RAPPORT NR. 4 – 2013

ERIK MØNNESS

Height curves based on the bivariate Power-Normal and the bivariate Johnson's System bounded distribution

Høgskolen i Hedmark

Erik Mønness

Height curves based on the bivariate Power-Normal and the bivariate Johnson's System bounded distribution

Høgskolen i Hedmark
Rapport nr. 4 – 2013

Høgskolen i **Hedmark**

Fulltekstutgave

Utgivelsessted: Elverum

Forfatteren er selv ansvarlig for sine konklusjoner. Innholdet gir
derfor ikke nødvendigvis uttrykk for Høgskolens syn.

I rapportserien fra Høgskolen i Hedmark publiseres FoU-arbeid
og utredninger. Dette omfatter kvalifiseringsarbeid, stoff av lokal
og nasjonal interesse, oppdragsvirksomhet, foreløpig publisering
før publisering i et vitenskapelig tidsskrift etc.

Rapport nr. 4 – 2013

ISBN: 978-82-7671-897-3
ISSN: 1501-8563

Hedmark University College

Title: Height curves based on the bivariate Power-Normal and the bivariate Johnson's System bounded distribution			
Author: Erik Mønness			
Number: 4	**Year:** 2013	**Pages:** 17	**ISBN:** 978-82-7671-897-3 **ISSN:** 1501-8563
Financed by:			
Keywords: height curve, bivariate Johnson's System bounded distribution, bivariate power-normal distribution			

Summary:

Often, a forest stand is modeled with a diameter distribution and a height curve as somehow separate tasks. A bivariate height and diameter distribution yield a unified model of a forest stand. The conditional median height given the diameter is a possible height curve. Here the bivariate Johnson's System bounded distribution and the bivariate power-normal distribution are evaluated and compared with a simple hyperbolic height curve.

Evaluated by the deviance, the hyperbolic function yields the best height prediction. A close second is the curve generated by the power-normal distribution. Johnson's System bounded distribution suffers from the sigmoid shape of the conditional median.

The bivariate power-normal is easy to estimate with good numerical properties. The bivariate power-normal is a good candidate model concerning forest stands.

Høgskolen i **Hedmark**

Tittel: Høydekurver basert på betinget median i bivariat Johnsons system begrenset fordeling og den potensnormale fordeling			
Forfatter: Erik Mønness			
Nummer: 4	**År:** 2013	**Sider:** 17	**ISBN:** 978-82-7671-897-3 **ISSN:** 1501-8563
Oppdragsgiver:			
Emneord: høydekurve, bivariat Johnsons system begrenset fordeling, bivariat potensnormal fordeling			

Sammendrag:

Diametre og høyder håndteres ofte separat, men en bivariate fordeling gir en enhetlig modell for et skogbestand. Den betingede median-høyden gitt diameter er en mulig høydekurve. Den bivariate Johnsons system B og den bivariate potens-normale fordeling er evaluert og sammenliknet med en enkel hyperbolsk høyde-kurve.

Vurdert med roten av midlere kvadratavvik: den hyperbolske funksjonen har minst avvik. Det er naturlig da den estimeres rett på diameter-høyde-relasjonen. Den potesnormale betingede medianen følger tett på. Johnsons system begrenset sin betingede median er funksjonelt en S-formet kurve som gjør den mindree god, selv om den også fungerer bra i mange tilfelle. Den bivariate potesnormale fordelingen har gode numeriske egenskaper og utgjør en god kandidat for å estimere skogbestand.

INNHOLD

INTRODUCTION

Predicting the height given the diameter is a common forestry task since diameter historically has been easy to measure and height harder to measure. A large set of mathematical functions have been explored, e.g. see Fang & Bailey (1998). However, using airborne laser height measurements (Gobakken & Næsset, 2004) it might be the other way around, so there might be a need for a diameter-given-height relationship. A bivariate distribution is a tool for a symmetric view of diameter and height.

This article is a follow up to Mønness (2011). There, the Power-Normal distribution (Freeman & Modarres, 2006) was explored, introduced to forestry and compared with Johnson's System bounded distribution (Johnson, 1949). This article will utilize the bivariate Power-Normal distribution and the bivariate Johnson's System bounded as a source of a height curve. The bivariate Johnson's System bounded, described in Johnson & Kotz (1972), was introduced to forestry by Schreuder & Hafley (1977) and has been much used (Wang & Rennolls, 2007). As a reference height curve, a simple hyperbolic function is used (Vestjordet, 1972).

This article does not cover the full properties and possibilities of bivariate distributions.

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 & 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. The S_B distribution is the distribution of X .

The Power-Normal distribution

The Box-Cox transformation (Box & Cox, 1964) 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 power-normal distribution (PN) is the distribution of X. The log() case [3] is identical to the Johnson's S_L so both PN and S_B have 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* can be a truncated normal.

The power-normal distribution (Freeman & Modarres, 2006) has density

[4] $f(x)=\frac{x^{\lambda-1}}{\Phi(sign(\lambda)k)\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x^{\lambda}-(\lambda\mu+1)}{\sigma\lambda}\right)^2}$, $x \geq 0$

With cumulative distribution

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

where $\Phi(\ )$ is the cumulative distribution function of the standard normal, with $k = \frac{\lambda\mu+1}{\sigma\lambda} = \frac{\mu}{\sigma}+\frac{1}{\sigma\lambda}$.

λ and k are shape parameters while (μ,σ) are location/scale parameters.

The properties and potential use in forestry are discussed in Mønness (2011). Since the inverse distribution is near normal (the truncation is in practice small, and diameters and heights are always positive) the bivariate normal distribution properties can be used to establish a bivariate PN-distribution.

Height curves

A large set of mathematical functions have been explored, e.g. see Fang & Bailey (1998) including the reference function below (Vestjordet, 1972), which is also function no 12 in Fang & Bailey (1998):

[6] $h = 1.3+\left(\frac{d}{A+Bd}\right)^2 = 1.3+\left(Ad^{-1}+B\right)^{-2}$ named Vestjordet2 herein ("2" for second power),

where h is in meters and d is diameter at breast height (1.3m). The function is a concave bounded hyperbolic curve with $h(0)=1.3$ and $h(\infty)=1.3+B^{-2}$. The estimation by a non-linear regression program (Systat Software, 2012) is straight forward and with stable numerical properties and seems to fit our data well.

Finding the conditional expected height given a diameter with S_B and PN is very complicated, but since the normal is symmetric, the next best thing is to find the conditional median (or any percentile) height given a diameter.

Bivariate S_B-distribution

Consider (diameter, height) data from every tree in a stand $(D_1,H_1),\ldots,(D_n,H_n)$. Estimate the S_B distribution separately on diameters and heights, transform to the normal scale and estimate the usual bivariate normal correlation ρ .(>0 in a forestry setting) (Schreuder & Hafley, 1977). The conditional median height given diameter is

$$[7] \quad h_{0.5}(d) = \tau_H + (\theta_H - \tau_H)\left(1 + \left(\frac{d-\tau_D}{\theta_D - d}\right)^{-\rho\frac{\sigma_H}{\sigma_D}} e^{\left(\rho\frac{\sigma_H}{\sigma_D}\mu_D - \mu_H\right)}\right)^{-1}$$

The curve is increasing and bounded by $h_{0.5}(\tau_D) = \tau_H$ and $h_{0.5}(\theta_D) = \theta_H$ .If $\rho\frac{\sigma_H}{\sigma_D} > 1$ the curve is sigmoid -shaped as seen from the d axis (Johnson & Kotz, 1972). The shape can be acceptable depending on which part of the sigmoid appears in the relevant domain of d and h values. When $0 < \rho\frac{\sigma_H}{\sigma_D} < 1$ the curve is sigmoid-shaped as seen from the H axis.

Bivariate PN-distribution

Estimate the PN distribution parameters $(\lambda_D, \mu_D, \sigma_D)$ and $(\lambda_H, \mu_H, \sigma_H)$ separately. Use the transformed data to estimate ρ the usual normal correlation (>0 in a forestry setting).

The conditional median height given the diameter d is given by

$$[8] \quad h_{0.5}(d) = \left\{1 + \lambda_H\mu_H - \frac{\rho\sigma_H\lambda_H}{\sigma_D\lambda_D}(1 + \mu_D\lambda_D) + \frac{\rho\sigma_H\lambda_H}{\sigma_D\lambda_D} d^{\lambda_D}\right\}^{\frac{1}{\lambda_H}}$$

λ_H and λ_D are larger than zero in every case (Mønness, 2011). Thus the curve is increasing in our setting. If $\frac{\lambda_D}{\lambda_H} < 1$ the shape is concave given that the other parameters are consistent and reasonable. The functional structure resembles that of Vestjordet2 [6] however with varying positive powers. The curve is unbounded.

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) and presented in Mønness (2011).

RESULTS

The estimation is documented in Mønness (2011). Here only the bivariate results are shown. The statistical software SAS (SAS, 2008) was used for programming and SYSTAT (Systat Software, 2012) for graphics. Graphs were enhanced using a metafile program (Companion-Software, 2008)

The height prediction fit was evaluated by the deviance[1], $\sqrt{\frac{\sum\left(H_{observed}-H_{predicted}\right)^2}{n}}$ and by visual inspection of the H/D plots of every stand. The measurement scales in the figures and tables are 0.1m for height and 0.01m for diameters.

Figure 1 depicts the deviance on each stand between the methods: Upper figures show PN and S_B vs. Vestjordet2. Lower left shows S_B vs. PN. Lower right shows the stand correlations with S_B and PN on the normal scale. Points on the diagonal have equal fit by the two methods in question.

The correlations themselves are highly correlated. The deviances are also highly correlated; some stands yield good prediction by all methods and other stands yields poorer prediction by all methods. It is evident that S_B has the highest deviance on most stands. Vestjordet2 and PN are very close Table 1 shows the residuals from all stands, “the entire forest”, taken together. Figure 1 and Table 1 tell the same story.

Vestjordet2 has the smallest mean residuals (also on nearly every stand, not shown). This is to be expected since Vestjordet2 works directly on the height-diameter data while PN and S_B work through the bivariate distribution and the transformed-retransformed values. Usually S_B underestimates the mean slightly while PN overestimates the mean slightly.

Figure 2 depicts sample stands. The circles are individual trees. The solid line is the Vestjordet2 curve. The long dashed line is S_B. The short dashed line is PN. The selected sample stands are chosen as the stands with best and worst prediction, as measured by the deviance. The lower right stand has the median fit with Vestjordet2. The ranking within each method of the chosen stands are given in Table 2. Rank=1 is best, rank 139 is worst.

The shape of S_B height curve is given by $\rho\frac{\sigma_H}{\sigma_D}$ while the shape of PN height curve is given by $\frac{\lambda_D}{\lambda_H}$. These parameters are plotted in Figure 3 for each stand, on a log scale. Most stands appear

1 The denominator has not been adjusted for the number of estimated parameters; 2 (Vestjordet2), 7 (PN) and 9 (SB).

with $\rho \frac{\sigma_H}{\sigma_D} > 1$ and $\frac{\lambda_D}{\lambda_H} < 1$. In those cases the S_B is sigmoid-shaped and PN are concave. In most other cases the curves appear close to straight lines on the plots (not shown). The higher $\rho \frac{\sigma_H}{\sigma_D}$, the more evident is the sigmoid-shape. The more evident the sigmoid-shape, the higher is the deviance of the S_B height curve.

Figure 4 depicts stand no. 91; diameter and height distributions and the height given diameter curve (distributions also depicted in Mønness (2011)). In this plot solid line is PN and dashed line is S_B.

DISCUSSION

Establishing an H/D relation working directly on the height and diameter values seems to yield the best fit. However, the bivariate distribution yields a unified model of distributions and the H/D relation. It is also symmetric between height and diameter. The PN has an overall fit close to Vestjordet2.

Judging the median H/D relation, the PN seems superior to S_B. The S_B has a drawback concerning the functional sigmoid-shape that also appears in many real cases. However, the distribution fit of PN was better than that of S_B on these data, especially for heights (Mønness, 2011). This might have an offset on the H/D relation too.

The PN is easy to estimate with good numerical properties. The bivariate PN is a good model concerning forest stands. A follow-up study would be to explore the full bivariate distribution properties.

REFERENCES

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B, 26*(2), 211-252.

Companion-Software. (2008). Metafile Companion (Version 1.2.0). Sunderland MA USA: Companion-Software. Retrieved from http://companionsoftware.com/Home.aspx

Fang, Zixing, & Bailey, R. L. (1998). Height-diameter models for tropical forests on Hainan Island in southern China. *Forest Ecology and Management, 110*(1-3), 315-327. doi: Doi: 10.1016/s0378-1127(98)00297-7

Freeman, J., & Modarres, R. (2006). Inverse Box-Cox: The power-normal distribution. *Statistics & Probability Letters, 76*(8), 764-772. doi: 10.1016/j.spl.2005.10.036

Gobakken, Terje, & Næsset, Erik. (2004). Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scandinavian Journal of Forest Research, 19*(6), 529-542. doi: 10.1080/02827580410019454

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrica, 36*, 149-176.

Johnson, N. L., & Kotz, S. (1972). Distributions in statistics. Continuous multiivariate distributions. New York: John Wiley.

Lambert, J.L. (1970). Estimation of Parameters in the four-parameter Lognormal Distribution. *Australian Journal of Statistics, 12*(1), 33-43.

Mønness, E. (2011). The Power-Normal Distribution: Application to forest stands. *Canadian Journal of Forest Research, 41*(4), 707-714. doi: 10.1139/X10-246

Rennolls, K., & Wang, M. (2005). A new parameterization of Johnson's SB distribution with application to fitting forest tree diameter data. *Canadian Journal of Forest Research, 35*, 575–579. doi: 10.1139/x05-006

SAS. (2008). SAS 9.1. Cary, NC, USA.: SAS. Retrieved from http://www.sas.com/

Schreuder, H. T., & Hafley, W. L. (1977). A Useful Bivariate Distribution for Describing Stand Structure of Tree Heights and Diameters. *Biometrics, 33*, 471-478.

Systat Software, Inc. (2012). SYSTAT (Version 13.1). Chicago, IL USA: SYSTAT. Retrieved from http://www.systat.com

Vestjordet, E. (1972). Diameter distribution and height curves for even-aged stand of Norway Spruce. *Meddr norske SkogforsVes., 29*, 472-557.

Wang, Mingliang, & Rennolls, Keith. (2007). Bivariate Distribution Modeling with Tree Diameter and Height Data. *Forest Science, 53*(1), 16-24.

	Height residuals		
	PN	S_B	Vestjordet2
N of Cases	16984	16984	16984
Minimum	-31.26	-35.13	-37.20
Maximum	36.97	40.61	32.11
Arithmetic Mean	0.28	-0.25	0.05
Standard Deviation	6.54	7.44	6.47
Deviance	6.55	7.44	6.47

Table 1 Residuals, entire "forest" (all stands).

Stand	PN		S_B		Vestjordet2	
	Rank	Deviance	Rank	Deviance	Rank	Deviance
13	3	2.89	20	4.58	**1**	2.90
14	**1**	2.54	**1**	2.48	3	3.06
16	113	7.72	**139**	12.10	115	7.69
55	137	10.21	135	10.76	**139**	10.19
61	**139**	10.58	136	10.83	138	10.18
123	54	5.45	56	6.33	70 (median)	5.77

Table 2 Rank and deviance of the sample stands, all three methods.

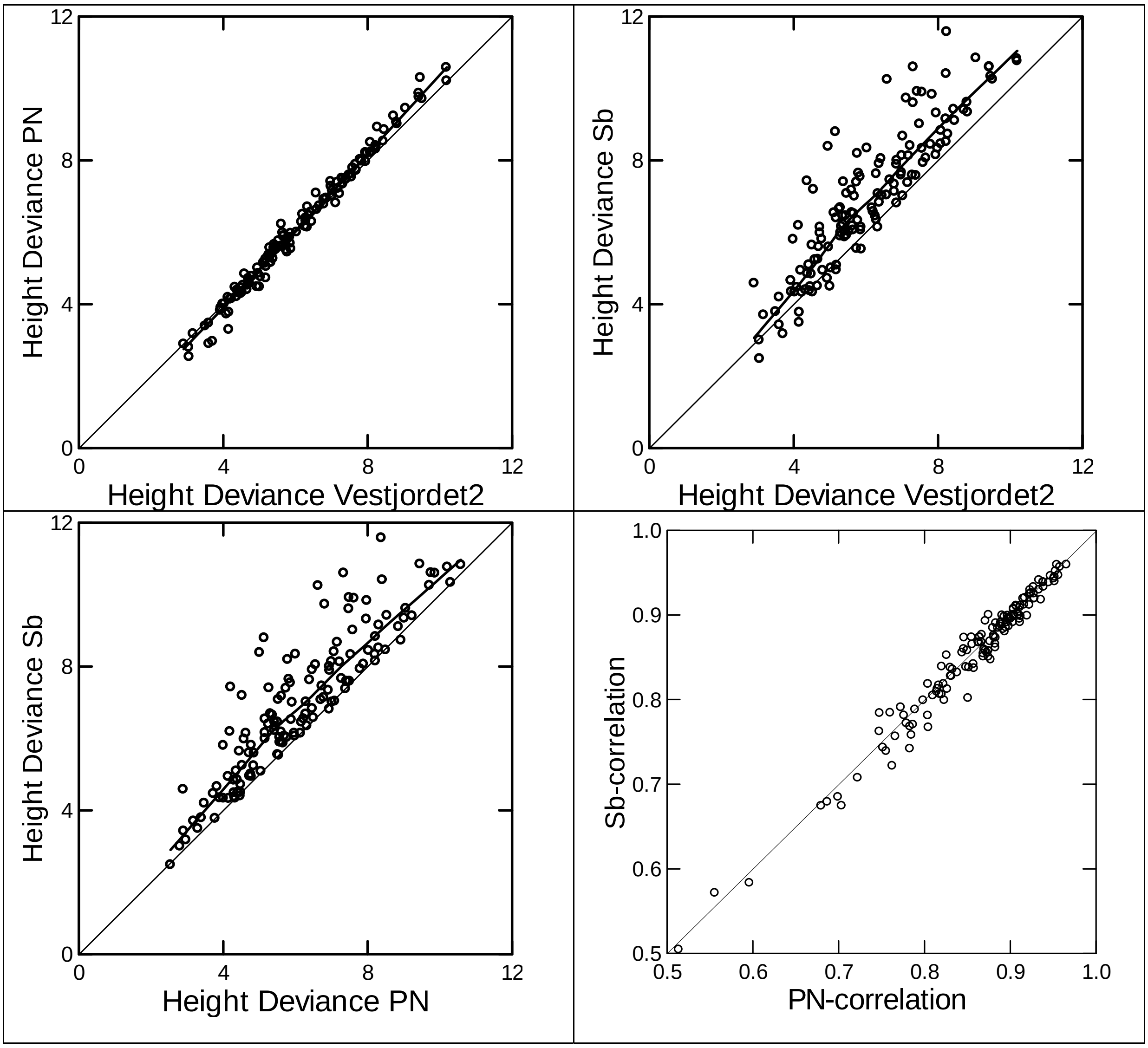
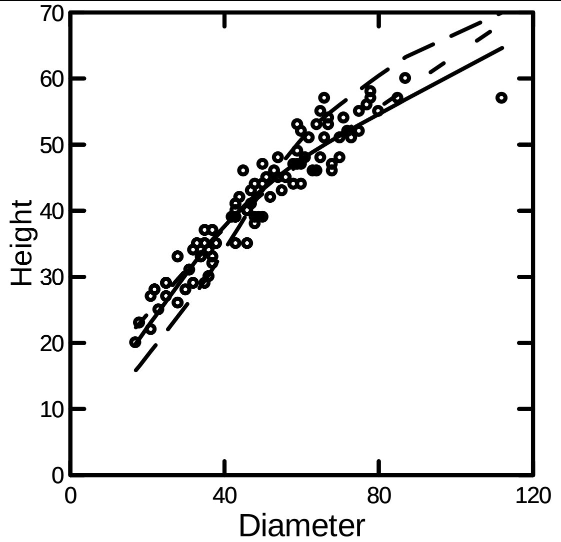
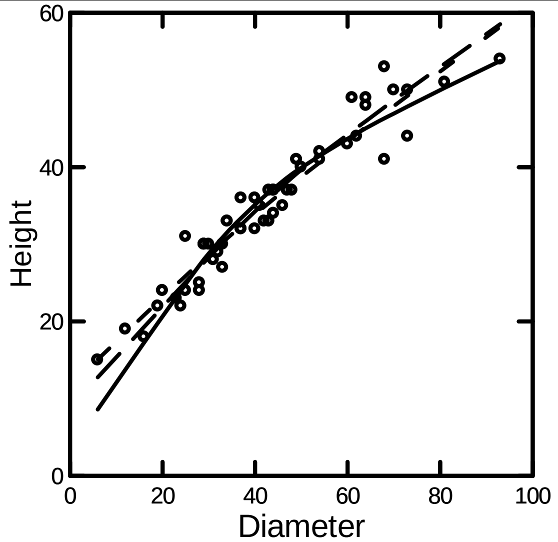


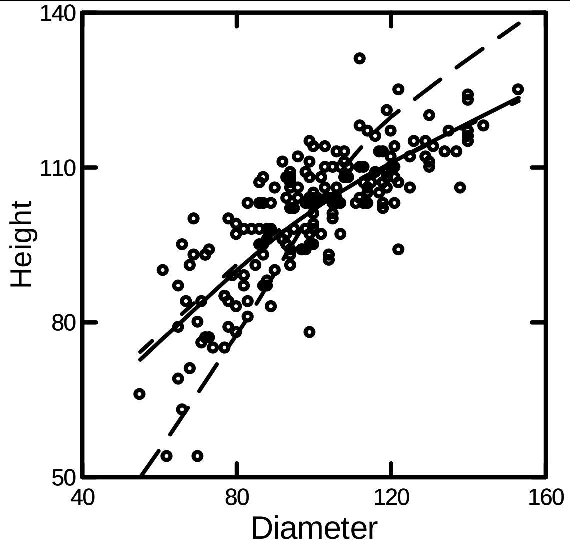
Figure 1 Deviance on each stand between the three predicted methods. Lower right: correlations.



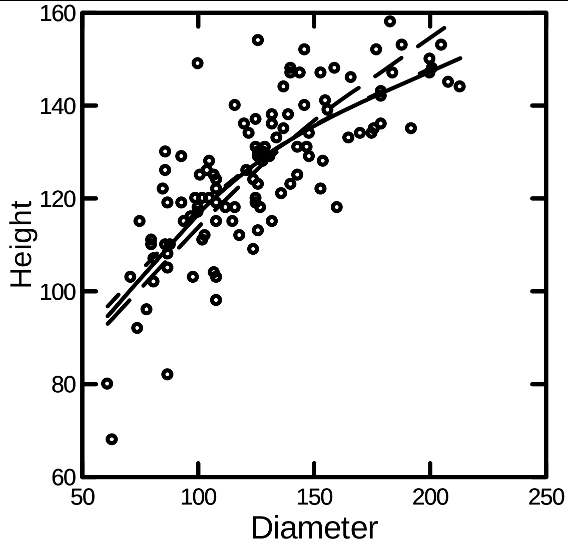
Stand 13



Stand 14



Stand 16



Stand 55

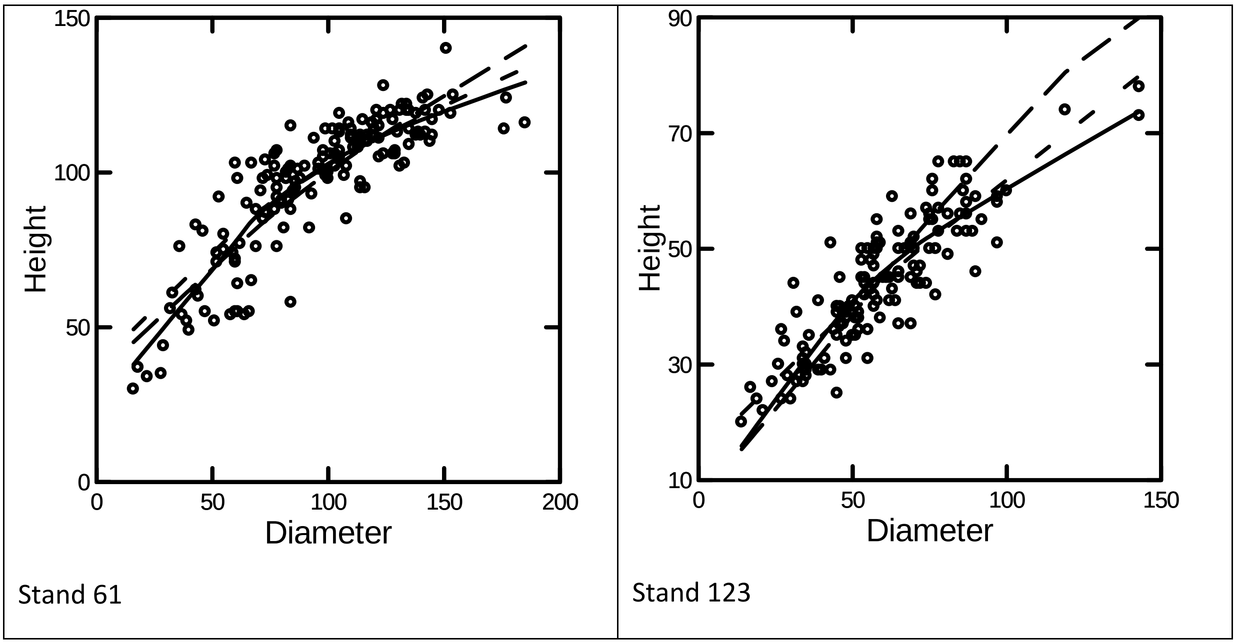


Figure 2 Sample Height/Diameter plots with predicted values

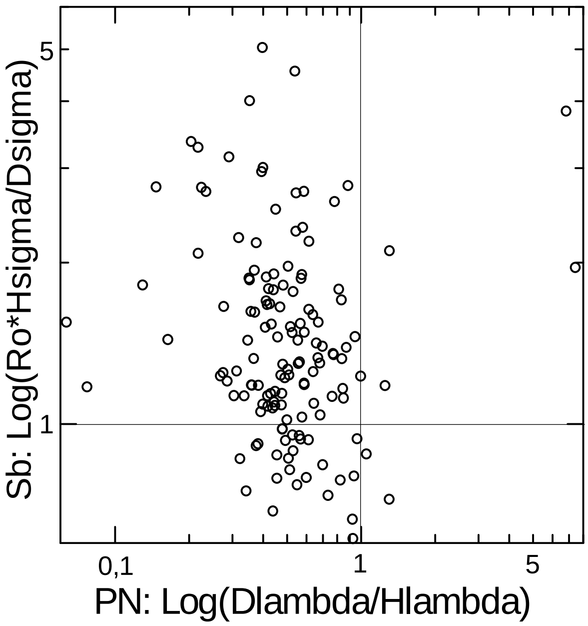


Figure 3 SB and PN shape parameters of the stands.

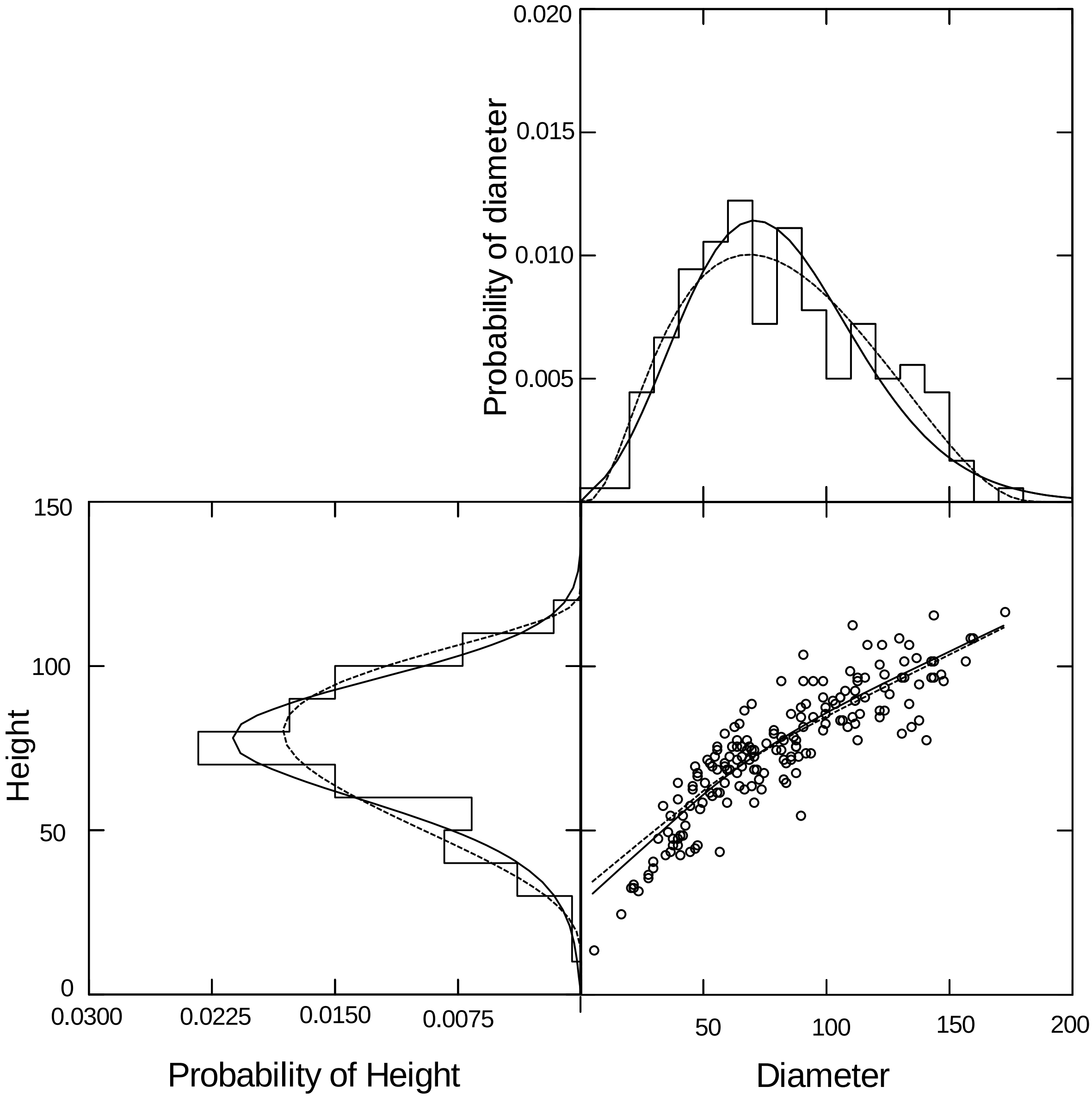


Figure 4 Diameter and height distribution with diameter*height curve. Stand 91.