A Model for the Diameter-Height Distribution in an Uneven-Aged Beech Forest and a Method to Assess the Fit of Such Models

Walter Zucchini, Matthias Schmidt and Klaus v. Gadow

Zucchini, W., Schmidt, M. & Gadow, K. von. 2001. A model for the diameter-height distribution in an uneven-aged beech forest and a method to assess the fit of such models. Silva Fennica 35(2): 169–183.

This paper illustrates the application of a mixture model to describe the bivariate diameter-height distribution of trees growing in a pure, uneven-aged beech forest. A mixture of two bivariate normal distributions is considered but the methodology is applicable to mixtures of other distributions. The model was fitted to diameter-height observations for 1242 beech trees in the protected forest Dreyberg (Solling, Germany). A considerable advantage of the model, apart from the fact that it happens to fit this large data set unusually well, is that the individual parameters all have familiar interpretations. The bivariate Johnson S_{BB} distribution was also fitted to the data for the purpose of comparing the fits.

A second issue discussed in this paper is concerned with the general question of assessing the fit of models for bivariate data. We show how a device called “pseudo-residual” enables one to investigate the fit of a bivariate model in new ways and in considerable detail. Attractive features of pseudo-residuals include the fact that they are not difficult to interpret; they can be computed using generally available statistical software and, most important of all, they enable one to examine the fit of a model by means of simple graphs.

Keywords diameter-height distribution, mixture models, bivariate normal distribution, S_{BB} distribution, goodness-of-fit, pseudo-residuals, beech forest.

Authors’ addresses *Zucchini*: Georg-August-Universität Göttingen, Institute for Statistics and Econometrics; *Schmidt*: Forest Research Station of Lower Saxony; *von Gadow*: Georg-August-Universität Göttingen, Institute for Forest Management and Yield Sciences
Fax +49 551 397 279 **E-mail** zucchini@wi-wiss.uni-goettingen.de

Received 10 November 1999 **Accepted** 22 January 2001

1 Introduction

One of the most important elements of forest structure is the relationship between tree diameters and heights. Information about size-class distributions of the trees within a forest stand is important for estimating product yields. The size-class distribution influences the growth potential and hence the current and future economic value of a forest stand (Knoebel and Burkhart 1991). More recently, a rather different need for diameter-height relations arose in response to an increasing interest in the structure of natural forests. Unmanaged forests are used as a standard for comparison of different types of managed stands, using indices such as the Shannon-Index for analysis of vertical structure (Shannon 1948, Weber 1998). Thus, detailed modelling of the variation of heights within diameter classes is required.

The product yield estimates are usually based on an assessment of the distribution of diameters and a common practice is to fit a model, such as the Weibull function, to the empirical diameter class frequencies (Clutter and Allison 1974, Gadow 1987, Saborowski 1994). For improved product estimates, tree heights are routinely assessed and a variety of models are used for defining the relationship between diameters and heights (Schmidt 1967, Curtis 1967). The traditional method does not quantify the distribution of heights for a given diameter and one approach for modelling the conditional height distribution for the different diameters is to use the height residuals (Gaffrey 1996, p. 258). In our experience, it is very seldom that the height residuals are homoscedastic and normally distributed. In most forest stands the variance about the diameter-height regression is heterogeneous.

Hence there has been considerable interest in identifying suitable bivariate distributions to describe diameter-height frequency data. It has been reported that the S_B distribution (Johnson 1949a) fits the marginal frequencies of both diameter and height consistently better than the Weibull, beta, gamma, lognormal and normal distributions (Johnson 1949b, Hafley and Schreuder 1976) and that the bivariate extension of the S_B distribution, the S_{BB}, is more realistic and provides more useful information than the currently accepted approach for describing forest diameter-height data (Schreuder and Hafley 1977).

Beech forests often develop a single layer structure when management stops, as a result of heavy shading and drought eliminating the smaller suppressed trees of this species. A variety of vertical structures may develop in unmanaged beech forests, depending on the successional stage. There are numerous examples showing that virgin beech forests exhibit structures which include more than one layer of tree heights (Korpel 1992, Košir 1966). In a managed beech forest, the vertical structure depends on the type of thinning that is applied. In a high thinning, which is generally practised in Germany, only bigger trees are removed while the smaller ones may survive for a very long time, resulting in a typical pattern with two subpopulations. Usually, the gradient of the diameter-height regression is much steeper in the smaller subpopulation. A possible explanation for this phenomenon may be the greater competition for light in the lower layer. Thus, trees in the lower layer have to support height growth more than diameter growth. On the other hand, because of static reasons, trees that have reached the upper layer have to invest more in diameter growth. The result of these different strategies are different diameter-height relations and, for most managed beech forests, it would be biologically plausible to find that the population of trees is composed of a mixture of two subpopulations having different diameter-height distributions. The particular model that is discussed here is specifically designed to describe such populations.

The purpose of this paper is to illustrate how one can go about fitting a mixture of bivariate normal distributions to diameter-height observations. The bivariate Johnson S_{BB} was also fitted to the data for the purpose of comparison. Secondly we show how "pseudo-residuals" can be used to assess the fit of this, or any other, model that one has fitted to univariate or bivariate observations.

2 Material and Methods

2.1 Study Material

The models investigated in this paper were fitted to diameter-height measurements on 1242 beech trees in the protected forest Dreyberg (Solling, Lower Saxony, Germany). The data were collected in 1995 by the Forest Research Station of Lower Saxony from 50 permanent 1000 m^2 circular plots on a 100 m × 100 m regular grid. The age of the stand, between 113 and 120 years, can be determined only approximately because the stand originated from natural regenerations. Some stand characteristics are given in Table 1.

The dominant species is beech and therefore the stand is close to the potential association of a *luzulo fagetum* which is typical for extensive areas in the Southern Lower Saxony hill country.

Table 1. Mean stand characteristics with standard error (s.e.) for the whole stand and for beech trees of the Dreyberg data.

	Whole stand	s.e.	Beech
Trees/ha	273.4	10.1	269.6
Stand basal area (m^2 / ha)	28.4	0.5	28.0
Standing volume (m^3 / ha)	383.3	9.2	377.9

A simple height-diameter curve as shown in Fig. 1, describes how the mean height varies with diameter (at breast height) but it does not quantify the complete distribution of heights for each diameter.

Fig. 2 shows a nonparametric (kernel) estimate of the bivariate diameter-height distribution (see for example Silverman 1986). This figure indicates that there may be two subpopulations of trees and that the height-diameter relationship differs in these subpopulations. The larger subpopulation (approximately 80% of the entire population) comprises larger trees in which the slope of the height-diameter regression is less steep than that for the smaller subpopulation (approximately 20% of the population).

One way of modelling such a distribution is to use a mixture of two bivariate normal distributions described in Section 2.2). The resulting regression curve (the conditional expectation of height given diameter) for such a model is a smooth function of the two straight lines: steeper for the smaller trees subpopulation and less steep for the dominant trees subpopulation. The general shape of this curve is similar to that shown in Fig. 1.

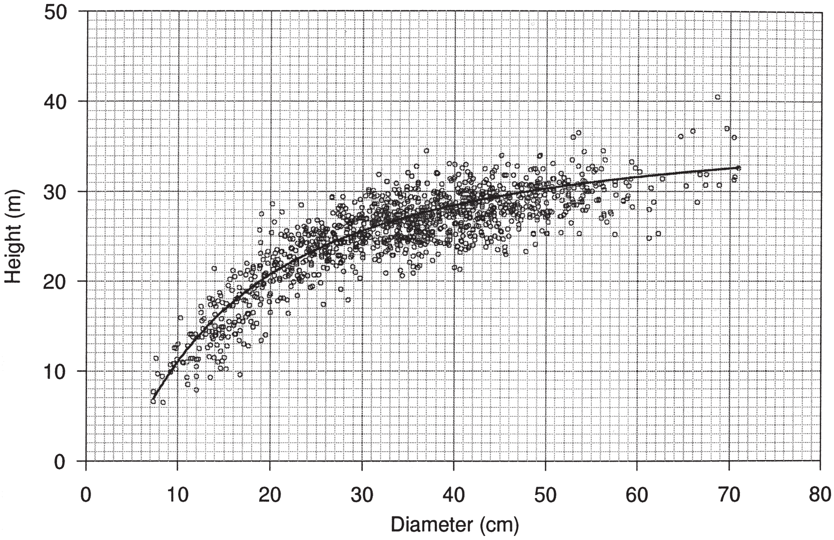


Fig. 1. Breast height diameter–height data for 1242 beech trees in the protected forest Dreyberg (Solling). The height-curve has been drawn using $h = e^{\alpha+\beta/d}$ with $\alpha = 3.6638$ and $\beta = -12.601$.

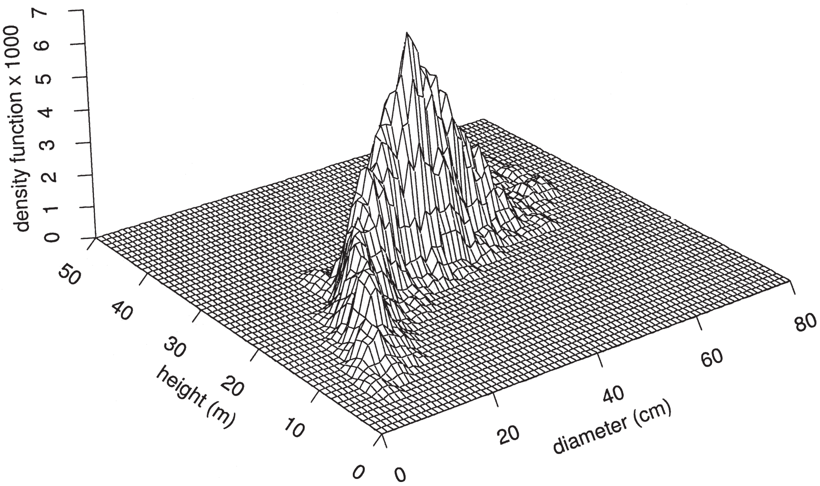


Fig. 2. Perspective plot of a nonparametric (kernel) estimate of the bivariate density function for the diameters and heights.

Section 2.3 gives an outline of the Johnson S_{BB} model which we also fitted to the data for the purpose of comparison.

2.2 A Mixture of Two Bivariate Normal Distributions

Let $f(d, h)$ denote the bivariate probability density function of diameter and height. The proposed model is

$$f(d,h) = \alpha n_1(d,h) + (1-\alpha) n_2(d,h) \tag{1}$$

where α, a parameter in the interval (0, 1), determines the proportion of trees belonging to each of the two component bivariate normal distributions $n_1(d, h)$ and $n_2(d, h)$. The parameters of $n_j(d, h)$ are the expectations μ_{dj}, μ_{hj}; the variances σ_{dj}^2 and σ_{hj}^2, and the correlation coefficient, ρ_j, $j = 1, 2$. Thus the full probability density function of diameter and height is given by (1) with

$$n_j(d,h) = \frac{J}{2\pi\sqrt{1-\rho_j^2}} \exp\left\{ -\frac{z_{dj}^2 - 2\rho_j z_{dj} z_{hj} + z_{hj}^2}{2(1-\rho_j^2)} \right\} \tag{2}$$

$z_{dj} = (d - \mu_{dj}) / \sigma_{dj}$, $z_{hj} = (h - \mu_{hj}) / \sigma_{hj}$ and
$J = (\sigma_{dj}\sigma_{hj})^{-1}$, $\quad j = 1,2$

The term J, the Jacobian of the transformation from the standard to the general bivariate normal, is identified explicitly here in order to contrast this distribution with the Johnson S_{BB} in Section 2.3.

The marginal distributions for the diameters and heights are also mixtures of the corresponding component marginal distributions. They are of the form

$$f(x) = \alpha n_1(x) + (1-\alpha) n_2(x) \tag{3}$$

The components for the diameters and heights are given, respectively, by

$$\begin{aligned} n_j(d) &= \frac{\exp(-z_{dj}^2/2)}{\sqrt{2\pi}\sigma_{dj}}, \\ n_j(h) &= \frac{\exp(-z_{hj}^2/2)}{\sqrt{2\pi}\,\sigma_{hj}}, \quad j = 1,2 \end{aligned} \tag{4}$$

The conditional distributions of height given diameter, and of diameter given height are not mixtures of normal distributions but they can be computed using expressions (1) to (4) from their definitions:

$$f(h|d) = f(d,h) / f(d),$$
$$f(d|h) = f(d,h) / f(h) \qquad (5)$$

The maximum likelihood estimators of the parameters of this model are not available in explicit form but the estimates are not difficult to compute by directly maximizing the likelihood function or (more conveniently) by minimizing the negative log-likelihood function:

$$\ell(Parameters;(d_i,h_i),i = 1,2,...,n) =$$
$$-\sum_{i=1}^{n} \log(f(d_i,h_i)) \qquad (6)$$

where (d_i, h_i), $i = 1, 2,..., n$, are the observed values of diameter and height, respectively. As some of the parameters are constrained ($0 < \alpha < 1$; $\sigma_{dj}, \sigma_{hj} > 0$; $-1 \leq \rho_j \leq 1$) it is advisable to use a constrained minimization procedure. If the software for this is not available then unconstrained minimization will also work if one uses starting values that are sufficiently close to the final parameter estimates. With a little practice it is not difficult to find suitable starting values by inspecting a scatterplot of the diameter-height observations.

2.3 Johnson's S_{BB} Distribution

The Johnson S_{BB} distribution (Johnson 1949b, Elderton and Johnson 1969) is a bivariate version of the S_B distribution (Johnson 1949a, see also Hafley and Schreuder 1976, 1977, Schreuder and Hafley 1977). It is based on a transformation of a single bivariate normal distribution and has density function given by

$$f(d,h) = \frac{J}{2\pi\sqrt{1-\rho^2}} \exp\left\{\frac{z_d^2 - 2\rho z_d z_h + z_h^2}{2\left(1-\rho^2\right)}\right\},$$

$$z_d = \gamma_d + \delta_d \log\left(\frac{d-\xi_d}{\lambda_d + \xi_d - d}\right),$$
$$z_h = \gamma_h + \delta_h \log\left(\frac{h-\xi_h}{\lambda_h + \xi_h - h}\right), \qquad (7)$$

$$J = \frac{\delta_d \, \lambda_d \, \delta_g \, \lambda_h}{(d-\xi_d)(\lambda_d + \xi_d - d)(h - \xi_h)(\lambda_h + \xi_h - h)}$$

where $\xi_d < d < \xi_d + \lambda_d$, $\xi_{\mathrm{h}} < \mathrm{h} < \xi_h + \lambda_h$, λ_d, δ_d, λ_h, $\delta_d > 0$ and $-1 \leq \rho \leq 1$.

The marginal distributions are Johnson S_B:

$$f(d) = \frac{\delta_d \lambda_d \exp\left(-z_d^2 / 2\right)}{\sqrt{2\pi}\left(d - \xi_d\right)\left(\lambda_d + \xi_d - d\right)}$$
$$f(h) = \frac{\delta_h d_h \exp\left(-z_h^2 / 2\right)}{\sqrt{2\pi}\left(h - \xi_h\right)\left(\lambda_h + \xi_h - h\right)} \qquad (8)$$

We note all the parameters in (8) are also present in (7). Thus once a bivariate S_{BB} distribution has been fitted the two S_B marginal distributions are immediately available. The conditional distributions $f(h|d)$ and $f(d|h)$ are not S_B but they can be computed using (7) and (8) in (5).

The maximum likelihood estimators of the parameters are not available in closed form but they can be computed by numerical minimization of (6) using the density (7). Numerical methods are needed even in the univariate case, the S_B (see Mønnes 1982, Siipilehto 1999). There are numerous constraints on the parameters and, unless software for constrained minimization is available, some trial and error experimentation with the starting values, step size and scale factors may be needed to achieve convergence. An alternative approach, outlined in Appendix 2, is to reparameterize the distribution in terms of *unconstrained* parameters.

3 Results and Evaluation of the Models

3.1 Results

The mixture of bivariate normals model and the Johnson S_{BB} model were each fitted to the observations described in 2.1 using the method of maximum likelihood. The S-PLUS functions "`nlmin`" and "`nlminb`" (Mathsoft 1995) were used to minimize expression (6) to compute the parameter estimates given in Tables 2 and 3.

Maximization of the likelihood for the S_{BB} distribution led to some numerical difficulties because, for these data, the maximum likelihood estimate of ξ_h is effectively “minus infinity” which leads to a limiting form of (S_B) marginal distribution for the heights. To achieve convergence it was found convenient to reparameterize the distribution as described in Appendix 2. We also fitted the S_{BB} using the additional parameter constraints ξ_d, $\xi_h \geq 0$ but this led to substantial deterioration in the fit – the negative log-likelihood increases from 7798.5 to 7848.8. The corresponding value for the mixture of bivariate normals model was 7678.5 which indicates a better fit than the S_{BB} model. However, this does not establish that the mixture model provides a good description of the data.

The contour plot of the fitted density and the observed values given in Fig. 3 shows that the model seems to capture the main features of the data but clearly it is preferable to have a somewhat more objective method to assess details of the fit.

The chi-squared and the Kolmogorov-Smirnov goodness-of-fit-tests, which are routinely applied in the context of univariate models, can be used to check the fit of the marginal distribu-

Table 2. Maximum likelihood estimates for the mixture of two bivariate normal distributions

Parameter	μ_{d_i}	μ_{h_i}	σ^2_{di}	σ^2_{hi}	ρ_i	α
Estimate (i = 1)	18.00	17.91	5.16	4.85	0.88	0.19
Estimate (i = 2)	39.06	27.27	9.69	2.90	0.63	

Table 3. Maximum likelihood estimates for the S_{BB} distribution

Parameter	ξ_d	λ_d	δ_d	γ_d	ξ_h	λ_h	δ_h	γ_h	ρ
Estimate	3.767	86.213	–1.408	0.901	–974.514	1027.804	5.723	–20.580	0.843

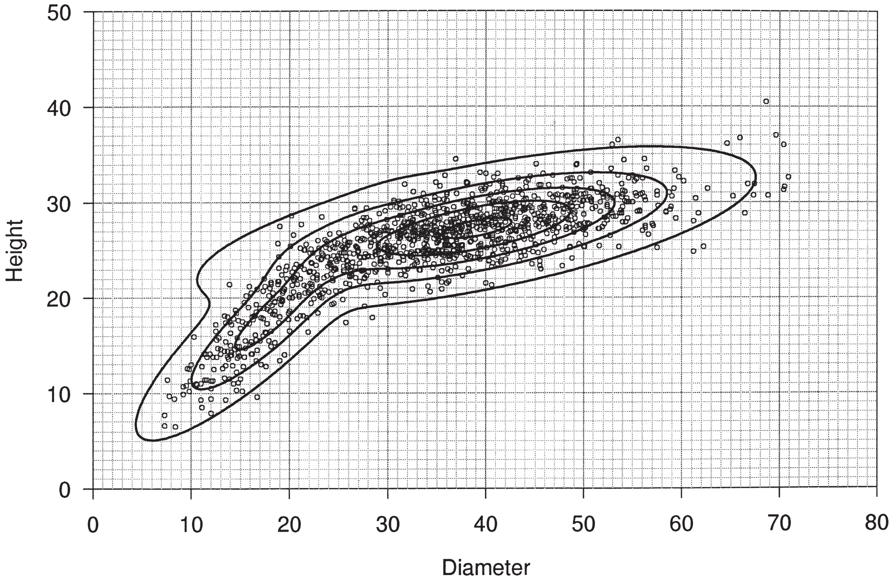


Fig. 3. Contour plot of the fitted density function for the diameters and heights. The observed values are shown as circles.

tions (see for example Reynolds et al. 1988). The chi-squared test is also applicable to bivariate distributions. One way to carry out the test is to partition the diameter-height plane by means of a rectangular grid, to count the number of observations in each rectangle (thus obtaining the "observed frequencies"), to compute the "expected frequencies" under the model, and then to carry out the well-known chi-squared test. The grid needs to be constructed in such a way that none of the expected frequencies are too small (five being generally regarded as large enough); otherwise the distribution of the test statistic is poorly approximated by the chi-squared and the test can be inaccurate. The main objection to this procedure, especially when it is applied to bivariate data, is that the value of the test-statistic depends strongly on how one partitions the plane, that is how one constructs the grid. Indeed it is not unusual to find that whereas one choice of grid leads to the rejection of the model an alternative grid does not.

An alternative, and somewhat more versatile tool for assessing the fit, is to use so-called pseudo-residuals (Zucchini and MacDonald 1999). These will now be defined for *univariate* continuous distributions and we will then illustrate how they can be used to assess the fit of bivariate distributions, such as the above diameter-height model.

3.2 Pseudo-Residuals

Suppose that the observations x_1, x_2,..., x_n are assumed to be independently distributed according to the probability density function $f(x)$ and denote the corresponding probability distribution function by $F(x)$. The issue here is to assess the fit of the model $f(x)$. The pseudo-residual, r_i, corresponding to the observation x_i is defined as

$$r_i = \Phi^{-1}(F(x_i)), \quad i = 1,2, ..., n$$

where Φ denotes the distribution function of the standard normal distribution:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Its inverse $\Phi^{-1}(p)$, where $0 < p < 1$, is the solution to the equation (in z): $p = \Phi(z)$. Most standard statistical software packages have a routine to compute these values – for example the S-PLUS function "`pnorm`" computes Φ and "`qnorm`" computes Φ^{-1}. Alternatively tables of the standard normal distribution can be used.

It is shown in the Appendix that, under the hypothesis that $f(x)$ is indeed the correct model for the observations, the pseudo-residuals r_i, $i =$ 1,2 ..., n, follow the standard normal distribution. Thus, no matter how complicated $f(x)$ might be, we can assess it's fit by computing the pseudo-residuals and checking how well they are fitted by a familiar $N(0,1)$ distribution. One of the most sensitive methods for doing this is to examine a "normal probability plot of the pseudo-residuals". This is constructed as follows:

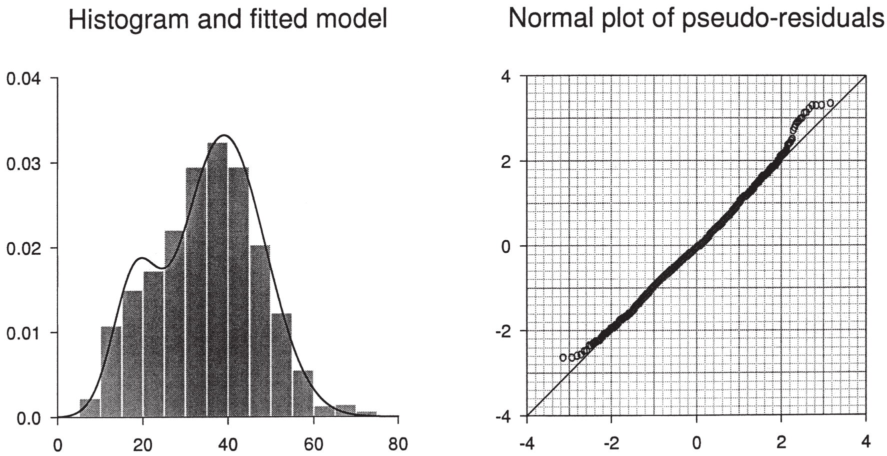
The pseudo-residuals are ordered from the smallest to largest. Let $r_{(i)}$ be the i-th largest value of r_1, r_2, ..., r_n. The $r_{(i)}$ are plotted against their so-called "plotting positions". The figures below are based on the plotting positions defined by

$$q_i = \Phi^{-1}\left(\frac{i}{n+1}\right), i = 1,2,...,n.$$

For alternative types of plots see, e.g., Chambers et al. (1983) or Hoaglin et al. (1983). The normal probability plot of the pseudo-residuals is a plot of the q_i (on the horizontal axis) against the $r_{(i)}$ (on the vertical axis).

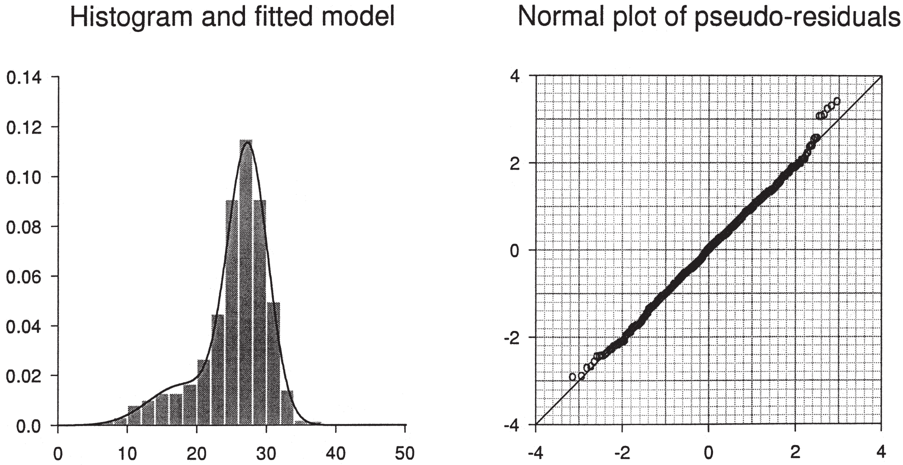
If the model fits the data then the points on this plot should lie close to the straight line of unit slope through the origin. Deviations from this line indicate lack of fit and also show which aspect of the model fails to fit the data.

For example the bivariate mixture distribution fitted to the diameter-height data implicitly models the marginal distribution of the diameter as a mixture of the two univariate normal distributions given in (3) and (4) with the parameter estimates given in Table 2. This density function together with a histogram of the diameters is shown in Fig. 4. Also shown is the normal probability plot of the corresponding pseudo-residuals. It can be seen that the fit is generally good but that about 14 trees (approximately 1% of the population) are somewhat larger than expected under the fitted model. These correspond to the


Histogram and fitted model

Normal plot of pseudo-residuals

Fig. 4. Histogram of diameters and the fitted marginal density (left); normal plot of the pseudo-residuals (right).


Histogram and fitted model

Normal plot of pseudo-residuals

Fig. 5. Histogram of heights and the fitted marginal density (left); normal plot of the pseudo-residuals (right).

two slightly raised boxes in the histogram over the intervals 65–70 and 70–75 [cm]. We note that in general the points on the extreme left and extreme right of a plot tend to be more variable than those in the middle even when the model is correct – they have a larger standard error.

Fig. 5 shows that the fitted marginal distribution of the heights marginal distribution fits unusually well. The 5 trees that are identified in the normal probability plot as being slightly higher than expected correspond to the small box in the histogram over the interval 36–38 [m].

We emphasise that the marginal distributions displayed in Fig. 4 and 5 were *not* fitted separately to the diameters and to the heights – they were derived from the fitted *bivariate* distribution thus making the accuracy of the fits to the two marginal distributions remarkable. By way of comparison Fig. 6 shows the normal probability plots of the pseudo-residuals for the marginal distributions of the diameters and heights associated with the S_{BB} distribution (7) with the parameter estimates given in Table 3. The lack of fit is evident.

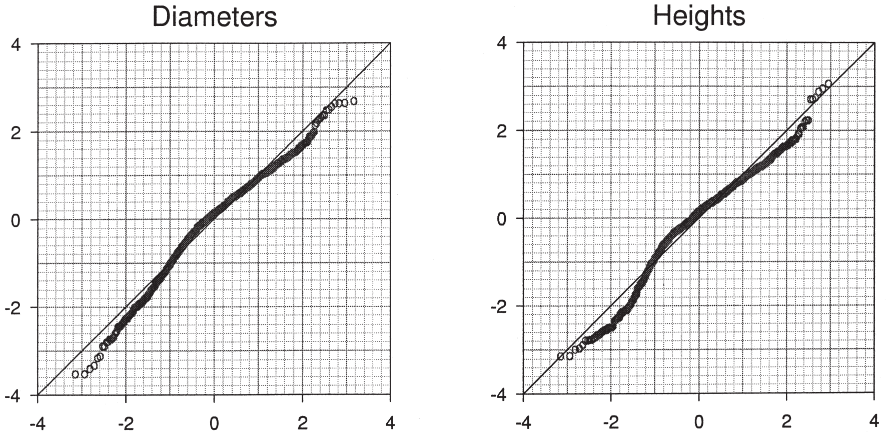


Fig. 6. Normal probability plot of the pseudo-residuals for the marginal S_B distribution of diameters (left) and of heights (right) obtained by fitting a bivariate S_{BB} model.

That the model fits the marginal distributions well is a necessary but not a sufficient condition that it describes the data as a whole well.

A difficulty that arises when assessing the fit of the conditional density, say $f(h|d)$ is that its parameters are a function of d. This is also the case in the familiar simple regression model based on a *single* bivariate normal distribution as opposed to the mixture of two such distributions as used here. However, in the simple regression case *only the conditional expectation* is a function of d; the conditional variance does not depend on d. (The term homoscedasticity is used to refer to this property.) Consequently the ordinary residuals – obtained by subtracting the observed heights from the regression line – are identically distributed. They can therefore be compared for the purposes of detecting outliers and for assessing the fit of the model.

This property does not hold for more complex bivariate distributions such as the Johnson S_{BB} or for the mixture model considered here. For such distributions the conditional mean, the conditional variance, and even the shape of the density function depends on d. In other words they are heteroscedastic. Consequently the ordinary residuals all have different distributions; they are not comparable and thus cannot be used to assess the fit. For example, it would not make sense to construct a histogram of these residuals because there is no single density with which it can be matched. Pseudo-residuals were developed to overcome this problem, they are comparable because they all have the same distribution, namely the standard normal.

Fig. 7 shows a histogram of the pseudo-residuals for the conditional distribution of height given diameter derived from the fitted bivariate model. Under the model these should have a standard normal distribution (whose density function is also shown in the figure). This, together with the normal probability plot also given in Fig. 7, confirms that this conditional distribution is fitted very well.

Pseudo-residuals also enable one to assess the fit of a bivariate model in additional detail. For example we can examine the fit of its projections onto lines other than the horizontal axis (marginal distribution of the diameter) or the vertical axis (marginal distribution of height). This enables us to examine the fit from additional points of view, thereby make it possible to detect outliers or lack of fit that might not be less evident if one only considers the usual marginal distributions. Such projections are illustrated in Figs 8 and 9.

Fig. 8 shows a plot of the standardized diameter and height data, that is after subtracting their respective means and then dividing by their respective standard deviations. It also shows a number of straight lines at various angles through the origin. By projecting the points onto one of

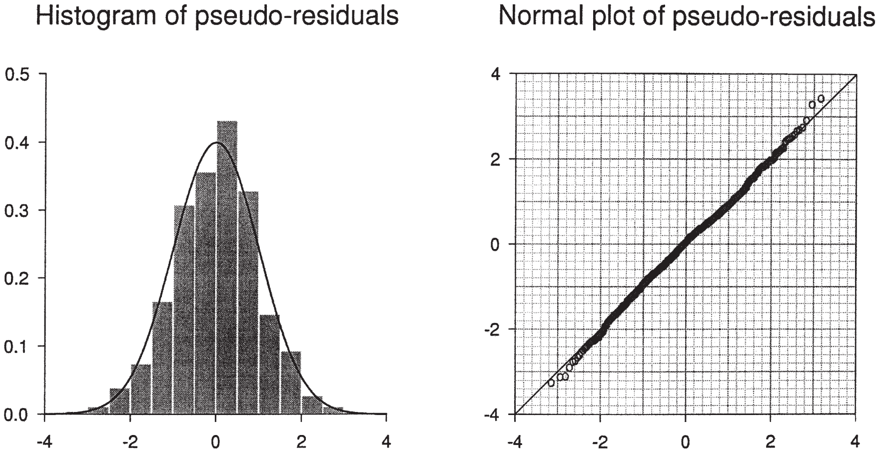


Fig. 7. Histogram of the pseudo-residuals and the standard normal density (left); normal probability plot of the pseudo-residuals (right) for the conditional distribution of heights given diameters.

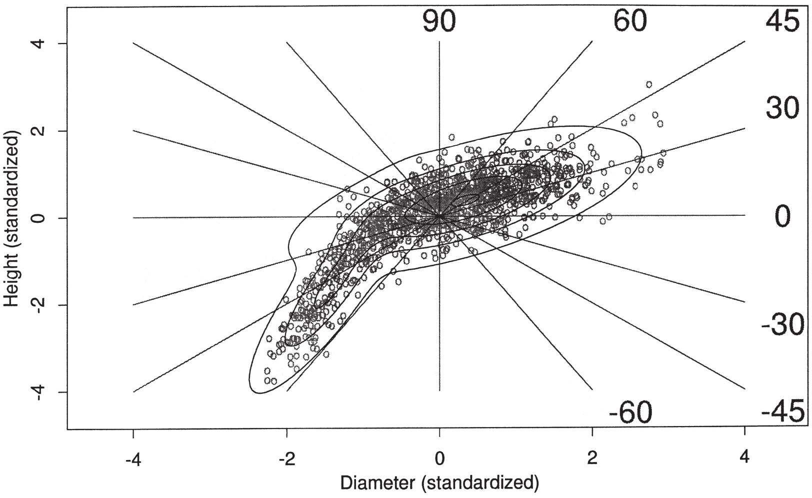


Fig. 8. Contour plot of the fitted bivariate density function for the standardized diameters and heights and selected lines through the origin.

these lines (for example that at 45%) we obtain the observed values of the marginal distribution of a linear function of the original observations: $\ell_i = \alpha_0 + \alpha_1 d_i + \alpha_2 h_i$, $i = 1,2, ...,n$, where α_0, α_1 and α_2 are determined by the particular projection that we wish to examine, determined by the angles shown in Fig. 9. We can now compare the distribution of the ℓ_i with that implied by the model fitted to the original data.

The normal plots of pseudo-residuals associated with the ℓ_i enable us to assess the fit of the model from a variety of different angles. Such plots are shown in Fig. 9. (The plots for 0° and for 90° have already been given in Fig. 4 and 5, respectively.)

The plots in Fig. 9 show that, on the whole, the model fits these data very well. (Compare these plots with those in Fig. 6.) The projection at the

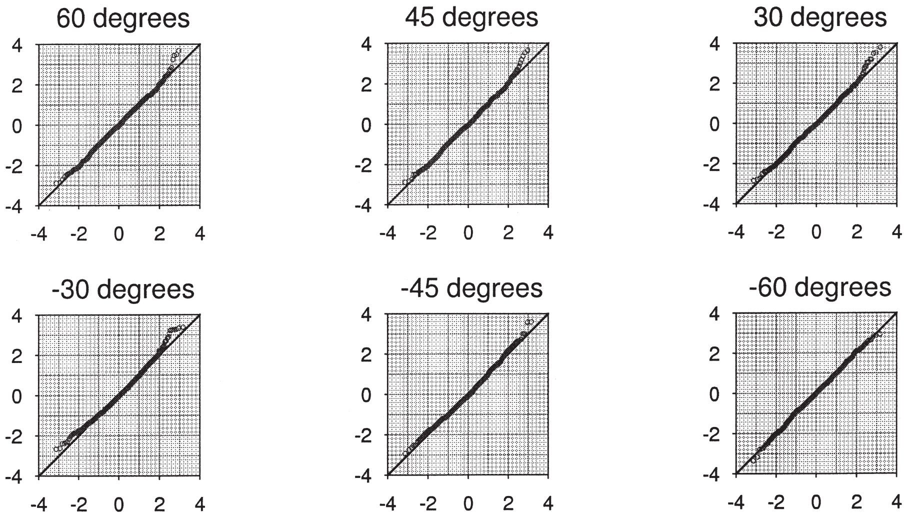


Fig. 9. Normal probability plots of pseudo-residuals for the marginal distributions of selected linear combinations of the diameters and heights (cf. Fig. 8).

angle –30° shows a slight lack of fit in the lower tail of the distribution. The plots reveal the 14 exceptionally large trees that were evident in the ordinary marginal distribution of diameter in Fig. 4. These show up clearly on the upper right-hand corners of the plots for the projections corresponding to the angles between 60° and –30°. Regarding the 14 exceptionally large trees the data were derived from a systematic sample with plots spread over a large area, covering several hectares, which includes sections exhibiting different tree dimensions. Two plots are situated in small section where the trees are much larger than the rest. Although these exceptions show up prominently on the plots, it needs to be kept in mind that they represent about 1% of the points on each plot, the remaining 99% (1228 points) fall either on, or very close to the main diagonal.

4 Discussion

The type of model discussed in this paper is appropriate in situations where there is reason to suppose that the population under consideration comprises two subpopulations. This was the case for the example considered here and it was found that a mixture of two bivariate normal distributions fits the data unusually well. The fitted model is easily interpretable and reflects something real about the population, namely that it comprises two distinct subpopulations whose diameter-height distributions differ. The larger one (approximately 80% of the trees) has a different diameter-height relationship to the smaller (approximately 20% of the trees). In the smaller subpopulation the diameter-height relationship is steeper. This finding would lead one to expect that the distribution of other tree attributes of this forest (for example maximum crown width, height to crown base or dead branch height) could also be appropriately modelled by a mixture of distributions.

Of course it is also possible to improve the fit of such models by using a mixture of three or more distributions instead of just two. However, unless there are persuasive biological grounds for believing that the population comprises three or more distinct subpopulations the parameters of the fitted model lose their simple interpretability. Furthermore the resulting increase in the number of parameters erodes the precision with which it is possible to estimate them.

For special investigations like structural analysis in natural and virgin forests it is generally no problem to obtain funds for measuring a sufficient number of heights to fit a bivariate diameter-height distribution. However, in regular management inventories, foresters cannot afford to measure more than say 20–30 heights per beech stand. Thus, if the application of bivariate techniques is to become common practice in forest inventories, it will be necessary to develop methods for estimating the parameters of the distribution from a relatively small number of measured trees (Schmidt and Gadow 1999). Such sample sizes are too small to fit the mixture of two bivariate normal distributions outlined here – one needs about 50 observations to obtain useable estimates.

One approach might be to investigate whether some of the parameters of the model might be estimated by relating them to available data, such as dominant height or basal area. A second approach, currently under investigation, is to develop a method of estimation from larger samples of approximate height measurements that would be less expensive to make than precise measurements.

The second issue discussed in this paper was the use of pseudo-residuals for assessing the fit of a model.

A convenient feature of pseudo-residuals is that one only needs to think of them as observations from a standard normal distribution – if the model fits. This applies irrespective of whether one is examining the fit of a marginal distribution, a conditional distribution, the distribution of a linear combination, or some more complex function, of the variables. Thus, for example, histograms of pseudo-residuals are always compared with the familiar bell-shaped $N(0,1)$ density function; normal probability plots of the pseudo-residuals are always plotted on the same standard scale.

Finally, the fit of the model can be assessed by simply examining plots. Of course such plots do not replace formal goodness-of-fit tests; they should be regarded as additional tools that are easy to apply and that can reveal not only whether or not a model fits, but also details regarding the lack-of-fit.

The above examples illustrate just some of the ways in which pseudo-residuals can be used to assess the fit of a particular bivariate model. Of course they can be applied to any of the other bivariate models that are of interest in a forestry context. Bivariate distributions have been used for some time to describe the relationship between diameters and heights (Hafley and Schreuder 1976, Warren et al. 1979, Tewari and Gadow 1999) and more recently also to include other tree attributes, such as height to maximum crown width or height to crown base (Uusitalo and Kivinen 1998). More exact estimation of height of maximum crown width, crown base and dead branch height can improve the estimation of branchiness and wood quality (Collin and Houllier 1992, Maguire et al. 1994, Seifert 1999). Another application of a bivariate distribution is to model height and density of natural regeneration in stands (Schweiger et al. 1997). Growth models are an important tool for long- and medium-term simulation studies and Knoebel and Burkhart (1991), for example, presented an approach for modelling the bivariate distribution of diameter at two points in time. A related application involves the simultaneously estimation of mechanical wood properties (Pearson 1980, Warren et al. 1979) and the estimation of concomitant strength wood properties related to known nondestructive wood properties (Pellicane 1993).

Acknowledgements

The data were provided by the Forest Research Station of Lower Saxony. The constructive suggestions of two anonymous referees are gratefully acknowledged.

References

Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. 1983. Graphical methods for data analysis. Wadsworth, Belmont, California.

Clutter, J.L. & Allison, B.J. 1974. A growth and yield model for Pinus radiata in New Zealand. In: Growth Models for Tree and Stand Simulation. J.Fries (ed), Royal College of Forestry, Stockholm, Sweden. Department of Forest Yield Research Note 30: 136–160.

Colin, F. & Houllier, F. 1992. Branchiness of Norway spruce in north-eastern France: predicting the main crown characteristics from usual tree measurements. Annales-des-Sciences-Forestieres 49: 511–538.

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

Droessler, T.D. & Burk, T.E. 1989. A test of nonparametric smoothing of diameter distributions. Scandinavian Journal of Forestry Research 4: 407–415.

Elderton, W.P. & Johnson, N.L. 1969. Systems of frequency curves. Cambridge University Press.

Gadow, K. v. 1987. Untersuchungen zur Konstruktion von Wuchsmodellen für schnellwüchsige. Plantagenbaumarten. Forstliche Forschungsberichte, University Munich 77: 147 p.

Gaffrey, D. 1996. Sortenorientiertes Bestandeswachstum-Simulationsmodell auf der Basis intraspezifischen, konkurrenzbedingten Einzelbaumwachstums – insbesondere hinsichtlich des Durchmessers – am Beispiel der Douglasie. In: Berichte des Forschungszentrums Waldökosysteme: Reihe A, 133: 413 p.

Hafley, W.L. & Schreuder, H.T. 1976. Some non-normal bivariate distributions and their potential for forestry applications. In: Proceedings of the XVI IUFRO World Congress Division VI (Oslo, Norway, 1976), J. Parde and L. Strand (coordinators), 104–114.

— 1977. Statistical distributions for fitting diameter and height data in even-aged stands. Canadian Journal of Forest Research 7: 481–487.

Hoaglin, D.C., Mosteller, F. & Tukey, J.W. 1983. Understanding robust and exploratory data analysis. Wiley, New York.

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

— 1949b. Bivariate distributions based on simple translation systems. Biometrika 36: 297–304.

Knoebel, B.R. & Burkhart, H.E. 1991. A bivariate distribution approach to modelling forest diameter distributions at two points of time. Biometrics 47: 241–253.

Korpel, S. 1992. Ergebnisse der Urwaldforschung für die Waldwirtschaft im Buchen-Ökosystem. In Allgemeine Forstzeitschrift 1992, 47(21): 1148–1152.

Košir, Z. 1966. Beitrag zur Erforschung der Urwaldstruktur reiner Buchenwälder. In: Gesellschaftsmorphologie, Ber. Int. Symp. Int. Ver. Veg.-Kunde, Rinteln. 4.–7.4. 1966. R. Tüxen (ed.) 1970.

Maguire, D.A., Moeur, M. & Bennett, W.S. 1994. Models for describing basal diameter and vertical distribution of primary branches in young Douglas-Fir. Forest Ecology and Management 63: 23–55.

Mathsoft 1995. S-PLUS Guide to statistical and mathematical analysis, Version 3.3, StatSci Division, Mathsoft, Inc., Seattle, Washington.

Mønnes, E.N. 1982. Diameter distributions and height curves in even-aged stands for Pinus sylvestris, L. Meddelelser fra Norsk Institutt General for Skogforsking 36(15): 1–43.

Pearson, R.G. 1980. Potential of the S_B and S_{BB} distributions for describing mechanical properties of lumber. Wood-and-Fiber. 1980, 12(4): 244–254.

Pellicane, P.J. 1993. Application of the S_B distribution to the prediction of concomitant wood properties. Wood Science and Technology 27(3): 161–172.

Reynolds, M.R.Jr., Burk, T.E. & Huang, W.-C. 1988. Goodness-of-fit tests and model selection procedures for diameter distribution models. Forest Sciences 34 (2): 373–399.

Saborowski, J. 1994. Schätzung von Durchmesserverteilungen. Vortrag beim Kolloquium der Sektion Forstl. Biometrie u. Informatik im Deutschen Verband Forstl. Forschungsanstalten in Ljubljana, 1994: 153–163.

Schmidt, A. 1967. Der rechnerische Ausgleich von Bestandeshöhenkurven. Forstwissenschaftliches Centralblatt 86: 370–382.

Schmidt, M. & Gadow, K. v. 1999. Baumhöhenschätzung mit Hilfe der bivariaten Johnson's S_{BB}-Funktion. Forstwissenschaftliches Centralblatt 118: 355–367.

Shannon, C.E. 1948. The mathematical theory of communication. In: Shannon, C.E. & Weaver, W. (ed.). The mathematical theory of communication. Urbana, Univ. Illinois Press: 3–91.

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

Schweiger, J., Sterba, H., Skovsgaard, J.P. & Vanclay, J.K. 1997. A model describing natural regeneration recruitment of Norway spruce (Picea abies (L.) Karst.) in Austria. In: Selected and edited papers from the IUFRO conference on modelling regeneration success and early growth of forest stands, Copenhagen, Denmark, 10–13 June 1996. Forest Ecology and Management 97(2): 107–118.

Seifert, T. 1999. Modelling wood quality of Norway spruce (Picea abies) depending on silvicultural treatment. In: Proceedings of the Third IUFRO-

Workshop on the Connection between Silviculture and Wood Quality through Modelling Approaches and Simulation Softwares. Nepveu, G. (ed.), 1999: La Londe les Maures, France.

Siipilehto, J. 1999. Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fennica 33(4): 281–301.

Silverman, B.W. 1986. Density estimation for statistics and data analysis. Chapman and Hall, New York, 175 p.

Tewari, V.P. & Gadow, K. v. 1999. Modelling the relationship between tree diameters and heights using S_{BB} distribution. Forest Ecology and Management 119: 171–176.

Uusitalo, J. & Kivinen, V.P. 1998. Constructing bivariate dbh/dead-branch height distribution of pines for use in sawing production planning. Scandinavian Journal of Forest Research 13(4): 509–514.

Warren, W.G., Ord, J.K., Patil, G.P. & Taillie, C. 1979. Some recent development relating to statistical distributions in forestry and forest products research. Statistical distributions in ecological work. 1979, recd. 1984: 247–260.

Weber, J. 1998. Beschreibung der Diversität von Bestandesstrukturen mit Hilfe von Teildiversitäten. Tagung des Deutschen Vereins Forstlicher Forschungsanstalten – Sektion Forstliche Biometrie und Informatik – und der AG Ökologie der Internationalen Biometrischen Gesellschaft in Freiburg: 25–37.

Zucchini, W. & MacDonald, I.L. 1999. Illustrations of the use of pseudo-residuals in assessing the fit of a model. In: Proceedings of the 14th International Workshop on Statistical Modelling, 19–23 July 1999, Graz. Friedl, H., Berghold, A. & Kauermann, G. (ed.): 409–416.

Total of 36 references

Appendix 1. Proofs of Claims Regarding Pseudo-Residuals.

We first show that if a continuous random variable X has distribution function $F(x)$ then the corresponding pseudo-residual $r = \Phi^{-1}(F(X))$ is distributed as $N(0,1)$.

We note that since X is a continuous random variable its distribution function $F(x)$ is monotone strictly increasing and therefore has an inverse $F^{-1}(x)$ which is also monotone strictly increasing. The distribution function Φ of a standard normal random variable and its inverse Φ^{-1} also have these properties.

The random variable $U = F(X)$ is uniformly distributed on the interval (0,1). This follows from

$$\begin{aligned} P\{U \le u\} &= P\{F(X) \le u\} = P\{X \le F^{-1}(u)\} \\ &= F(F^{-1}(u)) = u, \quad \text{for } 0 < u < 1, \end{aligned}$$

which is the distribution function of a uniformly distributed random variable.

The distribution function of the pseudo-residual, r, is given by

$$\begin{aligned} P\{r \le z\} &= P\{\Phi^{-1}(F(X)) \le z\} \\ &= P\{F(X) \le \Phi(z)\} \\ &= P\{U \le \Phi(z)\} = \Phi(z), \end{aligned}$$

which is the distribution function of a $N(0,1)$ random variable.

Suppose now that $X_1, X_2, ..., X_n$ are independently distributed and that distribution function of X_i is $F_i(x)$, $i = 1,2, ..., n$. (Note that the X_i are *not assumed to be identically distributed.*) Then, by the result above, *each* pseudo-residual $r_i = \Phi^{-1}(F_i(X_i))$, $i = 1,2, ..., n$, is distributed as $N(0,1)$. Finally, as the X_i are independently distributed, the r_i, are also independently distributed.

Appendix 2. A Reparameterization of the Johnson S_{BB} Distribution for Unconstrained Maximization of the Likelihood Function.

The nine parameters of the S_{BB} are subject to several constraints (see Section 2.3) which need to be respected in the numerical minimization of the negative log-likelihood function. If software for constrained numerical minimization is not available it is convenient to (temporarily) express the likelihood function in terms of different parameters that are unconstrained. Such a reparameterization and its inverse is given below.

New parameters	Original parameters
$\theta_1 = \log(d_{\min} - \xi_d)$	$\xi_d = d_{\min} - \exp(\theta_1)$
$\theta_2 = \log(\xi_d + \lambda_d - d_{\max})$	$\lambda_d = d_{\max} - \xi_d + \exp(\theta_2)$
$\theta_3 = \log(\delta_d)$	$\delta_d = \exp(\theta_3)$
$\theta_4 = \gamma_d$	$\gamma_d = \theta_4$
$\theta_5 = \log(h_{\min} - \xi_h)$	$\xi_h = h_{\min} - \exp(\theta_5)$
$\theta_6 = \log(\xi_h + \lambda_h - h_{\max})$	$\lambda_h = h_{\max} - \xi_h + \exp(\theta_6)$
$\theta_7 = \log(\delta_h)$	$\delta_h = \exp(\theta_7)$
$\theta_8 = \gamma_h$	$\gamma_h = \theta_8$
$\theta_9 = \mathrm{tan}h^{-1}(\rho)$	$\rho = \mathrm{tan}h(\theta_9)$

Here $d_{\min}$, $d_{\max}$, $h_{\min}$, $h_{\max}$ denote the smallest and largest of the observed diameters and heights, respectively.

If one wishes to impose the additional constraints $\xi_d > 0$ and $\xi_h > 0$ this can be achieved by replacing θ_1 and θ_5 in the above table with the following:

New parameters	Original parameters
$\theta_1 = \log(\xi_d \,/\, (d_{\min} - \xi_d))$	$\xi_d = d_{\min} \cdot \exp(\theta_1) \,/\, (1 + \exp(\theta_1))$
$\theta_5 = \log(\xi_h \,/\, (h_{\min} - \xi_h))$	$\xi_h = h_{\min} \cdot \exp(\theta_5) \,/\, (1 + \exp(\theta_5))$

Of course these transformations can also be applied for estimating the parameters in the univariate case, that is to fit a S_B distribution. For the data considered in this paper, and for the particular software that we used, namely the S-PLUS functions “`nlmin`” and “`nlminb`” (Mathsoft 1995), unconstrained maximization of the reparameterized likelihood required fewer iterations than did constrained maximization with respect to the original parameters.