Distributional Modelling in Forestry and Remote Sensing

Mingliang Wang

BSc (Shandong Agriculture University)

MSc (The Chinese Academy of Forestry)

A thesis submitted in partial fulfilment of the requirement of the University of G
the degree of Doctor of Philosophy

Statistics and Operational Research Group
School of Computing and Mathematical Sciences
The University of Greenwich

March 2005

Abstract

The use of distributional models in forestry is investigated, in terms of their capability of modelling distributions of forest mensurational attributes, for modelling and inventory purposes. Emphasis is put on: (i) the univariate and bivariate modelling of tree diameters and heights for stand-level modelling work, and (ii) heuristic methods for use and analysis of distributions which occur in multi-temporal EO imagery, (for the inventory-related tasks of land-use mapping, change detection and growth modelling).

In univariate distribution modelling, a new parameterization of the widely-used Johnson's S_B distribution is given, and new Logit-Logistic, generalized Weibull and the Burr system (XII, III, IV) models are introduced into forest modelling. The Logit-Logistic distribution is found to be the best among those compared. The use of regression-based methods of parameter estimation is also investigated.

In the domain of bivariate distribution modelling of tree diameters and heights the Plackett method (a particular form of copula) is used to construct Plackett-based bivariate Beta, S_B and Logit-Logistic distributions, (the latter two are new), which are compared with each other and the S_{BB} distribution. Other copula functions, including the normal copula, are further employed (for the first time in forest modelling) to construct bivariate distributional models. With the normal copula, the superiority of the Logit-Logistic in the univariate domain is extended into the bivariate domain.

To use multi-temporal EO imagery, two pre-processing procedures are necessary: image to image co-registration, and radiometric correction. A spectral correlation-based pixel-matching method is developed to "refine" manually selected control points to achieve very accurate image co-registration. A robust non-parametric method of spectral-distribution standardization is used for relative radiometric correction between images. Finally possibilities for further research are discussed.

Acknowledgements

I would like to thank my supervisors, Professor Keith Rennolls (the School of Computing and Mathematical Sciences, the University of Greenwich) and Professor Shouzheng Tang (the Research Institute of Forest Resources Information Techniques, the Chinese Academy of Forestry) for their guidance and support throughout my research, especially like to express on my first supervisor for his logical and intuitive guidance. I am grateful to the School of Computing and Mathematical Sciences, the University of Greenwich, for awarding me a three-year PhD scholarship to support my research.

I am grateful for the help from my colleagues in the Research Institute of Forest Resources Information Techniques, the Chinese Academy of Forestry. I would like to thank Professor Xifei Li, Professor Lingxia Hong, and Dr. Xiangdong Lei for providing me the Larix data used in forest growth and yield modelling. I also thank Professor Zengyuan Li, Professor Honggan Wu, Professor Bingxiang Tan, and Dr. Erxue Chen, for the permission to using their remote sensing imagery, and for the help from them and Professor Xianwen Zhao during my remote sensing studies.

Finally, I would like to thank my family, and especially my wife, Jing Hong, for their support during my PhD research.

Declaration

I certify that this work has not been accepted in substance for any degree, and is not concurrently submitted for any degree other than that of Doctor of Philosophy (PhD) of the University of Greenwich. I also declare that this work is the result of my own investigations except where otherwise stated.

Contents

Chapter 1: Introduction

1.1 Introduction

This thesis reports research on the development and application of distributional modelling techniques in Forestry. The application is considered to fall into two main areas.

First, *Forest Mensuration and Modelling* is concerned with the characterization of the distributions of measurements of individual tree attributes, typically measured on sample plots within a forest compartment. This form of data is characteristic of all field measurements of forest compartments.It is such a basic component of any forest inventory that plot measurements and models and estimates derived from them, are often regarded as an essential component of the area of *Forest Inventory*. Stand volume estimation is an important aspect of forest mensuration and inventory, and is usually based on estimates of individual tree volumes from a (double, triple, or a multi-phase) sample of tree diameters and heights (and volumes). In Section 1.2 we expand on this area, as a rationalé/justification for the distributional modelling research of this thesis.

The second main application area considered in this thesis is concerned with the use of Earth Observation (EO) imagery data in order to support the inventory and modelling of forest status, change and growth. The typically available data in this case consist of partial image coverage of a land area, obtained at a number of distinct times. In general, the sensing instrument could differ between recording times, and hence the pixel size and spectral structure of the imagery could differ between recording dates. However, this degree of generality has not been considered in the research reported in this thesis, and only multi-temporal imagery from the same sensing instrument has been considered. In particular the imagery used in this thesis is obtained by Landsat Thematic Mapper (TM). Use of EO imagery has been much researched since it has been seen as a means to rapid and cost-effective forest inventories of the future. Consequently, the literature on the topic, since the

Landsat satellite was launched in 1972 (Landgrebe 1997), is vast. This area of research is also highly complex, since modelling and estimation from complex samples from a spatio-temporal process are amongst the most difficult on-going statistical challenges in environmental process modelling.

Multi-temporal imagery data may be seen as "distributional data" in spectral space if the pixel spatial information is ignored. Use of unsupervised and supervised classification techniques in this "feature space" are the standard techniques in this domain of remotely sensed imagery analysis. Inclusion of the spatial information into extended clustering and classification algorithms has led to object-based classification methods (Kettig and Landgrebe 1976). The complexity and dimensionality of the data are such that parametric methods of analysis likely not be appropriate, and non-parametric methods such as the k-nearest-neighbour (KNN) estimator (Tomppo 1991, Hardin 1994, Franco-Lopez et al. 2001, Haapanen et al. 2004) and Neural Networks (Lee and Landgrebe 1997) are probably the only realistic way in which the full complexity of the data can be addressed. See section 1.3 for further general considerations.

Section 1.4 gives an overview of the thesis.

1.2 Distributional Modelling in Forest Mensuration and Modelling

1.2.1 Traditional Approach to Stand Volume Estimation

The traditional procedure to estimate stand volume may consist of three steps. First, a sample of n_d trees is selected on which diameters are measured: we include the subscript to indicate the sample on which the estimate is based. It is assumed here, for simplicity and convenience, that a fixed area sample plot is used. Hence, an estimate of the marginal distribution of diameter, $\hat{f}_{n_d}(d)$, may be obtained. Second, a sample of tree heights (of size

n_h say) is obtained, normally on a sub-sample from the n_d diameter-trees. The height on diameter regression,

$$E(h \mid d) = \hat{h}_{n_h}(d) \tag{1.1}$$

may be obtained on the n_h trees for which both height and diameter are measured, where E denotes statistical expectation, h is tree height, d is tree diameter. See section 1.2.4 for further discussion following equation (1.1).

Finally, an estimated individual volume (V) equation,

$$E(V \mid d,h) = \hat{V}_{n_v}(d,h) \tag{1.2}$$

is traditionally used to estimate the mean sample tree volume ($\hat{\bar{v}}$) in the population as:

$$\hat{\bar{v}} = \int_{d>0} f(d)E(V \mid d, \hat{h}_{n_h}(d))\, dd \tag{1.3}$$

where $f(d)$ is the marginal diameter distribution. Then the estimator of mean tree volume ($\hat{\bar{v}}_T$) is given as:

$$\hat{\bar{v}}_T = \int_{d>0} \hat{f}_{n_d}(d)\, \hat{V}_{n_v}(d, \hat{h}_{n_h}(d))dd \tag{1.4}$$

A discrete diameter-sample based version ($\hat{\bar{v}}'_T$) is

$$\hat{\bar{v}}'_T = \frac{1}{n_d} \sum_{i=1}^{n_d} \hat{V}_{n_v}(d_i, \hat{h}_{n_h}(d_i)) \tag{1.5}$$

The individual volume equation (1.2) is usually assumed to be generally applicable, that is, independent of stand attributes (age, site quality, density,...). It is often constructed prior to the inventory, due to the difficulty in measuring individual tree volume during forestry inventory. In contrast, the *h-d* relationship of equation (1.1) (traditionally termed the "H-D" relationship) is usually stand/plot specific since both diameter and height measurement are usually practically feasible. Therefore, stand volume estimation is usually made from a double sample of tree diameters and heights (Clutter and Allison 1974).

1.2.2 Univariate Distribution Modelling for Stand Volume Estimation

Since stand volume is the primary variable in which forest managers are interested, any improvement on the traditional approach would be valuable. One natural improvement, as seen in equations (1.3) and (1.4), is to use a more flexible univariate distribution model for describing the diameter data. The most frequently used distributional models are the Weibull (Bailey and Dell 1973) and Johnson's S_B (Johnson 1949a, Hafley and Schreuder 1977).

The diameter distribution model is a key component of many growth and yield models (Hyink and Moser 1983, Borders and Patterson 1990). Diameter distribution models are intermediate between stand-level models (Tang 1991, Avery and Burkhart 1994) and individual tree models (Mitchell 1975, Wykoff et al. 1982, Rennolls and Blackwell 1988).

Diameter-distribution based growth-and-yield models can forecast the range of products which might be expected from a stand (Rennolls et al. 1985). Use of flexible models to describe the diameter distribution, in conjunction with methods for the construction of appropriate bivariate distribution models (particularly by use of the copula method, for details see Chapter 6), provides the potential for more accurate estimation of stand volume than the traditional techniques. This is further discussed in the next section.

1.2.3 Bivariate Distribution Modelling for Improving Stand Volume Estimation

Stand volume estimation may be improved by modelling the joint distribution of tree diameter and height. Bivariate distribution modelling provides an alternative to the traditional approach to estimating stand volume.

The mean sample tree volume in the population is given by

$$\hat{\bar{v}} = \iint\limits_{d>0,h>0} f(d,h)\, V(d,h)\, dd.dh \qquad (1.6)$$

where $f(d, h)$ is the joint distribution. Re-writing (1.6), we have,

$$\hat{\bar{v}} = \iint_{d>0,h>0} f(d,h)\, V(d,h)\, dd\, dh = \iint_{d>0,h>0} f(h \mid d) f(d)\, V(d,h)\, dd\, dh$$

$$= \int_{d>0} f(d) \left\{ \int_{h>0} f(h \mid d)\, V(d,h)\, dh \right\} dd \tag{1.7}$$

$$= \int_{d>0} f(d)\, E(V \mid d)\, dd$$

where $f(h|d)$ is the conditional height distribution given diameter d. The traditional approach (as in (1.3)) incorrectly assumes (or approximates) $E(V|d)$ is given by the volume of the tree with the expected height for the given d. That is,

$$E(V \mid d) = E(V(d, E(h \mid d)) \tag{1.8}$$

Hence (1.4) or (1.5) follows.

The estimated-height bias effects in (1.4) or (1.5) may be avoided by evaluating $E(V|d)$ in (1.7) by using $f(h|d)$, the conditional distribution of h for given d. That is,

$$E(V \mid d) = \int_{h>0} f(h \mid d)\, V(d,h)\, dh \tag{1.9}$$

where the conditional distribution is estimated from the height-sample by:

$$\hat{f}_{n_h}(h \mid d) = \frac{\hat{f}_{n_h}(d,h)}{\hat{f}_{n_h}(d)} \tag{1.10}$$

Hence an unbiased volume estimator ($\hat{\bar{v}}_B$), based on the bivariate and marginal densities, $f(d,h)$ and $f(d)$ respectively, is

$$\hat{\bar{v}}_B = \int_{d>0} f(d)\, \hat{E}_{n_v}(V \mid d)\, dd$$

$$= \int_{d>0} f(d) \left\{ \int_{h>0} \frac{\hat{f}_{n_h}(d,h)}{\hat{f}_{n_h}(d)} \hat{V}_{n_v}(d,h)\, dh \right\} dd \tag{1.11}$$

$$= \iint_{d>0,h>0} \hat{f}_{n_d}(d) \frac{\hat{f}_{n_h}(d,h)}{\hat{f}_{n_h}(d)} \hat{V}_{n_v}(d,h)\, dd\, dh$$

where the sample dependence has been made explicit. The discrete diameter-sample based version ($\hat{\bar{v}}'_B$) is

$$\hat{\hat{v}}'_B = \frac{1}{n_d}\sum_{i=1}^{n_d}\left\{\int_{h>0}\frac{\hat{f}_{n_h}(d_i,h)}{\hat{f}_{n_h}(d_i)}\hat{V}_{n_v}(d_i,h)\,dh\right\} \tag{1.12}$$

It is normally assumed that the diameter and height samples may be regarded as independent simple random samples of the population of trees in the (infinite) stand, (even when a fixed size sample plot is used). Then the estimates of the population conditional distribution (from the joint distribution) and marginal distributions may be based on the height-sample and diameter-sample respectively, where a double sampling is implemented.

If all the diameters and heights are measured on the sample plot then double sampling collapses to a simple random sampling with $n_h = n_d$. Hence an unbiased volume estimator can be based on the bivariate density $f(d,h)$ alone, and equations (1.11) and (1.12) can be simplified to:

$$\hat{\hat{v}}_B = \iint_{d>0,h>0} \hat{f}_{n_d}(d,h)\,\hat{V}_{n_v}(d,h)\,dd\,dh \tag{1.13}$$

1.2.4 The H-D Regression

The height-diameter regression model (1.1) (i.e. $E(h\,|\,d) = \hat{h}_{n_h}(d)$) may also be reformulated in terms of conditional and joint distributions. In fact, this is an underlying reason for the early work on bivariate distributions (Schreuder and Hafley 1977). That is,

$$\begin{aligned} E(h\,|\,d) &= \int_{h>0} hf(h\,|\,d)\,dh \\ &= \int_{h>0} h\frac{f(d,h)}{f(d)}\,dh \end{aligned} \tag{1.14}$$

In contrast, the traditional approach to obtaining (1.1) is by regression methods: ordinary least squares (OLS), weighted least squares (WLS), or generalized least squares (GLS).

In this sense, we see that the bivariate distribution modelling of $(d,\ h)$ provides an alternative to the generally used regression method of obtaining the H-D regression model. It seems that model (1.1) obtained by the "bivariate distribution fitting" is the main justification for a number of studies on bivariate distribution modelling (Schreuder and Hafley 1977, Tewari and Gadow 1999, Li et al. 2002), in that the resulting H-D models may more reasonably express the H-D relationship, and thus may improve the traditional approach to volume estimation by estimating more accurately the expected height at a given diameter.

The main advantage of the regression methods lies in that the conditional expectations, i.e., $E(\mathrm{h}|\mathrm{d})$, are explicitly given by the regression model $h_{n_h}(d_i)$. The two methods are complementary (Ord 1972). We may further regard the bivariate distribution modelling as another way to accommodate heteroscedasity as does the WLS when the assumptions in using OLS are violated.

1.2.5 Multivariate (dimension >2) Distribution Modelling in Forestry

Suppose we adopt a 3-stage sampling of (D, H, V) where V is assumed to be measurable, by fitting a trivariate distribution to the sample of (D, H, V), we can subsequently obtain (1.2) from the fitted trivariate distribution by analogy with obtaining (1.1) from a fitted bivariate distribution as indicated in (1.14). One example was given by Schreuder et al. (1982). Note that this trivariate distribution improves the traditional approach to volume estimation by modifying the individual volume equation.

More generally, it is clear that for a generic tree, each of its mensurational attributes is just one coordinate in the multi-dimensional characterisation of that generic tree. The multivariate structure of a multivariate tree dataset may be studied by multivariate statistical methods which generally amount to a description of the multidimensional relationships of the attribute data. Such purely statistical methods of analysis come in a number of forms, with regression analysis possibly being the most widely applied statistical technique. A basic requirement of regression is that one variate be the primary measure of interest (the dependent variable) and the others are explanatory variables (the independent variables). This may be not appropriate, since we may be more interested in understanding the joint distribution of the multivariate data. A by-product of the joint multivariate distribution approach is the conditional expectations of the dependent variate.

1.3 Distributional Modelling in the Analysis of Multi-Temporal Imagery

Satellite remote sensing will play an ever increasing role in forest inventory. Especially at the large scale, it provides forest information at a lower cost but more quickly than by ground survey (Holmgren and Thuresson 1998).

The main methods for extracting forest information from satellite images include regression analysis and classification. Regression analysis is the most commonly used method to establish the relationship between forest measurements and image properties, and therefore to estimate forest variables such as stand volume, age, and species composition (Franklin 1986, Ardö 1992, Cohen and Spies 1992, Gemmell 1995, Trotter et al. 1997, Lefsky et al. 2001, Lu et al. 2004). Such regression based approaches normally use single-date imagery. More recently, Lefsky et al. (2001) showed that multi-date TM is superior to single-date TM, ADAR (a hyperspatial sensor), and AVIRIS (a hyperspectral sensor) in its ability to predict forest structure variables such as basal area and biomass. They recommended that multi-temporal TM should be considered as an alternative to either ADAR or AVIRIS.

Classification can be used in the analysis of single as well as multiple temporal images. The maximum likelihood (ML) based classification is probably the most frequently used supervised classifier in remote sensing, which is based upon the assumption of the multivariate normal distribution. Hence, we see that distributional modelling also plays a potential role in using remote sensing imagery as well as in growth and yield modelling. In particular, the distributional modelling in the analysis of multi-temporal imagery may be roughly compartmentalized into the following tasks:

(i), (a) Image co-registration (related to the geometric correction) and (b) radiometric correction,

(ii) Classification of imagery (pixels) into different land-uses to produce land-use maps,

(iii) Change detection and mapping from imagery at two or more times, and

(iv) "Growth" estimation from imagery at more than two times.

In the first task, multi-temporal image co-registration and radiometric correction are two outstanding requirements of the pre-processing necessary before change/growth detection/estimation can be conducted (Coppin and Bauer 1996).

The second task, the classification of land areas into different land cover types (forest/non-forest, forest cover types) has been extensively studied (Horler and Ahern 1986, Moore and Bauer 1990, Bolstad and Lillesand 1992, Bauer et al. 1994, Wolter et al. 1995). Use of the estimated land-use classes as a basis for stratification in large scale forest inventory can increase the precision of inventory estimates (McRoberts et al. 2002).

In the third task, the "change" refers to abrupt or rapid change. Multi-temporal satellite imagery has been effectively used to detect and monitor abrupt changes in forests, such as human induced clearcuts or thinning (Coppin and Bauer 1994, Olsson 1994, Franklin et al. 2000, Wilson and Sader 2002), insect and disease damage (Vogelmann and Rock 1988). A key review on forest change detection is Coppin and Bauer (1996).

In the fourth task, we purposely use "growth" to represent the gradual change due to the normal forest growth. Growth estimation through time-series of satellite imagery is of much interest to foresters, but also very challenging (Joyce and Olsson 1999).

We note that most satellite data analysis, regression or classification, is carried out on a pixel basis. Taking the spatial information into account, the analysis becomes object/polygon based (Kettig and Landgrebe 1976). Polygon-based analysis seems more appropriate in forestry application, as homogenous polygons represent forest stands/compartments. Polygon-related methods have been used in Tomppo (1987), Woodcock and Harward (1992), Bauer et al. (1994), Kilpeläinen and Tokola (1999), Rennolls (1999), and Wulder and Seemann (2003).

The combined use of regression, classification and change detection using satellite imagery, possibly together with use of growth and yield models, provides the basis to implement an annually updated forest inventory system (Bauer et al. 1994, Czapewski 1999, McRoberts 1999). The annual forest inventory system may possibly replace the traditional periodic inventory system at a large scale (regional or national).

All these aims, (i)-(iv), were part of the original objectives of the current research. However, in this thesis, only the first task is reported (Chapter 7 and 8).

1.4 Overview of this Thesis

This chapter, (Chapter 1: Introduction) provides an analysis of the use of distributional models in forestry in Forest Mensuration, Inventory, and Remote Sensing. This material, particularly on the Mensuration and Inventory side, contains new material which is not yet published.

In Chapter 2, a new parameterization of Johnson's S_B is presented.

Chapter 3 introduces the Logit-Logistic distribution, which is similar to Johnson's S_B but more flexible. Other distributional models are introduced as well, including the generalized Weibull and the Burr system (XII, III, IV) models, applied to forest diameter distribution modelling. Generally speaking, the Logit-Logistic is found to be the best univariate model among those compared.

The maximum likelihood method for the estimation of parameters of distribution models is the most common method used in this thesis. Regression-based methods for parameter estimation are also found to perform well, (Chapter 4).

Chapter 5 uses Plackett's method to obtain a Plackett-bivariate-S_B and Plackett-Logit-Logistic.

Chapter 6 employs copula functions, in particular the normal copula, to construct bivariate distributional models. This chapter includes the normal copula with Logit-Logistic marginals, which proves to be superior to S_{BB}.

Chapter 7 reports work on using multi-temporal TM imagery for image co-registration.

Chapter 8 tentatively considers some aspects of radiometric correction (spectral standardization). Most of the material in chapter 8 appeared in a joint paper with Professor Rennolls, presented at a Digital Forestry Workshop in Beijing in 2004 (unpublished). Many of the ideas expressed in Chapter 8 are from Professor Rennolls' contribution to this joint paper.

It should be noted that that the research work on this project has been conducted in a manner oriented towards facilitating the publication of research results. Accordingly, the project has been broken down into sub-problems (these correspond to our main chapters), and each has been addressed largely independently from the others, even though there is, of course, a (back-ward) sequential dependence. A consequence of this is that the literature reviews for each of the sub-problems appears separately in each of the chapters, rather in a single "Literature Review" chapter presented early in the thesis.

Much of the material in Chapter 1 and Chapter 8 is open-ended and discursive. Much of this material is also either new or with future research challenges. Similarly, each of Chapters 2-7 contains its own relevant discussion material. Much of this material could have been presented within a final "Chapter 9: Discussion". However, the material has been placed in

Chapters 1 and 8 and throughout the thesis, in order to provide a rationale, a direction, and a start and an end-point for the thesis linked together by the Chapters 2-7 of this thesis. As a result there is no "Chapter 9: Discussion" in this thesis.

Chapter 2: A New Parameterization of Johnson's S_B Distribution

Summary

The S_B distribution is widely used in forestry to represent the empirical distributions of forest tree variables such as diameter, height and volume. The parametric form of the S_B model that has invariably been used is the form originally put forward by Johnson, in the 1949 paper in which he introduced the S_B distribution. It is well known that the parameterization chosen for a distribution model can have important effects when the distribution is fitted to real data. One parameterization may yield parameter estimates that are highly correlated, while another 'natural' parameterization could yield parameter estimates that are essentially uncorrelated. The feature that makes a parameterization "natural" is that the parameter has a natural interpretation in terms of the observed data distribution. A more "natural" parameterization of S_B is suggested, and the performance of the alternative parameterization is compared empirically on a 20 plots dataset of Changbai larch (*Larix olgensis* Henry). It is found that the new parameterization is better than Johnson's original parameterization, for the data sets considered here.

2.1 Introduction

Normal L. Johnson is the man of the 20^{th} century in relation to the distributional models in the statistics, being famous for his Johnson's system of distributions. Hafley and Schreuder (1977) first introduced the four parameter Johnson's S_B distribution (Johnson 1949a) into forest literature, and since then it has been widely used in forest diameter (and height) distribution modelling (Hafley and Buford 1985, Knoebel and Burkhart 1991, Zhou and McTague 1996, Kamziah et al. 1999, Li et al. 2002, Scolforo et al. 2003, Zhang et al. 2003). Johnson's definition and parameterization of the S_B distribution is based upon a transformation to normality. However, in his original parameterization, the diagram aimed to help understanding the idea of transformation to normal, is rather difficult to comprehend. In this Chapter we consider the inverse transformation from normality to S_B: doing so suggests a new and more natural parameterization of S_B.

A model is considered to be "well-parameterized", with respect to a given dataset, if the estimated variance-covariance matrix of the parameter estimates is diagonal. That is, the correlations between the estimates are all zero. Well-parameterized models are likely to be more stably and speedily estimable than models that are not well-parameterized. Variances of the parameter estimates of well- parameterized models are likely to be smaller than for models that are not well-parameterized (given other things being equal). Well-parameterized models are likely to result if the parameters are chosen to reflect clearly identifiable features of the observed dataset (Ross 1990). Models that have parameters relating to underlying processes that generate the distribution, can also lead to well-parameterized models. Finally, we might expect a "natural" parameterization (for example in terms of canonical parameters mean (μ) and variance (σ^2) for the Normal distribution, rather than ($1/\mu$) and (σ/μ)), to turn out to be "better parameterized" than a model that is constructed with no concern for parametric form. In fact, the canonical parameterization for a distributional model belonging to the exponential family is necessarily "well-parameterized", as is well known for the Normal distribution.

Maximum Likelihood is a commonly used method of estimating the parameters of a distribution model, and we use this approach to compare the estimates of the alternative S_B parameterizations, and their statistical properties (i.e. standard errors and correlations).

2.2 Alternative Parameterization of the Johnson S_B Distribution

Johnson's S_B probability distribution (Johnson 1949a) specifies a bounded pdf (of variate x, say) with the minimum parameter ξ, range λ, and two shape parameters γ and δ as the following,

$$f(x) = \frac{\delta}{\sqrt{2\pi}} \frac{\lambda}{(\xi+\lambda-x)(x-\xi)} e^{-\frac{1}{2}[\gamma+\delta\ln(\frac{x-\xi}{\xi+\lambda-x})]^2} \tag{2.1}$$

where $\lambda > 0$ and $\delta > 0$. The distribution, as specified by Johnson (1949a), and all subsequent users of the distribution in forestry applications, is the result of the following sequence of transformations of x to normality. Variate x has the S_B distribution if z, as defined in (i), (ii) and (iii) below, is a standard Normal distribution, N(0,1), where we:

(i) Scale x to a unit range:

$$y = \frac{x-\xi}{\lambda} \tag{2.2}$$

(ii) Apply a logit transformation to y:

$$u = \ln\left(\frac{y}{1-y}\right) \tag{2.3}$$

(iii) Apply an affine/linear transformation to u, to give z, which is N(0,1):

$$z = \gamma + \delta\, u \tag{2.4}$$

(Note: a linear transformation to normality would usually be represented as the equivalent standardisation transformation.) Essentially, the S_B distribution is transformed to normality by the logit transformation, and by analogy with the log-normal distribution (as the distribution transformed to normality by the log transformation) might well have been named the logit-normal distribution.

The "inverted" definition of S_B given above makes the sequence of transformations rather hard to visualise. Certainly, the diagrams presented by Johnson (1949a) are not easy to comprehend. Inverting this definitional sequence of transformations gives us a constructive definition of the S_B distribution.

(i)′ z~N(0, 1). Scale z to u, by:

$$u = \left(-\frac{\gamma}{\delta}\right) + \left(\frac{1}{\delta}\right) z \tag{2.5}$$

So, u~N($-\gamma/\delta$, $1/\delta^2$). It is the parameterization of this scaling transformation (corresponding to the affine transformation of (iii) above) that seems rather "unnatural" to us.

(ii)′ Apply a standard logistic transformation to u to give y, in the (0, 1) range:

$$y = \frac{1}{1+\exp(-u)} \tag{2.6}$$

(iii)′ Scale y to x, with range λ and minimum ξ:

$$x = \xi + \lambda\, y \tag{2.7}$$

Though the affine transformation given in (2.4) is a natural choice in mathematics, we see, when it is re-expressed as a scaling transformation in (2.5), that it is not the form of transformation that is statistically 'natural'. The natural scaling transformation would be:

$$u = \gamma' + \delta' z \tag{2.8}$$

so that u~N(γ', δ'^2) (≡N(μ, σ^2)) where

$$\delta'(\equiv \sigma) = \frac{1}{\delta} \tag{2.9}$$

$$\gamma'(\equiv \mu) = -\frac{\gamma}{\delta} \tag{2.10}$$

This is our simple re-parameterization of Johnson's S_B and the two parameterizations are related by equation (2.9) and (2.10). We use parameter pairs (γ', δ') when we wish to compare with Johnson's original (γ, δ) parameters, but the equivalent (μ, σ) if we just work with the new parameterization. Equivalently, combining these re-parameterized transformations we obtain:

$$x = \xi + \frac{\lambda}{1+\exp(-(\gamma' + \delta' z))} \tag{2.11}$$

a four-parameter logistic transformation of the standard normal z which reveals the transformational simplicity of the S_B distribution. A similar model is used in Item Response Theory of psychological testing (Barton and Lord 1981). With the new parameterization, the S_B pdf is given as,

$$f(x)=\frac{1}{\delta'\sqrt{2\pi}}\frac{\lambda}{(\xi+\lambda-x)(x-\xi)}e^{-\frac{1}{2}[\frac{\ln(\frac{x-\xi}{\xi+\lambda-x})-\gamma'}{\delta'}]^2} \tag{2.12}$$

We note also, that (i)′, the scaling up from N(0,1), could be dropped if we just started the construction with a N(μ, σ^2). Alternatively but equivalently, we could retain the start of the construction with N(0,1), drop the scaling up to N(μ, σ^2), but apply a simple-linear-logistic regression model, $x = 1/(1+\exp(-(\mu+\sigma\ z)))$, to N(0,1), finally scaling up to the range (ξ, ξ +λ).

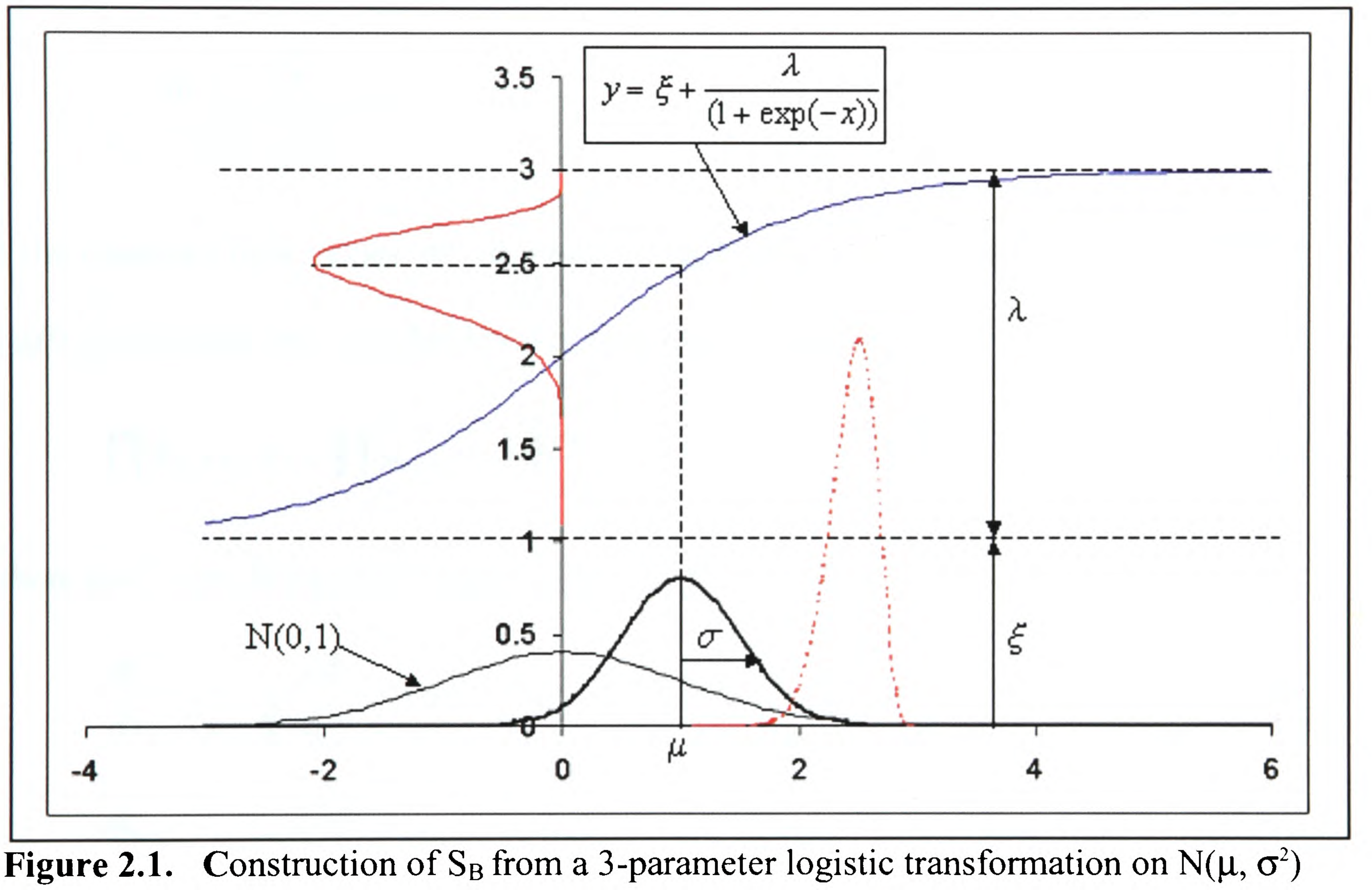


Figure 2.1. Construction of S_B from a 3-parameter logistic transformation on N(μ, σ^2)

Figure 2.1 illustrates the construction of S_B by transformation from a N(0,1) on the real *x*-axis, through N(μ, σ^2), followed by the transformation by $y = \xi + \lambda/(1+\exp(-x))$ (in blue), to the S_B on the *y*-axis (in red). The constructed S_B is also plotted (red-dashed) on the *x*-axis for comparison purposes. With such a diagram it is easy to see that S_B approaches the Log-Normal (with positive skew) as $\mu/\sigma \to -\infty$, (since the lower tail of the logistic is asymptotically exponential), while the (pseudo) Log-Normal with negative skew is obtained from $\mu/\sigma \to \infty$.

2.3 Maximum Likelihood Estimation of Johnson's S_B Distribution

For estimation it is most convenient to work with z, which is the standard normal. For the original and the new parameterizations of S_B, z is given by the following, respectively:

$$z = \gamma + \delta \ln \frac{x-\xi}{\xi+\lambda-x} \tag{2.13}$$

$$z = \frac{\ln \frac{x-\xi}{\xi+\lambda-x} - \gamma'}{\delta'} \tag{2.14}$$

If the observed data values are assumed *iid* (independently and identically distributed) from $S_B(\boldsymbol{\theta})$ distribution then the likelihood of the observed data is:

$$\prod p_{S_B}(x_i \mid \theta) = \prod p_{N(0,1)}(z_i \mid \theta) \frac{dz}{dx} \tag{2.15}$$

where $\boldsymbol{\theta}=(\xi, \lambda, \gamma, \delta)$ for (2.13) and $\boldsymbol{\theta} = (\xi, \lambda, \gamma', \delta')$ for (2.14), and

$$\frac{dz}{dx} = \frac{\lambda\delta}{(x-\xi)(\xi+\lambda-x)} \tag{2.16}$$

$$\frac{dz}{dx} = \frac{\lambda / \delta'}{(x-\xi)(\xi+\lambda-x)} \tag{2.17}$$

Using the right hand side of (2.15) gives the following minus-log-likelihood function:

$$-LL = \frac{n}{2} \ln(2\pi) + \frac{1}{2} \sum_{i=1}^{n} z_i^2 - \sum_{i=1}^{n} \ln \frac{dz_i}{dx_i} \tag{2.18}$$

Hence ML estimation of S_B amounts to the minimization of ($-LL$), with respect to the parameter vector $\boldsymbol{\theta}$.

2.4 The Variance-Covariance Matrix

After parameter estimation, the asymptotic information matrix, the variance-covariance matrix, and the correlation matrix (of the estimated parameter vector) can be computed. The $(i, j)^{\text{th}}$ element of Fisher's information matrix (of the parameter estimates) is

$$I(\hat{\theta}) = E\left(\frac{\partial L(\theta)}{\partial \theta_i \partial \theta_j}\right)_{\theta=\hat{\theta}} \tag{2.19}$$

The asymptotic variance-covariance matrix of parameter estimates, $V(\hat{\theta})$, is the inverse of the information matrix. The correlation matrix (of the parameter estimates) is obtained from the variance-covariance matrix in the usual way (Cox and Hinkley 1979).

Algebraic methods lead to very complex expressions for the correlations between the parameter estimates for the two parameterizations. It is not clear which of the parameterizations is more "well-formulated" in terms of having lower correlations between the parameter estimates. It may be that neither is generally better than the other, but that superiority depends on the data used. Hence we have to resort to an empirical evaluation of the performance of the two parameterizations.

2.4.1 Computing the Variance-Covariance Matrix

The function *nlminb* (local minimizer for smooth nonlinear functions subject to bound-constrained parameters) of S-Plus (Mathsoft 1999) is used for parameter estimation. To estimate ***V***, the following approximation was used,

$$I(\hat{\theta}) \cong Observed\left(\frac{\partial L(\theta)}{\partial \theta_i \partial \theta_j}\right)_{\theta=\hat{\theta}} \tag{2.20}$$

with the partial derivatives being evaluated symbolically using the S-Plus function "*deriv*". The approximate Information, variance-covariance and correlation matrices were then obtained. The standard errors of the parameter estimates are given by the square roots of the diagonal elements of the variance-covariance matrix.

2.5 The Forest Tree Diameter Data for the Empirical Comparison

The diameter data of 20 plots of Changbai larch plantations as provided by the Chinese Academy of Forestry were used in this empirical comparison. These plots were located at the Jingouling Farm of the Wangqing Forestry Bureau in north Changbai Mountains, north-eastern China, 130° 5′ to 130° 20′ E, 43° 17′ to 43°25′N. Figure A2.1, in the Appendix, shows the twenty diameter distributions which illustrate the range of shapes in the distributions of this empirical evaluation with the fitted frequency curves overlaid. A summary of the plot data is presented in Table 2.1, including age, plot size, number of trees in each plot, sample skewness ($\sqrt{b_1}$) and kurtosis (b_2), where the latter are defined as (see Johnson and Kotz 1970):

$$\sqrt{b_1} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2]^{\frac{3}{2}}} \tag{2.21}$$

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2]^2} \tag{2.22}$$

Figure 2.2 shows the variation of the skewness and kurtosis statistics for the twenty plots, together with some reference lines. The Weibull and Log-Normal lines represent the Weibull and Log-Normal distributions, respectively (more details presented in the next Chapter). The "lower-limit-line" is a line such that, below it, is the impossible region in terms of achievable ($\sqrt{b_1}$, b_2) pairs. The Normal is at the lowest point, (0, 1), on this lower limit line. The "Pseudo

Log-Normal" line is the counterpart of the usual positive-skew Log-Normal but with negative skew-coordinate, $\sqrt{b_1}$.

Table 2.1. Summary of age, plot size, number of trees/plot, skewness ($\sqrt{b_1}$) and kurtosis (b_2)

Plot	Age	Plot Size (m^2)	No. Trees in Plot	$\sqrt{b_1}$	b_2
301	37	775	75	0.33	3.29
302	37	775	110	0.62	2.50
303	37	1300	143	0.15	2.46
304	37	975	87	-0.15	2.31
305	37	2000	191	1.09	6.23
306	37	2000	273	0.74	3.81
307	37	2000	206	0.07	2.55
308	37	2000	124	-0.41	2.65
309	35	2500	273	0.50	2.34
310	35	2500	199	0.53	2.93
311	35	2500	184	0.57	3.23
312	35	2500	216	0.66	3.51
313	35	2025	140	0.42	4.42
314	35	2025	157	0.58	2.92
315	35	1125	148	0.92	4.81
316	35	1000	104	0.51	3.57
317	35	1000	128	0.04	2.37
318	36	1125	95	-0.05	3.19
319	36	1000	82	0.23	3.79
320	35	1000	132	0.43	2.54

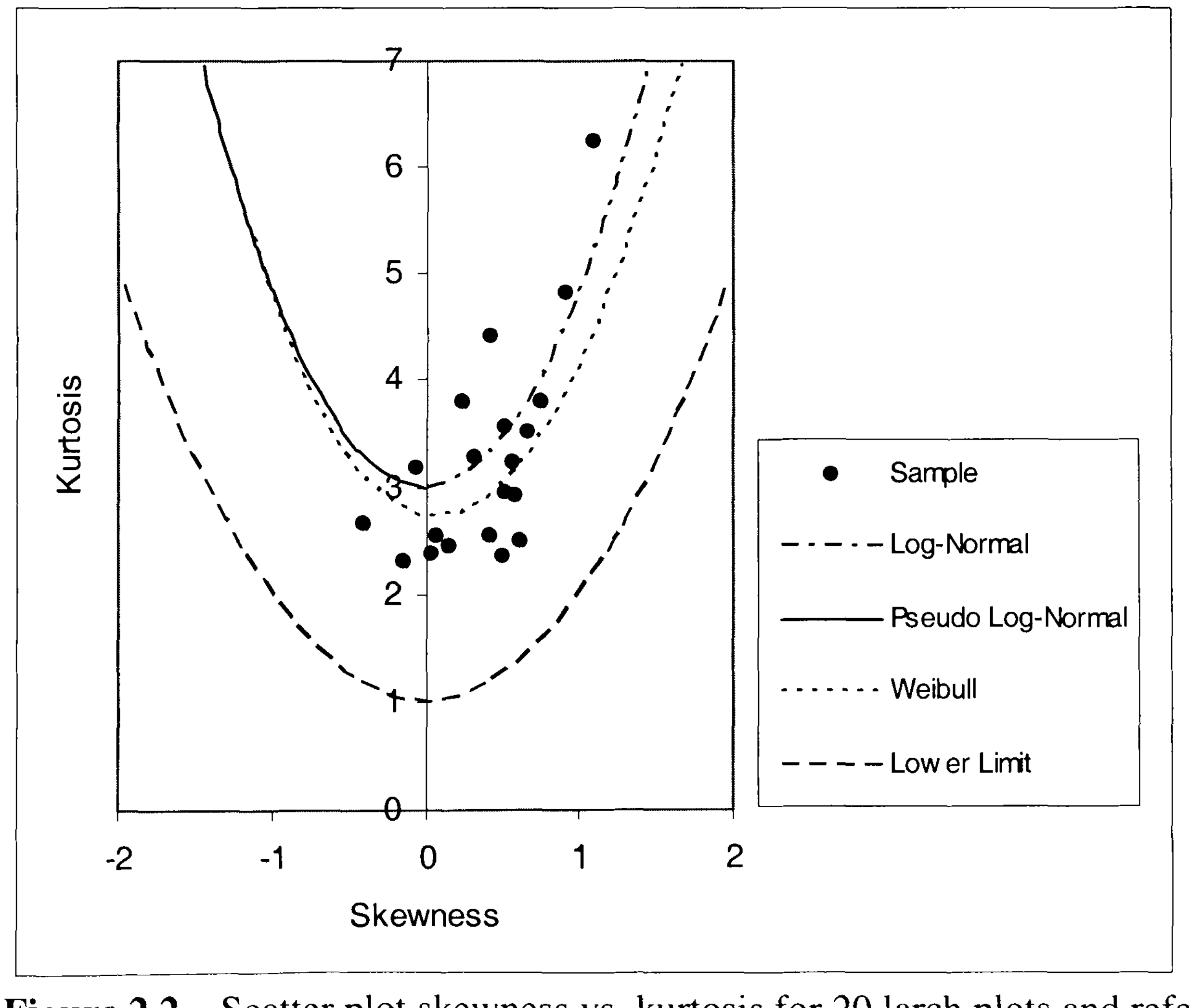


Figure 2.2. Scatter plot skewness vs. kurtosis for 20 larch plots and reference lines

The $(\sqrt{b_1}, b_2)$ coverage of Johnson's S_B is between the "lower-limit-line" and the Log-Normal/Pseudo-Log-Normal upper limit. We have seen from Figure 2.1 that this upper limit of S_B arises when the bulk of the initial generating N(0,1) is transformed by the lower tail of the logistic function, (i.e. as $\frac{\gamma'}{\delta'} \rightarrow -\infty$). From Table 2.1 or Figure 2.2, all but 3 (plot 304, 308 and 318) out of the 20 sample data exhibit positive skewness, which is in agreement with Assmann's indication that diameters usually have positively skewed distributions (Assmann 1970).

Also from Figure 2.2, four points are seen to lie well above the Log-Normal line indicating that the distributions are not of "S_B-form". Three points are just above the Log-Normal line and one point just below: the distributions concerned are of "Log-Normal" form. Two distributions are close to the Weibull line, and one is between the Weibull and Log-Normal lines, with the remaining nine lying between the Weibull and lower-limit-line. That S_B has such a variety of "forms" including those of the Log-Normal and Weibull, is probably one of the main reasons why S_B has been so much used since its introduction.

2.6 Results

2.6.1 Parameter Estimates and Standard Errors

Table 2.2 lists the estimates, standard errors (se) and |t| (t=estimate/se(estimate)) values for (γ, δ) and (γ', δ'). Estimates for ξ and λ are not listed, since, the parameterizations are the same in respect to these two parameters, and the estimates of the parameters and standard errors were found to be identical, as is necessarily so.

The main point to note in Table 2.2 is the |t| values for $\hat{\gamma}'$ are generally larger than those for $\hat{\gamma}$ (17 out of 20 cases), indicating that $\hat{\gamma}'$ is better parameterized than $\hat{\gamma}$. It is also noted that the |t| values for $\hat{\delta}$ and $\hat{\delta}'$ are identical, a result which follows from the invariance principle for ML estimates, since the parameters are reciprocals of each other.

Table 2.2. Estimates, Standard-Errors and |t|-values for (γ, δ) and (γ', δ')

Plot	$\hat{\gamma}$	se($\hat{\gamma}$)	|t|	$\hat{\delta}$	se($\hat{\delta}$)	|t|	$\hat{\gamma}'$	se($\hat{\gamma}'$)	|t|	$\hat{\delta}'$	se($\hat{\delta}'$)	|t|
301	0.99	0.73	**1.36**	1.64	0.61	2.70	-0.60	0.31	**1.92**	0.61	0.23	2.70
302	0.76	0.26	**2.91**	0.97	0.14	7.15	-0.78	0.20	**3.97**	1.03	0.14	7.15
303	0.43	0.22	**2.01**	1.26	0.25	4.99	-0.34	0.17	**2.06**	0.79	0.16	4.99
304	0.19	0.29	**0.67**	1.25	0.35	3.58	-0.15	0.25	**0.60**	0.80	0.22	3.59
305	4.33	3.72	**1.16**	2.35	0.87	2.71	-1.84	0.97	**1.90**	0.43	0.16	2.71
306	2.73	1.84	**1.48**	2.04	0.63	3.22	-1.34	0.53	**2.55**	0.49	0.15	3.22
307	0.28	0.31	**0.88**	1.70	0.43	3.93	-0.16	0.18	**0.89**	0.59	0.15	3.94
308	-0.56	0.42	**1.33**	1.29	0.38	3.42	0.44	0.24	**1.83**	0.78	0.23	3.42
309	0.84	0.09	**9.40**	0.78	0.05	15.97	-1.08	0.09	**11.33**	1.28	0.08	16.10
310	1.12	0.24	**4.59**	1.19	0.16	7.51	-0.94	0.15	**6.44**	0.84	0.11	7.49
311	1.11	0.59	**1.86**	1.41	0.40	3.51	-0.79	0.24	**3.22**	0.71	0.20	3.51
312	1.32	0.21	**6.35**	1.08	0.14	7.92	-1.22	0.13	**9.23**	0.92	0.12	7.94
313	1.77	0.75	**2.37**	1.95	0.46	4.22	-0.91	0.26	**3.51**	0.51	0.12	4.22
314	0.81	0.20	**4.08**	0.96	0.12	7.84	-0.84	0.15	**5.63**	1.04	0.13	7.84
315	3.61	4.54	**0.79**	2.19	1.27	1.72	-1.65	1.17	**1.41**	0.46	0.27	1.72
316	1.77	1.55	**1.14**	1.91	0.92	2.06	-0.93	0.45	**2.06**	0.52	0.25	2.06
317	0.21	0.14	**1.45**	0.95	0.16	5.82	-0.22	0.15	**1.44**	1.06	0.18	5.81
318	0.44	0.36	**1.24**	1.62	0.37	4.42	-0.27	0.21	**1.30**	0.62	0.14	4.42
319	1.27	1.59	**0.80**	2.59	1.12	2.30	-0.49	0.47	**1.05**	0.39	0.17	2.31
320	0.75	0.31	**2.47**	1.23	0.22	5.52	-0.61	0.19	**3.26**	0.81	0.15	5.53

Table 2.3. Correlation Coefficients Among Parameter Estimates

Plot	($\hat{\xi}$,$\hat{\gamma}$)	($\hat{\xi}$,$\hat{\gamma}'$)	d_1	($\hat{\lambda}$,$\hat{\gamma}$)	($\hat{\lambda}$,$\hat{\gamma}'$)	d_2	($\hat{\gamma}$,$\hat{\delta}$)	($\hat{\gamma}'$,$\hat{\delta}'$)	d_3
301	-0.4766	0.0168	0.4598	0.8673	-0.5459	0.3214	0.7434	0.3347	0.4087
302	-0.3245	0.0574	0.2671	0.8862	-0.7487	0.1375	0.7742	0.5019	0.2723
303	0.0794	-0.4525	-0.3731	0.3740	0.0066	0.3674	0.2621	-0.1447	0.1174
304	0.6984	-0.7907	-0.0923	-0.4640	0.5831	-0.1191	-0.4822	-0.6064	-0.1242
305	-0.8128	0.6659	0.1469	0.9971	-0.9834	0.0137	0.9456	0.8462	0.0994
306	-0.8122	0.6460	0.1662	0.9970	-0.9602	0.0368	0.9471	0.8348	0.1123
307	0.1627	-0.3753	-0.2126	0.2238	-0.0041	0.2197	0.1689	-0.0562	0.1127
308	0.9327	-0.7867	0.1460	-0.8701	0.6803	0.1898	-0.8011	-0.5677	0.2334
309	-0.1319	-0.2949	-0.1630	0.6062	-0.3354	0.2708	0.5606	-0.0263	0.5343
310	-0.3815	-0.2018	0.1797	0.8837	-0.5246	0.3591	0.7078	0.1352	0.5726
311	-0.7278	0.3932	0.3346	0.9600	-0.7771	0.1829	0.8917	0.6229	0.2688
312	-0.4656	-0.3553	0.1103	0.8848	-0.3921	0.4927	0.7290	-0.1046	0.6244
313	-0.4872	-0.0304	0.4568	0.9435	-0.6614	0.2821	0.7671	0.3069	0.4602
314	-0.3536	-0.0849	0.2687	0.8487	-0.5763	0.2724	0.7182	0.2731	0.4451
315	-0.8906	0.7834	0.1072	0.9986	-0.9825	0.0161	0.9682	0.8966	0.0716
316	-0.7537	0.4068	0.3469	0.9774	-0.8055	0.1719	0.9015	0.6270	0.2745
317	0.1986	-0.4136	-0.2150	0.1783	0.0520	0.1263	0.1231	-0.1255	-0.0024
318	0.1157	-0.3740	-0.2583	0.4333	-0.1804	0.2529	0.2994	0.0207	0.2787
319	-0.4389	0.1692	0.2697	0.8866	-0.7232	0.1634	0.7768	0.5624	0.2144
320	-0.3365	-0.0535	0.2830	0.8408	-0.5772	0.2636	0.7011	0.3367	0.3644

Note: d_1 = |corr($\hat{\xi}$, $\hat{\gamma}$)|- |corr($\hat{\xi}$, $\hat{\gamma}'$)| ; d_2 = |corr($\hat{\lambda}$, $\hat{\gamma}$)| - |corr($\hat{\lambda}$, $\hat{\gamma}'$)|; d_3 = |corr($\hat{\gamma}$,$\hat{\delta}$)| - |corr($\hat{\gamma}'$,$\hat{\delta}'$)|.

2.6.2 Correlation Coefficients Among Parameter Estimates

Table 2.3 lists the correlation coefficients among some parameter estimates. Those for $\hat{\xi}$ and $\hat{\lambda}$ are not listed, since their correlation coefficients are necessarily the same for each parameterization. It was also noticed that the correlation coefficient of ($\hat{\xi}, \hat{\delta}$) and that of ($\hat{\xi}, \hat{\delta}'$) are the same in absolute values but with opposite signs; the same applies to ($\hat{\lambda}, \hat{\delta}$) and ($\hat{\lambda}, \hat{\delta}'$). This is due to the invariance principle for ML estimates and the reciprocal relationship between parameter δ and δ'. Therefore, correlation coefficients of ($\hat{\xi}, \hat{\delta}$), ($\hat{\xi}, \hat{\delta}'$), ($\hat{\lambda}, \hat{\delta}$) and ($\hat{\lambda}, \hat{\delta}'$) are also not listed.

It was found from Table 2.3 that 14 out of 20 correlation coefficients (in absolute values) between $\hat{\xi}$ and $\hat{\gamma}'$ are less than those between $\hat{\xi}$ and $\hat{\gamma}$; that 19 out of 20 correlation coefficients between $\hat{\lambda}$ and $\hat{\gamma}'$ are less than those between $\hat{\lambda}$ and $\hat{\gamma}$, and that 18 out of 20 correlation coefficients between $\hat{\gamma}'$ and $\hat{\delta}'$ are less than those between $\hat{\gamma}$ and $\hat{\delta}$. The mean reduction in parameter estimate correlations is significant, using a paired t-test (Table 2.4). This indicates that the new parameterization considerably reduced interdependency among some parameter estimates.

Table 2.4. Paired t-test for difference in correlation coefficients (in absolute value) with old/new parameterization

Pair	Mean	Std. Dev.	t	df	p-value (1-tailed)
D_1	0.111	0.247	2.017	19	0.029
d_2	0.201	0.143	6.302	19	<0.0001
d_3	0.267	0.200	5.979	19	<0.0001

Therefore, from the above, we conclude that the new parameterization has superior statistical properties than the standard S_B parameterization, and hence (in the terms of Ross (1990)) may be considered to be a "better" parameterization than the one commonly used.

2.7 Discussion

In Table 2.2, parameter estimates of ξ and λ were intentionally left out for clarity. These two parameters are both restricted by $0 \leq \xi < D_{\min}$ and $\xi + \lambda > D_{\max}$ in fitting, where $D_{\min}$ and $D_{\max}$ are the minimum and maximum value of diameters in each plot respectively. Hafley and Schreuder (1977) set $\xi = 0$, avoiding any estimation problems associated with this bounding parameter. In this empirical study there were 6 plots (plot 301, 304, 305, 313, 318 and 319) whose estimates of ξ were zero, the lower bound for ξ. This may be taken to indicate that a lower bound of zero for parameter ξ may be not small enough as a lower bound, even though a negative value for the lower bound for ξ is "unphysical". The sample distribution for plot 319, shown in Figure A2.1, indicates that an excess relative frequency observed in the lowest diameter class (9-11cm) could be the reason for ξ being set to its lower bound of zero. Detailed examination of the other plots for which ξ is set to zero suggests the possibility that the minimal diameter measurement (about 6 cm) results in the sample being slightly truncated. This possibility, and the adaptations needed to estimation methods in such a situation, will be considered for further research.

For the 14 plots for which ξ is not set to zero, the estimated value of ξ is plotted against the minimum observed diameter, in Figure 2.3. The differences between ξ estimates and minimum diameters in plot data vary from 0.27cm to 6.51cm, with a mean of 3.03cm, and this indicates that the practice of the setting the "parameter estimate" for ξ to the minimum observed diameter minus some small constant (1.3cm, say) before fitting the other parameters (Zhou and McTague 1996, Zhang et al. 2003), may not be the best approach. However, it seems to be the case that the Maximum Likelihood method is not well suited for parameters such as ξ, which are the lower bounds of a distribution. An "order-statistic" based method for estimating ξ has been prepared by Professor Rennolls and the author.

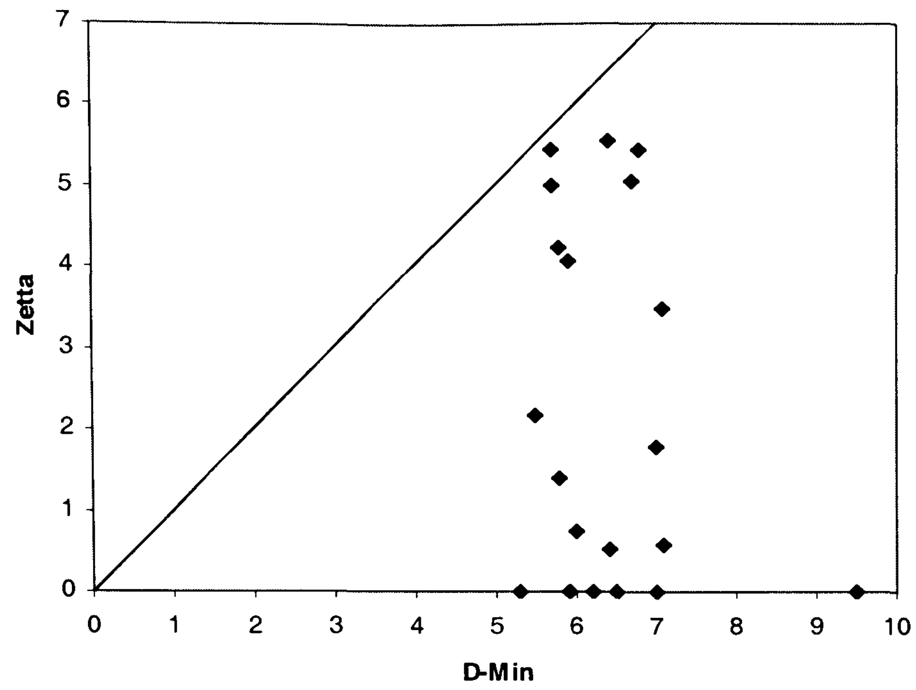


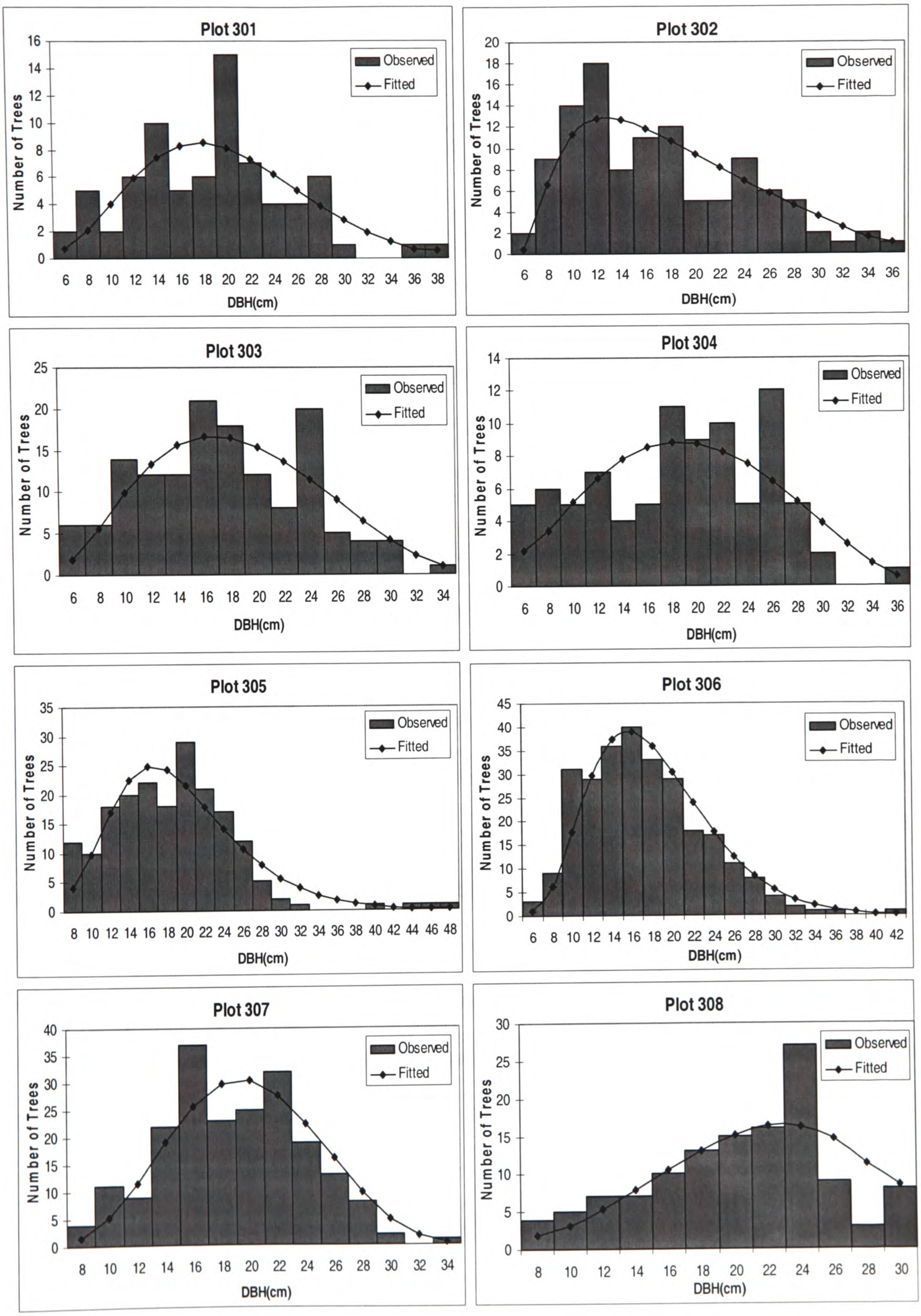
Figure 2.3. Scatter plot of $\hat{\xi}$ vs. minimum sample diameter, with reference line: $y=x$

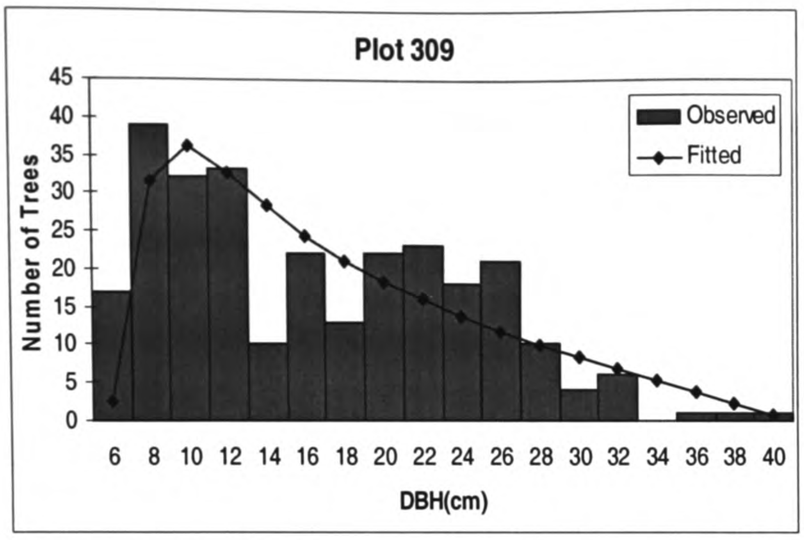
We have compared two parameterizations of Johnson's S_B in terms of standard errors of parameter estimates and correlation coefficients among parameter estimates. The fitted models under the two parameterizations have the same likelihood, since the underlying model is the same for the two parameterizations. Hence attempts to compare the "goodness of fit" of the two parameterizations would not be appropriate, or fruitful.

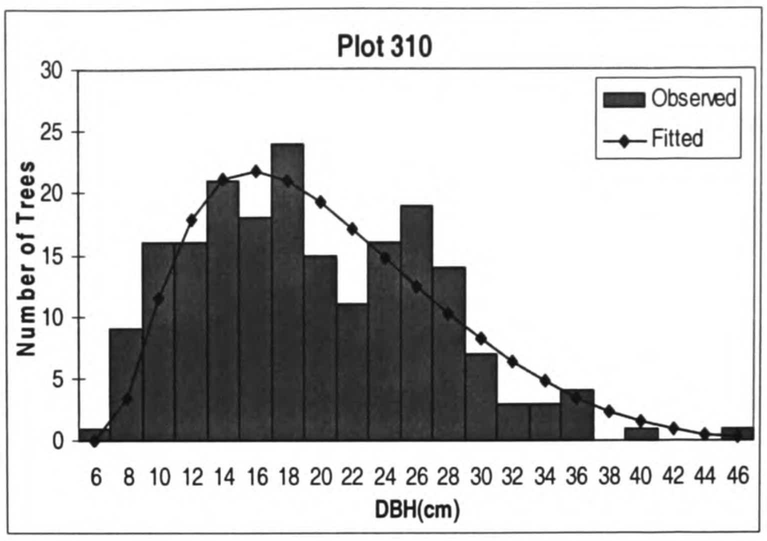
The new parameterization of Johnson's S_B, equation (2.12), can be readily applied to the bivariate version of Johnson's S_B, Johnson's S_{BB} (Johnson 1949b) for modelling the joint distribution of tree diameter and height (Schreuder and Hafley 1977, Hafley and Buford 1985, Tewari and Gadow 1997, 1999, Li et al. 2002), and to the trivariate Johnson's S_{BBB} for modelling the joint distribution of tree diameter, height and volume (Schreuder et al. 1982a, 1982b). We have used this new parameterization of Johnson's S_{BB} to characterise the joint tree diameter-height distribution (Wang and Rennolls 2005b).

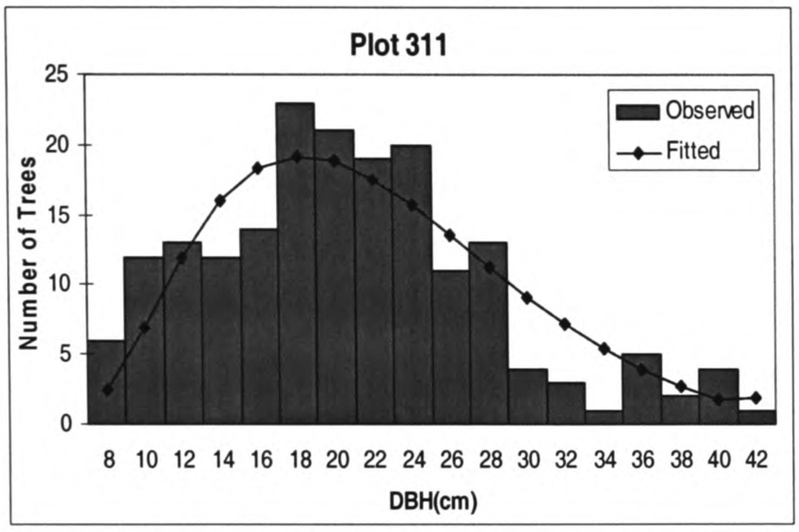
This new parameterization not only can be extended into the multivariate versions of Johnson's S_B, but is also used in the Logit-Logistic distribution (Wang and Rennolls 2005a) and its bivariate versions, in the following Chapters.

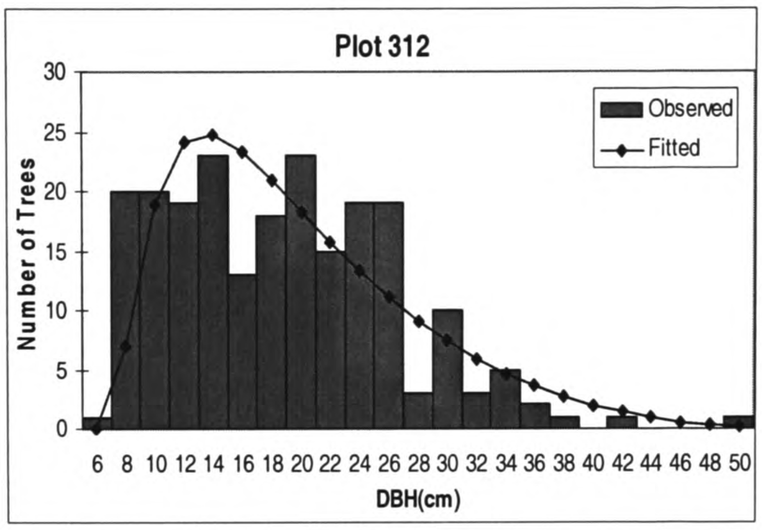
Appendix Figure A2.1. Histograms of diameter data for 20 Changbai larch sample plots and the fitted Johnson's S_B frequency curves. The mid-class diameters are given.

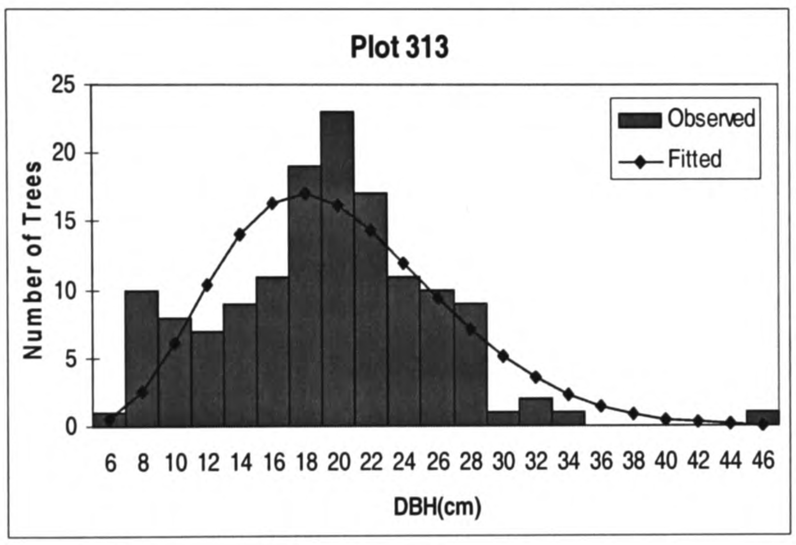


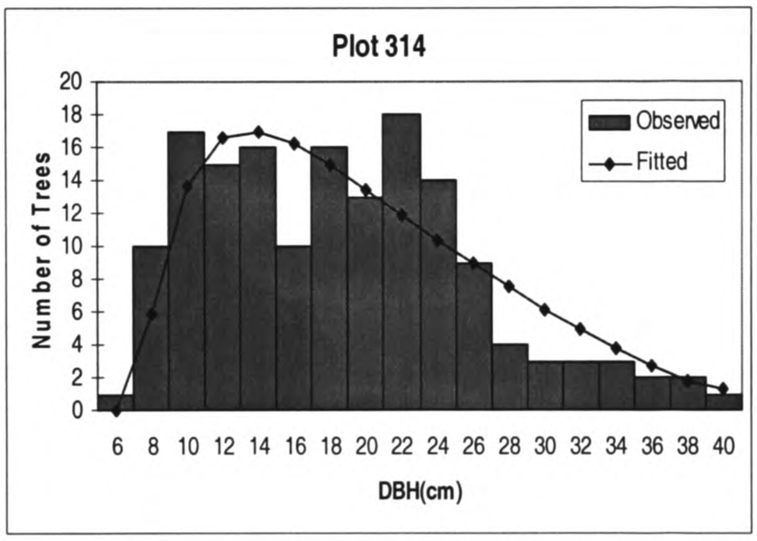

Plot 309
Observed
Fitted
Number of Trees
DBH(cm)

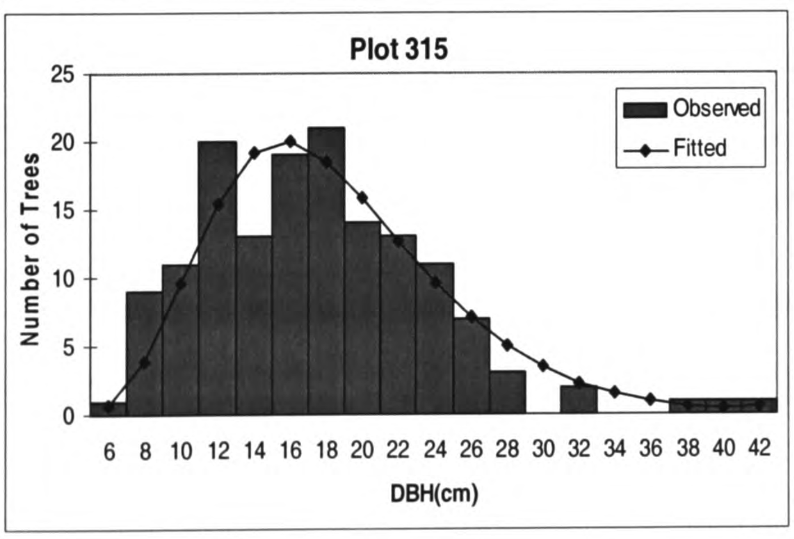

Plot 310
Observed
Fitted
Number of Trees
DBH(cm)

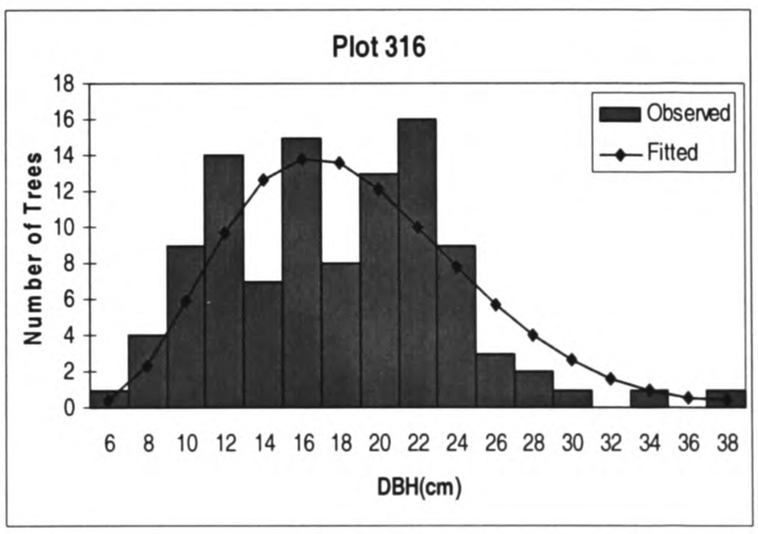

Plot 311
Observed
Fitted
Number of Trees
DBH(cm)

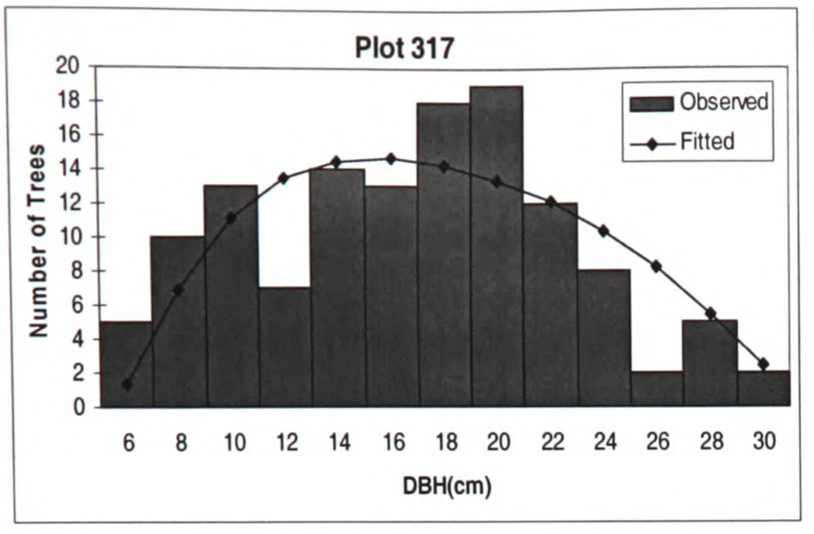

Plot 312
Observed
Fitted
Number of Trees
DBH(cm)

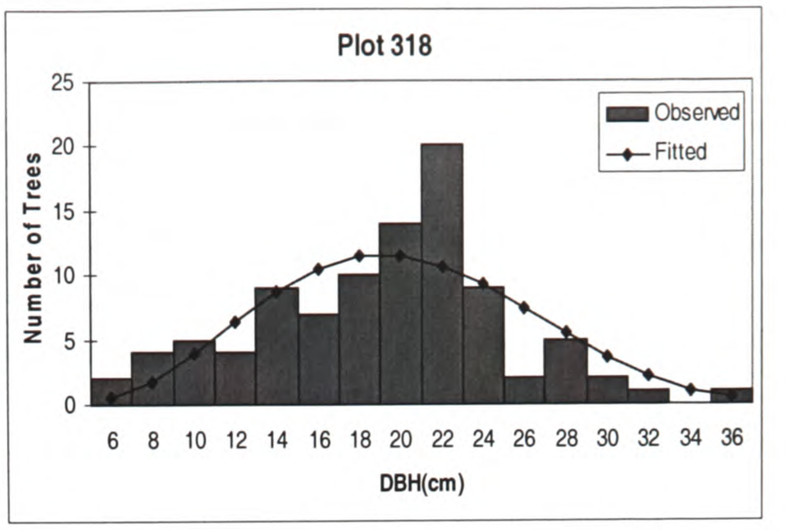

Plot 313
Observed
Fitted
Number of Trees
DBH(cm)

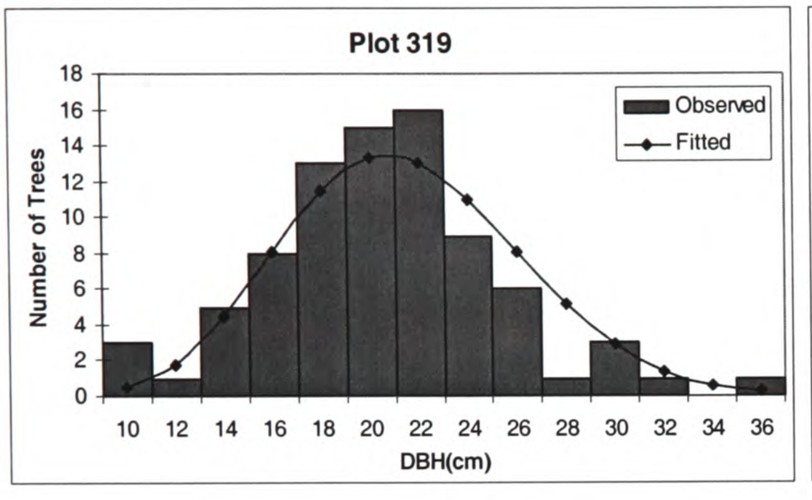

Plot 314
Observed
Fitted
Number of Trees
DBH(cm)

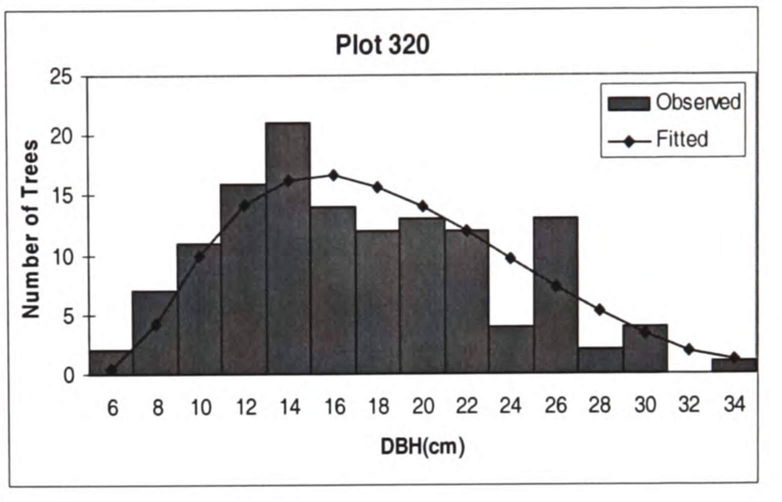

Plot 315
Observed
Fitted
Number of Trees
DBH(cm)


Plot 316
Observed
Fitted
Number of Trees
DBH(cm)


Plot 317
Observed
Fitted
Number of Trees
DBH(cm)


Plot 318
Observed
Fitted
Number of Trees
DBH(cm)


Plot 319
Observed
Fitted
Number of Trees
DBH(cm)


Plot 320
Observed
Fitted
Number of Trees
DBH(cm)

Chapter 3: The Logit-Logistic Distribution and Other New Models

Summary

The "Logistic" distribution is that tractable distribution which has a Logistic function as its cumulative distribution function: it is approximately the Normal distribution. By replacing the Normal distribution of Johnson's S_B with the Logistic distribution, a new distributional model which approximates S_B, is obtained. It is analytically tractable, and we name it the "Logit-Logistic" distribution. The "Log-Logistic" is a limiting form of the Logit-Logistic.

A 4-parameter "Generalized Weibull" distribution is introduced. It may also be seen as a generalization of a "Richards" distribution, which has been used previously in tree diameter distribution modelling.

Using the distribution "shape-plane" (with axes skew^2 and kurtosis), we compare the "coverage" properties of the Logit-Logistic and the Generalized Weibull with Johnson's S_B, the Beta, and the 3-parameter Weibull (the main distributions used in forest modelling), and the Burr III, and XII Distribution. The Logit-Logistic is found to have the largest range of shapes.

An empirical case-study of the distributional models is conducted on 107 sample-plots of Chinese fir. The Logit-Logistic performs best amongst 4-parameter models. The ($\xi \equiv 0$)-constrained S_B is best amongst ($\xi \equiv 0$)-constrained 3-parameter models.

3.1 Introduction

A wide range of probability density functions have been used in forestry to model tree diameter distributions. These have included the Log-Normal (Bliss and Reinker 1964), Gamma (Nelson 1964), Weibull (Bailey and Dell 1973, Rennolls et al. 1985), Beta (Clutter and Bennett 1965, Zöhrer 1972, Li et al. 2002), S_B (Hafley and Schreuder 1977), Logistic (Hui and Sheng 1995, Wang and Sun 1998), and the Normal. Among these models, the 4-parameter S_B and the 3-parameter Weibull models are possibly the most frequently used.

Hafley and Schreuder (1977) compared the Beta, Johnson's S_B, Weibull, Log-Normal, Gamma, and Normal distributions in terms of their coverage in the skewness-squared vs kurtosis (the β_1-β_2) plane. They concluded that Johnson's S_B gave the best performance in terms of the quality of fitting a variety of sample distributions (tree diameter and height data). Subsequently, the S_B and its bivariate version have been much used and compared with the other common distributional models (Schreuder and Hafley 1977, Hafley and Buford 1985, Knoebel and Burkhart 1991, Zhou and McTague 1996, Kamziah et al. 1999, Tewari and Gadow 1997, 1999, Li et al. 2002, Scolforo et al. 2003, and Zhang et al. 2003).

In considering how to generate new families of models of distributions, Johnson's approach (Johnson 1949) was to consider families of transformations which would result in normality. S_B is that distribution transformed into normality by (i) a linear scaling to the (0,1) range, then (ii) a Logit transformation ($y=\ln(x/(1-x))$) where $0<x<1$, and finally (iii) a linear scaling to the standard Normal. Johnson (1949) pointed out that the "transformation to normality" idea could be adapted to any other standard target distribution, such as the Laplace distribution as considered in Johnson 1954. Another choice is the Logistic distribution (the distribution obtained by using a logistic function as a cumulative distribution function (CDF)), especially considering the fact that the Logistic distribution has a shape similar to that of normal distribution. In fact, the close similarity in shape between the Logistic distribution and the normal allows, in suitable situations, to replace the Normal by the Logistic to simplify the analysis without too great discrepancies in the theory (Johnson and Kotz 1970). Berkson (1951) and Johnson and Kotz 1970, amongst many others, approximate the Normal distribution by the Logistic distribution. Mardia (1970a, b) suggested using the Logistic distribution function to approximate cumulative probability for Johnson's family of distributions when fitting contingency-type bivariate distributions. Tadikamalla and Johnson (1982) used the Logistic distribution as their standard distribution, giving what we call the Logit-Logistic distribution (following the naming convention for the Log-Normal).

The 3-parameter Weibull was developed by Weibull (1939, 1951) in studies of reliability of materials to evaluate the probability of material failure and was introduced by Bailey and Dell (1973) as a model for tree diameter distributions. The popularity of the Weibull in forestry is mainly due to two reasons. The first is its more flexibility than the Gamma and Log-Normal to take on a number of different shapes corresponding to many different observed unimodal tree diameter distributions. The second is that the CDF of the Weibull exists in closed form and thus allows for quick and estimation of the number of trees by diameter class without integration of the probability distribution function (PDF) once the parameters have been estimated. The CDF in closed form is the main advantage of the Weibull, which is represented by a line in the skewness-kurtosis shape plane, over the more flexible S_B and Beta, both of which cover an area in the plane. There are already many methods to add an additional parameter to the 3-parameter Weibull model to increase its flexibility. One natural way is to exponentiated its CDF, resulting what we call the Generalized Weibull.

Both the Logit-Logistic and the Generalized Weibull are flexible in shape and have simple CDF, which are thus possibly promising in modelling tree diameter distributions. There are also several other distributional models from the Burr system, namely, the Burr III, Burr XI and Burr IV, which are flexible and simple as well.

This chapter introduces the Logit-Logistic, the generalized Weibull, and the Burr III, XII, and IV models to forest diameter distribution modelling, and compares their performance with the other main distributions that have been used. We note that we did the work on the development and fitting of the Logit-Logistic and the generalized Weibull before we discovered the respective precedents of Tadikamalla and Johnson (1982) and Mudholkar and Srivastava (1993).

3.2 The Main Distributional Models Considered

Seven 4-parameter distributions were considered in detail in this study. The Beta and Johnson's S_B have been much used in forest distributional studies, because of their flexibility of distributional form (or shape), and their ability to represent equally well positive and negative skew distributions. We introduce the Logit-Logistic, generalized Weibull, and the Burr III, XII and IV for comparison. The seven distributions are defined in the following sections. However, as we discovered the Burr III and IV during the last stage of writing this thesis, we will not compare the Burr IV with the others in terms of its flexibility in the skewness and kurtosis shape-plane, because we do not know its coverage in the shape-plane, but included both of them in our empirical study.

3.2.1 The S_B in a New Parameterisation

Johnson's distribution system, based on "transformation to normality", includes the Log-Normal system (S_L), by use of the log transformation; the bounded system (S_B) by use of the Logit transformation; and the unbounded system (S_U) by use of the inverse hyperbolic sine transformation. In chapter 2 (see also Rennolls and Wang 2005), we present an inverse transformational definition of Johnson's S_B and a new parameterization. The S_B distribution is obtained, for X say, by a 4-parameter logistic transformation of a standard Normal variate, Z:

$$x = \xi + \frac{\lambda}{1 + \exp(-(\mu + \sigma z))} \qquad (3.1)$$

where, $-\infty < z < \infty$, $\xi \leq x \leq \xi + \lambda$. The S_B PDF is given by,

$$f_X(x) = f_Z(z)\left(\frac{dx}{dz}\right)^{-1} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \cdot \frac{\lambda}{\sigma(\xi + \lambda - x)(x - \xi)} \qquad (3.2)$$

where z is given by the inversion of (3.1) as,

$$z = \frac{\ln\left\{\frac{x - \xi}{\xi + \lambda - x}\right\} - \mu}{\sigma} \qquad (3.3)$$

There is no explicit form for the S_B CDF, in contrast to the Logit-Logistic distribution of the next section.

3.2.2 The Logit-Logistic Distribution Model

Substitution of the Normal in the Johnson system by the Logistic gives an alternative, but similar, set of families of distribution models to those based on the Normal. For the Logit transformation we call the resulting distribution the Logit-Logistic distribution rather than the L_B as in Tadikamalla and Johnson (1982), in analogy to the naming of the Log-Normal. Replacing the standard normal z in equation (3.1) with that of the standard Logistic, L(0,1), results in the Logit-Logistic (LL) distribution.

The standard Logistic distribution (SL) has CDF given by

$$F_{Z_{SL}}(z) = \frac{1}{1 + e^{-z}} \qquad (3.4)$$

and its PDF is,

$$f_{Z_{SL}}(z) = \frac{e^{-z}}{(1 + e^{-z})^2} = \frac{1}{e^z + e^{-z} + 2} \qquad (3.5)$$

The standard Logistic is symmetric and has a standard deviation of $\pi/\sqrt{3}(\approx 1.82)$, so is not a standardized distribution. We follow convention and work with the standard Logistic rather than the standardized Logistic. The kurtosis of the SL is 4.2. The transformation given by (3.1) relates z_{SL} to x_{LL}. It follows that:

$$F_{X_{LL}}(x) = F_{Z_{SL}}(z) = \frac{1}{1 + e^{\frac{\mu}{\sigma}} \left(\frac{x-\xi}{\xi+\lambda-x}\right)^{-\frac{1}{\sigma}}} \tag{3.6}$$

$$\begin{aligned} f_{X_{LL}}(x) &= f_{Z_{SL}}(z) \cdot \frac{\lambda}{\sigma(\xi+\lambda-x)(x-\xi)} \\ &= \frac{\lambda}{\sigma} \frac{1}{(x-\xi)(\xi+\lambda-x)} \cdot \frac{1}{e^{-\frac{\mu}{\sigma}} \left(\frac{x-\xi}{\xi+\lambda-x}\right)^{\frac{1}{\sigma}} + e^{\frac{\mu}{\sigma}} \left(\frac{x-\xi}{\xi+\lambda-x}\right)^{-\frac{1}{\sigma}} + 2} \end{aligned} \tag{3.7}$$

The fact that the Logit-Logistic CDF exists in a simple invertible closed form facilitates its practical application, compared with the use of S_B for which no closed form of CDF exists.

3.2.3 The Beta Distribution

The Beta PDF is,

$$f(x) = \frac{1}{B(p,q)} \frac{(x-a)^{p-1}(b-x)^{q-1}}{(b-a)^{p+q-1}} \tag{3.8}$$

where, $a \le x \le b$, $p, q > 0$ are two shape parameters.

3.2.4 The Generalized Weibull Distribution

The 3-parameter Weibull CDF is,

$$F(x) = (1 - e^{-(\frac{x-a}{b})^c}) \tag{3.9}$$

where, $x \ge a$, $b > 0$, $c > 0$. By adding an "exponentiated" parameter, k ($k > 0$), it can be generalized as,

$$F(x) = (1 - e^{-\left(\frac{x-a}{b}\right)^c})^k \tag{3.10}$$

Its PDF is given as,

$$f(x) = \frac{ck}{b}\left(\frac{x-a}{b}\right)^{c-1} e^{-\left(\frac{x-a}{b}\right)^c} \left(1 - e^{-\left(\frac{x-a}{b}\right)^c}\right)^{k-1} \tag{3.11}$$

The Generalized Weibull (GW) distribution is reversed J-shaped when $ck \leq 1$ and unimodal when $ck > 1$. The original reason for introducing this Generalized Weibull distribution was to be able to compare distributional models each having the same number of parameters as S_B and the Logit-Logistic. Mudholkar and Srivastava (1993) termed this model the "exponentiated Weibull". It has been studied extensively in the statistics (Mudholkar et al. 1995, Mudholkar and Hutson 1996, Jiang and Murthy 1999, Singh et al. 2002, Nassar and Eissa 2003, 2004). The applications of the GW distribution in reliability and survival studies were illustrated by Mudholkar et al. (1995).

This generalized Weibull includes one form of a "Chapman-Richards distribution" when $c = 1$, which has been used in forest diameter distribution modelling (Ishikawa 1987, 1991, 1996, 1997, 1998). The Chapman-Richards distribution corresponds to the well known Chapman-Richards growth function in forestry (Pienaar and Turnbull 1973), and was recognized as a distributional model by Ahuja and Nash (1967). Currently it is known as the "Generalized Exponential" or the "Exponentiated Exponential" (Gupta and Kundu 1999) and has received much attention (Gupta and Kundu 2001a, b, 2002, 2003, 2004, Ragab 2002, 2004, Kundu et al. 2005). Both the Chapman-Richards (generalized exponential) and the Weibull generalize the exponential distribution, but in different ways.

Another specific distribution with $c = 2$ is the Burr X distribution (Burr 1942).

3.2.5 The Burr III, XII, and IV Distributions

Burr (1942) introduced 12 families of distributions. All the 12 families CDFs exist in closed form, as Burr's objective was to fit cumulative distributions rather than density functions to frequency data, to avoid the problems of numerical integration which are encountered when probabilities are evaluated from Pearson curves. Among these families, Types III and XII are the simplest functionally and thus the most attractive for statistical modelling. Originally, only the Burr XII was studied in detail (Burr 1942, 1968, 1973, Burr and Cislak 1968, Rodriguez 1977), which is the reason we did not pay attention to the Burr III and Burr IV in the earlier stage of this study.

3.2.5.1 The Burr XII

The Burr XII CDF is,

$$F(x) = 1 - \frac{1}{[1 + (\frac{x-a}{b})^c]^k} \tag{3.12}$$

where, $x \geq a$, $c > 0$, $k > 0$. The PDF is given,

$$f(x) = \frac{kc}{b}(\frac{x-a}{b})^{c-1}[1 + (\frac{x-a}{b})^c]^{-k-1} \tag{3.13}$$

The 4-parameter Burr XII distribution does not seem to have been previously evaluated for forest modelling and is included in this comparative study because of the simple form of its CDF.

3.2.5.2 The Burr III

The Burr III CDF is,

$$F(x) = \frac{1}{(1 + (\frac{x-a}{b})^{-c})^k} \tag{3.14}$$

where, $x \geq a$, $c > 0$, $k > 0$. Its PDF is given as,

$$f(x) = \frac{kc}{b}(\frac{x-a}{b})^{-c-1}[1 + (\frac{x-a}{b})^{-c}]^{-k-1} \tag{3.15}$$

This family was studied in detail later on than the Burr XII (Rodriguez 1977, Tadikamalla 1980), which is more flexible than the Burr XII. Lindsay et al. (1996) investigated the Burr III in modelling diameter distributions. They found the Burr III outperforms the Weibull in fitting tree diameter distributions. The Burr XII and Burr III are related, in the sense that if X has a Burr XII distribution with parameters c and k, then $1/X$ has a Burr III distribution with parameters c and k (Tadikamalla 1980). However, we note that the relationship between the two models is only applicable to their standard forms (that is, with $a = 0$ and $b = 1$). If we consider the location and scale parameters explicitly, very interestingly, we found that if X ($a \le x < \infty$) has a Burr XII distribution with parameters c and k, then $Y = 1/X$ ($0 \le y < 1/a$) will probably have a Logit-Richards distribution with parameters b, c and k, which is being consideration for further research. That is,

$$\Pr(Y < y)=\Pr(1/X < y)=\Pr(X>1/y)=1-\Pr(X < 1/y)$$

$$=1-\{1-\frac{1}{[1+(\frac{1/y-a}{b})^c]^k}\}=\frac{1}{[1+(\frac{1/y-a}{b})^c]^k}$$

$$=\frac{1}{\{1+[\frac{b}{a}(\frac{y}{1/a-y})]^{-c}\}^k} \tag{3.16}$$

If we take $a = 1$ in (3.16), then $0 \le y < 1$ and (3.16) becomes,

$$F(y)=\frac{1}{\{1+[b(\frac{y}{1-y})]^{-c}\}^k} \tag{3.17}$$

Equation (3.17) is the standard form of the Logit-Richards, which includes the Logit-Logistic and the Burr IV ((3.20) in the subsequent section) as special cases with $k = 1$ and $b = 1$, respectively. It should be noted that the scale parameter b in the Burr XII distribution now becomes a shape parameter.

3.2.5.3 The Burr IV

The Burr IV CDF without location (minimum parameter) and scale parameters is,

$$F(y)=\frac{1}{[1+(\frac{c-y}{y})^{1/c}]^k} \tag{3.18}$$

where $0 < y < c$, $c > 0$, $k > 0$. Parameter c acts as a maximum location parameter and also a shape parameter as well. With $x = a + b\,y$ and some re-parameterizations, the Burr IV CDF is given as,

$$F(x) = \frac{1}{[1 + (\frac{x-a}{a+b-x})^{-c}]^k} \tag{3.19}$$

where $a < x < a + b$, and c is no longer related to location or range. Note that the standard form of the Burr IV CDF (where $a = 0$ and $b = 1$) becomes,

$$F(x) = \frac{1}{[1 + (\frac{x}{1-x})^{-c}]^k} = \frac{1}{[1 + (\frac{1-x}{x})^{c}]^k} \tag{3.20}$$

and is obviously different from that originally developed by Burr. However, we still term it the "Burr IV". Its PDF is given as,

$$f(x) = \frac{kcb}{(x-a)(a+b-x)}(\frac{x-a}{a+b-x})^{-c}[1 + (\frac{x-a}{a+b-x})^{-c}]^{-k-1} \tag{3.21}$$

Interestingly, the Burr IV CDF shows some similarity to the Logit-Logistic CDF, both being special cases of a more general distributional model, the Logit-Richards.

3.3 Comparison of the Range of Shapes of the Distributional Models

The mean and the standard deviation of a distribution are location and scale parameters and may be used to produce a standardized distribution with mean zero and standard deviation one. The shape of the distribution is therefore characterised by the (standardized) distribution's higher order moments. Usually skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) are taken to be adequate to represent distribution shape. Both are given as moment ratios,

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{1.5}}, \quad \beta_2 = \frac{\mu_4}{\mu_2^2} \tag{3.22}$$

where μ_k is the k^{th} central moment. Skewness is used for describing a departure from symmetry about the mean where negative values indicate a distribution with a long tail to the left (left-skewed) and positive values a long tail to the right (right-skewed). Kurtosis is generally considered to be a relative measure of flatness or peakedness of a distribution.

Although skewness and kurtosis do not uniquely define a distribution, they do characterise the general shape of the distribution and suggest potential models for consideration for a particular problem. If the data has skewness and kurtosis values outside the coverage of a particular model then the model can never fit the data well.

A graph of (β_1, β_2) [≡ (skewness^2, kurtosis)] is commonly used to demonstrate the range of shapes covered by various statistical distributions. Such a graph is very helpful in envisaging the representational strengths and weaknesses of distributions. All the 4-parameter distributional models considered have two shape parameters (μ and σ for the S_B and LL, p and q for the Beta, and c and k for the others), each of which covers an area in the (β_1, β_2) shape plane. Before we start to compare their areal coverage in the shape plane, we firstly introduce some shape lines which are used to delimit the boundaries of the 4-parameter models. These lines correspond to the 3-parameter distributions.

3.3.1 3-Parameter Distributions ("Lines" in Figure 3.1)

As the location parameter and the scale parameter do not affect the distribution shape, for simplicity, only the shape parameter ($c > 0$) is presented in the functional form of CDF or PDF for each of the 3-parameter distribution. The only exception is for the Log-Normal.

The Gamma (Pearson Type III) distribution is given by its PDF as,

$$f(x) = \frac{x^{c-1} e^{-x}}{\Gamma(c)} \tag{3.23}$$

where Γ represent the gamma function.

The Log-Normal is given by its PDF as,

$$f(x) = \frac{\delta}{\sqrt{2\pi}(x-a)} e^{-\frac{1}{2}(r+\delta \ln(x-a))^2} \tag{3.24}$$

where $x > a$, $\delta > 0$, and δ is the shape parameter.

The Log-Logistic distribution has the CDF,

$$F(x) = \frac{1}{1+x^{-c}} \tag{3.25}$$

The Weibull distribution has the CDF,

$$F(x) = 1 - e^{-x^c} \tag{3.26}$$

The Burr II distribution has the CDF,

$$F(x) = \frac{1}{(1+e^{-x})^c} \tag{3.27}$$

The Burr II distribution is one generalization of the Logistic distribution, and we call it the "Richards" distribution.

All of these one shape parameter distributions are represented by lines in the shape plane, they are,

Gamma Line, $2\beta_2 - 3\beta_1 - 6 = 0$ (3.28)

Log-Normal Line, $\beta_1 = (w-1)(w+2)^2$

$\beta_2 = w^4 + 2w^3 + 3w^2 - 3$ (3.29)

where, $w = e^{\delta^{-2}}$.

Log-Logistic Line, $\beta_1 = \dfrac{(B_3 - 3B_1B_2 + 2B_1^3)^2}{(B_2 - B_1^2)^3}$

$$\beta_2 = \frac{(B_4 - 4B_1B_3 + 6B_1^2B_2 - 3B_1^4)}{(B_2 - B_1^2)^2} \tag{3.30}$$

where $B_i = B(1 - i/c, 1 + i/c)$, $i = 1,2,3,4$, $c>4$, B is the beta function.

Weibull Line, $\sqrt{\beta_1} = \dfrac{\Gamma(1+\frac{3}{c}) - 3\Gamma(1+\frac{2}{c})\Gamma(1+\frac{1}{c}) + 2\Gamma^3(1+\frac{1}{c})}{[\Gamma(1+\frac{2}{c}) - \Gamma^2(1+\frac{1}{c})]^{\frac{3}{2}}}$

$$\beta_2 = \frac{\Gamma(1+\frac{4}{c}) - 4\Gamma(1+\frac{3}{c})\Gamma(1+\frac{1}{c}) + 6\Gamma(1+\frac{2}{c})\Gamma^2(1+\frac{1}{c}) - 3\Gamma^4(1+\frac{1}{c})}{[\Gamma(1+\frac{2}{c}) - \Gamma^2(1+\frac{1}{c})]^2} \tag{3.31}$$

Burr II Line, $$\sqrt{\beta_1} = \frac{2[\zeta(3,1) - \zeta(3,c)]}{[\zeta(2,1) + \zeta(2,c)]^{\frac{3}{2}}}$$

$$\beta_2 = \frac{6[\zeta(4,1) + \zeta(4,c)]}{[\zeta(2,1) + \zeta(2,c)]^2} + 3 \qquad (3.32)$$

where $\zeta(s,a) \equiv \sum_{n=0}^{\infty} \frac{1}{(n+a)^s}$ is the Hurwitz zeta function ($s > 1$).

Another special line is the so-called "lower limit" line, that is,

"Lower Limit" Line, $$\beta_2 - \beta_1 - 1 = 0 \qquad (3.33)$$

It is called the "lower limit" line, as for a given value of skewness β_1, there is a lower bound on the possible value of kurtosis β_2, which is determined by the equation (3.33). In other word, for any distributions, we have $\beta_2 \geq \beta_1 + 1$.

The literature for these lines are referred to Johnson and Kotz (1970), Ord (1972), and Ahuja and Nash (1967).

3.3.2. Comparison in the (β_1, β_2) Region

3.3.2.1 Logit-Logistic, S_B, Beta, and Burr XII

Figure 3.1 shows the known "shape domains" for some of the distributions considered in this study (Johnson and Kotz 1970, Ord 1972). The placement of the axes (with the y-axis for β_2 plotted downwards) in Figure 3.1 is conventional in such studies, but has the effect that what is called "the lower-limit line", the limit for all distributions, is shown as the upper line in Figure 3.1. Distributions on the lower-limit line are discrete (two) probability-mass distributions, the asymptotic limit of U-shape distributions. No distributions can exist above or to the right of it in Figure 3.1.

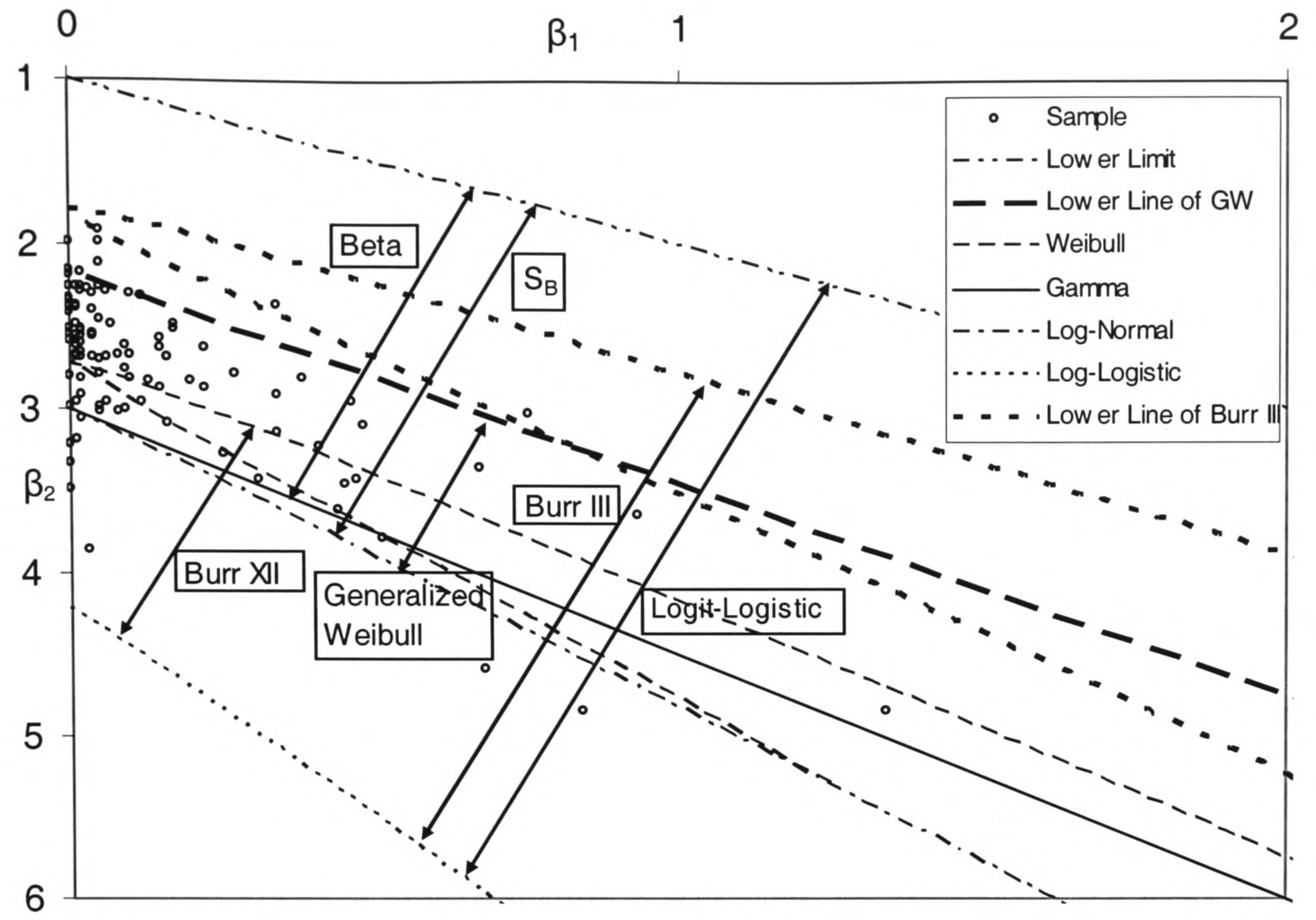


Figure 3.1. (β_1, β_2) of Distribution Families

The Normal distribution is represented by the point at $(\beta_1, \beta_2) = (0, 3)$. From this point two lines emanate, one corresponding to the 1-dimensional family of shapes covered by Gamma distributions, and the other, further out, corresponds to the Log-Normal. It can be seen that the 2-dimensional family of shapes covered by the Beta distributions is the domain between the lower-limit-line and the Gamma line. The S_B distributions cover the shape domain between the lower-limit-line and the Log-Normal line, the asymptotic limit of S_B. Hence, in terms of coverage of the positive (β_1, β_2) quadrant, the S_B shape-domain encompasses that of the Beta, and is more extensive. The Burr XII distribution covers the domain between the Weibull line, the asymptotic lower limit of Burr XII, and Log-Logistic line. The Burr III shape domain encompasses that of the Burr XII and extends much more towards the lower-limit-line. The Logit-Logistic shape-domain stretches from the lower-limit-line to the Log-Logistic line (the limit of the Logit-Logistic as $\mu \to -\infty$ and σ remains bounded), intersecting the β_2 axis at $(\beta_1, \beta_2) = (0, 4.2)$, which corresponds to the Logistic distribution. Hence the Logit-Logistic shape domain encompasses and extends (dominates) the Beta, S_B and Burr XII: LL $\supset$ S_B $\supset$ Beta; LL $\supset$ Burr III $\supset$ Burr XII, where the inclusion refers to the shape domains. However, both the Burr XII and Burr III distributions as parametrically specified have limited range in the negative skew domain, a problem which may be overcome by using a reflection transformation which will change the sign of the odd moments.

We note that the popular 3-parameter Weibull is represented by a line (a 1-dimensional family of shapes) which has two branches with a fold at the β_2 =2.7 point on the β_2 axis. This 3-parameter Weibull may be seen from Figure 3.1 to be capable of having shapes close to both the Gamma and Log-Normal distributions. However, the 1-dimensional shape coverage of all these 3-parameter models is encompassed by the 2-dimensional shape coverage of the Logit-Logistic, S_B , Burr XII, and the Burr III.

3.3.2.2 The Generalized Weibull

The generalized Weibull has a 2-dimensional shape-space coverage in Figure 3.1, in contrast to the 1-dimensional coverage line for the 3-parameter Weibull. We do not have analytical forms of the boundaries in shape space of the generalized Weibull, but the shape-space coverage as indicated by simulation methods is shown in Figure 3.2. It seems from Figure 3.2 that the GW upper-limit is fairly close to the Log-Normal line, the same upper-limit as S_B, but not as high as the Log-Logistic upper-limit of the Logit-Logistic. Using the simulated limit cases in Figure 3.2, the lower-limit line for the Generalized Weibull may be approximated by $\beta_2 = 2.17+1.29\beta_1$. This line is also indicated in Figure 3.1 for comparing the GW with the other distributions. Significantly, the GW cannot get near the lower-limit line and thus does not have the lower-limit shape-space coverage of the Logit-Logistic, S_B, and the Beta, all of which extend to the Lower-Limit line.

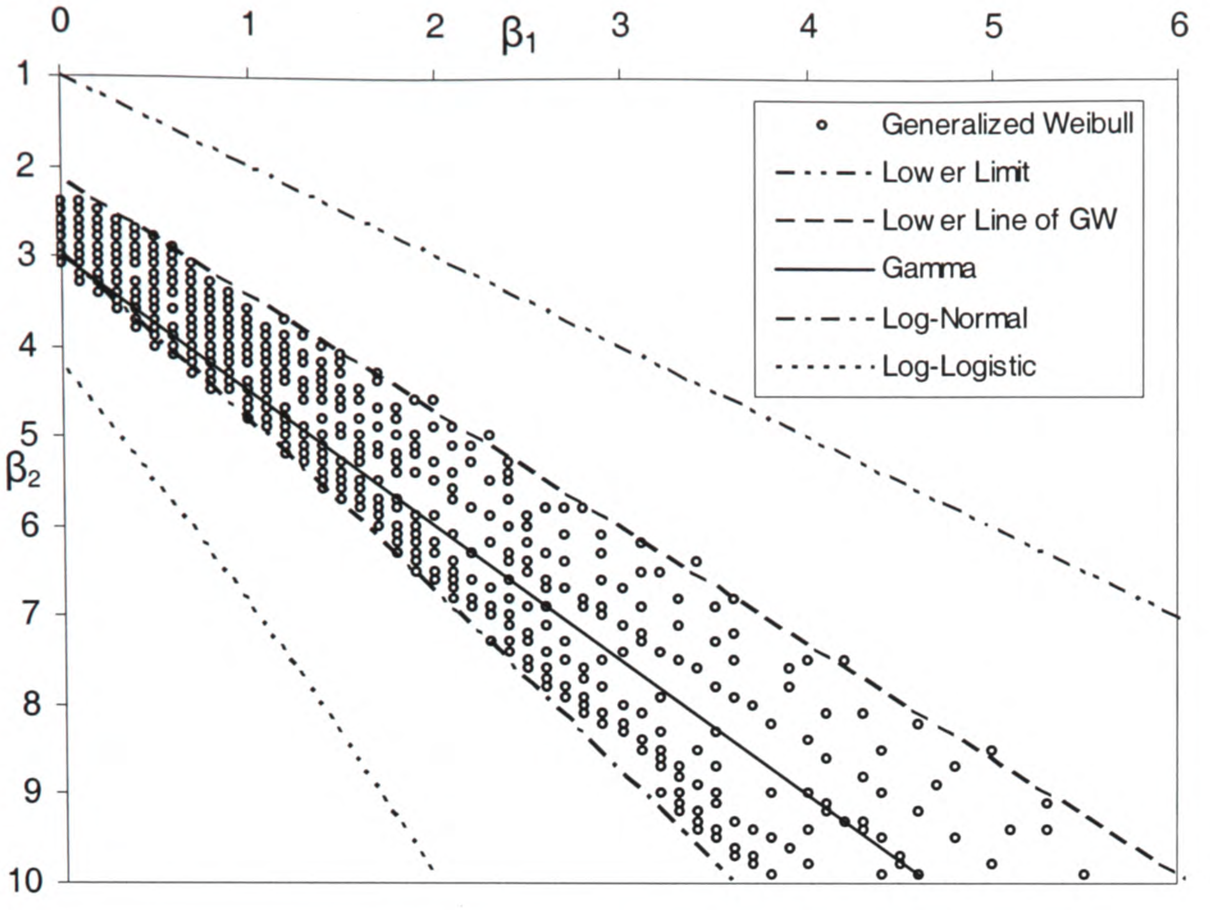


Figure 3.2. (β_1, β_2) coverage of the Generalized Weibull Distribution

3.3.3 Comparison in the ($\surd\beta_1$, β_2) Region

Traditionally, the parameter choice (β_1, β_2) = (skewness^2, kurtosis) is chosen, possibly because in this parameterisation, the shape-domain boundaries are linear, or close to it. However, this type of moment-ration diagrams suffers from the defect that information on the sign of μ_3 is lost. If we wish a diagrammatic representation of shape-space which retains this information, then we may use ($\surd\beta_1$, β_2) in stead of (β_1, β_2). In the following diagrams of $\surd\beta_1$ and β_2, $\surd\beta_1$ is plotted on the horizontal axis and β_2 on the vertical axis, in the usual way.

Figure 3.3, 3.4, and 3.5 show the results for the "symmetric" distributions (Logit-Logistic, S_B and Beta), the asymmetric Burr XII and III, and the GW, respectively.

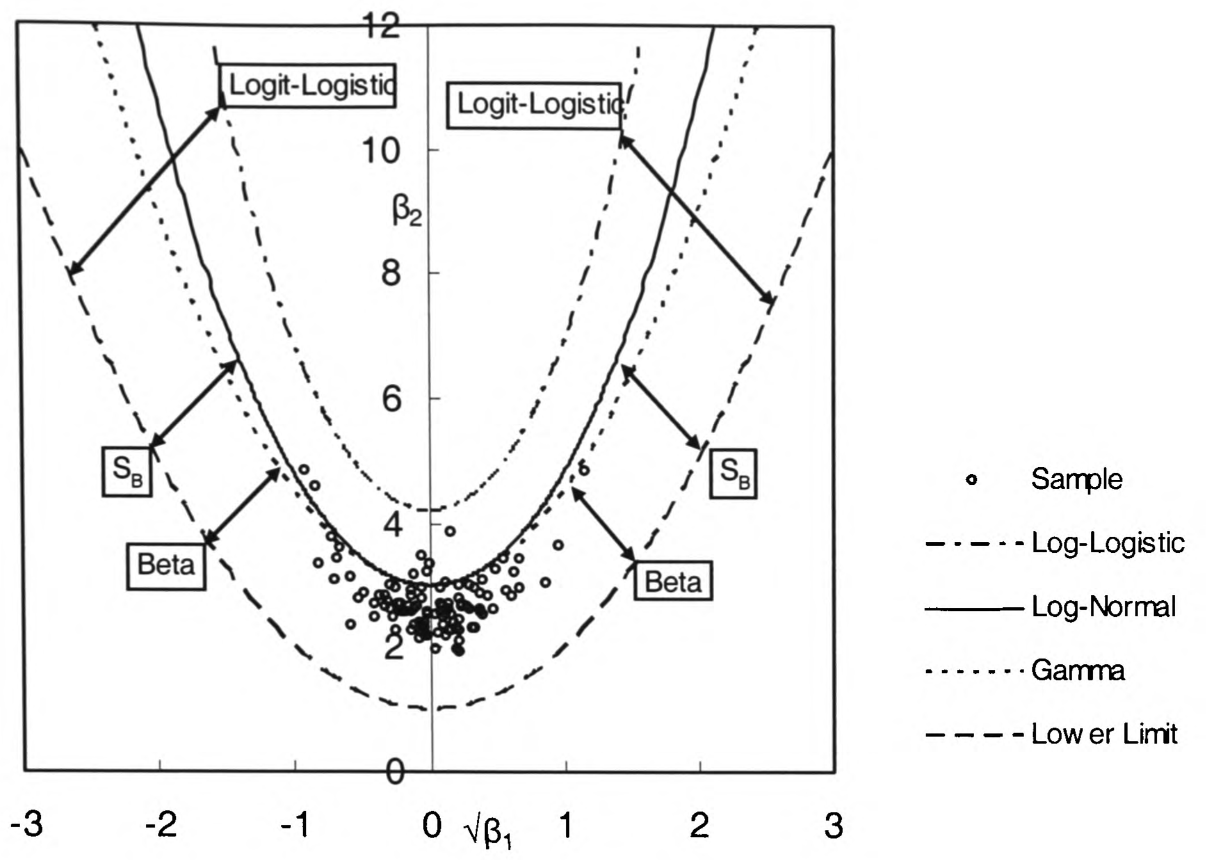


Figure 3.3. ($\sqrt{\beta_1}$, β_2) of the "skew-symmetric" Distribution Families

Figure 3.3 shows that the symmetric coverage by Beta, S_B and Logit-Logistic in both positive and negative skewness areas. This is obvious for the Beta (from application of the transformation $y = 1 - x$ to the standard Beta) and clear from the diagrammatic representation of the S_B and Logit-Logistic (see Chapter 2). The asymptotic limits of the Beta, S_B, and the Logistic-Logistic are the Gamma, Log-Normal and Log-Logistic of both positive and negative skewness, respectively. The Gamma, Log-Normal and Log-Logistic as specified are all limited to shapes that have positive skewness. However, a linear transformation with negative slope will yield distributions with negative skewness. Equivalently, a reflection and a change of scale will achieve the same result. The negative skew forms of the Gamma, Log-Normal and Log-Logistic distributions are, rather surprisingly, not mentioned in the statistical distribution literature. We may have called such negative-skew counterparts "pseudo-Gamma", "pseudo-Log-Normal", and "pseudo-Log-Logistic" respectively, but not shown them in Figure 3.3 for line clarity with black-white drawing.

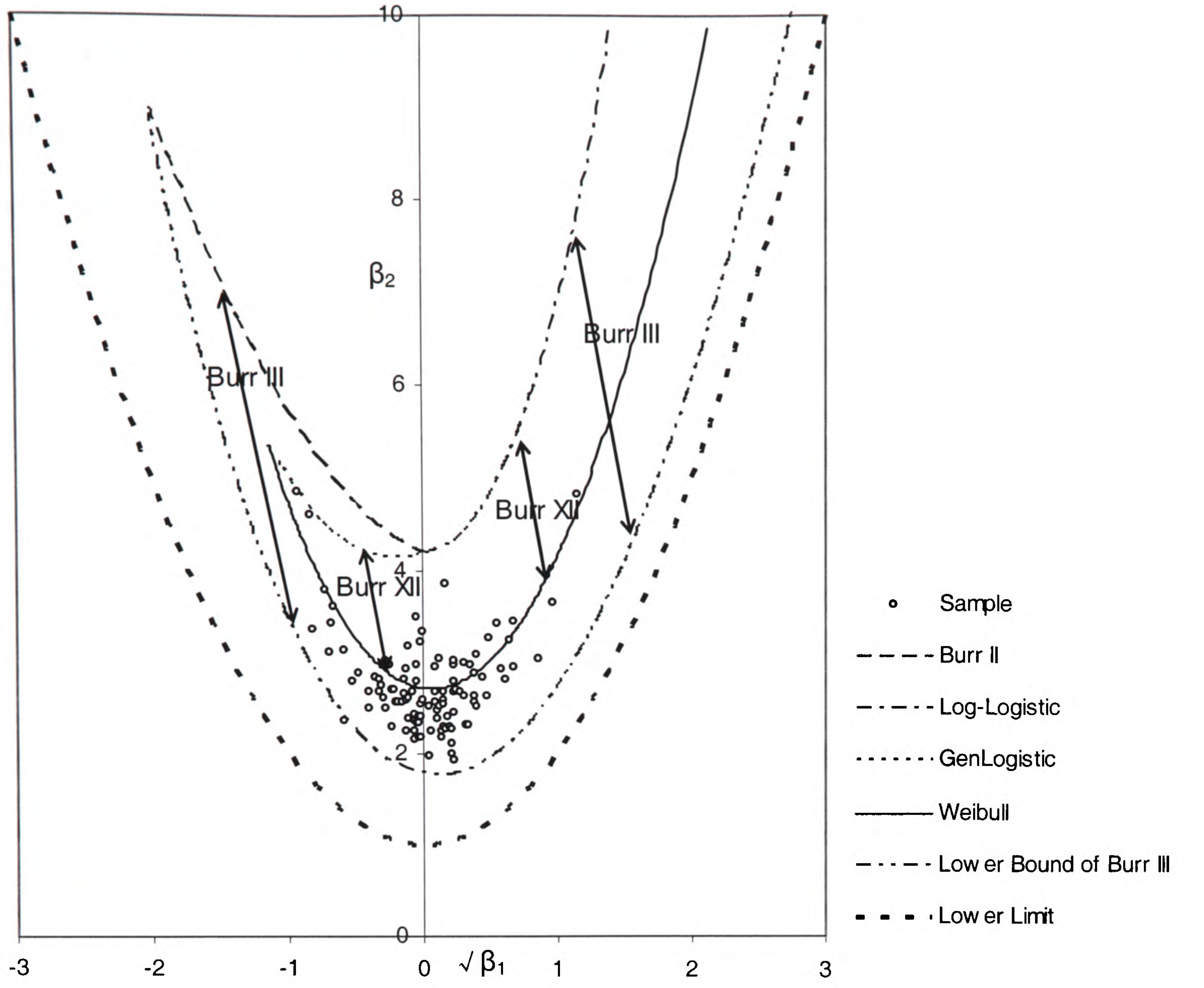


Figure 3.4. ($\sqrt{\beta_1}$, β_2) of the Burr XII and III Distribution

Figure 3.4 shows the asymmetric coverage by Burr XII and Burr III. Both the Burr XII and Burr III distributions have the capacity to describe both positive and negative skewness, but with a rather smaller range of negative skew than positive. The skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) for the Burr XII are given by

$$\sqrt{\beta_1} = \frac{\Gamma^2(k)\lambda_3 - 3\Gamma(k)\lambda_2\lambda_1 + 2\lambda_1^3}{[\Gamma(k)\lambda_2 - \lambda_1^2]^{3/2}}$$

$$\beta_2 = \frac{\Gamma^3(k)\lambda_4 - 4\Gamma^2(k)\lambda_3\lambda_1 + 6\Gamma(k)\lambda_2\lambda_1^2 - 3\lambda_1^4}{[\Gamma(k)\lambda_2 - \lambda_1^2]^2} \tag{3.34}$$

where $\lambda_i = \Gamma(1 + i/c)\Gamma(k - i/c)$, Γ is the gamma function, i = 1,2,3,4 and $ck > 4$.

The lower bound of Burr XII corresponds to the Weibull curve, which is realized as $k \rightarrow \infty$ with c fixed (Burr 1968, Rodriguez 1977). The upper bound in the positive $\sqrt{\beta_1}$ half-plane corresponds to the Burr XII distributions for which $k = 1$ and $c > 4$ (Rodriguez 1977), the Log-Logistic line. The upper bound in the negative $\sqrt{\beta_1}$ half-plane is part of a curve called "generalized logistic" (GenLogistic) corresponding to the generalized logistic distribution as defined below,

$$F(x) = 1 - \frac{1}{(1+e^x)^k} \tag{3.35}$$

which represents the limiting forms of Burr XII distributions as $c \rightarrow \infty$ with k fixed (Rodriguez 1977). This limiting Burr XII curve pass through the Logistic point (0, 4.2) and approaches the Weibull curve asymptotically as $k \rightarrow \infty$ at their end point of (-1.14, 5.4). Figure 3.4 shows that although the Burr XII family covers a large portion of the ($\sqrt{\beta_1}$, β_2) diagram, a much greater area is covered by the Burr III family. The skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) formula for the Burr III are the same as those given for Burr XII, that is, (3.34), but with parameter c replaced with $(-c)$ for calculating λ_i, that is $\lambda_i = \Gamma(1 - i/c)\Gamma(k + i/c)$, $c > 4$. The lower boundary of the Burr III corresponds to the limiting forms of Burr III as $k \rightarrow 0^+$ with c fixed. The upper boundary in the negative $\sqrt{\beta_1}$ corresponds to the Burr II distributions with $c \leq 1$. This limiting Burr XII curve pass through the Logistic point (0, 4.2) and approaches the lower bound of Burr III asymptotically as $k \rightarrow 0^+$ at their end point of (-2, 9). The upper boundary in the positive $\sqrt{\beta_1}$ corresponds to the Log-Logistic distributions, same as the Burr XII. It is important to note that Burr III covers all the space regions in the skewness-kurtosis plane occupied by Gamma, Weibull, Log-Normal, and the Burr XII distributions.

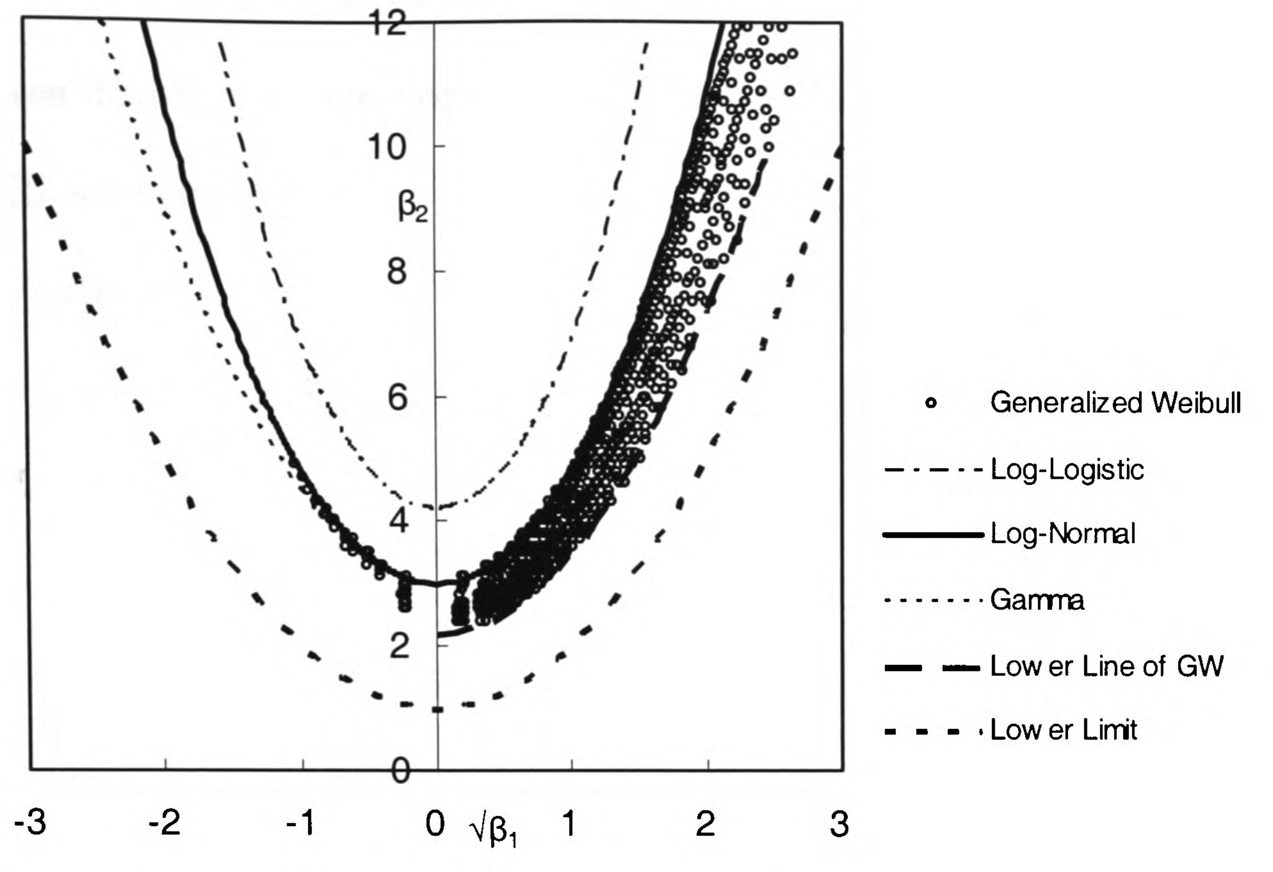


Figure 3.5. ($\sqrt{\beta_1}$, β_2) coverage of the Generalized Weibull Distribution

Figure 3.5 shows the coverage of the GW in the ($\sqrt{\beta_1}$, β_2) shape plane using simulation results. The upper and lower bound in the positive $\sqrt{\beta_1}$ half plane are approximated by the Log-Normal and the $\beta_2 = 2.17+1.29(\sqrt{\beta_1})^2$, respectively, but we do not know their counterparts in the negative positive $\sqrt{\beta_1}$ half plane, analytically or numerically. The main point of what we can draw is the GW is similar to the Burr XII or Burr III, both being asymmetric in the ($\sqrt{\beta_1}$, β_2) plane with much more capacity to describe positive skewness than negative skewness but in different area.

In summary, the Logit-Logistic distribution covers more area in the (β_1, β_2) or ($\sqrt{\beta_1}$, β_2) shape-space than the 3-parameter models, Log-Normal, Weibull, and Gamma, and 4-parameter models S_B, Beta, Generalized Weibull, and the Burr XII and III. We therefore expect that the Logit-Logistic distribution model would perform, empirically, rather better than most previously used distributional models.

3.4 Case-Study with Chinese Fir Diameter Distribution Data

Seven distributions, the Logit-Logistic, S_B, Generalized Weibull, Beta, and Burr III, IV, and XII were compared, all in 4-parameter form. The seven distributions are defined in the above sections.

In terms of skew-kurtosis coverage, the analytical results summarised in the last section come to the clear conclusion that the LL distribution is the best of the distributions considered. However, there are various criteria of goodness of fit that go beyond the third and fourth moments of the distributions fitted.

Ideally we would like to be able to compare the performance of the various new models that we have introduced with the performance of the more familiar models, using a range of criteria of fit and a standard database of empirical distributions which had previously been used by other authors. Unfortunately this is an unachievable ideal, since there is no such database established, and for example, the data used in the early studies of Hafley and Schreuder (1977) have been lost (personal communication). Hence, we do an empirical evaluation on datasets that are available to use. We do not claim they represent a perfect dataset covering all the forms of distributions that might arise in practice. However the data selected for the empirical study is chosen to include a fairly wide range of distributional forms.

In order to allow unambiguous comparison between models we consider in this paper only models with four parameters. We also use a common estimation method, maximum likelihood, and adopt the corresponding measure of goodness of fit, the deviance ($\cong$ (-2log-likelihood)), which, in our case, is equivalent to the AIC criterion (= (-2log-likelihood) + 2P; where P is the number of parameters of the model) for model identification (Akaike 1974), since all the models considered have the same number of parameters (that is, four).

3.4.1 The Case-Study Data

The diameter data of 107 plots for Chinese fir plantations were provided by the Chinese Academy of Forestry. These data have been extensively used in stand-level growth and yield modelling in the China (Wang and Tang 1997, Wang and Li 2000, Li and Wang 2001).

These plots were located at Kaihua forestry farm, Zhejiang province, South-eastern China. The plot size ranges from 400 to 600 m^2, age from 10 to 29 (years), density from 1000 to 4500 per hectare. The sample size ranges from 63 to 239, with mean of about 119. See Table A3.1, in the Appendix, for a detailed sample-plot summary. Figure 3.1 shows the sample-plot distribution shapes in terms of skewness and kurtosis.

From Figure 3.1 we see that LL and Burr III cover all of the sample distribution shapes, and would be expected to provide an adequate fit to all of the sample distributions. However, we see that 8 of the 107 sample distributions lie between the Log-Normal/Gamma line and the Log-Logistic line, and we would therefore expect that they would not be well fitted by the S_B or Beta. Conversely, we would expect the Burr XII to perform very poorly for the distribution of shapes in this case study. Since the GW covers the middle range of shapes we would expect them to perform reasonably well.

3.4.2 Model Fits to the Case-Study Diameter Distributions

Maximum likelihood estimation (MLE) was used, by minimizing the negative log-likelihood function ($-\Lambda\Lambda$) using S-Plus (Mathsoft 1999). ($-\Lambda\Lambda$) is essentially a deviance measure and is used as a goodness of fit criterion. Decreasing ($-\Lambda\Lambda$) indicates improved model fit. Significance tests using this statistic are only valid if the models compared are nested. However, we take this statistic as our common goodness of fit measure in comparing the various distributions, each having the same number of parameters.

We have fitted the Logit-Logistic, S_B, Generalized Weibull, Beta, and Burr III, IV, and XII to all the datasets. Table A3.2, given in the Appendix, lists the value of the ($-\Lambda\Lambda$) goodness of fit statistic for each of the models fitted to each of the sample plot distributions.

We already know from the considerations of the last section that some models will not be able to get close to fitting even the (β_1, β_2) statistics of a dataset, and hence non-convergence in fitting is a likely outcome.

For the 107 sample plots in the case study, the maximum likelihood estimation method converged for the LL, S_B, Beta, GW, and Burr III, IV and XII distributions 106, 106, 105, 88, 100, 88 and 106 times, respectively. The poor (technical) convergence percentage of the Burr IV (88/107≈82%) is not clear to us, since we do not know its coverage in the shape-plane, analytically or by simulation. The poor (technical) convergence percentage of the GW distribution (88/107≈82%) may be partly due to the limited shape coverage of the GW distribution; it cannot get near the lower-limit line. Most non-convergence for GW occurred for datasets that were near the GW lower line and this may possibly be a contributory factor in GW (technical) non-convergence. However, the effects of parameters c and k in GW are confounded, since kc is the parametric combination which determines GW model shape. For the empirical distributions for which there was non-convergence, the re-parameterization $k \rightarrow k/c$ was used, but no improvement in convergence performance was achieved.

We note that Burr XII also, cannot get close to the lower-limit line but attains convergence in 99% (=106/107) of the sample plots. We found (see Table A3.2) that there was only one empirical distribution (plot 73) for which there was non-convergence for all the compared distributional models, which is reverse J-shaped and thus the non-convergence is not unexpected due to the non-regular problems with MLE (Smith 1989, Cheng and Traylor 1995). However, for other empirical distributions for which there was (technical) non-convergence for some of the models, difference in the resulting (negative) log-likelihood between converged models and non-converged models is small, (being much less than that required to demonstrate a significant difference between nested 3 and 4 parameter models). Therefore, we only eliminate one sample (plot 73) from subsequent comparisons.

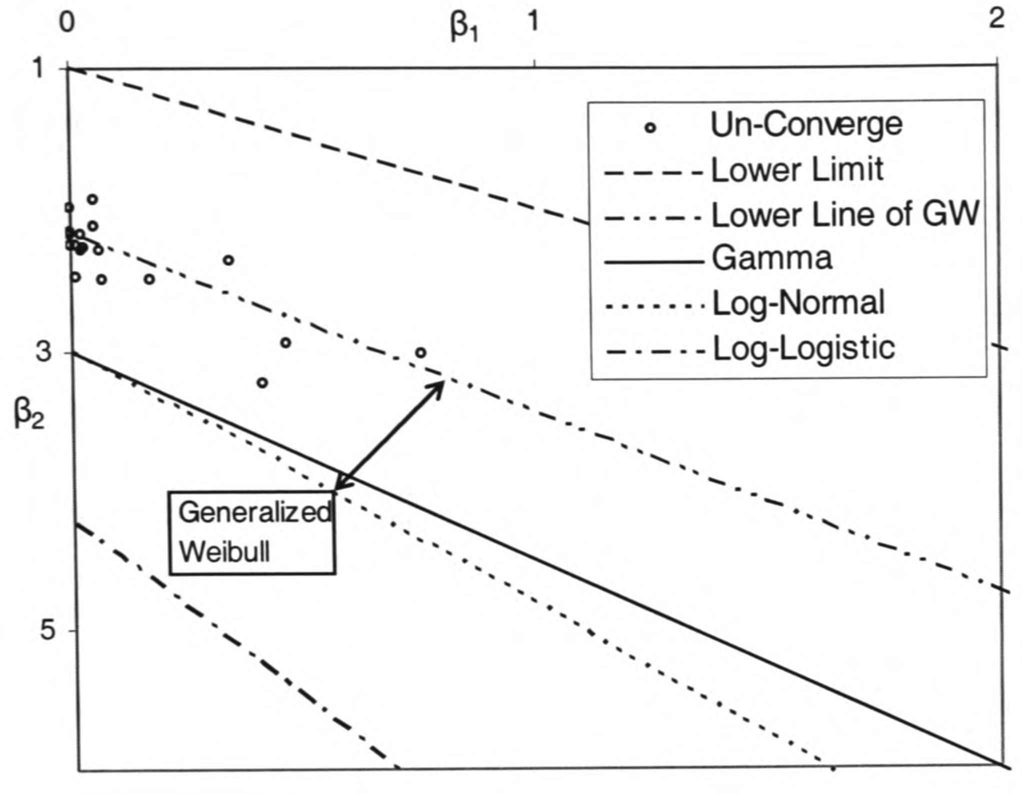


Figure 3.6. Un-converged Samples in Fitting GW

3.4.2.1 Comparison in terms of (-ΛΛ)

Table 3.1 shows the between-model comparative performance of the various models in terms of goodness-of-fit statistic (-ΛΛ).

Table 3.1. Comparison results based on {-ΛΛ}

Row \ Col	BurrIV	GW	Beta	S_B	BurrIII	BurrXII
LL	60/106	74/106	86/106	94/106	72/106	90/106
BurrIV		67/106	67/106	71/106	75/106	79/106
GW			66/106	73/106	71/106	89/106
Beta				84/106	59/106	70/106
S_B					55/106	65/106
BurrIII						63/106

Proportion of cases in which the row-distribution model had a lower {-ΛΛ} than the column-distribution.

From Table 3.1, the Logit-Logistic (LL) had a lower (-ΛΛ) than S_B for 94 of the 106 distributions (i.e. 89%). The LL was better than the Burr IV, GW, Beta, and Burr III and XII for 57%, 70%, 81%, 68% and 85% of cases, respectively. Hence, in terms of this criterion of comparison, LL dominates all the other models, at least in this case study.

These relative performances, based on the (-ΛΛ) criterion, may be summarized as: LL>Burr IV>GW>Beta>S_B>Burr III>Burr XII, where the inequality indicates the relative performance. The main conclusion is that the Logit-Logistic distribution performed better than all other alternative distributional models in these empirical comparisons, as was expected from the considerations of Section 3. The Burr IV ranked second, better than the Beta and S_B for 63% and 67% of cases, respectively. Although we do not know its coverage in the shape plane, we expect some similarity between this model and the LL. The worst performance of the Burr XII is not unexpected, since most of samples (about 85) lies below or on the Weibull line, the lower bound of this model, in the ($\sqrt{\beta_1}$, β_2) shape-plane as shown in Figure 3.4 or the Figure 3.7 below. For most of these empirical distributions and the two near the GenLogistic line, the estimates of parameter k were found to be rather large (with a minimum of about 319, a maximum of 42218, and the mean of 13263), and this indicated that for these samples the fitted Burr XII asymptotically approached the 3-parameter Weibull, the lower limit distribution of the Burr XII. We compared the Burr XII and the Weibull, and not surprisingly, we found that the Weibull performed better than the Burr XII for all these cases with large estimates of k, though the difference in the (-ΛΛ) was quite small (with a mean of difference of 0.001), while for all the other 21 cases (exclusive of plot 73) Burr XII performed better than the Weibull. Figure 3.7 further graphically showed the comparative performance between the 4-parameter Burr XII and the 3-parameter Weibull. Therefore, for this empirical study, the 4-parameter Burr XII performed worse than the 3-parameter Weibull, which can be explained by the coverage of our empirical distributions in the shape-plane and more importantly indicates that for those un-nested models, a model with an extra parameter does not necessarily fits the data better than a model with less number of parameters.

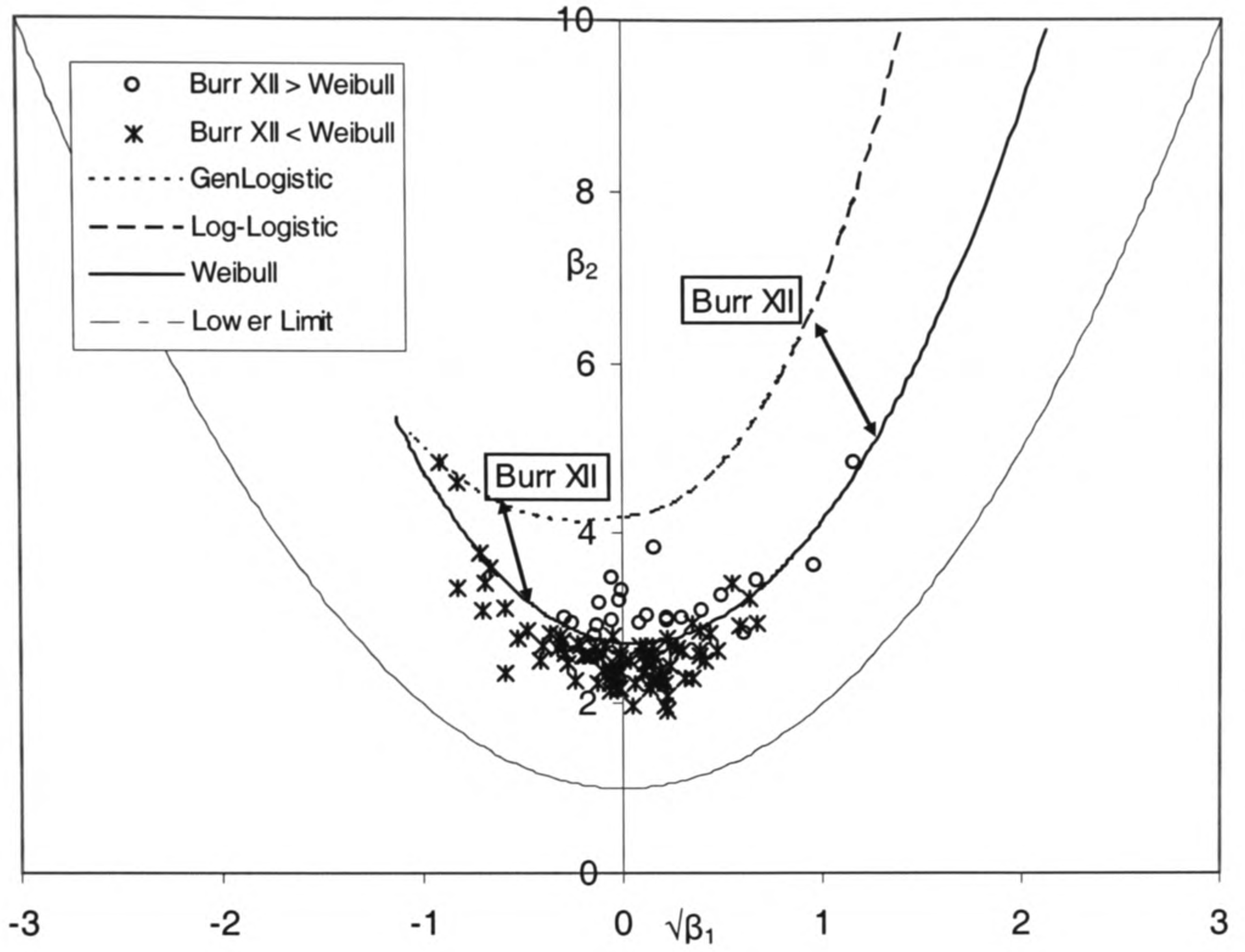


Figure 3.7. Comparative Performance of the Burr XII and 3-parameter Weibull

However, it is found that the Beta is better than the S_B in 84 of 106 (79%) of cases, confirming the results of Li et al. (2002), but not in accord with our expectations following from their shape-space coverage properties. The poor performance of the Burr III compared to the Beta (59/106 ≈ 56%) and S_B (55/106 ≈ 52%), though not differing as much as the other pairs of compared distributions, is also out of our expectation, as the Burr III covered all the empirical distributions in the shape-plane and we expected its performance similar to the Logit-Logistic. The main reason is possibly that as shown in Figure 3.1, most of the (β_1, β_2) points of our samples fall in the common area shared by Beta, S_B, and Burr III, with only seven sample points laying beyond the Gamma line as well as the Log-Normal line and thus outside the coverage area of Beta or S_B. Therefore, these comparison results may indicate that for those (empirical) distributions falling in the common area of these three models, Beta may fit them better than the S_B , and both models overall better than the Burr III, for which the reasons are not clear. Graphically, Figure 3.8 showed the comparison between Beta and S_B in the (β_1, β_2) shape-plane, while Figure 3.9 and Figure 3.10 showed the comparison between Burr III, and the Beta and S_B in the ($\sqrt{\beta_1}$, β_2) shape-plane, respectively. From Figure 3.9 and 3.10, not surprisingly, Burr III performed better than Beta and S_B for 6 and 7 out of the seven samples which fall in the shape area of Burr III but not in those of Beta and S_B, respectively. By removing these seven sample points, that is, for those samples in the common area of these three models, the better performance in percentage of Beta and S_B over the Burr III increased to 58/99 (59%) and 55/99 (56%) from 59/106 (56%) and 55/106 (52%), respectively.

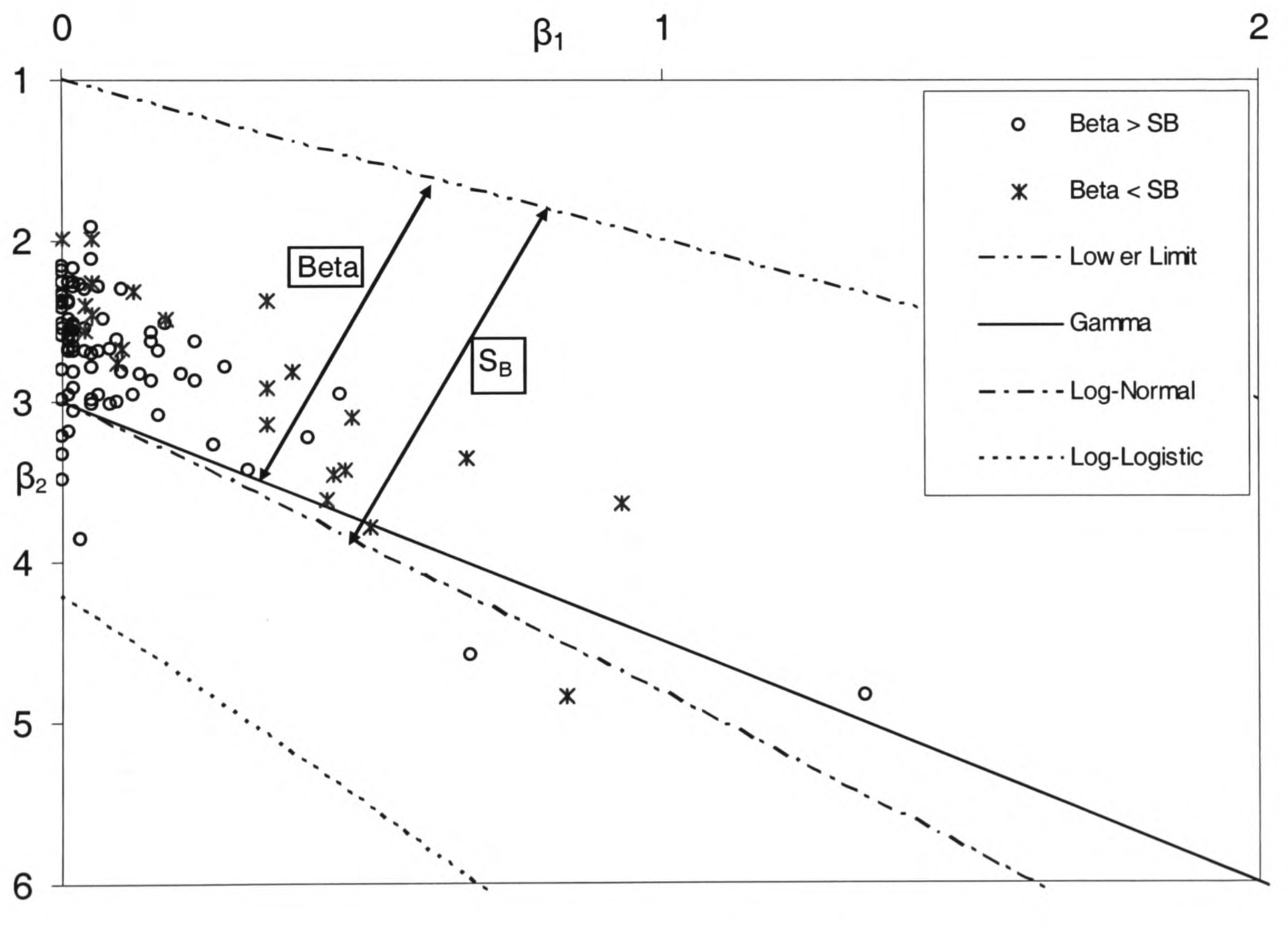


Figure 3.8. Comparative Performance of the Beta and S_B

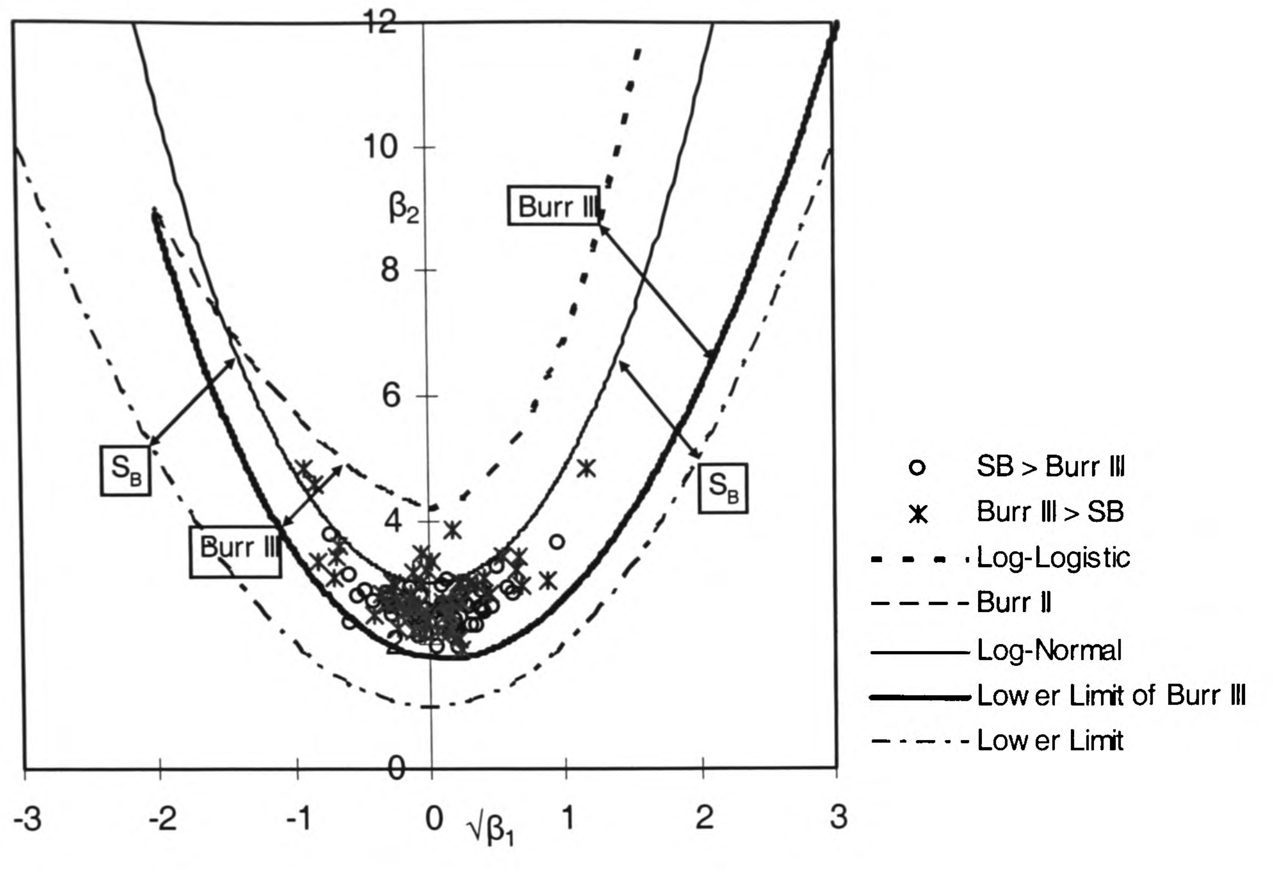


Figure 3.9. Comparative Performance of the Burr III and S_B

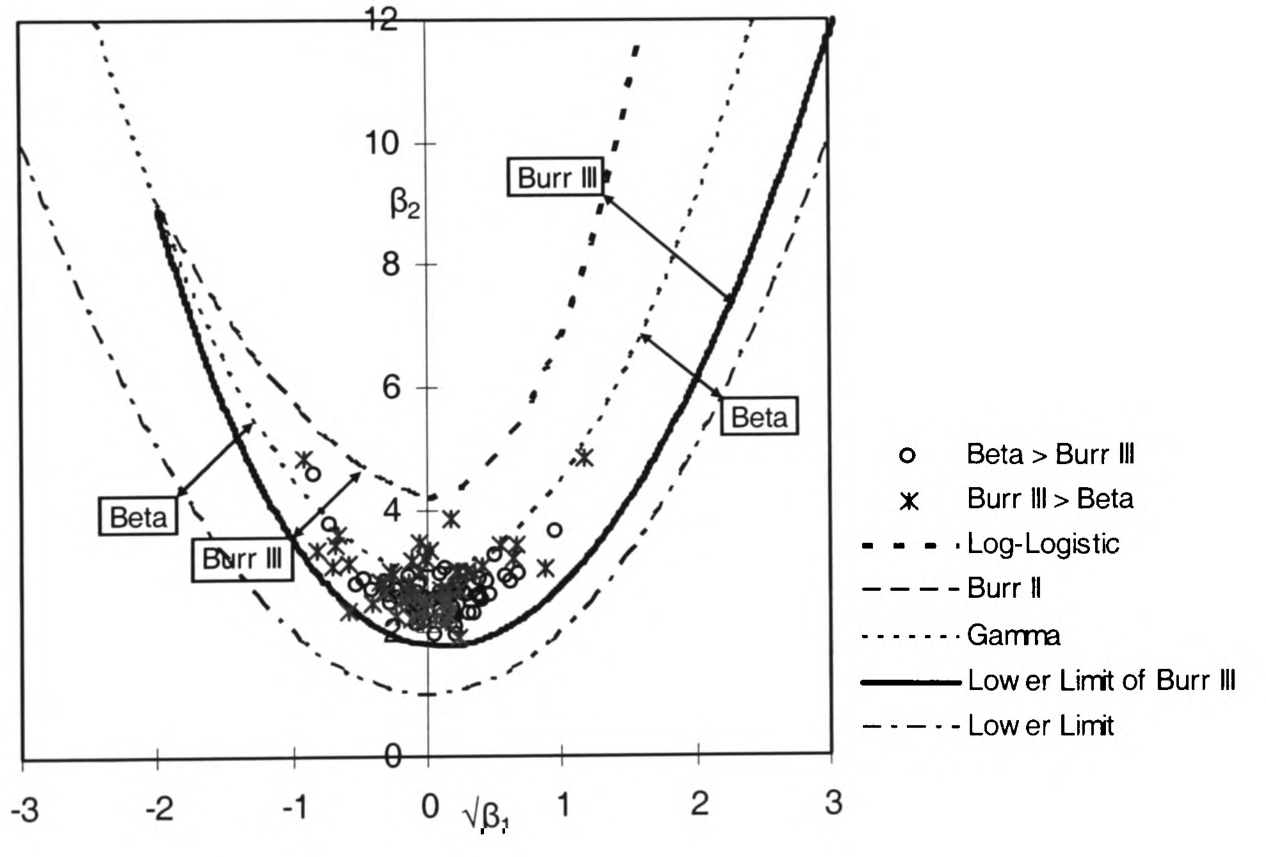


Figure 3.10. Comparative Performance of the Burr III and Beta

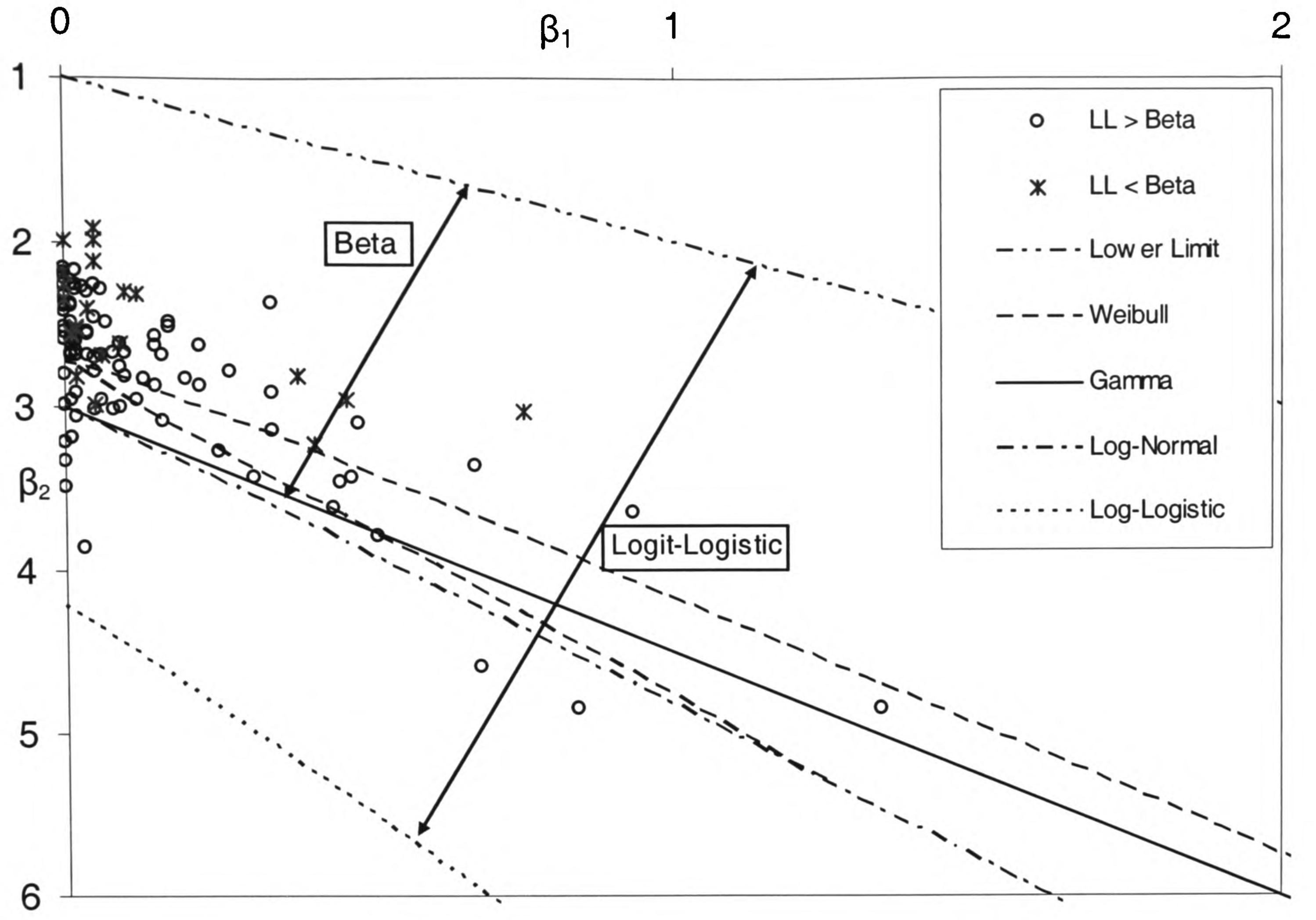


Figure 3.11. Comparative Performance of the Logit-Logistic and Beta

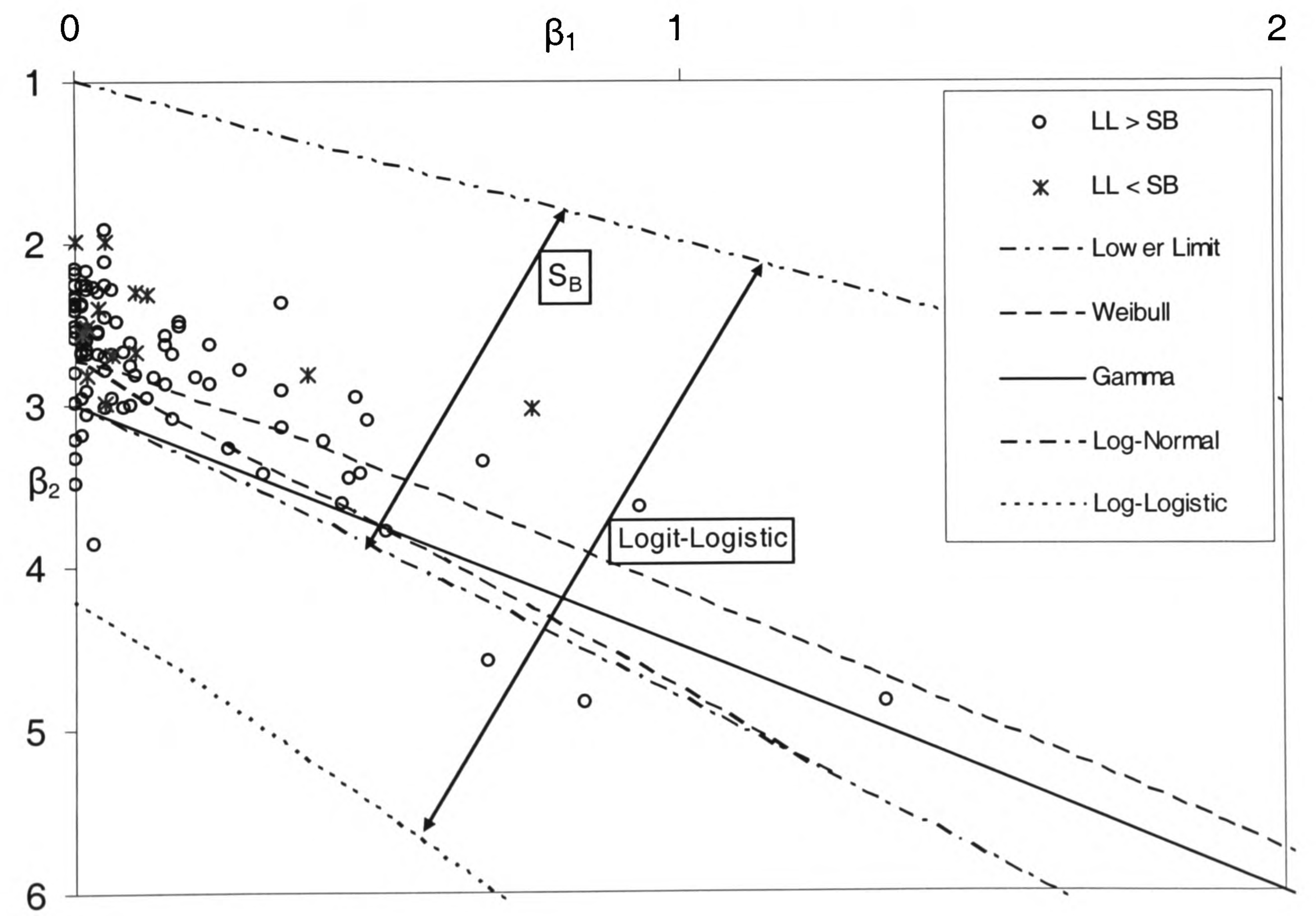


Figure 3.12. Comparative Performance of the Logit-Logistic and S_B

Therefore, by recognizing the fact that the common area of Beta, S_B and Burr III in which most of the samples fall is also shared by the Logit-Logistic, and possibly by the Burr IV and Generalized Weibull, we may conclude that: the better performance of the LL, Burr IV and GW over the others is not only because they cover more wider area in the shape plane (except the GW), but more importantly because they may fit the sample distributions in the common area more adequately, at least in this empirical study here. We have to admit that we have not understood the second point, but Figure 3.11 and 3.12 illustrated this point by comparing LL with Beta and S_B empirically in this study, respectively.

3.4.2.2 Comparison in terms of other Goodness-of-fit Criteria

During reviewing one of our submitted papers, it has been suggested by a reviewer and an Associate Editor that a comparison of the models using a different criterion of goodness-of-fit than ($-\Lambda\Lambda$) might be of interest. We have done this for several other possible criteria, including Kolmogorov-Smirnov (D), Cramér-von Mises (W^2), and Reynolds' "error index", both grouped and continuous versions. We did not include the Burr III and IV models here since we discovered both models much later and these criteria are all secondary as we already included them in the comparison using the ($-\Lambda\Lambda$) criterion. Let $z_i = \hat{F}(x_i)$ be the estimated CDF, we have

$$D^+ = \max_i\{\frac{i}{n} - z_i\},\ D^- = \max_i\{z_i - \frac{(i-1)}{n}\},\ D = \max(D^+, D^-) \tag{3.36}$$

$$W^2 = \sum_{i=1}^{n}\{z_i - \frac{(2i-1)}{2n}\}^2 + \frac{1}{12n} = \sum_{i=1}^{n}\{z_i - \frac{(i-0.5)}{n}\}^2 + \frac{1}{12n} \tag{3.37}$$

Reynolds et al. (1988) suggested an "error index" (EI) as a measure of fit, which is a weighted sum of the absolute differences between predicted and observed numbers of diameters in each diameter class, which is defined as,

$$EI = N\sum_{j=1}^{n}\left|\int_{I_j} w(x)d\hat{F}(x) - \int_{I_j} w(x)dF_n(x)\right| \tag{3.38}$$

where $\hat{F}(x)$ is the estimated CDF, $F_n(x)$ is the empirical CDF, $w(x)$ is a weight function of diameter, I_j is the jth diameter class (j=1,2,...,n) and N is the sample size. Let $w(x)$ = 1, this error index collapses to,

$$EI = \sum_{i=1}^{n}\left|O_i - E_i\right| \tag{3.39}$$

where O_i and E_i are the observed and predicted/expected numbers of trees respectively in the ith diameter class. An "un-grouped" version of EI can be defined as,

$$EI2 = \sum_{i=1}^{n}\left|\hat{F}(x_i) - F_n(x_i)\right| = \sum_{i=1}^{n}\left|\hat{F}(x_i) - \frac{i-0.5}{n}\right| \tag{3.40}$$

The comparison results based on these secondary criteria are shown in Table 3.2, 3.3, 3.4, and 3.5, respectively. It can be summarized as,

D: LL>Burr=GW>S_B>Beta

W^2: LL>Burr>GW>S_B>Beta

$EI2$: LL>GW>Burr>S_B>Beta

EI: S_B>LL, LL> (Beta, Burr, GW), GW>Burr>S_B>Beta

With the criteria of D, W^2, and $EI2$, it is seen that the LL model still dominates the other 4 distributions considered, while the Beta is dominated by all other distributions. However, with the EI, the set of non-transitive relations above makes it impossible to make clear conclusion, except that the Beta is dominated by all other distributions. The non-transitive relations occurred due to our "1 to 1" comparison logic, that is, we compare model A and B, model B and C, separately. Suppose we get A>B and B>C (> denoting relative better performance). If we further get A>C, then the performance among these three models is A>B>C. However, sometimes we may get C>A, under such cases, we will not draw conclusion clearly. As an alternative to this comparison logic and also for more confidence in our comparing the models, we rank all the compared models for each data set and the rank sum will determine the relative performance of the models. Table 3.6 presents the rank sums for the criteria we used (($-\Lambda\Lambda$), D, W^2, EI, and $EI2$) across the 106 data sets, in which smaller value of rank sums indicates relative better performance. From table 3.6, the comparison results based on ($-\Lambda\Lambda$), D, W^2, and $EI2$ under the "rank sum" logic can be regarded as the same as those under the "1 to 1" logic, considering that "GW > Burr XII" based on $EI2$ under the "1 to 1" comparison logic is because GW performed better than Burr XII for just 1 more than the half (53) of the samples. With the criterion of EI, we tend to accept the comparison results using the "rank sum" logic, that is, GW>S_B>LL>Burr XII>Beta.

From the analysis above, the main conclusion we would draw is that the LL model still dominates the other four 4-parameter distributions considered. However, the differences between these criteria (D, W^2, EI, and $EI2$) based performances of Burr XII, GW, S_B and Beta are so marginal that we do not feel a ranking of the model is justified from this empirical study. The first reason is that since the fitting has been done using the ($-\Lambda\Lambda$) criterion, examination of performance on a secondary criterion, such as D, amounts to an attempt to evaluate the models's fit to the data using two criteria simultaneously. The second is that if we do prefer to use these criteria (D, W^2, EI, and $EI2$) for model selection, we actually can use each of these criteria for fitting the models as well (like {$-\Lambda\Lambda$} used in MLE), which we will discuss further in some chapter later.

Table 3.2. Comparison results based on KS statistics (D)

Col / Row	Burr XII	GW	S_B	Beta
LL	66/106	74/106	78/106	84/106
Burr XII		53/106	62/106	68/106
GW			56/106	65/106
S_B				68/106

Proportion of cases in which the row-distribution model had a lower (D) than the column-distribution.

Table 3.3. Comparison results based on W^2

Row \ Col	Burr XII	GW	S_B	Beta
LL	65/106	70/106	70/106	82/106
Burr XII		55/106	61/106	64/106
GW			56/106	63/106
S_B				69/106

Proportion of cases in which the row-distribution model had a lower (W^2) than the col-distribution.

Table 3.4. Comparison results based on *EI*

Row \ Col	S_B	LL	GW	Burr XII	Beta
S_B		59/106	47/106	52/106	64/106
LL	47/106		55/106	56/106	58/106
GW	59/106	51/106		58/106	65/106
Burr XII	54/106	50/106	48/106		54/106
Beta	42/106	48/106	41/106	52/106	

Proportion of cases in which the row-distribution model had a lower (*EI*) than the col-distribution.

Table 3.5. Comparison results based on *EI2*

Row \ Col	GW	Burr XII	S_B	Beta
LL	64/106	60/106	59/106	77/106
GW		54/106	59/106	63/106
Burr XII			64/106	64/106
S_B				72/106

Proportion of cases in which the row-distribution model had a lower (*EI2*) than the col-distribution.

Table 3.6. Rank Sum of Compared Distributional Models based on 5 Measures of Fit

Measure \ Model	LL	S_B	Beta	Burr XII	GW	Compare Results
$(-\Lambda\Lambda)$	186	398	316	420	270	LL>GW>Beta>S_B>Burr XII
D	228	340	391	307	324	LL>Burr XII>GW>S_B>Beta
W^2	243	330	384	309	324	LL>Burr XII>GW>S_B>Beta
EI2	270	322	382	304	312	LL>Burr XII>GW>S_B>Beta
EI	314	308	347	324	297	GW>S_B>LL>Burr XII>Beta

3.4.3 Constrained Model Estimation

To complete the comparison of the various models considered, but in the context of the lower bound parameter being zero, we fitted all the seven distribution models, with the lower bound parameter, ξ or a, constrained to zero. Table 3.7 lists the comparative results.

Table 3.7. Comparison results based on $\{-\Lambda\Lambda\}$ (with ξ or $a \equiv 0$)

Col / Row	Beta	GW	BurrIV	BurrXII	LL	BurrIII
S_B	78/106	74/106	69/106	84/106	85/106	88/106
Beta		71/106	65/106	84/106	85/106	88/106
GW			54/106	96/106	93/106	92/106
BurrIV				66/106	73/106	79/106
BurrXII					85/106	85/106
LL						87/106

Proportion of cases in which the row-distribution model had a lower $\{-\Lambda\Lambda\}$ than the col-distribution.

From Table 3.7, it follows that S_B > Beta > GW >Burr IV > Burr XII > LL > Burr III, where ">" represents better performance. The Logit-Logistic performs the second worst, and except for the Burr III and XII distributions which are still among the worst, the results for the other five models were "inverted" compared with the results without the constraint on the lower boundary parameter! Hence, the alternate strategies, of imposition of a constraint on the location parameter (to zero) in Hafley and Schreuder (1977), but not in Li et al. (2002), explains, at least to some extend, why different conclusions about the comparative performance of S_B and Beta were reached in those studies. In this study it was found that imposing the zero-constraint on lower boundary parameter resulted in S_B performing better than Beta, consistent with the Hafley and Schreuder (1977). However, without this constraint, Table 3.1 confirms the conclusion of Li et al. that the Beta out-performs S_B. Of course, the studies of Hafley and Schreuder (1977) and Li et al. (2002) used different data, another possible reason for the differing conclusions.

3.4.4 Testing the Lower-bound Parameter Constraint (**ξ or $a \equiv 0$**)

We may ask, and test if adopting the constraint is reasonable or not for the case-study data. It is well known (McCullagh and Nelder 1989) that -2(ΛΛ(unconstrained)- ΛΛ(constrained)) is distributed approximately as Chi-Square distribution with one degree of freedom, where ΛΛ denotes log-likelihood. We have employed a Chi-Square test to test whether the constraint makes a significant difference, at the 5% probability level. However, as parameter on boundary is a non-standard case, testing the hypothesis (ξ or $a \equiv 0$) becomes a 50:50 mixture of Chi-Square on 0 and 1 degree of freedom (df), not the standard Chi-Sqaure on 1 df (Self and Liang 1987). For this mixture distribution, the 95% significant point is 2.7 (Ramesh 1995). Table 3.8 lists the results.

Table 3.8. Chi-Square test of H_0: "Location parameter = 0"

Beta	Burr XII	S_B	LL	GW	Burr IV	Burr III
38/106	38/106	25/106	65/106	43/106	47/106	67/106

Proportion of nulls rejected at 5% level.

For the Logit-Logistic and Burr III distributions, more than 50% of Chi-Square test rejected the "Location=0" null hypothesis. That is, the pre-setting of the location parameter to zero is **not** reasonable for both models. In contrast, pre-setting the location parameter to zero may be relatively reasonable for the other 4-parameter distributions considered, especially for S_B. This largely explains why Logit-Logistic performed best in the unconstrained situation, but the second worst in the constrained situation.

Use of $\xi = 0$ constrained models might be regarded as indicated, for example for S_B, for which 33 out of the 106 sample-plots estimated ξ as 0 in this study. Use of the Logit-Logistic model would avoid such constraint since only 7 out of 106 sample-plots have ξ-estimates of zero.

3.4.5 Comparison of 3 and 4 Parameter Weibull

Since the 3-parameter Weibull is a special case of the Generalized Weibull with $k = 1$, we tested the null hypothesis: $k = 1$, using a likelihood ratio test at the 0.05 significance level. It was found that for 24 out of 106 sample-plots the 4-parameter model was better. Hence, the Generalized Weibull, with one additional shape parameter, improves goodness-of-fit performance over the usual 3-parameter Weibull model, in this empirical study.

3.4.6 Comparison of the 3-parameter Weibull with Constrained 4-parameter Models

The 3-parameter Weibull distribution model, equation (3.9), is widely used, and because of this we have compared it with the seven 3-parameter models obtained from the 4-paramerer models when the lower bound parameters are set as zero. This conventional 3-parameter Weibull model (with unconstrained lower bound parameter) performed better (in $-\Lambda\Lambda$ terms) than all of the seven constrained models, evaluated over the 106 sample-plots: 60/106 for the S_B, 65/106 for Beta, 70/106 for the Generalized Weibull, 63/106 for Burr IV, 93/106 for Burr XII, 84/106 for the Logit-Logistic, and 84/106 for the Burr III.

3.5 Discussion

The Logit-Logistic distribution performed best in both the theoretical study of skew-kurtosis shape-space coverage, and in the empirical study, and would therefore seem to offer considerable potential for future practical usage, particularly in view of its tractability because of the availability of its CDF in explicit form.

We have also conducted a simulation study in which the performance of each of the distributional models is evaluated on data simulated from each of the other distribution models. It is from this simulation study that the shape-space points for the Generalized Weibull were obtained in Figure 3.2 and 3.5. The results are not simple, but overall, the

Logit-Logistic distribution is most often the best 4-parameter distribution at fitting data that is generated by another distribution, consistent with the results of the case study in this Chapter.

We do not claim to fully understand this result. On *a priori* grounds we would have expected the performance of the Logit-Logistic to be similar to the S_B, with the main comparative advantage being the tractability of the model.

The reversal of performance rankings, depending on defining the lower boundary parameter to be zero, is not fully understood. However, it may be noted that the estimation of boundary parameters has been a recurrent problematic issue in the fitting of diameter distributions, and is considered in detail in another paper being prepared.

The Beta, S_B and LL are all equally capable of representing positive or negative skewness, and they all extend up to the lower-limit line near which distributions are U-shaped. Figure 3.3 shows that the upper-limit lines for these three distributions form a fan-like arrangement, in which the main distinguishing feature is the largest kurtosis attainable for zero skew. For the Beta and the S_B this is 3 (the Normal), compared with 4.2 for Logit-Logistic (the Logistic). Can a new distributional model be found or devised with a higher (zero skew) maximal kurtosis than 4.2? If so, it may be even better than the Logit-Logistic!

Appendix: Chinese Fir case-study summaries: data and results.

Table A3.1: Plot Summaries: age (years), number of trees/plot, minimum and maximum diameters (cm), root-mean-square diameter (cm), skewness $\sqrt{b_1}$ and kurtosis b_2.

Plot	Age	Num	D_{min}	D_{max}	D_g	$\sqrt{b_1}$	b_2	Plot	Age	Num	D_{min}	D_{max}	D_g	$\sqrt{b_1}$	b_2
1	17	106	4.6	15.7	9.7	0.39	2.57	55	15	114	5.6	16.4	12.6	-0.82	3.35
2	15	110	4.9	18.6	13.4	-0.52	2.78	56	16	102	4.9	17.4	11.8	-0.30	2.75
3	15	117	3.8	21.4	11.5	0.22	2.98	57	18	99	8.0	18.6	13.4	0.04	2.52
4	13	146	3.8	19.3	8.9	0.64	3.23	58	17	138	5.4	16.8	10.9	0.28	2.67
5	15	113	4.8	18.9	12.0	0.09	2.96	59	13	99	4.3	16.5	11.5	-0.70	3.10
6	13	121	5.2	17.5	11.8	-0.05	2.80	60	15	91	5.9	19.9	13.7	-0.11	3.18
7	13	156	4.3	18.9	11.8	-0.05	2.98	61	15	105	5.0	18.9	13.7	-0.66	3.61
8	13	138	4.7	15.6	9.8	0.22	2.11	62	15	82	5.3	18.0	12.1	0.05	1.98
9	16	144	4.7	18.6	11.3	0.09	2.69	63	14	74	4.9	17.0	12.4	-0.29	2.61
10	12	100	4.1	15.7	10.6	-0.01	3.21	64	16	195	1.9	11.5	6.7	0.62	2.82
11	13	91	6.5	17.9	13.3	-0.05	2.34	65	15	104	4.0	15.4	9.9	-0.07	2.42
12	13	74	9.6	18.3	14.1	-0.24	2.29	66	15	103	5.9	19.0	11.6	0.32	2.30
13	16	85	10.1	20.5	14.5	0.38	2.63	67	15	118	4.0	16.3	10.6	-0.01	2.59
14	16	71	9.7	19.5	14.0	0.30	3.00	68	15	162	5.5	20.7	11.8	-0.13	2.81
15	15	101	4.5	17.7	12.4	-0.68	3.43	69	15	96	4.5	19.8	12.9	-0.21	2.56
16	16	95	8.3	18.0	13.3	-0.10	2.61	70	15	112	4.0	22.4	13.2	0.10	2.56
17	14	109	4.6	16.9	11.9	-0.32	2.81	71	15	156	4.3	17.2	11.0	0.22	2.26
18	14	162	4.0	15.5	10.1	-0.08	2.67	72	12	122	4.2	19.8	10.5	0.17	3.85
19	15	132	4.6	18.2	12.0	-0.07	2.15	73	15	90	4.0	16.3	7.9	0.87	3.03
20	15	167	4.1	17.6	9.4	0.50	3.27	74	14	95	3.7	11.3	6.3	0.96	3.64
21	15	71	4.8	18.0	11.4	0.13	2.17	75	14	96	4.2	13.7	8.6	0.15	2.29
22	15	95	6.5	16.9	11.1	0.67	3.45	76	17	77	4.0	15.0	11.2	-0.59	2.37
23	16	83	7.4	20.9	16.2	0.72	3.79	77	17	95	5.5	20.0	12.5	-0.02	2.41
24	16	116	8.1	17.6	13.4	-0.06	2.37	78	14	167	4.0	13.6	8.2	0.41	2.52
25	16	94	7.5	18.1	13.3	-0.12	2.58	79	16	190	4.3	17.6	10.2	0.15	2.68
26	16	90	5.4	19.1	12.2	-0.19	2.55	80	17	125	4.5	15.1	9.2	0.35	2.31
27	17	99	8.1	17.6	13.3	-0.13	2.92	81	17	103	4.0	16.3	10.1	0.18	2.27
28	12	133	4.0	14.5	9.3	-0.12	2.25	82	17	101	7.5	23.0	15.6	-0.05	3.48
29	15	86	4.0	20.8	13.6	-0.23	2.70	83	26	63	9.7	23.4	17.8	-0.40	2.68
30	16	118	6.0	17.2	10.8	0.24	2.68	84	16	81	4.6	22.9	14.4	-0.14	2.66
31	14	125	4.1	15.4	9.7	0.15	2.25	85	17	177	4.7	20.4	9.7	1.16	4.84
32	15	107	2.9	15.0	8.1	0.41	3.09	86	13	92	8.3	19.1	13.5	0.15	2.60
33	13	187	4.3	13.2	8.0	0.47	2.63	87	13	116	6.2	20.7	14.7	-0.41	2.49
34	15	110	3.5	15.0	9.2	-0.16	2.56	88	13	106	5.0	16.5	10.3	0.23	2.78
35	10	147	3.5	14.6	9.6	-0.21	2.69	89	16	100	4.9	18.7	14.2	-0.92	4.84
36	14	199	3.5	15.8	8.7	0.16	2.53	90	13	183	4.0	17.1	10.6	0.19	2.40
37	15	93	6.1	16.9	10.5	0.58	2.92	91	15	138	4.6	15.1	9.9	0.10	2.37
38	17	72	9.5	20.2	15.3	-0.06	2.25	92	26	65	4.5	27.5	20.5	-0.83	4.59
39	29	123	4.5	22.1	12.6	0.30	2.62	93	16	122	6.7	20.5	14.2	-0.32	2.67
40	14	173	3.9	19.5	11.2	0.23	1.92	94	26	93	10.8	25.9	17.1	0.44	2.83
41	14	131	3.5	17.2	11.4	-0.47	2.87	95	19	122	8.8	22.0	15.5	0.01	3.33
42	15	148	4.1	17.5	12.1	-0.58	3.14	96	15	109	4.1	17.2	10.9	-0.10	2.38
43	15	86	6.8	18.0	13.3	-0.27	2.49	97	15	142	4.0	14.9	8.0	0.68	2.95
44	17	162	3.5	17.3	10.5	0.20	2.30	98	15	147	4.6	17.7	10.5	0.39	2.87
45	16	211	3.5	18.0	10.2	0.06	2.25	99	15	129	5.9	18.6	12.1	0.23	2.45
46	15	148	4.7	18.5	10.7	0.13	2.51	100	15	173	4.0	11.7	8.1	-0.03	2.33
47	13	162	2.2	13.1	8.4	0.21	1.99	101	16	108	6.5	22.5	14.2	-0.25	2.96
48	12	109	4.9	19.5	12.4	-0.36	2.83	102	26	94	7.5	26.3	16.7	0.23	3.02
49	12	239	4.0	14.2	9.1	-0.03	2.18	103	26	110	9.9	24.9	18.2	-0.02	2.19
50	14	69	6.4	23.0	13.1	0.55	3.43	104	26	71	10.7	27.5	19.5	0.10	2.67
51	16	89	4.0	19.2	10.6	0.35	2.96	105	26	125	6.7	24.5	15.0	0.13	3.05
52	15	110	6.1	18.0	12.5	-0.28	3.01	106	26	78	12.1	24.7	17.8	0.12	2.48
53	13	123	7.2	17.3	12.6	-0.03	2.39	107	18	138	9.0	19.6	14.0	0.24	2.69
54	15	177	4.9	20.2	12.1	-0.01	2.54								

Table A3.2. Negative Log-Likelihood for Fitting Logit-Logistic (LL), Johnson's S_B (S_B), Generalized Weibull (GW), Beta, and Burr III, IV and XII Distributions.

Plot	LL	S_B	GW	Beta	Burr III	Burr IV	Burr XII
1	244.35	245.05	245.25	244.94	246.36	245.01	245.74
2	271.08	271.96	271.82	271.92	272.67	272.01	273.53
3	300.64	300.45	299.94	300.31	299.20	300.53	299.90
4	357.68	357.71	353.11*	355.25	349.75*	348.38*	356.36
5	276.21	276.89	276.63	276.74	277.09	276.26	276.62
6	285.22	285.92	285.84	285.88	286.79	285.27	285.85
7	379.63	380.63	379.78	380.30	378.80	378.80	379.53
8	323.34	324.29	324.21*	323.26	324.15*	322.50*	330.51
9	350.97	351.48	350.68	351.35	350.22	349.81	351.12
10	227.27	228.77	228.17	228.51	227.99	227.27	227.98
11	211.07	211.17	212.32	210.83	214.35	210.19	212.61
12	153.84	154.07	153.18*	154.04	152.52	153.85	155.03
13	183.89	184.06	184.08	184.01	185.17	184.63	184.18
14	147.98	148.20	147.98	148.16	147.52	147.94	147.97
15	233.52	235.08	233.53	235.25	232.62	236.93	234.35
16	205.82	206.62	206.04	206.52	206.12	205.77	206.71
17	250.25	250.45	250.44	250.43	251.19	250.63	250.56
18	380.19	381.78	381.38	381.60	381.72	380.33	381.88
19	335.01	335.50	335.82*	335.27	336.50	335.06	339.09
20	377.58	377.79	377.71	377.76	378.56	382.59	377.65
21	178.96	179.33	179.35*	179.16	179.04	178.86	180.90
22	200.19	201.01	200.81	201.18	200.22	201.55	200.08
23	189.79	190.40	190.09	190.49	190.80	191.44	190.23
24	250.14	250.75	251.65	250.44	253.44	249.85	252.45
25	214.39	215.40	215.26	215.21	215.44	214.50	215.75
26	227.63	228.08	226.54	228.07	226.18	224.01*	228.19
27	205.77	206.29	206.23	206.27	206.49	205.99	206.23
28	306.05	306.82	304.60*	306.27	304.20	304.98	310.28
29	226.98	227.36	227.13	227.27	227.31	227.29	227.27
30	249.98	248.90	248.82	248.88	248.77	250.41	248.82
31	290.36	290.60	291.06	290.54	291.92	290.47	292.11
32	240.61	240.76	240.52	240.68	240.62	243.23	240.48
33	376.53	377.03	377.37	376.90	379.79	379.03	377.95
34	250.46	250.69	249.33	250.69	248.72	247.95	250.53
35	331.53	332.13	332.13	332.04	333.57	331.92	332.46
36	468.99	469.16	463.90	467.83	462.33	461.90*	469.06
37	200.83	200.90	201.12	200.99	202.45	203.21	201.16
38	164.35	164.49	164.61	164.45	164.95	164.34	165.48
39	336.13	336.78	336.52	336.63	337.10	335.98	336.90
40	466.41	466.45	469.19*	465.30	465.01*	466.14*	477.54
41	327.70	328.87	328.23	328.83	329.21	329.26	330.57
42	351.32	352.42	351.53	352.61	352.47	354.13	353.09
43	204.99	206.56	206.75*	205.70	207.80	205.18	208.32
44	415.76	417.23	418.14	416.60	420.57	415.98	421.13
45	544.50	544.86	540.74*	543.92	540.27	536.70*	547.91
46	355.50	354.96	353.45	354.75	353.23	353.34	354.69
47	374.87	373.82	378.34	373.97	380.25	373.69	378.74
48	274.20	274.93	273.69	274.84	272.44	271.56	274.11
49	542.43	544.01	540.77*	542.50	541.05	540.65	551.07
50	182.19	182.59	181.66	182.30	180.70	180.39*	181.99
51	226.07	226.55	225.92	226.37	225.51	225.52	226.13
52	247.96	248.75	248.32	248.70	247.93	247.31	248.26
53	273.77	274.77	275.09	274.42	276.61	273.78	276.41
54	466.45	467.55	463.87*	467.08	463.91	459.79*	467.78
55	245.88	246.53	246.17	246.92	246.39	248.05	248.84
56	240.95	241.08	240.99	241.08	241.72	241.47	241.31
57	220.69	221.12	221.08	221.05	222.27	220.67	221.40
58	319.15	320.35	320.28	320.20	321.30	319.35	320.48
59	223.78	224.86	223.15	225.24	221.45	227.30	224.63
60	218.36	219.25	218.84	219.10	219.01	218.43	218.76
61	243.06	244.69	243.41	244.77	242.35	246.21	243.67
62	203.09	202.90	203.67*	202.95	204.19	203.07	205.62
63	173.38	173.50	174.16	173.23	174.89	172.87	174.72
64	397.94	396.42	396.41	396.72	399.54	404.39	398.90
65	243.48	244.15	243.73	243.98	244.51	243.42	245.09
66	251.85	251.69	251.80	251.61	252.36	252.26*	253.25
67	282.03	282.85	283.21	282.69	285.34	281.99	283.38
68	397.12	396.86	394.15	396.54	389.09	387.78*	395.02
69	246.58	246.64	246.57	246.66	247.10	246.82	247.14
70	302.34	302.30	302.44	302.30	304.20	302.50	302.47
71	384.16	384.30	386.19	384.44	389.12	384.53	386.78
72	287.59	289.86	288.26	289.05	286.06	287.96	287.06
73	191.09*	141.71*	173.55*	181.50*	94.53*	50.73*	199.83*
74	165.83	165.91	166.35	166.22	167.76	178.66	166.44
75	210.39	211.18	209.71*	210.45	209.88*	208.75*	213.03
76	177.87	178.57	180.39*	178.81*	178.80	178.81*	189.33
77	245.17	245.55	244.27	245.39	244.20	242.86*	246.06
78	355.14	355.82	356.27	355.67	358.41	356.47	357.21
79	467.61	468.98	465.28	468.08	464.55	461.58*	468.57
80	280.88	280.60	281.53	280.82	282.82	281.63	283.13
81	255.51	256.92	256.83*	255.93	257.03*	254.55*	259.94
82	241.41	242.60	241.94	242.38	241.39	241.38	241.38
83	157.56	157.91	157.97	157.87	158.58	157.87	158.61
84	223.20	223.60	223.30	223.53	223.40	223.32	223.53
85	413.17	414.16	414.03	414.12	413.45	429.81*	413.59
86	203.36	203.63	203.66	203.60	204.74	203.47	203.72
87	296.28	296.70	296.18*	296.86	296.42	297.25	298.98
88	244.69	245.53	245.40	245.48	246.19	244.46	245.44
89	215.09	217.43	215.75	217.71	214.10	220.15	215.96
90	438.44	438.32	439.38	438.42	442.63	439.03	439.53
91	313.97	315.15	315.05	314.71	316.60	313.85	316.90
92	181.75	183.29	182.68	182.96	183.07	182.74	183.32
93	293.53	293.45	292.88	293.56	293.23	294.52	293.51
94	237.26	237.65	237.37	237.50	237.60	237.65	237.58
95	266.92	267.35	267.04	267.24	266.86	266.89	266.70
96	270.42	271.02	270.26	270.84	271.11	270.27	272.36
97	311.82	312.97	310.42*	311.27	312.27*	308.42*	313.34
98	351.00	351.90	351.76	351.83	352.45	352.05	351.76
99	309.11	309.30	309.77	309.32	311.49	309.64	310.00
100	337.88	338.77	339.03	338.52	339.96	337.90	340.87
101	270.40	271.00	269.82	270.83	267.09	267.09	269.63
102	258.74	259.24	259.11	259.19	259.14	258.96	259.04
103	287.53	287.62	288.64	287.47	289.46	287.26	289.92
104	190.25	190.58	190.70	190.56	192.05	190.28	190.70
105	322.21	322.48	322.18	322.31	323.40	322.26	322.21
106	193.19	193.88	192.14*	193.36	191.72	187.23*	194.25
107	303.02	303.61	303.42	303.51	304.31	303.30	303.54

Note: * denotes un-convergence

Chapter 4: Least Squares Approaches to Estimating Parameters of Logit-Logistic and Johnson's S_B

Summary

In this chapter, following the fundamental theory of the order statistics, we describe two least squares (LS) methods in fitting distributions, the percentile-based regression method and the cumulative distribution function (CDF) based regression method. The performance of the two LS methods and the MLE is compared in terms of a number of goodness-of-fit statistics, for both Johnson's S_B and the Logit-Logistic using the Chinese fir data set. Meanwhile, comparison of the Logit-Logistic and S_B under each estimation method in terms of these measures of fit is made. It was shown that the CDF-based performed best among the three compared estimation methods, and that overall the percentile-based LS better than the MLE, but with the exception of the Logit-Logistic when both the lower bound and the scale parameters were predetermined with the Knoebel-Burkhart method. The overall out-performance of the Logit-Logistic over S_B is consistent with the result we obtained in Chapter 3. We suppose this is due to the more flexibility of the Logit-Logistic than S_B in terms of the area covered in the (skewness-kurtosis) shape plane.

4.1 Introduction

In Chapter 3, the Logit-Logistic was introduced into forest diameter distribution modelling and its performance was compared with the S_B and other distributional models. The overall superior performance of the Logit-Logistic over the other models was found not only analytically in terms of the model coverage in the skewness-kurtosis shape plane, but also empirically on a large dataset of Chinese fir, using the (log) likelihood (essentially equivalent to the AIC criterion) as the comparison criterion resulted from the maximum likelihood estimation (MLE) method adopted to estimate parameters.

MLE is generally considered the best as it is asymptotically the most efficient method, and thus it is the most frequently used method to estimate parameters of distributions. However, the MLE does not exist in cases where the likelihood function can be made arbitrarily large. This occurs, for example, to distributions whose range depends on their parameters, such as the three-parameter lognormal, Weibull, and gamma distributions (see Cheng and Amin 1983, Castillo and Hadi 1995) and the four-parameter S_B and Logit-Logistic as we found in our simulation study. A numerical example we encountered in Chapter 3 is plot 73 sample distribution whose frequency curve is inverse J-shaped, on which none of the compared distributional models converged.

On the other hand, many other methods have been proposed to estimate the parameters of distributions, such as the method of moment. Particularly taking the S_B for example, these methods can be found in the statistical literature as well as the forestry literature, including the moment method (Johnson 1949), the four percentile method (Slifker and Shapiro 1980), the Knoebel-Burkhart method (Knoebel and Burkhart 1991), the mode method (Hafley and Buford 1985), and the regression methods (Zhou and McTague 1996, Kamziah et al. 1999). The regression methods (linear or nonlinear) have been consistently found to be superior for estimating parameters of the S_B (Zhou and McTague 1996, Kamziah et al. 1999, Zhang et al. 2003) in forestry applications. Considering the similarity between the S_B and the Logit-Logistic, we may expect the regression methods to perform better than the MLE in estimating the parameters of the Logit-Logistic.

The linear regression method proposed by Zhou and McTague (1996) and the nonlinear method by Kamziah et al. (1999) can be regarded as percentile (quantile) based regression methods, based on the theory of order statistics. On the other hand, Wilson (1983) suggested the other least squares method to estimate parameters of the S_B, which is cumulative distribution function (CDF) based, following the theory of order statistics. Both types of regression methods provide alternatives to the MLE and have an advantage in computation that most of the statistical software packages currently available (S-Plus, SAS, SPSS,...) support the LS estimation but may not support the MLE, therefore it is worthwhile to introducing the LS methods for fitting the Logit-Logistic distribution and comparing their performance with the MLE. Furthermore, as the relative performance of different distributional models may depended on the estimation method used, one example being Zhang et al. (2003), it is then interesting to see if the superior performance of the Logit-Logistic over the other models still holds under the LS methods.

In this Chapter, firstly we briefly introduce the fundamental theory of the order statistics. Then based on order statistics theory, we describe the percentile-based regression method and introduce the CDF-based regression method. Subsequently we compare the performance of the two least squares (LS) methods with MLE in terms of several goodness-of-fit statistics, for both the S_B and the Logit-Logistic using the Chinese fir data sets. Finally, comparisons of the Logit-Logistic and S_B via these measures of fit and the sum of squared errors with the LS methods is made, which is a complementary to the comparison conducted based on the log-likelihood in Chapter 3. For simplicity, we limited our comparison of the Logit-Logistic with only the S_B.

We note that the two LS approaches are applicable to all continuous distributions in principle and inversely the applications in other distributions can be used to justify the use of the LS methods for the Logit-Logistic, however, in introducing the LS methods we put our emphases on the references related to the Weibull and S_B, as they are the most widely used distributional models in forestry.

4.2 Basic Properties of Order Statistics

Let x_1, x_2,..., x_n is an ordinary random sample (independent and identically distributed, i.i.d.) of size n from a given distribution with cumulative distribution function $F(x)$ and probability density function (PDF) $f(x)$, and $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ be the order statistics obtained by rearranging the ordinary sample in ascending order.

4.2.1 CDF, PDF and Moments of Order Statistics

The CDF of the ith order statistic, $x_{(i)}$ (i=1,2,...,n) is given by:

$$F_{X_{(i)}}(x) = \Pr[X_{(i)} \leq x] = \Pr[\text{at least } i \text{ of the } x_{(r)} \text{ are less than or equal to } x]$$

$$= \sum_{r=i}^{n} \binom{n}{r} [F(x)]^r [1 - F(x)]^{n-r} \tag{4.1}$$

and that the PDF is:

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} [1 - F(x)]^{n-i} f(x) \tag{4.2}$$

(see Cox and Hinkley 1979, David 1985).

The kth moment of the ith order statistic denoted by $\mu_{(i)}^{(k)}$ (k = 0,1,..., and $1 \leq i \leq n$) is given by:

$$\mu_{(i)}^{(k)} = E(x_{(i)}^k) = \frac{n!}{(i-1)!(n-i)!} \int_{-\infty}^{\infty} x^k [F(x)]^{i-1} [1 - F(x)]^{n-i} f(x) dx$$

$$= \frac{n!}{(i-1)!(n-i)!} \int_0^1 [F^{-1}(x)]^k x^{i-1} (1-x)^{n-i} dx \tag{4.3}$$

For the first moment, we have

$$\mu_{(i)}^{(1)} = E(x_{(i)}) = \frac{n!}{(i-1)!(n-i)!} \int_0^1 F^{-1}(x) x^{i-1} (1-x)^{n-i} dx \tag{4.4}$$

which is the expected value of the observed order statistics.

4.2.2 Uniformized Order Statistics

It is well known that the transformation $p = F(x)$ leads to variables $p_1, p_2, \ldots, p_n$ which are i.i.d. from the uniform distribution on (0,1), hence the transformation is sometimes termed uniformization. And clearly $p_{(i)}=F(x_{(i)})$. It then follows from (4.1) and (4.2) that the uniformized order statistics $p_{(i)}$ have beta distributions with parameters $(i, n\text{-}i\text{+}1)$. That is, the PDF of the $p_{(i)}$ is given as:

$$g(p_{(i)}) = \frac{n!}{(i-1)!(n-i)!} p_{(i)}^{i-1}(1-p_{(i)})^{n-i} \tag{4.5}$$

with the expectations, variance and covariance given by

$$E(p_{(i)}) = \frac{i}{n+1} \tag{4.6}$$

$$Var(p_{(i)}) = \frac{i(n-i+1)}{(n+1)^2(n+2)} \tag{4.7}$$

$$Cov(p_{(i)}, p_{(j)}) = \frac{i(n-j+1)}{(n+1)^2(n+2)}, \ (i<j) \tag{4.8}$$

(see Cox and Hinkley 1979). It is noted that this uniformized transformation is applicable to all the continuous distributions, with the advantage of simple closed forms of expectations, variance and covariance of the transformed order statistics being existed. Other transformations of the order statistics $x_{(i)}$ specific to distributional models may be considered, two examples given in the context of the following section.

4.3 Nonlinear LS Estimations Based on (Uniformized) Order Statistics

Based on the basic properties of order statistics and uniformized order statistics as introduced above, there are two ways of seemingly using least squares methods to estimate parameters, the first one based on the order statistics $x_{(i)}$ and their expectations, and the other based on the transformed order statistics $F(x_{(i)})$ and their expectations, termed as percentile-based LS and CDF-based LS respectively. We introduce these two approaches in detail in the following.

4.3.1 Percentile-Based LS Estimation

Based on equation (4.4), the percentile-based LS method estimates parameter by minimizing the sum of squares of the difference between the observed and expected values of the order statistics $x_{(i)}$, that is,

$$\sum_{i=1}^{n}[x_{(i)} - E(x_{(i)})]^2 \tag{4.9}$$

Or in terms of regression analysis, (4.9) be expressed as,

$$x_{(i)} = E(x_{(i)}) + e_i \tag{4.10}$$

where e_i are error terms. For the Weibull distribution defined as,

$$F(x) = 1 - e^{-(\frac{x-a}{b})^c} \tag{4.11}$$

where a, b, c are parameters, Weibull (1967) derived the expected values, variance and covariance of the transformed order statistics, $y = c\log_{10}(\frac{x-a}{b})$, then used this LS method (weighted or unweighted depending on using variance information or not) to estimate the parameters. Mykytka and Ramberg (1979) derived the kth moment (thus expectation, variance and covariance) of the order statistics for the generalized lambda distribution (GLD, Ramberg et al. 1979) as defined by,

$$x_p = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2} \tag{4.12}$$

where λ_i (i=1,2,3,4) are parameters, x_p is quantile corresponding to probability p. Later on Öztürk and Dale (1985) used this LS estimator to GLD.

However, it is noted that the expected values of the order statistics are rather complicated and always difficult to obtain for even simple distributions. Then in practice it is common procedure to approximate these expected values using theoretical (population) quantiles (Cox and Hinkley 1979, Castillo and Hadi 1995), which are given by the inverse function (percentile/quantile function) as,

$$E(x_{(i)}) \cong F^{-1}(p_{(i)}) \tag{4.13}$$

That is, to estimate parameters by minimizing

$$\sum_{i=1}^{n}[x_{(i)} - F^{-1}(p_{(i)})]^2 \tag{4.14}$$

Equivalently (4.13) or (4.14) be expressed as,

$$x_{(i)} = F^{-1}(p_{(i)}) + e_i \tag{4.15}$$

This approximation could be justified by considering the asymptotic distributions of order statistics, that is, for $i=np$, $0<p<1$, $n\rightarrow\infty$, $x_{(i)}$ is asymptotically normal with mean ξ_p given by the population percentile (quantile) function

$$\xi_p = F^{-1}(p) \tag{4.16}$$

and variance by

$$\frac{p(1-p)}{nf^2(\xi_p)} \tag{4.17}$$

(See Cox and Hinkley 1979, David 1985). Also, this approximation could be considered quite natural by noting the fact that a continuous probability distribution can be alternatively defined by its percentile (quantile) function rather than by its distribution function or by its density function as usually (Ramberg et al. 1979). Therefore, it is quite natural to use some nonlinear regression techniques for parameter estimation based on the percentile function by using (4.14) as objective function in stead of (4.9), that is, to estimate parameters by minimizing the sum of squared difference between sample percentiles and population percentiles. However, we may have to realize the fact that by doing so we actually use the theoretical percentiles as approximations to the expected values of order statistics (the sample percentiles), though in practice we may directly apply LS to model (4.15) without explicitly resorting to the theory of order statistics. This will help to understand what is behind of this percentile-based LS method defined as (4.14) or (4.15).

When referring to the percentile-based LS in the following, we mean (4.15), not (4.10) any more. For some distributions whose percentile functions exist in closed form, this estimation method has been used quite successfully, such as the Weibull in reliability analysis (Duffy et al. 1993, Gross 1996), generalized exponential distribution (Gupta and Kundu 2001), and generalized Rayleigh distribution (Kundu and Raqab 2005). More importantly, some more flexible families of distributions can be derived by using the transformation method based on quantile function, indicating a natural way of parameter estimation using such percentile-based LS. For example, the GLD families can be obtained from transformation on the uniform, for which the MLE may not apply. For the S_B, Chapter 2 (see also Rennolls and Wang 2005) presents an alternative transformational definition by applying a 4-parameter linear-logistic function,

$$x = \xi + \frac{\lambda}{1+\exp(-(\gamma+\delta z))} \tag{4.18}$$

to a standard normal distribution. Similarly the Logit-Logistic is readily obtained by replacing the standard normal with the standard Logistic. By replacing z with the standard normal percentile $\Phi^{-1}(p)$ or standard Logistic percentile $\ln\frac{p}{1-p}$, the quantile functions for the S_B and Logit-Logistic are given as,

$$x = \xi + \frac{\lambda}{1+\exp(-(\gamma+\delta\Phi^{-1}(p)))} \tag{4.19}$$

$$x = \xi + \frac{\lambda}{1+e^{-\gamma}(\frac{p}{1-p})^{-\delta}} \tag{4.20}$$

For these families of distributions, which are defined by their quantile functions, the use of quantile-based LS for fitting is then a natural choice. Öztürk and Dale (1985) suggested this percentile-based LS method for fitting the GLD. Kamziah et al. (1999) proposed this method for estimating the parameters of the S_B based on its percentile function,

$$x_{(i)} = \xi + \frac{\lambda}{1 + \exp(-(\gamma + \delta\Phi^{-1}(p_{(i)})))} + e_i \tag{4.21}$$

Zhou and McTague (1996) used linear LS to fit S_B with the location parameter ξ and range parameter λ predetermined, that is,

$$z_{(i)} = \frac{\ln \frac{x_{(i)} - \hat{\xi}}{\hat{\xi} + \hat{\lambda} - x_{(i)}} - \gamma}{\delta} + e_i = \frac{-\gamma}{\delta} + \frac{1}{\delta} \ln \frac{x_{(i)} - \hat{\xi}}{\hat{\xi} + \hat{\lambda} - x_{(i)}} + e_i \tag{4.22}$$

where $z_{(i)} = \Phi^{-1}(p_{(i)})$ and $p_{(i)}$ are selected at 0.1, 0.2, …, 0.9. This LS approach has been found to be superior for estimating the parameters of the S_B (Zhou and McTague 1996, Zhang et al. 2003). This method is actually based on the transformed order statistics $\frac{\ln \frac{x_{(i)} - \hat{\xi}}{\hat{\xi} + \hat{\lambda} - x_{(i)}} - \gamma}{\delta}$. If we consider the transformed order statistics $\ln \frac{x_{(i)} - \hat{\xi}}{\hat{\xi} + \hat{\lambda} - x_{(i)}}$, the following linear regression model may be used,

$$\ln \frac{x_{(i)} - \hat{\xi}}{\hat{\xi} + \hat{\lambda} - x_{(i)}} = \gamma + \delta z_{(i)} + e_i = \gamma + \delta\Phi^{-1}(p_{(i)}) + e_i \tag{4.23}$$

Both (4.22) and (4.23) can be regarded as some variants of (4.21), transformed from nonlinear to linear for utilizing the facility of the linear regression analysis. It is noted that in regression analysis, (4.22) and (4.23) differ in which variable is dependent and which is independent for the simple linear regression models. For the cases where the two variables are closely correlated, which occurs if "correct" distributional model is selected, the two regression lines would be almost equivalent. As Kamziah et al. (1999) found their nonlinear LS method outperformed all the other methods of parameter estimation (including Zhou and McTaugue's linear LS) for fitting S_B, therefore, for the S_B with the location parameter ξ and range parameter λ predetermined, the nonlinear LS approach may still be preferred.

For the percentile-based LS method to be employed, some estimate of the uniformized order statistic $p_{(i)}= F(x_{(i)})$ has to be used to obtain theoretical/population percentile $F^{-1}(p_{(i)})$. This is equivalent to choosing the probability plotting positions as in P-P plots or Q-Q plots. Various plotting positions have been proposed (Looney and Gulledge 1984), including,

$$p_{(i)} = \frac{i-0.5}{n} \tag{4.24}$$

$$p_{(i)} = \frac{i}{n+1} \tag{4.25}$$

$$p_{(i)} = \frac{i-0.375}{n+0.25} \tag{4.26}$$

$$p_{(i)} = \frac{i-0.3}{n+0.4} \tag{4.27}$$

All the above estimates can be obtained from a general equation given as,

$$p_i = \frac{i-c}{n+1-2c} \quad (0 \le c \le 1) \tag{4.28}$$

with different constant c assigned. In this study, we use $p_{(i)} = \frac{i}{n+1}$ as it is one of the most used estimators of $F(x_{(i)})$ and also the expected value of $F(x_{(i)})$.

The ordinary LS method (OLS) can then be used by minimizing the sum of squares defined by (4.14). It may be mentioned that in doing so, we tacitly but incorrectly assume that the order statistics are i.i.d. In fact, these order statistics, $x_{(i)}$, are neither independent nor identically distributed though the original x's are. Theoretically, the variance and covariance of the $x_{(i)}$'s and thus the $e_{(i)}$'s can be obtained by using the moment formula (4.3), but would be too complicated to limit such derivation. Some approximations to the variance and covariance may be made by considering the asymptotic distributions of order statistics, that is, for $r_i=np_i$ with $0<p_i<1$ ($i=1,\ldots,k$), as $n\to\infty$ the order statistics $x_{(r_1)}, \ldots, x_{(r_k)}$ are asymptotically multivariate normal with mean and covariance matrix determined by (4.16), (4.17) and

$$Cov(x_{(r_i)}, x_{(r_j)}) \sim \frac{p_i(1-p_j)}{nf(\xi_{p_i})f(\xi_{p_j})} \ (r_i<r_j) \tag{4.29}$$

(see Cox and Hinkley 1979). However, even this asymptotic approximation may be also too complicated to be used in LS estimation. Thus, in practice, we may ignore the heterogeneity and correlations of the order statistics, but nevertheless ordinary least squares estimation can be made, which is one common practice in regression analysis when the error terms are not i.i.d.

It is noted that this percentile-based LS method could be considered as the extension of the percentile-matching method, which has been used for fitting distributions for a long time. The percentile-matching method estimates parameters by equating observed (sample) percentiles to their theoretical (population) values $F^{-1}(p_{(i)})$, the number of percentiles depending on the number of parameters to be estimated. This percentile-matching method has been used for fitting the Weibull (Zarnoch and Dell 1985, Shiver 1988, Newberry et al. 1993), and the S_B (Johnson 1949, Bukac 1972, Mage 1980, Slifker and Shapiro 1980, Wheeler 1980, Newberry and Burk 1985, Shayib 1989, Knoebel and Burkhart 1991, Siekierski 1992, Newberry et al. 1993). However, in using such quantile-based estimators for fitting, the question can arise as to whether a highly-selective set, a more representative set, or all the quantiles should be used. This becomes the more general question of whether and how quantiles should be "weighted" in quantile-based procedures (Rayner and MacGillivray 2002). Meanwhile, for such percentile-based estimators, we have to realize that different percentile choices may always lead to different parameter estimates although we may choose some "special/important" percentiles to achieve better fit, as Newberry et al. (1993) evaluated such choices for the Weibull and S_B. Also, we may suppose that the "selective" percentile-based estimators may be partially due to the computation difficulty encountered before. Therefore, the percentile-based LS method could be regarded as a method using all of the percentile information and by which the parameter estimates are unique.

4.3.2 CDF-Based LS Estimation

Wilson (1983) proposed a least squares criterion which minimizes the sum of squared difference between the uniformized order statistics $p_{(i)}=F(x_{(i)})$ and their expected values given by equation (4.6) for fitting Johnson's systems of distribution. That is, to estimate parameters by minimizing

$$\sum_{i=1}^{n}[F(x_{(i)})-\frac{i}{n+1}]^2 \tag{4.30}$$

This LS approach is termed as CDF-based regression method in contrary to the percentile-based LS as discussed above. The regression model can be given as the following,

$$F(x_{(i)})=\frac{i}{n+1}+e_i \tag{4.31}$$

The error $e_i=F(x_{(i)})-\frac{i}{n+1}$ represents the random deviation between the observed and expected values of the ith uniformized order statistic, and the covariance between e_i and e_j is given by equation (4.8), or equivalently,

$$Cov(e_i,e_j)=\frac{i(n-j+1)}{(n+1)^2(n+2)},\cdots 1\le i\le j\le n \tag{4.32}$$

It is noted that this LS approach is obviously different from what are the traditional regression models in that the dependent variable, $F(x_{(i)})$, is not "observed" and that the expectations act as the "independent" variable! Nonetheless, the LS estimation can be made. Noting that these uniformized order statistics (thus the errors e_i's) are dependent and nonidentically Beta distributed, there are several variants of the LS methods. By ignoring the variance and covariance, the ordinary least squares estimator is given by minimizing (4.30).

Considering the heteroscedasticity only, the weighted LS (WLS) estimator is obtained by minimizing

$$\sum_{i=1}^{n} w_i [F(x_{(i)}) - \frac{i}{n+1}]^2 \tag{4.33}$$

where $w_i = \dfrac{1}{Var(F(x_{(i)}))} = \dfrac{(n+1)^2(n+2)}{i(n-i+1)}$ which is determined by (4.7). More generally, taking both the heteroscedasticity and correlation into account, the generalized LS (GLS) may be used with the "weights" (W) are given by the inverse of the variance and covariance matrix as, $W = V^{-1}$, $V = [Cov(e_i\ , e_j)]$, with the WLS as a special case of GLS in that $W = D^{-1}$, $D = diag\{Var(e_1),\ldots, Var(e_n)\}$. These two variants of the OLS method, WLS and GLS, have been proposed for fitting the Johnson cumulative probability distributions (Swain and Wilson 1985, Swain et al. 1988). However, for most i and moderate sample sizes (n > 30), the Beta distributions will be fairly normal and the correlation could be neglected, so that the OLS is suitable for estimation. In small and medium samples, the GLS approach can yield relatively large bias in the fitted CDF as well as in the GLS parameter estimators, while better results can be obtained by the WLS (Storer et al. 1988). Therefore, the OLS and WLS estimators have been used widely in simulation study, to fitting Johnson's systems of distribution (Swain and Wilson 1985, Swain et al. 1988, Storer et al. 1988, DeBrota et al. 1988), to Bézier distributions (Wagner and Wilson 1996), to generalized exponential distribution (Gupta and Kundu 2001), and to generalized Rayleigh distribution (Kundu and Raqab 2005).

It is noted that, intuitively, both percentile-based LS and CDF-based LS could be regarded as numerical refinement of the graphical methods in estimating parameters, corresponding to the well-known Quantile-Quantile (Q-Q) plots of the fitted/theoretical distribution quantiles versus the sample quantiles and Percentile-Percentile (P-P) plots of the fitted distribution probability versus the empirical probability (plotting positions), respectively. Both Q-Q plots and P-P plots are always used to graphically show whether the hypothesized distribution adequately fits the sample data or not, and these informal graphical techniques are sometimes used to estimate parameters of the alternative distributions, especially for those location-scale families of distribution.

4.4 Case-Study with Chinese Fir Diameter Distribution Data

The percentile-based regression models for the S_B and Logit-Logistic can be expressed as,

$$x_{(i)} = \xi + \frac{\lambda}{1+\exp(-(\gamma+\delta\Phi^{-1}(p_{(i)})))} + e_i \tag{4.21}$$

$$x_{(i)} = \xi + \frac{\lambda}{1+e^{-\gamma}(\frac{p_{(i)}}{1-p_{(i)}})^{-\delta}} + e_i \tag{4.34}$$

CDF-Based regression model is,

$$\Phi(z_{(i)}) = \frac{i}{n+1} + e_i \tag{4.35}$$

where $z_{(i)} = \dfrac{\ln\dfrac{x_{(i)}-\xi}{\xi+\lambda-x_{(i)}} - \gamma}{\delta}$,

and,

$$F(x_{(i)}) = \frac{1}{1+e^{\frac{r}{\delta}}(\frac{x_{(i)}-\xi}{\xi+\lambda-x_{(i)}})^{-\frac{1}{\delta}}} = \frac{i}{n+1} + e_i \tag{4.36}$$

4.4.1 The Case-Study Data

The Chinese fir data set consists of 107 plots collected from Chinese fir plantations, which have been used in Chapter 3.

4.4.2 Model Fits to the Case-Study Diameter Distributions

We use MLE and the two LS methods to estimate parameters of the S_B and Logit-Logistic. The MLE method is to fit both models by minimizing the minus log-likelihood (details see Chapter 3 or Wang and Rennolls 2005). For both percentile-based and CDF-based LS approaches, we use OLS to fit models (4.21), (4.34) to (4.36) by ignoring heterogeneity of and dependency between order statistics, since our sample sizes are generally large (ranging from 63 to 239, with mean of about 119). All the least squares fitting is carried out using the S-Plus (Mathsoft 1999) function *nlregb* (local minimizer for sums of squares of nonlinear functions subject to bound-constrained parameters), the S-Plus codes of fitting Logit-Logistic being given in the Appendix, Code 4.1, 4.2, and 4.3.

Rather than to test if one distributional model under specific parameter estimation method is adequate to fit the data or not, which is one way of ranking different parameter estimation methods for a specific distributional model as adopted by Kamziah et al. (1999), we emphasize to see which estimation method (MLE, percentile-based LS, CDF-based LS) gives better fit, in terms of several measures of fit. For the tree diameter distribution modelling, Reynolds et al. (1988) suggested an "error index" (*EI*) as a measure of fit for selecting and validating distributional models. This error index is a weighted sum of the absolute differences between predicted and observed numbers of diameters in each diameter class, which is defined as,

$$EI = N\sum_{j=1}^{n}\left|\int_{I_j} w(x)d\hat{F}(x) - \int_{I_j} w(x)dF_n(x)\right| \tag{4.37}$$

where x is DBH, $\hat{F}(x)$ is the estimated CDF, $F_n(x)$ is the empirical CDF, $w(x)$ is a weight function of diameter, I_j is the jth diameter class (j=1,2,...,n) and N is the sample size. Simply let $w(x)$=1, this error index collapses to,

$$EI = \sum_{i=1}^{n} |O_i - E_i| \tag{4.38}$$

where O_i and E_i are the observed and predicted/expected numbers of trees respectively in the ith diameter class. This simplified error index has been used to evaluate different methods of parameter estimation for Weibull and S_B (Zhang et al. 2003). Two EDF-based goodness-of-fit test statistics are also used as measures of fit, the Kolmogorov-Smirnov (D) and Cramér-von Mises (W^2). They are defined as,

$$D = \sup_x | F_n(x) - \hat{F}(x) |$$

$$W^2 = n\int_{-\infty}^{\infty} \{F_n(x) - \hat{F}(x)\}^2 d\hat{F}(x)$$

Let $z_i = \hat{F}(x_i)$, we have

$$D^+ = \max_i \{\frac{i}{n} - z_i\},\ D^- = \max_i \{z_i - \frac{(i-1)}{n}\},\ D = \max(D^+, D^-) \tag{4.39}$$

$$W^2 = \sum_{i=1}^{n} \{z_i - \frac{(2i-1)}{2n}\}^2 + \frac{1}{12n} = \sum_{i=1}^{n} \{z_i - \frac{(i-0.5)}{n}\}^2 + \frac{1}{12n} \tag{4.40}$$

It is clear that the Kolmogrov-Smirnov statistic is based on the maximum distance between the empirical CDF and the hypothesized CDF, while the Cramér-von Mises statistic is an overall measure of the squared distance between the EDF and the true CDF evaluated at all the observed data values. Compared to the Cramér-von Mises statistic, the error index is an essentially "grouped" measure of overall differences between fitted and empirical CDFs. We then consider an "un-grouped" version of EI, that is,

$$EI2 = \sum_{i=1}^{n} \left|\hat{F}(x_i) - F_n(x_i)\right| = \sum_{i=1}^{n} \left|\hat{F}(x_i) - \frac{i-0.5}{n}\right| \tag{4.41}$$

These four measures of fit were computed under these three estimation methods for each sample, and the performance of the three methods was compared under each fit measure. Table 4.1 lists the proportions of the first estimation method better than the second for each pair of compared estimation methods under each measure of fit, for fitting the S_B and the Logit-Logistic. Table 4.2 lists the proportions of the Logit-Logistic better than S_B compared on each measure of fit under the three estimation methods. Although we already compared the performance of both models in terms of these test statistics with the MLE used for parameter estimation in Chapter 3, we purposely included these results here for further comparison with the LS estimation methods.

Table 4.1. Superior Proportions of 1st Estimation Method than 2nd for Fitting Johnson's S_B and Logit-Logistic Compared on 4 Measures of Fit

Compared Methods (1st 2nd)	Compared Statistics	Superior Proportion: Johnson's S_B	Superior Proportion: Logit-Logistic
CLS-MLE	*D*	98/107	91/107
	W^2	106/107	104/107
	EI2	105/107	98/107
	EI	60/107	57/107
PLS-MLE	*D*	76/107	74/107
	W^2	75/107	75/107
	EI2	77/107	78/107
	EI	65/107	68/107
CLS-PLS	*D*	91/107	82/107
	W^2	104/107	103/107
	EI2	103/107	97/107
	EI	54/107	42/107

Note: MLE-Maximum Likelihood Estimation, CLS-CDF based LS, PLS-Percentile-based LS, *D*-Kolmogorov-Smirnov Statistic, W^2-Cramér-von Mises statistic, *EI*-Reynold's Error Index, *EI2*-ungrouped Error Index.

From Table 4.1, it was shown that in terms of statistics *D*, W^2, *EI2*, and *EI*, both LS methods overall performed better than the MLE, for fitting the S_B and the Logit-Logistic. For the two LS methods, in terms of *D*, W^2, *EI2*, the CDF-based LS outperformed the percentile-based LS for both models, while in terms of Reynolds et al.'s error index, the two methods performed almost equally (54 in contrast to 53) for the S_B and the percentile-based LS better than the

CDF-based (65 out of 107) for the Logit-Logistic. It is noted that the difference in performance of the compared three estimation methods in terms of *EI* are all smaller than those in terms of the other statistics, for both S_B and Logit-Logistic, and this may be due to the fact that this index is a "grouped" measure by which some difference may have been "smoothed". Also noted is that the best performance of the CDF-based LS (except in the case of the Logit-Logistic in terms of *EI*) is not unexpected, since this method uses the sum of squared differences between the uniformized order statistics (fitted probability) and their expected probabilities as the objective function in fitting which naturally has more close link to the measures of goodness-of-fit used for comparisons.

Table 4.2. Superior Proportions of Logit-Logistic than Johnson's S_B Compared on 4 Measures of Fit under Three Estimation Methods

Estimation Methods	Compared Statistics	Superior Proportion
MLE	*D*	79/107
	W^2	71/107
	EI2	60/107
	EI	48/107
CLS	*D*	53/107
	W^2	50/107
	EI2	45/107
	EI	52/107
PLS	*D*	62/107
	W^2	69/107
	EI2	64/107
	EI	58/107

Note: MLE-Maximum Likelihood Estimation, CLS-CDF based LS, PLS-Percentile-based LS, *D*-Kolmogorov-Smirnov Statistic, W^2-Cramér-von Mises statistic, *EI*-Reynold's Error Index, *EI2*-ungrouped Error Index.

From Table 4.2, with the MLE, the Logit-Logistic performed better than S_B in terms of *D*, W^2, and *EI2* (79, 71 and 60 out of 107, respectively), but worse in terms of *EI* (48 out of 107). With the CDF-based LS, the S_B performed a little better than the Logit-Logistic in terms of all the statistics used for comparisons here, while with the percentile-based LS, the Logit-Logistic performed better than S_B. This may indicate that different parameter estimation methods affect the comparative performance of the Logit-Logistic and S_B, but in general we may conclude that the Logit-Logistic outperformed than S_B.

When using the three methods of parameter estimation, the lower bound parameter ξ and the scale parameter λ are restricted by $0 \leq \xi < D_{\min}$ and $\xi + \lambda > D_{\max}$ in fitting, where $D_{\min}$ and $D_{\max}$ are the minimum and maximum value of diameters in each plot respectively. In Chapter 3, it was found that with the MLE, there are more proportions of estimates of ξ as "zero" for S_B than Logit-Logistic, 33 out of 107 sample-plots in contrast to 7 out of 107. For both LS methods we used, it was also found more proportions of estimates of ξ as "zero" for the S_B than Logit-Logistic, that is, 56 out of 107 in contrast to 36 out of 107 with the CDF-based LS, and 57 out of 107 in contrast to 20 out of 107 with the percentile-based LS. Therefore, compared to the MLE, there was much increase in such proportions with both LS approaches. Furthermore, it was found that both LS methods are likely to estimate the minimum parameter (ξ) as $D_{\min}$, 7 and 19 out of 107 for the S_B and Logit-Logistic respectively with the CDF LS, 3 and 7 out of 107 respectively with the percentile LS, and to estimate the maximum parameter ($\xi + \lambda$) as $D_{\max}$, 25 and 35 out of 107 for S_B and Logit-Logistic respectively with the CDF LS, 5 and 13 out of 107 respectively with the percentile LS, which is not observed with the MLE.

4.4.3 Model Fits with Parameter ξ, λ Predetermined

To complete the comparison of the three estimation methods and the comparison of the S_B and the Logit-Logistic under each of estimation methods, but in the context of the lower bound parameter ξ and the scale parameter λ being predetermined which is a common procedure in fitting the S_B in forestry practice, we predetermined these two parameters in the sense of Knoebel-Burkhart (1991), that is,

$$\xi = D_{\min} - 1.3 \tag{4.42}$$

$$\lambda = D_{\max} - D_{\min} + 5.1 \tag{4.43}$$

We fitted the two distribution models using three estimation methods. The performance of each model under the three methods was then compared, and Table 4.3 lists the comparative results. Table 4.4 lists the comparison results of both models under each estimation method.

Table 4.3. Superior Proportions of 1st Estimation Method than 2nd for Fitting Johnson's S_B and Logit-Logistic Compared on 4 Measures of Fit (ξ, λ predetermined)

Compared Methods 1st 2nd	Compared Statistics	Superior Proportion Johnson's S_B	Logit-Logistic
CLS-MLE	*D*	101/107	76/107
	W^2	105/107	101/107
	EI2	102/107	92/107
	EI	87/107	77/107
PLS-MLE	*D*	100/107	30/107
	W^2	101/107	33/107
	EI2	101/107	36/107
	EI	86/107	37/107
CLS-PLS	*D*	92/107	90/107
	W^2	102/107	105/107
	EI2	91/107	98/107
	EI	82/107	79/107

Note: MLE-Maximum Likelihood Estimation, CLS-CDF based LS, PLS-Percentile-based LS, *D*-Kolmogorov-Smirnov Statistic, W^2-Cramér-von Mises statistic, *EI*-Reynold's Error Index, *EI2*-ungrouped Error Index.

Table 4.4. Superior Proportions of Logit-Logistic than Johnson's S_B Compared on 4 Measures of Fit under Three Estimation Methods (ξ, λ predetermined)

Estimation Methods	Compared Statistics	Superior Proportion
MLE	*D*	75/107
	W^2	79/107
	EI2	81/107
	EI	67/107
CLS	*D*	55/107
	W^2	61/107
	EI2	59/107
	EI	53/107
PLS	*D*	51/107
	W^2	60/107
	EI2	62/107
	EI	57/107

Note: MLE-Maximum Likelihood Estimation, CLS-CDF based LS, PLS-Percentile-based LS, *D*-Kolmogorov-Smirnov Statistic, W^2-Cramér-von Mises statistic, *EI*-Reynold's Error Index, *EI2*-ungrouped Error Index.

From Table 4.3, it was shown that in terms of statistics D, W^2 and $EI2$ and EI, CDF-based LS overall performed best for fitting both S_B and Logit-Logistic, and percentile-based LS better than the MLE for the S_B. This is generally in agreement with the results with all the four parameters being estimated. However, it is noted that for the Logit-Logistic, the MLE performed better than the percentile-based LS, inverting the comparative result for this model when all parameters are to be estimated. The reason is not understood.

From Table 4.4, with all the three estimation methods, the Logit-Logistic overall performed better than the S_B in terms of all statistics considered. But with both LS methods, this out-performance decreased compared to the MLE: both models performed roughly equally in terms of D and EI.

4.4.4 Comparison of Logit-Logistic with S_B in terms of Sum of Squared Errors (SSE)

In the above comparisons, for comparing the performance of the LS methods with the MLE for fitting distributions, we had to adopt some commonly used goodness-of-fit measures as comparison criteria. Meanwhile, we also used these criteria to compare the two distributional models. However, as we argued in Chapter 3 that comparing different models under the same estimation method by using the secondary criteria (the goodness-of-fit measures we used, say) is so marginal that we prefer the criteria directly resulted from the estimation process, we then used the (log) likelihood obtained from the MLE for model comparison in Chapter 3. Therefore, it would be more reasonable to use the sum of squared errors (SSE) resulted straightforward from the LS fitting as the criterion for the comparison of the S_B and Logit-Logistic under the LS estimation methods, which is a common practice in regression analysis. Table 4.5 lists the results.

Table 4.5. Superior Proportions of Logit-Logistic than Johnson's S_B Compared on SSE under the LS Estimation Methods

Estimation Methods	Superior Proportion	
	(ξ, λ) Estimated	(ξ, λ) predetermined
CLS	55/107	65/107
PLS	69/107	55/107

Note:CLS-CDF based LS, PLS-Percentile-based LS.

From Table 4.5, with both LS methods, the Logit-Logistic overall performed better than the S_B. However, in the case of all the four parameters being estimated simultaneously with the CDF-based LS and in the case of the (ξ, λ) predetermined with the percentile-based LS, the two models performed roughly equivalently (that is, 55 to 52).

4.5 Discussion

4.5.1 Measures of Model Fit in Favour of CDF-based Method

In general, the CDF-based performed best among the three compared estimation methods. We suppose this is mainly due to the fact that this method uses the sum of squared differences between the uniformized order statistics (fitted probability) and their expected probabilities as the objective function (see equation (4.30)) in fitting and this objective function could be regarded as an alternative but also a similar measure of model fit to those used for comparisons in this paper. For example, comparing the objective function of equation (4.30) and the Cramér-von Mises (W^2) statistic defined by (40), the difference between them is essentially only the different choices of the "plotting position"! Therefore, in this sense, we might not "reasonably" conclude that this method did outperform the others, since the measures we used for comparison is in favour of this method though we did not find other more reasonable measures. Also noted is that one drawback of this method, we think, is that this estimator is more likely to be affected by the boundary limits imposed on the lower bound parameter ξ and the scale parameter λ (then the upper bound, $\xi + \lambda$).

The out-performance of the percentile-based LS over the MLE for fitting the S_B is consistent with the comparison result by Kamziah et al. (1999). However, for the Logit-Logistic, the performance of these two methods was affected by whether the lower bound parameter ξ and the scale parameter λ being predetermined (with the Knoebel-Burkhart method) or not: with the two parameters predetermined, MLE performed better.

4.5.2 Other Estimation Methods

The fact that the CDF-based LS method is actually to minimize an alternative measure of goodness-of-fit may indicate more general approaches to estimating parameters of distributional models by optimizing goodness-of-fit statistics directly. Such an idea has been suggested by Starlinger et al. (1993). One way is to minimize the Cramér-von Mises (W^2) statistic defined by (4.40) using LS method, or more generally to minimize the sum of squared differences between the uniformized order statistics $p_{(i)}=F(x_{(i)})$ and an empirical estimate of $p_{(i)}$ (plotting positions), that is,

$$\sum_{i=1}^{n}[F(x_{(i)}) - \hat{p}_{(i)}]^2 \tag{4.44}$$

Other than the LS methods which take the sum of squares as the objective function in fitting, many other criteria (generally the L_p norms) could be used. For example, the Least Absolute Deviation (the L_1) is defined as

$$\sum_{i=1}^{n}| F(x_{(i)}) - \hat{p}_{(i)} | \tag{4.45}$$

in which the un-grouped version of Reynolds et al.'s Error Index defined by (4.41) is a special case. The L_∞ norm is defined as

$$\| F_n - \hat{F} \|_\infty = \sup_x | F_n(x) - \hat{F}(x) | \tag{4.46}$$

which is the Kolmogorov-Smirnov statistic D. The use of L_1 and L_∞ norms in estimating parameters of the Johnson distributions have been employed (DeBrota et al. 1988, Wagner and Wilson 1996, Wilson et al. 1988).

Similarly, these non-LS criteria could also be applied to the percentile-based LS methods. All of these variants of parameter estimation would be worthwhile to be exploited for their potential uses.

4.5.3 Performance of Logit-Logistic over S_B

Under each of the three estimation methods, the Logit-Logistic overall performed better than the S_B, which is consistent with the result we got in Chapter 3. We suppose this is due to the more flexibility of the Logit-Logistic than S_B in terms of the area covered in the (skewness-kurtosis) shape plane, and then the Logit-Logistic provides a good replacement of the S_B in forest diameter distribution modelling.

Appendix Code 4.1: Maximum Likelihood Estimation of Logit-Logistic

```
strfolder<-"j:/distrpapers/estimation/"##folder to store data
datafilename<-"test.dbf"##Chinese Fir data
datafile<-paste(strfolder,sep="",datafilename)
import.data('test2',datafile,"DBASE")

##Negative Log-Likelihood Function
LogitLogistic4.neg.ll<-+
function(theta,x)
{ n<-length(x)
  zetta<-theta[1]##minimum parameter, 0<=zetta<Dmin
  lambda<-theta[2]##maximum parameter, lambda>Dmax
  gamma<-theta[3]
  delta<-theta[4]
  u1<-x-zetta
  u2<-lambda-x
  u<-u1/u2
  n*(log(delta)-log(lambda-zetta)-
     gamma/delta)+1/delta*sum(log(u))+sum(log(u1))+sum(log(u2))+2*sum(log(1+exp(gamma/delta)*u^(-
     1/delta)))
 }

 for (i in 1:107)
 {    d0<-test2$D[test2$PLOT= =i]
      d0<-sort(d0)
      n.sample<-length(d0)
      outfilename<-"Logit4MLE.txt"##message on convergence
      outfilename2<-"Logit4MLE_p.txt"##message on parameter estimates
      outfile<-paste(strfolder,sep="",outfilename)
      outfile2<-paste(strfolder,sep="",outfilename2)
      cat(i,",",file=outfile,append=T)
      cat(i,",",file=outfile2,append=T)
      min0<-min(d0)
      max0<-max(d0)
      mylist<-nlminb(start=c(min0-2,max0+2,1,1),objective=LogitLogistic4.neg.ll,control =
      nlminb.control(eval.max=10000,iter.max =10000),lower=c(0,max0,-Inf,0),upper=c(min0,Inf,Inf,Inf),x=d0)
      cat(mylist[2],mylist[3],"\n",file=outfile,append=T
      cat(unlist(mylist[1]),"\n",file=outfile2,append=T)
 }
```

Appendix Code 4.2: Nonlinear Least Squares (based on the cumulative probability function/CDF) Estimation of Logit-Logistic

```
strfolder<-"j:/distrpapers/estimation/"##folder to store data
datafilename<-"test.dbf"##Chinese fir data
datafile<-paste(strfolder,sep="",datafilename)
import.data('test2',datafile,"DBASE")

## Residual Function (CDF based)
LogitLogistic<-+
function(theta,x,y)
{ zetta<-theta[1]##minimum parameter, 0<=zetta<Dmin
  lambda<-theta[2]##maximum parameter,lambda>Dmax
  gamma<-theta[3]
  delta<-theta[4]
  u1<-y-zetta
  u2<-lambda-y
  z<-(log(u1/u2)-gamma)/delta
  x0<-plogis(z,0,1)
  x-x0 }

 for (i in 1:107)
 { d0<-test2$D[test2$PLOT= =i]
   d0<-sort(d0)
   d00<-d0
   n.sample<-length(d0)
   for (ii in 1:n.sample) { d00[ii]<-ii/(n.sample+1) }
   outfilename<-"Logit4NLR.txt"##message on convergence
   outfilename2<-"Logit4NLR_p.txt"##message on parameter estimates
   outfile<-paste(strfolder,sep="",outfilename)
   outfile2<-paste(strfolder,sep="",outfilename2)
   cat(i,",",file=outfile,append=T)
   cat(i,",",file=outfile2,append=T)
   min0<-min(d0)
   max0<-max(d0)
   mylist<-nlregb(n=n.sample,start=c(min0-2,max0+2,0,0.5),residuals=LogitLogistic,control =
     nlregb.control(eval.max=10000,iter.max =10000,lower=c(0,max0,-Inf,0), upper=c(min0,Inf,Inf,Inf),
     x=d00,y=d0)
   cat(mylist[2],mylist[3],"\n",file=outfile,append=T)
   cat(unlist(mylist[1]),"\n",file=outfile2,append=T)
 }
```

Appendix Code 4.3: Nonlinear Least Squares (percentiles based) Estimation of Logit-Logistic

```
strfolder<-"j:/distrpapers/estimation/"##folder to store data
datafilename<-"test.dbf"##Chinese fir data
datafile<-paste(strfolder,sep="",datafilename)
import.data('test2',datafile,"DBASE")

## Residual Function (percentile based)
LogitLogistic<-+
function(theta,x,y)
{ zetta<-theta[1]##minimum parameter, 0<=zetta<Dmin
  lambda<-theta[2]##maximum parameter,lambda>Dmax
  gamma<-theta[3]
  delta<-theta[4]
  z<-qlogis(x)
  y0<-zetta+(lambda-zetta)/(1+exp(-(gamma+delta*z)))
  y-y0 }

 for (i in 1:107)
 {  d0<-test2$D[test2$PLOT= =i]
    d0<-sort(d0)
    d00<-d0
    n.sample<-length(d0)
    for (ii in 1:n.sample)  { d00[ii]<-ii/(n.sample+1) }
    outfilename<-"Logit4NLR2.txt"##message on convergence
    outfilename2<-"Logit4NLR2_p.txt"##message on parameter estimates
    outfile<-paste(strfolder,sep="",outfilename)
    outfile2<-paste(strfolder,sep="",outfilename2)
    cat(i,",",file=outfile,append=T)
    cat(i,",",file=outfile2,append=T)
    min0<-min(d0)
    max0<-max(d0)
    mylist<-nlregb(n=n.sample,start=c(min0-2,max0+2,1,1),residuals=LogitLogistic,control =
      nlregb.control(eval.max=10000,iter.max =10000),lower=c(0,max0,-Inf,0), upper=c(min0,Inf,Inf,Inf),
      x=d00,y=d0)
    cat(mylist[2],mylist[3],"\n",file=outfile,append=T)
    cat(unlist(mylist[1]),"\n",file=outfile2,append=T)
 }
```

Chapter 5: Bivariate Distribution Modelling with Plackett's Method

Summary

This chapter compares four bivariate distributional models in terms of their adequacy in representing empirical diameter-height distributions from 102 sample plots of the Chinese fir data sets. The four bivariate models are: S_{BB}, the natural, well-known, and much-used bivariate generalization of S_B; and the bivariate distributions with the Logit-Logistic (LL), S_B and Beta (GBD) as marginals, constructed using Plackett's method (LL-2^P etc...). All models are fitted using maximum likelihood, and their goodness-of-fits are compared using model deviance (equivalent to Akaike's Information Criterion, the AIC). The performance ranking was: S_{BB}, LL-2^P, GBD-2^P, and S_B-2^P.

5.1 Introduction

As we discussed in the "Introduction" chapter, stand volume estimation is an important aspect of forest mensuration, and is usually based on estimates of individual tree volumes from samples of tree diameters and heights. The traditional approach consists of fitting a marginal diameter distribution and then using an empirical height-diameter regression model to estimate the average height per diameter class and hence volume (Clutter and Allison 1974). That is, the mean sample tree volume is traditionally (T) estimated as

$$\hat{\bar{v}}_T = \int_{d>0} \hat{f}(d)\hat{V}(d,\hat{h}(d))dd \cong \frac{1}{n}\sum_{i=1}^{n}\hat{V}(d_i,\hat{h}(d_i)) \qquad (5.1)$$

where $\hat{f}(d)$ is the marginal diameter distribution, $\hat{h}(d)$ is the height-diameter regression model, both obtained from the n diameter sample trees, and $\hat{v}(d,h)$ is an individual volume equation, usually determined previously.

This approach, in using estimated heights in the volume equation, rather than actual heights, ignores the fact that height can vary considerably for a given diameter, and therefore introduces biases into volume estimation and associated precision estimates (Schreuder and Hafley 1977). These estimated-height bias effects may be avoided by use of the empirical bivariate distribution of diameters and heights for the height-diameter sample trees, or the fitted bivariate (B) density $\hat{f}(d,h)$ say, to obtain mean tree volume estimate as:

$$\hat{\bar{v}}_B = \iint_{d>0,h>0} \hat{f}(d,h)\,\hat{V}(d,h)\,dd.dh$$
$$\cong \frac{1}{n}\sum_{i=1}^{n}\left\{\int_{h>0}\hat{f}(h\,|\,d_i)\,V(d_i,h).dh\right\} = \frac{1}{n}\sum_{i=1}^{n}\left\{\int_{h>0}\frac{\hat{f}(d_i,h)}{\hat{f}(d_i)}V(d_i,h).dh\right\} \qquad (5.2)$$

Furthermore, the fitted bivariate distribution provides an alternative to usually adopted regression analysis for obtaining the H-D model, another approach to improving volume estimation.

These considerations highlight the importance of estimating the joint and conditional distributions of tree diameter and height, as well as their marginal distributions. Therefore, in this chapter we introduce two new bivariate models which may be used by forest biometricians, which are resulted from Plackett's method in following the work of Li et al. (2002). Although Plackett's method was found to fall into a more general topic, the copula, which is to be introduced in the next Chapter, we did the work related to Plackett's method here much earlier than we discovered the "copula" and then we report it in an independent Chapter as a preliminary work on the general topic of copula.

5.2 Literature Review

The Farlie-Gumbel-Morgenstern (FGM) system (Conway 1983) is an approach to constructing bivariate distributions, but a major drawback of bivariate FGM distributions is that they are limited to describing only weak dependence between X and Y (Schucany et al., 1978). Hafley and Schreuder (1976) derived a bivariate Weibull distribution of the FGM form. They found this bivariate Weibull was not biologically reasonable for describing the bivariate height-diameter distribution.

Schreuder and Hafley (1977) introduced Johnson's S_{BB} (Johnson 1949b) for describing diameter-height frequency data. This bivariate distribution model has been used increasingly in forestry (Hafley and Buford 1985, Knoebel and Burkhart 1991, Tewari and Gadow 1997, 1999, Li et al. 2002). The S_{BB} is constructed by applying a 4-parameter Logistic transformation to each of the component variables of a standard bivariate normal distribution (see section 5.5 below and Johnson (1949b), Johnson and Kotz (1972), Schreuder and Hafley (1977), and Rennolls and Wang (2005)).

Construction of a satisfactory bivariate distribution without resorting to transformation of a bivariate normal distribution is surprisingly difficult. No completely satisfactory method has yet been found of determining bivariate distribution models as extensions of univariate distributions (Kendall and Stuart 1977). The marginal distributions do not uniquely determine the corresponding bivariate distribution: it can be shown that for given marginal distributions, there exist infinitely many bivariate distributions with these marginal distributions (Fréchet cited in Gumbel 1960, 1961). For example, two standard normal marginal distributions may correspond to the infinite family of bivariate normal distributions indexed by the correlation parameter ρ.

Some bivariate distributions that may be constructed have the unsatisfactory property that the correlation parameter cannot take the whole of the range form -1 to 1, for example, the FGM system. In contrast, Plackett (1965) introduced a method of constructing a bivariate distribution from given marginal distributions, in which the resulting distribution has a single parameter of association, and the whole range of correlation is available. Li et al. (2002) used Plackett's method to obtain the bivariate Generalized Beta Distribution (GBD-2) and used it to fit empirical distributions of tree diameter and height. They found, for their data, better performance using GBD-2 than Johnson's S_{BB}. Therefore, Plackett's method seems providing a promising way of constructing bivariate distribution models.

Various univariate distribution models have been used for describing (marginal) diameter distributions in forest stands, including Gamma (Nelson 1964), Lognormal (Bliss and Reinker 1964), Weibull (Bailey and Dell 1973, Rennolls et al. 1985), Beta (Zohrer 1972, Li et al. 2002), and Johnson's S_B (Hafley and Schreuder 1977, Zhou and McTague 1996, Kamziah et al. 1999). In chapter 3, we introduced the Logit-Logistic (LL) distribution for forest diameter distribution modelling. Among these distribution models, the LL, S_B, and Beta are the most flexible distributions in terms of the (skewness-kurtosis) region covered, ranked in the order: LL, S_B and Beta/GBD, while in Chapter 3 it was found in the empirical tree diameter distribution study that the LL and the Beta/GBD were both more flexible than S_B .

In this Chapter we construct the bivariate distributions with double-LL and double-S_B marginal distributions by Plackett's method: we denote them by LL-2^P and S_B-2^P following the notation of Li et al. (2002) (i.e. GBD-2), and indicating that Plackett's method has been used. We compare these two new bivariate distributions with the S_{BB} and GBD-2^P distributions, and examine if the superior univariate flexibility of LL extends into the bivariate domain. The newly constructed bivariate models provide more "model choices" in forest bivariate distribution modelling, beyond the well-known Johnson's S_{BB} , and the GBD-2 of Li et al. (2002).

The bivariate distribution models considered are compared in an empirical case study using the maximum-likelihood estimation and a deviance goodness-of-fit criterion, on the Chinese fir diameter and height data sets.

5.3 Plackett's Method

Plackett proposed a method of constructing a one-parameter class of bivariate distributions from given marginal distributions (Plackett 1965, Mardia 1967, 1970, Ord 1972). This method is based on the use of the odds-ratio as the measure of association for a 2×2 contingency table and the resulting distributions are called contingency-type bivariate distributions by Mardia (1967, 1970). Figure 5.1 illustrates the concepts underlying the method.

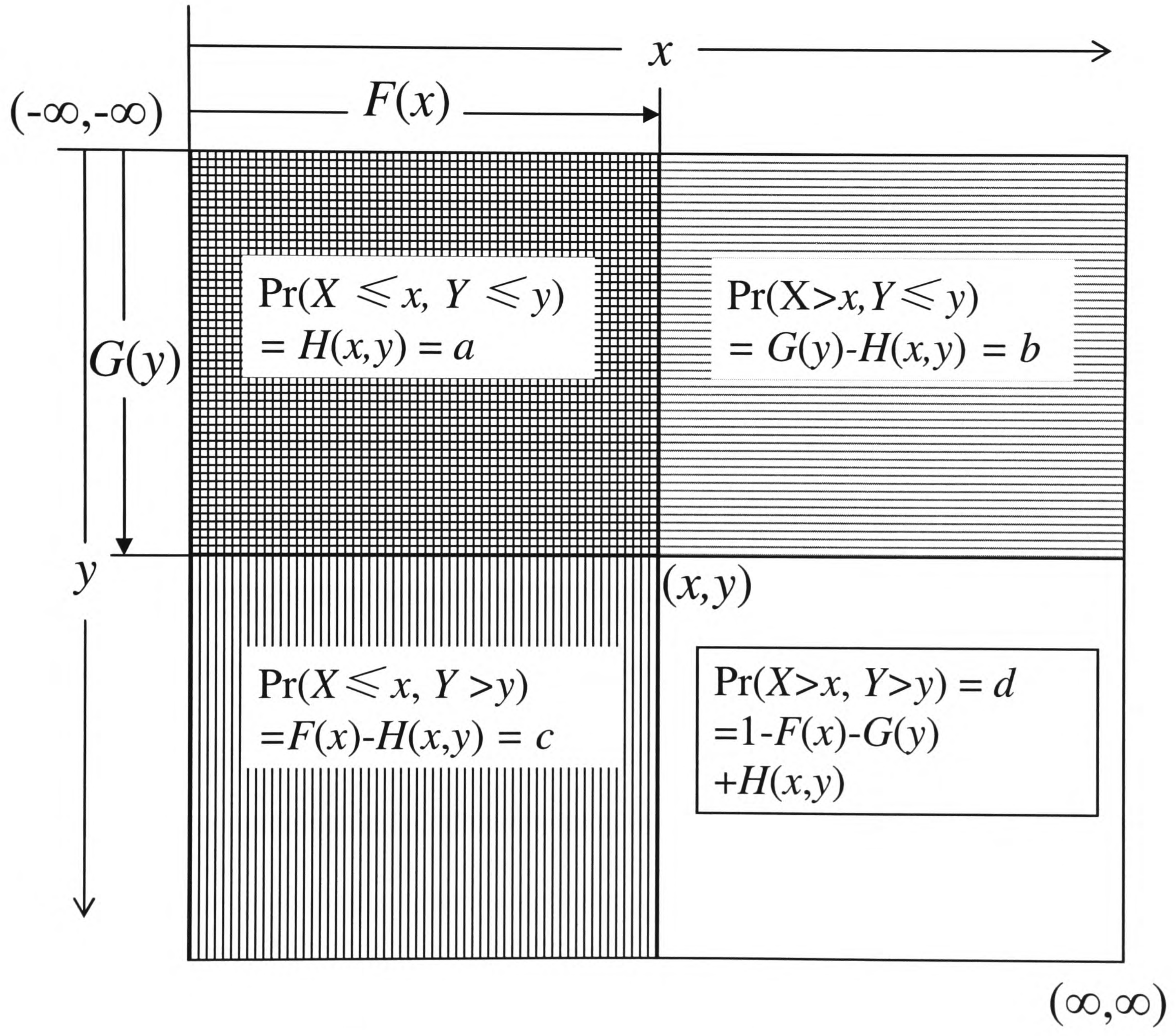


Figure 5.1. Plackett's method: 2 × 2 partition of the joint density H(x,y)

In Figure 5.1, $F(x)$ and $G(y)$ are the marginal cumulative distribution functions (CDFs) of random variables X and Y, respectively, and $H(x, y)$ is the bivariate CDF. The odds-ratio (ψ

= (ad/bc) ≥ 0) is a commonly used measure of the association between the row and column variables of a 2 × 2 contingency table. Evaluating this odds-ratio on the probabilities in the four quadrants around the point (x, y) in Figure 5.1 gives:

$$\psi = \frac{ad}{bc} = \frac{H[1-F-G+H]}{[F-H][G-H)]} \tag{5.3}$$

where the arguments of the CDFs have been suppressed. This identity may be expressed as a quadratic equation in H:

$$(\psi - 1)\,H^2 - S\,H + \psi\,F\,G = 0 \tag{5.4}$$

where $S = [1 + (\psi\text{-}1)(F + G)]$. $H(x,y)$ may be determined as a root of this equation,

$$H(x,y) = \frac{S(x,y) - \sqrt{S^2(x,y) - 4\psi(\psi-1)F(x)G(y)}}{2(\psi-1)}, \tag{5.5}$$

The other root is does not give a suitable CDF (Mardia 1967).

The bivariate probability density function (PDF), $h(x,y)$, is thus obtained by differentiating equation (5.3) with respect to x and y. Some manipulations give:

$$h(x,y) = \frac{\psi f(x)g(y)\{1+(\psi-1)[F(x)+G(y)-2F(x)G(y)]\}}{[S^2(x,y)-4\psi(\psi-1)F(x)G(y)]^{3/2}} \tag{5.6}$$

where $f(x)$ and $g(y)$ are the marginal densities.

The association parameter ψ characterizes the full range of dependence between random variables X and Y. Let $\rho(\psi)$ denote the correlation between X and Y for association ψ. When $\psi = 1$, $H(x, y) = F(x)G(y)$, and X and Y independent with $\rho(1) = 0$. With $\psi > 1$ and $\psi < 1$, X and Y are positively and negatively correlated, respectively, with $\rho(0) = -1$ and $\rho(\infty) = +1$. The conditional CDF of Y given $X = x$ is given by:

$$K_{Y|X}(y \mid x) = \frac{\psi G(y) - (\psi-1)H(x,y)}{1+(\psi-1)[F(x)+G(y)-2H(x,y)]} \tag{5.7}$$

(Mardia 1967), which may be used in (5.2) or used to derive a median regression equation.

5.4 Marginal Distribution Models

Three univariate distributions are considered as candidate marginal distributions from which to construct bivariate distributions using Plackett's method: the Beta (GBD), Johnson' S_B, and the Logit-Logistic (LL).

5.4.1 Beta (GBD)

The Beta PDF is defined by:

$$f(x) = \frac{1}{B(p,q)} \frac{(x-a)^{p-1}(b-x)^{q-1}}{(b-a)^{p+q-1}} \tag{5.8}$$

where $a \le x \le b$, a is the minimum parameter, b is the maximum parameter, and p, $q > 0$ are shape parameters.

5.4.2 Johnson's S_B

Johnson's S_B distribution (Johnson 1949a, Johnson and Kotz 1970) is that distribution which, when scaled to the range (0, 1), and then transformed by a logit transformation yields a Normal distribution. Chapter 2 (see also Rennolls and Wang 2005) present an inverse transformational definition of the S_B and a new parameterization. That is, $X \approx S_B$ distribution is obtained by applying a 4-parameter logistic transformation to a standard Normal variate, Z:

$$x = \xi + \frac{\lambda}{1 + \exp(-(\mu + \sigma\, z))} \tag{5.9}$$

where $-\infty < z < \infty$, $\xi \le x \le \xi + \lambda$. The parameters μ and σ are related to the parameters used by Johnson by $\mu = -\gamma_J / \delta_J$ and $\sigma = 1/\delta_J$. The S_B density is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \frac{\lambda}{(x-\xi)(\xi+\lambda-x)} e^{-\frac{1}{2}\left(\frac{\ln\left(\frac{x-\xi}{\xi+\lambda-x}\right)-\mu}{\sigma}\right)^2} \tag{5.10}$$

5.4.3 Logit-Logistic

Replacing the standard Normal in (5.9) with the standard Logistic results in the Logit-Logistic (LL) distribution (see Chapter 3). Its CDF and PDF are given by:

$$F(x)=\frac{1}{1+e^{\frac{\mu}{\sigma}}(\frac{x-\xi}{\xi+\lambda-x})^{-\frac{1}{\sigma}}} \tag{5.11}$$

$$f(x)=\frac{\lambda}{\sigma}\frac{1}{(x-\xi)(\xi+\lambda-x)}\cdot\frac{1}{e^{-\frac{\mu}{\sigma}}(\frac{x-\xi}{\xi+\lambda-x})^{\frac{1}{\sigma}}+e^{\frac{\mu}{\sigma}}(\frac{x-\xi}{\xi+\lambda-x})^{-\frac{1}{\sigma}}+2} \tag{5.12}$$

5.5 Bivariate Distributions

5.5.1 Johnson's S_{BB}

By analogy to the alternative transformational definition of S_B given in Chapter 2 (Rennolls and Wang 2005), the S_{BB} distribution (Johnson 1949b, Johnson and Kotz 1972) is the bivariate distribution of random variables x and y obtained by applying separate 4-parameter linear-logistic transformations, as (5.9), to each of the component variables, z_x and z_y, of a standard bivariate normal N(**0**, **Σ**), with correlation ρ . The z_x and z_y have the joint distribution:

$$f(z_x,z_y,\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{z_x^2-2\rho z_x z_y+z_y^2}{2(1-\rho^2)}} \tag{5.13}$$

By applying the linear-logistic transforms as,

$$x=\xi_x+\frac{\lambda_x}{1+e^{-(\mu_x+\sigma_x z_x)}} \qquad y=\xi_y+\frac{\lambda_y}{1+e^{-(\mu_y+\sigma_y z_y)}} \tag{5.14}$$

where, $\lambda_x > 0$, $\sigma_x > 0$, $\xi_x \le x \le \xi_x + \lambda_x$., and $\lambda_y > 0$, $\sigma_y > 0$, $\xi_y \le y \le \xi_y + \lambda_y$.

The S_{BB} joint density for x and y is thus given by

$$\begin{aligned} f(x,y,\rho) &= f(z_x,z_y,\rho)\left(\frac{dx}{dz_x}\right)^{-1}\left(\frac{dy}{dz_y}\right)^{-1} \\ &= \frac{J}{2\pi\sqrt{1-\rho^2}}e^{-\frac{z_x^2-2\rho\, z_x z_y+z_y^2}{2(1-\rho^2)}} \end{aligned} \tag{5.15}$$

where

$$J=\left(\frac{dx}{dz_x}\right)^{-1}\left(\frac{dy}{dz_y}\right)^{-1}=\frac{\lambda_x}{\sigma_x(x-\xi_x)(\xi_x+\lambda_x-x)}*\frac{\lambda_y}{\sigma_y(y-\xi_y)(\xi_y+\lambda_y-y)} \tag{5.16}$$

$$z_x = \frac{\ln(\frac{x-\xi_x}{\xi_x+\lambda_x-x})-\mu_x}{\sigma_x}, \quad z_y = \frac{\ln(\frac{y-\xi_y}{\xi_y+\lambda_y-y})-\mu_y}{\sigma_y} \tag{5.17}$$

5.5.2 Plackett's Bivariate Beta, S_B and Logit-Logistic

Although the two marginal distributions do not necessarily have to have the same distributional form, in practice, we always select them to be so. Use of (5.5), by replacing the marginals ($F(x)$ and $G(y)$) with the Logit-Logistic (LL), S_B and Beta (GBD) respectively, leads to the LL-2^P, S_B-2^P and GBD-2^P joint distributions. It is noted that the Plackett's method is based on CDFs. For the three univariate distributions considered here, only the Logit-Logistic has a closed form for its CDF (in (5.11)), thus for the S_B and Beta, numerical methods for evaluating their CDFs have to be used in the model fitting process.

5.6 Model Fitting and Goodness-of-Fit Criterion

Previous approaches in the forestry literature to fitting a bivariate distribution to tree diameter and height data have been as follows. First, fit the two marginal distributions separately using either the maximum likelihood method, a moment method, or other methods. It is also common practice to predetermine the location parameter (ξ of the S_B or a of the Beta) and possibly the range parameter (λ for S_B or the maximum parameter b for the Beta) to specific sample related values (Schreuder and Hafley 1977, Li et al. 2002). Second, estimate the dependence parameter, ρ for S_{BB} (Johnson 1949b, Schreuder and Hafley 1977) or ψ for the bivariate Plackett distributions, usually by heuristic methods (Plackett 1965, Li et al. 2002).

In contrast, we used the maximum likelihood estimation to fit the bivariate distributions directly; all the parameters are estimated simultaneously. We imposed the necessary constraints on ξ and λ for S_B or the LL ($0 \le \xi < x_{min}$ and $\xi + \lambda > x_{max}$), and on a and b for the Beta ($0 \le a < x_{min}$ and $b > x_{max}$) where x_{min} and x_{max} are the minimum and maximum values of the sampled x-values (x being diameter or height). The computation was carried out using the function *nlminb* (local minimizer for smooth nonlinear functions subject to bound-constrained parameters) of S-Plus (Mathsoft 1999) by minimizing the negative log-likelihood (-ΛΛ) function of density (5.15) for the S_{BB}, or (5.6) for the three Plackett-based bivariate models.

Assuming the sample is *iid*, then for the S_{BB} the (-ΛΛ) is the sum over all trees of:

$$-\ln f = -\ln J + \ln 2\pi + 0.5\ln(1-\rho^2) + \frac{z_x^2 - 2\rho z_x z_y + z_y^2}{2(1-\rho^2)} \tag{5.18}$$

where J, z_x and z_y are as defined in (5.16) and (5.17). For the Plackett-based bivariate models, (-ΛΛ) is the sum of single-tree terms of the form:

$$-\ln h = -\ln\psi - \ln f - \ln g - \ln\{1+(\psi-1)[F+G-2FG]\} + \left(\tfrac{3}{2}\right)\ln[S^2 - 4\psi(\psi-1)FG] \tag{5.19}$$

The marginal CDFs (F and G) of the Beta and Normal are evaluated using S-plus's *pbeta* and *pnorm* functions, respectively.

The goodness-of-fit criterion which we use for comparison purposes is the negative log-likelihood (-ΛΛ), the deviance statistic, which is equivalent to the AIC (Akaike 1974) in this case study, as each bivariate model considered here has the same number of parameters, i.e. 9. The smaller is (-ΛΛ) of a fitted model, the better the model fits.

5.7 Data and Results

Among the 107 plots of the Chinese fir data, there are 102 plots, in which each tree in a plot was measured for diameter and height. Thus these 102 plots were used in this study.

Table 5.1 lists the value of the (-ΛΛ) goodness-of-fit statistics for each of the models fitted to each of the sample-plot diameter and height distributions. For the 102 sample plots in the case study, the maximum likelihood estimation method did not converge for only 2 plots for the LL-2^P model (plot 73 and 76) and for only four plots for the GBD-2^P model (plots 40, 47, 73 and 76). The reason for non-convergence is not completely clear to us, but is probably related to these plots having J-shaped marginal distributions. However, we found from Table 5.1 that three of the four empirical distributions for which there was non-convergence for either LL-2^P or GBD-2^P or both, the differences in the resulting $(-\Lambda\Lambda)$ is quite small between converged models and non-converged models. However, this was not the case for plot 73, and therefore, we have eliminated plot 73 from subsequent comparisons.

Table 5.2 indicates the between-model comparative performance of these four models in terms of their goodness of fit statistics. The S_{BB} has out-performed all three bivariate Plackett-based distributional models in our empirical comparisons. The S_{BB} distribution had a lower value of the $(-\Lambda\Lambda)$ statistic than the LL-2^P, GBD-2^P, and S_B-2^P for 74, 85 and 86 of the 101 distributions, respectively.

The observed relative performance of S_{BB} and GBD-2^P is different from that observed by Li et al. (2002). There are a number of possible reasons for this: (i) We have used different case-study data than they used, (ii) We have used a different estimation method, (iii) We have used the log-likelihood/deviance/AIC/$(-\Lambda\Lambda)$ goodness-of-fit criterion obtained directly from the maximum likelihood estimation process, while Li et al. (2002) used a χ^2 criterion relating to volume predictions.

From Table 5.2, for the three bivariate Plackett-based models, LL-2^P was better than GBD-2^P (88 out of 101) and S_B-2^P (95 out of 101), and GBD-2^P better than S_B-2^P (71 out of 101). We may summarize the results from Table 5.2 as follows:

$S_{BB} >$ (LL-2^P) $>$ (GBD-2^P) $>$ (S_B-2^P)

where ">" represents "better than" in the sense used in Table 5.2.

The better performance of LL-2^P over both S_B-2^P and GBD-2^P is not unexpected, since the Logit-Logistic (LL) is the most flexible (in terms of skewness-kurtosis coverage) among these three distributional models in the univariate domain, and the bivariate models have been constructed using the same method.

Table 5.1. Negative log-likelihood (-ΛΛ) for four bivariate distribution models

Plot	S_{BB}	LL-2^P	S_B-2^P	GBD-2^P	Plot	S_{BB}	LL-2^P	S_B-2^P	GBD-2^P
1	335.22	334.20	336.67	335.51	52	337.64	344.75	345.80	345.82
2	400.37	404.64	406.16	405.44	53	444.25	442.29	443.20	442.80
3	415.55	422.05	421.36	421.29	54	669.32	677.64	678.76	677.15
4	492.02	496.93	496.64	494.75	55	364.91	366.59	367.41	367.66
5	422.01	430.09	431.09	430.92	56	366.02	365.52	366.98	366.96
6	390.74	396.80	398.52	398.30	57	329.75	330.76	331.91	331.72
7	570.96	581.87	582.48	582.38	58	462.10	463.98	465.83	465.56
8	468.81	472.21	473.32	472.64	59	346.20	351.57	352.52	352.74
9	470.60	469.05	469.79	469.54	60	340.19	349.02	349.86	349.79
10	328.49	328.88	330.74	330.48	61	361.60	364.61	366.08	366.16
11	305.73	305.35	307.09	306.02	62	260.95	266.07	266.42	266.46
12	246.64	245.42	246.28	246.16	63	252.76	254.11	255.11	254.83
13	278.51	278.20	278.75	278.75	64	521.55	519.54	521.85	520.71
14	218.86	219.43	220.14	220.22	65	362.62	363.86	365.32	365.12
15	342.63	349.76	351.94	352.18	66	367.41	367.31	367.75	367.47
16	316.56	316.70	317.74	317.62	67	428.86	438.42	439.62	439.36
17	353.72	359.13	360.83	360.67	68	561.45	561.26	562.21	562.47
18	551.41	555.90	558.68	558.07	69	355.03	362.10	362.88	362.82
19	469.15	469.55	470.56	470.67	70	445.40	448.22	448.88	448.89
20	548.22	553.42	556.02	555.80	71	548.71	551.14	550.96	551.29
21	252.82	246.42	247.04	246.81	72	406.89	409.17	410.71	410.73
22	286.91	291.43	292.63	292.78	73	241.25	244.07*	237.19	297.36*
23	286.80	292.00	292.50	292.54	74	220.70	222.57	223.12	223.28
24	385.85	384.59	384.90	384.82	75	282.28	284.64	286.76	285.12
25	317.54	319.05	320.70	320.26	76	260.80	262.90*	266.00	266.91*
26	348.33	346.15	347.65	346.91	77	354.59	353.49	353.66	353.38
27	299.98	307.96	307.97	307.96	78	516.47	516.22	516.82	516.76
28	420.42	418.27	419.52	418.38	79	697.00	696.88	699.82	698.22
29	336.03	344.27	344.73	344.59	80	418.69	426.61	426.87	426.60
30	339.28	339.18	339.04	339.06	81	385.57	390.74	392.47	391.30
31	404.80	403.85	404.50	404.15	82	358.46	357.45	359.98	359.73
32	335.39	340.33	341.03	339.38	83	246.23	248.26	248.75	248.48
33	528.21	534.71	535.79	535.79	84	314.11	318.00	319.64	318.78
34	374.01	378.27	379.23	379.12	85	647.25	651.28	651.32	651.71
35	474.49	476.60	477.92	477.71	86	303.06	305.33	305.45	305.42
36	644.06	649.60	649.77	648.55	87	465.30	470.34	471.37	471.37
37	278.66	286.77	288.10	287.47	88	355.34	355.80	357.67	357.30
38	246.99	240.38	240.70	240.76	89	321.45	327.32	328.68	328.82
39	539.39	550.58	551.69	551.97	90	629.12	638.22	637.68	637.90
40	648.84	649.73	650.09	650.47*	91	429.81	430.73	432.37	431.97
41	482.02	476.86	478.92	478.79	92	262.17	266.50	267.22	266.70
42	512.31	513.62	514.80	514.74	93	444.55	451.93	452.05	452.09
43	312.79	316.29	318.33	317.28	96	390.72	397.96	398.84	398.49
44	621.38	622.59	624.27	623.77	97	463.77	473.93	474.81	473.46
45	801.59	801.41	802.62	802.68	98	515.41	519.23	520.37	520.43
46	536.36	540.88	541.16	541.21	99	451.72	457.33	457.80	457.65
47	526.30	529.08	531.26	529.72*	100	490.36	478.37	480.23	479.85
48	391.40	398.42	400.32	400.11	101	418.53	423.34	423.07	422.89
49	758.15	763.95	765.37	763.70	102	361.98	369.30	370.73	370.36
50	268.72	268.10	268.80	268.73	104	253.75	255.92	256.49	256.56
51	329.62	328.62	329.00	328.76	106	292.60	289.81	290.93	290.50

Note: * indicates that a non-convergence notification was obtained in fitting.

Table 5.2. Comparison results of four bivariate distribution models based on ($-\Lambda\Lambda$): Proportion of cases in which the row-distribution model had a lower ($-\Lambda\Lambda$) than the column-distribution.

row \ col	LL-2^P	GBD-2^P	S_B-2^P
S_{BB}	74/101	85/101	86/101
LL-2^P		88/101	95/101
GBD-2^P			71/101

It is noted that the ranking of the bivariate Plackett-based distributions is consistent with the rankings obtained for the marginal distributions, for both diameter and height. The results for height distributions are listed in Table 5.3, while the details for diameter distributions are given in Chapter 3 (also Wang and Rennolls 2005). The empirical superiority of GBD-2^P over S_B-2^P (and Beta over S_B) in this case-study is slightly unexpected, since (theoretically) S_B is more flexible than the Beta, by virtue of larger coverage of the skewness-kurtosis shape-plane.

Table 5.3. Comparison results of Logit-Logistic (LL), S_B and Beta (GBD) in fitting marginal heights based on ($-\Lambda\Lambda$): Proportion of cases in which the row-distribution model had a lower ($-\Lambda\Lambda$) than the column-distribution.

row \ col	Beta	S_B
LL	89/102	89/102
Beta		61/102

5.8 Discussion

Besides the transformation-translation method of construction of Johnson's S_{BB}, and the Plackett's method by which the other three bivariate distribution models are constructed in this paper, there are other ways to construct bivariate distributions. With the Logit-Logistic as the marginals, we also constructed and fitted two bivariate models using two methods of

Gumbel (1961) (LL-2^{GI} and LL-2^{GII} say). The results were poor and are not presented here. The poor performance is due to the fact that both Gumbel methods can only accommodate weak dependency, which is not the case of bivariate diameter and height modelling. Table 5.4 listed descriptive statistics of correlations between tree diameters and heights for the 102 bivariate samples. The Pearson's correlation coeeficient (r^2) is the traditional linear correlation measure. Since correlation between tree diameters and heights is intrinsically non-linear, two other rank based correlation measures, Kendall's τ and Spearman's ρ, are also provided. Both measures will be discussed in detail in the next chapter.

For a good bivariate distribution, not only must the marginal distributions have flexible shape, but the method of construction of the bivariate distribution has to be able to model both high and low levels of dependency/correlation. The Logit-Logistic (LL) distribution seems to be the most suitable of the available choices for univariate distribution modelling while the Plackett's method is able to model dependencies with correlations between -1 and 1. In the case of bivariate diameter-height modelling, tree heights are usually positively highly correlated with tree diameters (see Table 5.4), what we are concerned is, therefore, that the Plackett's method is capable of modelling positive correlation ranging from zero to one. Hence, we recommend the LL-2^P bivariate model for future use in diameter-height modelling studies in forestry.

Table 5.4. Descriptive Statistics of Three Correlation Measures of H-D Samples

Correlation	N	Minimum	Maximum	Mean	Std. Dev.
Pearson's r^2	102	0.65	0.96	0.8478	0.05846
Kendall's τ	102	0.43	0.84	0.6544	0.07128
Spearman's ρ	102	0.59	0.97	0.8353	0.06553

Chapter 6: The Copula Approach to Bivariate Distribution Modelling

Summary

By a theorem due to Sklar in 1959, a bivariate distribution can be represented in terms of its underlying marginals by binding them together using a copula function. This copula representation of the bivariate distribution allows different specifications for the marginals but models the dependence with the copula function itself, thus providing a general way of constructing bivariate distributions. The basics of the copula and some popular copula functions are introduced in this Chapter. The well-known Johnson's S_{BB} is obtained from the normal copula using the S_B as marginal model. Using the normal copula, we construct bivariate Logit-Logistic and bivariate Beta distributions. An empirical case-study on tree heights and diameters of 102 sample plots of Chinese Fir demonstrates that the bivariate Logit-Logistic can have superior performance over S_{BB}, extending similar univariate results for diameter distributions. An empirical comparison of the normal copula with the others shows that the normal copula is the best for modelling the joint distribution of tree diameters and heights.

6.1 Introduction

As an alternative but also an improvement to the traditional approach to stand volume estimation, fitting a joint bivariate distribution model, *f(d,h) say,* to the height-diameter sample is generally considered more satisfactory for estimating stand volume (Schreuder and Hafley 1977, Li et al. 2002). The bivariate distribution modelling approach considers not only the fact that height can vary considerably for a given diameter, but also provides an alternative way of obtaining the H-D regression rather than the traditional regression methods. Also, the bivariate *f(d,h)* provides a means of describing diameter-height dynamics over time in the context of spatial modelling of a forest stand (Li et al. 2002).

Schreuder and Hafley (1977) introduced Johnson's S_{BB} (Johnson 1949b) into the forestry literature for modelling the joint distribution of tree diameter and height, and since then this bivariate distributional model has been widely used in forestry (Hafley and Buford 1985, Knoebel and Burkhart 1991, Tewari and Gadow 1997, 1999, Li et al. 2002). In fact, the S_{BB} seems the only bivariate model successfully used in describing the joint distribution of tree diameter and height for a long time.

Li et al. (2002) used Plackett's method to obtain a bivariate Beta distribution. In Chapter 5, we constructed a bivariate Logit-Logistic using the same method, as well as a bivariate S_B (not equivalent to S_{BB}) for comparative purposes. Plackett's method allows for arbitrarily marginals to be employed in constructing a bivariate distributon model. However, given marginal distributions do not uniquely determine a bivariate ditribution: the bivariate distribution determined by Plackett's method is only one of a family of possible bivariate distributions with the given marginals.

Mardia (1970) in extending Johnson's method of "transformation to bivariate normal" allows for the construction of a bivariate distribution with arbitrary given marginals. Mardia's approach can produce the S_{BB} as the resulting bivariate model when both the marginals are S_B.

Both Plackett's and Mardia's methods for the construction of a bivariate distribution model from given marginals are special cases of a more general copula-based approach. The copula approach to multivariate modelling follows from a theorem in Sklar (1959) (see also Sklar 1973, Nelson 1999) and provides a constructive method for building a bivariate model from given marginals, using a copula function. By considering a family of copula functions, a family of bivariate distributions can be obtained. Recently, practical use of copulas in constructing multivariate models has been increasing, especially in the financial and actuarial fields (Frees and Valdez 1998, Klugman and Parsa 1999, Embrechts et al. 1999, 2001, Bouyé et al. 2000, Li 2000, Smith 2003, Wüthrich 2004, Hürlimann 2004), also in decision analysis or management science (Jouini and Clemen 1996, Clemen and Reilly 1999). However, the concept of "copula" is new to forestry, though the normal copula and the Plackett copula, with which Johnson's S_{BB} and Plackett's families can be obtained respectively, have been implicitly used. So, it would be much of interest to forest researchers to introduce the copula approach into forestry.

In this Chapter, firstly we briefly describe the basic properties of copulas useful to our study. Then we introduce some copula models, not surprisingly with emphasis on the normal copula. Finally we conduct an empirical study of comparing these copula models for their potential usages in forestry.

6.2 Bivariate Copula

6.2.1 Basics of Copula Functions

The idea behind the concept of the copula is to separate a multivariate distribution function into two parts, one describing the dependence structure and the other describing marginal behaviors. This concept was introduced by Sklar in 1959 to answer a question about the relationship between a multivariate distribution function and its univariate marginals in the framework of "Probabilistic Metric Spaces". As noted by Micocci and Masala (2003), since 1986 copula functions are intensively investigated from a statistical point of view due to the impulse of Genest and MacKay's work "The joy of copulas" (1986). A historical review and major developments can be found in Joe (1997) and Nelsen (1999).

Broadly speaking, a copula is a function that joins or "couples" a multivariate distribution function to their one-dimensional marginal distributions (Nelson 1999), or alternatively a copula is a joint multivariate distribution function with uniform marginals. For a bivariate copula, we have,

$$C(u,v)=\Pr(U\leq u,V\leq v) \quad (6.1)$$

where u and v are uniform random variables. Suppose we have two random variables x and y which follow arbitrary marginal distribution functions $F(x)$ and $G(y)$, respectively. The copula function C combines these two marginals to give the joint distribution function as

$$C(F(x),G(y))=H(x,y) \quad (6.2)$$

simply due to the fact that the probability integral transformed variables $u=F(x)$ and $v=G(y)$ are uniform random variables. This copula function defines a new bivariate distribution, which can be easily shown as follows:

$$\begin{aligned}C(F(x),G(y)) &= \Pr[U\leq F(x),V\leq G(y)]\\ &= \Pr[F^{-1}(U)\leq x,G^{-1}(V)\leq y]\\ &= \Pr[X\leq x,Y\leq y]\\ &= H(x,y)\end{aligned}$$

Conversely, any bivariate distribution function H can be written in terms of its marginals using a copula representation as,

$$H(x,y)=C(F(x),G(y)) \quad (6.3)$$

If the marginal distributions F and G are both continuous, then H has a unique copula C. The above equations (6.2) and (6.3) comprise of the theoretical basis of multivariate modeling by copulas, which is known as Sklar's theorem. Clearly the copula function C is a re-formulation of the joint cumulative distribution function (CDF) in terms of its marginal CDF, for which a copula is sometimes called a "uniform representation" (Kimeldorf and Sampson 1975). Furthermore, given that marginal distribution functions (F and G) and C are differentiable, the joint density $f(x, y)$ can be expressed as,

$$h(x,y)=f(x)\times g(y)\times c(u,v) \quad (6.4)$$

where $f(x)$, $g(y)$ are the density corresponding to F and G, respectively, and where

$$c(u,v)=\frac{\partial^2 C(u,v)}{\partial u\partial v} \quad (6.5)$$

is the called the copula density.

By contrast to the traditional modeling approach that decomposes the joint density as a product of marginal and conditional densities, equation (6.4) shows that the joint density can be expressed as a product of the marginal densities and the copula density. Then from the representation in (6.4), it is clear that the density $c(u,v)$ thus the copula $C(u,v)$ encodes information about the dependence between x and y, for which the copula is also known as the "dependence function" (Galambos 1978), while the densities $f(x)$ and $g(y)$ control the marginal behaviors.

Therefore, the separation of a bivariate distribution into its marginal part and its dependence part with the copula approach provides much facility in statistical modeling: firstly as a way of studying scale-free measures of dependence and secondly as a tool of constructing families of bivariate distributions (Fisher 1997). As equation (6.2) shows that various bivariate distributions can be constructed with copula functions and marginals, equation (6.3) indicates such copula functions can be obtained from some existing bivariate distributions by expressing them in terms of CDF, one example being the normal copula which we will describe in detail later. For our purposes, we are interested in the "copula" approach mainly because it provides a general way of setting up multivariate distributions given the specific copula and arbitrary marginals. However, the dependence via some measure(s) described by the copula can act as a criterion for prescreening copula functions, then we will briefly introduce the use of copulas in studying measures of dependence in the subsequent section.

6.2.2 Dependence Measures

A measure of dependence commonly used in statistics is (Pearson's) linear correlation (ρ). However, it is known that this measure cannot capture the nonlinear dependence relationships exhibited by random variables, is not invariant under strictly increasing transformations of these variables, and the possible values of correlation depend on the marginal distributions. To avoid such drawbacks, some alternative measures of dependence, known as rank correlations, have been introduced in statistics. The most famous rank correlations include Kendall's tau (τ) and Spearman's rho (ρ_s). Let X and Y be random variables with distribution functions F and G and joint distribution H. Spearman's ρ can be interpreted as a correlation coefficient between the CDFs of the two variables. That is,

$$\rho_s(X, Y) = \rho\,(F(x), G(y)) \tag{6.6}$$

where ρ is the usual linear correlation. Kendall's tau is defined as the probability of concordance minus the probability of discordance:

$$\tau(X, Y)=\mathrm{P}[(X_1-X_2)(Y_1-Y_2)>0]-\mathrm{P}[(X_1-X_2)(Y_1-Y_2)<0] \tag{6.7}$$

where (X_1,Y_1), (X_2,Y_2) are independent random vectors from the common joint H.

Schweizer and Wolff (1981) established that the copula accounts for all the dependence between these two variables, X and Y, and also showed that these two measures could be expressed solely in terms of the copula function, they are:

$$\rho_s = 12\int_0^1\int_0^1[C(u,v)-uv]dudv \tag{6.8}$$

$$\tau = 4\int_0^1\int_0^1 C(u,v)dC(u,v)-1 \tag{6.9}$$

In contrast, the linear correlation depends not only on the copula but also on the marginal distributions.

For continuous random variables the above measures are measures of concordance, and both can take values in [-1, 1]. Both measures take the value zero when we have independence between two variables, that is, when we have the independent copula, uv. Two other important copulas are the lower Fréchet bound, max $(0, u+v-1)$, and the upper Fréchet bound, $\min(u,v)$, corresponding to the perfect negative dependence ($\tau = \rho_s = -1$) and the perfect positive dependence ($\tau = \rho_s = 1$), respectively. As Fréchet (1951, see Nelson 1999) showed that any copula function satisfies that $\max(0, u+v-1) \le C(u,v) \le \min(u,v)$, but does not necessarily cover the whole interval, the copula which can range from the lower Fréchet bound to the upper Fréchet bound is said to be *comprehensive*. Though a copula is not necessarily comprehensive, the capability of its accommodating dependence in terms of the range of Kendall's tau or Spearman's rho it can achieve would be undoubtedly useful to decide whether this copula is suitable or not.

6.2.3 Copula Functions

If we have a collection of copulas, then using Sklar's theorem, we can construct bivariate distributions with arbitrary marginals. Although many copula families are available with one or more parameters, we now only present a very brief overview of some one-parameter families of copulas with much attention put on the normal copula. Extensive surveys of families of copulas can be found in Hutchinson and Lai (1990), Joe (1997) and Nelson (1999). Table 6.1 provides a brief description of seven families most commonly used in biostatistics, actuarial science, or even management science.

Table 6.1: Families of One-Parameter Copulas

Family	Function Form $C(u,v)$	Parameter Range	Kendall's tau Range	Spearman's rho Range
Normal	$N_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$	[-1,1]	[-1,1]	[-1,1]
Plackett	$\frac{1+(\theta-1)(u+v)}{2(\theta-1)} - \frac{\sqrt{[1+(\theta-1)(u+v)]^2 - 4uv\theta(1-\theta)}}{2(\theta-1)}$	$(0,\infty)$	[-1,1]	
Frank	$-\frac{1}{\theta}\ln[1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}]$	$(-\infty,\infty)$	[-1,1]	[-1,1]
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$(0,\infty)$	[0,1]	
Gumbel	$e^{-[(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{1}{\theta}}}$	$[1,\infty)$	[0,1]	
AMH	$\frac{uv}{1-\theta(1-u)(1-v)}$	[-1,1]	[-0.18, 1/3]	[-0.27,0.48]
FGM	$uv[1+\theta(1-u)(1-v)]$	[-1,1]	[-2/9, 2/9]	[-1/3, 1/3]

Note: Φ and N_ρ are the standard normal or bivariate normal respectively.

In each copula, the parameter θ (for the normal copula we purposely use ρ indicating the link with the normal distribution) measures the degree of association/dependence. For the Normal (Mardia 1970), Frank (Frank 1979), AMH (Ali et al. 1978) and FGM (Conway 1983), the larger θ is in absolute value, the stronger the association between variables: $\theta > 0$ implies positive dependence, $\theta < 0$ implies negative dependence, and $\theta = 0$ or $\theta \to 0$ implies independence. For Gumbel (Gumbel 1960, 1961), $\theta = 1$ leads to independence and $\theta \to \infty$ leads to perfect positive dependence. For the Clayton copula (Clayton 1978, Cook and Johnson 1981, 1983, Oaks 1982), $\theta \to 0$ and $\theta \to \infty$ lead to independence and perfect positive dependence, respectively. For Plackett (Plackett 1965), $\theta = 1$ leads to independence, and $\theta \to 0$ and $\theta \to \infty$ lead to perfect negative or positive dependence, respectively.

As shown by Schucany et al. (1978) that the FGM copula allows only for a limited degree of dependence (Kendall's τ is restricted to [-2/9, 2/9] and Spearman's ρ to [-1/3, 1/3]), which reduces its appeal for use in applications. Similar considerations hold also for the AMH, whose range for Kendall's τ is restricted to [-0.18, 1/3] and for Spearman's ρ to [-0.27, 0.48]. Schreuder and Hafely (1977) found that the bivariate distribution constructed by using the FGM copula with Weibull as marginals is not satisfactory in modelling joint distribution of tree diameters and heights. We also used this copula but with the Logit-Logistic as marginals as mentioned in Chapter 5 and concluded the poor performance of this copula in modelling joint distribution of tree diameters and heights is due to the lower correlation restricted to this copula. For the AMH copula, it would be expected that it is not suitable for modelling joint distributions of tree diameters and heights because of the limited degree of dependence it allows. Therefore, we may not consider this copula as well as the FGM in this study. However, in order to clearly show that the performance of bivariate distributions with the same given marginals will be dependent on copulas used to construct such bivariate distributions, we will compare these two copulas with the others.

In contrast, the Normal, Frank and Plackett copulas are comprehensive in that they can accommodate both positive and negative dependence ranging from the lower Fréchet bound (perfect negative dependence) to the upper Fréchet bound (perfect positive dependence). The Clayton copula and Gumbel copula are not comprehensive but they can accommodate all the possible positive dependence, which is the case of joint distributions of tree diameters and heights.

As shown later, the widely used Johnson's S_{BB} can be obtained from the Normal copula. We already used the Plackett copula to construct bivariate models in Chapter 5 and found the S_{BB} performed better than Plackett's families of bivariate distributions. Then we will pay much attention to the Normal copula, and also investigate the Gumbel, Frank, and Clayton copulas, since recently they have been widely used in other disciplines.

6.2.3.1 The Normal Copula (Mardia's Extension of Johnson's System)

Mardia (1970) extended the normal marginals of a bivariate normal distribution to any given marginals, by transforming specified marginals into normal distributions by means of the inverse standard normal distribution function so that a standard bivariate normal can be used to model the joint distribution. The CDF and PDF of the standard bivariate normal are given as,

$$C(z_x, z_y) = \int_{-\infty}^{z_x}\int_{-\infty}^{z_y} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{z_x^2 - 2\rho z_x z_y + z_y^2}{2(1-\rho^2)}} dxdy \tag{6.10}$$

$$f(z_x, z_y; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{z_x^2 - 2\rho z_x z_y + z_y^2}{2(1-\rho^2)}} \tag{6.11}$$

For any given marginals $u=F(x)$ and $v=G(y)$, the inverse function of the standard normal distribution function (Φ^{-1}), actually being the well-known normalization, will transform the uniformized variables u and v (thus x and y) into standard normal, that is,

$$z_x = \Phi^{-1}(u) = \Phi^{-1}(F(x)) \; , \; z_y = \Phi^{-1}(v) = \Phi^{-1}(G(y)) \tag{6.12}$$

which is due to the fact that the probability integral transformed random variable from any given (continuous) distribution will be uniformly distributed. Therefore a standard bivariate normal can be used to model the joint distribution. The bivariate density $h(x, y)$ is then given as the following,

$$h(x,y) = \frac{dz_x}{dx}\frac{dz_y}{dy}\frac{\partial^2 C(z_x, z_y)}{\partial z_x \partial z_y} = \frac{dz_x}{dx}\frac{dz_y}{dy} f(z_x, z_y; \rho)$$

$$= \frac{f(x)g(y)}{\sqrt{1-\rho^2}} e^{-\frac{\rho}{2(1-\rho^2)}[\rho(z_x^2+z_y^2)-2z_x z_y]} \tag{6.13}$$

In general, the density function does not exist in closed form since there is no closed form expression for the inverse standard normal (Φ^{-1}). However, for example, if the marginals are normal or from Johnson's system of univariate distributions (S_B, S_L, and S_U), the involved transformations (to normal) as defined by equation (6.12) exist in closed form thus does the joint density. For example, with normal marginals where the usual bivariate normal is obtained, equation (6.12) simplifies to the "standardization", that is,

$$z_x = \frac{x - \mu_x}{\sigma_x}, \; z_y = \frac{y - \mu_y}{\sigma_y} \tag{6.14}$$

With the S_B where Johnson's S_{BB} is the resulting bivariate distribution, we have,

$$z_x = \frac{\ln(\frac{x-\xi_x}{\xi_x+\lambda_x-x})-r_x}{\delta_x} \; , \; z_y = \frac{\ln(\frac{y-\xi_y}{\xi_y+\lambda_y-y})-r_y}{\delta_y} \tag{6.15}$$

So we see the inverse standard normal function method of "transformation to normal" collapses to Johnson's method of "transformation to normal" (given as (6.15) for the S_B) when marginals from Johnson's families of univariate distributions. And in this sense, the bivariate normal distributions are extended to allow for marginals from other than Johnson's univariate system.

Lee (1983) used Mardia's method to formulate models with given marginal distributions for econometric applications. This method is now known as "the normal copula" approach and has been increasingly used in financial and actuarial modelling (Li 2000, Carrière 2000, Ané and Kharoubi 2003, Mendes and Souza 2004). Also Clemen and Reilly (1999) used it for decision and risk analysis. It is called the normal copula because it encodes dependence in precisely the same way as the standard bivariate normal distribution does, but it does so for variables with arbitrary marginals. By rewriting equation (6.10), the standard normal, in terms of u and v rather than z_x and z_y, the bivariate normal copula is given as,

$$C(u,v) = N_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$$

$$= \int_{-\infty}^{z_x}\int_{-\infty}^{z_y} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{z_x^2 - 2\rho z_x z_y + z_y^2}{2(1-\rho^2)}} dxdy \tag{6.16}$$

where $z_x = \Phi^{-1}(u)$, $z_y = \Phi^{-1}(v)$. Obviously, this copula is obtained by the "inversion of marginals" method" (Nelson 1999) from the standard bivariate distribution, equation (6.10). The copula density is given as,

$$c(u,v) = \frac{dz_x}{du}\frac{dz_y}{dv}\frac{\partial^2 C(z_x, z_y)}{\partial z_x \partial z_y} = \frac{dz_x}{du}\frac{dz_y}{dv} f(z_x, z_y; \rho)$$

$$= \frac{1}{\sqrt{1-\rho^2}} e^{-\frac{\rho}{2(1-\rho^2)}[\rho(z_x^2+z_y^2)-2z_x z_y]} \tag{6.17}$$

and the bivariate density $h(x, y)$ is then given as the following,

$$h(x,y) = \frac{du}{dx}\frac{dv}{dy}\frac{\partial^2 C(u,v)}{\partial u \partial v}$$

$$= \frac{f(x)g(y)}{\sqrt{1-\rho^2}} e^{-\frac{\rho}{2(1-\rho^2)}[\rho(z_x^2+z_y^2)-2z_x z_y]} \tag{6.18}$$

The equivalence of Mardia's method and the copula approach in constructing the joint distribution can be easily shown by noting that,

$$H(x,y) = C(z_x, z_y) = C(u,v) \tag{6.19}$$

$$h(x,y) = \frac{dz_x}{dx}\frac{dz_y}{dy}\frac{\partial^2 C(z_x, z_y)}{\partial z_x \partial z_y}$$

$$= \left(\frac{dz_x}{du}\frac{du}{dx}\frac{dz_y}{dv}\frac{dv}{dy}\right)\frac{\partial^2 C(z_x, z_y)}{\partial z_x \partial z_y}$$

$$= \frac{du}{dx}\frac{dv}{dy}\left(\frac{dz_x}{du}\frac{dz_y}{dv}\frac{\partial^2 C(z_x, z_y)}{\partial z_x \partial z_y}\right)$$

$$= \frac{du}{dx}\frac{dv}{dy}\frac{\partial^2 C(u,v)}{\partial u \partial v} \tag{6.20}$$

We note that this equivalence essentially indicates an important property of the copula, that is, with strictly increasing transformations on the random variables, the copula of the transformed variables is the same as that before transformation, or in other words, the copula function is invariant under strictly increasing transformations of the variables. The advantage of the copula approach mainly lies that it clearly separates the joint distribution density into its marginal densities and the copula densities while the transformation method does not explicitly. Also, we note that the normal copula (correlation) parameter ρ is not the correlation coefficient between random variables x and y any more, but the correlation coefficient between these two variables is much smaller than the copula parameter ρ (equal to only if both variables are normal).

Very interestingly, we note that the inverse function of the standard normal distribution function involved in the Normal copula approach to constructing bivariate distributions, is actually the well-known normalization, and in doing so to each component variable, the Normal copula method may be regarded as the normalization in the multi-dimension domain.

6.2.3.2 Archimedean Families

The Frank, Clayton, Gumbel, and AMH belong to an important class of copulas known as Archimedean (Genest and MacKay 1986), which is of the form as,

$$C(u,v) = \varphi^{-1}(\varphi(u),\varphi(v)) \tag{6.21}$$

where φ, called the generator function, is a continuous strictly decreasing convex function, defined on [0,1] and satisfying that $\varphi(0) = \infty$ and $\varphi(1) = 0$. Table 6.2 lists generators of these four Archimedean members. These families of this class are capable of representing different types of dependencies. Frank copula implies symmetric dependence patterns. Gumbel and Clayton copulas are both asymmetric, implying higher dependency at right tails and left tails, respectively. These Archimedean copulas are among some popular copula models and have been increasingly used in applications, especially in financial and actuarial fields (Frees and Valdez 1998, Klugman and Parsa 1999, Embrechts et al. 1999, 2001, Carrière 2000, Li 2000, Henessey and Lapan 2002, Ané and Kharoubi 2003, Hürlimann 2004, Mendes and Souza 2004, Wüthrich 2004) as a tool for investigating problems such as risk measurement. Procedures also exist for choosing a particular member of a given family of Archimedean copulas to fit a data set (Genest and Rivest 1993). Another attractive point of the Archimedean class of copulas is their rather simple closed form expressions for both the copula function and copula density function, compared to the normal copula.

Table 6.2: Archimedean Families of One-Parameter Copulas

Family	Function Form $C(u,v)$	Generator Function $\varphi(t)$
Frank	$-\frac{1}{\theta}\ln[1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}]$	$-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$\frac{t^{-\theta}-1}{\theta}$
Gumbel	$e^{-[(-\ln u)^{\theta}+(-\ln v)^{\theta}]^{\frac{1}{\theta}}}$	$(-\ln t)^{\theta}$
AMH	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\ln\frac{1-\theta(1-t)}{t}$

6.3 Parameter Estimation

There are usually two approaches to fitting a copula model. The first one is to estimate marginal distributions and the copula jointly. The second is a two-step procedure, estimating the marginal and the copula parameter separately.

6.3.1 Joint Estimation

For the joint estimation, the Maximum Likelihood Estimation (MLE) is usually employed to estimate all the parameters simultaneously. With the density function of a bivariate distribution given as,

$$h(x,y)=f(x)\times g(y)\times c(u,v) \qquad (6.4)$$

For a sample of size n, the likelihood function will be given as,

$$L=\sum_{i=1}^{n} f(x_i)g(y_i)c(u_i,v_i)=\sum_{i=1}^{n} f(x_i)g(y_i)c(F(x_i),G(y_i)) \qquad (6.22)$$

The one-step procedure estimates parameters by maximizing the log-likelihood for the joint distribution, that is,

$$\ln L=\sum_{i=1}^{n}[\ln f(x_i)+\ln g(y_i)+\ln c(u_i,v_i)]$$

$$=\sum_{i=1}^{n}[\ln f(x_i)+\ln g(y_i)+\ln c(F(x_i),G(y_i))] \qquad (6.23)$$

The resulting estimate for the copula parameter would be marginal-dependent, just as the estimates of the parameters involved in the marginal distributions would be indirectly affected by the copula.

6.3.2 Two-Step Estimation

Note that the log likelihood of the whole system maybe very complicated thus the optimization may be rather difficult to solve (MLE computationally too difficult or infeasible). In this case, some two-stage estimation can be adopted, as the copula representation splits the parameters into marginal parameters and the copula parameter(s). By re-expressing the log-likelihood function as,

$$\ln L = \sum_{i=1}^{n} \ln f(x_i) + \sum_{i=1}^{n} \ln g(y_i) + \sum_{i=1}^{n} \ln c(F(x_i), G(y_i)) \tag{6.24}$$

the marginal distributions F and G can be estimated firstly using MLE by maximizing the marginal log-likelihood for F and G respectively, that is,

$$\ln L_x = \sum_{i=1}^{n} \ln f(x_i),\ \ln L_y = \sum_{i=1}^{n} \ln g(y_i) \tag{6.25}$$

After obtaining the MLE estimates of marginal parameters, substitute them into (6.24) then estimate the copula parameter by maximizing the log-likelihood for the copula,

$$\ln L_c = \sum_{i=1}^{n} \ln c(\hat{F}(x_i), \hat{G}(y_i)) \tag{6.26}$$

where $\hat{F}(x)$ and $\hat{G}(y)$ denote the MLE estimates of the marginals from the first step. This two-step MLE method is referred to as the method of Inference Functions for Margins or IFM method (Joe and Xu 1996, Joe 1997, 2004). The advantage of the IFM over the joint ML is primarily in computational ease, but these estimators are less efficient than ML (Joe and Xu 1996, Joe 2004). Note that this IFM method is the same as the MLE for the multivariate normal distribution. Obviously, with the IFM method, the estimation of the copula parameter depends on the choice of univariate distributions and the misspecification of marginals may affect the estimation of the copula parameter(s).

Other than the MLE used in the IFM, the two-step procedure allows for other alternatives of estimating parameters of the marginal distributions as well as the copula parameter. For the marginal distributions, (linear or nonlinear) least squares methods can be employed, which have been found superior for fitting Johnson's S_B (Zhou and McTague 1996, Kamziah et al. 1999, Zhang et al. 1999) and the Logit-Logistic in the forestry literature, for details see Chapter 4.

For the correlation parameter, it can be estimated "completely independently of" the marginal fitting without specifying marginals. One approach is to use the empirical marginal distributions (EDF) $F_n(x)$ and $G_n(y)$ as the "estimated" marginals, that is, to estimate the copula by maximizing the log-likelihood for the copula,

$$\ln L_c = \sum_{i=1}^{n} \ln c(F_n(x_i), G_n(y_i)) \qquad (6.27)$$

Several estimates of the EDF could be obtained from, for example, i/n, i/(n+1), or the others (see Chapter 4). This approach to estimating copula parameter was probably originally proposed by Genest et al. (1995) and is known as the Canonical maximum Likelihood method (CML). The CML method is sometimes called "semiparametric two-stage estimation" in that the marginals are estimated as the empirical distributions. It is appropriate when one does not want to specify any parametric model to describe the marginal distribution and in such cases the inference about the dependence parameter should be marginal-free. For each copula model, we should obtain the maximum likelihood estimation (MLE) of the dependence parameter. In this case, the model selection problem could be addressed using the AIC criteria through the resulting values of pseudo log-likelihood of the copula model. We note that the estimate of the copula parameter using CML method has a larger asymptotic variance than the maximum likelihood estimator of this parameter assuming marginals known (Genest et al. 1995).

The other approach is to match the sample correlation measure (Kendall's tau or Spearman's rho) to the theoretical dependence measure which can be expressed as a function of the correlation parameter, independent of marginal distributions. This method is in the spirit of Pearson's method of moments (Genst and Rivest 1993). For example, we could estimate the copula parameter by using a relationship between Kendall's tau and the copula (Genest 1987, Genest and Rivest 1993). This method was reported by Genest (1987) to perform reasonably well in the case of Frank's family for samples of size 50 or larger. This method's efficiency was also investigated by Oaks (1986) for Clayton's system of distributions. Both approaches to estimating the copula parameter are marginal free, thus we can evaluate model performance based on the choice of copula families rather than the joint distributions if we accept the same marginal models due to some prior knowledge on the univariate domain.

Clearly, combinations of marginal estimation methods with the copula estimation methods will provide many choices for the two-step approach. As contrast to the IFM, we may firstly estimate the copula parameter independently of the marginals, and then estimate the marginal parameters.

In this Chapter, we mainly use the joint MLE and the CML methods to estimate parameters. The resulting log-likelihood from both methods can be subsequently used as a measure of fit to rank alternative bivariate distributional models, which is essentially Akaike's information criterion (AIC) due to the fixity of the marginals thus the same number of parameters across estimated models. The AIC is defined by

$$\text{AIC}=2(\text{negative log-likelihood})+2p \qquad (6.28)$$

where p is the number of parameters of the model. The lower is the AIC, the better the model. In our case all models have the same number of parameters, thus equivalently the lower the negative log-likelihood, the better the model. This advantage makes it easier to compare different bivariate distributions obtained from the same copula with different marginals or different copulas with the same marginals. It is to be mentioned that for the Archimedean Families, Genest and Rivest (1993) proposed a procedure for identifying the particular family of Archimedean copulas that provides the best possible fit to the data. However, for our purposes, we have to compare the Archimedean families with the other copulas thus we do not explore Genest and Rivest's procedure in this study.

We have to admit that there are many other selection criteria in choosing specific copula from a family of them, such as the distance between the considered copula and the empirical copula and the generalized likelihood ratio test (Vuong 1989), however, the AIC seems much simple and straightforward with the joint MLE estimation or the CML method. The CML resulting AIC may be especially convenient for the comparison of different copulas with the same set of univariate marginals.

6.4 Case Study

6.4.1 Data

The same 102 of the plots of the Chinese fir as used in Chapter 5, were used in this study.

6.4.2 Normal Copula with Different Marginals

Various univariate distributions have been used in modelling diameter distributions. However, not all of these univariate models have their satisfactory counterparts in the bivariate domain which are suitable for modelling the joint distribution of tree diameters and heights, one example being the Weibull. Considering the successful applications of them in forestry, which is obtained from the normal copula approach, we then well reasonably conclude that the normal copula may provide a satisfactory way of constructing bivariate models with given marginals which we may already know are suitable in the univariate domain. Such marginal models include the Logit-Logistic, S_B, and Beta, for details see Chapter 5.

In Chapter 5, we used the Plackett copula to construct bivariate distribution models, with the Logit-Logistic, S_B and Beta as the marginal distributions, respectively. The three bivariate distribution models and Johnson's S_{BB} are fitted to the same data set as used in this study, and it was found that the S_{BB} performed best, and for the three Plackett-based models, Logit-Logistic was better than Beta and S_B, and Beta better than S_B. We argued that the better performance of the S_{BB} over the Plackett's models is mainly due to the construction methods, and now with the concept of the copula, we may have reasons to prefer the normal copula for coining bivariate models. It was also found that the comparison results for the three Plackett-based models are consistent with those results made for marginal distributions, diameter and height, respectively, for details see Chapter 3 and 5. This fact may indicate that with the same copula, the performance of the resulting bivariate distributions will depend on the marginal models. In Chapter 3, it was already found that the Logit-Logistic was more flexible than the S_B and Beta in the univariate distribution modelling situation. Hence with the normal copula, we may expect that the superior univariate flexibility extends into the bivariate domain, that is, better performance of normal-copula based bivariate Logit-Logistic over the S_{BB}.

The normal copula is then employed to set up bivariate distributional models, with the Logit-Logistic, the S_B (Johnson's S_{BB} resulting) and Beta as the marginal distributions, respectively. The CDF and PDF of the Logit-Logistic are given as,

$$F(x) = \frac{1}{1 + e^{\frac{r}{\delta}} \left(\frac{x-\xi}{\xi+\lambda-x}\right)^{-\frac{1}{\delta}}} \tag{6.29}$$

$$f(x) = \frac{\lambda}{\delta} \frac{1}{(x-\xi)(\xi+\lambda-x)} \frac{1}{e^{-\frac{r}{\delta}} \left(\frac{x-\xi}{\xi+\lambda-x}\right)^{\frac{1}{\delta}} + e^{\frac{r}{\delta}} \left(\frac{x-\xi}{\xi+\lambda-x}\right)^{-\frac{1}{\delta}} + 2} \tag{6.30}$$

The S_B PDF is given by:

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} \frac{\lambda}{(\xi+\lambda-x)(x-\xi)} e^{-\frac{1}{2}\left[\frac{\ln\left(\frac{x-\xi}{\xi+\lambda-x}\right)-r}{\delta}\right]^2}, \xi \le x \le \xi + \lambda \tag{6.31}$$

The Beta PDF is,

$$f(x) = \frac{1}{B(p,q)} \frac{(x-a)^{p-1}(b-x)^{q-1}}{(b-a)^{p+q-1}} \quad \text{where, } a \le x \le b, p > 0, q > 0 \tag{6.32}$$

The three bivariate distribution models and the S_{BB} are fitted to each sample of our data set with the joint estimation approach using S-Plus (Mathsoft 1999), and the resulting negative log-likelihood is used to compare model performance.

Table 6.3 listed fitting results, the negative log-likelihood ($-\Lambda\Lambda$) for the three bivariate distributions which are constructed using the normal copula but with three different marginal models, in fitting each plot diameter and height data. It is found from this table that the bivariate model with the Logit-Logistic as marginals did not converge on a relatively large proportion of sample plots, 23 out of 102, in contrast to 0 and 7 out of 102 for the S_{BB} and Beta models, respectively. The reason is not clear to us, since we already found in Chapter 3 that all the three univariate models converged on almost all the sample marginal distributions, diameter and height, respectively, and no improvement was found at all even we have used these marginal parameter estimates as the starting points for the joint estimation.

Table 6.4 indicates the between-model comparative performance of the three models in terms of their goodness of fit statistics on the sample-plot diameter-height joint distributions on which both models converged. For example, the bivariate Logit-Logistic (LL) distribution had a lower of ($-\Lambda\Lambda$) than the bivariate Beta for 72 out of the 79 distributions for which joint convergence of bivariate Logit-Logistic and bivariate Beta were obtained. From Table 6.4, it follows that LL>Beta>S_B where ">" represents "better than". Both bivariate Logit-Logistic and Beta performed better than bivariate S_B (Johnson's S_{BB}), which is consistent with the comparison results of the three marginal distributions we got in Chapter 3. Together with the comparison results obtained for the Plackett-based bivariate models in Chapter 5, we see empirically that with the same copula (method of constructing bivariate models), the choice of marginal distributional models will play an important role in deciding the superior bivariate distributional model.

Table 6.3: minus log-likelihood of three bivariate distributions based on normal copula

Keys: LL-Bivariate Logit-Logistic, Betaa-Bivariate Beta

Plot	LL	S_{BB}	Beta	Plot	LL	S_{BB}	Beta
1	333.79	335.22	335.00	52	334.46*	337.64	337.49
2	399.54	400.37	400.05	53	443.58	444.25	444.04
3	414.66	415.55	414.59	54	667.79	669.32	668.09
4	491.73*	492.02	491.31	55	362.80	364.91	365.16
5	421.21	422.01	421.97	56	363.17	366.02	365.58
6	388.96	390.74	390.48	57	328.86	329.75	329.68
7	569.77	570.96	570.86	58	460.59	462.10	462.00
8	468.05	468.81	468.47	59	344.67	346.20	346.11
9	469.35	470.60	470.44	60	339.00	340.19	340.03
10	325.51	328.49	328.06	61	360.33	361.60	361.59
11	299.94*	305.73	298.43*	62	253.34*	260.95	260.88
12	242.41*	246.64	246.49	63	252.05	252.76	252.69
13	278.12	278.51	278.51	64	519.76	521.55	519.71
14	218.17	218.86	218.94	65	361.84	362.62	362.55
15	339.92	342.63	342.46	66	367.09	367.41	367.19
16	316.02	316.56	316.53	67	427.16	428.86	428.53
17	352.14	353.72	353.63	68	558.53*	561.45	561.25
18	548.96	551.41	551.14	69	353.80	355.03	354.81
19	467.89	469.15	469.16	70	442.03*	445.40	445.20
20	546.01	548.22	548.11	71	547.27	548.71	548.58
21	252.58	252.82	252.67	72	400.42*	406.89	406.66
22	285.58	286.91	287.36	73	241.21	241.25	238.46
23	286.07	286.80	286.72	74	219.98	220.70	220.74
24	385.83	385.85	385.83	75	277.33*	282.28	280.95*
25	316.54	317.54	317.39	76	259.00	260.80	260.14
26	347.33	348.33	348.23	77	354.89	354.59	354.44
27	299.74	299.98	299.99	78	516.40	516.47	516.58
28	419.19	420.42	419.95	79	693.96	697.00	695.90
29	333.71*	336.03	335.89	80	413.83*	418.69	418.10
30	338.22	339.28	339.29	81	384.11	385.57	384.84
31	405.49	404.80	405.05	82	353.92	358.46	357.98
32	334.82	335.39	334.63	83	245.81	246.23	246.11
33	523.83*	528.21	527.87	84	303.67*	314.11	310.15*
34	372.93	374.01	373.87	85	646.74	647.25	647.48
35	473.50	474.49	474.45	86	302.64	303.06	303.02
36	643.22	644.06	643.50	87	461.35*	465.30	465.26
37	273.30*	278.66	272.47*	88	353.65	355.34	355.10
38	246.85	246.99	247.04	89	320.57	321.45	321.55
39	534.77*	539.39	539.51*	90	628.66	629.12	629.32
40	649.21	648.84	649.29	91	428.77	429.81	429.63
41	477.76	482.02	481.58	92	257.95*	262.17	261.78
42	510.67	512.31	512.08	93	444.02	444.55	444.58
43	310.98	312.79	312.34	96	387.35*	390.72	389.96
44	620.69	621.38	621.34	97	460.75*	463.77	463.43
45	800.28	801.59	801.42	98	513.15	515.41	514.97
46	535.32	536.36	536.58	99	450.17*	451.72	451.58
47	519.60*	526.30	520.26*	100	488.52	490.36	490.09
48	388.82	391.40	391.04	101	418.76	418.53	418.28
49	756.94	758.15	757.14	102	351.48*	361.98	349.06*
50	267.49	268.72	268.42	104	253.79	253.75	253.80
51	328.16*	329.62	329.48	106	291.48	292.60	292.39

Note: * denotes un-convergence in the joint MLE fitting

Table 6.4. Comparison of 3 normal copula based bivariate distribution models based on $\{-\Lambda\Lambda\}$

Keys: LL-Bivariate Logit-Logistic, Beta-Bivariate Beta

row \ col	Beta	S_{BB}
LL	72/79	74/79
Beta		77/95

Proportion of cases (in which there was joint fitting convergence) in which the row-distribution model had a lower {-LL} than the col-distribution.

Because of the better performance of the bivariate Logit-Logistic over the S_{BB}, we may desire to estimate parameters of this bivariate model properly for those samples where the one-step procedure failed. For this purpose, some two-step procedure would be worthwhile. Meanwhile, it is interesting to see whether the joint MLE estimates of the copula parameter differ significantly from those obtained from the two-step procedure. For simplicity, we estimated the copula parameter by matching the sample Kendall's tau or Spearman's rho to its theoretical correspondence expressed in function of this copula parameter, that is, $\tau = \frac{2}{\pi}\arcsin(\rho)$ and $\rho_s = \frac{6}{\pi}\arcsin(\frac{\rho}{2})$, respectively. Table 6.5 lists estimates of the copula parameter (ρ) obtained from the one-step procedure with the three marginal distributional models and from the two-step procedure. Rather than comparing these two approaches in terms of goodness-of-fit statistics, we compared these estimates of the copula parameter ρ with those obtained from the above one-step MLE. Table 6.6 listed the paired t test results. Clearly, all the one-step estimates of ρ for the bivariate models with different marginals do not show significant differences from those obtained by matching Kendall's tau, that is, $\hat{\rho} = \sin(\frac{\pi\tau}{2})$. Also, no significant differences is found between the one-step estimates for the bivariate Logit-Logistic and those obtained by matching Spearman's rho, that is, $\hat{\rho} = 2\sin(\frac{\pi\rho_s}{6})$, while there are significant differences between the one-step estimates for the bivariate Beta or S_B and those by matching Spearman's rho, which may indicate that the estimation of the copula parameter may be affected by the choice of the marginal distributions. Furthermore, significant difference is shown between tau based and rho based estimates. Therefore, if the one-step MLE does not converge in fitting, we may independently estimate the copula parameter by matching Kendall's tau, and then fit the marginals (separately or possibly simultaneously given the estimated copula parameter).

Table 6.5: Normal Copula Parameter Estimates ($\hat{\rho}$)

Keys: LL-Bivariate Logit-Logistic, Beta-Bivariate Beta

	Joint MLE			Macthing			Joint MLE			Macthing	
plot	LL	Beta	S_{BB}	tau	rho	plot	LL	Beta	S_{BB}	tau	rho
1	0.88	0.88	0.88	0.89	0.89	52	0.88*	0.88	0.88	0.86	0.85
2	0.84	0.85	0.85	0.84	0.83	53	0.72	0.72	0.72	0.72	0.74
3	0.93	0.92	0.92	0.94	0.94	54	0.89	0.90	0.90	0.88	0.87
4	0.90*	0.90	0.90	0.90	0.90	55	0.84	0.84	0.84	0.85	0.84
5	0.86	0.86	0.86	0.85	0.86	56	0.89	0.88	0.88	0.90	0.89
6	0.83	0.84	0.84	0.81	0.80	57	0.83	0.83	0.83	0.84	0.84
7	0.84	0.84	0.84	0.81	0.80	58	0.89	0.89	0.89	0.89	0.89
8	0.87	0.87	0.87	0.88	0.89	59	0.81	0.81	0.81	0.75	0.76
9	0.91	0.91	0.91	0.92	0.91	60	0.82	0.83	0.82	0.77	0.75
10	0.85	0.85	0.85	0.86	0.86	61	0.77	0.77	0.77	0.76	0.77
11	0.83*	0.87*	0.83	0.84	0.83	62	0.92*	0.92	0.92	0.93	0.93
12	0.72*	0.72	0.72	0.71	0.69	63	0.88	0.88	0.88	0.86	0.86
13	0.65	0.65	0.66	0.70	0.70	64	0.85	0.86	0.86	0.87	0.87
14	0.77	0.77	0.77	0.79	0.77	65	0.88	0.88	0.88	0.89	0.89
15	0.88	0.89	0.89	0.86	0.84	66	0.90	0.90	0.90	0.91	0.92
16	0.79	0.80	0.80	0.80	0.80	67	0.88	0.88	0.88	0.88	0.88
17	0.87	0.87	0.87	0.87	0.87	68	0.85*	0.84	0.84	0.87	0.87
18	0.89	0.90	0.90	0.89	0.88	69	0.90	0.91	0.91	0.90	0.90
19	0.91	0.91	0.91	0.91	0.91	70	0.86*	0.86	0.85	0.85	0.84
20	0.86	0.86	0.86	0.86	0.85	71	0.91	0.91	0.91	0.92	0.92
21	0.89	0.89	0.89	0.92	0.90	72	0.84*	0.84	0.84	0.80	0.77
22	0.81	0.81	0.81	0.80	0.80	73	0.96	0.95	0.96	0.97	0.97
23	0.68	0.68	0.68	0.62	0.61	74	0.85	0.85	0.85	0.85	0.84
24	0.72	0.72	0.72	0.76	0.75	75	0.90*	0.92*	0.92	0.92	0.92
25	0.78	0.79	0.79	0.80	0.78	76	0.89	0.90	0.90	0.88	0.88
26	0.85	0.85	0.85	0.86	0.85	77	0.79	0.79	0.79	0.82	0.82
27	0.74	0.74	0.74	0.73	0.72	78	0.86	0.86	0.86	0.87	0.86
28	0.91	0.91	0.91	0.91	0.91	79	0.85	0.85	0.85	0.85	0.85
29	0.91*	0.90	0.90	0.89	0.88	80	0.87*	0.88	0.87	0.87	0.88
30	0.82	0.81	0.81	0.85	0.85	81	0.89	0.89	0.89	0.90	0.91
31	0.92	0.92	0.92	0.93	0.93	82	0.77	0.76	0.76	0.79	0.80
32	0.88	0.88	0.88	0.88	0.87	83	0.80	0.80	0.80	0.79	0.78
33	0.86*	0.86	0.86	0.86	0.86	84	0.93*	0.94*	0.93	0.93	0.92
34	0.87	0.87	0.87	0.86	0.86	85	0.84	0.84	0.84	0.83	0.81
35	0.87	0.87	0.87	0.87	0.86	86	0.73	0.73	0.73	0.75	0.75
36	0.90	0.90	0.90	0.90	0.91	87	0.78*	0.78	0.78	0.73	0.72
37	0.89*	0.89*	0.89	0.89	0.90	88	0.82	0.83	0.83	0.82	0.80
38	0.76	0.76	0.76	0.81	0.80	89	0.81	0.82	0.82	0.79	0.78
39	0.83*	0.83*	0.83	0.82	0.81	90	0.89	0.88	0.88	0.89	0.89
40	0.87	0.87	0.87	0.88	0.88	91	0.87	0.87	0.87	0.89	0.88
41	0.82	0.83	0.82	0.83	0.82	92	0.83*	0.80	0.80	0.76	0.76
42	0.85	0.86	0.86	0.85	0.84	93	0.81	0.81	0.81	0.79	0.79
43	0.85	0.86	0.86	0.85	0.85	96	0.90*	0.90	0.90	0.90	0.90
44	0.90	0.90	0.90	0.91	0.91	97	0.91*	0.91	0.91	0.91	0.91
45	0.91	0.91	0.91	0.92	0.93	98	0.89	0.89	0.89	0.89	0.88
46	0.89	0.89	0.89	0.90	0.90	99	0.89*	0.88	0.88	0.89	0.89
47	0.89*	0.89*	0.90	0.92	0.92	100	0.79	0.80	0.80	0.82	0.81
48	0.90	0.90	0.90	0.87	0.87	101	0.82	0.83	0.83	0.81	0.79
49	0.86	0.86	0.86	0.86	0.86	102	0.88*	0.88*	0.86	0.83	0.83
50	0.86	0.85	0.85	0.88	0.89	104	0.91	0.91	0.91	0.90	0.90
51	0.90*	0.90	0.90	0.90	0.89	106	0.79	0.78	0.78	0.78	0.78

Note: * denotes un-convergence in the joint MLE fitting

Table 6.6. Paired t-test for difference in the normal copula $\hat{\rho}$ with different approaches

Pair	Paired Differences Mean	Std. Dev.	t	df	p-value (2-tailed)
$\hat{\rho}_{\tau}$, $\hat{\rho}_{r_s}$	0.00436	0.00711	6.186	101	<0.0001
$\hat{\rho}_{LL}$, $\hat{\rho}_{\tau}$	-0.00236	0.01935	-1.082	78	0.282
$\hat{\rho}_{Beta}$, $\hat{\rho}_{\tau}$	0.00077	0.02072	0.365	94	0.716
$\hat{\rho}_{S_{BB}}$, $\hat{\rho}_{\tau}$	0.00074	0.01989	0.375	101	0.709
$\hat{\rho}_{LL}$, $\hat{\rho}_{r_s}$	-0.00176	0.02175	0.720	78	0.472
$\hat{\rho}_{Beta}$, $\hat{\rho}_{r_s}$	0.00521	0.02394	2.119	94	0.037
$\hat{\rho}_{S_{BB}}$, $\hat{\rho}_{r_s}$	0.00510	0.02309	2.229	101	0.028
$\hat{\rho}_{LL}$, $\hat{\rho}_{Beta}$	-0.00151	0.00338	-3.965	78	<0.0001
$\hat{\rho}_{LL}$, $\hat{\rho}_{S_{BB}}$	-0.00145	0.00343	-3.756	78	<0.0001
$\hat{\rho}_{Beta}$, $\hat{\rho}_{S_{BB}}$	0.00028	0.00126	2.127	94	0.036

Note: $\hat{\rho}_{\tau}$-matching sample Kendall's tau with the theoretical, $\hat{\rho}_{r_s}$-matching sample Spearman's rho with the theoretical, $\hat{\rho}_{LL}$-joint MLE with LL marginals, $\hat{\rho}_{Beta}$-joint MLE with Beta marginals, $\hat{\rho}_{S_{BB}}$-MLE with the S_{BB}.

6.4.3 Comparison of Copula Models

Since the Logit-Logistic is the most flexible distributional model among the three we compared and the resulting normal copula bivariate model consistently showed the best performance, we then use this model as marginals and employ other copulas to construct bivariate models, thus comparing the performance of these copulas in modelling the correlation structure between tree diameters and heights. Table 6.7 lists copula density functions in facilitating fitting these bivariate distributions.

In Chapter 5, we have already used the Plackett copula and the resulting negative log-likelihood can be found there. However, for convenience, we listed these results together with the results for the other copulas in Table 6.8. The following table 6.9 indicates the between-model comparative performance of these seven bivariate distributional models. From table 6.9, overall, it follows that Normal > Frank > Plackett > Gumbel > Clayton >

AMH > FGM where ">" represents "better than". The best performance of the normal copula based bivariate distributional model indicates that it is the best way of constructing bivariate distributions for our tree diameter and height data, then if other marginal distributions are to be used, the normal copula would be the first choice to construct the corresponding bivariate models. All the other six models performed better than the FGM model in fitting each sample plot, and all the other five better than the AMH almost in fitting each sample. This is due to the fact that both copulas can only be suitable for describing "weak dependency", thus not adequate for modelling the joint distribution of tree diameter and height as regularly they are both highly dependent.

Table 6.7: Density Function of Copulas

Family	Function Form $C(u,v)$	density function $C_{12}(u,v)=\frac{\partial^2 C(u,v)}{\partial u \partial v}$
Normal	$N_\rho(\Phi^{-1}(u),\Phi^{-1}(v))$	$\frac{1}{\sqrt{1-\rho^2}} e^{-\frac{\rho}{2(1-\rho^2)}[\rho(z_x^2+z_y^2)-2z_x z_y]}$
Plackett	$\frac{1+(\theta-1)(u+v)}{2(\theta-1)}-\frac{\sqrt{[1+(\theta-1)(u+v)]^2-4uv\theta(1-\theta)}}{2(\theta-1)}$	$\frac{\theta[1+(\theta-1)(u+v-2uv)]}{\{[1+(\theta-1)(u+v)]^2-4uv\theta(1-\theta)\}^{3/2}}$
Frank	$-\frac{1}{\theta}\ln[1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}]$	$\frac{\theta(e^\theta-1)e^{\theta(1+u+v)}}{[e^\theta-e^{\theta(1+u)}-e^{\theta(1+v)}+e^{\theta(u+v)}]^2}$
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$(1+\theta)(uv)^{-1-\theta}(u^{-\theta}+v^{-\theta}-1)^{-2-\frac{1}{\theta}}$
Gumbel	$e^{-[(-\ln u)^\theta+(-\ln v)^\theta]^{\frac{1}{\theta}}}$	$\frac{(\ln u \ln v)^{\theta-1}[(-\ln u)^\theta+(-\ln v)^\theta]^{1/\theta-2}[(-\ln u)^\theta+(-\ln v)^\theta+\theta-1]^{1/\theta}}{uve^{[(-\ln u)^\theta+(-\ln v)^\theta]^{1/\theta}}}$
AMH	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\frac{1+\theta(uv+u+v-2)-\theta^2(1-u)(1-v)}{[1-\theta(1-u)(1-v)]^3}$
FGM	$uv[1+\theta(1-u)(1-v)]$	$1+\theta(1-2u)(1-2v)$

As we know, the normal, Frank and Plackett copulas all imply symmetric dependence patterns, while Gumbel and Clayton copulas are both asymmetric, implying higher dependency at right tails and left tails, respectively. Meanwhile as we know that "in the even-aged forest stand, the tallest (dominant) trees and the shortest (suppressed) trees are associated, respectively, with the largest and smallest diameters" (Schreuder and Hafley 1977), the better performance of these three symmetric copulas (normal, Frank, and Plackett) over the other two asymmetric copulas (Gumbel and Clayton) may further indicate that the dependence between tree diameters and heights is somewhat symmetric, that is, the association between the tallest trees and the largest diameters may behave in the similar way as the association between the shortest trees and the smallest diameters. The out-performance of the normal copula over the Frank, Plackett may indicate that this copula can be more capable of capturing the "symmetric dependence" between tree diameters and heights than the other two alternatives.

Furthermore, we calculated the "fitted" Kendall's tau with correlation parameter estimates obtained for each copula except Plackett's copula for which no closed tau formula exists, see table 6.10. We then compared the "fitted" Kendall's tau resulted from each copula with the sample Kendall's tau by using paired t test. Table 6.11 showed the results: there is no significant difference between the sample tau and the fitted tau for the normal and Gumbel copulas, while there is for Frank and Clayton. So, in terms of Kendall's tau, the normal and Gumbel copulas may be regarded as more "successfully" reflecting the dependency for our data while Frank and Clayton not. The inefficiency of Frank and Clayton in expressing dependency may also be recognized by the observation that 92 out of 99 (number of sample with fitting converged) fitted tau with Frank copula is bigger than the sample Tau while 98 out of 98 fitted tau is smaller than the sample tau for the Clayton copula. We have to admit that Kendall's tau is only one measure of dependence and there should be other aspects of dependence for which tau can not capture, but comparison of the fitted tau with the sample tau may provide a criterion to decide which copula is not well suited.

In fact, for selecting one particular copula from a collection of them, we may use the CML method to fit the copula itself without specifying any univariate model but using empirical univariate distributions ($i/(n+1)$, we adopted here) for approximating the marginals. The resulting pseudo log-likelihood is then used as the comparing criterion. Table 6.12 lists the pseudo log-likelihood for each copula, and table 6.13 shows the comparison results for the seven copulas. Not surprising, the CML based comparison results are consistent with those based on the joint log-likelihood above. This marginal free selection procedure may be preferred for choosing an appropriate copula from a given family.

In addition, we compared the estimates of the normal copula parameter ρ with those obtained from the above one-step MLE. Table 6.14 listed the paired t test results. Clearly, all the two-step estimates of ρ do not show significant differences. Also, no significant differences is found between the CML estimates and those from fitting the bivariate (normal) Logit-Logistic, while there are significant differences between the two-step estimates and those for the bivariate (normal) Beta or S_{BB}. Therefore, for the bivariate normal copula based Logit-Logistic, if the one-step MLE does not converge in fitting, we may independently estimate the copula parameter by matching Kendall's tau or using the CML method, and then fit the marginals.

Table 6.8: Negative log-likelihood of 7 bivariate distributions based on 7 copulas with logit-logistic as marginals

Plot	Normal	Frank	Plackett	Gumbel	Clayton	AMH	FGM
1	333.79	330.07	334.20	333.36	351.87	365.40*	386.06
2	399.54	403.97	404.64	407.52	399.90	415.00	441.47
3	414.66	419.33	422.05	417.54	435.56	463.02	495.61
4	491.73*	496.19	496.93	484.61*	521.18	546.49	572.13
5	421.21	429.68	430.09	428.77	429.49	447.06	472.79
6	388.96	398.81	396.80	393.22	400.26	410.28	434.68
7	569.77	583.21	581.87	581.30	576.73	593.42	629.03
8	468.05	466.96	472.21	473.64	479.95	502.11	528.35
9	469.35	466.53	469.05	478.16	468.87	515.40	559.27
10	325.51	327.23	328.88	329.65	330.68	345.05	366.98
11	299.94*	306.43	305.35	307.65	294.44*	317.09*	337.71
12	242.41*	247.07	245.42	248.03	245.46	242.89*	259.77
13	278.12	277.20	278.20	280.77	281.17	281.40	286.72
14	218.17	219.17	219.43	221.19	221.79	224.46	235.19
15	339.92	351.38	349.76	347.04	347.60	362.88	391.69
16	316.02	314.40	316.70	321.25	317.77	325.36	341.44
17	352.14	357.77	359.13	359.28	361.11	377.20	404.58
18	548.96	554.36	555.90	560.50	555.55	595.87	642.04
19	467.89	466.73	469.55	476.57	465.04	504.52	543.08
20	546.01	555.08	553.42	553.95	558.44	583.87	621.14
21	252.58	250.25	246.42	248.95*	258.50	270.48*	290.20
22	285.58	291.18	291.43	289.86	293.20	300.93	316.68
23	286.07	292.56	292.00	284.26*	294.04	294.10	299.06
24	385.83	385.18	384.59	387.88	389.90	393.73	405.59
25	316.54	316.33	319.05	322.88	317.81	324.93	341.51
26	347.33	344.80	346.15	352.67	346.42	360.90	381.55
27	299.74	306.22	307.96	307.37	301.91	305.36	319.71
28	419.19	413.60	418.27	422.47	422.86	460.67	496.50
29	333.71*	345.51	344.27	341.60	334.90*	360.02	385.91
30	338.22	336.49	339.18	339.36	350.90	360.17	376.27
31	405.49	402.72	403.85	411.60	406.17	454.68	486.71
32	334.82	339.60	340.33	340.56	345.72	363.32	388.70
33	523.83*	535.49	534.71	528.77	555.27	578.81	608.90
34	372.93	376.13	378.27	379.64	383.92	399.64	426.70
35	473.50	477.61	476.60	482.46	471.72	502.03	540.67
36	643.22	650.19	649.60	646.06*	665.95	707.16	757.79
37	273.30*	285.15	286.77	275.78*	295.54	309.29	328.12
38	246.85	241.60	240.38	247.70	244.94	251.04	260.78
39	534.77*	549.53	550.58	548.57	540.49	554.48	580.81
40	649.21	648.73	649.73	654.50	662.44	693.07	729.35
41	477.76	481.43	476.86	478.68	490.58	501.17	525.03
42	510.67	514.87	513.62	521.07	504.95	531.00	569.63
43	310.98	313.47	316.29	318.01	316.70	328.46	348.16
44	620.69	616.83	622.59	629.12	629.30	675.31	712.89
45	800.28	797.67	801.41	804.11	821.92	875.48	930.58
46	535.32	539.76	540.88	538.00	555.21	580.55	614.88
47	519.60*	520.33*	529.08	525.74*	548.73	579.52*	615.83
48	388.82	397.65	398.42	400.19	388.99	414.80	449.02
49	756.94	762.47	763.95	771.92	762.96	799.79	857.73
50	267.49	264.77	268.10	268.95	278.17	285.97	294.70
51	328.16*	329.70	328.62	330.60	340.55	359.32	380.39

Plot	Normal	Frank	Plackett	Gumbel	Clayton	AMH	FGM
52	334.46*	348.25	344.75	337.74*	348.75	364.05	391.10
53	443.58	441.99	442.29	445.93	449.51	452.64	463.39
54	667.79	679.98	677.64	677.05	681.27	717.80	769.73
55	362.80	366.33	366.59	370.34	367.27	381.11	406.65
56	363.17	364.77	365.52	365.62	374.16	392.42	416.40
57	328.86	329.23	330.76	334.74	330.06	342.69	362.21
58	460.59	464.15	463.98	466.73	474.10	503.67	536.91
59	344.67	349.90	351.57	354.68	341.29	348.49	371.99
60	339.00	350.83	349.02	342.08	348.70	354.56	371.88
61	360.33	364.44	364.61	364.62	364.40	368.51	385.05
62	253.34*	264.38	266.07	261.17*	269.95	289.84*	311.85
63	252.05	253.28	254.11	257.58	250.87	267.53	287.05
64	519.76	516.81	519.54	518.09	551.55	574.37	600.74
65	361.84	359.52	363.86	370.19	360.75	383.58	411.06
66	367.09	363.25	367.31	370.87	376.67	403.10	424.57
67	427.16	436.98	438.42	437.07	436.42	457.10	488.37
68	558.53*	552.42	561.26	568.31*	566.06	587.43	617.37
69	353.80	361.45	362.10	361.84	358.03	382.72	410.07
70	442.03*	450.22	448.22	449.08	451.58	464.48*	491.33
71	547.27	547.88	551.14	553.62	563.41	608.39	644.90
72	400.42*	414.96	409.17	411.10	404.78	417.28	449.47
73	241.21	239.46*	276.40*	237.65	255.29	293.20*	310.73*
74	219.98	224.10	222.57	217.51	236.81	246.03	258.58
75	277.33*	283.36	284.64	270.89*	289.79	317.88	345.09
76	259.00	261.87*	263.75*	266.83*	252.05	274.48	297.92
77	354.89	353.10	353.49	354.04	366.44	370.48	378.79
78	516.40	516.47	516.22	519.18	533.18	557.43	587.52
79	693.96	694.48	696.88	704.77	693.42	720.49	770.89
80	413.83*	425.02	426.61	421.55	438.17	454.24*	478.72
81	384.11	387.37	390.74	393.39	388.71	412.23	439.32
82	353.92	354.75	357.45	357.78*	366.28	368.92	378.75
83	245.81	247.33	248.26	248.80	249.07	254.07	264.41
84	303.67*	319.18	318.00	308.47*	322.93	349.51*	376.02
85	646.74	657.66	651.28	643.63	671.42	687.63	716.01
86	302.64	303.19	305.33	305.59	309.18	311.28	319.40
87	461.35*	471.40	470.34	472.50	460.36	466.02*	492.92
88	353.65	356.07	355.80	359.69	356.04	367.31	390.89
89	320.57	329.32	327.32	324.27	328.97	334.14	353.25
90	628.66	638.71	638.22	636.38	649.74	684.09	724.22
91	428.77	428.38	430.73	434.90	436.58	461.90	491.93
92	257.95*	266.08	266.50	266.20	253.11*	266.33*	280.33
93	444.02	450.80	451.93	451.59	450.50	459.26	481.29
96	387.35*	396.55	397.96	385.66*	406.36	428.76	453.53
97	460.75*	470.48	473.93	460.61*	490.49	519.49	548.60
98	513.15	520.54	519.23	516.64	532.02	560.51	593.02
99	450.17*	455.67	457.33	458.93	459.57	484.86*	515.22
100	488.52	482.44	478.37	488.55	489.11	506.55	537.24
101	418.76	424.39	423.34	425.13	418.86	430.83	455.59
102	351.48*	371.01	369.30	360.98*	342.68*	375.78*	399.09
104	253.79	254.64	255.92	253.18	266.69	282.24	294.55
106	291.48	287.23	289.81	295.99	292.54	298.61	311.24

Note: * denotes un-convergence in the joint MLE fitting

Table 6.10: Kendall's tau by Sample and Copulas

plot	sample	normal	Frank	Gumbel	Clayton	Plot	sample	normal	Frank	Gumbel	Clayton
1	0.70	0.68	0.72	0.68	0.66	52	0.66	0.69*	0.69	0.67*	0.60
2	0.63	0.64	0.65	0.63	0.59	53	0.52	0.51	0.55	0.52	0.47
3	0.77	0.75	0.77	0.75	0.66	54	0.68	0.70	0.73	0.70	0.65
4	0.71	0.72*	0.73	0.70*	0.64	55	0.65	0.64	0.66	0.64	0.57
5	0.65	0.66	0.67	0.65	0.61	56	0.71	0.69	0.72	0.70	0.63
6	0.60	0.62	0.63	0.61	0.54	57	0.64	0.62	0.65	0.62	0.59
7	0.60	0.63	0.63	0.62	0.57	58	0.70	0.70	0.73	0.70	0.67
8	0.69	0.67	0.69	0.66	0.62	59	0.54	0.60	0.59	0.59	0.52
9	0.74	0.74	0.76	0.75	0.70	60	0.56	0.62	0.60	0.58	0.52
10	0.66	0.65	0.68	0.65	0.61	61	0.55	0.56	0.56	0.56	0.48
11	0.63	0.62*	0.66	0.63	0.58*	62	0.76	0.75*	0.76	0.77*	0.69
12	0.50	0.51*	0.54	0.54	0.43	63	0.66	0.69	0.71	0.69	0.66
13	0.50	0.45	0.49	0.45	0.43	64	0.67	0.65	0.69	0.65	0.61
14	0.57	0.56	0.59	0.57	0.49	65	0.70	0.69	0.71	0.69	0.65
15	0.65	0.69	0.69	0.68	0.60	66	0.74	0.72	0.75	0.72	0.69
16	0.59	0.58	0.62	0.58	0.54	67	0.69	0.69	0.69	0.68	0.63
17	0.67	0.67	0.70	0.67	0.61	68	0.67	0.64*	0.67	0.64*	0.59
18	0.70	0.70	0.74	0.71	0.68	69	0.71	0.72	0.73	0.71	0.66
19	0.73	0.73	0.75	0.74	0.69	70	0.65	0.66*	0.67	0.65	0.58
20	0.66	0.66	0.68	0.65	0.61	71	0.75	0.73	0.75	0.73	0.69
21	0.74	0.69	0.75	0.73*	0.68	72	0.59	0.64*	0.64	0.64	0.57
22	0.59	0.60	0.61	0.59	0.55	73	0.84	0.82	0.74*	0.82	0.80
23	0.43	0.48	0.45	0.48*	0.35	74	0.65	0.65	0.67	0.64	0.58
24	0.55	0.52	0.54	0.52	0.48	75	0.74	0.72*	0.76	0.73*	0.70
25	0.59	0.57	0.61	0.56	0.54	76	0.69	0.70	0.69*	0.71*	0.68
26	0.66	0.64	0.68	0.65	0.62	77	0.61	0.58	0.62	0.59	0.50
27	0.52	0.53	0.51	0.50	0.47	78	0.68	0.66	0.69	0.66	0.62
28	0.73	0.72	0.75	0.73	0.70	79	0.65	0.65	0.68	0.66	0.61
29	0.70	0.73*	0.72	0.71	0.66*	80	0.68	0.68*	0.69	0.67	0.60
30	0.65	0.62	0.64	0.62	0.54	81	0.71	0.70	0.72	0.69	0.66
31	0.76	0.75	0.78	0.75	0.73	82	0.58	0.56	0.59	0.55*	0.46
32	0.69	0.68	0.70	0.68	0.63	83	0.58	0.59	0.61	0.58	0.55
33	0.67	0.66*	0.68	0.65	0.60	84	0.75	0.76*	0.79	0.75	0.72
34	0.66	0.67	0.70	0.67	0.60	85	0.63	0.64	0.65	0.63	0.57
35	0.67	0.67	0.70	0.68	0.64	86	0.54	0.52	0.54	0.52	0.46
36	0.72	0.71	0.74	0.70*	0.65	87	0.52	0.57*	0.57	0.55	0.51
37	0.70	0.69*	0.70	0.69*	0.63	88	0.61	0.61	0.65	0.62	0.58
38	0.60	0.55	0.61	0.59	0.56	89	0.58	0.60	0.60	0.60	0.50
39	0.61	0.62*	0.62	0.59	0.58	90	0.70	0.69	0.70	0.69	0.63
40	0.69	0.68	0.70	0.68	0.63	91	0.69	0.67	0.70	0.67	0.64
41	0.62	0.61	0.65	0.62	0.55	92	0.55	0.62*	0.59	0.57	0.53*
42	0.64	0 .65	0.68	0.66	0.62	93	0.58	0.60	0.60	0.59	0.53
43	0.65	0.65	0.68	0.64	0.60	96	0.71	0.71*	0.72	0.70*	0.66
44	0.73	0.71	0.74	0.72	0.70	97	0.73	0.72*	0.74	0.70*	0.66
45	0.75	0.73	0.76	0.74	0.68	98	0.69	0.69	0.72	0.69	0.65
46	0.71	0.70	0.72	0.70	0.62	99	0.70	0.70*	0.71	0.69	0.64
47	0.74	0.70*	0.74*	0.72*	0.68	100	0.62	0.58	0.64	0.61	0.57
48	0.68	0.71	0.73	0.71	0.66	101	0.60	0.62	0.63	0.62	0.57
49	0.66	0.66	0.68	0.66	0.61	102	0.62	0.68*	0.66	0.67*	0.59*
50	0.68	0.65	0.68	0.64	0.58	104	0.72	0.72	0.76	0.72	0.67
51	0.71	0.71*	0.74	0.71	0.66	106	0.57	0.58	0.63	0.59	0.54

Note: * denotes un-convergence in the joint MLE fitting

Table 6.12: pseudo log-likelihood of 7 copulas

Plot	Normal	Frank	Plackett	Gumbel	Clayton	AMH	FGM
1	75.67	74.95	72.88	75.65	51.55	37.79	24.28
2	62.44	57.17	56.47	52.30	65.75	43.86	23.23
3	118.37	110.19	109.47	111.68	93.46	51.44	28.51
4	108.62	107.09	106.55	116.02	67.54	51.33	33.72
5	73.85	66.95	65.58	64.28	64.29	44.37	24.59
6	66.13	57.70	59.85	61.26	58.57	44.12	24.12
7	86.37	74.72	77.33	74.46	80.07	58.45	31.25
8	89.63	93.09	87.44	81.48	69.11	51.28	31.77
9	118.79	120.11	121.74	103.45	118.40	64.09	34.31
10	63.55	60.88	59.43	57.22	57.37	39.23	21.99
11	51.84	48.66	49.52	45.69	46.17	33.11	18.76
12	25.33	23.02	24.92	22.88	26.61	22.31	12.02
13	25.37	24.74	23.46	21.50	20.92	19.91	14.60
14	30.68	29.48	29.15	27.31	27.10	22.44	13.69
15	70.05	57.48	60.78	64.72	65.47	41.40	21.35
16	43.98	43.87	42.20	36.92	40.02	32.39	19.48
17	73.86	67.48	66.63	63.92	66.10	43.79	24.16
18	122.67	112.72	115.20	103.80	120.28	69.66	36.60
19	95.07	102.44	98.89	78.60	100.44	56.47	30.85
20	107.83	99.36	101.78	100.50	91.78	64.05	36.03
21	57.26	56.39	59.75	58.33	43.75	27.53	16.14
22	46.99	42.90	42.21	43.77	37.00	29.75	19.02
23	21.99	17.39	17.79	22.62	16.49	16.35	11.51
24	41.78	43.51	42.90	38.58	34.94	31.36	22.20
25	44.09	40.93	40.71	37.03	43.00	33.32	18.80
26	49.64	51.86	51.73	42.22	50.24	34.54	19.47
27	36.51	31.53	29.60	28.06	36.20	30.62	17.46
28	102.51	105.41	101.18	90.07	87.74	54.53	31.26
29	66.29	57.83	58.91	59.93	64.46	36.93	19.14
30	66.53	67.67	64.71	63.19	45.35	37.78	24.99
31	110.15	112.24	111.87	96.49	101.88	55.40	29.99
32	78.33	69.00	69.99	69.59	68.48	43.32	23.72
33	123.38	114.00	115.30	121.48	90.12	66.81	40.73
34	70.62	65.90	64.13	62.01	57.38	42.01	24.21
35	94.66	90.84	93.35	82.61	102.09	61.51	32.03
36	150.71	151.55	149.37	135.72	121.32	78.57	46.78
37	72.44	66.10	64.49	71.53	51.88	35.58	21.31
38	29.64	33.31	33.48	25.11	27.30	22.12	14.33
39	66.79	55.53	55.47	58.48	64.10	44.53	23.84
40	118.39	117.71	116.49	102.94	98.50	67.39	39.54
41	73.25	68.36	72.48	70.42	61.36	45.72	27.38
42	85.53	82.84	86.50	71.85	98.33	60.53	31.39
43	53.73	49.91	48.67	44.69	48.11	33.07	18.55
44	128.58	129.47	125.80	112.52	107.63	66.83	38.16
45	175.61	181.89	177.28	159.19	145.87	88.59	50.36
46	108.40	107.41	103.47	97.83	88.11	58.51	34.49
47	128.22	131.92	123.25	115.90	92.68	63.88	38.46
48	74.11	67.30	68.71	62.56	78.16	45.08	23.72
49	146.91	141.41	142.10	125.13	145.30	95.47	52.17
50	43.32	45.63	42.59	40.71	29.48	22.99	15.69
51	66.90	65.75	67.21	64.72	49.99	33.93	20.21

Plot	Normal	Frank	Plackett	Gumbel	Clayton	AMH	FGM
52	73.16	63.81	66.46	68.63	62.53	42.80	23.57
53	43.09	43.43	43.36	39.38	37.00	33.64	23.08
54	125.47	114.21	119.10	113.38	110.46	71.69	39.03
55	70.01	65.00	65.77	58.21	66.84	44.90	24.41
56	77.41	73.67	73.47	73.08	59.20	40.36	23.30
57	54.01	52.77	51.44	44.88	53.13	36.94	20.88
58	106.28	98.52	101.76	96.46	90.10	57.00	31.31
59	39.77	36.47	35.51	28.94	48.53	36.19	18.97
60	44.28	35.36	37.82	44.42	38.64	30.48	16.35
61	44.38	41.57	40.08	39.78	39.33	33.33	20.53
62	70.62	67.47	65.70	65.90	54.23	31.86	18.34
63	47.03	44.60	44.57	39.06	49.04	29.55	15.71
64	121.99	124.14	125.62	122.75	81.30	63.16	42.71
65	69.91	72.88	69.69	57.10	66.75	42.72	24.05
66	81.89	84.25	79.01	73.70	62.17	41.25	24.72
67	89.14	79.43	79.25	76.97	82.98	49.57	26.85
68	93.03	98.81	91.69	75.19	79.51	59.61	36.19
69	76.50	69.83	68.74	66.68	68.44	40.60	22.20
70	66.98	63.26	65.25	62.76	58.45	41.89	23.34
71	135.70	134.09	130.57	122.27	106.44	66.11	37.44
72	60.31	53.45	59.46	53.45	70.07	46.38	23.18
73	102.06	111.08	107.60	106.84	68.98	35.38	21.93
74	57.75	54.60	56.18	63.96	38.96	30.37	20.22
75	81.46	79.09	78.75	71.76	75.11	41.60	22.69
76	49.93	49.56	47.29	36.95	60.24	32.81	17.45
77	44.80	45.31	44.22	45.07	27.04	25.27	18.92
78	108.91	104.61	106.87	103.05	78.25	58.86	35.57
79	113.27	109.52	107.47	93.44	115.82	76.25	41.66
80	85.07	80.84	77.84	82.74	59.70	44.29	28.09
81	77.66	76.30	72.72	62.43	78.66	43.70	24.05
82	47.16	45.60	42.73	42.45	32.23	29.06	20.66
83	29.41	26.62	26.93	26.19	26.80	20.79	11.85
84	79.08	70.13	74.02	75.05	70.34	36.28	19.02
85	104.69	93.95	101.48	110.09	81.46	62.26	35.73
86	34.20	32.98	31.16	30.31	25.87	24.06	17.03
87	45.01	38.70	39.71	36.18	51.40	39.57	20.26
88	55.35	51.41	53.27	49.07	52.15	38.13	21.00
89	51.36	43.49	44.96	46.55	46.74	35.63	19.81
90	138.38	130.61	130.49	130.52	110.43	73.94	42.04
91	95.80	91.43	91.51	86.45	83.31	54.53	30.45
92	28.23	24.68	23.76	23.07	26.84	21.30	11.98
93	58.41	51.52	50.46	49.80	51.82	41.05	23.63
96	84.81	78.58	76.71	77.70	65.79	43.50	24.90
97	117.86	110.03	108.80	113.13	83.18	55.88	32.69
98	112.79	100.30	104.01	108.09	91.00	58.68	32.63
99	91.09	88.37	86.49	77.08	81.45	52.29	29.09
100	82.34	86.96	90.58	77.35	80.52	60.35	34.89
101	56.84	51.32	53.82	49.30	57.97	40.49	20.95
102	52.69	45.23	45.39	49.99	46.65	32.27	18.18
104	54.40	49.81	50.78	55.11	37.69	25.50	15.32
106	32.20	33.62	33.28	26.25	30.99	24.75	15.18

Table 6.9: Comparison results of 7 copula-resulting bivariate distribution models based on joint $\{-\Lambda\Lambda\}$

row \ col	Frank	Plackett	Gumbel	Clayton	AMH	FGM
Normal	51/77	63/77	67/74	68/79	76/76	78/78
Frank		64/99	62/85	71/95	85/87	99/99
Plackett			59/85	68/96	85/87	100/100
Gumbel				49/83	75/77	85/85
Clayton					87/87	97/97
AMH						88/88

Proportion of cases (in which there was joint fitting convergence) in which the row-distribution model had a lower $\{-\Lambda\Lambda\}$ than the col-distribution.

Table 6.11. Paired t-test for difference between sample tau and those of copula resulted

Pair	Paired Differences Mean	Std. Dev.	t	df	p-value (2-tailed)
τ_N, τ_s	-0.00353	0.02122	-1.479	78	0.143
τ_F, τ_s	0.02014	0.01421	14.099	98	<0.0001
τ_G, τ_s	-0.00239	0.01770	-1.251	85	0.214
τ_C, τ_s	-0.05398	0.02254	-23.707	97	<0.0001

Note: τ_s-sample Kendall's tau, τ_N, τ_F, τ_G, τ_C-Kendall's tau calculated from the fitted Normal, Frank, Gumbel and Clayton copula, respectively.

Table 6.13. Comparison results of 7 copula models based on the CML pseudo $\{-\Lambda\Lambda\}$

row \ col	Frank	Plackett	Gumbel	Clayton	AMH	FGM
Normal	79/102	87/102	92/102	87/102	102/102	102/102
Frank		60/102	75/102	69/102	101/102	102/102
Plackett			76/102	72/102	100/102	102/102
Gumbel				55/102	99/102	102/102
Clayton					102/102	102/102
AMH						102/102

Table 6.14. Paired t-test for difference in the normal copula $\hat{\rho}$ between CML resulted and other approaches

Pair	Paired Differences Mean	Std. Dev.	t	df	p-value (2-tailed)
$\hat{\rho}_\tau$, $\hat{\rho}_{CML}$	0.0023	0.0140	1.694	101	0.093
$\hat{\rho}_{r_s}$, $\hat{\rho}_{CML}$	-0.0020	0.0178	-1.138	101	0.258
$\hat{\rho}_{LL}$, $\hat{\rho}_{CML}$	0.0013	0.0124	0.948	78	0.346
$\hat{\rho}_{Beta}$, $\hat{\rho}_{CML}$	0.0036	0.0131	2.676	94	0.009
$\hat{\rho}_{S_{BB}}$, $\hat{\rho}_{CML}$	0.0031	0.0126	2.484	101	0.015

Note: $\hat{\rho}_\tau$-matching sample Kendall's tau with the theoretical, $\hat{\rho}_{r_s}$-matching sample Spearman's rho with the theoretical, $\hat{\rho}_{CML}$-CML resulted, $\hat{\rho}_{LL}$-joint MLE with LL marginals, $\hat{\rho}_{Beta}$-joint MLE with Beta marginals, $\hat{\rho}_{S_{BB}}$-MLE with the S_{BB}.

6.5 Conclusions

Many statistical distribution functions have been used to describe diameter distributions in forest stands, including the Weibull, Beta, Johnson's S_B and Logit-Logistic. Then very naturally one quite interesting question arises that: is there such a method that we can use it to construct "promising" bivariate distributions with these given marginal distributions? The answer to this question would be much of interest to forestry researchers and practitioners, as in many cases we already have prior knowledge about the distributional model of the marginals. The copula approach provides such a general way of extending these univariate distributions into their bivariate domain.

By a theorem due to Sklar in 1959, a multivariate distribution can be represented in terms of its underlying marginals by binding them together using a copula function. A copula function then separates a bivariate distribution into its marginal component and the dependence structure between two variables. This separation allows for any specified marginals to be used for constructing a bivariate distribution with a copula function while the copula captures the "nonparametric", "distribution-free" or "scale-invariant" nature of the association between random variables. Using the normal copula, we constructed the bivariate (normal) Logit-Logistic model, thus extending its superior performance in the univariate domain to the bivariate domain. The normal copula may be regarded as normalization in the multivariate domain, while the Logit-Logistic may represent the empirical marginals much more adequately than the other univariate models we used.

We also compared other copulas (mainly of the Archimedean families) available in the statistical literature with the normal copula. Our results showed that the normal copula is the best one used for modelling the joint distribution of tree diameters and heights. As argued by Hafley and Schreuder (1976) that for a suitable bivariate distribution in modelling the joint distribution of tree diameters and heights, not only the marginals should be flexible for fitting the marginal frequencies of heights and diameters satisfactorily, but also this bivariate distribution provides a reasonable relationship between these two marginal variables. Based on this study we may tend to say that the normal copula approach is such a method of setting up such a suitable bivariate distribution, though we mainly put attention to some measures of dependence in this study while the relationship is characterized in terms of (median) regression in Hafley and Schreuder (1976) and thereafter. With the copula approach, the traditional mean regression would be too complicated but the median regression of Y to X is easily obtained since the conditional distribution of $Y|X=x$ is given as,

$$F_{Y|X}(y|x)=C_1[F(x),G(y)] \tag{6.33}$$

where $C_1(u,v)=\dfrac{\partial C(u,v)}{\partial u}$. For the normal copula, the median regression is given as,

$$\Phi^{-1}(G(y))=\rho\Phi^{-1}(F(x)) \tag{6.34}$$

which usually has no closed form expression, but Mardia (1970) gave an approximation by approximating the CDF of the normal with the Logistic as the following,

$$G(y)=\frac{\{F(x)\}^{\rho}}{\{F(x)\}^{\rho}+\{1-F(x)\}^{\rho}} \tag{6.35}$$

The median regression height-diameter model implied by the S_{BB} as in Schreuder and Hafley (1977) is thus a special case of equation (6.34), we may then expect that with more flexible marginal model, the resulting median regression model would be more adequate for describing the tree height-diameter relationship.

Our comparisons of different bivariate distributions, with the same copula but different marginals or with the same marginal but different copulas, were mainly based on the likelihood resulting directly from the MLE fitting the models, which is actually the AIC criterion. Other goodness-of-fit measures may be explored, especially when we use the two-step estimation in fitting for which we think further research on other methods of parameter estimation (LS method, say) would be worthwhile and in these cases the AIC may not apply.

Finally, both the bivariate normal copula and the bivariate Archimedean families can be extended into the multivariate cases. However, the primary advantage of the multivariate normal copula over the Archimedean families may lie that the normal copula permits the use of any positive-definite correlation matrix as the ordinary multivariate normal distribution but the class of Archimedean copulas is limited to intra-class correlation matrices.

Chapter 7: Spectral Refinement of Control Points for Co-Registration of Remotely Sensed Imagery

Summary

The traditional method of co-registering multi-temporal images to each other is to independently geometrically correct each image to a set of ground control points (GCPs). However, image-to-image co-registration is possible without the need for ground control points. Image co-registration control points (CPs) selected manually are subject to locational measurement errors. Spectral/intensity information is used to "refine" the estimate of the location of CPs for image-to-image registration. Three date Landsat Images (1987 Thematic Mapper (TM), 1997 TM, 2000 Enhanced Thematic Mapper (ETM)) covering a forest area in Northeastern China were used as a case study.

7.1 Introduction

Remotely sensed data contain both errors derived from the sensor instrument, and geometric representation errors. Most commercially available remote sensor data (e.g., TM images) already have much of the instrumental error removed (Jensen 1996). However, the non-instrumental error remains in the image, and therefore geometric correction usually has to be carried out to remove this error before actually analyzing remotely sensed (RS) data.

There are two common geometric correction procedures often used, image-to-map rectification and image-to-image registration. The difference between these two procedures is that in image-to-map rectification the reference is a map in a standard map projection, while in image-to-image registration the reference is another image. Both procedures normally involve selecting a number of control points (CPs). The selection of CPs is not an easy task when done manually, which is the norm.

In the image to image case, even for an experienced RS interpreter, the locations of selected CPs are always subject to some measurement error. For example, the interpreter may know that the selected CP is within a small area on the image, but not exactly which pixel.

As we know residual geometric error in a production map produce a "component" of the apparent classification error (Rennolls 2002), it is of utmost importance to reduce the rectification error as much as possible. The quality of the classification stage will then be enhanced, and the classification accuracy statistics will really refer to "pure" classification error rather than classification error due to pixel mismatch.

In our study, we are concerned with a sequence of images, and are interested primarily in estimating change and growth. For this, sub-pixel accuracy of image co-registration is desired (Coppin and Bauer 1994). In this chapter we go into some detail on the pixel-matching problem, and present a simple heuristic spectrally based pixel matching algorithm which seems to offer considerable scope for very accurate image-to-image co-registration.

7.2 Literature Review

There is a very full and technical history and literature in the area of registration and geometric correction. Two comprehensive surveys of image registration methods are referred to Brown (1992) and Zitová and Flusser (2003). The image registration normally consists of four steps, (1) control points extraction from images (2) control points matching (3) transformation model estimation using matched CPs, and (4) image re-sampling. With the first two steps completed, the last two steps are straightforward in that both can be done automatically with support of some commercially available packages (e.g. Erdas). Therefore, the first two steps are two key steps in the process of image registration, and not surprisingly, much research has involved in automating control points/feature selection as well as matching. The control points frequently used include line intersections (Stockman et al. 1982), centroids of closed contours or salient points of open contours (Li et al. 1995), centers of closed-boundary regions (Goshtasby et al. 1986, Ton and Jain 1989, Dowman and Dare 1999), and feature points detected from maxima of wavelet coefficients (Le Moigne et al. 2002).

The extracted CPs can then be matched to find the feature correspondence, through feature-based matching methods or the area-based (intensity-based) methods. Feature-based methods are to establish geometric correspondence between two sets of salient features in the reference and sensed images using their spatial relations or various descriptors of features after the features have been detected (Zitová and Flusser 2003). In contrast, area-based methods compare directly pixel intensity values in small image subsets/windows (Pratt 1974, Li et al. 1995), not necessarily to detect feature points before matching. There are many similarity metrics used in area-based methods, including correlation (Bernstein 1983, Pratt 1991, Brown 1992, Zitová and Flusser 2003), mutual information (Cole-Rhodes et al. 2003, Chen et al. 2003, Bardera et al. 2004), and sum of absolute or squared differences of the image intensity values (Zitová and Flusser 2003). Among them, the correlation metric is probably the most common used (Igbokwe 1999, Kenneday and Cohen 2003).

We note that area-based methods put emphasis on the CPs matching rather than on detecting them (Zitová and Flusser 2003), nonetheless it is the case of our study in which CPs are manually but roughly selected. Another point is that area-based methods may work with the whole images without requiring CPs, for example, Yao and Chern(2001) is particularly interesting in the way that combines the estimation of the registration transformation with the estimation of a shading function, with a robust weighted least squared method.

7.3 The Case Study Data

Three Landsat images were provided by the Chinese Academy of Forestry. They are 1987 Landsat5 TM, 1997 Landsat5 TM, Landsat7 ETM. All of them have been systematically corrected, geometrically and radiometrically. The 1987 and 1997 images are in the TM/Krasovsky projection with pixel size of 35m, and the 2000 image in UTM/WGS 84 with pixel size of 33m.

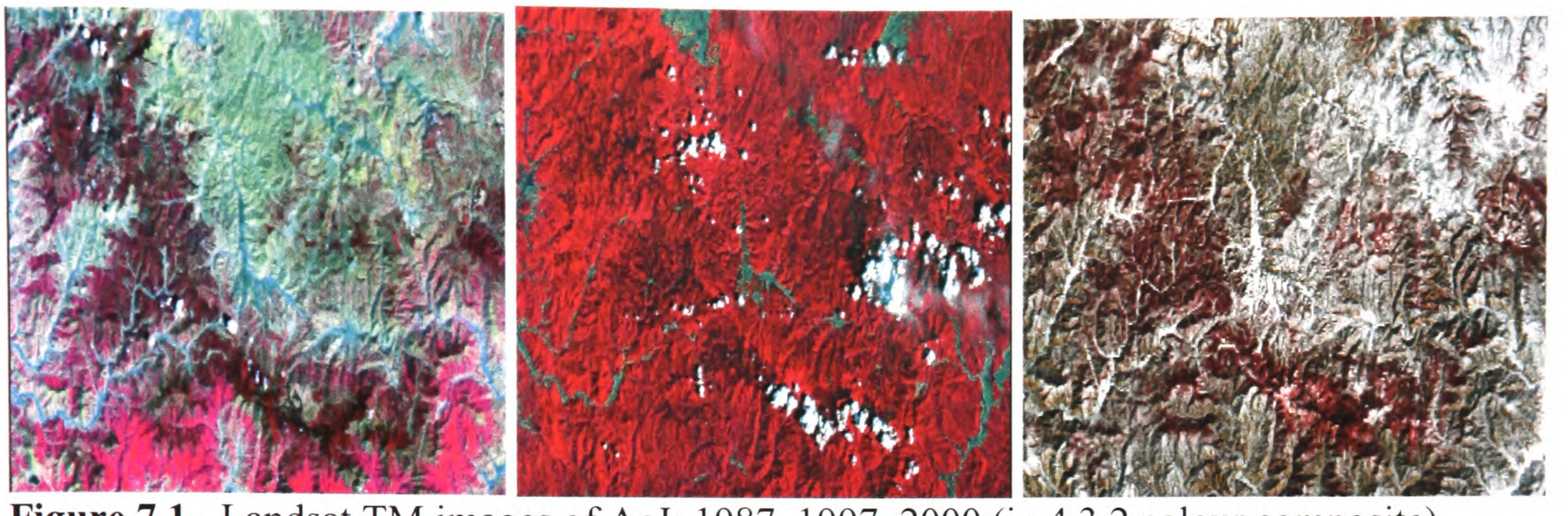


Figure 7.1. Landsat TM images of AoI: 1987, 1997, 2000 (in 4,3,2 colour composite)

7.4 Co-registration: Image to Image Matching

7.4.1 Stage 1 Matching

If the matching of two or more images is to be done completely automatically, the problem is "hard" (Keysers and Unger 2003). The rectification/geometric-correction approach to matching is usually "spatially" based. Control points have to be selected on each of the images and their coordinates determined as accurately as possible. If this determination is without error then a rubber-sheet transformation, such as a thin plate spline interpolates the match across the whole image. In such an approach the root-mean-square (RMS) error is zero and the matching accuracy (or equivalently the matching model-error) has to be determined from a validation set of control points that have been reserved from the rectification process.

Ironically this image matching is largely trivial for the human eye-brain complex. From observation of the images in Figure 7.1, we immediately "see" the lower "trident foot" of the common "snakes" in our images, whether they be ridges, valleys or roads. Our mind's-eye immediately (i) matches the images just with a handful of corresponding features, and mentally (ii) superimposes the intervening regions on the images.

We bypass the hard problem of whole image matching, and follow the human eye-mind heuristic through two phases, as indicated above. We use the (actual) eye to select a ***"couple"*** of corresponding control points, and ***roughly*** determine their coordinates in their own image frames.

7.4.2 Selecting Control Points

For the 1987 and 1997 TM images, 41 CPs were selected manually. They are mainly road or river intersection points, which are relatively easily recognized in images. Suppose each of these points is within 5 pixels area with itself as the center. That is, we know for sure this CP is within a 11×11 square lattice of pixels, but are not quite sure which cell/pixel it is in.

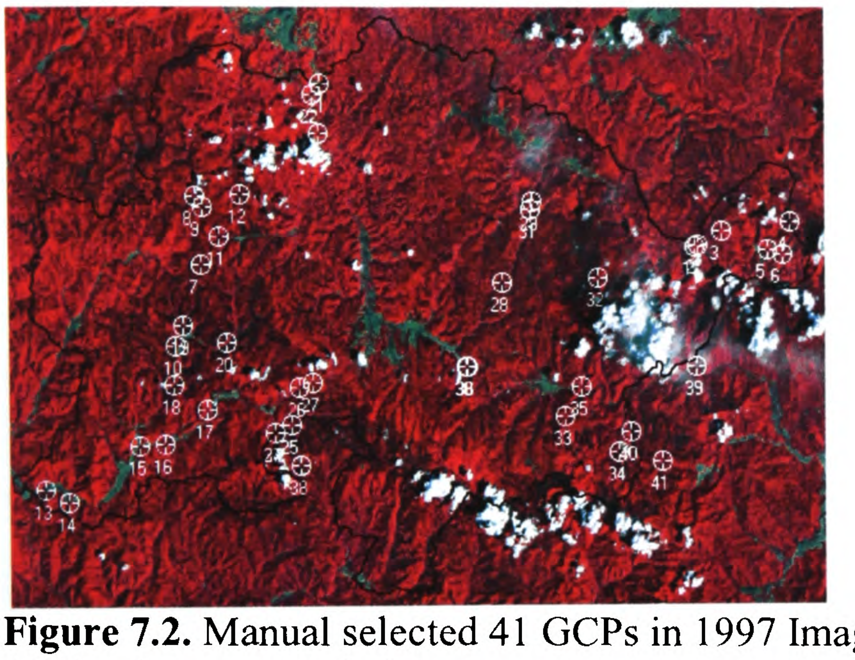


Figure 7.2. Manual selected 41 GCPs in 1997 Image for 1987-1997 image co-registration (in 4,3,2 colour composite)
Note: Line in black is Bureau boundary

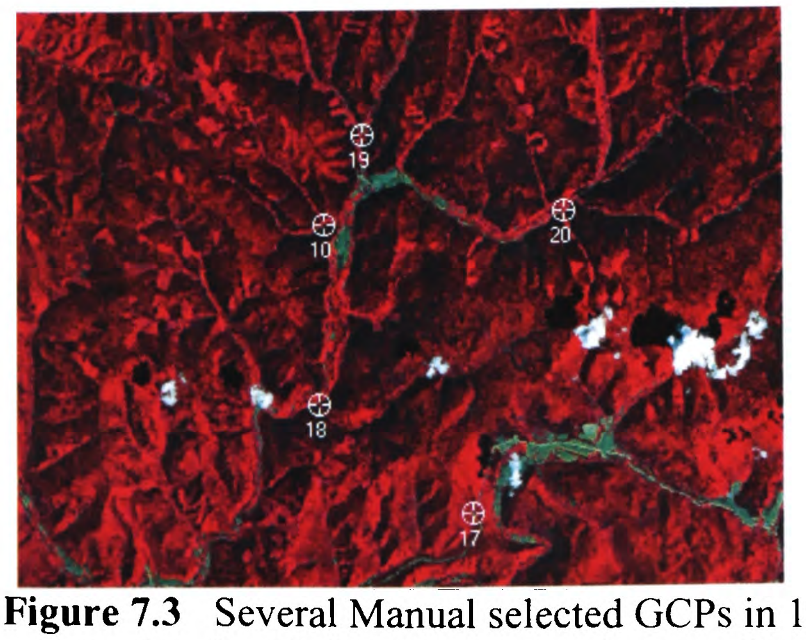


Figure 7.3 Several Manual selected GCPs in 1997 Image for 1987-1997 image co-registration (in 4,3,2 colour composite), road/road and road/river intersections

7.4.3 The Perspective Transformation

Various transformation models have been used for geometric correction, including affine (linear) transformation, bilinear transformation, perspective transformation, radial basis functions, and so on. For details see Fogel and Tinney (1996) and Glasbey and Mardia (1998). In this study, we use the perspective transformation given as,

$$u = \frac{a_0 + a_1 x + a_2 y}{1 + c_1 x + c_2 y} \qquad v = \frac{b_0 + b_1 x + b_2 y}{1 + c_1 x + c_2 y} \tag{7.1}$$

where (u, v) are 1997 image coordinates, (x, y) 1987 or 2000 image coordinates.

7.4.3.1 The 1987-1997 Perspective Transformation

The 41 control points are used to fit the perspective transformation model using least squares method. Table 7.1 lists the parameter estimates and Figure 7.4 shows the residuals diagrams.

Table 7.1. Perspective transformation 1987→1997, on 41 "rough" points

a_0=	10882.83622
a_1=	0.975338626
a_2=	-0.002058339
b_0=	112205.3522
b_1=	-0.005667355
b_2=	0.953838592
c_1=	-1.287E-09
c_2=	-4.749E-09

From Figure 7.4, there are about 7 points whose registration errors are bigger than one pixel size, 35 meters in this case, in the x-direction or y-direction. Usually we may remove these points with large errors and refit the transformation model (Jensen 1996).

Note that c_1 and c_2 are essentially zero, so the perspective form of the transformation is not necessary. Affine transformation is sufficient. Note also that for the affine transformation, we essentially have $a_1 = 1$ and $a_2 = 0$. Similarly we have b_1=0 and $b_2 = 1$. Hence the (x, y) and (u, v) may already have the same orientation (no rotation is needed) and the same scale, since these two 1987 and 1997 images are in the same coordinate system (TM/Krasovsky) with the same remote sensor and for the same area (Xie et al. 2003). **We should have known this!** Under this situation, the transformation involved in mapping the 1987 image to the 1997 reference reduces to a simple translation of the origin, given by the (a_0, b_0) vector, or in other words the pixels coincide perfectly. However, with the control points manually selected, such information has been hidden, see Figure 7.5.

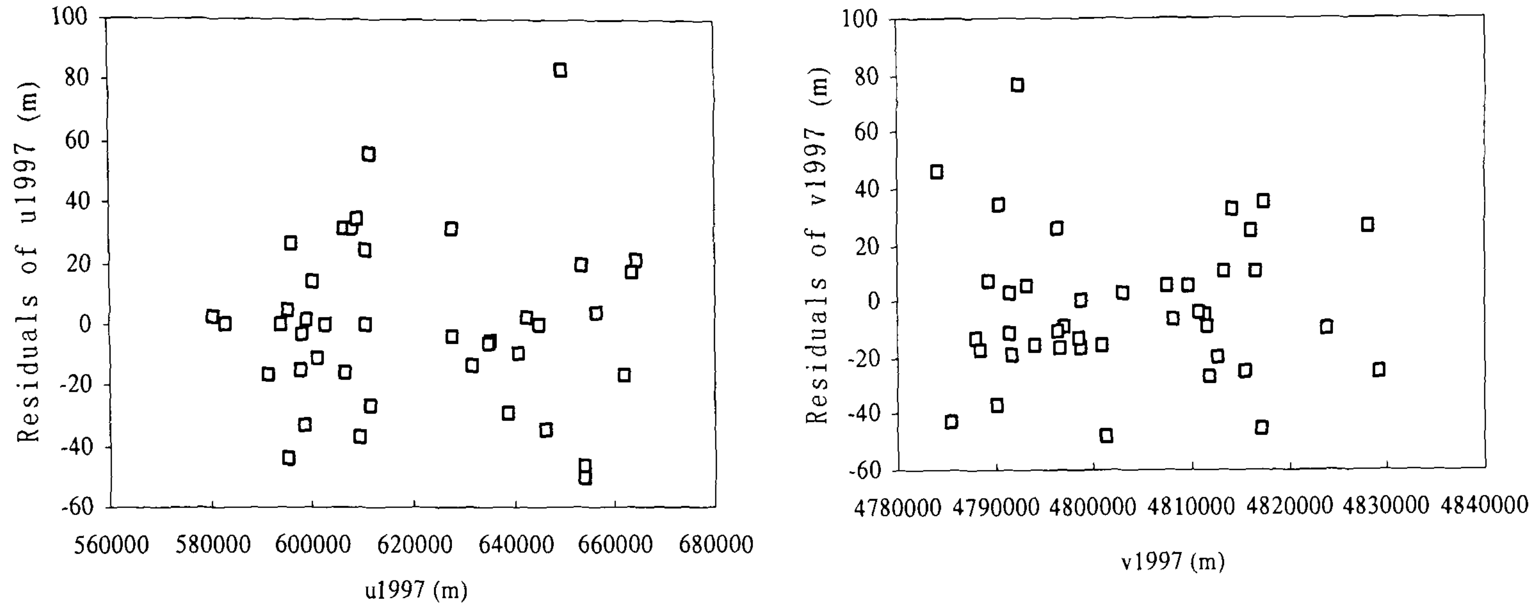


Figure 7.4. Residual plots for the 1987→1997 perspective transformation on 41 "rough" points

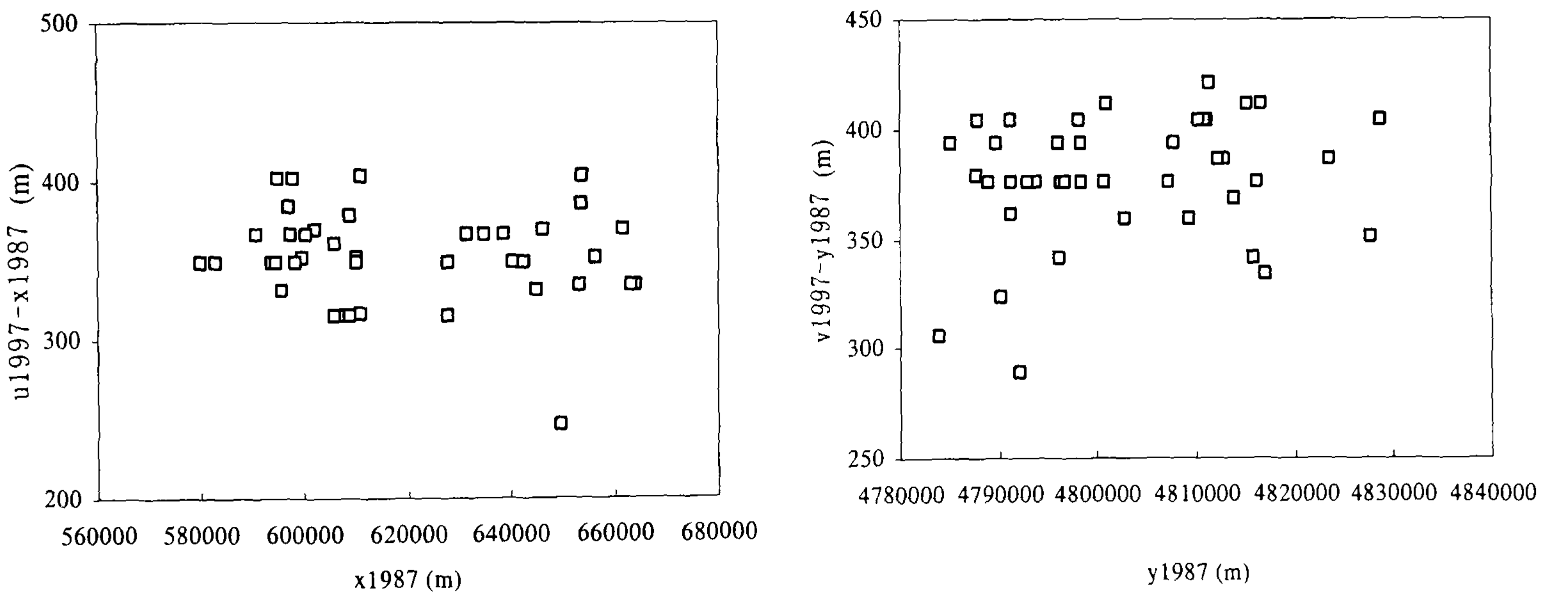


Figure 7.5. Shift differences for 1987→1997 images on 41 "rough" points

7.4.4 Spectral Pixel Matching for Refinement of Image-to Image Co-Registration

Our spectral refinement method also uses intensity information with the correlation coefficient (CC) as similarity metric to refine CPs. The difference between our refinement method and the automatic image registration lies in the fact that we have Stage 1 image to image registration already established and are just seeking to refine the accuracy of this pixel to pixel registration. Figure 6 shows the methodology.

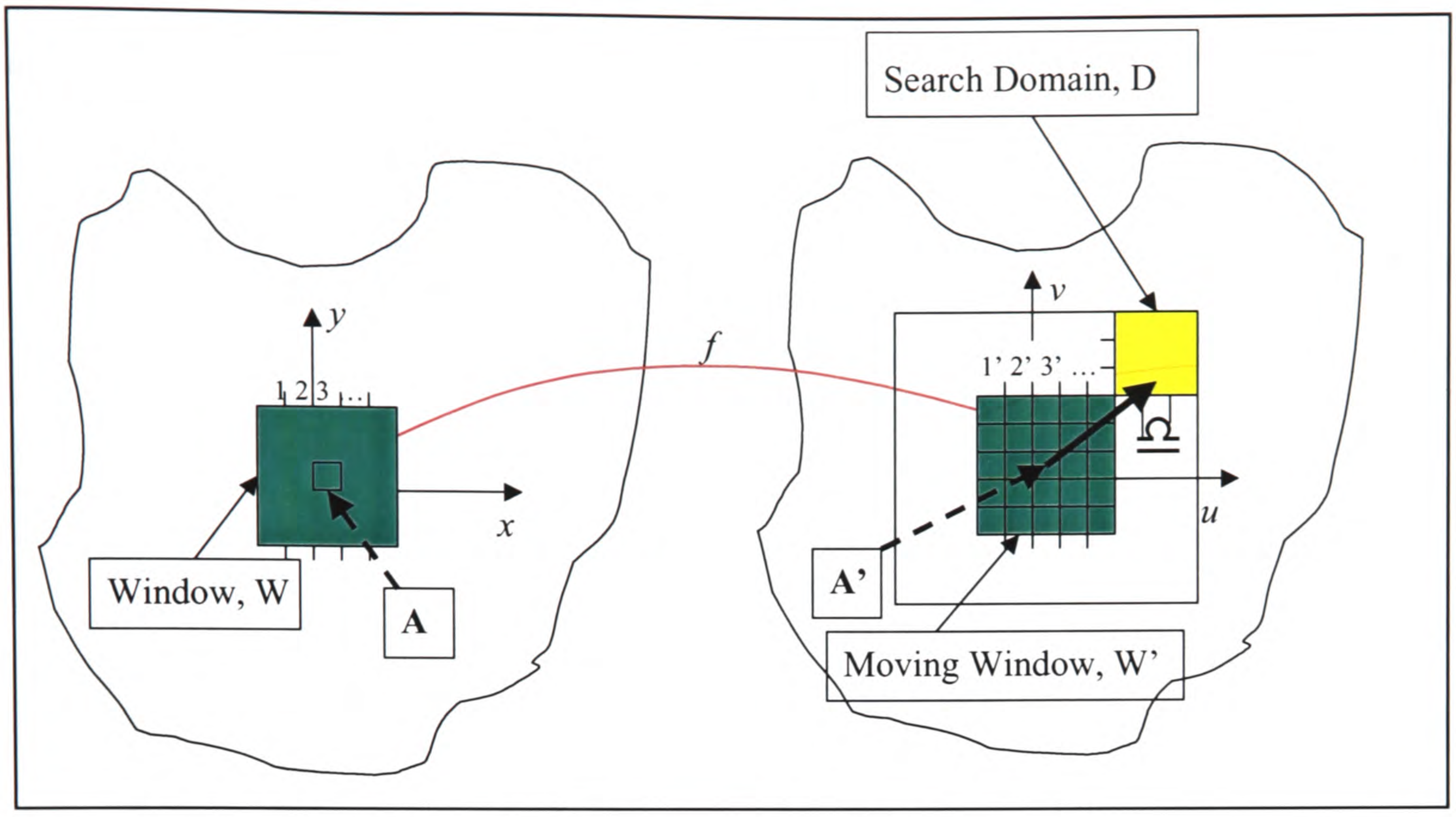


Figure 7.6. Spectral refinement of pixel-to -pixel matching

Let A be a pixel in image 1. Let A'=*f*(A) be the corresponding pixel in image 2, under the Stage 1 image-to-image co-registration transformation *f*. Let $\delta = (\delta_u , \delta_v)$ be an off-set from A', where δ_u and δ_v are the off-sets from A' in pixels in the directions u and v, the directions of the coordinate axes of image 2. Denote the pixel in image 2 which is δ from A' by B(δ). If δ is small, then B(δ) is a pixel corresponding to the neighbourhood of the image of A, in image 2. We define a square neighbourhood of pixels around A' which we call the Search Domain, Δ say. We are interested in how "similar" the environments of A and B(δ) are. We define an equal size square Sensing Window (of size m×m) around A and B(δ), in images 1 and 2 respectively, denoted by W and W'(δ) respectively. In each of these two sensing windows we label the pixels in the same sequence; 1, ... , M (=m×m) (relative to the reference axes in each image). We then use the correlation between the Band-I intensities in the pixel-sequences (1, ..., M) in each image as the measure of local similarity, (I= 1, 2, ...7).

These measures may be regarded as cross-image (structured) spatial correlation functions. The "structured" term is used to distinguish from the more often used cross-($|\delta|$-dependent)-correlation function.

We search over δ in $\Delta(A')$ to find the δ such that $Corr(W_I(A), W'(B(\delta))$ is maximum. We take this point $B(\delta)$, in Image 2, to be the "spectrally refined" matching point of A in Image 1. This is an intuitively obvious local correlation search algorithm, which we imagine is similar to the usual algorithms for matching whole images using correlation search. However, we have not seen this local form of search for refined registration defined in the literature.

We have tried using some other spectral distance measures, but to date the correlation measure seems best. When calculating these cross-correlational measures the search window size, we have evaluated the use of 3x3, 5x5, 7x7, 9x9, and 11x11 sensing windows. Since we found that decreasing size of sensing windows decreased the number of "Good Match Point" as defined in the following section, we report only on the use of an 11x11 search domain (that is, $\delta_u = \delta_v = 5$), and an 11x11 sensing window, (the sensing window may extend beyond the search window).

Figure 7.7 shows the calculated correlation table for the 11x11 search domain for CP1 using the band 1 spectral information, between the 1987 and 1997 images. The A' pixel is the centre, where the blue row and column intersect. We see the two highest values of the local correlation measure are to the immediate right of A'. From this Figure, we would conclude there is **(about)** one pixel error in the original "eye" matching of CP1. Figure 7.8 shows a contour map of the local correlation structure constructed using S-Plus.

Figure 7.8 suggests that by fitting a suitable response surface model to this correlational map, it would be possible to estimate the pixel-to-pixel match with considerable sub-pixel resolution. This is indeed reassuring, since one of the main aims of searching for accurate pixel-to-pixel spectral registration improvement was to obtain sub-pixel accuracy across most of the image to image matching exercise, so that the subsequent phases, of classification and growth and change estimation could be dissociated from confounded registration error.

	1	2	3	4	5	6	7	8	9	10	11
1	-0.18	-0.2	-0.18	-0.15	-0.13	-0.1	-0.13	-0.13	-0.12	-0.14	-0.13
2	-0.1	-0.12	-0.09	-0.13	-0.12	-0.09	-0.11	-0.1	-0.1	-0.09	-0.01
3	0.09	0.06	-0.01	-0.07	-0.09	-0.06	-0.03	-0.03	0	0.03	0.11
4	0.24	0.22	0.18	0.17	0.14	0.07	0.1	0.16	0.19	0.21	0.25
5	0.32	0.32	0.38	0.44	0.46	0.49	0.49	0.43	0.41	0.39	0.3
6	0.2	0.18	0.25	0.41	0.48	0.6	0.71	0.7	0.52	0.21	-0.03
7	-0.06	-0.03	0.12	0.24	0.27	0.23	0.18	0.15	0.11	0.06	-0.03
8	-0.13	-0.13	-0.07	-0.08	-0.12	-0.11	-0.13	-0.15	-0.15	-0.14	0.02
9	-0.24	-0.16	-0.21	-0.26	-0.33	-0.34	-0.21	-0.15	-0.14	-0.12	-0.1
10	-0.19	-0.25	-0.21	-0.18	-0.16	-0.17	-0.21	-0.19	-0.16	-0.08	-0.05
11	-0.06	-0.13	-0.18	-0.16	-0.04	0	-0.02	-0.11	-0.19	-0.14	-0.02

Figure 7.7. 11×11 Correlation Map for Control Point 1 (Band 1) using a 11×11 search window

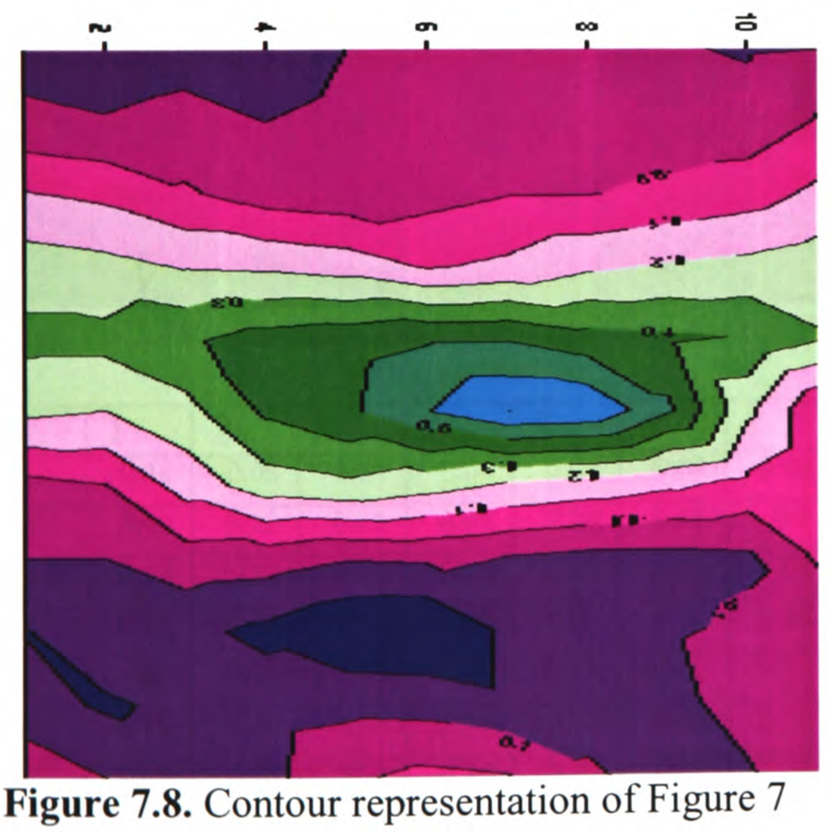


Figure 7.8. Contour representation of Figure 7

In contrast, Figure 7.9 shows that the thermal band 6 does not perform well in this approach, and should not be used.

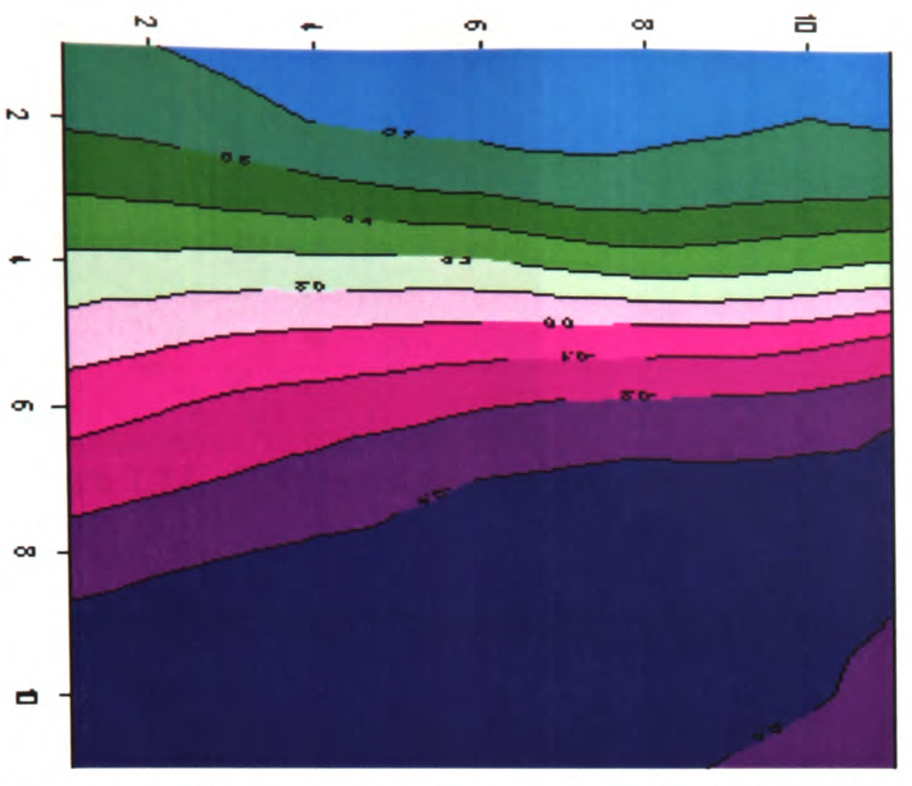


Figure 7.9. Correlation Contour Map for Control Point 1 (Band 6/Thermal)

Figure 7.10 shows the correlational maps for each of the bands (excluding the thermal band 6) for CP1 in the 1987-1997 matching exercise. We see that we have substantial mutual support in each of the bands for determination of the improved pixel to pixel match.

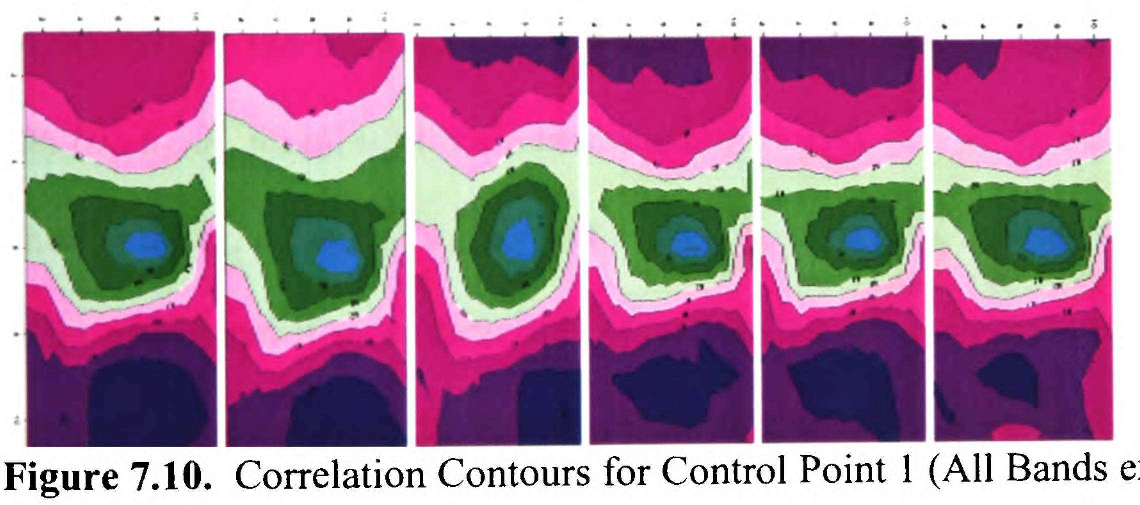


Figure 7.10. Correlation Contours for Control Point 1 (All Bands excluding Thermal)

	1	2	3	4	5	6	7	8	9	10	11
1	0.14	0.06	0.01	0.04	0.04	0.09	0.08	0.06	-0.03	-0.06	-0.11
2	0.14	0.13	0.01	0.06	0.09	0.08	0.1	0.09	-0.01	-0.09	-0.13
3	0.19	0.2	0.13	0.06	0.14	0.11	0.09	0.15	-0.06	-0.17	-0.22
4	0.25	0.19	0.15	0.06	0.09	0.11	0.16	0.1	0	-0.11	-0.18
5	0.3	0.32	0.29	0.23	0.23	0.36	0.32	0.14	0.12	-0.07	-0.12
6	0.36	0.41	0.36	0.38	0.33	0.36	0.37	0.23	0.03	-0.11	-0.14
7	0.18	0.16	0.2	0.25	0.25	0.24	0.34	0.27	0.05	-0.06	-0.11
8	0.1	0.05	0.12	0.11	0.06	0.01	0.08	0.23	0.07	-0.12	-0.08
9	0.12	0.07	0.02	0.08	0.01	-0.02	-0.03	0.19	0	-0.14	-0.12
10	0.05	0.08	0.04	0.06	0.03	-0.02	-0.01	0.06	0.02	-0.06	-0.18
11	0.02	0.06	0.1	0.13	0.08	-0.01	0.02	0.12	0.05	-0.05	-0.11

Figure 7.11. 11×11 Correlation Map for Control Point 4 (Band 1) using a 11×11 search window

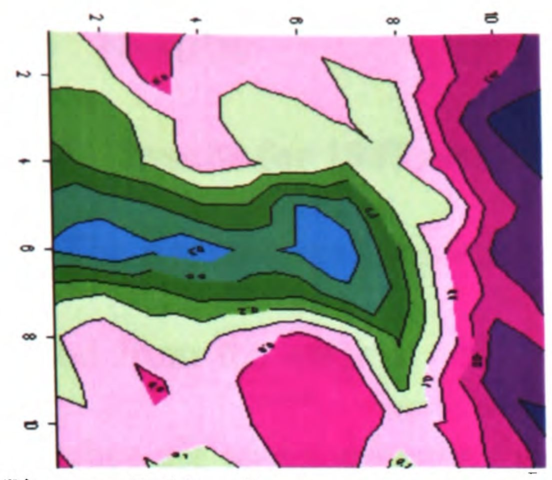


Figure 7.12. Contour representation of Figure 11

Figures 7.11 and 7.12 show the correlational map for a pixel-pixel mach which cannot be well defined since the maximum of the correlation measure is not well defined in the Search Window. There is a correlational ridge along the negative u-axis.

7.4.5 A Multi-Spectral Pixel Matching Criterion

Let (x, y) be one image coordinates, (u, v) the other. For each band of 6 bands (b1, b2, b3, b4, b5, b7) of TM images, we got 6 matching results using CC as similarity measure for each CP: $(u_1, u_2, u_3, u_4, u_5, u_7)$ for u corresponding to x, $(v_1, v_2, v_3, v_4, v_5, v_7)$ for v corresponding to y. Then with $(u_1, u_2, u_3, u_4, u_5, u_7)$ and $(v_1, v_2, v_3, v_4, v_5, v_7)$, how to get the refined u and v? We adopt a rule that if more than half of $(u_1,u_2,u_3,u_4,u_5,u_7)$ and more than half of $(v_1,v_2,v_3,v_4,v_5,v_7)$ are the of same numerical values, respectively, then the same "u" and "v" are taken as the refined results for u and v, expressed as (u_r, v_r). If not, we may think the refinement failed.

Moreover, the "round trip refinement" we called here as Ton and Jain (1989) termed the "two-way matching" was carried out. That is, we refine (x, y) from (u_r, v_r) following the same procedure for refining (u, v). Suppose (x_r, y_r) is the refinement, we have,

$$(x, y) \rightarrow (u_r, v_r) \rightarrow (x_r, y_r)$$

If $x = x_r$, and $y = y_r$, the paired CPs are very stable and defined as a "Good Match Point" (GMP).

7.5 Results

7.5.1 Results for 1987 and 1997 Images

The rigorous multi-spectral criterion defined in the last section resulted in 20 GMPs being found from the original 41 "rough" point-pairs for the (87↔97) image matching. These matches are so accurate they should make use of a rubber-sheeting representation, such as a thin plate spline, to represent and interpolate them. However, for convenience in comparing with the previous perspective transformation of the 41 original rough control points the fit of a similar transformation is given in Table 7.2. Figure 7.13 shows the residuals.

Table 7.2 Parameter estimates for the perspective transformation on the 20 GMPs(87↔97)

a_0=	-6649.383485
a_1=	1.005822864
a_2=	0.002181113
b_0=	-53477.76408
b_1=	-0.043547679
b_2=	1.028246821
c_1=	-9.062E-09
c_2=	3.545E-09

From Figure 7.13, registration precision is much improved in that all the residuals are less than one pixel size (35 meters), the maximum error being about 25m, in both the x-direction and y-direction. However, both scatter plots show some error trends. Check of parameter estimates from Table 7.2 indicates that as similar to our findings from Table 7.1, we essentially have $a_1 = 1$ and $a_2 = 0$, and $b_1 = 0$ and $b_2 = 1$. Therefore, the transformation involved in mapping the 1987 image to the 1997 reference reduces to a simple translation, and the error trends may be due to some outliers of the control points. Figure 7.14 shows the translation differences for the 20 GMPs (in x and y directions).

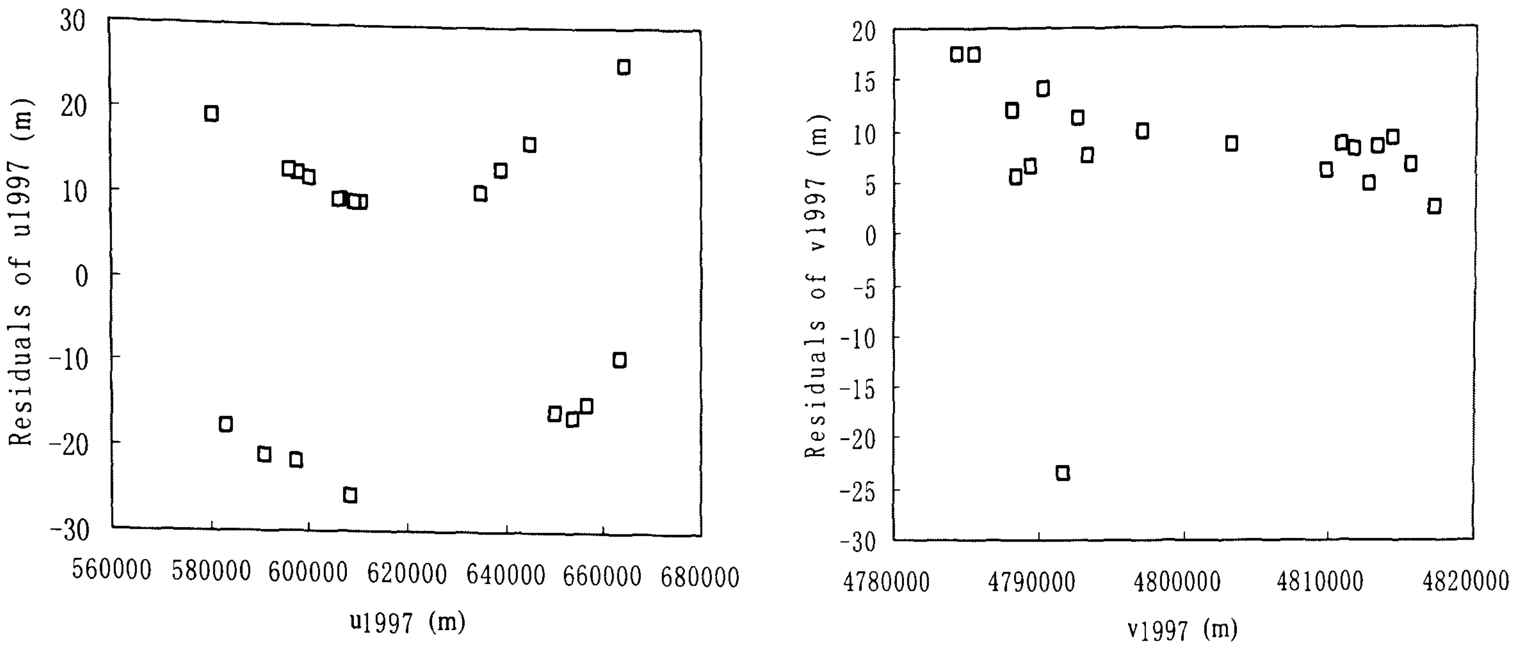


Figure 7.13. Residual plots for the 1987↔1997 perspective transformation on 20 good match points

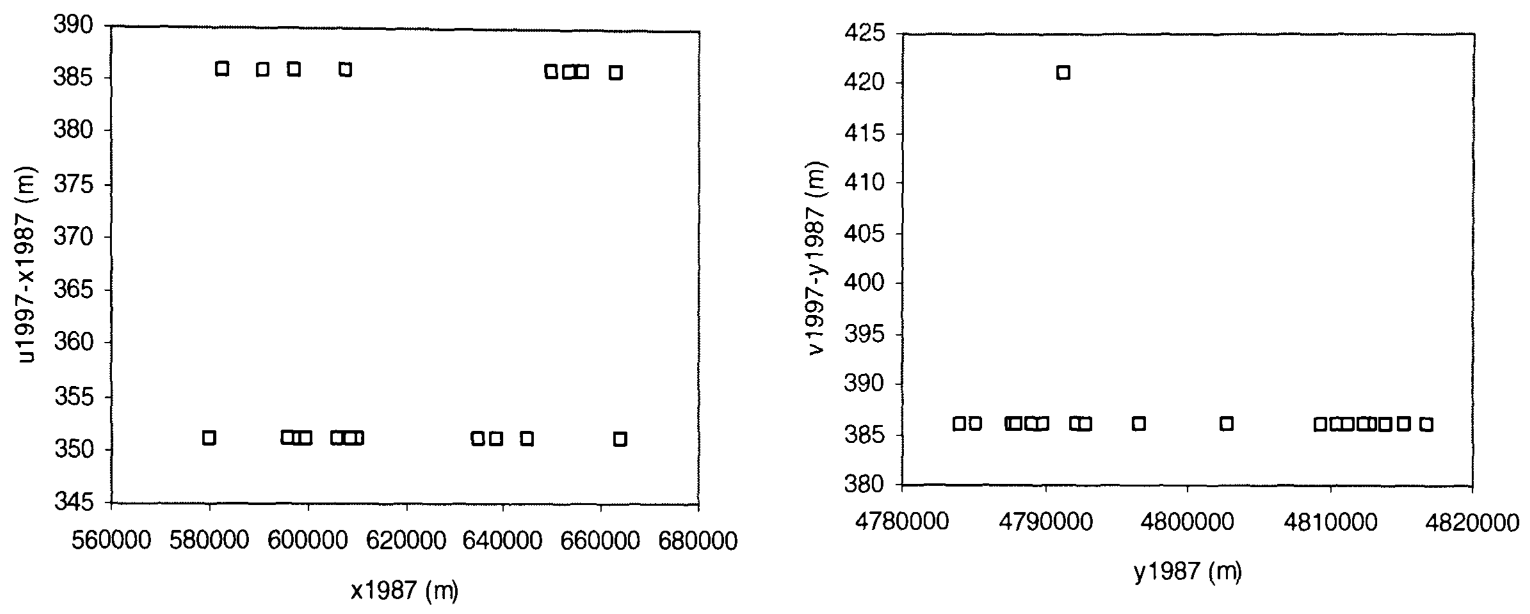


Figure 7.14. Shift differences for 1987↔1997 images on 20 good match points

The translation (shift) differences in meter as shown in Figure 7.14 can be transformed into pixel units. In terms of the *x*-pixel difference, 12 CPs have a particular pixel count difference, and the other eight have an extra pixel difference. This indicates that the pixels do not in general "line-up" in terms of the *x*-*u* axes. The results would support a sub-pixel analysis. However, in terms of the pixel count differences in the *y*-direction the vast majority do have a fixed difference, with just 1 out-lying points (assumed erroneous). So it does look as if the *y* and *v* axes do "line-up" to an exact pixel translation and therefore pixels match up exactly in this direction.

7.5.2 Results for 2000 and 1997 Images

Similarly, for the 2000 and 1997 images, we roughly select 94 control points and obtain 14 GMPs. Since the two images are of different scales (pixel size of 33m and 35m, respectively), we will not expect the simple translation result. However, the perspective transformation is found to collapse to the affine transformation as in the case of 1987 and 1997 images. Therefore, the affine transformation is used. Figure 7.14 and 7.15 show the residuals before and after refining, and the contrast shows the scale of residuals has been decreased significantly. In fact, the RMS decreased from about 36m to about 10m (0.3 pixel) and from about 43m to about 14m (0.4 pixel) in x-direction and y-direction, respectively.

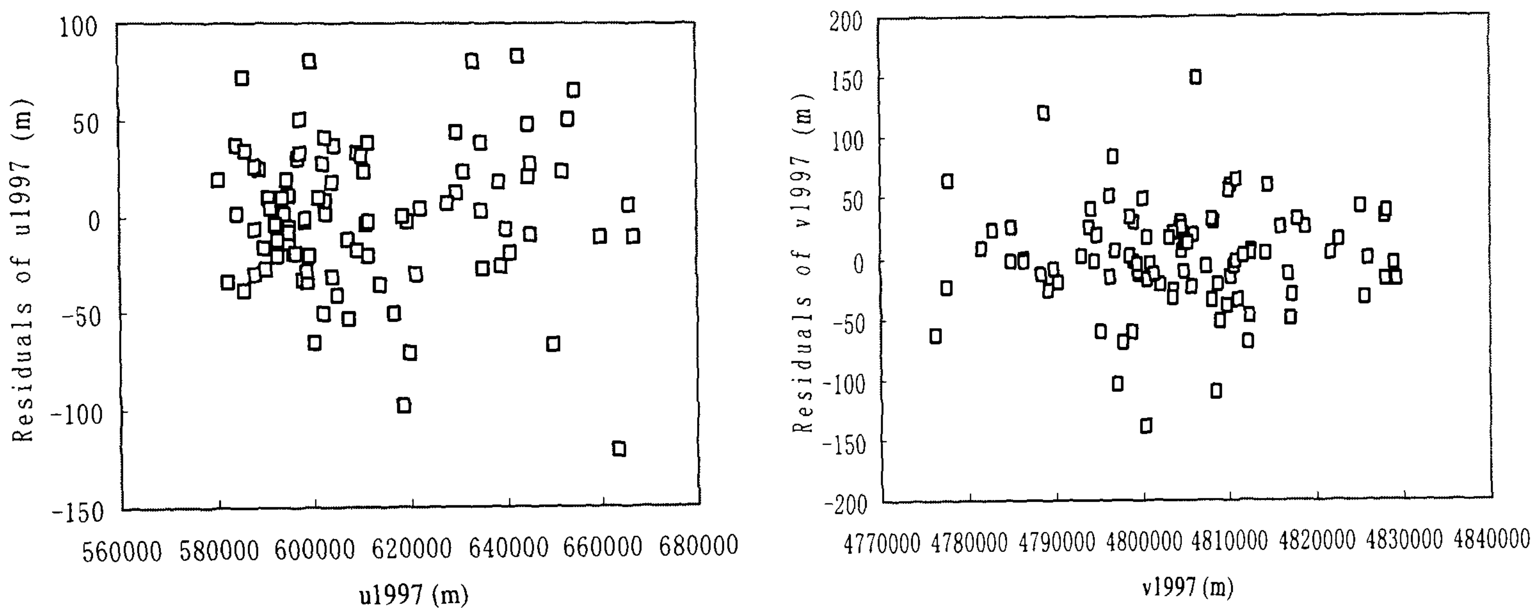


Figure 7.15. Residual plots for the 2000→1997 affine transformation on 94 "rough" points

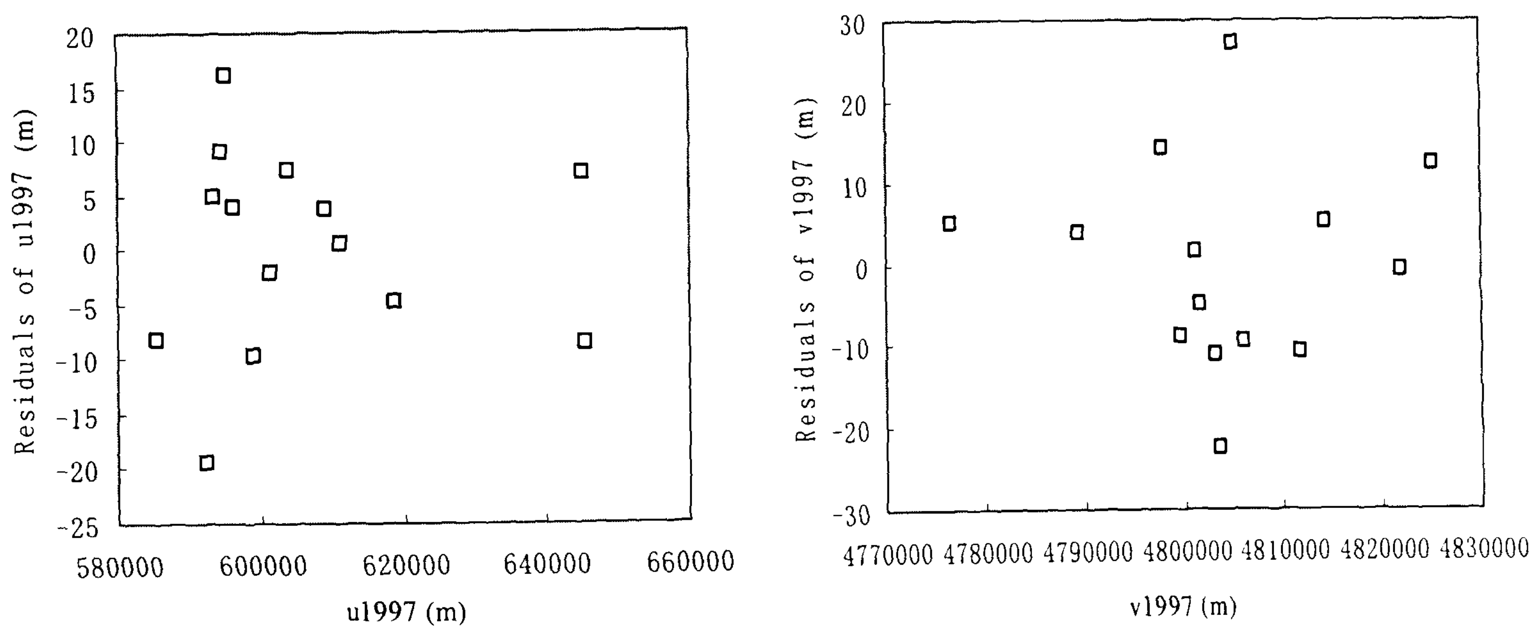


Figure 7.16. Residual plots for the 2000↔1997 affine transformation on 14 good match points

Chapter 8: Radiometric Correction and Spectral Standardization

…"Spectral Evolution Model" and

…"Multi-temporal Classification"…?

Summary

Radiometric correction is another pre-processing procedure which must be completed prior to using multi-temporal remotely sensed images for change detection and growth modelling. Truncation and rank-statistic based spectral standardization methods are used to achieve robust relative radiometric rectification on the case-study TM images.

8.1 Introduction

This chapter has a title and a tentative subtitle. The title refers to some preliminary work and ideas which lead on to the more tentative topics in the subtitle. These later topics should be taken as possible areas for future research.

Multi-temporal remotely sensed images have been used for change detection in forestry (Coppin and Bauer 1994, 1996, Collins and Woodcock 1996, Tokola et al. 1999, Wilson and Sader 2002). The basic idea behind change detection is that a difference exists in the spectral response of a pixel on two or multiple dates if the biophysical materials within the same area have changed (Jensen 1996). Therefore, when performing change detection, it is desirable to eliminate the radiometric noise caused by satellite sensors, environmental factors (atmospheric conditions, solar angles, etc.) and the phenological disparities (different growth seasons) as much as possible.

Radiometric correction of remote sensed images generally falls into two broad categories; absolute and relative. Absolute correction is usually not applicable to multi-temporal images, since the required atmospheric measurements at the time of data acquisition are very difficult to obtain. Relative correction is then generally to be recommended in order to radiometrically/spectrally normalize the images to each other (Song et al. 2001).

The most widely used relative method is probably the one developed by Hall et al. (1991), which involves in selecting pseudo-invariant features (also known as radiometric control points, compared with geometric control points for geometric correction/registration) which are expected to be spectrally constant from image to image and then applying a linear transformation (obtained by regression) to adjust spectral values between images. However, the typical pseudo-invariant features like lakes, concrete, and gravel which are commonly used (Coppin and Bauer 1994, Elvidge et al. 1995, Yuan and Elvidge 1996, Yang and Lo 2000), may not be easily obtained especially in mainly forest occupied area. In the other hand, for change detection in forest stands, Olsson (1993) argues that the forest pixels themselves should be used as the spectral stable targets for radiometric correction.

In making an attempt at integrated classification and growth modelling on multi-temporal imagery, Joyce and Olsen (1999) suggested a "spectral standardization" based on a symmetric radiometric correction method between successive images (using means and standard deviations). In this Chapter, we apply a similar method using medians and inter-quartile ranges.

8.2 The Case Study Data

We selected a subset of our study area, the 10th forest farm, for this study, since the size of total study area is quite large. Figure 8.1 shows the three images. It is seen that the spectral difference in 1987 and 1997 is probably due to different growth seasons taken for data acquisition (May and July for 1987 and 1997, respectively), while the difference between 1987 and 2000 both taken in May is mainly due to the different sensors, that is, TM and ETM.

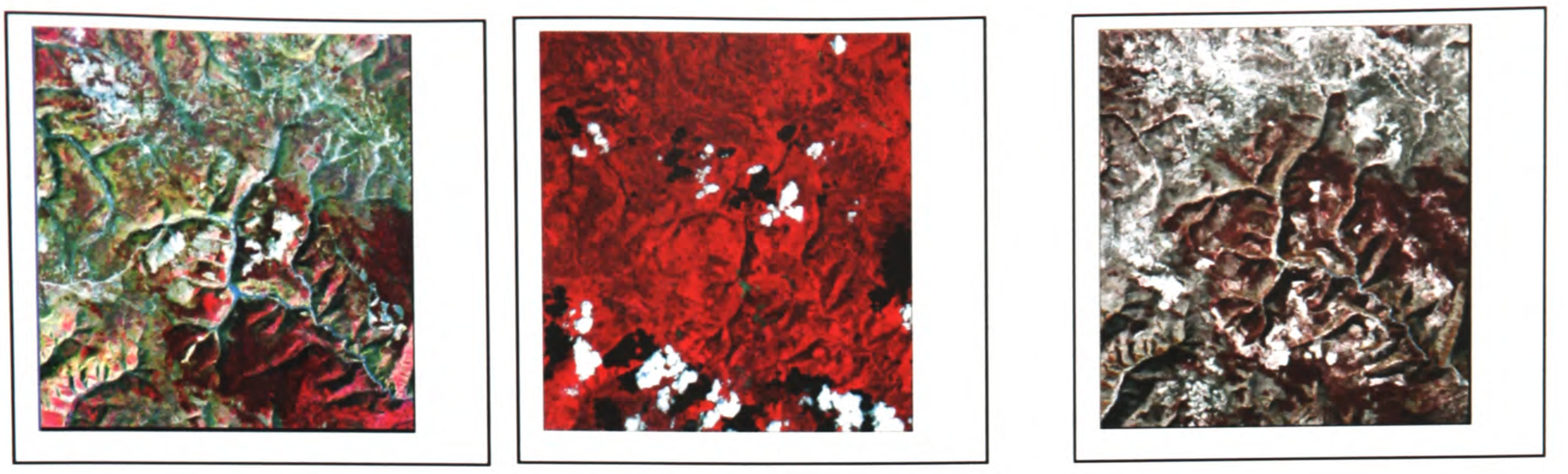


Figure 8.1. Landsat TM subset images of 10th Farm: 1987, 1997, 2000

8.3 Spectral Standardization

In Figure 8.2 are shown the frequency disributions over the possible intensity range, for Bands 1 (visible blue), 4 (near infrared) and 5 (middle infrared) for the three subset images. For Bands 1 and 4, for the 1987 and 1987 image distributions it is easy to see that the distributions have shapes that can be morphed into each other by a suitable scaling transfomration, in the same "direction" for both of these bands.

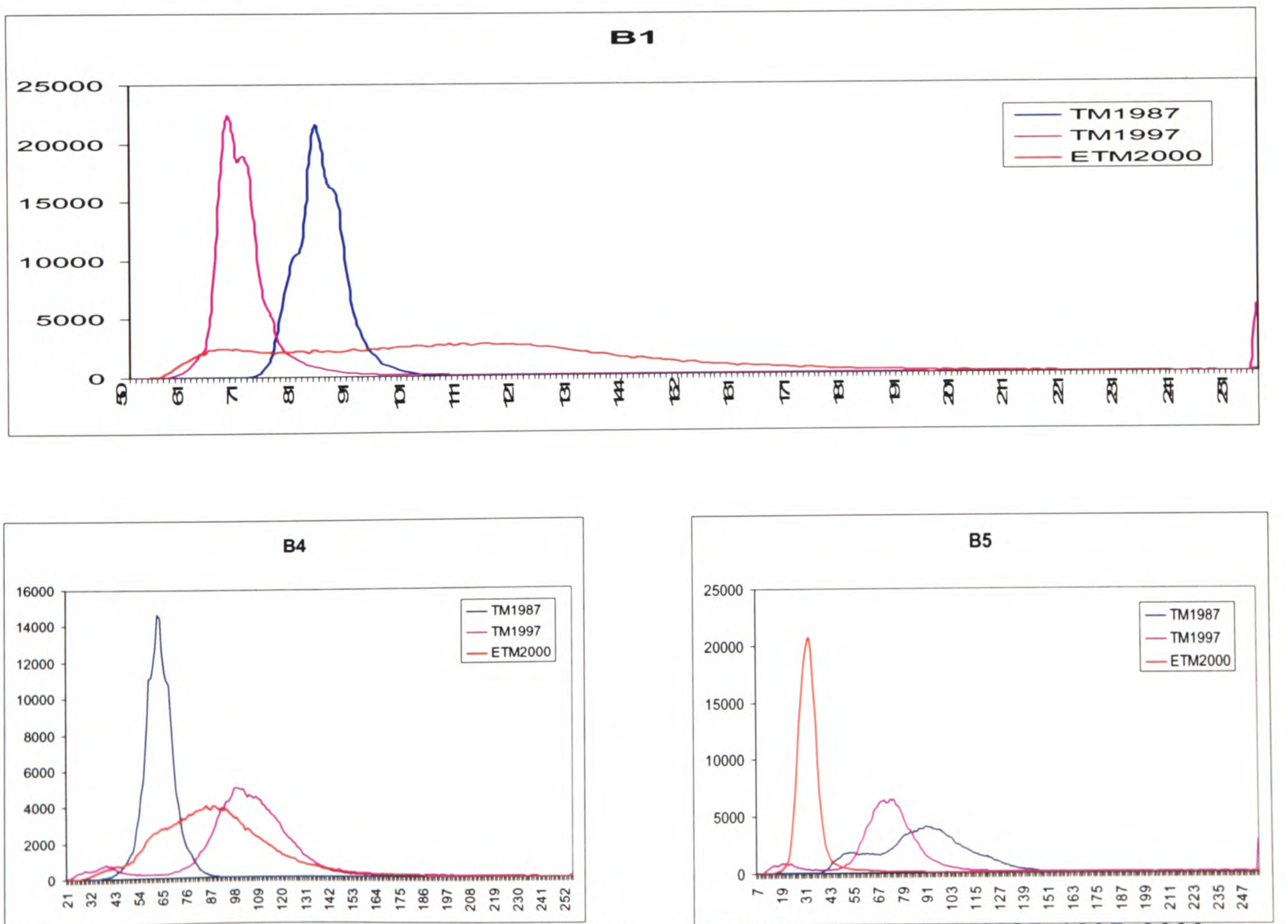


Figure 8.2. Intensity distribution for farm 10, Landsata TM Band 1, 4, 5; 1987, 1997, 2000

Joyce and Olsen (2000) used a symmetric and transitive model for scaling between two such spectral intensity distributions, that is,

$$\frac{x_2 - \bar{x}_2}{x_1 - \bar{x}_1} = \frac{s.d.(x_2)}{s.d.(x_1)} \qquad (8.1)$$

where x_i , $\bar{x}_i$ and $s.d.(x_i)$ are spectral intensity values, the mean and standard deviation of the intensity distribution for a specific spectral band of the i^{th} image. Then if we scale the first image (x_1) to the unit of the second (x_2), we got,

$$x_1' = (x_1 - \bar{x}_1)\frac{s.d.(x_2)}{s.d.(x_1)} + \bar{x}_2 \qquad (8.2)$$

Such a scaling transformation assumes that there is no "significant" change in the spectral shape/signature of the forest mask of the interested image; rather inappropriate if one is actually trying to estimate change or growth from the information in the spectral signatures.

Notice there is a spike at the maximal intensity for the 1997 image for bands 1 and 5. Furhter check indicates that the spike corresponds to the cloud/shadow pixels as shown in Figure 8.1. We may mask out the cloud/shadow pixels before we carry out our radiometric correction. Alternatively, use of the robust median and interquartile range would be a better choice.

$$x_1' = (x_1 - m_1)\frac{i.q.(x_2)}{i.q.(x_1)} + m_2 \qquad (8.3)$$

where m_i and $i.q.(x_i)$ are the median and interquartile range of the intensity distribution for a specific spectral band of the i^{th} image.

Taking band 1 for example, Figure 8.3 shows intensity distributions after the median-interquartile range scaling of 1987 and 2000 images to the 1997 reference, but only over the range between the 5 percentile and the 95 percentile for the purpose of removing noises such as clouds/shadows. From Figure 8.3, the three spectral distributions are all multi-modal Mixture distribution models, with each mixture componenet having its own mode, would seem to be appropriate. The TM87 curve has a clear dominant mode just over 70, with bossible sub-modes at about 73 (upper sub-mode) and 67 (lower sub-mode) and and possibly another sub-dode at 78. The TM97 curve in Figure 8.3 has a dominant mode at about 68, 2 down from the dominant mode of TM87. The upper sub-mode in TM97 is about 71, also 2 units down from the upper sub-mode of TM87 (at 73). The lower sub-mode of TM87 has dissappeared in TM97. It seems as if more of the salient features of the distributions could be used to obtain a non-linear scaling transformation which achieved a clearer supposition of the bulk of the intensity distributions in these cases. We do not go into detail here.

Interpretation of the differences between TM97 and ETM2000 are more problematic, since the two sensors differ.

Also, Hayes and Sader (2004) and many other researchers use NDVI as a basis for change detection, rather than the raw band data. Hence, in further research we should probably be looking at distributions on the NDVI scale rather than the raw bands.

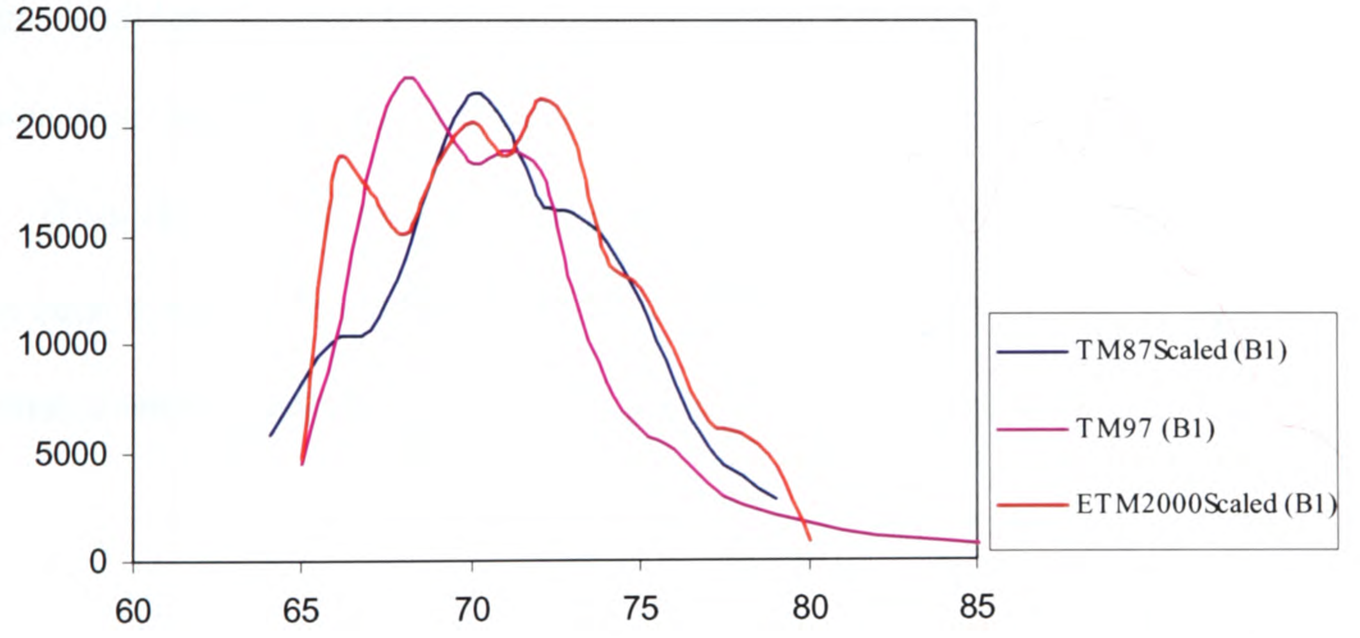


Figure 8.3: The median-interquartile range scaling of 1987, 2000 (B1) distribution to the 1997

"Spectral Standardization" makes an implicit assumption of "no change" and "no growth" in the spectral distribution shape/signature, contrary to the basis of the whole exercise, which is attempting to detect and characterise change and growth from the image sequence. Indeed our interpretation of the changes between TM87 and TM97 are in this mode of thought.

What is really needed is a "spectral evolution model", which allows, for a given forest compartment of a given forest type, to change ("continuously") in spectral intensity distribution with increasing age (assuming that images were taken by the same instrument/sensor, at the same time in the season each year, under the same lighting conditions). We would expect that for two similar stands, of the same type, but differing in age by (for example) ten years would have spectral intensity distributions such that the younger crop's distribution evolved into the older crop's distribution after a period corresponding to the age difference.

We therefore see that in such an attempt to overcome the inappropriateness of the assumptions of spectral distribution shape/signature standardization we find that spectral characterizations and spectral evolution models are confounded with the forest type and forest age variables. Note that the classificaiton status of a pixel would normally be obtained at a later stage, when image classification is done.

If we are in a supervised learning situation (for example, we might have ground truth data on such things as forest type, age, etc...) it seems as if modelling of spectral evolution and forest classification/aging should take place simultaneously.

References

Forest Biometrcs

Ahuja, J. C. and Nash, S. W. 1967. The generalized Gompertz-Verhulst family of distributions. Sankhya, A, 29:141-156.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6):716 - 723.

Ali, M. M., Mikhail, N.N., and Haq, M.S. 1978. A class of bivariate distributions including the bivariate logistic. Journal of Multivariate Analysis, 8:405–412.

Ané, T. and Kharoubi, C. 2003. Dependence structure and risk measure. Journal of Business, 76(3):411-438.

Assmann, E. 1970. The principles of forest yield study. Pergamon, Oxford.

Avery, T.E. and Burkhart, H.E. 1994. Forest measurements (4^{th} ed). McGraw-Hill, New York.

Bailey, R. L. and Dell, T. R., 1973. Quantifying diameter distributions with the Weibull function. For. Sci. 19: 97-104.

Barton, M.A. and Lord, F.M. 1981. An upper asymptote for the three-parameter logistic item-response model. *Research Bulletin* 81-20. Princeton Educational Testing Service.

Berkson, J., 1951. Why I prefer logits to probits. Biometrics, 7: 327-339.

Bliss, C.I. and Reinker, K.A. 1964. A lognormal approach to diameter distributions in even-aged stands. For. Sci. 10:350-360.

Bouyé E., Durrelman V., Nikeghbali A., Riboulet G., and Roncalli T. 2000. Copulas for Finance: A Reading Guide and some Applications. working paper, Credit Lyonnais, Paris (download from http://gro.creditlyonnais.fr/content/wp/copula-survey.pdf).

Bukac, J. 1972. Fitting S_B curves using symmetrical percentile points. Biometrika, 59:688-690.

Burr, I. W., 1942. Cumulative frequency functions. Annals of Mathematical Statistics, 13: 215-232.

Burr, I. W., 1968. On a general system of distributions. III. The sample range. Journal of the American Statistical Association, 63: 636-643.

Burr, I. W. and Cislak, P. J., 1968. On a general system of distributions. I. Its curve-shaped characteristics. II. The sample median. Journal of the American Statistical Association, 63: 627-635.

Burr, I. W., 1973. Parameters for a general system of distributions to match a grid of α_3 and α_4. Communications in Statistics, 2(1): 1-21.

Carrière J. F. 2000. Bivariate survival models for coupled lives. Scandinavian Actuarial Journal, 100(1):17-32.

Castillo, E. and Hadi, A.S. 1995. A method for estimating parameters and quantiles of distributions of continuous random variables. Computational Statistics and Data Analysis, 20: 421-439.

Cheng, R.C.H. and Traylor, L. 1995. Non-regular maximum likelihood problems (with discussion). Journal of the Royal Statistical Society, Series B, 57(1):3-44.

Clayton, D.G. 1978. A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65:141–151.

Clemen, R. T. and Reilly, T. 1999. Correlations and copulas for decision and risk analysis. Management Sci. 45:208–224.

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

Conway, D. 1983. Farlie-Gumbel-Morgenstern distributions. in: S. Kotz, N.L. Johnson (Eds.), Encyclopedia of Statistical Sciences, 3, Wiley, New York, pp 28-31.

Cook, R. and Johnson, M.E. 1981. A Family of Distributions for Modeling Non-elliptically Symmetric Multivariate Data. Journal of the Royal Statistical Society, Series B, 43(2):210-218.

Cook, R. and Johnson, M.E. 1986. Generalized Burr-Pareto-Logistic distributions with applications to a uranium exploration data set. Technometrics, 28:81-89.

Cox, D.R., and Hinkley, D.V. 1979. Theoretical statistics. Chapman and Hall, London.

Dall'Aglio, G. (1991) Fréchet classes: the beginnings. Advances in probability distributions with given marginals (Rome, 1990), 1--12, Math. Appl., 67, Kluwer Acad. Publ., Dordrecht.

David, H.A. 1985. Order statistics. In: Encyclopedia of Statistical Sciences, Volume 6 (ed: S. Kotz, N.L. Johnson & C.B. Read). Wiley, New York. 6:504-510.

DeBrota D.J., Roberts S.D., Swain J.J., Dittus R.S., Wilson J.R., and Venkatraman S. 1988. Input modeling with the Johnson system of distributions. Proceedings of the 1988 Winter simulation conference (Abrams M., Haigh P., and Comfort J. eds.), pp165-177.

Devroye (1986) Non-uniform random variate generation. NY, Springer-Verlag.

Duffy, S.F., Powers, L.M., and Starlinger, A. 1993. Reliability analysis of structural ceramic components using a three-parameter Weibull distribution. Journal of Engineering for Gas Turbines and Power, 115(1):109-116.

Embrechts, P., McNeil, A, Straumann, D., 1999. Correlation and Dependence in Risk Management: Properties and Pitfalls, Working Paper, Risklab ETH Zurich. (Available at http://www.risklab.ch/ftp/papers/CorrelationPitfalls.pdf)

Embrechts, P., Lindskog, F., McNeil, A, 2001. Modeling Dependence with Copulas and Applications to Risk Management, Working Paper, Risklab ETH Zurich. (Available at http://www.risklab.ch/ftp/papers/DependenceWithCopulas.pdf)

Fisher, N.I. 1997. Copulas. In: Encyclopedia of Statistical Science, Update Vol. 1:159-163. John Wiley Sons, New York.

Frank, M.J., 1979. On the Simultaneous Associativity of F(x, y) and x+y−F(x, y). Aequationes Mathematicae, 19:194–226.

Frees, E.W. and Valdez, E. (1998) Understanding relationships using copulas, North American Actuarial Journal, 2: 1-25.

Galambos, J. 1978. The Asymptotic Theory of Extreme Order Statistics. Wiley, New York.

Genest, C. 1987. Frank's family of bivariate distributions. Biometrika, 74:549-55.

Genest, C. and MacKay, J. 1986. The joy of copulas: bivariate distributions with uniform marginals. The American Statistician, 40:280-283.

Genest, C. and Rivest, L.P. 1993. Statistical inference procedures for bivariate Archimedian copulas. Journal of American Statistical Association, 88(423):1034-1043.

Genest, C., Ghoudi, K. and Rivest, L.-P., 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika, 82(3):543–552.

Gross B. 1996. Least squares best fit method for the three parameter Weibull distribution: Analysis of tensile and bend specimens with volume or surface flaw failure. NASA Technical Memorandum 4721.

Gumbel, E.J. 1960. Bivariate exponential distributions. Journal of American Statistical Association, 55: 698-707.

Gumbel, E.J. 1961. Bivariate logistic distributions. Journal of American Statistical Association, 56:335-349.

Gupta, R.D. and Kundu, D., 1999. Generalized exponential distributions. Austral. N. Z. J. Statist. 41(2): 173-188.

Gupta, R.D. and Kundu, D., 2001a. Exponentiated exponential distribution: an alternative to gamma and Weibull distributions. Biometrical J. 43 (1): 117-130.

Gupta, R.D. and Kundu, D. 2001b. Generalized exponential distribution: different methods of estimations. Journal of Statistical Computation and Simulation, 69(4):315-338.

Gupta R. D. and Kundu D., 2002. Generalized exponential distribution: Statistical Inferences. Journal of Applied Statistical Sciences, 1: 101-118.

Gupta, R.D. and Kundu, D., 2003. Discriminating between Weibull and generalized exponential distributions. Computational Statistics and Data Analysis, 43:179-196.

Gupta, R.D. and Kundu, D., 2004. Discriminating between Gamma and generalized exponential distributions. Journal of Statistical Computation and Simulation, 74(2):107-121.

Hafley, W.L. and Schreuder, H.T. 1976. Some non-normal bivariate distributions and their potential for forest application. pp 104-114 in the XVI World Congress Proc., Div. VI. IUFRO.

Hafley, W. L. and Schreuder, H. T., 1977. Statistical distributions for fitting diameter and height data in even-aged stands. Can. J. For.Res. 7: 481-487.

Hafley W.L., and Buford, M.A. 1985. A bivariate model for growth and yield prediction. For. Sci. 31:237-247.

Henessey, D.A. and Lapan, H.E. 2002. The use of Archimedean copulas to model portfolio allocations. Mathematical Finance, 12:143-154.

Hougaard, P. 1986. A Class of Multivariate Failure Time Distributions. Biometrika, 73:671-678.

Hui, G. and Sheng, W. 1995. Study on stand diameter structure model. Forest Research, 8(2):127-131. in Chinese with English abstract.

Hürlimann, W. 2004. Fitting bivariate cumulative returns with copulas. Computational Statistics and Data Analysis, 45:355-372.

Hutchinson, T.P. and Lai, C.D. 1990. Continuous bivariate distributions, emphasising applications. Rumsby Scientific Publishing, Adelaide.

Ishikawa, Y., 1987. The Richards growth function in terms of distribution. J. Jpn. For. Soc. 69(11):445-449. in Japanese.

Ishikawa, Y., 1991. Application of Richards' growth function to the frequency distribution of DBH in stands. J. Jpn. For. Soc. 73(5):365-369. in Japanese.

Ishikawa, Y., 1996. Analysis of the diameter distribution using the Richards distribution. Jpn. J. Plann. 27:7-18. in Japanese with English abstract.

Ishikawa, Y., 1997. Analysis of the diameter distribution by the Richards distribution function (II) Fitting from relationship between mean diameter or diameter variance and parameter k. Jpn. J. Plann. 29:41-47. in Japanese with English abstract.

Ishikawa, Y., 1998. Analysis of the diameter distribution using the Richards distribution function (III) Relationship between mean diameter or diameter variance and parameter m or of uniform and even-aged stands. Jpn. J. Plann. 31:15-18. in Japanese with English abstract.

Jiang, R. and Murthy, D.N.P. 1999. The exponentiated Weibull family: a graphical approach. IEEE Transactions on Reliability, 48(1):68-72.

Joe, H. and Xu, J.J. 1996. The estimation method of inference functions for margins for multivariate models. Technical Report, no. 166, Department of Statistics, University of British Columbia. (download from http://hajek.stat.ubc.ca/~harry/ifm.pdf)

Joe, H. 1997. Multivariate Models and Dependence Concepts. Chapman and Hall, London.

Joe, H. 2004. Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analysis (in press).

Johnson, M.E. 1987. Multivariate statistical simulation. John Wiley & Sons, New York.

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., 1954. Systems of frequency curves derived from the first law of Laplace. Trab. Estad. 5: 283-291.

Johnson, N. L. and Kotz, S., 1970. Continuous univariate distributions (2 volumes). Houghton Mifflin, New York.

Johnson, N.L., Kotz, S. 1972. Distribution in statistics: Continuous multivariate distributions. Wiley, New York.

Jouini, M.N. and Clemen, R.T. 1996. Copula models for aggregating expert opinions. Operations Research, 44(3):444-457.

Kamziah, A.K., Ahmad M.I., and Lapongan J. 1999. Nonlinear regression approach to estimating Johnson S_B parameters for diameter data. Can. J. For. Res. 29:310-314.

Kendall, M. and Stuart, A. 1977. The advanced theory of statistics-Vol.1: Distribution theory. Griffin, London.

Kimeldorf, G., and Sampson, A., (1975). Uniform Representations of Bivariate Distributions, Communications in Statistics, Theory and Method 4(7): 617-627.

Klugman, S.A. and Parsa, R. 1999. Fitting bivariate loss distributions with copulas. Insurance: Mathematics and Economics, 24: 139-148.

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

Kotz, S. (1997). Some remarks on copulas in relation to modern multivariate analysis, prepint, 1997 International Symposium on Contemporary Multivariate Analysis and Its Applications, Hong Kong.

Kundu, D. and Raqab, M.Z. 2005. Generalized Rayleigh distribution: different methods of estimations. Computational Statistics and Data Analysis, 49(1): 187-200.

Lee, L.F. 1983. Generalized Econometric Models with Selectivity. Econometrica, 51:507–512.

Li, D.X. 2000. On Default Correlation: A Copula Function Approach. Journal of Fixed Income, 9(4):43-54.

Li, F., Zhang L., and Davis C.J., 2002. Modeling the joint distribution of tree diameters and heights by bivariate generalized beta distribution. For. Sci. 48(1):47-58.

Lindsay, S.R., Wood, G.R., and Woollons, R.C. 1996. Modelling the diameter distribution of forest stands using the Burr distribution. Journal of Applied Statistics, 23(6):609-619.

Li, X. and Wang, M. 2001. Research on whole-stand volume models. Forest Research 14(1):50-53 (in Chinese with English abstracts).

Mage, D.T. 1980. An explicit solution for SB parameters using four percentile points. Technometrics, 22(2):247-251.

Mardia, K.V. 1967. Some contributions to contingency-type bivariate distributions. Biometrika, 54: 235-249. (Corrections in Biometrika, 55: 597).

Mardia, K. V., 1970a. Some problems of fitting for contingency-type bivariate distributions. Journal of the Royal Statistical Society, B. 32, 254-264.

Mardia, K. V., 1970b. Families of bivariate distributions. Griffin, London.

Mathsoft, 1999. S-plus 2000 guide to statistics, Vol.1. Data analysis products division, Mathsoft, Inc., Seattle, Washington.

McCullagh, P. and Nelder, J.A., 1989. Generalized Linear Models. Second Edition. Chapman and Hall, London.

Mendes, B.V.M. and Souza, R.M. 2004. Measuring financial risks with copulas. International Review of Financial Analysis, 13:27-45.

Micocci, M. and Masala, G. 2003. Pricing pension funds guarantees using a copula approach. (Download from http://www.afir2003.nl/).

Mitchell, K.J., 1975. Dynamics and simulated yield of Douglas-fir. Forest Science Monograph 17, Society of American Foresters.

Mudholkar, G.S. and Srivastava, D.K. 1993. Exponentiated Weibull family for analyzing bathtub failure-rate data. IEEE Trans. Reliability, 42:299–302.

Mudholkar, G.S., Srivastava, D.K. and Freimer, M. 1995. Exponentiated Weibull family a reanalysis of the bus motor failure rate data. Technometrics, 37:436–445.

Mudholkar, G.S. and Hutson, A.D. 1996. The exponentiated Weibull family: some properties and a flood data application. Communications in Statistics: Theory and Methods, 25(12):3059-3083.

Mykytka, E.F. and Ramberg J.S., 1979. Fitting a distribution to data using an alternative to moments. In: IEEE Proceedings of the 1979 Winter Simulation Conference, Association for Computing Machinery, New York, pp. 361–374.

Nassar, M.M. and Eissa, F.H. 2003. On the exponentiated Weibull distribution. Communications in Statistics: Theory and Methods, 32(7):1317-1336.

Nelsen, R.B. 1999. An Introduction to Copulas. Springer, New York.

Nelson, T.C., 1964. Diameter distribution and growth of loblolly pine. For. Sci. 10:105-115.

Newberry, J.D. and Burk, T.E. 1985. S_B distribution-based models for individual tree merchantable volume-total volume ratios. For. Sci. 31:389-398.

Oakes, D. 1982. A model for association in bivariate survival data. Journal of the Royal Statistical Society - B, 44(3):414-422.

Oakes, D. 1986. Semiparametric inference in a model for association in bivariate survival data. Biometrika, 73: 353-361.

Oakes, D. 1989. Bivariate survival models induced by frailties. J. Am. Statist. Ass., 84: 487-493.

Ord, J. K., 1972. Families of frequency distributions. Griffin, London.

Öztürk A. and Dale R.F. 1985. Least squares estimation of the parameters of the generalised lambda distribution. Technometrics, 27:81–84.

Pienaar, L.V. and Turnbull, K.J. 1973. The Chapman-Richards generalization of Von Bertalanffy's growth model for basal area growth and yield in even-aged stands. For. Sci. 19:2-22.

Phelps, A.L. and Weissfeld, L.A. (1997). A comparison of dependence estimators in bivariate copula models. Comm. in Statist., Part B -- Simul. Computat. 26: 1583-1597.

Plackett, R.L. 1965. A class of bivariate distributions. Journal of American Statistical Association, 60:516-522.

Ramberg J.S., Tadikamalla P.R., Dudewicz E.J. and Mykytka E.F. 1979. A probability distribution and its uses in fitting data. Technometrics, 21:201–214.

Ramesh, N.I. 1995. Statistical analysis on Markov-modulated Poisson processes. Environmetrics, 6: 165-179.

Raqab M. Z. 2002. Inferences for generalized exponential distribution based on record statistics. Journal of Statistical Planning and Inference 104: 339-350

Raqab M. Z., 2004. Generalized exponential distribution: moments of order statistics. Statistics: A Journal of Theoretical and Applied Statistics. 38(1):29-41.

Raqab M. Z., Ahsanullah M. 2001. Estimation of location and scale parameters of generalized exponential distribution based on order statistics. J. Statist. Comput. Simul. 69: 109-124

Rayner, G.D. and MacGillivray, H.L. 2002. Weighted quantile-based estimation for a class of transformation distributions. Computational Statistics and Data Analysis, 39(4):401-433.

Rennolls, K. and Blackwell, P., 1988. An integrated forest process model: Its calibration and its predictive performance. For. Ecol. Manage. 25: 31–58.

Rennolls, K., Geary, D. N., and Rollinson, T. J. D., 1985. Characterizing diameter distributions by the use of the Weibull distribution. Forestry, 58(1):57-66.

Rennolls, K. and Wang, M., 2005. A new parameterization of Johnson's S_B distribution with application to fitting forest tree diameter data. Can. J. For. Res., 35:575-579.

Reynolds, M.R., Burk, T.E. and Huang, W.C. 1988. Goodness-of-fit tests and model selection procedures for diameter distribution models. For. Sci. 34:373-399.

Rodriguez, R.N., 1977. A guide to the Burr type XII distributions. Biometrika, 64(1):129-134.

Rodríguez-Lallena, J.A. and úbeda-Flores, M. (2004) A new class of bivariate couplas. Statist. Probab. Lett. 66: 315-325.

Ross, G.J.S. 1990. Nonlinear estimation. New York : Springer-Verlag.

Schreuder, H.T., Bhattacharya, H.T., and McClure, J.P. 1982a. Towards a unified distribution theory for stand variables using S_{BBB} distribution. Biometrics, 38:137-142.

Schreuder, H.T., Bhattacharya, H.T., and McClure, J.P. 1982b. The S_{BBB} distribution: a potentially useful trivariate distribution. Can. J. For. Res. 12:641-645.

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

Schucany, W.R., Parr, W.C., and Boyer, J.E. 1978. Correlation structure in Farlie-Gumbel-Morgenstern Distributions. Biometrika, 65:650-653.

Schweizer, B. and Sklar, A. 1983. Probabilistic metric spaces. North-Holland Publishing Co., New York.

Schweizer, B. and Wolff, E. F. 1981. On nonparametric measures of dependence for random variables. Ann. Statist. 9(4):879-885.

Scolforo, J.R.S., Tabai, F.C.V., Macedo, R.L.G., Acerbi, F.W., and Assis, A.L. 2003. S_B distribution's accuracy to represent the diameter distribution of Pinus taeda, through five fitting methods. For. Ecol. Manage. 175:489-496.

Self, S.G. and Liang, K. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of American Statistical Association, 82: 605-610.

Shayib M.A. 1989. The procedure for selection of transformations from the Johnson system. Comm. Statist. B, Simulation 18:1457–1464.

Shi, J.H. and Louis, T.A. 1995. Inferences on the association parameter in copula models for bivariate survival data. Biometrics, 51:1384-1399.

Shiver, B.D. 1988. Sample sizes and estimation methods for the Weibull distribution for unthinned slash pine plantation diameter distributions. For. Sci. 34:809-814.

Siekierski, K. 1992. Comparison and evaluation of three methods of estimation of the Johnson S_B distribution. Biometrical Journal, 7:879-895.

Singh, U., Gupta, P.K. and Upadhyay, S.K. 2002. Estimation of exponentiated Weibull shape parameters under LINEX loss function. Commun. Statist. Simulation Comput. 31 (4):523–537.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. (French) Publ. Inst. Statist. Univ. Paris 8, 229-231.

Sklar, A. 1973. Random variables, joint distribution functions and copulas. Kybernetika 9:449-460.

Slifker J.F. and Shapiro S.S. 1980. The Johnson system: Selection and parameter estimation. Technometrics, 22:239–246.

Smith M.D. 2003. Modelling sample selection using Archimedean copulas. Econometrics Journal, 6(1): 99-123.

Smith, R.L. 1985. Maximum likelihood estimation in a class of nonregular cases. Biometrika, 72(1):67-90.

Starlinger A, Duffy S.F., Palko J.L. 1993. Parameter estimation techniques based on optimizing goodness-of-fit statistics for structural reliability. Proceedings of the 10th Biennial Conference on Reliability, Stress Analysis, and Failure Prevention, DE-Vol. 55, R.J. Schaller ed., ASME, Albuquerque, NM, pp. 65-72.

Storer, R.H., Swain, J.J., Venkatraman, S. and Wilson, J.R. 1988. Comparison of Methods for Fitting Data Using Johnson Translation Distributions. In Proceedings of the 1988 Winter Simulation Conference, M. Abrams, P. Haigh, and J. Comfort (eds.), 476-481. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Swain J.J. and Wilson J.R. 1985. Fitting Johnson distributions using least squares: simulation applications. Proceedings of the 1985 Winter Simulation Conference, p.150-157 (Gantz D.T., Solomon S.L., and Blais G.C. eds.), Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Swain, J., Venkatraman, S. and Wilson, J. 1988. Least-squares estimation of distribution function in Johnson's translation system. Journal of Statistical Computation and Simulation, 29: 271-297.

Tadikamalla, P.R. and Johnson, N.L., 1982. Systems of frequency curves generated by transformations of logistic variables. Biometrika, 69(2):461-465.

Tang, S.Z. 1991. Establishment and application of an integrated stand growth model of masson pine in Daqingshan. Forest Research, 4(monograph):8-13(in Chinese).

Tewari, V.P. and Gadow, K.V., 1997. Fitting a bivariate distribution to diameter–height data of forest trees. Indian Forester 123: 815–820.

Tewari, V.P., and Gadow, K.V. 1999. Modelling the relationship between tree diameters and heights using S_{BB} distribution. For. Ecol. Manage. 119:171-176.

Vuong, Q.H. 1989. Likelihood ratio tests for model selection and non-nested hypothesis. Econometrica, 57(2): 307-333.

Wagner, M.A.F. and Wilson, J.R. 1996. Using univariate Bézier distributions to model simulation input processes. IIE Transactions 28:699-711.

Wang, M. and Li, X. 2000. Research on nonlinear height-diameter models. Forest Research 13(1): 75-79 (in Chinese with English abstracts).

Wang, M., and Rennolls, K. 2005a. Introduing the Logit-Logistic Distribution. In Press, Can. J. For. Res.

Wang, M., and Rennolls, K. 2005b. Bivariate Distribution Modelling with Tree Diameter and Height Data. For. Sci., in Review.

Wang, M. and Sun, D., 1998. Quantifying and predicting stand diameter structure with the Logistic distribution. Forest Research 11(5):537-541 (in Chinese with English abstracts).

Wang, M. and Tang, S. 1997. Research on universal Height-Diameter curves. Forest Research 10(3):259-264 (in Chinese with English abstracts).

Weibull, W. 1939. A Statistical Theory of the Strength of Materials. Ing. Vetenskaps Akad. Handl. (Royal Swedish Institute Engineering Research Proceedings) NR151.

Weibull, W. 1967. The order statistics yi=log(zim), their properties and use for parameter estimation (z=standardized Weibull variate). USAF Report AFML-TR-67-161.

Wheeler R.E. 1980. Quantile estimators of Johnson curve parameters. Biometrika, 67:725–728.

Whitt, W. 1976. Bivariate distributions with given marginals. Ann. Statist. 4 (6): 1280-1289.

Wilson J.R. 1983. Fitting Johnson curves to univariate and multivariate data, Proceedings of the 1983 Winter Simulation Conference. (Roberts S.D., Banks, J. and Schmeiser B.W. eds.) p.115, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Wüthrich, 2004. Extreme value theory and Archimedean copulas. Scandinavian Actuarial Journal, 104:211-228.

Wykoff, W.R., Crookston N.L. and Stage A.R. 1982. User's guide to the stand prognosis model. USDA For. Serv. Gen. Tech. Rep. INT-133. 112 p.

Yi, W., and Bier, V. 1998. An Application of Copulas to Accident Precursor Analysis. Management Science, 44:257-270.

Zarnoch, S.J. and Dell, T.R. 1985. An evaluation of percentile and maximum likelihood estimators of Weibull parameters. For. Sci. 31:260-268.

Zhang, L., Packard, K.C., and Liu, C. 2003. A comparison of estimation methods for fitting Weibull and Johnson's SB distributions to mixed spruce-fir stands in northeastern North America. Can. J. For. Res. 33:1340-1347.

Zhou, B., and McTague J.P. 1996. Comparison and evaluation of five methods of estimation of the Johnson system parameters. Can. J. For. Res. 26:928-935.

Zohrer, F. 1972. The beta-distribution for best fit of stem-diameter-distribution. 3rd Conf. Advisory Group For. Stat. Proc. IUFRO, Institut National Recherche Agronomique, Paris, Publ. 72-3. pp. 91-106.

Remote Sensing

Ardö, J., 1992. Volume quantification of coniferous forest compartments using spectral radiance recorded by LANDSAT Thematic Mapper. International Journal of Remote Sensing, 13: 1779–1786

Bardera, A., Feixas, M., and Boada, I. 2004. Normalized similarity measures for medical image registration. http://ima.udg.es/iiia/GGG/UsersDocs/feixas/SPIE04_final.pdf

Bauer, M.E., Burk, T.E., Ek, A.R., Coppin, P.R., Lime, S.D., Walsh, T.A., Walters, D.K., Befort, W., and Heinzen, D.F. 1997. Satellite inventory of Minnesota forest resources. Photogrammetric Engineering and Remote Sensing, 60(3):287-298.

Bernstein, R. 1983. Image geometry and rectification. Chapter 21 in Manual of Remote Sensing, 2nd Edition (Colwell, R.N. ed.), Falls Church, Virginia: American Societry of Photogrammetry, Vol. 1:875-881.

Bolstad, P.V. and Lillesand, T.M. 1992. Improved classification of forest vegetation in northern Wisconsin through rule-based combination of soils, terrain, and Landsat thematic mapper data. For. Sci. 38:5-20.

Brown L.G. 1992. A Survey of Image Registration Techniques. ACM Computing Surveys, 24(4):325-376.

Chen, H.M., Arora, M.K., and Varshney, P.K. 2003. Mutual information-based image registration for remote sensing data. International Journal of Remote Sensing, 24(18):3701-3706.

Cohen, W.B. and T.A. Spies. 1992. Estimating structural attributes of Douglas-fir/western hemlock forest stands from Landsat and SPOT imagery, Remote Sensing of Environment, 41:1-17.

Cole-Rhodes, A.A., Johnson, K.L., LeMoigne, J., and Zavorin, I. 2003. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. IEEE Transactions on Image Processing, 12(12):1495-1511.

Coppin, P.R. and M.E. Bauer. 1994. Processing of multitemporal Landsat TM imagery to optimize extraction of forest cover change features. IEEE Transactions on Geoscience and Remote Sensing, 32(4):918-927.

Coppin, P.R. and Bauer M.E. 1996. Digital change detection in forest ecosystems with remote sensing imagery. Remote Sensing Reviews, 13:207-234.

Czaplewski, R.L. 1999. Multistage remote sensing: toward an annual national inventory. Journal of Forestry, 97(12):44-48.

Dowman, I. and Dare, P. 1999. Automated procedures for Multisensor registration and orthorectification of satellite images. International Archives of Photogrammetry and remote Sensing, 32(7-4-3 W6):37-44.

Fogel, D.N., and Tinney, L. 1996. Image registration using multiquadric functions, the finite element method, bivariate mapping polynomials and thin plate spline. Technical Report Number 96-01, National Center for Geographic Information and Analysis, University of California, Santa Barbara.

Franco-Lopez, H., Ek, A. R., and Bauer, M. E. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of Environment, 77(3): 251-274.

Franklin, J., 1986. Thematic Mapper analysis of coniferous forest structure and composition. Internat. J. Remote Sensing, 7: 1287–1588.

Franklin, S.E., Moskal, L.M., Lavigne, M.B. and Pugh, K., 2000. Interpretation and classification of partially harvested forest stands in the Fundy model forest using multitemporal Landsat TM digital data. Canadian Journal of Remote Sensing, 26 (4): 318–333.

Gemmell, F.M., 1995. Effects of forest cover, terrain, and scale on timber volume estimation with Thematic Mapper data in the rocky mountain site. Rem. Sens. Environ. 51: 291–305.

Glasbey, C.A. and Mardia, K.V. (1998). A review of image warping methods. Journal of Applied Statistics, 25: 155-171. Available on http://www.bioss.sari.ac.uk/image/warp.pdf

Goshtasby, A., Stockman, G.C., and Page, C. 1986. A region-based approach to digital image registration with subpixel accuracy. IEEE Transactions on Geoscience and Remote Sensing, GE-24(3):390-399.

Haapanen, R., Ek, A.R., Bauer, M.E., and Finley, A.O. 2004. Delineation of forest/non-forest land use classes using nearest neighbor methods. Remote Sensing of Environment, 89(3):265-271.

Hall, F.G., Strebel D.E., Nickeson, J.E., and Goetz, S.J. 1991. Radiometric rectification: Toward a common radiometric response among multidate, multisensor images. Remote Sensing of Environment, 35(1):11-27.

Hayes, D.J., and Sader, S.A. 2004. Change Detection techniques for Monitoring Forest Clearing and Regrowth in a Tropical Moist Forest. http://www.ghcc.msfc.nasa.gov/corredor/change_detection.pdf

Holmgren, P. and Thuresson, T. 1998. Satellite remote sensing for forestry planning – a review. Scand. J. For. Res. 13:90-110.

Horler, D.N.H. and Ahern, F.J., 1986. Forestry information content of Thematic Mapper data. International Journal of Remote Sensing, 7(3): 405–428.

Igbokwe, J.I. 1999. Geometrical processing of multi-sensoral multi-temporal satellite images for change detection studies. International Journal of Remote Sensing, 20(6):1141-1148.

Jensen, J.R. 1996. Introductory digital image processing: a remote sensing perspective. 2nd ed. Prentice Hall, New Jersey

Joyce, S. and Olssen, H., 2000. Long-term forecast monitoring with temporal-spectral trajectories from Landsat TM data. pp477-489 in Zawila-Niedzwiecki and Brack (eds), 2000. http://rogow99.sggw.waw.pl/09_session_5/04/

Kennedy, R.E. and Cohen, W.B. 2003. Automated designation of tie points for image-to-image coregistration. International Journal of Remote Sensing, 24(17):3467-3490.

Kettig R. L. and Landgrebe, D. A. 1976. Classification of multispectral image data by extraction and classification of homogeneous objects. IEEE Transactions on Geoscience Electronics, GE-14(1): 19-26.

Keysers, K. and Unger, W. 2002. Elastic image matching is NP-complete. Pattern Recognition Letters, 24:445-453.

Kilpeläinen, P. and Tokola, T. 1999. Gain to be achieved from stand delineation in Landsat TM image-based estimates of stand volume. Forest Ecology and Management, 124:105-111.

Landgrebe, D. 1997. The evolution of Landsat data analysis. Photogrammetric Engineering and Remote Sensing, 63(7): 859-867.

Lee, C. and Landgrebe, D.A. 1997. Decision boundary feature extraction for neural networks. IEEE Transactions on Neural Networks, 8(1): 75-83.

Lefsky, M.A., Cohen, W.B., and Spies, T.A. 2001. An evaluation of alternative remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. Can. J. For. Res. 31:78-87.

Le Moigne, J., Campbell, W.J., and Cromp, R.F. 2002. An automated parallel image registration technique based on the correlation of wavelet features. IEEE Transactions on Geoscience and Remote Sensing, 40(8):1849-1864.

Li H., and Manjunath, B.S., and Mitra, S.K. 1995. A contour-based approach to multisensor image registration. IEEE Transactions on Image Processing, 4(3):320-334.

Lu, D., Mausel, P., Brondizio, E., and Moran, E. 2004. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon basin. Forest Ecology and Management, 198:149-167.

McRoberts, R.E. 1999. Joint annual forest inventory and monitoring system: the North Central perspective. Journal of Forestry, 97(12): 27-31.

McRoberts, R.E., Wendt, D.G., Nelson, M.D., and Hansen, M.H. 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. Remote Sensing of Environment, 81:36-44.

Moore M.M. and Bauer, M.E. 1990. Classification of forest vegetation in north-central Minnesota using Landsat multispectral scanner and thematic mapper data. Forest Science, 36(2):330-342.

Olsson, H., 1994. Changes in satellite-measured reflectances caused by thinning cuttings in boreal forests. Remote Sensing of Environment, 50: 221–230.

Pratt, W.K. 1974. Correlation techniques of image registration. IEEE Transactions on Aerospace and Electronic Systems, 10(3):353-358.

Pratt, W.K. 1991. Digital image processing. 2nd ed., Wiley, New York.

Rennolls, K. 1999. Algorithms and software for Landsat image classification with a case study on a tropical rain forest in Sumatra. http://rogow99.sggw.waw.pl/09_session_5/08/

Rennolls, K. 2002. Components of Error Of Maps From Remotely Sensed Images. Spatial Accuracy 2002.

Rennolls, K. 2004. Satellite "Image-Banks" in Forest Information Systems? Prospects for Accurate Estimation of Forest Change and Growth from Data-Fusion Models. Spatial Accuracy Symposium, July 2004. Portland Maine, USA.

Rennolls, K., Richards, T., Fedorec, A., Ibrahim, M.T., McManus, K., Butler, A. 2004a. Requirements and Design of an Integrated European Environmental Information Communication System, (IEEICS). In FEIDSS'04: First International Workshop on Forest and Environmental Information and Decision Support Systems. DEXA'04. In Press, IEEE.

Rennolls, K., Richards, T., Fedorec, A., Ibrahim, M.T., McManus, K., Butler, A. 2004b. Models and Tools for an Integrated European Environmental Management and Decision

Support System, (IEEMDSS). In FEIDSS'04: First International Workshop on Forest and Environmental Information and Decision Support Systems. DEXA'04. In Press, IEEE.

Schott, J., Salvaggio, C. and Volchok, W. 1988. Radiometric scene normalization using pseudoinvariant features. Remote Sensing of Environment, 26:1-16

Secker, J., Staenz, K., Gauthier, R.P., and Budkewitsch, P. 2001. Vicarious calibration of airborne hyperspectral sensors in operational environments. Remote Sensing of Environment 76(1): 81-92.

Song, C., Woodcock, C.E., Seto, K.C., Pax-Lenney, M., and Macomber, S.A. 2001. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? Remote Sensing of Environment, 75(2):230-244.

Stockman, G., Kopstein, S., and Benett, S. 1982. Matching images to models for registration and object detection via clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 4:229-241.

Teillet, P. M., Fedosejevs, G., Gauthier, R. P., O'Neill, N. T., Thome, Kurtis J., Biggar, Stuart F., Ripley, H., and Meygret, A. 2001. A generalized approach to the vicarious calibration of multiple Earth observation sensors in hyperspectral data. Remote Sensing of Environment, 77(3): 304-327.

Tomppo, E., 1987. Stand delineation and estimation of stands variates by means of satellite images. University of Helsinki, Department of Forest Mensuration and Management, Research Notes 19:60-76.

Tomppo, E., 1991. Satellite image-based national forest inventory of Finland. International Archives of Photogrammetry and Remote Sensing, 28 (7-1): 419–424.

Ton, J., and Jain, A.K. 1989. Registering Landsat images by point matching. IEEE Transactions on Geoscience and Remote Sensing, 27(5):642-651.

Tokola, T., Löfman, S., and Erkkilä, A. 1999. Relative calibration of multitemporal Landsat data for forest cover change detection. Remote Sensing of Environment, 68:1-11.

Trotter, C.M., Dymond, J.R. and Coulding, C.J., 1997. Estimation of timber volume in a coniferous plantation forest using LANDSAT TM. Internat. J. Remote Sensing, 18 (10): 2209–2223.

Vogelmann J. E, and Rock B. N. 1988. Assessing forest damage in high-elevation coniferous forests in Vermont and New Hampshire using Thematic Mapper data. Remote Sensing of Environment, 24: 227-246

Wilson, E.H., and Sader, S.A. 2002. Detection of forest harvest type using multiple dates of landsat TM imagery. Remote Sensing of Environment, 80:385-396.

Wolter, P.T., Mladenoff, D.J., Host, G.E., and Crow, T.R. 1995. Improved forest classification in the northern Lake states using multi-temporal Landsat imagery. Photogrammetric Engineering and Remote Sensing, 61(9):1129-1143.

Woodcock, C.E., and Harward, V.J. 1992. Nested-hierarchical scene models and image segmentation, International Journal of Remote Sensing, 13:3167-3187.

Wulder, M.A. and Seemann, D. 2003. Forest inventory height update through the integration of Lidar data with segmented Landsat imagery. Canadian Journal of Remote Sensing, 29(5):536-543.

Xie, H., Hicks, N., Keller, G.R., Huang, H., and Kreinovich, V. An IDL/ENVI implementation of the FFT-based algorithm for automatic image registration. Computers & Geosciences, 29:1045-1055.

Yang, X., and Lo, C.P. 2000. Relative radiometric normalization performance for change detection from multi-date satellite images. Photogrammetric Eng. Remote Sens. 66:967-980.

Yao, J., and Chern, C.T. 2001. The Practice of Automatic satellite Image registration. Presented at the 22nd Asian Conference on Remote Sensing, 5-9 Nov. 2001, Singapore. http://www.crisp.nus.edu.sg/~acrs2001/pdf/072yao.pdf

Yuan, D., and Elvidge, C.D. 1996. Comparison of relative radiometric normalization techniques. ISPRS J. Remote Sens. 51:117-126.

Zawila-Niedzwiecki, T., and Brack (eds), 2000. Remote Sensing and Forest Monitoring; Poceedings of a IUFRO Conference, Rogow, Poland, 1999. Warsaw University Press. http://rogow99.sggw.waw.pl/

Zitová B. and Flusser J. 2003. Image registration methods: a survey. Image and Vision Computing, 21:977-1000.