Weibull Function for Modelling Diameter Distribution. Forest Science 50: 682-685.

Cohen J, Cohen P, West SG, Aiken LS. 2003. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Third Edition. New York, London, Routledge, 726 pp.

Eerikäinen K, Maltamo M. 2003. A percentile based basal area diameter distribution model for predicting the stand development of *Pinus kesiya* plantations in Zambia and Zimbabwe. Forest Ecology and Management 172: 109-124.

Fries C, Johansson O, Pettersson B, Simonsson P. 1997. Silvicultural Models to Maintain and Restore Natural Stand Structures in Swedish Boreal Forest. Forest Ecology and Management 94: 89-103.

Hafley WL, Schreuder HT. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. Canadian Journal of Forest Research 7: 481-487.

Hyink DM, Moser JW. 1983. A Generalized Framework for Projecting Forest Yield and Stand Structure using Diameter Distributions. Forest Science 29: 85-95.

Kilkki P, Paivinen R. 1986. Weibull Function in the Estimation of Basal Area Dbh Distribution. Silva Fennica 20: 149-156.

Knowe SA, Radosevich SR, Shula RG. 2005. Basal Area and Diameter Distribution Prediction Equations for Young Douglas-Fir Plantations with Hardwood Competition: Coast Ranges. Western Journal of Applied Forestry 20: 77-93.

Koesmarno HK, Whyte AGD, Mason EG. 1995. Dynamics of Size-Class Distributions of Young Radiata Pine in Response to Vegetation Management Treatments. In: Recent Advances in Forest Mensuration and Growth and Yield Research (Skovsgaard JP, Burkhart HE, eds.) Ministry of Environment and Energy Danish Forest and Landscape Research Institute, Denmark, pp 45-61.

Koesmarno HK. 1996. Class-Size Percentile Transformation for Reconstructing a Distribution Function. Journal of Applied Statistics 23: 423-434.

Lohrey RE, Bailey RL. 1977. Yield Tables and Stand Structure for Unthinned Long Leaf Pine Plantations in Louisiana and Texas. USDA Forest Service, Research Paper SO-133, 53 pp.

Lumbres RIC, Lee YJ. 2014. Percentile-Based Weibull Diameter Distribution Model for *Pinus kesiya* Stands in Benguet Province, Philippines. Southern Forests: A Journal of Forest Science 76: 117-123.

Maltamo M, Kangas A, Uuttera J, Torniainen T, Saramäki J. 2000. Comparison of Percentile Based Prediction Method and the Weibull Distribution in Describing the Diameter Distribution of Heterogeneous Scots Pine Stands. Forest Ecology and Management 133: 263-274.

Maltamo M. 1997. Comparing Basal Area Diameter Distributions Estimated by Tree Species and for the Entire Growing Stock in a Mixed Stand. Silva Fennica 31: 53-65.

Mc Elhinny C, Gibbons P, Brack C, Bauhus J. 2005. Forest and Woodland Stand Structural Complexity: Its Definition and Measurement. Forest Ecology and Management 218: 1-24.

McDonald JH. 2014. Handbook of Biological Statistics. Third Edition. Sparky House Publishing, Baltimore, Maryland, USA. 305 pp.

Mehtätalo L, Gregoire TG, Burkhart HE. 2008. Comparing Strategies for Modelling Tree Diameter Percentiles from Remeasured Plots. Environmetrics 19: 529-548.

Mehtatalo L., Maltamo M., Packalen P. 2007. Recovering Plot-Specific Diameter Distribution and Height-Diameter Curve using ALS Based Stand Characteristics. ISPRS Workshop on Laser Scanning and SilviLaser ESPOO; 2007 Sep 12-14, pp 288-293.

Ola-Adams BA. 1999. Biodiversity Inventory of Omo Biosphere Reserve, Nigeria. Country Report on Biosphere Reserves for Biodiversity Conservation and Sustainable Development in Anglophone Africa (BRAAF) Project. 315p.

Oyebade BA. 2014. Models for Growth Characteristics and their Applications in Yield Studies for *Pins caribaea* Morelet in 1851 in Southwestern Nigeria. Ph.D. Dissertation in the Department of Forest Resources Management, University of Ibadan, Ibadan, Nigeria.

Parresol BR. 2003. Recovering Parameters of Johnson's SB Distribution. Department of Agriculture, Forest Service, Southern Research Station. Asheville, NC: U.S, Research Station Research Paper SRS-31, 9 pp.

Poudel KP, Cao QV. 2013. Evaluation of Methods to Predict Weibull Parameters for Characterizing Diameter Distributions. Forest Science 59: 243-252.

Royall R. 1997. Statistical Evidence: A Likelihood Paradigm.

Chapman and Hall, London.
Siipilehto J, Mehtätalo L. 2013. Parameter Recovery vs. Parameter Prediction for the Weibull Distribution Validated for Scots Pine Stands in Finland. *Silva Fennica* doi: 10.14214/sf.1057
Tewari VP, Gadow K. 1997. Fitting a Bivariate Distribution to Diameter-Height Data of Forest Trees. Indian Forester 123: 815-820.
Yatich SK. 2009. Diameter Distribution Prediction Models for Thinned Slash and Loblolly Pine Plantations in the Southeast. Electronic Thesis & Dissertation Collection. Graduate Faculty of the University of Georgia. https://getd.libs.uga.edu/pdfs/yatich_sammy_k_2009 12_phd.pdf [Cited Feb, 2016]. 177 pp.
Zeide B, Zhang Y. 2000. Diameter Variability in Loblolly Pine Plantations. Forest Ecology and Management 128: 139-143.