The Power-Normal Distribution: Application to forest stands.

Erik Mønness

Hedmark University College, ØSIR

P.O. Box 104, 2450 Rena, Norway

Phone: +47 62430598
Fax: +47 62430500
e-mail: erik.monness@hihm.no

THIS IS A COPY OF THE PUBLISHED ARTICLE Published on the web 08 April 2011.in

Canadian Journal of Forest Research, 2011, 41:(4) 707-714, 10.1139/x10-246

(There might be minor text differences.)

Abstract

The Power-Normal distribution (PN) originates from the inverse Box-Cox transformation and is used in this study to fit the frequency distributions of tree diameter and height. PN is flexible in describing different shapes of observed distributions as indicated by the certain areas in the skewness x kurtosis shape plane. The estimation of the parameters using maximum likelihood is straightforward and the resulting numerical properties are desirable. The shapes achieved by PN are very diverse, even though only three parameters are used. In forestry, Johnson's System Bounded distribution (S_B) has been widely used to fit tree diameter distributions, but it is often susceptible to numerical problems when fitted by maximum likelihood estimation. Our results indicate that the performance of PN is superior to Johnson's S_B, as shown by the Kolmogorov-Smirnov statistic and visual inspection, particularly for fitting the tree height distributions.

Key words: Power-Normal distribution; Diameter and height distributions; Johnson's System bounded distribution; Maximum likelihood estimation; Percentiles.

Introduction

A major task in forestry is to predict the distribution of diameters or heights of a forest stand. The consideration of the possible shapes of a theoretical distribution was addressed by Hafley and Schreuder (1977) who introduced the Johnson System of distributions (Johnson 1949) as well as the Weibull and the Gamma distributions into forestry. Prediction functions that use a maximum likelihood (ML) estimation of the Johnson's System Bounded (S_B) distribution were proposed by Mønness (1982). Some of the issues involved in estimation using S_B were discussed in Lambert (1970), who introduced an improved parametrization, Siekierski (1992), and Rennolls and Wang (2005). The Weibull distribution was reconsidered by Maltamo et al. (2000) and Merganič and Sterba (2006). A new distribution, the logit-logistic, was introduced by Wang and Rennolls (2005). This paper introduces the Power-Normal distribution (PN) into forestry. Properties of PN are explored, and are compared with the Johnson's S_B.

PN has its origins in the Box-Cox transformation (Box and Cox 1964). The idea behind the Box-Cox transformation is to apply a non-linear transformation to normalize the data before the actual analysis takes place. This procedure has now gained wide acceptance and is discussed in many text books on applied statistics e.g. Box et al. (2005). The Box-Cox procedure assumes that the data distribution can be transformed to the normal distribution (Hernandez and Johnson 1980). Thus the inverse Box-Cox transformation can give rise to a family of distributions (PN), shown by Goto and Inoue (1980) and Freeman and Modarres (2006). With respect to forestry applications, Garcia (1983) has described the application of Box-Cox transformation to forestry growth curves.

Methods

The Johnson Distributions

The Johnson's distribution system consists of three non-linear transformations of a normal variate that cover the entire skewness x kurtosis space of shapes. Johnson himself referred to these three transformations as System bounded (S_B), System lognormal (S_L), and system unbounded (S_U). Lambert (1970) introduced a statistically improved parameterization of S_B that was revisited by Rennolls and Wang (2005) and will be used herein.

Let *Z* be a standard normal variate and *X* be the observed data. S_B is represented by the non-linear transformation:

[1] $$Z = \frac{\log\left(\frac{X-\tau}{\theta-X}\right)-\mu}{\sigma}$$

where (τ,θ) are the lower and upper bounds on the *X* scale, whereas (μ,σ) are the expectation and standard deviation on the Z scale.

Estimating S_B by ML is rather complex. However, once (τ,θ) have been found, (μ,σ) are readily available. The S_B log-likelihood surface as a function of (τ,θ) is rather flat around the maximizing value, which can result in problems of convergence.

Numerical convergence problems with S_B ML estimation were found here in some cases as also reported by Siekierski (1992). It can be seen that fixing the lower bound τ improves the numerical convergence. In this study, the lower bound of the parameter τ of Johnson's S_B was set to zero (τ = 0) for the diameter distributions, while the lower bound of the parameter τ of Johnson's S_B for the height distributions was computed as follows $\tau = H_{min} - \left(H_{max} - H_{min}\right)/\sqrt{n}$ where H_{min} is the minimum height and H_{max} is the maximum height.

The Power-Normal distribution

The Box-Cox transformation is applicable to positive data, i.e. *X≥0.*

[2] $$Z = \frac{\frac{X^{\lambda}-1}{\lambda}-\mu}{\sigma}$$ when λ≠0

[3] $$Z = \frac{\log(X)-\mu}{\sigma}$$ when λ=0

The log() case [3] is identical to the Johnson's S_L so both SP and S_B has S_L as a limiting case. Transformation [2] is always possible, but Z can only be N(0,1) when λ=1 and in the log() case. However, Z is a truncated normal (Freeman and Modarres 2006).

The truncation point on the z scale, when x=0 is

[4] $$k = \frac{\lambda\mu+1}{\sigma\lambda} = \frac{\mu}{\sigma} + \frac{1}{\sigma\lambda}$$

The truncation is to the left or right of k depending on the sign of λ. Define $K_k = \Phi\left(\text{sign}(\lambda)k\right)$, where $\Phi(\ )$ is the cumulative distribution function of the standard normal.
The cumulative distribution function of PN is

[5] $$F(x) = \frac{1}{K_k}\Phi\left(\frac{x^{\lambda}-(\lambda\mu+1)}{\sigma\lambda}\right) \qquad , x \geq 0$$

and the probability density function of PN is

[6] $f\left(x\right)=\frac{x^{\lambda-1}}{K_k\sqrt{2\pi\sigma}}e^{-\frac{1}{2}\left(\frac{x^{\lambda}-(\lambda\mu+1)}{\sigma\lambda}\right)^2} \quad , \ x\geq 0$.

When λ>0, K_k is typically close to 1 (often in real data μ>>σ so k is large). When λ<0, but close to zero, K_k is still close to 1. If λ<<0, K_k becomes small and the truncation has an influence. However, large absolute values of λ are of little practical value because then the transformation tends towards a vertical and a horizontal line that have a junction at $\left(1,-\frac{\mu}{\sigma}\right)$ in the *X*Z* plane. Both PN and S_B have the log-normal and the normal distribution as their limiting cases.

The shape of the density function is given by λ and k (Goto and Inoue 1980). If λ≤0 or λ≥1 the density is unimodal. A distribution with $k\leq 2\sqrt{\lambda^{-1}-1}$ & λ>0 has a density with a maximum at x=0. If $k>2\sqrt{\lambda^{-1}-1}$ & λ<1 the density has a local minimum close to x=0, but the size of the local minimum is negligible with most k found in practice and the shape appears to be unimodal. A plot of λ x k with the data in use is shown later in Figure 6.

The shape of PN is herein evaluated in terms of its skewness x kurtosis, and compared with S_B in the same space. Goto and Inoue (1980) provide some formulas for the skewness and kurtosis of the PN. Freeman and Modarres (2006) provide explicit formulas for some special cases, but these formulas are rather complicated. Herein, simulated skewness and kurtosis values are used as evidence of the capabilities of PN.

Five independent draws of 1000 independent random N(0,1) values were carried out (*Z* values). For each set of draws, *λ* was varied from -3 to 3 in steps of 0.1, *μ* was varied from -1 to 100 in steps of 2, σ was varied from 0.1 to 20 in steps of 2 (σ cannot be zero). Only X values with $X=\left(1+\lambda\left(Z\sigma+\mu\right)\right)>0$ were included in the "observed data set of X values". This represented the truncation effect. The skewness and kurtosis of the X data were then calculated.

The PN showed both left and right skewness (see Figure 1). There are three cases to be discussed:

1) $\lambda\in\left(-3,0\right)$. Theoretically, skewness and kurtosis do not always exist: "*No moments exist for any order more than* $\left|\lambda\right|$" (Goto and Inoue 1980), although estimates do always exist. The truncation is a critical factor in this case because most draws yield a negative out-of range value. Some of the estimated skewness x kurtosis points are in the "theoretically impossible region", above the upper lines. However, the points are statistics, not theoretical values. These irregular points typically arise from data sets that have very few data points (less than 10 out of 1000, due to the truncation).

2) $\lambda\in\left(0,1\right)$: The shapes are in a well defined area with right skewness around the log-normal line.

3) $\lambda\in\left(1,3\right)$. The points indicating right skewness in Figure 1 occur when ((λ<≈ 1.5) & (k<≈ 2.0)), otherwise the distribution is skewed to the left. Generally, in practice, k is large, thus $\lambda\in\left(1,3\right)$ will usually consist of left skewed distributions within a well defined area.

The negative values of μ will cause partial reversal of the plots shown. A specific skewness x kurtosis point may be achieved using different combinations of the three parameters.

The PN transformation [2] is continuous in λ thus estimation involves only one functional form. If the estimated value of λ is close to zero, the log-normal form [3] may be used instead. Typically, K_k is ignored in the estimation step. A ML procedure is described in Madansky (1988). The estimation is well established and has good numerical properties.

The statistical software SAS (2008) was used for programming and Systat Software (2004) for graphics. Graphs were enhanced using a metafile program (Companion-Software 2008)

Data

The data sets were obtained from 139 young stands in South East Norway, which comprised both Scots Pine and Norway Spruce. The fields were established in 1954 and thereafter. The diameter and height of each tree in the stands (16984 in total) were measured. The data are described in Vestjordet (1977) (In Norwegian, with a summary in English). The mean size of the plots was 420 m^2 for Scots Pine and 370 m^2 for Norway Spruce. The elevation (the sites height above sea level) varied from 25 to 510 m. Some stands were located on or near the coast, whereas others were located further inland. They were not intended to be a representative sample of young forests in southern Norway. The reasons for this were: 1) the mobility of researchers was low at the time of the study, and 2) usable areas of even-aged young forest were concentrated in a few locations because clear-cutting was not common in Norway at the time the stands were established. On the other hand, this was at the time considered a benefit, because several stands in the same area could be considered as replicates. The stands were established originally to explore the effects of pre-commercial thinning (via an early regulation of spacing, which was designed to be carried out before the stand had achieved a mean height of 5 m). Both un-thinned and thinned stands are included in the data. In the thinned stands, a regime was in place under which: 1) the remaining trees should be spaced evenly where possible; 2) the arithmetic heights of stands in the same area should have a small variation; 3) the height distribution within a stand should be small, and the canopy should be smooth; 4) deciduous trees should be removed; 5) the remaining trees should be of good quality; and 6) the mean height should be as high as possible. A summary of the stand data is given in Table 1. The range of the data values is shown in Figure 2. Histograms of the data (Some examples are shown in Figure 7) showed that the stands had right- or left-skewed unimodal distributions of diameter and height, respectively.

Results

The skewness x kurtosis of the original data, after the PN and S_B transformations, is depicted in Figure 3 . Each cross represents a stand, whereas the squares represent those stands where the S_B ML estimation did not converge properly. The most common reason for S_B non-convergence was that the upper bound θ iterated below H_{max} or D_{max} . In those cases the initial values of the upper bound was used.

Stands with skewness x kurtosis points close to normal (0,0) on the transformed scale had estimated parameters that enabled the shape to be modeled on the original scale. PN was successful in transforming to near normality in all cases because the skewness x kurtosis points were concentrated around (0,0). S_B was similarly successful in most cases when the ML estimation converged. There were S_B ML convergence problems with respect to diameters in 27 stands and with respect to heights in 79 stands. (If both upper and lower bound were to be estimated, there were convergence problems with respect to diameter in 84 stands. For heights, the equivalent number was 101).

The Kolmogorov–Smirnov distance statistic between the data and the estimated distributions is shown in Figure 4; PN and S_B are shown on the horizontal and vertical axes, respectively. Thus, both the actual values and a comparison can be seen. If a given point lies above the diagonal, PN provides a better fit than S_B. The curve is a LOWESS regression. Most Kolmogorov–Smirnov statistic values were between 0.025 and 0.075. PN appears to be a better distribution, particularly for heights.

There is a slight tendency that S_B underestimates the mean of the distribution (by about 0.5 %), and in some lesser degree, PN overestimates the mean with about 0.5 %. See Figure 5. But for most cases, both estimated distributions have a mean very close to the true mean. (The means of the estimated distribution is approximated by the sum of value x probability)

Figure 6 shows the estimated PN k*λ values for the data. No stands have a $k \le 2\sqrt{\lambda^{-1} - 1}$ or a λ estimate less than zero. The points between the curve and the vertical line λ=1 have densities with a local minimum close to X=0, but the size of the local minimum was negligible with these data. From Figure 6, it may also be seen that the truncation point k was greater than 2 in all cases. The median k values were 3.8 (for diameter) and 3.2 (for height). Thus the distribution truncation due to K_k was always less than $1 - \Phi(2) \approx 0.02$ and usually much smaller. The case ((λ<≈ 1.5) & (k<≈ 2.0)) did not appear: Thus with λ <1, the densities are right skewed, and with λ>1 they are left skewed. Most of the diameter distributions are right skewed and most of the height distributions are left skewed.

Figure 7 shows both the actual data distribution (histogram), the estimated PN distribution (solid line) and the estimated S_B (dashed line) for a selection of stands. In general (although not shown apart from Figure 4) PN and S_B provided a good fit for all the stands analyzed. The S_B fit is reasonable even in the cases where the initial upper border was in use due to non-convergence. The case (Height, Stand=55) shows that PN could be left skewed even if the data is not close to zero. An interesting feature is that the PN is always more spiky than the S_B. This is the case for all stands, both diameter and height (only examples are shown).

Discussion

The discussion is about shape, estimation and suggestions for further work.

Shape.

A theoretical distribution should be flexible enough to model the observed set of tree distribution shapes. Both S_B and PN fulfill this requirement. For S_B, there is a theoretical basis for this statement; however, only simulated evidence for PN has been provided herein. PN seems to yield a smaller range of shapes than S_B. PN supports shapes that lie below the log-normal line, whereas S_B does not. PN has been shown to achieve left skewness even if the data are at some distance from zero. An interesting feature of PN is that it covers an area of the skewness x kurtosis space using only three parameters. Other three parameter distributions like the Weibull and the Log-normal only covers a line.

Estimation.

PN works without the estimation of an upper bound, a convenient property, whereas S_B is bounded both from below and above. PN seems to be slightly better than S_B and have better numerical properties. Provided that convergence is achieved, S_B also fits the data quite well. For our data, the number of cases where S_B failed to converge was larger than expected from earlier work with S_B. Given that many stands were rather similar, non-convergence for one stand could imply non-convergence for similar stands. Also, these stands are young and dense. However, as seen by the plotted histograms, S_B seems to yield a reasonable fit even when using elementary (initial) estimates of (τ,θ).

Further work.
A typical task in forestry is, in a subsequent stage, to use the estimated distribution as a dependent variable in a regression, with the characteristics of stands being the independent variables. Typical independent characteristics are site index, age, density, location/climate, and basal area mean diameter. (Basal area mean diameter, $\sqrt{E(D^2)}$, is often considered to be an independent characteristic, but it is also itself a property of the diameter distribution).

The estimated distribution parameters often perform rather poorly as dependent variables in a linear model of stand characteristics. An alternative method is to calculate a certain set of percentiles from the estimated distributions that can be used as dependent observations in the regression. After this second estimation, the calculations must be inverted to obtain predicted parameters. PN requires three percentiles (the median and two more percentiles chosen symmetrically around the median) and will always yield a solution (Madansky 1988), whereas S_B requires four percentiles (the same three as for SP but in addition the minimum value) but might not always yield a solution (Mønness 1982), revisited in Siekierski (1992).

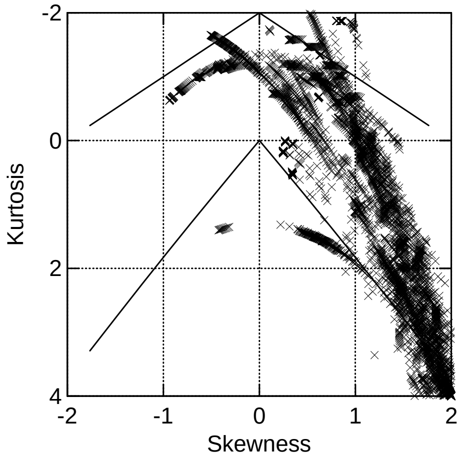
Conclusion
The Power-Normal distribution has properties that are well suited to the estimation and modeling of the distributions of tree diameter and height. Figure 7 shows how well it might fit. The PN can yield a large set of shapes with only three parameters, and is quite easy to use. Its excellent numerical properties imply that further exploration of its theoretical and practical properties would be worthwhile.

References:

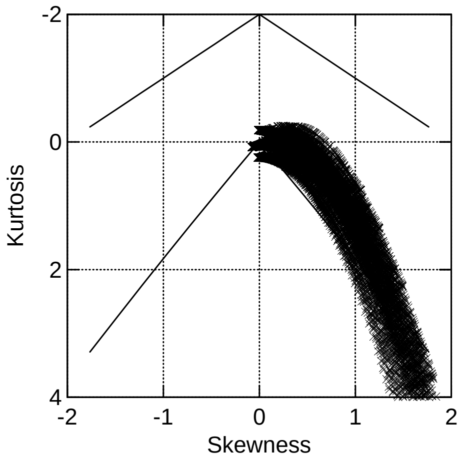
Box, G., and Cox, D. 1964. An Analysis of Transformations. Journal of the Royal Statistical Society, Series B **26**(2): 211-252.
Box, G., Hunter, J., and Hunter, W. 2005. *Statistics for Experimenters Design, Innovation and Discovery* 2ed. John Wiley & Sons.
Companion-Software. 2008. Metafile Companion Companion-Software, Sunderland MA USA.
Freeman, J., and Modarres, R. 2006. Inverse Box-Cox: The power-normal distribution. Stat. Probab. Lett. **76**(8): 764-772.
Garcia, O. 1983. A Stochastic Differential Equation Model for the Height Growth of Forest Stands. Biometrics **39**(4): 1059-1072.
Goto, M., and Inoue, T. 1980. Some properties of the Power Normal Distribution.. Bulletin Biometric Society of Japan **1**: 28-54.
Hafley, W.L., and Schreuder, H.T. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. Canadian Journal of Forest Research **7**: 481-487.
Hernandez, F., and Johnson, R.A. 1980. The Large-Sample Behavior of Transformations to Normality. Journal of the American Statistical Association **75**(372): 855-861.
Johnson, N.L. 1949. Systems of frequency curves generated by methods of translation. Biometrica **36**: 149-176.
Lambert, J.L. 1970. Estimation of Parameters in the four-parameter Lognormal Distribution. Australian Journal of Statistics **12**(1): 33-43.
Madansky, A. 1988. Prescriptions for Working Statisticians. Springer Verlag.
Maltamo, M., Kangas, A., Uuttera, J., Torniainen, T., and Saramäki, J. 2000. Comparison of percentile based prediction methods and the Weibull distribution in describing the diameter distribution of heterogeneous Scots pine stands. Forest Ecology and Management **133**(3): 263-274.
Merganič, J., and Sterba, H. 2006. Characterisation of diameter distribution using the Weibull function: method of moments. European Journal of Forest Research **125**(4): 427-439.
Mønness, E. 1982. Diameter distributions and height curves in even-aged stands of Pinus Sylvestris L. Medd. Nor. inst. skogforsk **36**(15): 1-43.
Rennolls, K., and Wang, M. 2005. A new parameterization of Johnson's S_B distribution with application to fitting forest tree diameter data. Canadian Journal of Forest Research **35**: 575–579.
SAS. 2008. SAS 9.1. SAS Institute, Inc. Cary, NC, USA.
Siekierski, K. 1992. Comparison and Evaluation of Three Methods of Estimation of the Johnson S_b Distribution. Biometrical Journal **34**(7): 16.
Systat Software, I. 2004. SYSTAT SYSTAT, Chicago, IL USA.
Vestjordet, E. 1977. Precommercial thinning of young stands of Scots Pine and Norway Spruce I: Data stability, dimension distribution etc. Meddelelser fra Norsk institutt for skogforskning **33**(9): 1-436.
Wang, M., and Rennolls, K. 2005. Tree diameter distribution modelling: introducing the logit–logistic distribution. Canadian Journal of Forest Research **35**: 1305–1313.

Number of forest stands =139	Minimum	Maximum	Mean	Standard Deviation
Total age	17.0	36.0	23.6	4.2
Upper Height at 15 year breast age, m	4.6	9.5	7.3	9.3
Upper Height, m	4.6	14.5	8.5	20.1
No. Trees pr. hectare	1000.0	6918.9	2831.9	997.2
Basal area mean diameter Dg, cm	4.2	15.4	8.5	2.1
Mean height, Lorey's formula Hl, m	3.7	13.3	7.4	2.0

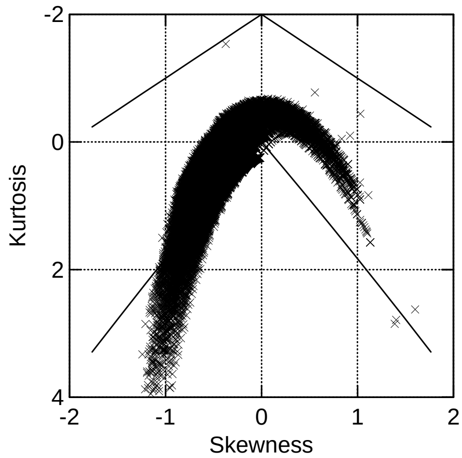
Table 1. Stand characteristics of the 139 stands.



$\lambda \in \langle -3, 0 \rangle$



$\lambda \in \langle 0, 1 \rangle$



$\lambda \in \langle 1, 3 \rangle$

Figure 1 Simulated skewness x kurtosis values of the PN distribution. The lines show the border of the (theoretical) impossible region, and the log-normal line. S_B covers the region between the lines. The normal distribution has a value (0,0).

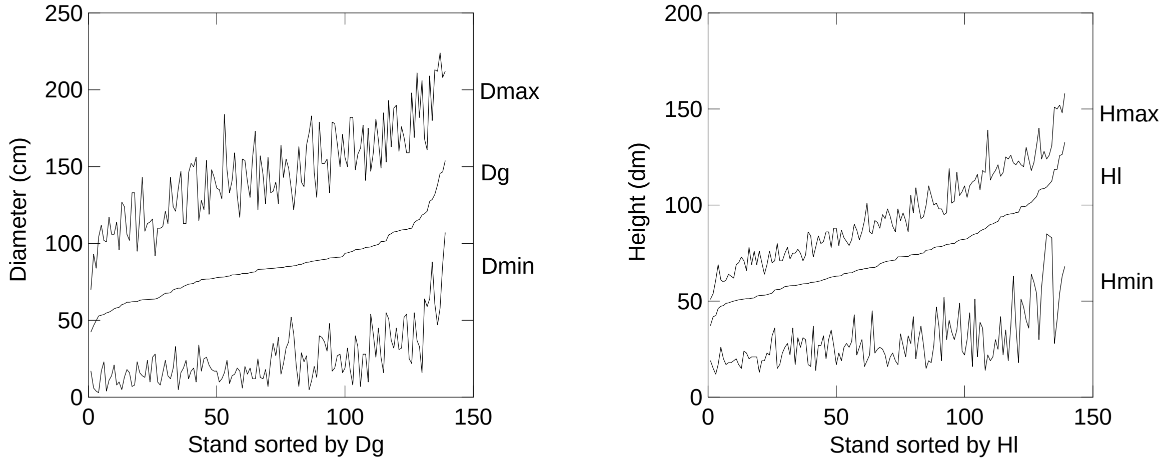


Figure 2 Diameters and heights of the stands. The stands were sorted by basal area mean diameter (Dg) and Lorey's height (Hl), respectively.

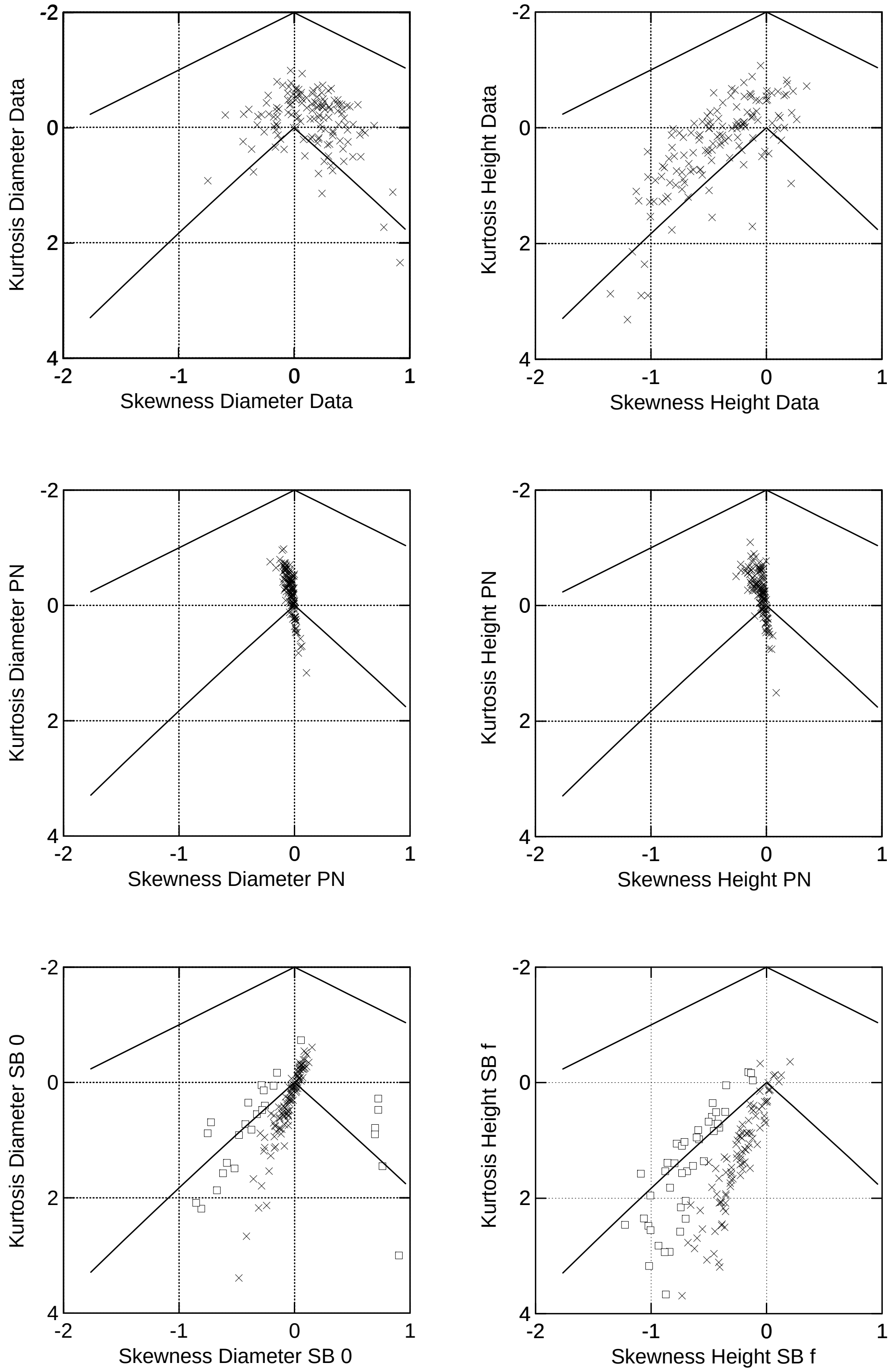
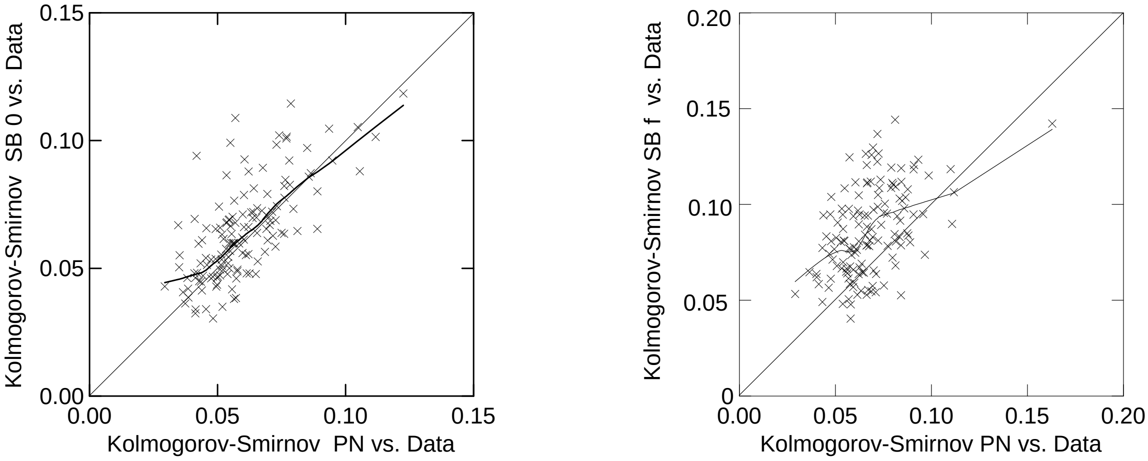


Figure 3. Skewness and Kurtosis values for diameters and heights. Values for the original data (upper), after the PN transformation (middle), and after the S_B transformation (bottom) are shown. For S_B, τ was fixed; $\tau = 0$ (S_B 0) on diameters and $\tau = H_{min} - \left(H_{max} - H_{min}\right)/\sqrt{n}$ on heights (S_B f). Each cross/square represent one stand of trees. A square indicates a stand where convergence did not occur and the starting upper value of θ was used.



Diameter Height

Figure 4. Kolmogorov–Smirnov statistics. PN is shown on the horizontal axis and S_B is shown on the vertical. For S_B, τ was fixed; $\tau = 0$ (S_B 0) on diameters and $\tau = H_{min} - \left(H_{max} - H_{min}\right)/\sqrt{n}$ on heights (S_B f).

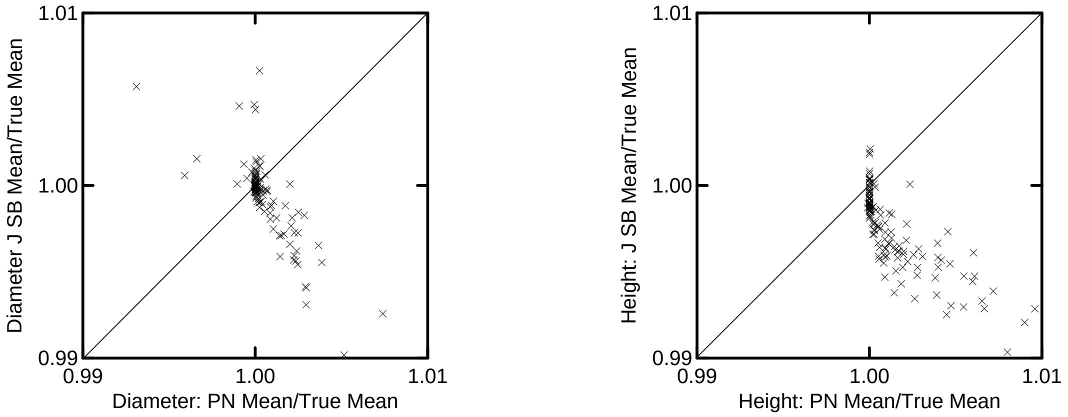


Figure 5. The means of the fitted distributions S_B and PN, divided by the true means. A point at (1,1) is a stand where both S_B and PN estimates the mean correctly. Points in the right-bottom corner are stands where S_B underestimates the mean while PN overestimates the mean. A value at 1.01 or 0.99 is a 1 % discrepancy.

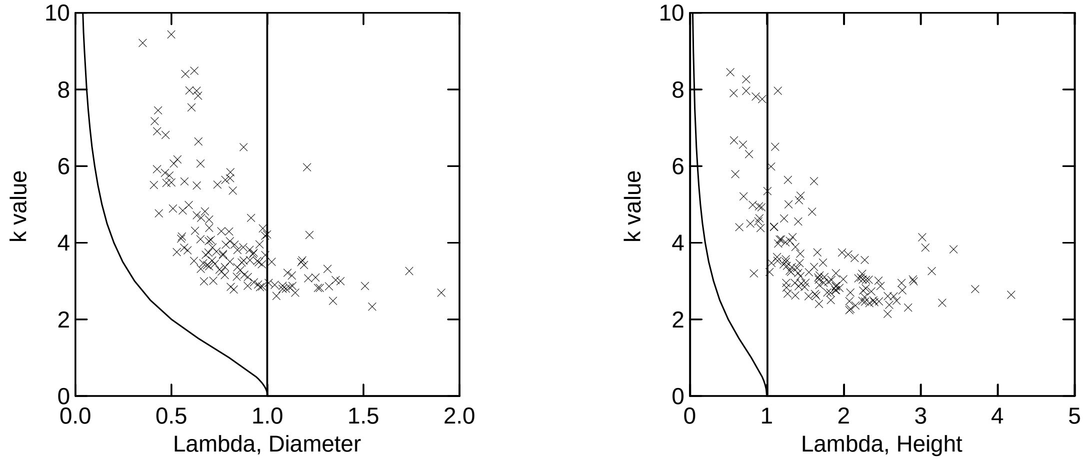
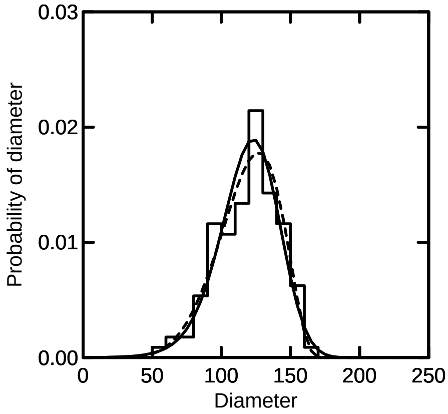
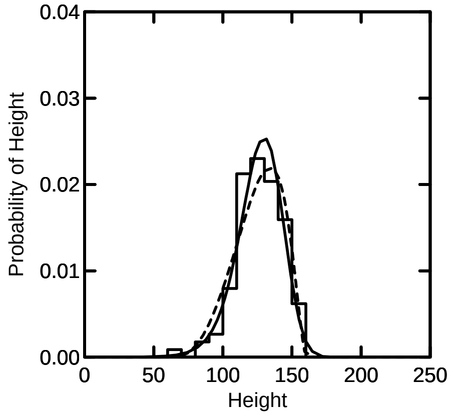


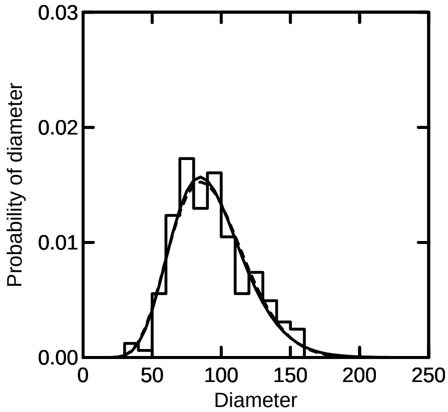
Figure 6. The space of λ * k. Each cross represents a stand of trees. The area between the curve $k = 2\sqrt{\lambda^{-1} - 1}$ and λ=1 yields right skewed distributions whereas the area where λ>1 and k>≈2 yields left skewed distributions.



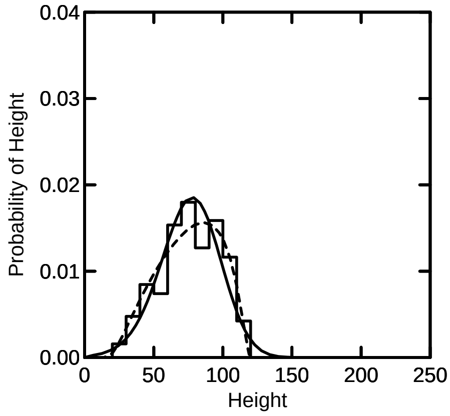
Diameter, Stand=18



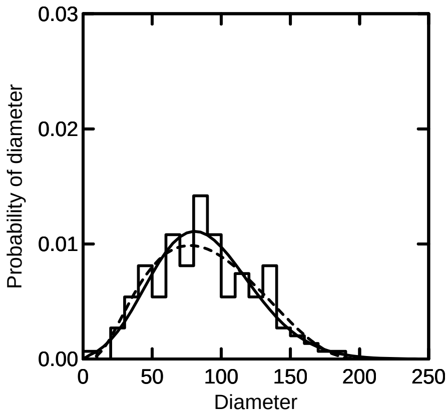
Height, Stand=55



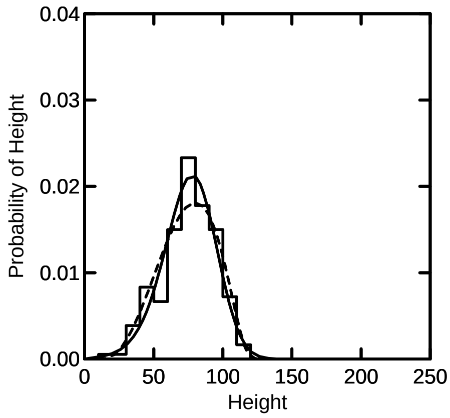
Diameter,Stand=59



Height, Stand=70



Diameter,Stand=91



Height, Stand=91

Figure 7. Examples of fitted diameter and height distributions. The histograms are the observed distribution. The fitted PN distribution (solid line) and the fitted SB distribution (dashed line). The areas under the curves are 1 in each case. The horizontal scale is cm for diameters and dm for heights.