STATISTICAL DISTRIBUTIONS FOR MODELLING STAND STRUCTURE OF NEEM (*AZADIRACHTA INDICA*) PLANTATIONS

D. M. Nanang

Department of Rural Economy, University of Alberta, Edmonton, Alberta, Canada T6G 2H.
E-mail: dmnanang@yahoo.ca

Received December 1999

NANANG, D. M. 2002. Statistical distributions for modelling stand structure of neem (*Azadirachta indica*) plantations. Individual and community plantations of neem (*Azadirachta indica*) and other exotic tree species were expanded in northern Ghana following the introduction of a Rural Afforestation Programme in 1989. This paper describes a study aimed at assessing the suitability of the univariate (normal, lognormal, Johnson S_B and gamma) and bivariate (normal (S_{NN}), lognormal (S_{LL}) and Johnson S_B (S_{BB})) distributions for modelling diameter and height distributions of the neem plantations. The four univariate and three bivariate distributions were fitted to seven age groups of diameter and height data collected from 120 temporary sample plots. In general, all four univariate distributions provided good fits to both the diameter and height data. However, based on the ranking of the Kolmogorov-Smirnov (KS) statistics between observed and predicted frequencies for each age group, the gamma and lognormal were judged the best for fitting the diameter and height data respectively. For the three bivariate distributions, the S_{LL} gave the best performance in terms of quality of fit to the seven age groups using the KS criterion. Height distributions were indirectly derived from each of the four univariate diameter distributions based on the relationship between diameter and height. However, these indirectly derived height distributions did not satisfactorily describe the observed height frequencies.

Key words: Bivariate distributions - Kolmogorov-Smirnov statistic - northern Ghana - univariate distributions

NANANG, D. M. 2002. Taburan statistik untuk model struktur dirian di ladang neem (*Azadirachta indica*). Ladang neem milik komuniti dan perseorangan serta spesies pokok dagang diperluaskan di Ghana Utara berikutan pengenalan Program Penghutanan Luar Bandar pada tahun 1989. Artikel ini menerangkan tentang kajian untuk menilai kesesuaian taburan satu-pengubah (normal, log normal, Johnson SB serta gamma) dan taburan dwi-pengubah (normal (S_{NN}), log normal (S_{LL}) dan Johnson SB (S_{BB})) bagi model taburan diameter dan ketinggian di ladang neem. Taburan empat satu-pengubah dan tiga dwi-pengubah dipadankan dengan tujuh kumpulan umur bagi data diameter dan ketinggian yang diambil daripada 120 petak sampel sementara. Secara amnya keempat-empat taburan satu-pengubah memberikan padanan yang baik bagi kedua-dua data diameter dan ketinggian. Bagaimanapun, berdasarkan ranking statistik Kolmogorov-Smirnov (KS) antara kekerapan yang dicerap dengan kekerapan yang dijangka bagi setiap kumpulan umur, didapati taburan gamma dan log normal merupakan padanan terbaik masing-masing bagi data diameter dan ketinggian. Bagi ketiga-tiga taburan dwi-pengubah, taburan S_{LL} menunjukkan prestasi terbaik dari segi kualiti padanan terhadap tujuh kumpulan umur menggunakan kriteria KS. Taburan ketinggian diperoleh secara tidak langsung daripada keempat-empat taburan diameter satu-pengubah berdasarkan hubungan antara diameter dan ketinggian. Bagaimanapun taburan ketinggian yang diperoleh secara tidak langsung ini tidak menerangkan dengan jelas tentang kekerapan ketinggian yang dicerap.

Introduction

A Rural Forestry Division was established within the Ghana Forestry Department in 1989 to encourage the establishment of plantations in order to mitigate the effects of rising consumption and diminishing supply of fuelwood and other forest products. To implement the Rural Afforestation Program (RAP), the new Rural Forestry Division was given a mandate to establish and expand existing tree nurseries, initiate and expand community and individual plantations, and to provide technical advice to farmers on establishment, management and protection of the trees. In addition, the new Division was to provide extension services and education on rural forestry and agroforestry.

Before the advent of the RAP, the Forestry Department, working in conjunction with some communities and individuals, had established plantations of exotic tree species. The most widely planted species before and during the implementation of the RAP was neem (*Azadirachta indica*), which has acclimatised well throughout northern Ghana. In addition to its use as fuelwood, neem has many desirable qualities (e.g. toughness and durability) which make it more suitable for use as poles and rafters for building construction by local communities than other species. Consequently, the ability to describe the stand structure of these plantations by suitable theoretical distributions will not only be important for estimating the type of products that can be obtained at various ages of stand development, but also useful for their effective management. Nanang (1998) developed diameter distribution models from temporary sample plot data for these plantations using a subset of the current data set. However, information on height distributions, diameter/height relationships, the possibility of satisfactorily deriving height distributions indirectly from diameter distributions, and fitting bivariate distributions of diameter and height were not investigated. The purpose of this paper is to address these issues.

Diameter and height are the two most important distributions for describing the horizontal and vertical structural characteristics of even-aged single species stands respectively (Chen & Rose 1978). However, there has been more extensive discussion of diameter distributions in the literature than height distributions. The most common continuous univariate distribution functions that have been used to describe diameter distributions are the Weibull distribution (Weibull 1951, Bailey & Dell 1973), gamma distribution (Nelson, 1964), lognormal distribution (Bliss & Reinker 1964), beta distribution (Clutter & Bennett 1965, Zöhrer 1969) and Johnson S_B distribution (Hafley & Schreuder 1977). Height distributions have been studied and reported by Hafley and Schreuder (1977), Chen and Rose (1978) and Kassier and Bredenkamp (1994), among others.

Forest managers and communities may also be interested in the distribution in size of both diameter and height. This information can be generated using bivariate distributions (Schreuder & Hafley 1977). Bivariate distributions are also useful for determining regression relationships between diameter and height. In the literature, Hafley and Schreuder (1976) compared several bivariate distributions which are potentially useful for describing joint frequency distributions of tree diameters and heights. Their results showed that the Johnson S_{BB} distribution was more flexible for describing joint frequencies of diameter and height compared with the other bivariate distributions they examined.

In this paper, temporary sample plot data were used to develop theoretical diameter and height distribution models for neem plantations in northern Ghana. Height distributions were modelled using direct and indirect approaches. The direct method involved directly fitting height observations to the normal, lognormal, Johnson S_B and gamma distributions. The indirect method used a relationship between diameter and height to indirectly derive a height distribution from a diameter distribution. Since diameters are more frequently measured (due to their ease of measurement) than heights, the ability to satisfactorily predict a height distribution from a diameter distribution will be of importance to local foresters. Finally, three bivariate distributions, the normal (S_{NN}), lognormal (S_{LL}), and Johnson S_{BB} (Johnson 1949b), were compared with the aim at determining which one best described the joint diameter and height frequency distributions of these plantations.

Materials and methods

Study area

The data used in the empirical analyses in this paper were collected from neem plantations in the northern region of the Republic of Ghana. This region is located approximately between longitude 2° 45" W and 0° 15" E and latitude 8° and 10° N. The vegetation type is guinea savannah, which is characterised by distinct wet (rainy) and dry seasons of about equal duration. There is a moderate mean annual rainfall of 960 to 1200 mm falling in one season from March/April to October and showing a very irregular distribution within a rainy season and great differences from year to year. Maximum rainfall during the year is achieved in July and August. The mean annual temperature of about 28 °C does not vary significantly during the seasons (Fisher 1984).

The characteristic vegetation is made up of short deciduous, widely spaced and heavily branched fire-resistant trees. They seldom form a closed canopy and overtop an abundant ground flora of grasses and shrubs of varying heights (Taylor 1952). The soils of the guinea savannah zone are varied because of the varied nature of the underlying geology. In general, however, two broad groups of soils are recognised: the savannah ochrosols and the groundwater laterites. The savannah ochrosols are found on the Voltaian sandstones (Boateng 1966). The groundwater laterites are the most extensive and are found on the Voltaian shales and granites.

Sampling procedure

Data were collected from 120 temporary sample plots selected from 30 plantations within the study area using a stratified two-stage sampling design. All plantations of neem planted from 1986 to 1994 within the study area were stratified by age into five groups (one-, two-, three-, four- and five-year age groups). Five plantations were randomly selected for measurement in each stratum. In addition, three and two plantations from six- and nine-year age groups respectively were also selected randomly for measurement. In all plantations selected for study, four square 10 × 10 m plots [1/100 ha] were established in each stand. To avoid the effect of errors in locating plot boundaries, plot boundaries were laid exactly halfway between the rows of trees. Since the usual initial spacing for neem and other plantation species in the study area is 2 × 2 m, in the absence of mortality and selective harvesting, each sample plot contained 25 trees. Secondly, all trees in each plot were measured for diameter at 50 cm above ground and for diameter at breast height (dbh) with either digital callipers or a diameter tape (for the larger trees) to the nearest mm, and for total height with height poles to the nearest cm. For trees ≤ 130 cm in height, only diameter at 50 cm and height were measured. However, in all estimations in the rest of this paper, diameter refers to the dbh (i.e. 1.3 m). Summary of the relevant statistics on diameter and height for each age group are given in Table 1. Note that not all trees measured were used in the analysis. Trees below dbh were eliminated. The older age classes (four years and older) had lower densities than the younger ages because of selective harvesting and mortality. Variations in diameter and height within age classes are expected to be influenced by site quality variation among plantations. However, for this particular study, differences in site quality is not considered a serious problem in influencing stand density since at least 50% of all plantations in each age group was on site class III (Nanang *et al.*, 1997). Parameter estimation and fitting of distributions were carried out using the MS EXCEL spreadsheet programme.

Direct method of fitting univariate diameter and height distributions

The classical approaches to diameter and height modelling were based on the assumption that at any point in time, the underlying diameter or height distribution of a stand under study can be adequately characterised by a probability distribution function from which the number of trees between specified diameters or height can be obtained (Knoebel & Burkhart 1991). The direct method, therefore, involved the use of observed measurements of either diameter or height to estimate frequencies directly from the four theoretical distributions under study. Preliminary examination of the beta and Weibull distributions produced predicted diameters and heights that were statistically significant from the observed in all age classes; so these two distributions were eliminated from further consideration.

The probability density functions (pdf) of the univariate gamma and S_B distributions are described below. The normal and lognormal distributions have been widely discussed in the statistics literature (e.g. Bury 1975, Johnson *et al.* 1994) and so the estimation procedures for these two distributions will not be repeated here. The estimated parameters for each distribution were used with the respective cumulative distribution functions to recover the frequencies by diameter or height class. In the following brief discussion of the distributions, x refers to either diameter at 1.3 m or total height, depending on the distribution of interest.

Gamma distribution

The pdf of the two-parameter gamma model with σ and λ as scale and shape parameters respectively is given as:

$$f(x) = \frac{1}{\sigma\Gamma(\lambda)}\left(\frac{x}{\sigma}\right)^{\lambda-1}\exp\left[-\left(\frac{x}{\sigma}\right)\right] \tag{1}$$

Bury (1975) provided maximum likelihood equations for estimating the two parameters as follows:

$$\hat{\sigma}\hat{\lambda} = \bar{x} \tag{2}$$

and

$$\ln\hat{\sigma} + \psi\left(\hat{\lambda}\right) = \ln G \tag{3}$$

where

$G = \prod_{i=1}^{n} x_i^{1/n}$ = geometric mean of x and

ψ = diagamma function.

The estimated parameters of the gamma distribution using equations (2) and (3) for each age group are given in Table 2.

Johnson S_B distribution

The Johnson S_B distribution (Johnson 1949a) has four parameters, two of which are the lower limit (ξ) and the range (ε) respectively. The pdf for the Johnson S_B distribution is:

$$f(x) = \frac{\delta}{\sqrt{2\pi}}\frac{\varepsilon}{(x-\xi)(\xi+\varepsilon-x)}\exp\left[-\frac{1}{2}\left(\gamma+\delta\ln\left(\frac{(x-\xi)}{\xi+\varepsilon-x}\right)\right)^2\right] \tag{4}$$

$\xi < x < \xi + \varepsilon,\ \delta > 0,\ -\infty < \gamma < \infty,\ \varepsilon > 0,\ -\infty < \xi < \infty = 0$ elsewhere

where

$$\gamma + \delta \ln\left(\frac{x-\xi}{\xi+\varepsilon-x}\right) = z_x \approx N(0,1)$$

The estimation procedures for the remaining two parameters are discussed later. In the fitting procedure used in this study, the smallest diameter and height were set as the lower bound for each age group. These lower bounds are represented by ξ in Table 3.

Indirect method of fitting univariate height distributions

If a relationship exists between two variables and a distribution is known for one of the variables, then a distribution can be derived for the other variable provided that the function describing the relationship can be inverted (Chen & Rose 1978). For a given site and stand age, an indirect height distribution can be generated from a diameter distribution based on the relationship between height and diameter. In this study, the relationship given by $D = (H/\theta)^{1/\phi}$ was used, where H is the height, D is the diameter and θ and ϕ are the positive regression constants related to species, age, site and stand density. Height distributions were, therefore, indirectly derived from the diameter distributions using $D = (H/\theta)^{1/\phi}$ in the cumulative density functions of the four univariate distributions. This equation was used because it was the best among those developed using the three bivariate distributions given in Table 8. Chen and Rose (1978) also found this relationship to be suitable for indirectly deriving a height distribution from a diameter distribution for a 32-year-old red pine plantation.

Fitting the bivariate distributions

The bivariate distributions considered in this paper were those for which both marginals are normal, lognormal and S_B distributions. Johnson and Kotz (1972) provided formulae for approximating bivariate probabilities of these distributions. Hafley and Schreuder (1976) noted that fitting the bivariate gamma distribution is extremely complicated; neither the parameter estimation nor the calculation of probabilities is easy. Knoebel and Burkhart (1991) also noted that the bivariate gamma distribution does not fit positively skewed and symmetrical distributions well. For the first reason, the bivariate gamma distribution was not pursued despite the good performance of its univariate counterpart in describing the diameter and height data.

If z_1 and z_2 are the standard normal variates for the Johnson S_{BB} distribution where

$$z_1 = \gamma_1 + \delta_1 \ln\left(\frac{y_1}{1-y_1}\right) \qquad (5)$$

and

$$z_2 = \gamma_2 + \delta_2 \ln\left(\frac{y_2}{1-y_2}\right) \tag{6}$$

then z_1 and z_2 have the joint bivariate normal distribution (pdf) with correlation coefficient ρ given as:

$$p(z_1,z_2;\rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(z_1^2 - 2\rho z_1 z_2 + z_2^2\right)\right] \tag{7}$$

In terms of diameter and height, $y_1 = \left(\frac{D-\xi_1}{\varepsilon_1}\right)$ and $y_2 = \left(\frac{D-\xi_2}{\varepsilon_2}\right)$ where D and H represent diameter and height respectively, and ε_1, ε_2 and ξ_1, ξ_2 are as defined above. The two remaining parameters of the S_B distribution are estimated as: $\hat{\gamma} = -\bar{f}/s_f$ and $\hat{\delta} = 1/s_f$ where $\bar{f}$ is the mean and s_f is the standard deviation of the transformation $f = \ln\left(\frac{y}{1-y}\right)$ with y defined as above.

Johnson (1949b) and Schreuder and Hafley (1977) discussed the properties of the S_{BB} distribution. One of the properties of interest is the regression relationship between height and diameter. Johnson (1949b) derived the median regression for the S_{BB} as follows:

$$\left(\frac{H-\xi_2}{\varepsilon_2}\right) = \theta\left[\left(\frac{\xi_1+\varepsilon_1-D}{D-\xi_1}\right)^{\phi} + \theta\right]^{-1} \tag{8}$$

where $\theta = \exp\left(\frac{\rho\gamma_1-\gamma_2}{\delta_2}\right)$ and $\phi = \rho\frac{\delta_1}{\delta_2}$ with $\phi > 0$.

In general, the mean regression is complicated and so the median regression is often used (Johnson 1949b, Schreuder & Hafley 1977).

In the case of S_{NN} distribution, the pdf is also given by equation (7), with z defined as:

$$z_1 = \gamma_1 + \delta_1 D,\ \ z_2 = \gamma_2 + \delta_2 H;\ \ \hat{\gamma} = -\bar{x}/s_x,\ \ \hat{\delta} = 1/s_x$$

where

$\bar{x}$ = mean of diameter or height and
s_x = the standard deviation.

For the lognormal bivariate distribution (S_{LL}), the pdf is given as (Bury 1975):

$$p(z_1,z_2;\rho)=\frac{1}{2\pi\left(\sqrt{1-\rho^2}\right)DH}\exp\left[-\frac{1}{2\left(1-\rho^2\right)}\left(z_1^{\ 2}-2\rho z_1 z_2+z_2^{\ 2}\right)\right] \quad (9)$$

where

$z_1=\gamma_1+\delta_1\ln D,$

$z_2=\gamma_2+\delta_2\ln H,$

$\hat{\gamma}=-\bar{g}/s_g,$

$\hat{\delta}=1/s_g$ and

$g=\ln D$ or $g=\ln H$.

The mean and standard deviation of the log-transformed diameter or height are $\bar{g}$ and s_g respectively. For the bivariate normal and lognormal distributions, the median regressions are (Johnson 1949b):

$$H=\ln\theta+\phi D \quad (10)$$

and

$$H=\theta D^{\phi} \text{ respectively,} \quad (11)$$

where

H = mean or expected total height in m for a given dbh,
D = the tree dbh in cm and
θ and ϕ = positive regression coefficients as defined under equation (8) and are related to species, site and density.

Results

Univariate distributions

Figures 1 and 2 show the predicted and observed diameter and height frequencies from all four univariate distributions respectively (for all age groups combined). In general, all four distributions predicted diameter and height that were close to the observed. The figures show that the Johnson S_B performed well on the lower diameter and height classes, but overestimated both diameter and height in all classes above 5 cm and 7 m respectively.

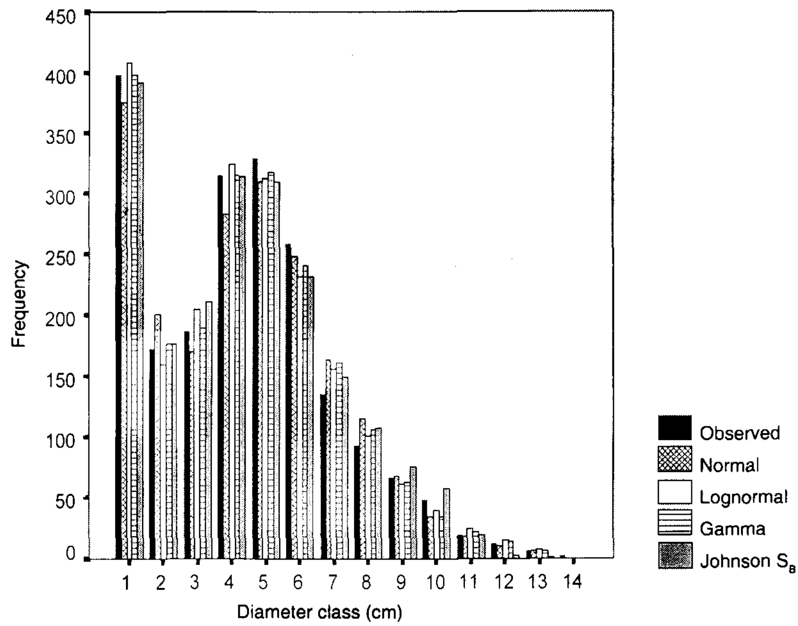

Figure 1 Comparison of observed and predicted diameter frequencies from the normal, lognormal, gamma and Johnson S_B distributions for the neem plantations (all age groups combined)

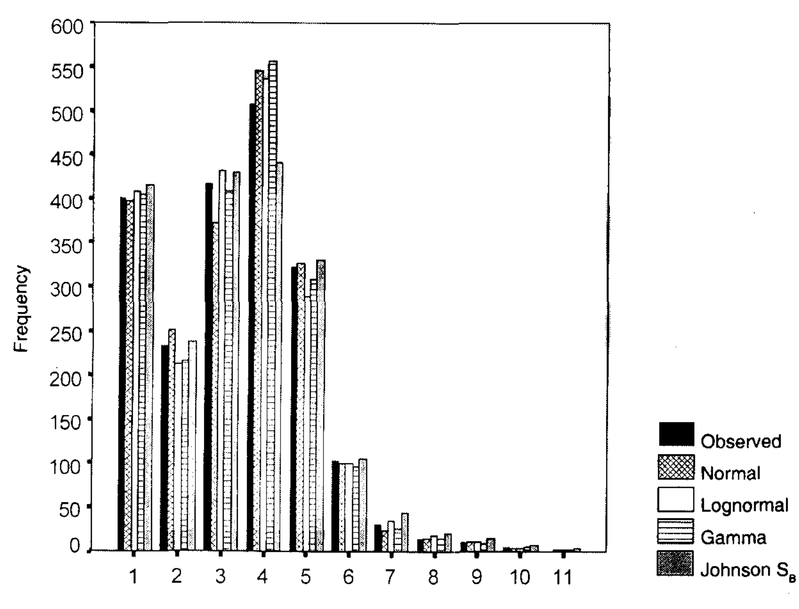

Figure 2 Comparison of observed and predicted height frequencies from the normal, lognormal, gamma and Johnson S_B distributions for the neem plantations (all age groups combined)

The quality of fit of the four univariate distributions for each age group was assessed using the Kolmogorov-Smirnov (KS) statistic (Sokal & Rohlf 1981). The ranking of each distribution based on the KS statistic for the seven age groups is summarised in Table 4. Ranks are in ascending order; the distribution with the smallest KS statistic between observed and predicted frequencies is given a rank of 1. The KS criterion compares the absolute difference between the cumulative frequency of the observed and expected frequencies. The null hypothesis is rejected when the greatest absolute difference between the observed and expected cumulative frequency is greater than the critical value. At $\alpha = 0.05$, the critical value for $n > 30$ is $1.36/\sqrt{n}$. The general problems with the use of the KS statistic as a goodness-of-fit test that are discussed in Reynolds *et al.* (1988) apply to this study; in particular, the fact that the KS test tends to be conservative. Almost all the univariate distributions predicted statistically insignificant differences between the observed and predicted frequencies for both diameter and height when the actual measured sample size was used to calculate the critical value (except those marked with asterisks in Table 4). However, when the data were converted to per hectare basis, most of these were significant because of the increase in sample size, which consequently lowered the critical value from 0.06094 to 0.02725. Based on the rank sums for the seven age groups, the gamma distribution was judged the best in describing the diameter data, whilst the lognormal was best for the height data (Table 4).

Table 1 Summary of diameter and height data used in this study

Age (year)	n	Diameter						Height					
		Min	Mean	Max	SD	Skewness	Kurtosis	Min	Mean	Max	SD	Skewness	Kurtosis
1	187	0.50	0.84	2.64	0.42	1.71	3.40	1.50	1.64	1.80	0.09	0.05	-1.16
2	369	0.50	1.42	4.37	0.70	1.23	1.92	1.50	2.38	3.13	0.37	0.32	-1.02
3	498	1.14	3.99	9.43	1.41	0.47	0.86	2.11	3.96	4.51	0.37	0.43	-0.89
4	400	2.00	5.05	11.00	1.21	1.12	2.16	3.11	4.88	5.25	0.21	0.01	-1.12
5	450	2.29	6.53	11.51	1.55	0.59	-0.21	2.70	5.82	6.67	0.39	0.39	-0.89
6	200	4.71	8.61	13.19	1.92	0.08	-0.69	3.71	7.21	8.17	0.45	0.68	-0.73
9	100	7.50	11.40	16.80	1.87	0.88	1.54	6.72	9.45	10.80	0.85	0.06	-1.35

Table 2 Estimated parameters of the gamma distribution

Age (year)	Diameter		Height	
	λ_1	σ_1	λ_2	σ_2
1	7.020	0.143	35.936	0.054
2	5.995	0.229	22.389	0.106
3	12.743	0.331	24.774	0.189
4	12.753	0.426	38.107	0.124
5	10.080	0.602	27.702	0.195
6	38.194	0.214	47.837	0.122
9	49.193	0.232	45.627	0.202

Table 3 Estimated parameters of the Johnson S_B distribution

Age (year)	Diameter				Height			
	ξ_1	ε_1	γ_1	δ_1	ξ_2	ε_2	γ_2	δ_2
1	0.500	2.150	1.129	0.780	1.500	1.620	0.921	0.747
2	0.500	3.100	1.000	0.872	1.500	2.460	0.641	0.804
3	1.140	6.870	0.311	1.276	2.110	5.650	0.257	1.282
4	2.000	9.510	0.789	1.226	3.110	4.530	0.717	1.023
5	2.290	10.90	0.804	1.086	2.700	6.450	0.475	1.319
6	4.710	6.580	- 0.128	0.992	3.710	4.430	0.110	0.838
9	7.500	6.500	0.455	1.306	6.720	4.080	- 0.501	0.707

Table 4 Ranking of the normal, lognormal, S_B and gamma distributions for diameter and height (rank in parenthesis) based on the KS statistic

Age (year)	Diameter				Height			
	Normal	Lognormal	S_B	Gamma	Normal	Lognormal	S_B	Gamma
1	1	2	3	2	2	2	1	3
2	4*	3	2	1	3	2	2	1
3	1	4	3	2	4	1	3	2
4	3	1	4	2	3	1	1	2
5	4	1	3	2	1	4	3	2
6	1	3	4	2	1	2	3	3
9	1	1	3*	2	2	1	2	2
Rank sum	15(2)#	15(2)	22(3)	13 (1)	16 (3)	13(1)	15(2)	15(2)

*Significant at the 5% level, #rank of distribution for all ages

The observed and predicted diameter and height frequencies for the three-year age group are presented in Tables 5 and 6 respectively. This age group was chosen for presentation because it had the highest stocking among the age classes studied as a result of selective harvesting in plantations older than three years. All the four distributions performed well in predicting diameter and height in this age group, except in the tails of the distributions. In particular, the normal distribution overestimated both diameter and height in the lower tails, whilst the remaining three distributions overestimated these two variables in the upper tails.

Given that the lognormal was the best in fitting the observed height data, followed by the gamma distribution, the possibility of predicting parameters of the lognormal and gamma distributions for height from other stand attributes was considered useful. A regression method was used to predict the height parameters of the gamma and lognormal distributions from the estimated parameters of the diameter distribution and stand age. In this procedure, the parameters of both diameter and height of the gamma and lognormal distributions were estimated for all 30 plantations and linear regressions of the height parameters on stand age and diameter parameters estimated. The predicted height parameters were then used to fit a height distribution for each age class and the predicted frequencies

Table 5 Observed and predicted diameter frequencies (number of trees ha^{-1}) from the normal, lognormal, S_B and gamma distributions for the three-year age group

Diameter class (cm)	Observed	Normal	Lognormal	S_B	Gamma
0.5–1.49	15	25	0	0	5
1.5–2.49	140	150	135	170	125
2.5–3.49	520	495	645	580	590
3.5–4.49	825	810	805	740	840
4.5–5.49	610	670	520	610	580
5.5–6.49	310	275	240	315	250
6.5–7.49	65	60	95	65	75
7.5–8.49	5	5	35	10	20
8.5–9.49	0	0	15	0	5
Total	2490	2490	2490	2490	2490
KS statistic		0.01606	0.04217*	0.03012*	0.01807

* Significant at 5% level; the critical value of the KS statistic at this level of significance is 0.02725

Table 6 Observed and predicted height frequencies (number of trees ha^{-1}) from the normal, lognormal, S_B and gamma distributions for the three-year age group

Height class	Observed	Normal	Lognormal	S_B	Gamma
1.5–2.49	5	30	5	5	5
2.5–3.49	280	245	245	290	230
3.5–4.49	830	780	905	800	880
4.5–5.49	865	935	850	835	910
5.5–6.49	405	425	360	460	375
6.5–7.49	100	70	100	80	80
7.5–8.49	5	5	20	20	10
8.5–9.49	0	0	5	0	0
Total	2490	2490	2490	2490	2490
KS statistic		0.024096	0.01606	0.02008	0.01807

Note: The critical value of the KS statistic at 5% level is 0.02725

compared with the observed using the KS statistic. The four regressions estimated to predict height parameters for the gamma (equations (12) and (13)) and lognormal (equations (14) and (15)) distributions are:

$$\hat{\lambda}_H = 27.755 + 0.693\hat{\lambda}_D - 1.537A \qquad R^2 = 0.686 \qquad (12)$$

$$\hat{\sigma}_H = 0.036 + 0.167\hat{\sigma}_D - 0.013A \qquad R^2 = 0.676 \qquad (13)$$

$$\hat{\overline{H}} = 0.631 + 0.456\overline{D} + 0.047A \qquad R^2 = 0.983 \qquad (14)$$

$$\hat{s}_H = 0.148 + 0.162s_D - 0.004A \qquad R^2 = 0.523 \qquad (15)$$

where

D and H identify diameter and height parameters respectively,
A = stand age and
s_H and s_D represent standard deviations of height and diameter respectively.

The predicted height frequencies for the gamma distribution using parameters from equations (12) and (13) showed that the KS statistic for the four-, five-, and nine-year age groups were not significant, whilst the remaining age groups had frequencies significantly different from the observed. In the case of the lognormal distribution, all predicted height frequencies using parameters from equations (14) and (15) were significantly different from the observed.

Indirect method

The results from the indirect method of fitting height distributions were a bit disappointing, as none of the four indirectly derived height distributions were satisfactory based on the KS criterion. For the normal distribution, all indirectly derived height frequencies were significantly different from the observed; whilst for the lognormal, Johnson S_B and gamma distribution, only the two- and five-year age groups were insignificant. However, the gamma distribution was best overall with the smallest KS statistic in all age groups, though five of these statistics were significant at the 5% level. Comparisons of height frequencies from the direct and indirect methods were made using the KS statistic. The results showed that for the normal distribution, the statistic was significant for all age groups. For the lognormal, only the two- and five-year age groups were not significant, whilst for the Johnson S_B, the one- and two-year age groups were non-significant. Only the five-year age group was not significant for the gamma distribution.

Bivariate distributions

The lognormal bivariate distribution was the most appropriate one to fit the diameter and height frequencies. This distribution provided predicted frequencies that were not statistically significant from the observed for six out of the seven age groups studied. The bivariate Johnson S_B was also found to provide a satisfactory fit, as four out of the seven age groups were not significant. The normal distribution was less satisfactory, with only one age group (age 2) providing a good fit. The observed and predicted frequencies for all age groups are not presented here, however, as before, the three-year age group is used as an example to present the observed and predicted bivariate frequencies for the S_{LL} distribution (Table 7).

Regressions from the three bivariate distributions

Median regressions of height on diameter from the three bivariate distributions (equations (8), (10) and (11)) were estimated and compared based on the χ^2 test and the sums of square deviations from the observed heights. Table 8 gives the estimated regression equations and the relevant statistics.

The regression from the S_{LL} distribution provided the best fit for the criteria used in the comparisons. The regressions in Table 8 were developed using the data from all age groups combined. The three regressions were used to predict heights of all observed diameters in the study. The χ^2- statistic showed that the observed heights were not significantly different from the predicted heights.

Similar regressions were estimated for each age group and subsequently, developed height distributions from the predicted heights. This was done to determine whether, given the diameter data alone, and the height/diameter relationship, it would be possible to generate a height distribution that was close to the observed. Based on the χ^2 test, the predicted heights were not significantly different from observed heights for all three-equation types. However, when these predicted heights were classified into height classes, the resulting distributions

Table 7 Observed and predicted (O/P) frequencies (number of trees ha^{-1}) from the S_{LL} distribution of the neem plantations for the three-year age group

Diameter class (cm)	Height class (m) 2	3	4	5	6	7	8	Total
1	5/2	10/13						15/15
2		115/120	25/18					140/138
3		120/105	335/312	60/73	5/9			520/499
4		30/25	370/395	370/362	55/43			825/825
5		5/4	90/85	325/315	155/159	30/63	5/3	610/629
6			10/5	105/120	160/152	35/28	0/2	310/307
7				5/8	30/48	30/20		65/76
8						5/1		5/1
Total	5/2	280/267	830/815	865/878	405/411	100/112	5/5	2490

Table 8 Comparison of the three regression equations in terms of the sums of squares and χ^2 values between observed and predicted heights (all ages combined)

Bivariate distribution	Equation	Sum of squares	χ^2- values
Normal (S_{NN})	$H = 1.8327 + 0.5688D$	1130.60	237.82
Lognormal (S_{LL})	$H = 2.078D^{0.5252}$	993.53	194.81
Johnson (S_{BB})	$(H - 1.50) = 7.732\left[\left(\frac{14.0 - D}{D - 0.5}\right)^{0.7245} + 0.8314\right]^{-1}$	1046.67	216.26

Note: Significant at 5%; critical value for the $\chi^2_{(2204)}$ - distribution at 5% is 2313.03

for all age groups were significantly different from the observed for the regression from the S_{NN} distribution. For the S_{LL} and S_{BB} regressions, three out of the seven age groups gave frequencies that were not significant. The frequencies from these regressions were also compared with predicted frequencies from the normal, lognormal and S_B distributions (direct method). That is, frequencies from the S_{NN} regression were compared with frequencies from the normal distribution for theseven age groups, the S_{LL} to the lognormal, and the S_{BB} to the S_B. Height frequencies from three out of the seven age groups from the S_{LL} and S_{BB} regressions were not significant, whilst for the S_{NN} regression, only the four-year age group was not significant.

Discussion

Univariate distributions

Comparison of the four univariate distributions used to fit diameter and height

In general, all four univariate distributions gave good fits to both diameter and height data. A comparison of the ranking of the KS statistic showed that the gamma and lognormal distributions were the best for describing the diameter and height data respectively. The fact that the Weibull distribution was not suitable for fitting these data is rather surprising. The range of kurtosis and skewness covered by the gamma, lognormal and Weibull distributions were close and so these three distributions tend to either fit data equally well or poorly (Hafley & Schreuder 1977). The Johnson S_B distribution is also shown to be more flexible than the normal, lognormal, Weibull, beta and gamma distributions for fitting diameter and height data (Hafley & Schreuder 1977). Given that these plantations were young, the observed distributions were generally positively skewed (Table 1). Therefore, the lognormal distribution was the appropriate choice for the diameter and height distributions if a single distribution was needed. Since this distribution is easy to fit even on a spreadsheet, it is a definite advantage for many developing countries where fast computers and expensive programmes may often be unavailable. A drawback of the univariate lognormal distribution is that because it is limited to describing data that is positively skewed, it generally approximates symmetrically distributed data poorly (Hafley & Schreuder 1976). As these plantations age beyond nine years, it is possible that the lognormal distribution will fail to satisfactorily describe the height distributions. This may not be a serious concern as most of these plantations are selectively harvested after the third year and final harvesting carried out by the time the plantations are 10 years of age.

Table 4 shows that the normal distribution had the smallest KS statistic between observed and predicted diameter frequencies in four out of the seven cases. This shows that the normal distribution may be preferred to the gamma distribution for fitting the diameter data, more especially because its parameters are easier to estimate and hence the distribution is easier to fit.

With the univariate gamma and lognormal distributions, it is possible to predict the number of trees ha^{-1} in any given diameter or height class. However, the use of neem for poles and rafters require that minimum diameter and height standards be met. The use of two separate univariate distributions does not allow for the simultaneous determination of frequencies that satisfy these diameter and height requirements. This is because for a given height (diameter), diameter (height) can vary considerably, depending on stocking and other factors.

Comparison of direct and indirect methods of estimating height distributions

It was observed from the results that whilst the direct method predicted height frequencies that were not different from the observed in most cases, the indirectly derived height distributions were not satisfactory. This result is contrary to that found by Chen and Rose (1978). They found that for a 32-year-old red pine plantation, both direct and indirectly derived height frequencies from the Weibull distribution were not significantly different from the observed frequencies. The accuracy with which the indirect method can predict observed frequencies depends on the diameter/height relationship used, which in turn depends on the strength of the correlation between the diameter and height observations. The parameters of equation (11) depend on the correlation coefficient between diameter and height. The correlation coefficients between diameter and height for the seven age groups ranged from 0.207 to 0.828. The highest correlation was observed in the two- and five-year age groups, which happened to produce insignificant differences. Since the same relationship was used in all age groups, the only plausible explanation is the differences in correlation. It is not clear why the highest correlation occurred in these two age classes. The accuracy of predicting heights could be improved if a regression that describes the diameter/height relationship better than the relationship used here is employed. In general, it is possible to derive an indirect height distribution from the diameter observations of these plantations; however, these should be considered as approximations. The linear regressions given by equations (12)–(15) can be used to estimate height parameters for the gamma and lognormal distributions if height observations are unavailable. However, given that the KS statistics between observed and predicted frequencies derived from the parameters estimated from these equations were significant, these regressions should be used in places where only approximate predictions of height frequency distributions are needed.

Comparison of bivariate distributions

The S_{LL} distribution seemed to fit the data better than the other two distributions, and resulted in a regression model that was a better predictor of observed heights than the S_{BB} and S_{NN}. This result is contrary to that of Hafley and Schreuder (1976), who found that the S_{BB} was not only more flexible in fitting diameter and height data than the S_{LL} and S_{NN}, but also resulted in a better diameter/height relationship than the other bivariate distributions. Given the range of the data used in this study, it is possible for the S_{BB} to perform better if older age classes were included. Expansion of the data set to include older age stands may not be useful in this case since the biologically optimum rotation ages of these plantations are five and seven years for site classes I and II respectively (Nanang *et al.* 1997), and economic rotations may even be shorter. Furthermore, most of these plantations are selectively harvested for poles and rafters as early as the third year; so that after the ninth year, any empirical investigation of stand structure will likely yield unreliable results. The regression relationship from the S_{LL} distribution, which was used by Chen and Rose (1978), predicted heights better than the other two regressions (Table 8). However, Curtis (1967) found the model from the S_{LL} distribution, which was first proposed by Greenhill (1881), to be less desirable for the Douglas fir data compared with other model forms examined.

With bivariate distributions, it is possible to estimate the number of trees ha^{-1} that meet the utilisation standards for poles and rafters. Field observations during data collection and sample measurements of poles and rafters often used in the study area show that the minimum diameters (at breast height) for poles and rafters are at least 5 and 10 cm respectively. The minimum height requirements are at least 3 m for poles and 4 m for rafters. Using the information in Tables 4 and 5, the number of trees ha^{-1} that satisfy the diameter and height requirements separately can be determined. Table 7 can be used to estimate the bivariate frequencies that satisfy these requirements simultaneously for age group three. Since the lognormal distribution is a transformation of the normal distribution, bivariate probabilities of the lognormal distribution can be calculated using the formulae by Johnson and Kotz (1972), or from charts and formulae provided by Owen and Wiesen (1959) or from statistical tables, e.g. by Pearson (1931).

Conclusions

The results of this study show that diameter distributions were best described by the gamma distribution. The best height model was the lognormal while the best bivariate model was the bivariate lognormal (S_{LL}). Regressions of height on diameter from the S_{LL} distribution predicted heights that were closer to the observed than regressions from the S_{NN} and S_{BB} based on the χ^2 test. When tree utilisation requirements include limits on both diameter and height, a bivariate distribution should be used instead of fitting two separate univariate distributions.

Acknowledgement

Thanks to T. K. Nunifu, Department of Renewable Resources, University of Alberta.

References

BAILEY, R. L. & DELL, T. R. 1973. Quantifying diameter distributions with the Weibull function. *Forest Science* 19(2): 97–104.

BLISS, C. I. & REINKER, K. A. 1964. A lognormal approach to diameter distributions in even-aged stands. *Forest Science* 10: 350–360.

BOATENG, E. A. 1966. *A Geography of Ghana.* 2nd edition. Cambridge University Press, Cambridge. 212 pp.

BURY, K. V. 1975. *Statistical Models in Applied Science.* John Wiley & Sons Inc., New York. 625 pp.

CHEN, C. M. & ROSE, D. W. 1978. *Direct and Indirect Estimation of Height Distributions in Even-Aged Stands.* Minnesota Forestry Research Notes. No. 267. 3 pp.

CLUTTER, J. L. & BENNETT, F. A. 1965. *Diameter Distributions in Old-Field Slash Pine Plantations.* Georgia Forestry Research Council Report 13. 9 pp.

CURTIS, R. O. 1967. Height-diameter and height-diameter-age equations for second growth Douglas fir. *Forest Science* 13(4): 365–375.

FISHER, N. M. 1984. The impact of climate and soil on cropping systems and the effect of cropping systems and weather on the stability of yield. Pp. 55–70 in Steinner, K. G. (Ed.) *Report on the On-Farm-Experimentation Training Workshop.* Nyankpala, Ghana. 3–13 July 1984. 120 pp.

GREENHILL, A. G. 1881. Determination of the greatest height consistent with stability that a vertical pole or post can be made, and of the greatest height to which a tree of given properties can grow. Pp. 65–73 in *Proceedings of the Cambridge Philosophical Society IV.* Part II. Cambridge.

HAFLEY, W. L. & SCHREUDER, H. T. 1976. Some non-normal bivariate distributions and their potential for forest application. Pp. 104–114 in *International Union of Forest Research Organisations. XVI World Congress Proceedings,* Div. VI. Oslo, Norway. 20 June–2 July 1976.

HAFLEY, W. L. & SCHREUDER, H. T. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. *Canadian Journal of Forest Research* 7: 481–487.

JOHNSON, N. L. 1949a. Systems of frequency curves generated by methods of translation. *Biometrika* 36: 149–176.

JOHNSON, N. L. 1949b. Bivariate distributions based on simple translation systems. *Biometrika* 36: 297–304.

JOHNSON, N. L. & KOTZ, S. 1972. *Distributions in Statistics: Continuous Multivariate Distributions.* John Wiley & Sons Inc., New York. 333 pp.

JOHNSON, N. L. KOTZ, S. & BALAKRISHNAN, N. 1994. *Continuous Univariate Distributions.* Volume I. 2nd edition. John Wiley & Sons Inc., New York. 756 pp.

KASSIER, H. W. & BREDENKAMP, B. V. 1994. Modelling diameter and height distributions through dispersion statistics in even-aged pine plantations. *South African Forestry Journal* 171: 21–27.

KNOEBEL, B. & BURKHART, H. 1991. A bivariate distribution approach to modeling forest diameter distributions at two points in time. *Biometrics* 47: 241–298.

NANANG, D. M., DAY, R. J. & AMALIGO, J. N. 1997. Growth and yield of neem (*Azadirachta indica* A. Juss.) plantations in Northern Ghana. *Commonwealth Forestry Review* 76(2): 103–106.

NANANG, D. M. 1998. Suitability of the normal, lognormal and Weibull distributions for fitting diameter distributions of neem (*Azadirachta indica* A. Juss.) plantations in Northern Ghana. *Forest Ecology and Management* 103(1): 1–7.

NELSON, T. C. 1964. Diameter distribution and growth of loblolly pine. *Forest Science* 10(1): 105–114.

OWEN, D. B. & WIESEN, J. M. 1959. A method of computing bivariate normal probabilities with application to handling errors in testing and measuring. *The Bell System Technical Journal* 38: 553–572.

PEARSON, K. 1931. *Tables for Statisticians and Biometricians.* Part II. Cambridge University Press, Cambridge. 262 pp.

REYNOLDS, M. R., BURK, T. E. & HAUNG, W. C. 1988. Goodness-of-fit tests and model selection procedures for diameter distribution models. *Forest Science* 34(2): 373–399.

SOKAL, R. R. & ROHLF, F. J. 1981. *Biometry.* 2nd edition. W. H. Freeman & Co., San Francisco. 857 pp.

SCHREUDER, H. T. & HAFLEY, W. L. 1977. A useful bivariate distribution for describing stand structure of tree heights and diameters. *Biometrics* 33: 471–478.

TAYLOR, C. J. 1952. *The Vegetation Zones of the Gold Coast.* Government Printer, Accra. 57 pp.

WEIBULL, W. 1951. A statistical distribution function of wide applicability. *Journal of Applied Mechanics* 18: 293–297.

ZÖHRER, F. 1969. Ausgleich von Haufigkeitsverteilungen mit Hilfe der Beta-Funktion. *Forstarchiv* 40(3): 37–42.