Investigation of the application of Best Linear Prediction for breeding and clonal production purposes in a *Eucalyptus grandis* population

By

ANDREA KATE LOUW

"... to make choices among the huge number of artificial selection procedures that the human mind can invent"

KEMPTHORNE, 1988

PREFACE

The experimental work described in this thesis was carried out in the School of Biochemistry, Genetics, Microbiology and Plant Pathology, University of KwaZulu-Natal, Pietermaritzburg, under the supervision of Dr Carolyn Hancock.

These studies represent original work by the author and have not otherwise been submitted in any other form to another university. Where use has been made of the work of others it is duly acknowledged in the text.

Andrea Louw

Andrea Louw

I hereby certify that this statement is correct.

CE Hancock

Dr Carolyn Hancock
Supervisor

Pietermaritzburg
December
2006

iii

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

a	Relative economic weighting
ANOVA	Analysis of variance
b	Index coefficient weighting
BLP	Best linear prediction
BV	Breeding value
CSIR	Council for Scientific and Industrial Research
dbf	Database file
DBH	Diameter at breast height
E.grandis	*Eucalyptus grandis*
GEI	Genotype-by-environment interaction
h^2	narrow-sense heritability
i	Selection intensity
LS-means	Least-Square means
n	Population size
n_o	Coefficient of sample size per group
p-values	Probability values
PROC	Procedure
r	Coefficient of relationship
R	Response to selection
r_A	Additive genetic correlation
r_p	Phenotypic correlation
REML	Restricted maximum likelihood
rep	Replication
rep*fam	Replication-by-family interaction
S	Selection differential
SA	South Africa
SAS	Statistical Analysis Software
μ	Population mean
σ	Standard deviation
σ_B	Between group standard deviation
σ_W	Within group standard deviation
σ_A^2	Additive genetic variance
σ_B^2	Between group variance
σ_E^2	Environmental variance
σ_P^2	Phenotypic variance
σ_T^2	Total observed phenotypic variance
σ_W^2	Within group variance

ABSTRACT

The genus *Eucalyptus* has been planted extensively throughout the world in tropical and subtropical regions, primarily because of its economic importance and use in wood and pulp production. Due to the growing demands for timber, forestry companies need to increase the productivity of available forest land. The genetic improvement of forest trees through selection and breeding involves a lengthy process of scientifically controlled trials focused on short-term and long-term goals using breeding and production populations. This investigation focused on the use of Best Linear Prediction (BLP) and its application to: (1) the prediction of genetic gains for a breeding population and, (2) the selection of superior individuals for clonal production of *E. grandis*.

A CSIR dataset for a 20-year-old progeny trial involving 90 open-pollinated families was obtained. Four traits, namely, diameter at breast height (DBH), stem form, splitting and density were identified for use in this investigation. Relevant data were extracted and a file termed, *Dataset* created. *Dataset* was edited, standardized and corrected for the fixed effect of replication using SAS® procedures.

Precise and accurate population parameter estimates are fundamental in determining breeding strategies and thus, heritabilities of each trait and phenotypic correlations between traits in *Dataset* were estimated using SAS® procedures. DBH was found to have the highest heritability (0.600), followed by density (0.492). The estimated heritability for stem form was 0.401 and splitting had the lowest heritability at 0.214. A high positive phenotypic correlation of 0.83 was estimated between DBH and stem form. The phenotypic correlations between other traits were close to zero.

An index provides a weighted score for individuals, which takes all relevant information into account and allows individuals or families to be chosen for breeding and production purposes. Consequently, Best Linear Prediction (BLP) of individual breeding values were calculated using MATGEN® (2003). Thereafter, BLP values were used to determine the rankings of individual trees for 15 different selection indices.

In order to determine the effect of selection on the change in the population mean of a trait, the breeding population's response to selection was predicted and compared across three selection strategies, namely: (1) individual selection, (2) single-trait index selection, and (3) multiple-trait index selection. The top 8% of individuals in the breeding population were selected for and the genetic gains were predicted. It was found that the response to selection was greatest when using individual selection. Furthermore, DBH had the best selection response for all three strategies as compared to the other traits under investigation.

Fifteen indices, considering different numbers and choice of traits, were compared for commonality among rankings of the top 30 individuals. Two methods, namely, (1) a rank-correlation matrix and (2) a manual assessment, were used. The commonality between indices showed that a simple index, considering two traits (DBH and density) was equally effective (93%) in identifying genetically superior individuals as the more complex index that considered four traits. Furthermore, it was possible to select for only three traits (DBH, splitting, density) and identify the same top 30 individuals as using the index that considered four traits.

The researcher's goal was to find the most desirable individuals in the population to be used for production purposes, such as clonal forestry. Consequently, various selection options, specifying certain trait requirements, were used to select superior individuals for use in production and deployment. The "commercial selection" option was the only option successful in obtaining an individual that met the required criteria for the four traits in the population of 475 individuals. The results suggested that breeders should consider large populations and only a few important traits in order to obtain a greater number of individuals suitable for mass propagation in clonal forestry.

In order to further investigate the effect of population size on the number of individuals suitable for clonal forestry, a hypothetical population was generated. This was accomplished using between family and within family standard deviation values obtained from *Dataset*. The large hypothetical population of 1000 individuals produced twelve individuals suitable for production purposes, as opposed to only one in the real population of 475 individuals. This result further indicates that a larger population provides a greater number of individuals appropriate for use in production and deployment.

This investigation successfully addressed the aims by: (1) calculating individual breeding values (BLP) and ranking individuals, (2) predicting the breeding population's response to selection, according to three strategies, for the four traits under investigation, and (3) identifying superior individuals for use in commercial clonal forestry.

As the work of tree breeders is aimed at improving the growth and quality of trees by increasing the frequency of desirable genotypes in the population, further research could focus on (1) the effect of different sets of economic weightings on index rankings in a population and (2) the influence that population structure has on the optimal genetic gains obtained.

1 CHAPTER ONE: CONTEXT AND AIMS OF THE INVESTIGATION

1.1 INTRODUCTION TO TREE BREEDING

1.1.1   The forestry industry

Forestry refers to the use and management of forests, and also includes the further processing of wood products into pulp for paper and packaging industries, sawn timber, furniture, shelving, flooring, and so forth. Forest woods are also used for fuel, charcoal production and construction materials, such as poles, beams, and thatching (Department of Water Affairs and Forestry, 2005). Forest resources are comprised of three components namely, plantations, natural or indigenous forests, and woodlands or savannas.

Due to agricultural demands, city expansion and road development, the land existing for forestry has diminished. This combined with the fact that pulp and paper production account for one per cent of the world's total economic output (Pot *et al.*, 2002) means that investments, particularly in the pulp and paper industry, and further growth in the demand for timber, is likely to continue to increase (Molony, 1999). Hence, forestry companies need to increase the productivity of available forest land.

Forestry operates with both short-term and long-term objectives. Short-term objectives include improvement in growth and survival of timber trees or increasing pulp yield for paper and cellulose products. Long-term objectives consist of enhancing the yield of several products for a range of possible forest sites, as well as maintaining and developing resistance to disease and insects (Namkoong *et al.*, 1988). There are often different ecological and economic goals within a forest region hence, different breeding objectives between forestry species and for various forest products (Namkoong, 1979).

A commercial plantation is made up of compartments (blocks) of trees where the trees of one compartment are mostly the same species and age and have all been planted at a fixed spacing (Department of Water Affairs and Forestry, 2005). South Africa (SA) supports a vibrant forestry industry where the mostly exotic species occupy 1, 333, 563 hectares of land, which is roughly 1.1% of SA's land area (Department of Water Affairs and Forestry, 2006). Forestry regions in SA stretch from the Eastern Cape along the eastern escarpment through KwaZulu-Natal, Mpumalanga and into the Limpopo Province. This includes areas with large differences in site qualities, altitudes,

temperature regimes and rainfall conditions. Ninty-two percent of SA's commercial plantation forestry is found in the three provinces, namely, Mpumalanga, KwaZulu-Natal and the Eastern Cape (Department of Water Affairs and Forestry, 2006). In order to increase the productivity of available forest land, many companies worldwide are involved with forest tree improvement programmes, through the application of genetic principles for the enhancement and management of forest trees.

One way of improving production of forest land is by changing the genetic parameters or properties of the tree populations (Fins *et al.*, 1992). This is known as tree breeding. The aim of the tree breeder is to produce a maximum of economically usable wood per tree per hectare. The objectives of a tree breeding programme must be clearly defined by the breeder, as the nature of forestry and forest tree improvement is long term and costly.

1.1.2   Traits of importance in tree breeding programmes

Forestry breeding, like crop and animal breeding, implies strong selection for one or more useful characteristics or traits (Libby *et al.*, 1969). Most traits of interest in tree improvement are economically important and are quantitatively inherited, indicating the influence of many genes and environmental factors (Wricke and Eberhard Weber, 1986). However, many of the most economically-important features of a forest tree are traits of the mature tree (Libby *et al.*, 1969). Genetic variances for a trait change over time, thus the trait may not be expressed in the same manner at different ages in the individual. Consequently, the trait's effects on other traits at different ages also changes (Namkoong and Kang, 1990). Furthermore, in many tree breeding programmes, the traits that need improvement will change from generation to generation because utilization standards of usable wood for a particular species may change in the future (Namkoong *et al.*, 1988).

There are many traits of high economic importance that breeders focus on, as these properties are confirmed to be major influences on the final value and quality of wood products. Wood density is one of the most important characteristics in determining the suitability of a piece of wood for a particular end use. Wood density is a good indicator of the conformability characteristics of fibres which in turn influence strength, surface properties and opacity of paper. There are direct relationships between all the strength properties of wood and its density (Verryn and Turner, 1999). Economic studies have shown that density has a major impact on mill profits because it affects harvesting, transportation and milling costs. Hence, breeding programmes integrating density have great consequences for forest owners' profits (Pot *et al.*, 2002).

Splitting is used as a selection trait to ensure that the largest trees with the lowest splitting scores are selected. Environmental conditions are thought to play a role in splitting. These include climate and soil conditions but can also relate to the way the tree is felled, transported and processed (Verryn and Turner, 1999). The highest splitting in *Eucalyptus grandis* (*E. grandis*) occurs in the

lowest 12 m of the tree which is the proportion of the tree with the highest economic value (Malan, 1979). In addition, diameter at breast height (DBH), splitting and stem straightness (or stem form) all have a large influence on veneer production value, and DBH and splitting also have a great influence on sawmill value. DBH is measured directly off standing trees at about 1.3 meters above ground (Cotterill and Dean, 1990), and stem straightness is visually assessed and assigned a subjective score such that the higher the score the straighter the tree. For example, an 8 point scale, where 8 indicates a straight tree and 1 indicates a very crooked tree. Good stem form reduces harvesting and processing costs. Furthermore, Mayo (1987) noted that breeding for resistance is one of the major activities of plant breeders in all countries. Tolerance to disease can be assessed on a 5 point scale where 0 equals no disease and 4 equals a heavily infested tree resulting in the death of the tree.

To advance genetic gains in various populations of commercial trees, an understanding of the genetic architecture of quantitative traits in the populations is required (Wu *et al.*, 2000). In order to improve the efficiency of breeding programmes knowledge of: (1) the heritability (h^2) of a trait; (2) the genetic and phenotypic correlations between traits; (3) the relative economic value of the trait; (4) the standard deviation (σ); and (5) the selection intensity (*i*) is required.

1.1.3   Role of quantitative genetics in tree breeding

Quantitative genetics is the branch of genetics studying the inheritance of measurable traits, known as multifactorial or quantitative traits. Their expression is influenced by combinations of several genetic and environmental factors (Lynch and Walsh, 1998). The phenotypes (observable characteristics) of quantitative traits show continuous (numerical) variation and their inheritance is complex (Snustad and Simmons, 2000) as the traits are controlled by several genes whose individual effects are too small to be detected by conventional Mendelian principles (Hill *et al.*, 1998).

The phenotype of quantitative traits is regarded as the sum of two components, the genetic and the non-genetic (environment) effects, represented as deviations from the overall population mean (Wricke and Eberhard Weber, 1986),

$$P = \mu + G + E$$

where

      P is the phenotypic value
      μ is the overall population mean
      G is the genotypic value
      E is the environmental effect

Therefore, the basis of quantitative genetics, as stated by Mayo (1987), is the partitioning of the phenotypic value of a quantitative trait for an individual into components attributable to the influence of genes and to the influence of the environment.

The analysis of quantitative traits is based on statistical descriptions, such as the means and variances of the phenotypes in a population or generation (Hill *et al.,* 1998; Snustad and Simmons, 2000). When geneticists consider a population of trees, they consider the variation within the population and the extent to which individual trees differ from the mean of the population. This measure is called variance (van Wyk, 1983). The genetic components of variation are important as they determine the rates at which traits in the population respond to selection (Lynch and Walsh, 1998). Breeders must be aware of the factors determining the genetic variation in a population so that efficient breeding strategies can be developed, as a lack of knowledge underlying variation in tree growth and development is the most important obstacle to successful tree breeding (Wu *et al.,* 2000).

Most decisions that are made in implementing a tree improvement programme depend on quantitative genetic models, particularly when designing breeding programmes and analyzing data. These quantitative models can subsequently be applied to many different levels of decision-making in forestry tree breeding. Tree improvement decisions fall into six major categories, namely:

- Programme initiation or continuation decisions
- Selection decisions
- Breeding decisions
- Testing decisions
- Production decisions
- Commercial deployment or management decisions

Various types of decisions are required in each of these categories. Table 1.1 shows how quantitative genetics is used to guide tree improvement decisions and indicates the ways in which quantitative analysis plays a key role in forestry management decisions.

Table 1.1 Application of quantitative genetics to tree improvement decisions (extracted from Fins *et al.*, 1992).

Tree improvement Decision	Applications of Quantitative Genetics
Programme initiation / continuation decisions	• Estimation of **potential genetic gains** to justify programme initiation and assess whether this gain is worth the costs and efforts.
Selection decisions	• Decision on **which traits to improve** through selection – depends on the goals of the organization. • Choice of **which candidates to select** – the species, the provenance (source), the choice of individuals or families within a source. • Assessment of **the number of candidates to select** – various population size alternatives are compared.
Breeding decisions	Determination of **a design strategy** to optimize: • genetic variation • population size • inbreeding level • projected genetic gain
Testing decisions	• **Comparisons among various genetic entries** – such as provenances, families, clones. • Classification of **treatment comparisons.** • Determination of **genetic and environmental effects.** • Evaluation of **genotype by environment interaction (GEI).**
Production decisions	• Verification of **the type of propagation system** – seed or vegetative propagation (clonal forestry) programmes. • Decision on **which genetic material to propagate.** • Assessment of **the numbers of families and individual genotypes to propagate.**
Commercial deployment / management decisions	• Organization of **the production and packaging of genetic material.** • Determination of **allocation procedures** – considering environmental pressures and risks, planting decisions, site-allocation, competitive behaviour of genotypes, stability analysis. Allows matching of specific material to specific site types. • **Development of improved stock** – empirical trials are critical for the development and testing of yield and growth hypotheses about families and clones.

1.1.4 Nature of tree improvement programmes

A tree improvement programme is an ongoing and recurrent process that incorporates cycles of selection, mating and testing for continued improvement over time. Due to the long generation interval of trees, breeding activities are usually focused on one cycle of improvement. However, in any tree improvement programme there is usually a need for a breeding (research and development) and a production (operational) phase (Zobel and Talbert, 1984). Where possible, these two phases are started simultaneously in an effort to minimize the delay before 'improved' material is made available (Hettasch *et al.*, 2006).

The objective of the breeding programme is to ensure a broad genetic base whilst still improving the breeding population. The objective of the production programme is for rapid deployment of the best genetic stock to maximize both yields and product quality (Hettasch *et al.*, 2006). A representation of the populations involved in a typical tree improvement programme is illustrated in Figure 1.1.
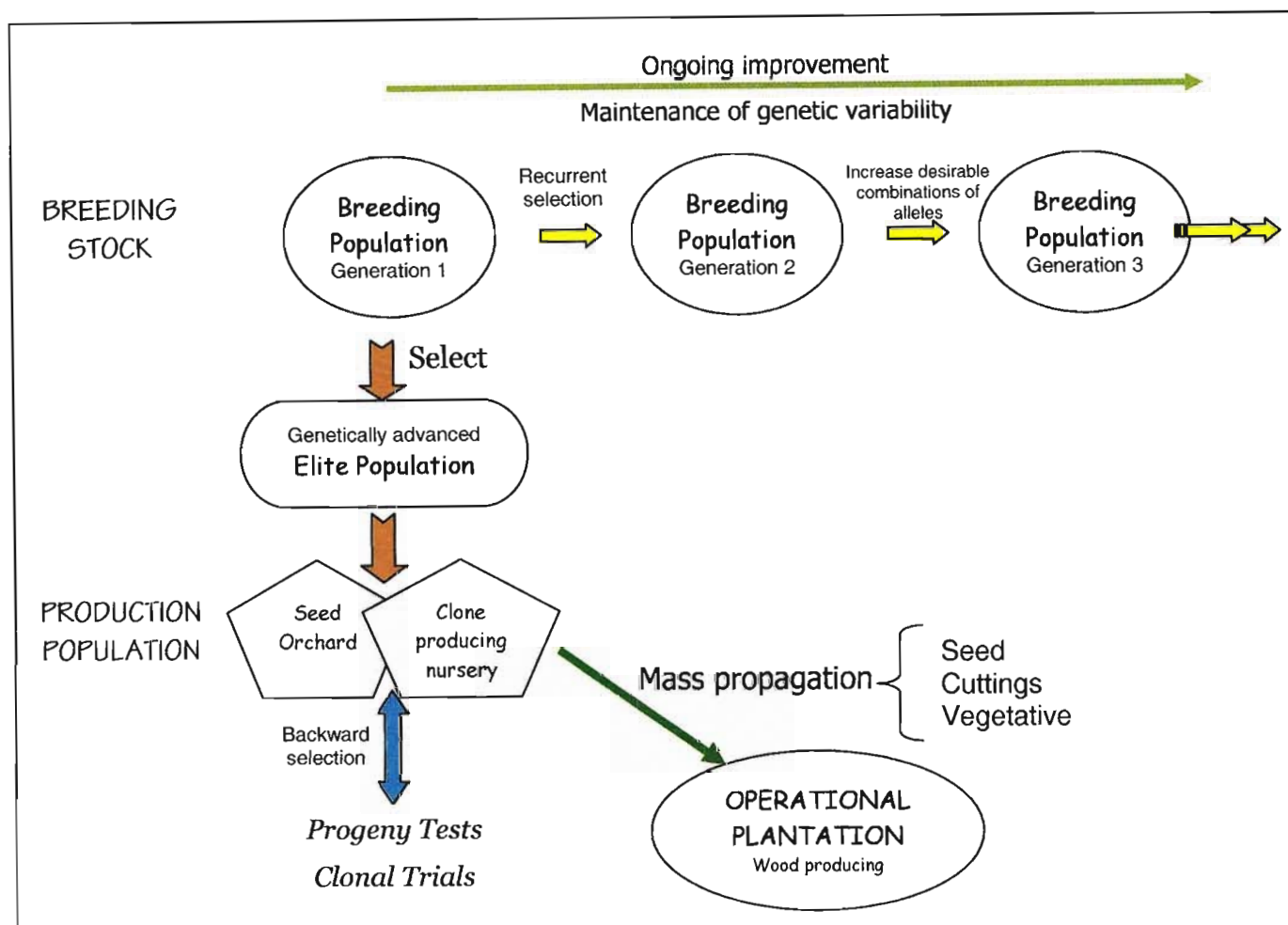
Figure 1.1 Representation of populations involved in a typical tree improvement programme.

The creation of elite breeding stock is regarded as a long-term breeding activity. To accomplish this, the breeding population undergoes recurrent selection which consists of selection of individuals within each generation and among the offspring of parents selected from the preceding generation (Hettasch *et al.*, 2006). The continuing improvement of the breeding population, over generations, is important for:

- The maintenance of sufficient genetic diversity to enable the breeder to achieve desired gains.
- The organization of the breeding population and generation turn-over.
- Gene conservation and the increase of desirable combinations of alleles.

- The removal of deleterious alleles in the population.

These factors will have profound impacts on future tree breeding generations.

A breeding population is used to form a new production population each generation. This is achieved via an elite population that is genetically advanced and intensively managed. The elite population consists of those individuals that are selected as the most desirable for seed orchards and clone producing nurseries.

The long-term activities (breeding population) provide the genetic resources that can be used in short-term breeding to account for environmental changes and changes in societal demands (Kang and Nienstaedt, 1987). The production population is focused on short-term breeding objectives and short-term activities are used in response to immediate commercial needs, such as:

- Genetically improved seeds or plants used in an operational plantation.
- Knowledge and techniques, gained from progeny tests and clonal trials.
- The requirement of disease resistance genotypes after disease infestation.

In each generation the benefits are captured for use in plantations by mass propagating the best, currently available genetic material through seed, cuttings or other methods of vegetative propagation.

1.2 IMPROVEMENT OF THE GENETIC PROPERTIES OF A BREEDING POPULATION

1.2.1 Genetic parameters of populations

Genetic parameters allow breeders to characterize a population of trees of interest (White and Hodge, 1989). For example, a population of trees has an average (mean) height and variance associated with the distribution of tree height. In forest genetics and tree improvement, there are many parameters that geneticists are interested in estimating, such as means, variances, covariances and heritabilities. Information on heritabilities of traits and the genetic and phenotypic correlations between these traits is fundamental knowledge required for planning efficient tree breeding strategies (Cotterill and Dean, 1990). Gathering information on the heritability and genetic correlations of each trait, and assigning economic weights to traits allows for the construction of an efficient breeding strategy which will maximize genetic gain (van Wyk and Verryn, 2000).

One of the fundamental parameters used to measure variation is the standard deviation (σ), which is a property of the trait and the population (Falconer and Mackay, 1996). Standard deviations are used for much the same purposes as variances but are easier to conceptualize as they are measured in trait units and have meaningful numerical values. The standard deviation also

provides a common scale unit which is particularly useful when comparing different traits, or when dealing with several traits at the same time (Fins *et al.*, 1992).

The measure of selection applied to a population in order to change the population mean is called the *selection differential* (*S*). The selection differential is the mean phenotypic value of the individuals selected as parents, expressed as a deviation from the mean phenotypic value of all the individuals in the parental generation before selection was made (Falconer and Mackay, 1996). The selection differential is often more conveniently expressed in terms of the σ (Wricke and Eberhard Weber, 1986),

$$ i = \frac{S}{\sigma_p} $$

where

i is the selection intensity
S is the selection differential
σ_p is the phenotypic standard deviation

If a breeder has some idea of the proportion of individuals to be selected, the selection intensity (i) can be estimated from tables constructed for this purpose (Becker, 1975).

(a) Heritability

Heritability is a genetic parameter defined as the ratio of the additive genetic variance to total phenotypic variance (Namkoong and Kang, 1990). This is expressed mathematically in terms of variance as (Wright, 1976),

$$ h^2 = \sigma^2_A \Big/ \sigma^2_P $$

where

h^2 is the narrow-sense heritability
σ^2_A is the additive genetic variance
σ^2_P is the phenotypic variance

Heritability is a useful statistic to describe relative contributions of the genotype and the environment to the phenotype (Namkoong, 1979). Thus, the h^2 of a particular trait is a measure of how strongly the observed variation of a trait is influenced by genetics and by the environment in a particular population (Hazel and Lush, 1943). Knowledge of the h^2 of traits allows the breeder to develop goal-orientated selective breeding programmes (Snustad and Simmons, 2000) and to predict genetic gain from simple selection procedures (Namkoong, 1979). In tree breeding, heritabilities of between 0.10 and 0.30 are considered intermediate heritabilities because moderate gains are usually expected from individual tree selection. Heritabilities less than 0.10 are considered low because poor gains are expected from selection, and heritabilities greater than 0.30 are considered high (Cotterill and Dean, 1990). Each heritability estimate is specific to the population, the trait and the environment on which the estimate is based (Fins *et al.*, 1992).

Heritability also represents the regression of individual tree's breeding values (A) on their measured phenotypic value (P) for the same trait, denoted as β_{AP} (Falconer and Mackay, 1996). The additive genetic variance component (σ^2_A) is the primary determinant of the degree to which offspring resemble their parents and thus, σ^2_A governs the response to selection for a particular trait. Hence, h^2 tells breeders whether selection for particular individuals will be effective due to the correlation between breeding values and phenotypic values. For a trait of high h^2, the phenotypic value of the trait can be expected to reliably reflect the true breeding value of an individual (Cotterill and Dean, 1990).

(b)   Correlations

The phenotypic correlation (r_P) is the statistical association between the measured phenotypic values for two traits in a population of trees (Cotterill and Dean, 1990). It is calculated as the covariance between the two traits divided by the product of the standard deviations of the individual traits (Lynch and Walsh, 1998):

$$r_P(x,y) = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$$

The additive genetic correlation (r_A) represents the correlation between breeding values for different traits (Cotterill and Dean, 1990). Knowledge of genetic correlations is necessary to predict correlated gains in one trait as a consequence of selection on another trait. Genetic correlations and hence correlated responses to selection are extremely variable from generation to generation. If the correlations are due to non-genetic sources they can change with environmental variations and if the correlations are partly genetic, they may change as linkage or pleiotrophic effects alter (Namkoong, 1979).

Genetic correlations between traits are of great interest to foresters since many plantations are subject to the simultaneous demands and selection for multiple sets of traits (Namkoong, 1979). If traits are correlated genetically, any changes in one might be accompanied by a deterioration of others (Young, 1961). It is not possible to build up reliable information concerning trait relationships unless forest geneticists perform the necessary calculations of genetic correlations as part of inheritance studies of tree populations.

(c)   Economic importance

In the same manner that nature selects for adaptive value, man selects for traits based on their economic value (Arbez et al., 1974). Economic values of a particular trait can be defined as the amount by which each unit of variation in the trait rises or lowers an individual's practical value (Hazel and Lush, 1943). Arbez et al. (1974) stated that economic weights must be estimated as a function of the influence the trait is supposed to have on the economic value of the mature trees. The relative economic weights of particular traits are important for the breeder to know as financial

constraints may dictate what breeding strategy to follow and hence what traits to select for (van Wyk and Verryn, 2000). However, economic weights are difficult to obtain and are subject to fluctuations since the economy and forest product values are uncertain. Hence, Namkoong (1969) stated that economic values are only predictable within broad limits. In addition, White and Hodge (1989) reported that parameters are never exactly known and estimates developed for any particular parameter will be associated with some error.

(d)   Environmental factors

All quantitative traits are influenced not only by a number of genes but also by a number of environmental factors. Environmental factors such as soil, moisture, temperature, and pH, may affect particular traits of interest and as a result influence tree breeding. A breeding programme is based on the principle that an individual's phenotype provides insight into its underlying genotypic value. However, if there is a substantial environmental influence, the amount of genetic information conveyed by a single phenotypic measurement may not be high (Lynch and Walsh, 1998). This is because individual phenotypes deviate from underlying genetic values due to confounding influences of the environment. The greater the environmental influence, the less accurate the phenotype becomes as a measure of genetic value. Different traits considered in tree breeding are more or less confounded by environmental effects (Fins et al., 1992). For example, environmental effects are often far less important for form traits, such as stem straightness or crown form, than for growth traits such as DBH and height (Cotterill and Dean, 1990).

Environments are composed of many distinct effects that breeders may not be able to distinguish. The environment also includes the tree's own physiological state and other organisms affecting trait expression. Many unknown microsite factors of the environment vary so rapidly over time, and over such small distances in unapparent patterns from tree to tree, that breeders are unable to recognize distinct factors or to effectively manage them (Namkoong et al., 1988). However, when some factors are sufficiently strong and direct in effect it becomes useful to examine these environmental effects on a set of genotypes.

Genes act in the context of an environment. A genotype by environment interaction (GEI) results when the relative performances of genotypes differ when grown in different environments (Romagosa and Fox, 1993). Thus, a GEI confounds a genotype's observed mean performance with its true value (Crossa, 1990), and as a result reduces the association between phenotypic and genotypic values (Romagosa and Fox, 1993). This is expressed mathematically as (Falconer and Mackay, 1996),

$$P = G + E + I_{GE}$$

where

I_{GE} is an effect attributed to genotype by environment interaction

Large differences between various environmental sites suggest that GEI may be present, as indicated in a previous study by Snedden and Verryn (1999). Consequently, an understanding of GEI for a particular species can have important management implications. Hence, knowledge of GEI allows breeders to determine an optimal breeding strategy for releasing genotypes with adequate adaptation to a target environment (Romagosa and Fox, 1993). A large amount of research on GEI has been done, with two reviews focusing on GEI in a South African context (van Wyk and Falkenhagen, 1984; Falkenhagen, 1985). van Wyk and Falkenhagen (1984) use examples of GEI for progeny trials from five species to illustrate the possible occurrence of GEI, and the need for close attention to this phenomenon when developing breeding populations for South Africa. The presence of GEI both within and between sites creates difficult and costly problems (Falkenhagen, 1985).

Breeders are faced with uncertainties in the environmental distribution of plantations and the kinds of genotypic responses to those conditions. When breeding for quantitative traits both the genetics of the trees under consideration as well as the environmental conditions where they will be grown must be taken into consideration. The relative effects of the genetic and environmental factors will determine the most effective method of improving the performance of the population. With knowledge of genetic parameters, breeders need to minimize environmental influences as much as possible in order to phenotypically identify those trees with the best genetic potential and select individual trees, families or clones for breeding and production purposes.

1.2.2   Selection strategies

Selection refers to the differential survival and reproduction among genotypes within a population (Snustad and Simmons, 2000). Selection does not create new genes, but rather increases the frequency of desirable of genes (Fins *et al.*, 1992). Charles Darwin hypothesized that variation provides the raw material for a species to change gradually over time (Snustad and Simmons, 2000). The most important type of selection practiced in plant breeding is directional selection (Wricke and Eberhard Weber, 1986), where the aim of the breeder is to change the mean in one direction to maximize the genetic worth of the selected population (Fins *et al.*, 1992). Even though Darwin's explanation that changes were brought about by natural selection is accurate, today the changes are accelerated by man applying artificial directional selection (van Wyk and Verryn, 2000).

In principle, no difference exists between natural and artificial selection (Wricke and Eberhard Weber, 1986). However, artificial selection is under human control rather than nature (Wright, 1976), and is the practice of choosing individuals from a population for reproduction, to improve or alter the average genotype of the population, usually because these individuals possess one or more desirable traits (Snustad and Simmons, 2000). Selection is aimed at increasing one or more components of a quantitative trait (e.g. resistance to insect A) without allowing a compensating decrease in other components (e.g. resistance to insect B) (Mayo, 1987).

Existing plantations form a source from which parent trees are selected for a breeding population (the elite population) of above average trees, known as plus trees (van Wyk and Verryn, 2000). These outstanding individuals are selected, tested, cross-bred and brought together in new environments, a process which would not normally have taken place under natural conditions (van Wyk, 1983). However, selection decisions are never uniform. Any particular selected tree may be superior because of the site on which it is growing rather than the genes that it contains. Nevertheless, if a trait is under genetic control, offspring of selected trees should out-perform offspring of average trees (Wright, 1976), hence indicating the ability to change the phenotype by selective breeding. A breeding strategy is designed for a specific population according to the method of selection used by the breeder.

There are a number of different methods available for the identification and selection of the best individuals in a population. The most efficient method of selection is that which results in the maximum genetic improvement per unit of time and effort expended (Hazel and Lush, 1943). All methods of selection can be applied more efficiently when reliable estimates of genetic and economic parameters are available, in order to achieve maximum genetic gain (Cotterill and Dean, 1990). These parameters include the number of traits selected, relative economic values of the traits, heritabilities, phenotypic and genetic correlations between traits, and selection intensity. Although estimates of these parameters are becoming available for some populations, these estimates are in general from young material and in many cases they lack reliability (Stonecypher, 1970) due to the difficulty in the prediction of 'mature' performances for relatively juvenile characters (Arbez *et al.*, 1974). The methods of selection that will be discussed in the following sections are: (a) single trait selection and (b) multiple trait selection. Multiple trait selection may be conducted using one of three different methods, namely, (i) tandem selection, (ii) independent culling and (iii) index selection.

(a)   Single trait selection

Single trait or individual selection is a method where a breeder selects for one trait at a time, and is the process of choosing trees based solely on their phenotypic performance. For example, when considering growth rate, it means the breeder will choose the tallest trees to be parents of the next generation (Wright, 1976). The simplest and least expensive approach many breeders adopt in selecting plus trees is to look at the visual appearance of the trees. However, this may be inefficient because the visual appearance of a tree is often a poor guide to its breeding value (or the genes the tree is carrying) (Cotterill and Dean, 1990) due to environmental factors.

A population's response to selection (R) for a particular trait measures how much the mean of the trait has changed in one generation (Snustad and Simmons, 2000). The population mean ($\overline{X}p$) shifts in the direction of the mean of the selected group ($\overline{X}s$). The difference between these two means is called the selection differential (S) as shown in Figure 1.2.
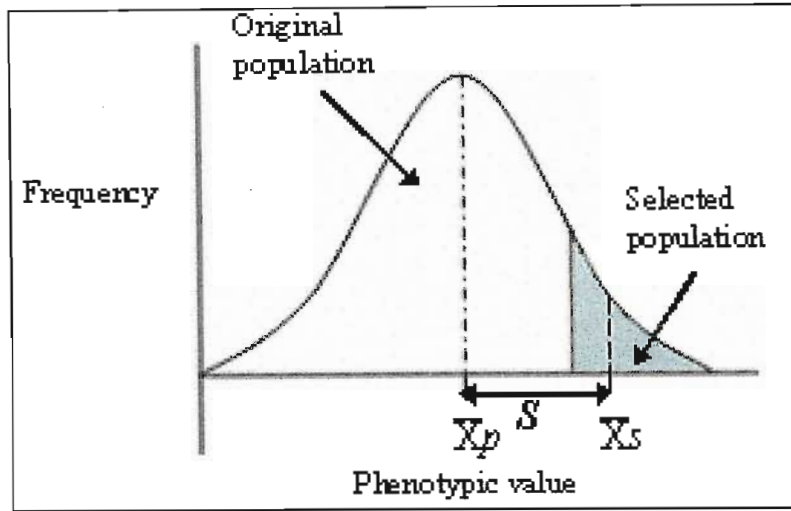
Figure 1.2 Depiction of the selection differential (S) as the difference between the mean of the selected parents ($\overline{X}s$) and the mean of the population ($\overline{X}p$) (modified from Cotterill and Dean, 1990).

Hence, the selection differential (S) is the difference between the mean of the selected group ($\overline{X}s$) and the mean of the population ($\overline{X}p$), represented as (Cotterill and Dean, 1990),

$$S = \overline{X}s - \overline{X}p$$

It is the intention of every breeder to make the selection differential as large as possible.

The general formula for the response to selection (R) is calculated as the product of the additive genetic effects (h^2) and S (Wright, 1976):

$$R = h^2 \times S$$

where, h^2 expresses the proportion of the selection differential that is due to additive genetic effects and thus, should be passed on to the next generation (Cotterill and Dean, 1990). Genetic gain for a single trait in a given population is therefore a function of the amount of variation **present** (σ_p), how much of this variation is genetic (h^2), and how intensively one selects in the population (*i*), represented as (Fins *et al.*, 1992):

$$R = i\,\sigma_p h^2$$

The rate of response to selection may be improved in the following manner:
- Increasing the h^2 by reducing the environmental variation (V_E) through attention to techniques of management, by multiple measurements (when possible), and to a small extent by assortative matings (partners are chosen because they are phenotypically similar) (Falconer and Mackay, 1996).

- Increasing *i* by decreasing the proportion of individuals selected.
- Maintaining the σ of the population. Traits with higher phenotypic standard deviations (σ_p) will show greater responses to selection.
- Increasing the accuracy of selection by using information from relatives.

Cotterill and Dean (1990) stated that it is possible to increase genetic gain for a single trait by using information from relatives and combining all the relevant information into an index. Combined index selection refers to the sum of an individual's own measured value for a trait plus its weighted family mean for the same trait. The index co-efficient weighting (b) placed on the family mean depends on: (1) the relationship of the family (i.e. half- or full-sib relatives) to the individual tree under selection (coefficient of relationship, (*r*)), (2) the accuracy of the family mean (as reflected by the number of progeny per family), and (3) the h^2 of the trait. Individuals having the highest combined index value would be chosen as parents of the next generation. Combined index selection is advantageous for use in selection for traits with low heritability (Stonecypher and Arbez, 1976).

Whether intended or not, the selection of trees for one trait will inevitably cause changes in other traits (Namkoong *et al.*, 1988), as a result of correlations between traits. A study by Pot *et al.* (2002) concluded that single trait genetic gains are possible in *Pinus pinaster*, although improvement of the average of one trait will have consequences for other traits, which may possibly impact pulp and timber production and quality. Hence, correlations between traits tend to complicate single trait selection and it then becomes necessary to include more than one trait in the selection strategy leading to multiple trait selection (Falconer and Mackay, 1996).

(b)   Multiple trait selection

A forest breeder is rarely faced with a situation in which improvement for only a single trait is desired, as the practical value of a tree is almost always affected by several traits. Thus, most tree breeding programmes involve selection for more than one trait as breeders wish to change several traits simultaneously (Stonecypher, 1970). This is known as multiple trait selection, where breeders consider numerous aspects of the tree when selecting parents for the next generation.

The different traits selected for are not likely to be equally important to the breeder or independent of each other (Hazel and Lush, 1943). The dependence of the response of one trait to another and effectiveness of multiple trait selection depends on genetic correlations between traits. The structures of trait correlation and the presence of genetic pleiotropies and linkage will require investigation, as these factors may limit selection progress (Namkoong and Kang, 1990). Furthermore, the rate of progress in each trait is slower as breeders increase the number of traits in the breeding strategy (Fins *et al.*, 1992).

There are three fundamental methods of selection for multiple traits: (i) tandem selection, (ii) independent culling and (iii) index selection.

i.   Tandem Selection

The method of tandem selection involves selecting for several traits, one at a time over several generations, until the desired level of improvement is reached in all traits (Hazel and Lush, 1943). This would mean that the breeder, for example, would ignore wood or form traits for one or two generations of selection while concentrating exclusively on growth. Young (1961) stated that the expected genetic gains by tandem selection are the same as the gains achieved by single trait selection.

Tandem selection is generally not appealing under the long generation intervals common in most forest tree species as it is a prolonged period of time before the desired result for a particular trait can be achieved (Cotterill and Dean, 1990). The cost, effort and time put into the method may even have no significant result at the end of the selection process. However, there are circumstances in which the use of this method may be justified, for example where a single characteristic limits the economic usefulness of a species. Examples of such situations include improvement of disease or frost resistance or rooting ability of cuttings, where these characteristics are fundamental to the economic viability of a species.

A weakness of tandem selection arises when long-term selection on one trait may lead to unacceptable deterioration in other negatively correlated traits (Cotterill and Dean, 1990). In an animal study by Elgin *et al.* (1970) it was concluded that much of the gain made in selection for the traits with the tandem method was lost in selection for other traits in later generations. Consequently, Elgin *et al.* (1970) recommended that tandem selection be used only in cases where one of the other multiple trait selection methods would not be appropriate or where the traits under consideration were positively correlated.

ii.   Independent Culling

In this method a certain level of merit is established for each trait, and all individuals below that level are not selected as parents, regardless of the superiority or inferiority of their other traits (Hazel and Lush, 1943). Therefore, a tree is rejected when its measured value for a certain trait falls below a predetermined standard or culling level.

Independent culling can be carried out either simultaneously or at different times in the one generation (Cotterill and Dean, 1990). Culling at different times may be convenient where traits are expressed at differing ages, for example, in trees an initial culling might be on traits that are expressed at the seedling stage with a second culling on traits expressed later.

Independent culling is seen as an inflexible approach in the sense that trees are culled if their value for one trait falls below a certain standard (culling level), regardless of the individual's merits in other traits (Hettasch *et al.*, 2006). Outstanding trees may be culled which are marginal for some other trait. The practical difficulty with independent culling selection is deciding on appropriate

culling levels for each trait. The possibilities are endless and in practice the culling levels eventually chosen are a matter of trial and error. Traits having the greatest economic importance would usually be subject to more intensive selection and a more severe culling level (Cotterill and Dean, 1990).

iii.   Index Selection

A breeder may be interested in the net genetic worth of an individual aggregated across a number of different traits with varying economic importance. Thus, the breeder would need to combine measurements on all these traits into a single prediction of the aggregate genetic worth of an individual (Fins et al., 1992). This is known as an index. The selection index employed in plant and animal breeding generally refers to a linear combination of observations (phenotypic values), from multiple sources, used to compute a measure for selection (one index value) for each individual available for choice (Henderson, 1963). Several traits are selected for and the simultaneous improvement of all traits is often desired (Namkoong, 1979).

An example of a linear function which may be used to generate an index, integrating the phenotypic or measured values for two traits, is given as (Cotterill and Dean, 1990),

$$I = b_1 P_A + b_2 P_B$$

where

I	is the index value for an individual tree
P_A	is the phenotypic measurements of an individual tree for trait A
P_B	is the phenotypic measurements of an individual tree for trait B
b_1	is the index coefficient for trait A
b_2	is the index coefficient for trait B

The b values or index coefficients used to weight each trait are calculated using heritabilities (h^2), the relative economic weight (a) of each trait, genetic and phenotypic correlations among traits, the number of progeny per family and the coefficient of relationship. Essentially, the measured values of several traits are reduced to one index value so that selection becomes equivalent to selecting for one trait (Cotterill and Dean, 1990). Each individual tree is scored according to this index. Trees having the highest index value are selected as parents of the next generation.

Standard parameters such as heritabilities and genetic and phenotypic correlations as well as juvenile-mature correlations are calculated to present values which may be applicable and useful for constructing selection indices. However, the use of standard parameter values may reduce the efficiency of index selection if the estimated parameters vary substantially from one site to another, as estimates determined for each individual site will be sensitive to changes in heritabilities and correlations with changing environments. For many sites it is not possible to estimate genetic parameters and there is no choice but to use standard estimates. Furthermore, it is important to revise parameter estimates after each generation of selection since levels of genetic variance and covariance may have changed in the new population (Cotterill and Dean, 1990).

The expected genetic response from one generation of selection using an index is given below as (Harvey and Townsend, 1985),

$$R = r_{IG} \, i \, \sigma_g$$

where

R	is the response to selection
r_{IG}	is the correlation between the index and the true breeding value
i	is the selection intensity
σ_g	is the genetic standard deviation

Genetic progress for each individual trait in the index can be estimated, as the magnitude and direction of these changes are not always immediately apparent by inspection. It is necessary to continually monitor the genetic progress made in each trait to ensure that selection on the basis of the index has not resulted in unacceptable change, for example, change in the wrong direction in one or more of the traits (White and Hodge, 1989).

All selection processes require assumptions. As Henderson (1963) pointed out, use of conventional selection index procedures assume that the fixed effects, phenotypic correlations and heritabilities are known. Given certain assumptions, the use of a selection index will result in the best linear prediction (BLP) of genetic worth of the individuals under selection. The following assumptions are also made (White and Hodge, 1989):

1. All individuals for selection have equal amounts of information and equal quality (i.e. balanced data). By equal information, it means that the same genetic parameters (heritability, variances, and correlations) are applicable to all measurements on all individuals and that all individuals are equally tested in every sense (Fins *et al.*, 1992).
2. Economic weights are accurate.
3. Genetic variances and covariances are known.

In any kind of multiple trait situations, a BLP is needed to make correct selection decisions and maximize and predict genetic gain (Fins *et al.*, 1992). This allows the breeder to consider all information available when ranking individuals for selection in a system which is objective and consistent. It is often useful to calculate many alternative indices to test different scenarios about the above assumptions, specifically, to explore ranges in the quality and quantity of information, differences in economic weights and differences in estimates of variances and covariances. White and Hodge (1989) noted that this allows realistic appraisal of the sensitivity of the index to the assumptions made and hence of the selections made based on index values.

Selection indices have several advantages, particularly when:

- The number of traits (*n*) to be improved increases (Hazel and Lush, 1943).
- Negative correlations exist among some of the traits (Stonecypher, 1970; Lin, 1978).
- Selection intensity is low (Young, 1961).

- The index can be revised and improved as new estimates of variances and covariances and other data becomes available (White and Hodge, 1989). Thus, the updated index can be used to examine the possible "mistakes" made using the old index.

Limitations of Index Selection, for the genetic improvement of one or more quantitative characters, still exist. Lin (1978) discussed the limitations of selection indices and noted that although index selection has been used extensively in plant or animal breeding, there are some problems frequently associated with it. These are discussed below.

Index selection reduces the genetic variance and covariances between traits over time and hence changes the phenotypic variances and covariances (Lin, 1978). This could lead to misleading results and incorrect interpretation of the results when the same selection index is used throughout the selection trial.

The index is derived by use of sample estimates. The influence of errors of parameter estimation on the accuracy of the selection index has been investigated by Heidhues (1961), Williams (1962), and Harris (1964). They concluded that errors of parameter estimation would affect the accuracy of the selection index. Similarly, Elgin *et al.* (1970) concluded that difficulties in accurate estimation of various parameters used in the index could cause the index to be less effective than theory would indicate. Furthermore, Namkoong (1969) noted that the error of estimate for genetic parameters is usually high and hence the error of estimate of true genetic value can be very high. Consequently, Namkoong (1969) demonstrated the risks inherent in using poorly estimated parameters for calculating indices. However, Arbez *et al.* (1974) concluded from their study that the selection index method is applicable to forest trees provided that reliable estimates of genetic parameters are available.

Hazel and Lush (1943) stated that the greatest obstacle to index selection is the difficultly of knowing how much importance should be given to each trait making up the index. The availability of reasonably accurate economic information could well be the limiting factor in applying successful multiple trait breeding procedures (Stonecypher, 1970). As a result, the inability to determine precise estimates of economic weights is sometimes cited as a reason to avoid using selection indices (White and Hodge, 1989). Moreover, economic values may change from time to time or vary from one location to another. As Lin (1978) suggested, this shows the necessity for reconstructing the index to manage economic changes. Hence, economic weights should be recalculated by the breeder for each new population, environment or time in which selection is to take place (Cotterill and Jackson, 1985). Change in economic value denotes a change in net merit (the selection goal) thereby reducing overall selection progress based on index selection (Lin, 1978).

In addition to these limitations, as discussed by Lin (1978), a further restriction includes that of negative correlations between traits. If two traits in an index are adversely correlated, then gain in one trait may be made while the other declines in genetic value. In a study on *Pinus caribaea* in

Australia, Dean *et al.* (1986) found that in an index incorporating stem diameter (growth trait), stem straightness (form trait) and branch diameter (form trait), placing a high weight on growth resulted in genetic deterioration in the form traits. White and Hodge (1989) stated that it may be unacceptable to allow deterioration in any trait.

iv.   Comparison between the three methods of multiple trait selection

Most of the literature proclaims that index selection is generally more efficient than independent culling which, in turn, is more efficient than tandem selection (Hazel and Lush, 1943; Young, 1961; Arbez *et al.*, 1974; Stonecypher and Arbez, 1976)

To compare the relative efficiency of the three multiple trait selection methods, the expected genetic gain from each must be calculated (Young, 1961). Falkenhagen (1986) stated that the realized gains are compared with the expected gains, calculated by theory, to check the reliability of an index. Under simplified conditions in which n traits are independently and equally important, and the heritability and standard deviations for each trait are equal, using index selection, the average improvement per generation in any one trait would only be $\frac{1}{\sqrt{n}}$ times as much if selection were directed for that trait alone (Hazel and Lush, 1943). Hazel and Lush (1943) noted that this is probably the basis for the belief that selection is most effective when applied to only one trait at a time. Furthermore, Young (1961) stated that the relative efficiency can be affected by changes in both phenotypic and genetic correlations between traits under selection, since changes in correlations affect genetic gains. In addition, Stonecypher and Arbez (1976) noted that the magnitude of genotype-environment interactions will also influence selection efficiency.

Selection experiments in pigs, for ten years in Norway, showed that there is little evidence that the index selection theory works well, since the index failed to improve two traits simultaneously as was predicted by the hypothesis (Vangen, 1979). Traditional multiple trait selection indices have been less than successful in many forestry cases and Falkenhagen (1986) stated that there is no evidence in forest trees that index selection lives up to its theoretical superiority even to tandem selection.

The success of tree breeding programmes will depend on the breeder's ability to pay attention to several traits. There is always the danger that selection will fall below its maximum efficiency because too much attention is paid to some traits and too little to others (Hazel and Lush, 1943). Breeders should avoid those traits which are unimportant and emphasize those which are most important in producing maximum usable wood per hectare (Stonecypher, 1970). The success of multiple trait selection is also dependent on genetic correlations between traits. When two traits are linked by a strong and unfavourable genetic correlation, simultaneous gains even by index selection will be very low (Arbez *et al.*, 1974). Hence special situations may justify the use of independent culling. Hazel and Lush (1943) stated that independent culling has another practical advantage over index selection in that individuals may be culled for each trait whenever that trait

becomes evident, without waiting until all the traits can be measured. If the selection intensity is high, then independent culling may be more appropriate than index selection because of the relative simplicity of operation (Young, 1961).

Mixtures of tandem, independent culling and index selection may generally be required in breeding strategies. However, tests of these different breeding methods are still needed (Namkoong and Kang, 1990). Stonecypher and Arbez (1976) concluded that the breeding and selection methods available need to be compared in an economic context and prioritized in an overall strategy.

Tree improvement programmes are becoming more and more complex with large amounts of data already collected and available for analysis. Advances in tree improvement and forest genetics have made it possible to make genetic improvement, using various selection strategies, happen more quickly, efficiently and effectively. There are many possible modifications to various breeding strategies and each organization will tailor a breeding strategy to meet the particular objectives subject to the organizational capabilities and constraints.

1.3 LARGE SCALE PRODUCTION OF SUPERIOR GENOTYPES

Tree breeding leads to the identification of individuals, families or clones with the best genetic potential to be selected for production purposes. Improved trees are developed, followed by the mass production of this improved stock on an operational scale. Seed orchards represent the traditional mode of transfer, however clonal forestry production is becoming more common.

1.3.1 Seed orchards

Seed orchards are established to produce seed of a particular origin or source. As described by Hettasch et al. (2006), a seed orchard is a plantation of selected clones or seedling progenies assumed or proven genetically superior, that has been isolated and managed to reduce pollination from genetically inferior outside sources. This seed orchard is intensively managed for mass seed production. Seed orchards are established not only for genetic improvement of specific traits but also to provide quantities of genetically improved seed for operational planting.

Any particular seed orchard has several advantages and disadvantages. The main advantages of clonal seed orchards are (Hettasch et al., 2006):
- From progeny tests, genotypes of seed-producing trees are known.
- Possibility of related mating among seed orchard trees is minimal.
- Seed production begins soon after orchard establishment.

However, the disadvantages of a clonal seed orchard may include:
- Grafting and incompatibility problems.
- Progeny testing must be carried out in a separate operation.

The advantage of a seedling seed orchard is that it first serves as a progeny trial for previously untested families, and after rogueing, it serves as a seed production stand. A further advantage is that since seedlings are used, grafting is unnecessary, making it simpler and cheaper. However, in seedling seed orchards, the genetic gain cannot be expected to be as high as with clonal seed orchards. Furthermore, superior selections may be excluded from this type of seed orchard as they fail to produce seed.

1.3.2 Clonal or vegetative propagation

After many years of seed orchard production, the full realization of genetic improvement is now arriving in the form of clonal forestry (Pait, 2005). Individual trees of many species can be replicated using various forms of vegetative propagation as opposed to seedling stock. Hence, it is possible to replicate the same genotype over many environments. Tropical eucalypts, such as *E. grandis*, and various superior hybrid lines are now routinely propagated as rooted cuttings for plantation establishment in many parts of the world (Doran *et al.*, 2000). In South Africa, approximately 40 % of all eucalypt plantations are clonal, the majority of which are hybrids (Hettasch *et al.*, 2006).

Each generation, the breeding population is used to produce superior individual's genotypes. These individuals must rank extremely high and appear good enough to add to a suit of clones for deployment in operational plantations. Potential new clones should out-perform a standard set of the previous generation's operational clones.

Eucalyptus breeding programmes have aimed to develop clonal forestry to enhance plantation productivity through product uniformity (Denison and Kietzka, 1993). Pait (2005) stated that clonal trials to date indicated that substantial gains in productivity were possible through superior trees being asexually propagated, thus resulting in a more uniform crop. Furthermore, Pait (2005) noted that high heritabilities may be achieved since the additive genetic effects make a considerable contribution towards the phenotypic variance in a clonal population, and the environmental variance may be reduced. Additionally, low levels of GEI may be achieved as clones provide a better estimate of "site by genotype" interaction (Hettasch *et al.*, 2006).

Clonal forestry offers several advantages over traditional seedling establishment practices (Libby *et al.*, 1969; Tuskan, 1997; Pait, 2005):
- Product uniformity, where quality traits are expressed. Uniformity is a great advantage, especially if uniform dimensions are important such as required for mining timber (Hettasch *et al.*, 2006).

- Gains in yield and improved productivity, as the more uniform crop results in a larger percentage of the plantation being at an optimal level for a particular trait (Hettasch *et al.*, 2006).
- Alternative disease management strategies and gains in disease resistance.
- Potential to capture greater amounts of the genetic variation as the clonal technique allows for the estimation of total genetic variation in the population.
- Speed of deployment per breeding cycle.
- Potential to evaluate GEI.

Breeding strategies, selection methods, natural variability and an appropriate usage of wood are of great importance in clonal forestry (Ferreira and Santos, 1995). Hence, there are a number of issues that need to be considered in clonal forestry:

1. The estimation of basic genetic parameters is crucial in determining appropriate strategies for clonal breeding and to predict genetic gains from deploying the best clones (White, 1996).

2. The numbers of clones deployed operationally, as well as the use of single or clonal mixtures, are important as they directly impact the breeding strategy (White, 1995).

3. The fate of an individual genotype is of paramount importance with a clonal stand of trees, since the clone is a single genotype (Fins *et al.*, 1992). Furthermore, Osorio (1999) concluded from an evaluation of seedling and clonal series tests that a lack of genetic correlation and correspondence between seedlings and clones from the same ortets was evident. The results suggested that clones of all promising individuals must be widely tested in the target environment prior to operational deployment.

4. Many traits are measured and considered before selecting individuals as potential clones. The rooting ability of the individual is very important and, other traits considered include, disease score and total radial shrinkage (TRS).

   Clones are required to have a successful record of rooting ability. Juvenility, hedge health and consistent environmental parameters are all critical to rooting success (McRae *et al.*, 1993). Even though operational production of *E. grandis* by rooted cuttings has been done for many years, the biological, genetic and physiological factors that control the sprouting and rooting processes of the cuttings are still unknown (Ikemori, 1990). After a few years of provisional deployment, the screening process on plus selections eliminate particular individuals on characteristics of wood density, rooting capacity requirements and growth and form characteristics.

5. The length of time (normally five years) required to go from a selected tree of seedling origin to a selected clone in sufficient numbers for operational use (Wright, 1995).

6. Disease susceptibility results (especially between 6 and 24 months of age) as the occasional clone which had no disease in the trials of row plot designs, may develop disease in operational plantations (Wright, 1995).

7. There are further costs associated with vegetative propagation as compared to seedling stock. Additionally, clones need more care during planting than routine seedling stock (Doran *et al.*, 2000).

These limitations have restricted the usefulness of cloning for most forest trees. Hence, a current key driver for clonal forestry at present is the development of cost-effective production systems (Pait, 2005).

1.4 CONCLUDING REMARKS

In a tree improvement programme the breeder is first required to improve the genetic properties of a breeding population. This is achieved by estimating the genetic parameters of the population, and subsequently calculating BLP values to improve the mean value, for a particular trait(s), in the breeding population. Superior genotypes in the breeding population are then identified and consequently selected for use in a production population, where mass propagation is used, resulting in the individuals becoming components of commercial plantations.

1.5 AIMS OF INVESTIGATION

This study focused on the use of multiple trait selection, specifically Best Linear Prediction (BLP). The study aimed to examine the following research objectives:

1. To edit a dataset so that a detailed study of the data could be made and for the creation of a standardized dataset, corrected for fixed effects, which could consequently be analyzed.

2. To estimate the population parameters, namely, heritabilities and phenotypic correlations, for use in index calculations.

3. To determine a BLP of individual breeding values in order to rank individuals according to an index value.

4. To predict a breeding population's response to selection in order to observe the change in the population mean for a particular trait for different selection strategies.

5. To compare selection indices, considering different numbers and choice of traits, for commonality among rankings of the top 30 individuals. This was used to evaluate whether a simple index, considering only a few traits, could be equally effective in identifying genetically superior individuals as a complex index considering many traits.

6. To select individual trees, using various selection options, for production and deployment purposes.

7. To determine the effect of population size on the number of individuals that could be selected for a clonal trial.

A representation of the aims of the investigation is illustrated in Figure 1.3.
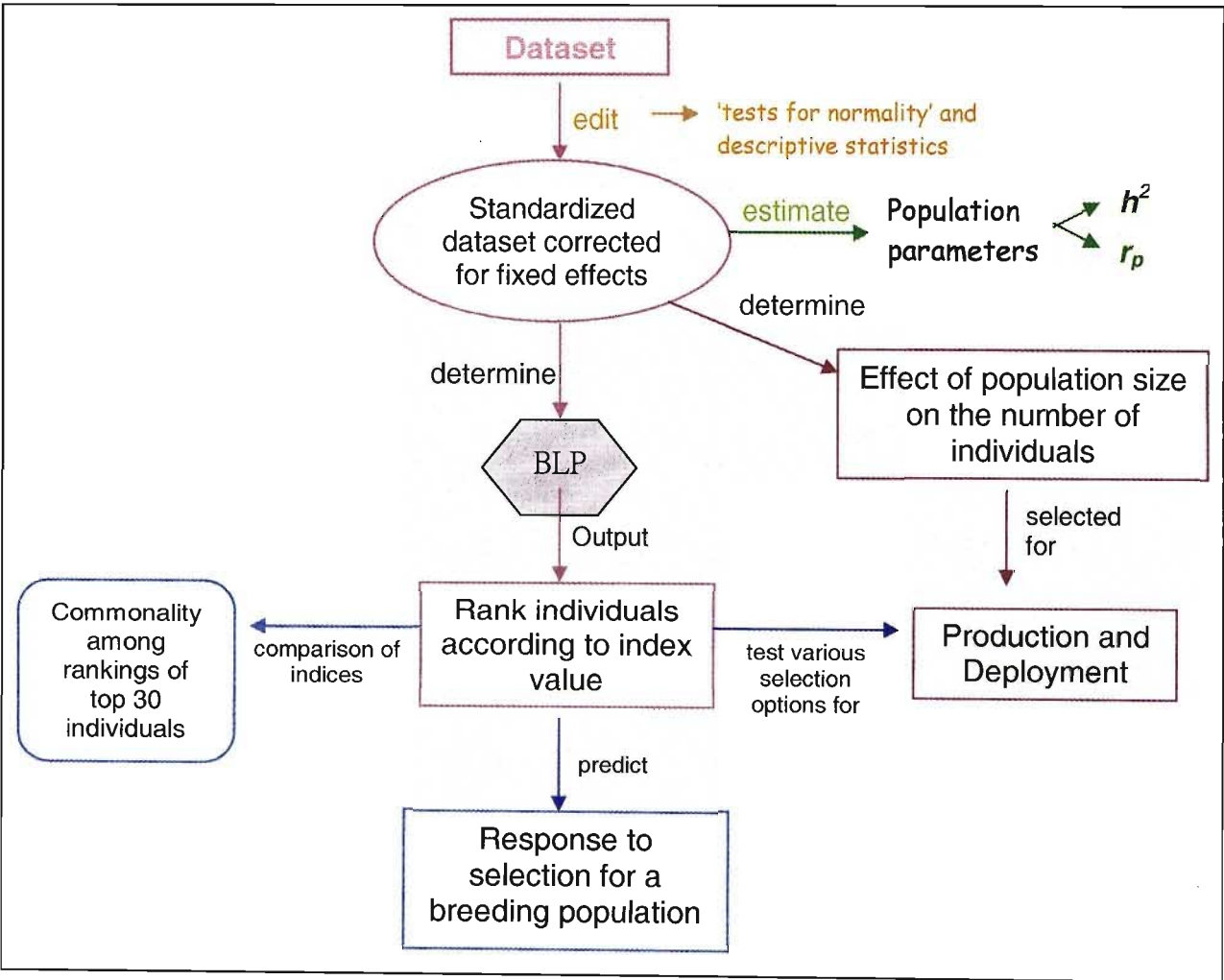
Figure 1.3 Illustration of the aims of the investigation.

2 CHAPTER TWO: MATERIALS AND METHODS

2.1 INTRODUCTION

There are various methods of selection that can be applied to identify desirable individuals in a population. The method used is dependent on whether the breeder is selecting for single or multiple traits. Usually, in a genetic improvement programme, there are several traits that require consideration. The forestry industry has recognized that selection indices are valuable tools that can be used by breeders, to combine all sources of information about the individual, into an index value, which is then used as the basis on which individuals are selected. This research project investigated index selection, using a Best Linear Prediction (BLP) software package, MATGEN® (2003). Subsequently, various selection options, based on particular trait requirements, were applied to the population to select individual trees for production purposes.

2.2 MATERIALS

A dataset, compiled for a wood quality study in 1999, was obtained from The Council for Scientific and Industrial Research (CSIR). The CSIR study investigated the mature wood properties of *Eucalyptus grandis* in a 20-year-old South African progeny trial. The trial was planted with seed imported from Florida and included 773 trees of which there were 90 families and 3 replications.

Relevant data, specifically, the tree number, plot number, replication (rep), family numbers, and the particular traits chosen for this research project, namely, DBH (mm), stem form (1-8 score), splitting (regressed score) and density (kg.m^{-3}), were extracted from the CSIR Wood Quality database and a new Excel® spreadsheet, entitled *Dataset*, was created. The four traits chosen for this investigation were noted in the literature as being important traits used in selection indices for various commercial tree improvement programmes. It should be noted that the traits used in this investigation were measured in the following manner:

➤ *DBH*: measured directly off standing trees at about 1.3 meters above ground using a diameter tape.

➤ *Stem Form*: visually assessed and assigned a subjective score such that the higher the score the straighter the tree.

➤ *Splitting*: assessed in field 72 hours after felling. The higher the corrected regressed values for split score, the lower the split. Conversely, the lower or more negative the regressed split score the greater the splitting observed in that tree (Verryn and Turner, 1999).

➢ *Density*: a gamma ray densitometer was used to measure wood density (Verryn and Turner, 1999).

The data in this study were analyzed using four software packages:

1) The SAS® Institute Inc. Software 9.1. This system is widely used and generally available to forest genetic researchers.

2) The Mixed Model Least-Squares and Maximum Likelihood (LSMLMW) computer programme developed by Harvey (1990).

3) MATGEN® (2003) BLP Version 6.1, developed by Dr Steve Verryn and Prof. Peter Geerthsen (Verryn and Geerthsen, 2003).

4) Microsoft Excel® 2003.

An illustration of the four software packages and their use in this study is shown in Table 2.1.

Table 2.1   Representation of the software packages and their purpose in this study.

Software	Function	Application in this study
SAS®	Statistical analysis software package	Data editing and the creation of a standardized dataset, corrected for fixed effects
Harvey's LSMLMW (1990) programme	Statistical analysis software package	Generates a listing file of the genetic parameters in the population
MATGEN® (2003) BLP	Multi-trait Best Linear Prediction package for unbalanced selection	Simulates forward selection for a single generation, to obtain an index ranking of individuals
Excel®	Performs calculations, analyses information and manages lists in spreadsheets	Data recording and filtering

2.3 METHODS

2.3.1   Data editing for the creation of a standardized dataset corrected for fixed effects

Preceding the analysis of *Dataset*, it was necessary that the data were first formatted and subsequently edited. The aim of data editing was to create a standardized dataset, corrected for fixed effects. The dataset was tested for normality, outliers, missing data and fixed effects. The four stages of data editing used are explained in this section and an overview of the process is illustrated in Figure 2.1.
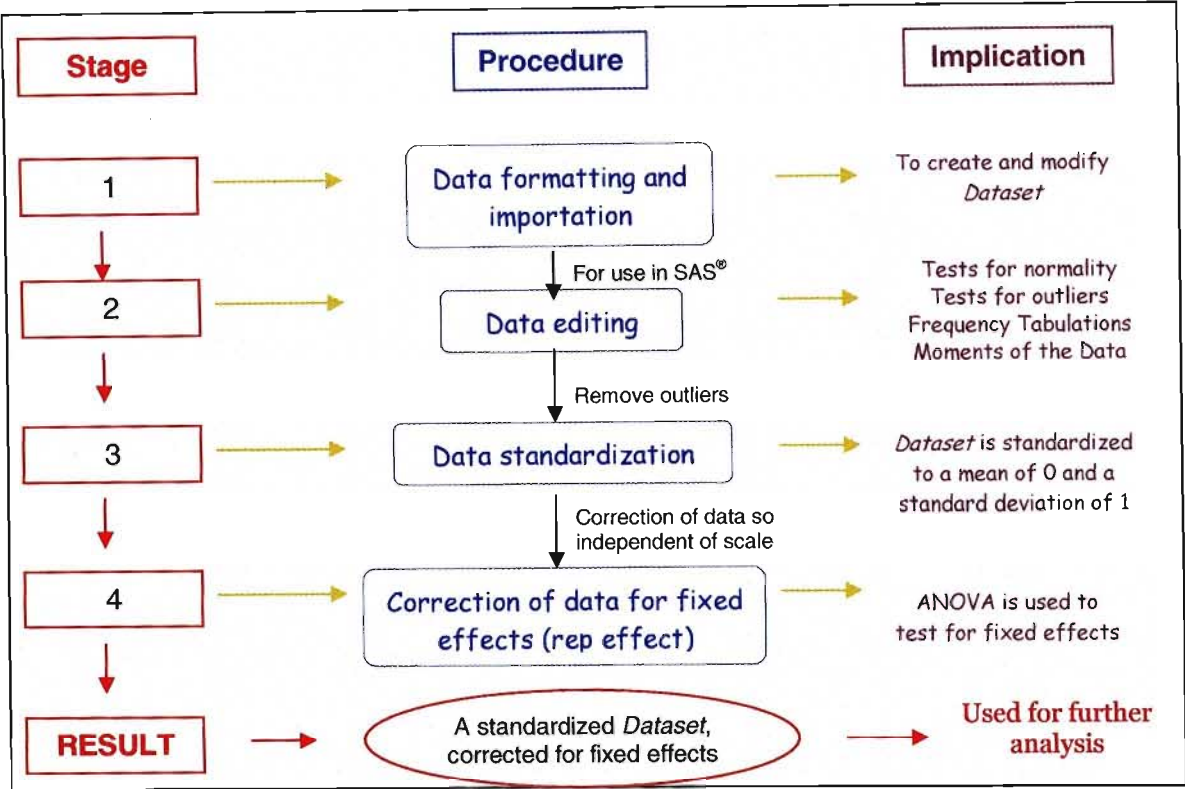
Figure 2.1 Illustration of the four stages in data editing, which result in a standardized dataset corrected for fixed effects.

Stage 1: Data formatting and importation

The Excel® spreadsheet, *Dataset*, was altered into a useable format so that the file could be successfully imported into SAS®. This entailed modifying the spreadsheet to: *(1)* include only data in *columns* and *rows*, *(2)* ensure the headings of particular variables were in the top row, and *(3)* ensure there were no spaces between the columns. The Excel® spreadsheet was saved as a txt file (tab delimited) ready to be imported into SAS® using the procedure outlined in Figure 2.2.

The *Import Data Wizard* in SAS® is an easy way to import the txt file into a SAS® dataset. The following procedure was used:
~   In the SAS® window, select the "**Import Data**" option in the **File** drop down menu.
~   Select the **Tab Delimited File** from the data source list (*Standard data source*).
~   Continue by following instructions from the *Wizard*.

Figure 2.2   Procedure to import spreadsheet into SAS®.

Stage 2: Data editing

A series of tests were conducted on the data before analyses could be performed. These included tests for normality, tests for outliers, frequency tabulations and interpretation of the moments of the data.

(a)   Tests for Normality

The statistical analyses used in this investigation are based on the assumption that the population under analysis is normally distributed. This assumes that the dataset is a random sample from a normal distribution and subsequent results obtained will therefore be valid and reliable (Hettasch *et al.*, 2004). In this study, statistical parameters such as skewness and kurtosis were used to give an indication of the shape of the distribution of the *Dataset's* variables. Furthermore, normality plots including stem-and-leaf plots (histograms), box plots, and normal probability plots were investigated.

(b)   Tests for Outliers

Outliers are observations that are far from the vast majority of observations. Outliers bias the mean and inflate the variance and should be investigated as to why they deviate from the norm. It was thus deemed important to remove outliers that were identified in the dataset during the data editing procedures. It was assumed that these outliers may have been due to incorrect measurement or data capture errors. Missing values were also removed from the dataset so that the values did not obscure the data while trying to calculate genetic parameters.

(c)   Frequency tabulations

Frequency tables depict the number of observations in each unit, for example, each family, rep or plot. This may give an indication of mistakes that could have occurred when breeders capture the raw data. In this study, if the number of observations for a particular unit was incorrect, then the cause of error could be investigated by reviewing the raw data.

(d)   The moments of the data

The moments of the data characterize and summarize values within the dataset. In this investigation, several parameters were calculated which were used to evaluate the data distribution. These parameters included: the number of observations (n), the mean and standard deviation, the skewness and kurtosis, uncorrected and corrected sums of squares, coefficient of variation, variance and the standard error of the mean. Several test statistics surrounding the hypotheses of a zero mean and median were also employed. These statistics were associated with probability values (p-values) describing the weight of the evidence on which to reject or accept the null hypothesis.

The SAS® procedures used for editing the dataset are shown in Figure 2.3.

1.	**PROC FREQ was used to produce frequency tables for any unit. The PROC FREQ statement was written in the *Programme Editor Window* in SAS® in the following format:** PROC FREQ Data=one; tables rep fam/nopercent; Run;
2.	**PROC UNIVARIATE gave simple descriptive statistics for numerical variables.** PROC UNIVARIATE plot plots; Var DBH stem_form den split; Run;
3.	***DATA* steps were used to delete missing values and remove outliers that were identified in the dataset.** DATA one; Set one; If tree_number = 405 then delete; If tree_number = 661 then delete; If split = . then delete; Run;

Figure 2.3 Procedure for data editing in SAS®.

Stage 3: Data standardization

Standardization is a means of transforming data to a format that is independent of scale. In this study, the values of the dataset were transformed to a distribution with a mean of 0 and a standard deviation of 1. This was beneficial as standardizing various trait scores to the same (or similar) mean and variance enabled the researcher, as suggested by Hettasch *et al.* (2004), to assign relative economic weights in selection indices without calculation of economic value per scaled unit. Another advantage of standardizing data, as shown by Snedden and Verryn (2003), was that it became easy to compare relative rankings of the individuals in a selection index, with a score of 0 being equal to the average and +1 indicating one standard deviation more than the average. The SAS® procedures used in the standardization of the dataset are given in Figure 2.4.

1.   **The original variable was first copied, using a DATA step, before running PROC STANDARD in order to keep the original value of the trait. The copy that was made of the variable (in this case, _ssplit_) was then standardized.**

DATA Three;
Set one;
Ssplit=split;
Run;

2.   **The statement specified that the mean of the variable must be equal to zero and the standard deviation must be equal to 1.**

PROC STANDARD mean=0 std=1 out=stand;
Var ssplit;
Run;

3.   **A second DATA step was used to select those variables from the dataset which were of importance and would be needed in succeeding analyses.**

DATA stand;
Set stand;
Keep rep plot tree_number family split ssplit;
Run;

Each of the four traits in this study was standardized according to the same procedures outlined above.

Figure 2.4 Procedure for data standardization in SAS®.

Stage 4: Correction of data for fixed effects

With fixed effects, the assumption was made, as suggested by Hettasch _et al._ (2004), that the _Dataset_ represented a complete sample of the entire population. In this investigation, the role that replication played in the dataset was estimated hence, the replication effect was corrected for. To enable selection for any individual from the entire trial without the preference for a specific replication, which may have given on average higher yields than the other replications, standardized values (mean = 0) were corrected for fixed effects using the following equation:

$$Y_{corrected} = Y_{measured} - \textbf{Replication Mean}$$

In this study, procedures were used to correct for a fixed effect (replication) when working with standardized values. Tree breeding trials are generally considered to be unbalanced datasets due to tree losses, defects and breakages, to name but a few (Snedden and Verryn, 2003). SAS® PROC GLM was used for this fixed effect linear model and was deemed best for unbalanced design, as stated by Snedden and Verryn (2003). The procedure used in this study is shown in Figure 2.5. The objective was to determine whether there was a significant difference between the means of the dependent variable (the standardized trait value) of the replications (replication effect).

---

1. **The least square means (LS-means) of the replication were calculated. The model statement defined the model for ANOVA (one-way model I) and named the dependent variable (*ssplit*) and independent effect (*rep*).**

    PROC GLM data=stand;
    Class rep;
    Model ssplit=rep;
    Lsmeans rep / out=lsmean;
    Run;

2. **Data was required to be sorted prior to merging.**

    PROC SORT;
    By rep;
    Run;

3. **The LS-mean dataset was merged with the original trial dataset *by* replication. The drop statement was used to specify variables that were to be dropped from the dataset. The corrected value was calculated by subtracting the replication LS-mean from the individual standardized trait values.**

    DATA three;
    Merge stand lsmean;
    *drop _name_ stderr;
    By rep;
    Csplit = ssplit - lsmean;
    Keep rep plot tree_number family split ssplit csplit;
    Run;

Each of the four traits in this study was corrected for the replication fixed effect according to the same procedures outlined above.

Figure 2.5 Procedure for correcting for the fixed effects in SAS®.

Consequently, each trait ended up with three values, namely, the raw value (e.g. split), the standardized value (e.g. ssplit) and the corrected value (e.g. csplit). This resulted in a new dataset that had been standardized and corrected for the fixed effect of replication, and could consequently be used in further analyses.

2.3.2   Estimation of population parameters

Obtaining precise and accurate population parameter estimates, such as heritabilities and phenotypic correlations among traits, is fundamental in determining breeding strategies, and for choosing individuals or genotypes for propagation purposes (White, 1987). In this investigation, these population parameters were calculated using SAS® procedures.

(a)   Heritability

Heritability is expressed mathematically as (Falconer and Mackay, 1996):

$$h^2 = \sigma^2{}_A \Big/ \sigma^2{}_P$$

In this investigation, there was a need to estimate the variance components and their ratios (heritabilities) for selection purposes. The estimates of genetic and environmental variances and covariances, as suggested by Fins *et al.* (1992), were required for the development of selection criteria and gain prediction.

i. Testing for replication and family effects

The narrow sense heritability was calculated from among (between) and within family variances. It was first established whether the replication effect was significant, whether families differed significantly and whether there was a significant replication by family interaction (rep*family). Replication is a fixed effect; Family is a random effect; and rep*family is a random effect [Generally, when an interaction involves a random effect, the interaction is declared as a random effect (Hettasch *et al.*, 2004)]. Testing for replication and family effects were done using a two-way ANOVA mixed model, as outlined by the procedure in Figure 2.6.

SAS® PROC GLM was used to test for replication and family effects.

There were two factors (rep and family). The 'family' factor had 90 levels (n=90) while the 'rep' factor had three levels (m=3). The model statement included random effects (family and rep*family). The test option in the random statement requested that PROC GLM determine the appropriate F-test on family and rep*family, treated as random effects.

```
PROC GLM;
Class rep family;
Model csplit = rep family rep*family;
Random family rep*family / test;
Run;
```

Figure 2.6 Procedure for testing for replication and family effects using SAS®.

ii. Estimating variance components

The statistical procedure Restricted Maximum Likelihood (REML) allowed estimation of variance components based on residuals calculated after fitting the fixed effects of the model by generalized least squares. REML estimation is the preferred choice in animal breeding (Henderson, 1984) and has also proven to have better properties for unbalanced data in forestry genetic tests, than other estimators (Huber *et al.*, 1994). The variance components in this investigation were estimated using the SAS® procedure as illustrated in Figure 2.7.

SAS® **PROC VARCOMP estimates the contribution of each of the random effects (family and rep*family) to the variance of the dependent variable (the trait being measured).**

> PROC VARCOMP method=reml;
> Class family;
> Model csplit = family;
> Run;

If the rep*family effect is not significant it does not have to be included in the model.

Figure 2.7 Procedure for estimation of variance components using SAS®.

iii. Calculation of narrow sense heritability (h^2)

In order to calculate the h^2, the additive variance had to be estimated. However, as noted by Squillace (1974), relatedness among individuals can bias the estimate of the additive variance component (σ^2_A) in open-pollinated populations of forest trees. Thus, in this study, the recommended coefficient of relationship in *E. grandis* was increased to ⅓ (0.3 in practice) for this open pollinated population under the assumption of 20% increased "relatedness", as suggested by Verryn (1993).

The family variance was estimated as,

$$\sigma^2_f = 0.3\sigma^2_A$$

Hence, the additive genetic variance was calculated as,

$$\sigma^2_A = \frac{1}{0.3}\sigma^2_f$$

Thus, the heritability of each trait was calculated using the family and error variance estimates as shown by the formula:

$$h^2 = \frac{\frac{1}{0.3}\sigma^2_f}{\sigma^2_f + \sigma^2_E}$$

(b)   Phenotypic Correlations

The phenotypic correlations between traits are parameters that are required for use in index calculations and, in this study, were estimated using the SAS® procedure outlined in Figure 2.8.

1.	**PROC SORT was used to sort the traits by specified variables, namely, rep, plot, family and tree number.**
	PROC SORT data=three; By rep plot family tree_number; Run;
2.	**A DATA step was subsequently used to merge the four datasets into one complete dataset.**
	DATA All; Merge three four five six; By rep plot family tree_number; Run;
3.	**PROC CORR was the SAS® procedure used to calculate the correlation coefficients between corrected variables.**
	PROC CORR data=all; Var csplit cdbh cstem cden; Run;

Figure 2.8 Procedure for estimation of phenotypic correlations between traits using SAS®.

The procedures described were used to obtain estimates of the population parameters in the *Dataset*, namely heritabilities for the four traits and phenotypic correlations between the traits. These were used in the determination of BLP of individual breeding values.

2.3.3   Determination of Best Linear Predictions

With regard to choosing individuals, genotypes or families for breeding and propagation purposes, there has been an implementation of new analytical tools in forestry, based on mixed linear models (Borralho, 1995) that are well suited to handling data from different sources, quality and quantity (White and Hodge, 1989). The sources of information required for selection indices include economic weights, different trait parameters, information from different sites, and information from relatives (Snedden and Verryn, 2003). Best Linear Prediction is regarded highly by breeders as it theoretically has the highest correlation with the (unknown) true genetic value, the prediction of random effects is unbiased and the error of the predicted genetic value is minimized (Henderson, 1984).

The software used for BLP of individual breeding values in this investigation was MATGEN® (2003) (Table 2.1). In this study, fifteen selection indices were run using MATGEN® (2003). These indices varied according to the number and choice of traits included in the index. This was done in order to compare the index ranking of individuals across the various selection indices. The four traits used in the study were numbered for ease of script, as follows:

DBH	=	1	
Stem form	=	2	(abbreviated to stem)
Splitting	=	3	(abbreviated to split)
Density	=	4	(abbreviated to den)

It should be noted that before BLP could be estimated using the MATGEN® (2003) software, Harvey's LSMLMW (1990) programme (Table 2.1) had to be run for each of the fifteen selection indices in order to generate listing files (output files) of the genetic parameters in the population. A summary of the fifteen selection indices that were evaluated in this study is illustrated in Table 2.2. Table 2.2 shows the abbreviations used for the Harvey (1990) output files (H) and MATGEN® (2003) outputs (M), according to the number and choice of traits in the index. For example, *run 9* would be the selection index for a combination of two traits, specifically, stem form (2) and density (4).

The original population of 773 trees, extracted from the CSIR wood quality study, only had 475 individual's measurements for density values. Consequently, any of the MATGEN® (2003) selection indices incorporating *density* could only be applied to the population of 475 individuals. Indices that excluded *density* had a population size of 748, due to missing records removed during SAS® data editing.

Table 2.2   **The 15 selection indices, according to the various combinations of traits, analyzed in this study. The abbreviations, used for the Harvey (1990) [H] and MATGEN® (2003) [M] outputs are shown.**

Run	Selection Index	Trait(s)	Harvey Output	MATGEN® (2003) Output
1	1	DBH	H1	M1
2	2	stem	H2	M2
3	3	split	H3	M3
4	4	den	H4	M4
5	1+2	DBH, stem	H12	M12
6	1+3	DBH, split	H13	M13
7	1+4	DBH, den	H14	M14
8	2+3	stem, split	H23	M23
9	2+4	stem, den	H24	M24
10	3+4	split, den	H34	M34
11	1+2+3	DBH, stem, split	H123	M123
12	1+2+4	DBH, stem, den	H124	M124
13	1+3+4	DBH, split, den	H134	M134
14	2+3+4	stem, split, den	H234	M234
15	1+2+3+4	DBH, stem, split, den	H1234	M1234

The following steps were followed in the creation of index values for individuals in the population:

    a)   Creation of a database file (dbf).
    b)   Creation of a Harvey (1990) output file.
    c)   Initiation of MATGEN® (2003).

(a)   Database file (dbf)

To create the dbf file, the SAS® output containing all the necessary variables and parameters [from the Data editing *(Section 2.3.1)*] was exported into Excel® using the *Export Wizard.* Once, the file had been exported successfully, the numerical columns in the spreadsheet were changed to *number format* with specific decimal places. The spreadsheet was consequently saved as a DBase IV.dbf file. This dbf file was subsequently used in all fifteen MATGEN® (2003) indices.

(b)   Harvey output file

In the SAS® **Programme Editor Window** the *Harvey macro* (obtained from Dr. Verryn of the CSIR) was included with the SAS® input statements so that the macro was run simultaneously with the SAS® analysis. The Harvey macro was used to get the data into a suitable format required by Harvey's LSMLMW (1990) programme as MATGEN® (2003) used the genetic parameters that were accessed directly from the LSMLMW listing file of the data in the calculation of the variance and covariance components as well as the heritability for the trait(s). The variance-covariance information was used, as proposed by Snedden and Verryn (2003), to solve indices and make best linear predictions.

In this study, each MATGEN® (2003) index required a unique Harvey (1990) output file. The procedure used to create the Harvey output files is shown in Figure 2.9. Fifteen Harvey programmes were run, according to the specific MATGEN® (2003) index (see Table 2.2).

To create the Harvey (1990) output file, specific to the particular selection index:

    1. On the main Windows toolbar: *Start > Run.*
    2. A command script file 'cmd' was opened.
    3. Type in: cd\lsmlmw.
    4. At the DOS prompt *C:\LSMLMW>*, **edit all.cnt** was typed in.
    5. The title of the all.cnt file was edited. Save. Exit.
    6. At the DOS prompt type in: **mmod all** to execute the calculations.
    7. The Harvey (1990) output file was saved in the LSMLMW folder and could be viewed in MS Word® (2003).

Figure 2.9 Procedure to create a Harvey (1990) output file required for BLP.

(c)   Initiation of MATGEN® (2003) Version 6.1

MATGEN® (2003) was used to obtain index values for individuals in the population. The procedure followed is explained using the example of M1234. Once the programme had been initiated, the researcher proceeded to the first input screen as shown in Figure 2.10.

The task was to create a new analysis according to the number of traits used in the selection index. In this example, the selection index for four traits (M1234), namely DBH, splitting, stem form and density was created. All parameters and options were entered in interactive screens or windows.

There were three sectors in the first input screen which required information to create a new analysis.

1. *Database to use*: The database (dbf) file that was created from the SAS® output using the *Export Wizard* was selected.

2. *Save file as*: The file was saved as a txt file according to the particular MATGEN® (2003) run i.e. M1234, indicating the analysis for the four traits.

3. *Title*: A title was given to the analysis to indicate what traits were being analyzed in the index. In this example, a suitable title was, BLP for *DBH + Stem + Splitting + Density*.

Figure 2.10   The first input screen in MATGEN® (2003) BLP Version 6.

After successfully creating a new analysis for the four traits, the *Analysis Variables* were entered and determined. In this investigation the three *Analysis Variables* used were as follows:

      i.   Analysis options
      ii.   Phenotypic and genetic matrices
      iii.   Economic weights

All interactive screens were selected sequentially before initiating a run, as suggested by Snedden and Verryn (2003), in order to ensure that all data required was captured.

      i.   Analysis options

The following parameters were inserted using the screen illustrated in Figure 2.11, thereby inputting the details required for the analysis:

a. The analysis indicated was forward selection as individual and family means were considered.
b. The coefficient of relationship of the population was 0.3. This was used to account for the degree of selfing in open-pollinated *E. grandis* (Verryn, 1993).
c. The number of sites considered was 1.

d. The number of traits used in the selection index was entered. In this example, four traits were being analyzed.

e. The Family field name refers to the variable name to identify the families representing the data. In this study, the variable name was 'FAMILY'. This was identified from the dbf file.

f. Field names such as rep, plot, and tree, to identify individuals, as well as the trait values (raw and corrected) were selected to be included in the MATGEN® (2003) listing output. The field names were selected from a list of possible variables in a drop down menu.

g. Fields to test for missing records included those variables where a missing value would indicate an inaccurate record and would therefore be dropped from the dataset.
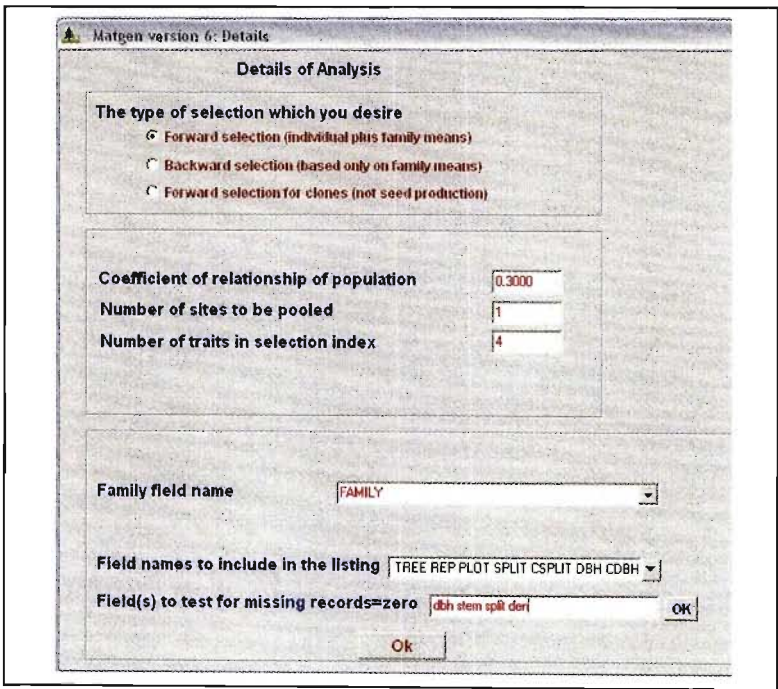
Figure 2.11   Screen for the Analysis Options in MATGEN® (2003) BLP Ver. 6.

ii.   Phenotypic and genetic matrices

MATGEN® (2003) offers two options for the input of genetic parameters. The parameters are either read directly from a Harvey's LSMLMW and MIXMDL (1990) programme output, in which case the user is prompted for the directory and name of the output file. Alternatively, the parameters may be input manually (Snedden and Verryn, 2003). In this study, Harvey's LSMLMW (1990) programme outputs were used. The steps followed to input the information required for the phenotypic and genetic matrices into the three screens is presented in Figure 2.12.

The researcher was prompted for the directory and name of the output file. The Harvey (1990) output file that was specifically created for the MATGEN® (2003) index was selected. In this case H1234 indicates the Harvey file for traits 1 (*DBH*) + trait 2 (*stem form*) + trait 3 (*splitting*) + trait 4 (*density*).
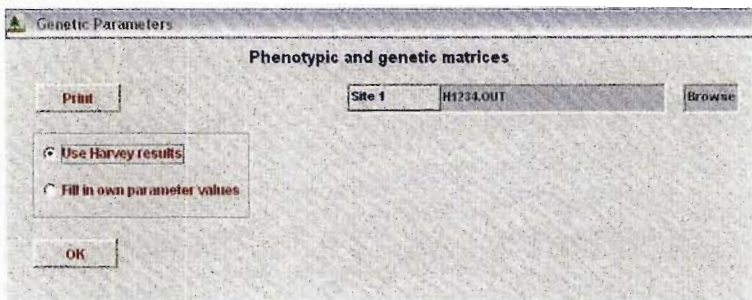
The traits used in the selection index from the database file were selected and matched to the corresponding trait names in the Harvey (1990) file.
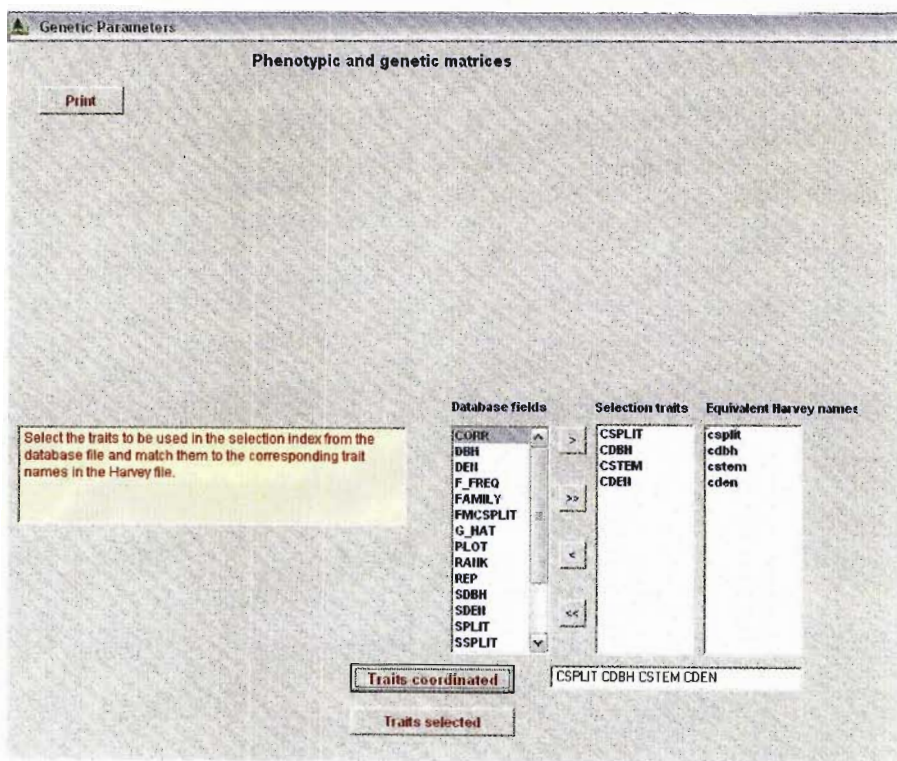
The 'Traits coordinated' button confirmed that the selection traits had been correctly matched.

This resulted in a table showing among and within (co)variance parameters that were extracted from the selected Harvey (1990) listing file.

Figure 2.12   Screens for the phenotypic and genetic matrices.

iii.   Economic weightings

Economic values may reflect the market situation, preferences, retrospective results or simply arbitrarily fixed values (Magnussen, 1990). Assignment of economic weightings are notoriously difficult in tree breeding where long rotations and ever-changing market and technological conditions make predictions of future economic values next to impossible (Namkoong, 1976). Economic weightings are applicable to only the particular population, environment and point in time for which they are estimated. The economic weights in this study were chosen subjectively after some discussion with Dr Steve Verryn. The relative economic weights were as follows:

DBH	0.50
Stem form	0.18
Splitting	0.25
Density	0.07

The economic weights, allocated for each trait were inserted into the interactive screen, as shown in Figure 2.13. These economic weights were then used in the calculation of the selection index.
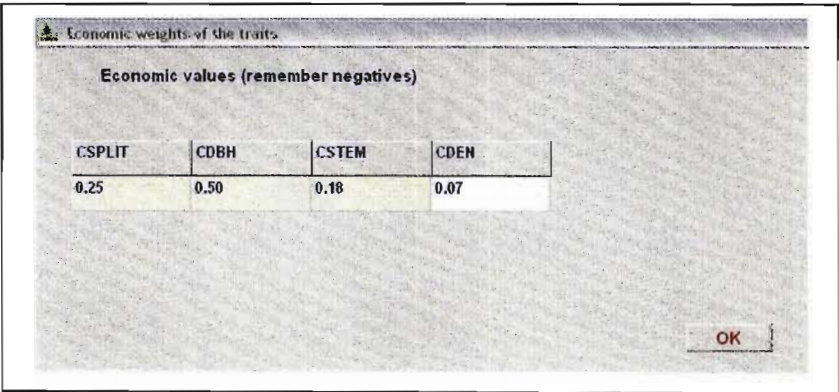
Figure 2.13   Screen for economic weights of the traits.

After the three *Analysis Variables* had been entered the Output was computed by clicking 'Compute...' on the first input screen (Figure 2.10). The Output was viewed by selecting the 'View Listing' button which generated a MATGEN® (2003) output file. The output file was automatically saved as a PRN (extension) file and when the file was selected it automatically opened in MS Word® (2003).

The procedure described, using the MATGEN® (2003) software, was used to carry out forward selection, for a single generation, to obtain an index value for each tree (BLP), for various selection indices applied in the population.

2.3.4 Determination of a breeding population's response to selection for a particular trait

The observed change produced by selection that interests breeders is the change of the population mean. This is the *response* to selection (R). In this study, the response to selection for each trait was estimated and compared across three selection strategies, over one generation. The top 8% of individuals in the breeding population was selected for, according to each of the three selection strategies, namely:

1. **Individual selection** –
   Selection of the top 8% of individuals, for one particular trait in the population, based purely on the individuals' phenotypic values. Hence, individuals with the highest trait values were selected.

2. **Single-trait index selection** –
   Selection of the top 8% of individuals for a particular trait from an index implemented for that particular single trait (*M1, M2, M3* or *M4*). These indices were different to the individual selections as they included family and individual weightings for the particular trait.

3. **Multiple-trait index selection** –
   Selection of the top 8% of individuals for one particular trait from the index implemented for all four traits, *M1234*, in the population. The desired trait was singled out from the other traits in the index and the response to selection for that trait was calculated, based on that trait having been selected for simultaneously with another three traits in the index (multiple-trait selection). This index ranking included: family and individual weightings and economic weightings for all four traits in the population.

Each trait's response to selection (*R*) for the three selection strategies was calculated as follows:

Firstly, the selection differential (*S*) was calculated as (Falconer and Mackay, 1996):

$$S = M_s - M_o$$

where M_s = the mean of the selected population

M_o = the mean of the original population

The response to selection (*R*) was subsequently predicted as (Snustad and Simmons, 2000):

$$R = h^2 S$$

where h^2 = the estimated heritability of the trait.

It should be noted that any prediction of response is valid for only one generation of selection, as the response depends on the heritability of the trait in the generation from which the parents are selected (Falconer and Mackay, 1996).

2.3.5   Comparison between indices for commonality among ranking of the top 30 individuals

The purpose of a selection index is to rank individuals so that the most desirable may be chosen for future breeding (Cotterill and Jackson, 1985). Two methods, namely, *(a)* a rank-correlation matrix and *(b)* a manual assessment, were used to determine the commonality among the ranking of the top 30 individuals between the various selection indices. This was used to determine whether a simple index, with fewer traits, could be used as opposed to a complex index with many traits when selecting individuals for breeding.

(a)   A *rank-correlation matrix* between the rankings of the top 30 individuals from all fifteen indices (Table 2.2) was established in order to determine if the same 30 trees result as the top individuals across all indices. The rank-correlation matrix was generated using SAS® PROC CORR.

(b)   The number of individuals common to the top 30 rankings between the fifteen indices (Table 2.2) was *manually assessed* in order to determine how many of the same individuals were found to be the top individuals in the population when applying various selection indices, according to an assortment of trait combinations (indices for 1, 2, 3 or 4 traits).

2.3.6   Selection of trees for production and deployment purposes

The breeders' goal in tree improvement is to find the most desirable individuals in the population and consequently use them for purposes, such as breeding or production. This requires that the individual trees meet certain trait requirements. Having already looked at the selection of top individuals in a breeding population and the consequent response to selection *(Section 2.3.4)*, the focus was now shifted to the selection of superior individuals for use in clonal forestry.

Clonal forestry, as described by Hettasch *et al.* (2006), refers to the selection of one or more selected and tested individuals which have been bulked up or cloned through vegetative propagation methods such as grafting or cuttings. According to Lindgren (2002), the three main reasons for using clones in forestry are:

- To produce a more uniform product.
- To improve the forest by using genetically better planting stock.
- To offer customer-tailored improved material.

In this study, selection options were implemented in the population according to various trait requirements. This was done in order to determine how many trees meet the trait requirements stipulated, according to the selection option. Those individuals that did succeed in meeting the requirements for all traits could consequently be classed as the superior individuals of the population and were selected by the researcher as potential clones for clonal forestry.

(a)   Selection for all four traits in the population

The MATGEN® (2003) index *M1234* for the four traits, specifically, DBH, stem form, splitting and density (population size = 475) was filtered in Excel®. It is interesting to note that on the **Data** menu in Excel® the user can select the **Filter** option. This allows the user to apply a filter to a range of values in the spreadsheet. Various filters are available; for example, the user may choose to filter for the smallest or largest number in the range, or to filter the range of values for numbers greater than or less than another number.

In this study, each trait was filtered according to a specific percentage specified by a particular selection option. The three selection options applied to the MATGEN® (2003) index M1234 for the four traits were as follows:

   a. Selection of the top 10% of individuals for all four traits in the population.
   b. Selection of the top 20% of individuals for all four traits in the population.
   c. Selection for a 'commercial situation' for the four available traits. This selection strategy was based on what breeders are likely to specify as requirements for the most desirable trees within a particular population. The specific percentages used for the traits are as follows:

DBH	-	the top 15% of individuals in the population
Stem form	-	the top 20% of individuals in the population
Splitting	-	the top 15% of individuals in the population
Density	-	the top 30% of individuals in the population

   It is to be noted however, that in practice, 'a commercial situation' selection would consider more than the four traits assessed in this population.

(b)   Selection for three traits in the population

The four MATGEN® (2003) indices, namely, M123, M124, M134, and M234, each selecting for three traits, were also filtered in Excel® according to a certain percentage. For every index, each of the three traits was filtered to show selection for the top 10% of individuals and a comparison was made across the four indices to determine the number of individuals satisfying the required criteria in each index.

2.3.7 Determination of the effect of population size on the number of individuals who would be selected for production and deployment purposes

It was proposed that a model could be formulated to show a trend between the size of a population and the number of individuals within that population meeting specific trait requirements. The original population of 773 trees, extracted from the CSIR wood quality study, only had 475 individuals' measurements for density values. Consequently, the selection index, selecting for four traits, *M1234*, could only be applied to the population of 475 individuals. After the 'commercial situation' selection option was implemented on the population of 475 trees *(see Method 2.3.6)*, only one individual was found to satisfy the requirements for all four traits. In order to compare the results of this population to a larger population, it was thus deemed necessary to generate a hypothetical population and evaluate whether a larger population would result in more individuals satisfying the trait requirements for a 'commercial situation' selection option.

(a) Generation of a hypothetical population

A hypothetical population of 1000 individuals, consisting of 100 families with 10 individuals per family was generated. This population mimicked the real dataset as it was simulated using the family variance components of trait means from the real data. The populations could then be assumed to have the same genetic parameters and environmental conditions, and could thus be compared on the basis of population size.

i. Determination of family variance components

The total observed phenotypic variance (σ^2_T) within a population is partitioned into two components (Lessells and Boag, 1987):
1. Between families, σ^2_b – Variance of means of groups (families) about the population mean, μ.
2. Within families, σ^2_w – Variance of individuals about a group (family) mean.

The hypothetical population was generated using the between (σ^2_b) and within (σ^2_w) family variance values, for each trait, from *Dataset*. Variance components were derived from a one-way ANOVA using SAS® procedures. For each of the four traits, the procedure in Figure 2.14 was implemented on *Dataset* to determine the family variance components.

1. The replication standard deviation was calculated in order to standardize the data and remove the replication effect.
2. The adjusted means (for the effect of replication) for the trait were calculated.
3. The arithmetic means for all effects specified, in particular the mean effects for the variable 'family' were computed using the SAS® procedure PROC GLM (ANOVA).
4. The ANOVA output was analyzed to determine the variance components, and the subsequent standard deviations. These variance components were calculated from the mean squares in the ANOVA as:

$$\sigma^2_w = MS_w$$

MS_b is the mean square between groups
MS_w is the mean square within groups

and

$$\sigma^2_b = (MS_b - MS_w) / n_o$$

n_o is a coefficient related to the sample size per group in the ANOVA. If group sizes are not equal, as in this case, then n_o is smaller than the mean group size, $\bar{n}$

The value of n_o is calculated as:

a is the number of groups
n_i is the sample size in the ith group
(Lessells and Boag, 1987).

$$n_o = [1/(a-1)] \cdot \left[\sum_{i=1}^{a} n_i - \left(\sum_{i=1}^{a} n_i^2 \Big/ \sum_{i=1}^{a} n_i \right) \right]$$

Figure 2.14   Procedure to determine the family variance components for each trait in the population.

ii.   Using between and within standard deviations to construct the population

For each trait in the hypothetical population, 100 family means were created, using the between standard deviation (σ_b), and then, based on the particular family mean, 10 individuals within that family were generated using the within standard deviation (σ_w). This may be represented in Figure 2.15, using *DBH* as an example:

| (1) Adjusted mean (DBH) | + | σ_b | → | Generation of random numbers representing **100 family means** |
| (2) Mean for Family #1 | + | σ_w | → | Generation of random numbers representing 10 individual trait values with that family |

Figure 2.15   Diagrammatic representation of the construction of the hypothetical population using the between and within standard deviations.

An Excel® function (*NORMINV*) was used to create the population, based on the adjusted means for the traits. The formula is represented as:

NORMINV (probability, mean, standard deviation)

The spreadsheet drew a random cumulative probability [*RAND()*] from a distribution whose mean was given by a particular cell and whose standard deviation was given in another cell. The probability was converted into an actual data point by the *NORMINV* function. This formula was then copied down the column, consequently computing values for any particular number of individuals.

Generation of Families, using the between standard deviation

An extract from the spreadsheet *Families* is shown in Figure 2.16. The 100 family means for each trait were computed, based on the adjusted mean of the trait. *Column A* represented the trait DBH; *Column C*: Stem form; *Column E*: Splitting; and *Column G*: Density. *Cell A1* (the mean for family #1) was calculated using the formula =*NORMINV(RAND(),L2,M2)*. The formula indicated that the probability was specified to be a random number; the mean was specified in *cell L2*; and the between standard deviation was given in *cell M2*. This function was then copied down for 100 rows, thus representing the 100 family means for *DBH*. Similarly, the 100 family means for the other traits were created.
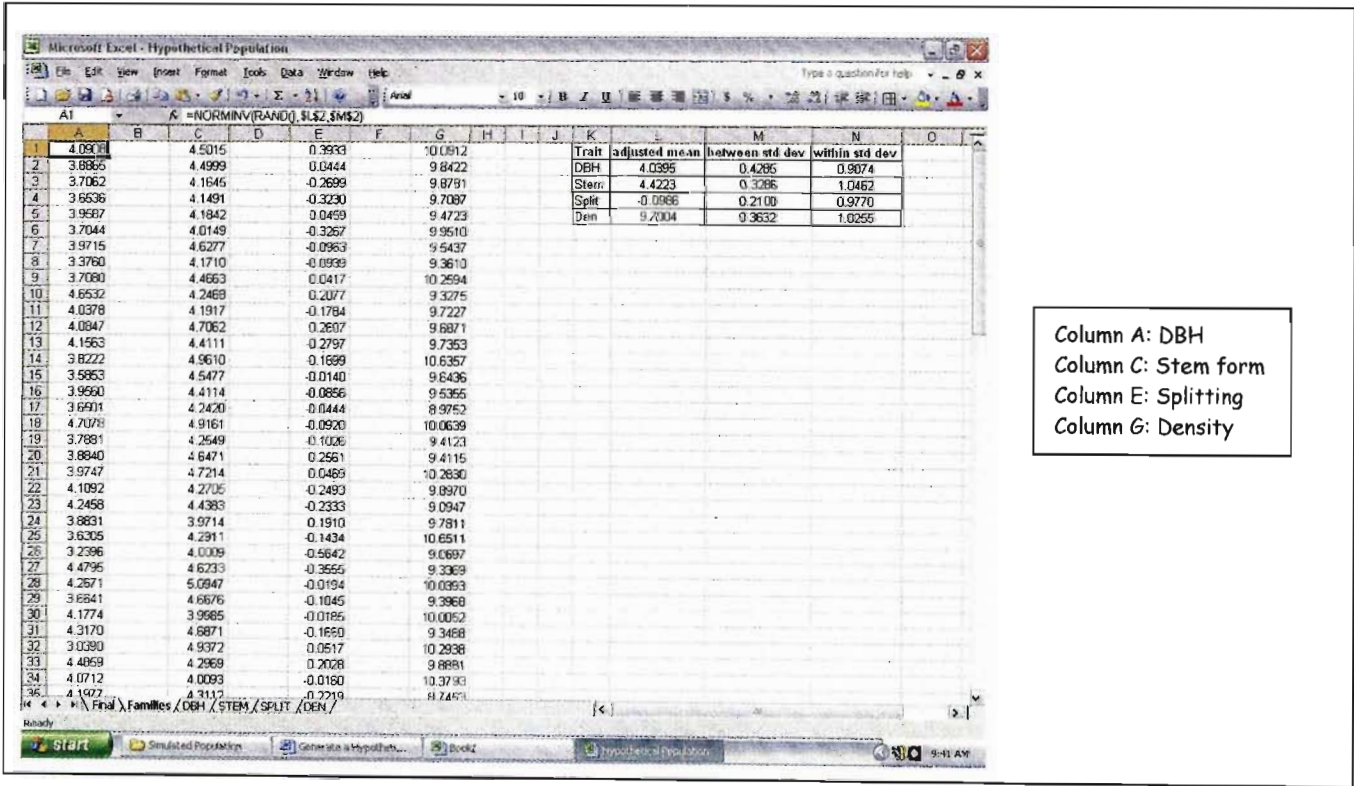
Figure 2.16 **Extract from the spreadsheet *Families* showing the generated family means for the four traits.**

Generation of Individuals within families, using the within standard deviation

An extract from the spreadsheet *DBH* is shown in Figure 2.17. The final phenotypic values of DBH (measured in mm) for the 1000 individuals in the hypothetical population were generated. *Column A* represented the family means for each family. *Column B* showed the randomly generated values for the individuals within each family. *Cell B1* (an individual's value for the trait, within family #1) was calculated using the formula =NORMINV(RAND(),A1,M2). As the formula indicated, the probability was specified to be a random number; the mean was specified in *cell A1* (the family mean for 10 individuals within that family); and the within standard deviation for DBH was given in *cell M2*. This formula was then copied down 10 rows, to account for 10 individuals within the family. For every family mean the formula to generate the 10 individuals within that family was edited to ensure that the correct family mean was used. For example, in *cell B11*, the probability was specified to be a random number; the mean was specified in *cell A11* (the family mean for those 10 individuals); and the within standard deviation was given in *cell M2*.
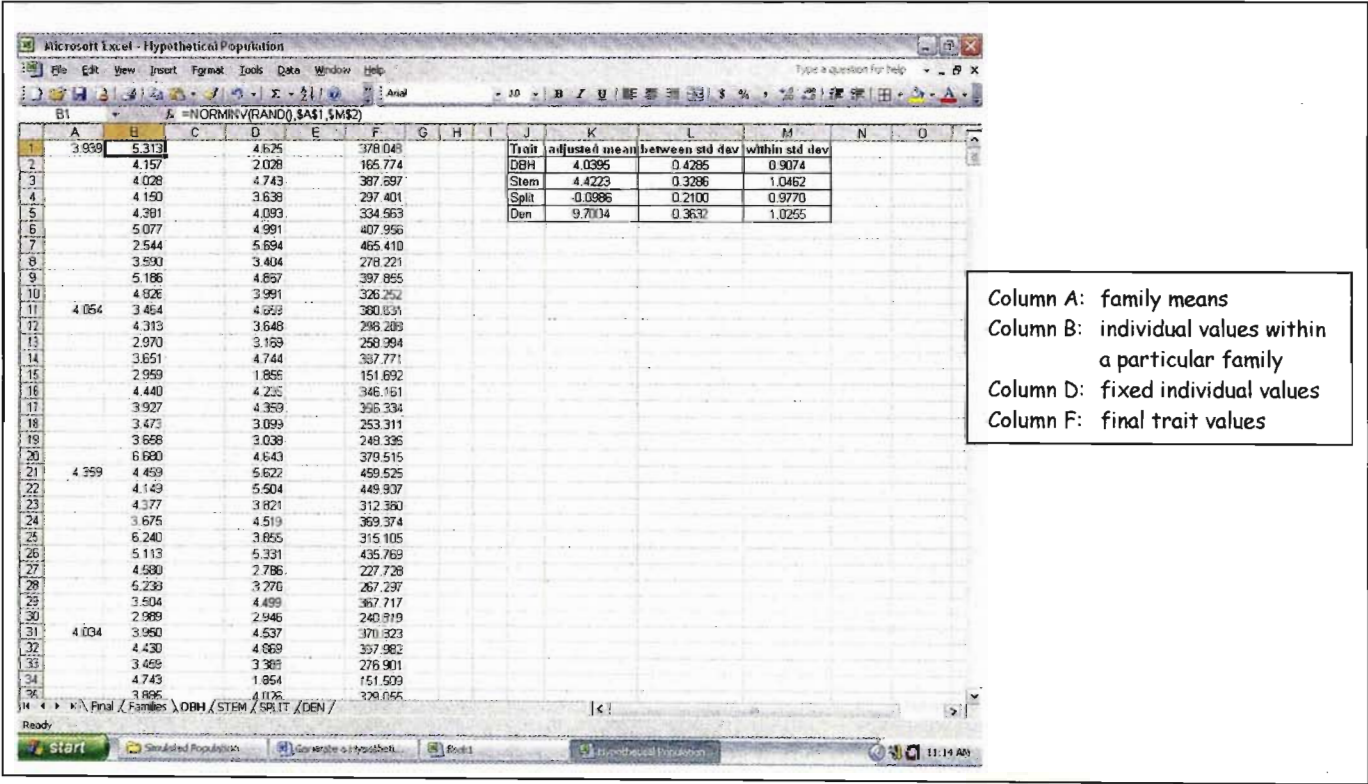
Figure 2.17 **Extract from the spreadsheet *DBH*, showing the generated individual phenotypic values for the trait DBH.**

When all 1000 DBH values had been generated, *column B* was copied, fixed and pasted as *column D*. As shown in Figure 2.17, *column D* represented the individual DBH values for 1000 individuals, based on the adjusted trait mean (adjusted for the effect of replication). The mean of the adjusted population for the particular trait DBH was subsequently calculated and compared to the mean of DBH in the *Dataset*. The computed DBH values were then adjusted by a factor to

relate back to the original population mean in the *Dataset*. The final hypothetical population values for DBH were represented in *column F*. Similarly, individuals within families were generated for the three other traits in the population. This resulted in the final hypothetical population of 1000 individuals, each with phenotypic values for the four traits, namely, DBH, stem form, splitting and density, as shown in Figure 2.18.
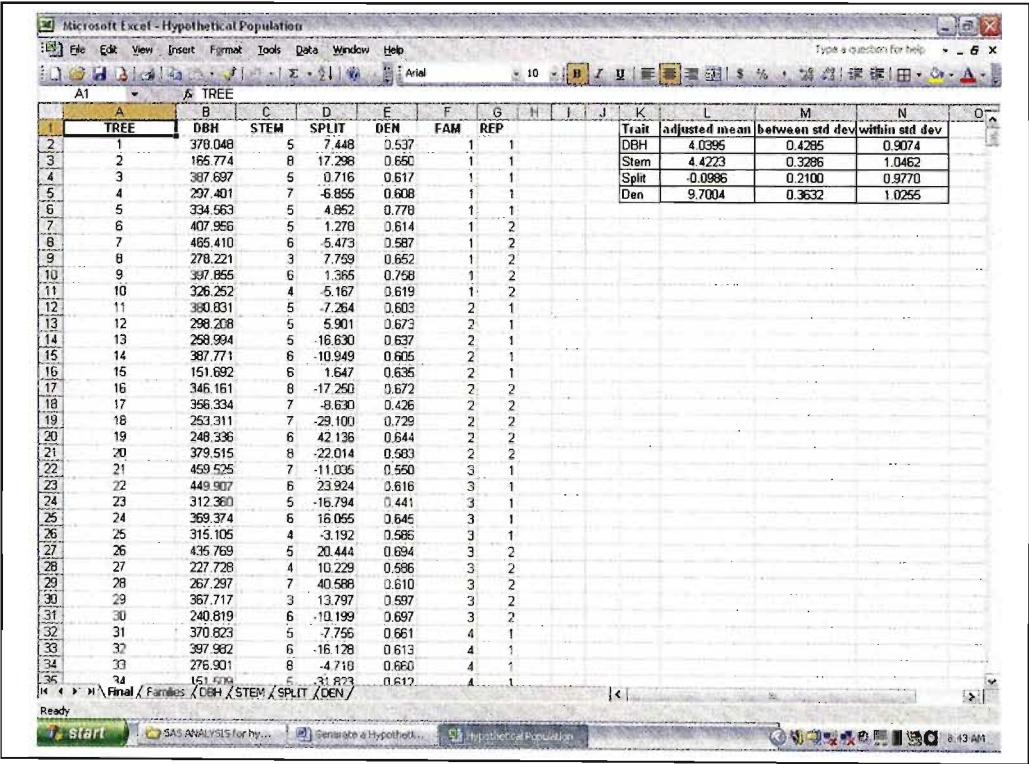
Figure 2.18   **Extract from the spreadsheet *Final* showing the final trait values for all the traits in the hypothetical population.**

A *replication* variable was required in the hypothetical population for further analyses using the statistical packages. Subsequently, individuals were categorized into their families under the variable *FAM*, as well as being allocated as a replication within a family, under the variable *REP* (Figure 2.18). The first five individuals within the family were assigned as Rep 1 and the last five individuals were assigned as Rep 2.

The hypothetical population underwent the same process of analysis as *Dataset*. The procedure is summarized in Figure 2.19 and details may be found in Sections 2.3.1 – 2.3.3.

---

1. Data editing for creation of a standardized dataset corrected for fixed effects
   Formatting and importation
   Editing
   Standardization
   Correcting for fixed effects

2. Estimation of population parameters
   Phenotypic correlations
   Heritabilities

3. Determination of BLP
   DBF file was created
   Harvey LSMLMW (1990) output was created
   Index selection (BLP) for the four traits in the population

---

Figure 2.19   Summarized process of analysis for the hypothetical population.

(b)   Selection of trees for production and deployment purposes

In contrast to the *Dataset*, only one selection option was implemented in the hypothetical population. The hypothetical population (M1234HP) was filtered in Excel®, in an equivalent manner as the Dataset, according to the requirements for the 'commercial situation' selection option, specifically:

DBH	-	the top 15% of individuals in the population
Stem form	-	the top 20% of individuals in the population
Splitting	-	the top 15% of individuals in the population
Density	-	the top 30% of individuals in the population

This allowed for a comparison to be made between the two populations on the basis of population size.

3 CHAPTER THREE: RESULTS

3.1 INTRODUCTION

This investigation focused, firstly, on the statistical analysis of the dataset from the CSIR, and secondly, on index selection, using MATGEN® (2003) BLP. After individuals were ranked using the index, selection options according to various trait requirements were implemented, and the selection of trees for production purposes was carried out. The parameters estimated from the CSIR dataset were then used to generate a hypothetical population, which was used to determine the effect of population size on the number of individuals that could be selected for production purposes.

3.2 DATA EDITING

The four traits (DBH, stem form, splitting and density) in this study were tested for normality and descriptive statistics were calculated to characterize the observations made on the population. This population was described by its phenotypic mean (μ) and standard deviation (σ) for each trait under consideration in the *E. grandis* population (Table 3.1).

Table 3.1 **Estimates of phenotypic means (μ) and phenotypic standard deviations (σ) for *E. grandis*.**

Trait	Phenotypic mean (μ)	Phenotypic Standard Deviation (σ)
DBH (mm)	344.89	85.52
Stem form (1-8 score)	5.52	1.27
Splitting (regressed score)	-2.85	28.50
Density (kg.m^{-3})	600.24	61.99

An extract of the SAS® output from the PROC UNIVARIATE statement, for the trait DBH, is shown in Figure 3.1. As interpreted from the moments of the data in the output (Figure 3.1), the variable DBH deviated slightly from normality since, for normal data both skewness and kurtosis would be zero. Skewness is a measure of the tendency for the distribution values to be more spread out on

one side than the other, thus, deviating from the normal distribution bell-shaped curve. A positive value for skewness, as in this case, indicated that the data values greater than the mean (on the right hand side of the mean if the data is visualized in a frequency distribution) were more spread out than the values that were smaller than the mean (on the left of the mean). Kurtosis is a measure of the shape of the distribution of the values. The negative value for kurtosis in this case, signified that the majority of the values were situated around the mean (a platykuric data distribution).
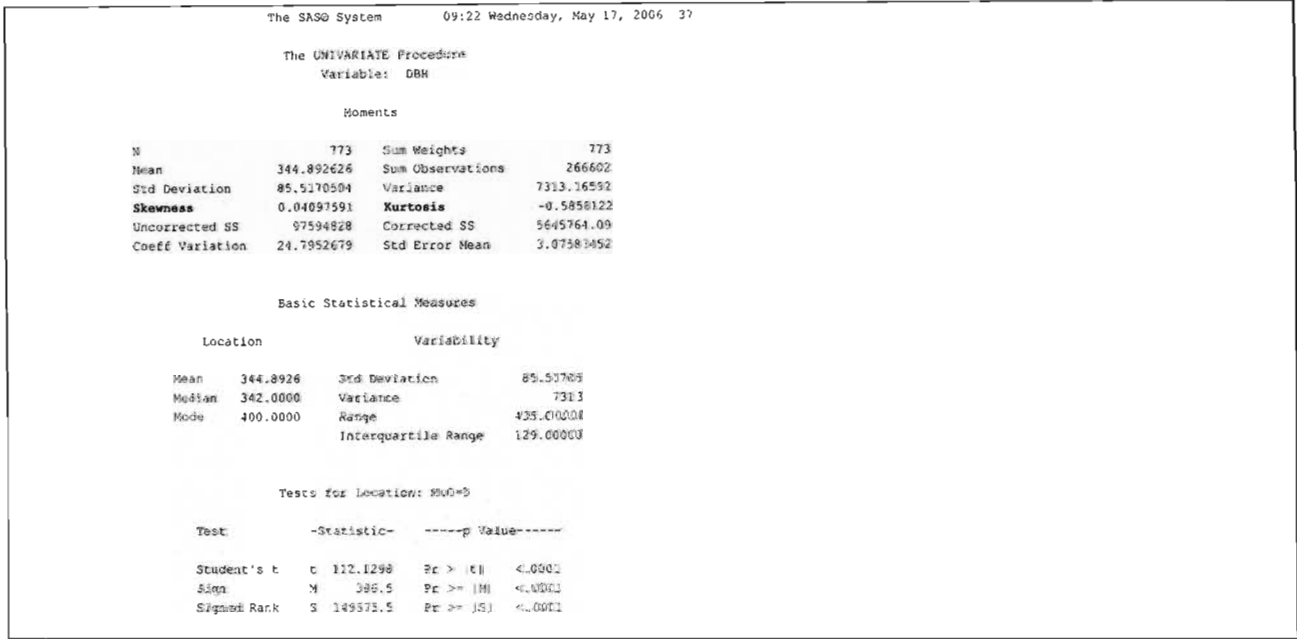
Figure 3.1 Illustration of the moments of the data and basic statistical measures for the trait DBH from an extract of the SAS® output.

The distribution of a population is a descriptive measure of the variability of individual data values around the mean of the population. In a dataset that has been standardized, the values of the dataset are transformed into a format that is independent of scale, with a mean of 0 and a standard deviation of 1 (Hettasch *et al.*, 2004). At the 5% level of significance, the hypothesis that the population mean and median is zero, is rejected if the p-value is smaller than 0.05. In this investigation, as shown in Figure 3.1, the p-values were <0.0001 indicating that the mean and median of the population for the trait DBH were not zero. This stands to reason as the data had not yet been standardized.

Outliers in a dataset can be identified from a listing of the five lowest and five highest observed values and, consequently, the range for the trait can be determined. These values are shown in Figure 3.2.
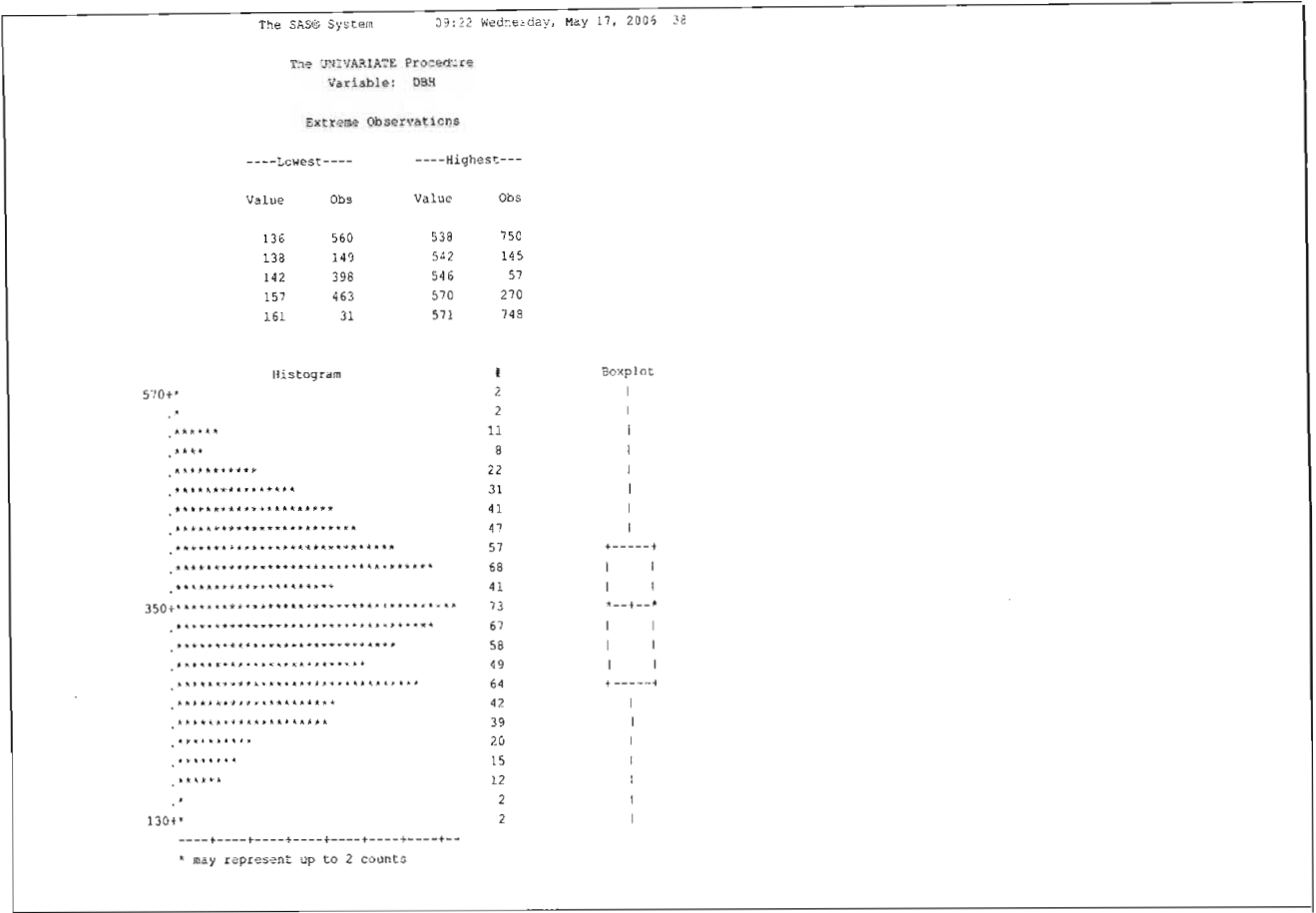
Figure 3.2 Illustration of the five lowest and five highest observations, a histogram and a boxplot, for the trait DBH, from an extract of the SAS® output.

The histogram (stem-and-left plot) represents a frequency distribution of the data (Figure 3.2). The boxplot complements the histogram by providing information on the shape, symmetry, and variability of the data distribution. The histogram and the boxplot are illustrations used to represent 'tests for normality' (Hettasch *et al.*, 2004). In this study, no outliers were identified and a very similar mean and median were documented in the boxplot for the trait DBH. The normal probability plot (Figure 3.3) represents data values by asterisks (*), while the standard normal distribution is represented by plus signs (+). If a large number of plus signs are visible, it indicates a non-normal distribution. These results indicated that the trait DBH deviated slightly from normality.
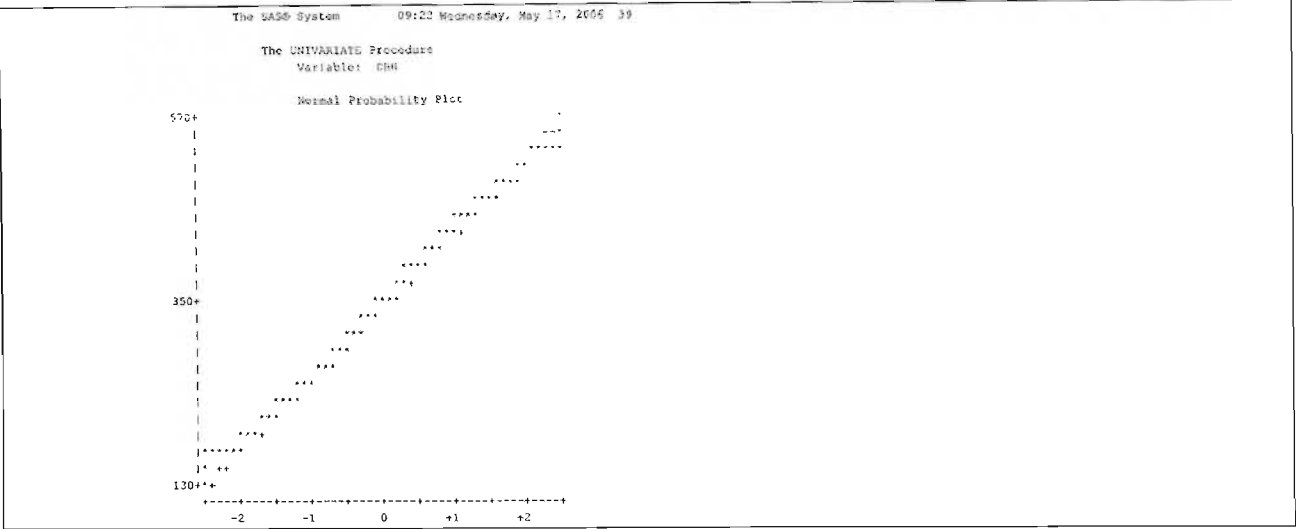
Figure 3.3 Illustration of the normal probability plot for the trait DBH, from an extract of the SAS® output.

The other three traits in *Dataset*, namely, stem form, density and splitting were analyzed in the same way as the trait DBH. The output for *splitting* identified two extreme outliers. On review of *Dataset*, these values were consequently removed as it was predicted that the outliers would be likely to obscure the estimation of the population parameters that were still to be calculated. Negative skewness and kurtosis values were noted for *stem form* indicating that the data were skewed to the left of the mean and that the majority of values were situated around the mean. This consequently indicated a deviation from normality. Positive skewness and kurtosis values were observed for *density*, indicating that the data were skewed to the right of the mean and that, compared with the majority of the density values, some values were situated far from the mean (a leptokurtic data distribution). This therefore indicated a deviation from normality.

The data for all four traits in this population were successfully edited, standardized and corrected for the fixed effect of replication. *Dataset* could consequently be used in further analyses, such as the estimation of population parameters.

3.3 ESTIMATION OF POPULATION PARAMETERS

Population parameters, specifically, the heritability of each trait and the phenotypic correlations between traits, were estimated. These estimates were required in order to calculate BLP values for individual trees.

3.3.1 Heritability

The heritability (σ^2_A/σ^2_P) describes the degree of resemblance between relatives hence, expressing the extent to which phenotypes are determined by the alleles transmitted to the offspring from the parents (Falconer and Mackay, 1996). Heritability is therefore of the greatest importance in breeding programmes. The family and error variance components for each of the traits in *Dataset* were estimated using the SAS® procedure PROC VARCOMP. Thereafter, the narrow-sense heritability for each trait was calculated (Table 3.2).

Table 3.2   The narrow-sense heritability estimates for the four traits in *Dataset*.

Trait	Estimated heritability (h^2)
DBH	0.600
Stem Form	0.401
Splitting	0.214
Density	0.492

Since DBH had the highest heritability, it was expected to show the greatest response to selection followed by density which had the second highest heritability. Splitting was noted to have the lowest heritability and was thus expected to show the smallest response to selection in this investigation. In this population, DBH had a higher proportion of its phenotypic variance made up of additive genetic variance than the other three traits. It would thus be possible to confidently select individuals with good phenotypes because a good phenotype would be a reliable predictor of a good genotype (Falconer and Mackay, 1996).

3.3.2 Phenotypic correlations

The standardized association between two traits that can be directly observed is the correlation of phenotypic values. Many quantitative traits are correlated with others and this is determined from measurements of the two characters in a number of individuals of the population. Pearson

correlation coefficients, between the corrected values for the four traits in *Dataset*, were estimated using the SAS® procedure PROC CORR (Table 3.3).

Table 3.3   **The phenotypic correlations (r_P) between the corrected values for the four traits in the population.**

Trait	Phenotypic Correlations			
	DBH	*Stem*	*Split*	*Den*
DBH	1	0.834	-0.086	0.016
Stem		1	-0.153	0.060
Split			1	0.033
Den				1

The results (Table 3.3) indicated a very high positive phenotypic correlation (0.834) between DBH and stem form. This implies that a tree with a large DBH would be likely to have a high score for stem form. However, it should be noted that these results apply to this *E. grandis* population alone. The results also illustrated that the phenotypic correlations between the other traits were close to zero. A zero value for a correlation means that the two traits under consideration vary independently and that there is no association between them (Hettasch *et al.*, 2006).

3.4 DETERMINATION OF BEST LINEAR PREDICTION

An index, which may be used in a selection process, provides a weighted score for individuals, combining economic values, heritabilities of several traits and information from relatives. The selection index is obtained through a multiple regression equation (see Chapter One, Section 1.2.2, pp 16-20).

The MATGEN® (2003) outputs displayed the following information:
   a. *Genetic Input Summary* which accounted for individual heritabilities, among and within family (co)variances and economic weights.
   b. *Summary Statistics* which included a summary of the parameters used for the index. This included the standardized means of the traits, which were very close to zero, and the total number of observations in the analysis.
   c. *Listing of ranked individuals* that comprised of the default variables (a summary of the parameter inputs) and a ranked selection list.

Forward selection for a single generation was conducted using MATGEN® (2003) and resulted in an index value for each tree (BLP), for the various selection indices applied to the population (see Table 2.1). For each selection index, the *listing output* ranked individuals according to their index

value (BLP). Individuals with the highest ranking were assumed by the researcher to be the most desirable trees in this population. An example of a MATGEN® (2003) output, for the index M1234, is shown in Appendix A.

3.5 THE BREEDING POPULATION'S ESTIMATED RESPONSE TO SELECTION

Selection for a certain trait over a number of generations will result in a gradual improvement in the population mean of that trait, provided all parameters, such as the heritability (h^2), the genetic variation of the population (σ) and the selection intensity (i) are favourable. It will, however, also result in a decrease in the additive genetic variance in the population, because selection acts on and decreases the genetic variation of a trait. This will slow the response to selection until eventually there will be too little genetic variation left to gain any significant response (Lynch and Walsh, 1998).

In this study, each trait's response to selection for one generation was calculated according to three selection strategies, namely, individual selection, single-trait index selection and multiple-trait selection. Table 3.4 shows the estimated response to selection for each trait in the population in relation to the particular selection strategy.

Table 3.4   Estimated responses to selection, R, for the four traits in the population, using three selection strategies.

	Trait			
Selection Strategy	*DBH*	*Stem Form*	*Splitting*	*Density*
Individual	92.2160	0.8004	6.2048	0.0625
Single-trait index	84.6244	0.7472	4.9676	0.0497
Multiple-trait index	84.6644	0.5618	2.9680	0.0229

The results showed that the expected response to selection for each trait was greatest when using the individual selection strategy as opposed to the trait being selected in either type of index selection strategy. Furthermore, for three of the traits in *Dataset* (excluding DBH), the selection responses to the single-trait index strategy were better than the responses from the multiple-trait index selecting for four traits (Table 3.4).

Each trait in this population was measured in different units and it was therefore difficult to compare the relative response to selection for the four traits. For example, it was difficult to compare the selection response of DBH with the response of density because *mm* was being

compared to $kg.m^{-3}$. The percentage increases in each trait for each of the three selection strategies, is illustrated in Figure 3.4. Due to the nature of the values for splitting, the percentage increase for the three strategies was unable to be calculated. Consequently, the graph (Figure 3.4) provides a perspective of the trend for the traits in this investigation, over the range of selection strategies. This allowed for comparisons to be made between the traits, because the response for each trait was standardized into a percentage and these percentages were then comparable.
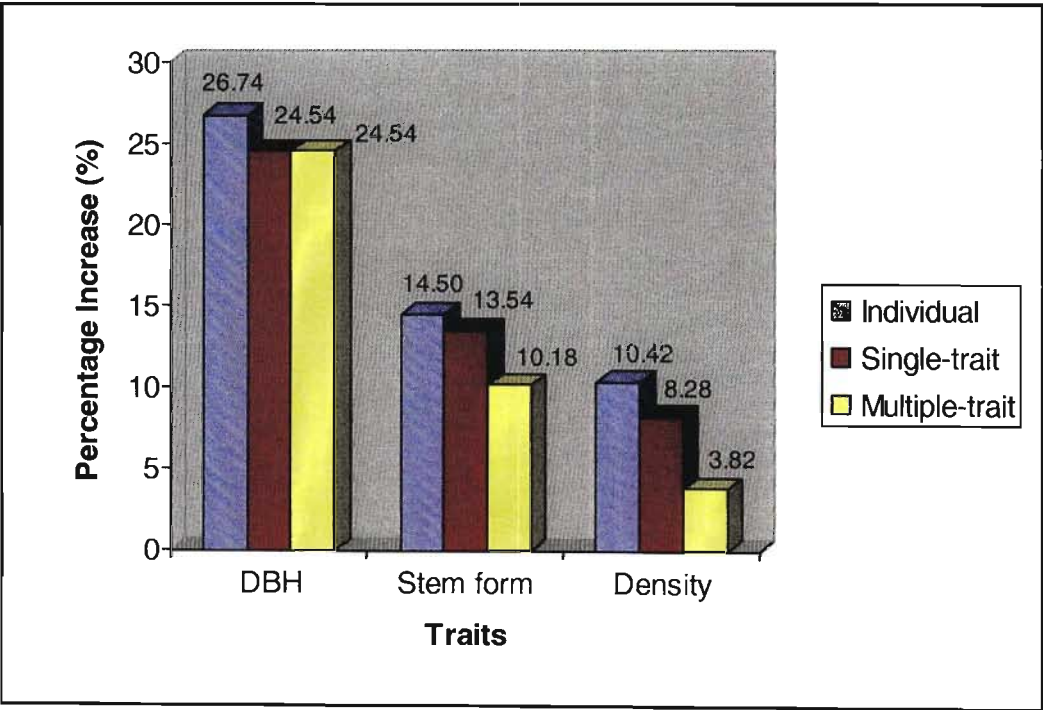
Figure 3.4 Percentage increases for three of the traits in *Dataset* for the three selection strategies, namely, individual selection, single-trait index selection and multiple-trait selection.

As illustrated in Figure 3.4, in this investigation DBH was found to have the greatest response to selection, across all three selection strategies, followed by stem form and then density.

In conclusion, the response to selection for this population was estimated and compared across three selection strategies for the four traits in *Dataset*. The expected response to selection for each trait was greatest when using the individual selection strategy. The trait DBH was noted to have the greatest response to selection.

3.6 COMPARISON BETWEEN INDICES FOR COMMONALITY AMONG THE TOP 30 RANKED INDIVIDUALS

The top 30 ranked individuals for each of the fifteen MATGEN® (2003) indices were compared. On assessment, it was observed that some individuals were the highest performers across all the indices.

An indication of commonality between the indices was anticipated to show the researcher that, in a tree improvement programme, a breeder could use a simple index, selecting for few traits, to be equally effective in identifying genetically superior individuals as a complex index selecting for more traits. This is because the researcher's goal is to improve all four traits in this population with the least effort and expense.

A representation of the top 30 ranked individuals for each index is shown in Figure 3.5. As an example, six individuals were highlighted in various colours to show the commonality that was observed. It was thus assumed by the researcher that a strong positive correlation between the indices would exist. It was however noted that, although these individuals were all in the top 30, their ranking was not the same in each index.

Rank	M1	M2	M3	M4	M12	M13	M14	M23	M24	M34	M123	M124	M134	M234	M1234
1	270	145	685	184	270	270	575	49	656	184	270	575	427	427	773
2	145	300	17	344	143	143	270	17	575	427	143	425	773	183	575
3	143	298	123	427	145	172	425	685	199	183	172	773	270	426	427
4	172	653	256	183	57	773	773	270	86	426	773	270	425	184	425
5	443	504	579	426	172	145	427	48	137	773	145	173	355	773	270
6	575	49	411	355	748	627	355	771	34	19	627	355	575	19	355
7	173	86	409	468	773	21	173	3	427	355	21	137	19	355	653
8	57	443	410	757	173	575	137	190	298	190	57	427	21	142	173
9	748	575	122	19	21	443	573	409	653	684	748	653	627	575	19
10	626	371	771	727	443	748	653	773	142	616	443	443	34	34	34
11	773	34	190	372	573	57	143	616	371	142	575	34	653	190	573
12	653	199	121	321	626	653	19	172	355	685	653	172	143	616	443
13	21	198	254	469	627	212	443	287	425	632	212	573	573	684	137
14	764	656	440	588	425	214	172	241	291	410	214	143	172	425	143
15	425	57	270	656	764	701	21	259	426	258	425	145	443	291	21
16	627	55	580	142	575	425	34	34	19	191	701	19	426	191	172
17	573	259	633	425	653	48	627	587	231	771	747	57	173	434	627
18	212	291	569	258	212	685	145	653	727	757	48	748	701	48	57
19	263	71	773	773	263	427	57	684	588	772	573	21	137	653	145
20	214	241	581	231	116	747	748	145	705	123	427	171	171	258	701
21	701	3	684	705	214	34	171	227	300	285	263	627	48	771	426
22	700	173	616	99	701	269	764	198	773	468	685	764	57	574	171
23	34	172	255	199	338	263	701	575	372	48	700	212	145	685	748
24	116	171	585	137	531	573	212	191	183	434	355	701	212	588	212
25	171	502	18	575	700	700	214	436	344	344	34	231	214	344	231
26	502	503	427	43	747	3	231	617	341	532	531	71	748	199	214
27	504	108	49	87	355	355	426	585	173	627	269	214	2	656	48
28	268	48	48	285	2	171	71	443	38	252	3	298	574	285	574
29	71	764	570	589	502	502	2	291	443	346	502	574	231	582	2
30	269	38	16	177	762	17	574	411	109	425	683	341	700	772	700

Figure 3.5 Illustration of the top 30 individuals for each index, highlighting six individuals in various colours to show commonality between the indices.

3.6.1   Rank-correlation matrix

A rank-correlation matrix was generated to show the correlations between the top 30 ranked individuals of the fifteen indices (Figure 3.6). However, the rank-correlations calculated between the fifteen indices indicated very weak positive and negative correlations. This was an unexpected result. The highest positive correlation was 0.392 and the highest negative correlation was -0.413, as shown in Figure 3.6 (highlighted in red). This was concerning for the researcher for the reason that when two indices were manually assessed it was found that many individuals were common to both in the top 30 ranking (Figure 3.5).

	M1	M2	M3	M4	M12	M13	M14	M23	M24	M34	M123	M124	M134	M234	M1234
M1	1.000	-0.106	0.151	-0.092	0.194	-0.141	-0.016	0.130	-0.085	0.172	0.392	-0.211	-0.023	-0.027	-0.190
M2		1.000	0.007	0.072	-0.248	0.237	0.107	-0.062	0.063	0.250	-0.034	0.112	-0.009	-0.091	-0.376
M3			1.000	0.268	0.016	0.110	0.054	-0.059	0.193	0.138	0.110	0.074	-0.371	-0.045	-0.171
M4				1.000	0.027	0.119	-0.399	0.068	-0.059	0.000	-0.357	-0.102	-0.398	-0.052	-0.221
M12					1.000	-0.110	-0.167	0.068	-0.118	-0.021	0.251	-0.266	-0.019	0.015	-0.071
M13						1.000	0.001	0.067	0.072	-0.024	0.095	0.100	-0.177	-0.383	-0.411
M14							1.000	0.104	0.104	0.032	0.362	0.200	-0.079	0.107	0.313
M23								1.000	-0.179	0.109	0.208	-0.018	0.109	-0.201	-0.173
M24									1.000	0.153	-0.050	0.292	0.022	0.062	-0.032
M34										1.000	0.039	0.014	-0.006	0.299	-0.412
M123											1.000	-0.213	0.073	0.120	0.050
M124												1.000	0.039	0.121	-0.209
M134													1.000	-0.280	0.087
M234														1.000	0.050
M1234															1.000

Figure 3.6 Rank-correlation matrix, for 15 indices.

In order to further investigate the reason for the values obtained in the rank-correlation matrix, two indices, M123 and M124, were compared. Table 3.5 shows the two indices, M123 and M124, as ranked according to their top 30 individuals (extracted from Figure 3.5). The columns of M123 and M124 represent the unique individual tree number for each of the 30 individuals. When the researcher counted the number of trees common to both, the answer obtained was twenty (20) individuals. This implies that if the researcher chose to select individuals by either using index M123 *(DBH, stem, split)* or index M124 *(DBH, stem, den)*, there would be twenty individuals common in the top 30 ranking. It was therefore implicit that a positive correlation should exist between these two indices. However, the rank-correlation was calculated as -0.213 (highlighted in purple, Figure 3.6), thus indicating a weak negative correlation. This same problem was acknowledged by the researcher throughout the rank-correlation matrix. Hence, the calculated rank-correlations did not portray the results that were expected.

This was explained once the nature of a *correlation* was reviewed. A correlation defines how two variables vary with respect to one another. Hence, even though the two indices in the example discussed previously had 20 individuals that were in common, the same individual tree in each index was not necessarily ranked in the same position. Therefore, the two indices did not vary in the same way thus, resulting in a weak negative correlation. Consequently, it was considered more appropriate to manually count the number of individuals common between the fifteen indices, rather than using rank-correlations, in order to approximate the number of individuals similar in the top 30 ranking across the fifteen indices.

Table 3.5   **An example, using M123 and M124, of two MATGEN® (2003) indices ranked according to the top 30 individuals.**

Rank	M123	M124
1	270	575
2	143	425
3	172	773
4	773	270
5	145	173
6	627	355
7	21	137
8	57	427
9	748	653
10	443	443
11	575	34
12	653	172
13	212	573
14	214	143
15	425	145
16	701	19
17	747	57
18	48	748
19	573	21
20	427	171
21	263	627
22	685	764
23	700	212
24	355	701
25	34	231
26	531	71
27	269	214
28	3	298
29	502	574
30	683	341
Correlation	**-0.213**	

3.6.2   Manual assessment for commonality

The number of individuals common, in the top 30 ranking, between the fifteen indices was manually counted. As shown in Table 3.6, the number of individuals common between two indices was represented as a number out of 30 (above the diagonal) and as a percentage (below the diagonal). Different colours were used to highlight various observations:

- One or two individuals common between indices were highlighted in purple and blue respectively.
- Pink was used to highlight the commonality between M1 and M12.
- Green highlighted indices that had 93% commonality.
- Red was used to show the 100% commonality between index M134 and M1234.

Table 3.6 The number of individuals common between the 15 MATGEN® (2003) indices.

	M1	M2	M3	M4	M12	M13	M14	M23	M24	M34	M123	M124	M134	M234	M1234
M1		13	2	3	24	23	22	8	7	3	22	22	21	5	21
M2	43%		2	3	9	11	11	13	13	1	10	12	10	7	10
M3	7%	7%		2	3	6	3	13	2	10	5	3	4	8	4
M4	10%	10%	7%		4	5	9	2	18	14	5	8	9	16	9
M12	80%	30%	10%	13%		22	21	7	7	4	23	20	21	5	21
M13	77%	37%	20%	17%	73%		21	12	8	7	28	21	23	9	23
M14	73%	37%	10%	30%	70%	70%		8	13	7	20	28	28	10	28
M23	27%	43%	43%	7%	23%	40%	27%		6	8	11	8	9	12	9
M24	23%	43%	7%	60%	23%	27%	43%	20%		9	8	14	13	16	13
M34	10%	3%	33%	47%	13%	23%	23%	27%	30%		7	6	8	21	8
M123	73%	33%	17%	17%	77%	93%	67%	37%	27%	23%		20	22	9	22
M124	73%	40%	10%	27%	67%	70%	93%	27%	47%	20%	67%		26	9	26
M134	70%	33%	13%	30%	70%	77%	93%	30%	43%	27%	73%	87%		11	30
M234	17%	23%	27%	53%	17%	30%	33%	40%	53%	70%	30%	30%	37%		11
M1234	70%	33%	13%	30%	70%	77%	93%	30%	43%	27%	73%	87%	100%	37%	

Above the diagonal Represented as a number out of 30 individuals
Below the diagonal Represented as a percentage

An overview of the results observed from Table 3.6 yielded the explanations given in Table 3.7.

Table 3.7 Analysis of the results obtained from Table 3.6.

Number of individuals common	MATGEN® (2003) Indices	Comparison of the number of individuals common to the top 30 ranking for various indices
1 (3%)	M2 (*stem*) – M34 (*split, density*)	Not many individuals were found to be common between these indices. The reason for this was due to the single-trait selection indices selecting for one trait only and being compared to another index selecting for a different trait(s).
2 (7%)	M1 (*DBH*) – M3 (*split*) M2 (*stem*) – M3 (*split*) M3 (*split*) – M4 (*density*) M4 (*density*) – M23 (*stem, split*) M3 (*split*) – M24 (*stem, density*)	
24 (80%)	M1 (*DBH*) – M12 (*DBH, stem*)	80% of the trees selected for a good DBH were the same trees found when selecting for DBH and stem form.
28 (93%)	M13 (*DBH, split*) – M123 (*DBH, stem, split*) M14 (*DBH, density*) – M124 (*DBH, stem, density*) M14 (*DBH, density*) – M134 (*DBH, split, density*)	*Stem form (2)* was added to the two-trait indices to select for three traits, but the trait did not seem to make a difference to the three-trait selection for the top 30 ranking as compared to selecting for two traits alone. Similarly, *splitting (3)* did not seem to make a difference when added to M14.
	M14 (*DBH, density*) – M1234 (*DBH, stem, split, density*)	This result implied that the researcher could have selected for only two traits using index M14 (*DBH and density*) and it would have resulted in 93% of the same individuals as when selecting for four traits (*M1234*).
30 (100%)	M134 (*DBH, split, density*) – M1234 (*DBH, stem, split, density*)	Exactly the same individuals were selected when using index M134 (*DBH, split, density*) as compared to selecting for four traits (M1234 – which includes *stem form*).

The result of 80% commonality based on expected rankings, for the selection indices M1 and M12 (Table 3.6 and Table 3.7), reinforced the earlier estimate of the phenotypic correlation (0.834) between DBH and stem form (Table 3.3). Furthermore, the results indicated, that it was possible to select for only two traits (M14 - *DBH and density*) in order to achieve 93% of the same individuals as when selecting for four traits (*DBH, stem, split, density*). Additionally, it was also possible to select for three traits (M134 - *DBH, split, density*) to obtain exactly the same individuals (100%) as those individuals selected when using the index for four traits (*M1234*).

In conclusion, the top 30 ranked individuals for each of the fifteen indices in this investigation were compared using two different methods. The manual assessment for commonality allowed the researcher to observe that a simple index, selecting for few traits (in this study, selection for DBH and density), resulted in a similar response (93%) to selection as that obtained from an index selecting for four traits. In a tree improvement programme this knowledge would benefit a breeder trying to improve four traits in the population with the least effort and expense.

3.7 SELECTION OF INDIVIDUALS FOR PRODUCTION PURPOSES

Selection for production purposes has different requirements to selection for breeding. When selecting for clonal production purposes the entire genotype of the individual is replicated thus, the non-additive genetic variance will also play a role together with the additive genetic variation. Consequently, the selection index for production purposes will place more emphasis on the phenotype of the individual tree (Hettasch *et al.*, 2006).

In this investigation, the researcher was required to look for a particular tree or a small group of trees that could be used for clonal production. Three selection options, according to specific trait requirements, were implemented on the indices that selected for three or four traits. These selection options were compared according to the number of individuals that would potentially meet the trait requirements for production purposes, such as clonal forestry.

3.7.1   Consideration of four traits in the population

The index *M1234* for the four traits analyzed in this investigation, namely, DBH, stem form, splitting and density was filtered in Excel® (refer to *page 43* for an explanation on filtering), according to the specific trait requirements specified by one of the three selection options, as follows:

   a.   Selection of individuals with phenotypic values in the top 10% for all traits

No individuals were identified as having phenotypic values that fitted into the top 10% for all four traits in the population of 475 individuals. However, tree 773 (ranked 1st in the index) did meet the requirements for three of the four traits. Its stem form value did not make the top 10%.

Nonetheless, its stem score was adequately high at 7, and this tree was consequently considered by the researcher as a potential candidate for use in mass propagation.

These results indicate that there is a very low probability of obtaining trees, which could be vegetatively propagated for production purposes that would meet the phenotypic values in the top 10% for all four traits under investigation.

b.   Selection of individuals with phenotypic values in the top 20% for all traits

As observed in the selection of individuals with phenotypic values in the top 10% for all traits, similarly, no individuals were identified as having phenotypic values in the top 20% for all four traits in this study. However, in this instance, six individuals were considered acceptable for selection for all four traits (Table 3.8), as the phenotypic values for three of the traits fell into the top 20% and they all had a stem score of above 6, which was considered acceptable for selection purposes and subsequent production.

Table 3.8   Six trees were identified as possibly having phenotypic values in the top 20% for all four traits.

Tree	Rank in M1234	DBH (mm)	Stem Form (1-8 score)	Splitting (score)	Density (kg.m^{-3})
773	1	520	7	32.04	678
427	3	451	6	30.46	798
21	15	490	6	21.50	648
627	17	474	6	28.97	651
532	34	438	6	16.03	706
616	74	440	7	32.21	685

Selection of individuals with phenotypic values in the top 20% was a more relaxed approach than selection for the top 10%. However, when breeders start phenotypically selecting more than the top 20% for a particular trait, the selection intensity (i) and thus, the response to selection (R) is expected to decline.

c.   Selection of individuals suitable for use in a commercial situation

In the population of 475 individuals, tree 48 (ranked 27th in the index) was the only tree acknowledged as satisfying the 'commercial requirements' as laid out by the researcher (See Chapter 2, Section 2.3.6, p43). Additionally, as shown in Table 3.9, a further four individuals were recognized as possibly satisfying the requirements for all four traits, since three of the traits met the trait criteria and they had a stem score of above 6. Fewer individuals satisfied the criteria for this

selection option, as compared to the selection of individuals having phenotypic values in the top 20%, as in some traits the selection intensity was less than 20%.

Table 3.9   **Four trees were identified as possibly having phenotypic values that met the trait criteria in the commercial situation for all four traits.**

Tree	Rank in M1234	DBH (mm)	Stem Form (1-8 score)	Splitting (score)	Density (kg.m^{-3})
773	1	520	7	32.04	678
427	3	451	6	30.46	798
21	15	490	6	21.50	648
627	17	474	6	28.97	651

Table 3.10 summarizes the selection for all four traits in the population, using the index M1234, and shows a comparison of the three selection options implemented in this investigation. Only one individual in this study (n=475) was found to entirely satisfy the trait requirements that were specified by the researcher, using the commercial situation selection option.

From the results in Table 3.10, the following was noted:

1. The commercial situation was the only selection option successful in obtaining at least one individual (tree 48) which met the requirements for all four traits under investigation when a population of 475 individuals was considered.
2. As the selection criteria for the traits were increased from 10%, it became noticeable that more individuals were available for selection as potential candidates for vegetative propagation.
3. The number of individuals that met the phenotypic trait requirements for production purposes changed between the 20% and commercial selection options due to changes in the trait criteria. For this reason, particular trees were either lost or gained from the list of prospective candidates.
4. The selection option providing the most individuals possibly available for production purposes was found using the 20% selection option.
5. When considering the commercial selection option in this study, the researcher could assume that, in total, five individuals (trees 48, 773, 427, 21 and 627) could have been suitable for further production purposes and thus be included in a clonal selection trial.

Table 3.10 **Selection for all four traits in the population, for clonal production purposes, using the three selection options.**

Selection option	Ranking	Tree No.	Explanation	Number of trees possibly available for production purposes
Select individuals in the top 10% for each trait	-	-	No trees met the requirements for all four traits in the top 10% of the population.	0
	1	773	Tree 773 met the requirements for three of the four traits in this selection option. Its stem form value did not fall into the top 10% requirements. Nonetheless, its stem score was adequately high at 7, and was considered as a suitable individual for production purposes.	1
Select individuals in the top 20% for each trait	-	-	No trees met the requirements for all four traits in the top 20% of the population	0
	1	773	Six individuals were considered as satisfying the requirements for all four traits using this selection option as they had a stem score of 6 and above, which was considered acceptable by the researcher for vegetative propagation.	
	3	427		
	15	21		
	17	627		6
	34	532		
	74	616		
Select for a commercial situation	27	48	Tree 48 was the only individual in this population acknowledged as satisfying the commercial trait requirements considered necessary by the researcher.	1
DBH – top 15% Stem – top 20% Split – top 15% Density– top 30%	1	773	Four individuals were considered as satisfying the requirements for all four traits using this selection option as they met the criteria for 3 of 4 traits and had a stem score of 6 and above, which was considered acceptable for clonal production.	
	3	427		
	15	21		4
	17	627		

3.7.2 Consideration of three traits in the population

There were four indices, namely, M123, M124, M134 and M234, that selected for various combinations of three traits. When selecting for only three traits, as compared to four traits, it was anticipated by the researcher that more individuals would be found to meet the phenotypic requirements for clonal production purposes. Only one selection option was investigated, under selection for three traits, namely, selection of individuals with phenotypic values in the top 10%.

A comparison between the four indices, when selecting for individuals with phenotypic values in the top 10%, is shown in Table 3.11. From the results, the following was noted:

1. The index that resulted in the highest number of individuals satisfying the three trait criteria was M123, showing thirteen individuals. This index was however applied on a larger population size (n=748) as compared to the other indices.
2. Tree 773 was selected in all the indices. However, its stem form did not fit the phenotypic value requirement for a top 10% in this population. Nevertheless, its stem form value was adequate at 7 and was thus considered satisfactory.
3. The same trees were not selected in the different indices, for example, trees exclusive to M124 included trees 137, 355, 425 and 575.
4. Tree 427 was common to three of the four indices.
5. The individuals selected had varied rankings between the indices, for example, tree 773 was ranked 4th, 3rd, 2nd, and 5th. However, tree 427 maintained its ranking for two indices, namely M134 and M234.

Table 3.11   **Comparison between the four indices which included three traits in the population.**

Index	Ranking	Tree number	Consideration of three traits	Number of trees satisfying all three traits
M123 DBH, stem, split n = 748	1 3 18 44 61 121	270 172 48 241 481 276	This index did not select for the trait density. Hence, it was found that more individuals satisfied the trait requirements when considering fewer traits in the index [as compared to index M1234 (which includes density)]. The corrected density values for these trees were below average. Consequently, if this index had included *density* these individuals would not have met the trait requirements for phenotypic values in the top 10%.	6
	2 4 6 7 14 20 22	143 773 627 21 747 427 171	These six individuals were considered to satisfy the requirements for all three traits in this index as they have a stem score of 6 and above.	7
M124 DBH, stem, density n = 475	1 7	575 137	Tree 575 had a lower DBH value than tree 773, but had a higher stem form and density value and was therefore ranked 1st in this index. Tree 137 was a bad splitter but had good values for the three traits selected here hence, is ranked 7th in this index and meets the requirements for the top 10% for all three traits.	2
	2 3 6	425 773 355	These three individuals were considered to satisfy the requirements for the three traits in this index as they had a stem score of 7.	3
M134 DBH, split, density n = 475	2	773	DBH was included in this index and tree 773 had an above average DBH value which fell into the top 10% of the population.	1
	1	427	Tree 427 was ranked 1st in this index however its corrected DBH value (1.316) did not fall into the phenotypic values for the top 10% of the population. However, due to the corrected DBH value being above the average it was considered as a suitable candidate.	1
M234 stem, split, density n = 475	-	-	No trees met the requirements for all three traits in the top 10% of the population.	0
	1 2 5	427 183 773	These three individuals were considered to satisfy the requirements for the three traits in this index as they had a stem score of 6 and above. Tree 183 does not have a good DBH value (-0.137). However, in this index the trait DBH was not considered. Consequently, in this index, a high ranking resulted for this individual.	3

3.7.3 Influence of the number of traits included in an index and population size on the number of individuals suitable for production purposes

The results of this investigation seemed to indicate that the number of individuals identified as suitable candidates for production purposes was influenced by:

a)   The number of traits included in an index, and

b)   Population size.

Thus, specific comparisons between indices constructed in this study were made.

a.   The number of traits included in an index

A comparison between four indices, M1234 (inclusion of four traits) and other indices, namely, M124, M134 and M234 (inclusion of three traits) was made. The results are shown in Table 3.12. It should be noted that these indices all had the same population size (n=475) and individuals with phenotypic values in the top 10% for the traits under investigation were identified.

Table 3.12   Comparison between four indices to establish the influence of the number of traits included in an index on the number of individuals suitable for production purposes.

	Index			
	M1234	M124	M134	M234
Number of traits included in the index	4	3	3	3
Number of individuals satisfying requirements for top 10%	0	5	2	3

The results show that the indices, namely, M124, M134 and M234 that considered three traits resulted in some individuals that met the requirements of phenotypic values in the top 10% of the population. However, index M1234 (which considered all four traits) was unsuccessful in identifying any individuals that had values in the top 10%. The results imply that as the number of traits included in an index increases, so the number of individuals satisfying the trait requirements decreases.

b.   Population size

A comparison between four indices based on different population sizes was made. Table 3.13 shows a comparison between the indices considering three traits in this investigation, namely, M123, M124, M134 and M234 and provides the number of individuals with phenotypic values in the top 10% for all traits.

Table 3.13   Comparison between four indices to establish the influence of population size on the number of individuals suitable for production purposes.

	Index			
	M123	*M124*	*M134*	*M234*
Population size	748	475	475	475
Number of individuals satisfying requirements for top 10%	13	5	2	3

As shown in Table 3.13, the difference between M123 and the other indices was population size. M123 *(DBH, stem form, splitting)* was analyzed on a larger population of 748 individuals, as compared to the other indices that considered only 475 individuals due to the inclusion of the trait *density* (abbreviated by the *number 4*), which only had measurements for 475 individuals in *Dataset*. As shown by the results, in the case of the index that considered a larger population (M123) thirteen individuals were identified as having phenotypic values that met the requirements for the top 10% for the three traits. In contrast, when the population size was only 475 between two and five individuals were identified as being suitable for production purposes. The results thus suggest that one is more likely to find outstanding individuals that have values in the top 10% for all traits under consideration when measurements are taken on a large number of trees.

In conclusion the results imply that when fewer traits were considered in the index, there was more chance of finding a greater number of individuals appropriate for clonal forestry. Similarly, when a larger population was considered, there was evidently more chance of obtaining a greater number of trees as suitable candidates for production purposes.

3.8   DETERMINATION OF THE EFFECT OF POPULATION SIZE ON THE NUMBER OF INDIVIDUALS WHO WOULD BE SELECTED FOR PRODUCTION PURPOSES

From previous results obtained in this investigation, population size was found to influence the number of individuals suitable for production purposes. In order to observe this effect further, the researcher generated a larger hypothetical population in order to evaluate whether a larger population would result in more individuals satisfying the trait requirements for a 'commercial situation' selection option.

3.8.1   Formation of the hypothetical population

A hypothetical population of 1000 individuals was generated using the adjusted means (for the effect of replication) of the four traits in the population, together with the estimated between (σ_b) and within (σ_w) standard deviation values. Table 3.14 shows the values for the adjusted means of the traits, calculated using SAS® procedures, and the family variance components, and subsequent standard deviation components, for each trait.

Table 3.14   **Estimated family variance (σ^2) and standard deviation (σ) values, based on the adjusted means for each trait in the population.**

Trait	Adjusted mean	Family Variance		Standard Deviation	
		Between (σ^2_b)	Within (σ^2_w)	Between (σ_b)	Within (σ_w)
DBH	4.040	0.1836	0.8234	0.4285	0.9074
Stem Form	4.422	0.1080	1.0946	0.3286	1.0462
Splitting	-0.099	0.0441	0.9545	0.2100	0.9770
Density	9.700	0.1319	1.0516	0.3632	1.0255

3.8.2   Estimation of the population parameters

The family and error variance components for each of the traits in the hypothetical dataset were estimated and, subsequently, the heritability for each trait was calculated, as shown in Table 3.15.

Table 3.15   **Estimated heritabilities for the four traits in the hypothetical population.**

Trait	Estimated heritability (h^2)
DBH	0.47
Stem Form	0.31
Splitting	0.18
Density	0.41

All the heritability estimates in the hypothetical population were slightly lower than those calculated in *Dataset* (Table 3.2). The heritabilities between the two populations were expected to be very similar due to the fact that the hypothetical population was constructed using the population parameters of the *Dataset*. An index for the four traits in the hypothetical population was run using MATGEN® (2003).

3.8.3   Commercial selection of trees for production purposes

A selection index for the four traits in the hypothetical population (M1234HP) was filtered in Excel®
according to the trait percentages specified by the commercial situation selection option, namely:

DBH	-	the top 15% of individuals in the population
Stem form	-	the top 20% of individuals in the population
Splitting	-	the top 15% of individuals in the population
Density	-	the top 30% of individuals in the population

The number of trees expected to meet the commercial requirements for all four traits in the
hypothetical population was predicted to be much higher than the single individual obtained in
Dataset (Table 3.10, p65). This was due to the hypothetical population being of a much larger size
(n=1000 individuals) as compared to *Dataset* (n=475). Thus, as expected, based on the index
ranking, 12 trees in the hypothetical population of 1000 individuals were found to have phenotypic
values that met the trait requirements for all four traits in the commercial selection option, further
indicating the effect of population size on the number of individuals suitable for selection purposes.
These 12 individuals were noted to have good corrected trait values for all four traits.

Noting that the hypothetical population was approximately double in size as compared to *Dataset*,
the hypothetical population was shown to have greatly increased the number of individuals found
to meet the percentage requirements for the four traits in the commercial selection option, as is
illustrated in Figure 3.7. The graph (Figure 3.7) shows that as the population size increased the
number of individuals that the researcher found acceptable, according to the trait criteria specified
by the commercial selection option, also increased. This seemed logical, as a larger population
offers a greater choice of individuals from which the researcher can choose for production
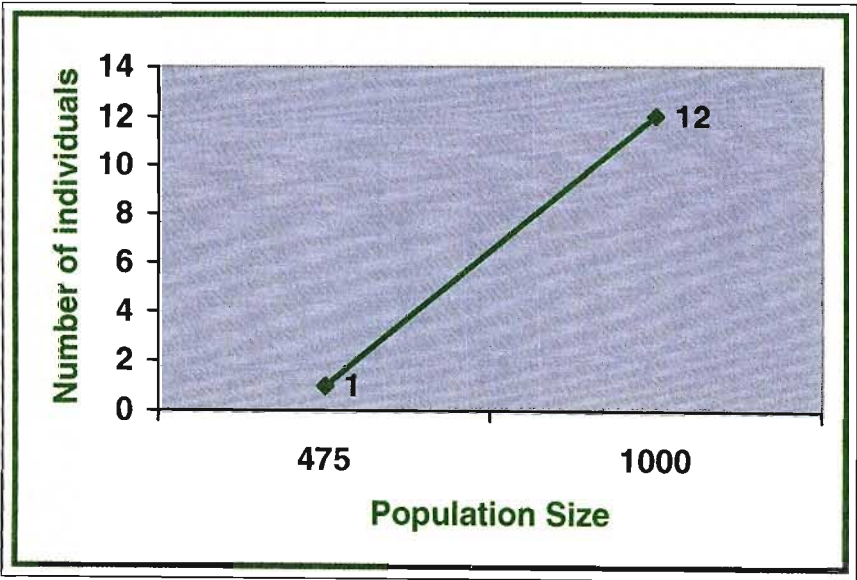purposes.

**Figure 3.7   Illustration between population size and the number of individuals that would
meet the trait criteria for the commercial selection option for all four traits.**

The phenotypic correlations estimated for *Dataset* and hypothetical population were compared and the following was noted:
- The phenotypic correlation between DBH and stem form, in *Dataset* was particularly high at 0.83 (Table 3.3, p55), in contrast to 0.66 in the hypothetical population.
- The phenotypic correlations between other traits, in both of the populations, were similar. In *Dataset* the correlations between other traits were of no particular magnitude and a similar result was observed in the hypothetical population as values were close to zero.

It was predicted that if the phenotypic correlation between DBH and stem form in the hypothetical population was as high as that in the *Dataset* (0.83), there would have been even more individuals in the hypothetical population that had phenotypic values that met the trait requirements specified by the commercial selection option.

It may be concluded that the larger the population size, the more individuals available to be identified as candidates for use in commercial forestry production further indicating the effect of population size on the number of individuals suitable for production purposes.

3.9   CONCLUDING REMARKS

In summary, the following results were obtained in this investigation:

➤ The *Dataset* was edited, standardized and corrected for the fixed effect of replication.

➤ The population parameters of the *Dataset*, namely, the heritabilities of the traits and the phenotypic correlations, were estimated for use in index calculations.

➤ An index value for each tree (BLP) for various selection indices was obtained.

➤ The response to selection for each trait was estimated and compared across three selection strategies, namely:
- Individual selection
- Single-trait index selection
- Multiple-trait index selection.

➤ The commonality of individuals in the top 30 ranking between the various selection indices was determined and showed that a simple index, considering two traits was equally effective in identifying genetically superior individuals as a complex index, considering four traits.

➤ The number of individuals meeting the trait requirements, stipulated by various selection options, namely, (1) selection of individuals with phenotypic values in the top 10% for all traits in the population, (2) selection of individuals with phenotypic values in the top 20%

and, (3) selection for a 'commercial situation' for the traits under investigation, were used to select superior individuals for use in production and deployment as potential clones for clonal forestry. The results suggested that breeders should consider large populations and only a few important traits in order to obtain a greater number of individuals suitable for mass propagation in clonal forestry.

➢ A hypothetical population was generated. The size of a population was shown to have an effect on the number of individuals who were selected for production and deployment options. The larger the population size, the more individuals available to be identified as candidates for use in commercial forestry production.

4 CHAPTER FOUR: GENERAL DISCUSSION AND SUGGESTIONS FOR FUTURE RESEARCH

4.1 INTRODUCTION

Forestry companies need to focus on global trends in tree improvement programmes in order to be successful in meeting the demand for forest products. Companies are expected to grow superior trees faster. Hence, tree breeders are needed: (1) to scientifically stock plantations with genetically superior trees and, (2) to anticipate market changes.

This investigation successfully addressed the aims by firstly calculating individual breeding values (BLP) and ranking individuals accordingly. Thereafter, this study investigated the use of BLP for (1) predicting the breeding population's response to selection for the traits under investigation and, (2) identifying genetically superior individuals for use in commercial clonal forestry production.

4.2 SUMMARY AND DISCUSSION OF RESEARCH FINDINGS

4.2.1 Research Aim 1

To edit a dataset so that a detailed study of the data could be made and for the creation of a standardized dataset, corrected for fixed effects, which could consequently be analyzed.

As stated by Hettasch *et al.* (2006), the quality of any analytical procedure will depend on the quality of the data. Consequently, an initial analysis of *Dataset* was performed so that accurate information about the trial could be obtained. The analysis included the determination of 'tests for normality', missing information, identification of outliers, ANOVA's, regressions, means and standard deviations. During the analysis of the data, accuracy and reliability of data were seen as extremely important. In this investigation, the data were edited, standardized and corrected for the fixed effect of replication using SAS® procedures.

4.2.2   Research Aim 2

To estimate the population parameters, namely, heritabilities and phenotypic correlations, for use in index calculations.

Quantification of the genetic control of a particular trait and an understanding of the relationships among relevant characteristics are important for developing breeding and production strategies to improve the quality of wood and its usefulness. Furthermore, knowledge of population parameters assist the researcher in breeding and production decisions and enable the prediction of genetic gain in various traits. Consequently, population parameters, namely, the heritabilities of each trait and the phenotypic correlations between traits in *Dataset* were estimated using SAS® procedures.

In this study, DBH was found to have the highest heritability (0.600), followed by density (0.492). The estimated heritability for stem form was 0.401 and splitting had the lowest heritability at 0.214. The narrow-sense heritability of numerous traits in many *Eucalyptus* populations has been estimated. For example, in a study conducted in Brazil by Santos *et al.* (2004), the heritability estimates of wood traits in open-pollinated *E. grandis* progenies varied from 0.34 to 0.61. In South Africa, the heritability for splitting in *E. grandis* is commonly recorded around 0.4, for stem straightness 0.2, and for density 0.3 (Hettasch *et al.*, 2006). However, the heritability estimates obtained in this investigation were expected to differ from such published estimates, as each heritability estimate is specific to the population, the trait and the environment on which the estimate is based (Fins *et al.*, 1992; Hettasch *et al.*, 2006). This was illustrated in a study conducted by Shelbourne and Low (1980) that assessed the heritability of DBH in *Pinus radiata* progenies at age seven years, at five sites across New Zealand. The heritability estimates for the trait DBH showed a wide range in values across the various regions. This was concluded to be due to extreme site variability at particular sites emphasizing the importance of estimating heritabilities for a particular population for the environment in which selection will be undertaken. As explained by Jacquard (1983), the heritability of the trait does therefore not characterize in absolute terms the trait itself, but rather the structure of the population in which it was studied. Consequently, heritability values are expected to vary from trial to trial.

The phenotypic correlations between traits obtained in this investigation were close to zero. Only the correlation between DBH and stem form had distinct magnitude (0.83). This was a much larger value than that noted by Dean *et al.* (1983) who obtained a phenotypic correlation of 0.01 for DBH and stem form. However, this study was conducted on radiata pine of age four-and-a-half to six years at two sites in eastern Victoria, Australia. The above-mentioned discrepancy in values for population parameters such as correlations further indicates that population parameters vary according to the species, environment, age and genetic-makeup of the trial.

A further difference between correlations obtained in this investigation and published literature was seen when considering the traits of DBH and wood density. In this investigation, the phenotypic correlation between DBH and wood density was noted as slightly positive, yet hardly noteworthy as

the value was nearly zero (0.016). However, in a study of *Picea abies* (Norway Spruce), Costa E Silva *et al.* (2000a) found an adverse relationship between DBH and wood properties from three groups of 15 to 18-year-old progeny tests. This was consequently found to restrict simultaneous genetic gains in growth rate and wood quality. Similarly, Dean *et al.* (1983) revealed that it was not possible to achieve substantial improvement simultaneously in wood density and growth, due to negative genetic correlations in *Pinus radiata*. These results emphasize the necessity to calculate population parameters for specific populations, especially for different species, as the results will influence breeding decisions.

It should be noted that tree selection is usually conducted before rotation age with the purpose of minimizing the generation interval and, consequently increasing genetic gains per unit of time (Osorio, 1999). Half rotation age is a common age for final assessment of families and individuals because half and full rotation data are highly correlated (Zobel and Talbert, 1984) thus allowing for selections to be done at half rotation. In most tree breeding trials the age at which selection traits are assessed is 5 to 8 years. However, this investigation used data obtained from a 20-year-old progeny trial and parameters were calculated using this data. Therefore, the heritabilities and phenotypic correlations calculated in this study may not be the same as estimates published for a younger progeny trial, as changes in correlations between traits at different ages (Lambeth, 1980; Magnussen, 1991) or changes in variance estimates occur over time (Franklin, 1979). Consequently, if this investigation had used data from a younger progeny trial, it was anticipated by the researcher that better comparisons could have been made to other studies. However, it must be pointed out that older trials provide more accurate parameter estimates than data from younger trials and thus the estimates from this investigation may be considered as accurate for the population under consideration.

In conclusion, it was difficult to compare the results in this investigation to other published data since most parameters such as heritabilities and correlations have been published for species such as *Pinus*. Furthermore, the eucalypts data to date refer to the additive genetic variation within provenances (sources) and have been calculated based on open-pollinated progeny trials using data from younger trees (Potts, 2004).

4.2.3   Research Aim 3

To determine a BLP of individual breeding values in order to rank individuals according to an index value.

An index provides a weighted score for each tree thus allowing individual trees or families to be chosen for breeding and production purposes. Index values such as BLP are presumed to have statistical advantages over traditional methods of selection that include increased accuracies of selection (a measure of how close the estimated breeding value is to the true breeding value of a tree), and a better ability to compare trees in different populations, generations or sites (Borralho, 1995). Consequently, in this study, forward selection for a single generation was conducted using

MATGEN® (2003). This resulted in an index value for each tree (BLP) for the various selection indices applied to this population. Individuals with the highest ranking in each index were assumed by the researcher to be the most desirable trees for the traits under consideration in this population, under the assumption that the population parameters used in the construction of the indices in this study were considered accurate and reliable. This is an important consideration as before using any type of index selection, breeders are advised by Cotterill (1985) to check the consequences in terms of genetic gains expected from selection. This is because studies such as those conducted by Hazel and Lush (1943); Elston, (1963); Cotterill, (1985); Dean *et al.*, (1986) and Borralho, (1995) that have analyzed and compared indices in an attempt to demonstrate which are the most efficient, have found that their efficiency is restricted by the reliability of estimates of population parameters.

4.2.4   Research Aim 4

To predict a breeding population's response to selection in order to observe the change in the population mean for a particular trait for different selection strategies.

In order to determine the effect of selection on the change in the population mean of a trait, the breeding population's response to selection was predicted and compared across three selection strategies, namely: (1) individual selection, (2) single-trait index selection, and (3) multiple-trait index selection. The top 8% of individuals in the breeding population were selected for and the genetic gains were predicted. The results illustrated that: (1) in this study, individual selection strategy yielded the best selection response and, (2) DBH showed the greatest response to selection across all three selection strategies.

The response to selection being the greatest for individual selection was an unexpected result. The researcher had anticipated that the single-trait index selection strategy would yield the greatest response to selection. This is because with single-trait index selection, an individual's own phenotypic value is not the only source of information contributing to its breeding value; additional information is provided by the phenotypic values of relatives (Falconer and Mackay, 1996). The most desirable individuals are selected on the basis of information about the individual as well as a weighted index of the between and the within family components for each individual (Hettasch *et al.*, 2006). However, as noted by Falconer and Mackay (1996) and Stonecypher and Arbez (1976) index selection is recognized to be most advantageous for use in selection for traits with low heritability. This is because, for traits with a low heritability, the environmental deviations constitute a large part of the phenotypic variance (Lynch and Walsh, 1998). A low heritability for a trait therefore guides the breeder to place greater emphasis on family means and the use of index selection (Hettasch *et al.*, 2006).

A reason for the individual selection strategy yielding the greatest responses to selection may be explained by the relatively high heritabilities (ranging from 0.21 – 0.60) obtained for the four traits considered in this investigation. As noted by Hettasch *et al.* (2006) individual selection works best for highly heritable traits where the phenotype is a good expression of the genotype, and consequently trees with the best breeding values are selected. For this reason, it was assumed that the results in this investigation were due to the high heritabilities of the traits in *Dataset*. The generally high heritability values provided the conditions whereby individual selection showed the greatest response to selection.

It should however be noted that individual selection does result in similar individuals being selected each generation, meaning that the variation will drop rapidly over the generations until there is little to no variation left in the population to gain any significant response. The implication of this rapid decrease in the standard deviation is that the response to individual selection is going to slow down each generation and eventually level off. Thus, as the standard deviation tends to zero, as highlighted by the equation $R=i.\sigma_p.h^2$, so does the selection response. However, in the case of forestry, it could be argued that as generation intervals are so long, breeders need to select individuals with the best breeding values to obtain maximum genetic gains in the short term.

A comparison between the selection responses for single-trait index selection and multiple-trait index selection for three of the traits in *Dataset* (excluding DBH), showed that single-trait index strategies produced better responses for a single trait than the multiple-trait index selecting for four traits. This was due to the single-trait index focusing on the parameters of one trait only as compared to the multiple-trait index that focused on four traits simultaneously. Hazel and Lush (1943) showed that the more traits that are selected for, the less the response in each trait, as the average improvement in any one trait would only be $\frac{1}{\sqrt{n}}$ times as much if selection were directed for that trait alone. This relationship was successfully shown to be the case in this investigation. Using *density* as an example, the researcher noted that the response to selection for the trait *density* in the multiple-trait index strategy, using four traits (n=4), was only half (0.5) the selection response of that estimated for the *density* single-trait index strategy.

It was noted by the researcher that DBH had the same selection response values for the two index selection strategies. The reason for this may be because the response to selection was being predicted, using the same population, but for two different selection strategies. In the single-trait strategy, DBH was the only trait selected for in the index, and in the multiple-trait strategy DBH was one of four traits that were selected. However, DBH had a high economic weighting (0.50) in the multiple-trait index and it was found that the top 8% of individuals in the breeding population equated to the same individuals in each selection hence, the same response to selection was predicted. As stated by Dean *et al.* (1983), if one or a few traits dominate the index [as measured by the product of the economic weight *(a)* and the heritability for the trait], the other traits will individually be of less importance. In this investigation, this was found to be the case for the trait DBH.

DBH was found to have the greatest response to selection, across all three selection strategies, followed by stem form. Reasons for this result were due to the following:

- DBH had the highest heritability estimate (0.60) in this investigation. The greater the heritability estimate, the larger the proportion of additive genetic variance and consequently, the greater the response to selection. High heritabilities generally indicate successful breeding for improvement in the trait hence, a better response to selection than traits with low heritabilities (Lynch and Walsh, 1998).

- DBH had the largest amount of variation σ_p present as compared to the other three traits under investigation. Traits with higher standard deviations will have a wider variety of individuals to choose from. There is therefore a better chance of achieving a significant response to selection when there is a greater amount of variation available in the population, as it is variation that provides the 'means' for selection to work on.

- DBH had the largest economic weighting in the index (0.50) and this contributed to DBH having the greatest selection response as compared to the other traits in the index. In a study conducted by Costa E Silva *et al.* (2000b) for the prediction of breeding values and expected genetic gains in Norway spruce, diameter growth (DBH) was generally noted as the most weighted trait under multiple-trait selection.

Stem form was observed to have a greater selection response than *density*. This was a little unexpected as *stem form* had a lower estimated heritability (0.40) than that of *density* (0.49). Possible reasons for a greater response to selection for stem form may be attributed to:

- Stem form had a greater amount of variation σ_p relative to that of density.
- Stem form had a higher economic weighting (0.18) in the index as compared to density (0.07).

In may be concluded that the response to selection for each trait was successfully estimated and compared across three selection strategies.

4.2.5 Research Aim 5

To compare selection indices, considering different numbers and choice of traits, for commonality among rankings of the top 30 individuals. This was used to evaluate whether a simple index considering only a few traits could be equally effective in identifying genetically superior individuals as a complex index considering many traits.

Fifteen indices, considering different numbers and choice of traits, were compared for commonality among rankings of the top 30 individuals. Two methods, namely, (1) a rank-correlation matrix and (2) a manual assessment, were used. The commonality between indices showed that a simple index, considering two traits (DBH and density) was equally effective (93%) in identifying genetically superior individuals as the more complex index that considered four traits (DBH, stem form, splitting and density). Furthermore, it was possible to select for three traits (DBH, splitting, density) to identify the same individuals (100%) as those individuals selected using the index that

considered four traits (DBH, stem form, splitting and density). The result implies that it may be possible to exclude *stem form* as a trait considered in a selection index, as the trait did not make an impact with respect to the genetically superior individuals selected in this *E. grandis* population. This was thought to be due to the high phenotypic correlation between DBH and stem form (0.83). It was therefore proposed that after selections have been made, based on index rankings, the trees should then be reviewed to determine if their *stem form* is, in reality, acceptable. These results imply that in some instances the breeder could use a simpler index in order to identify the same genetically superior individuals in a population.

There are many articles on index selection (Hazel, 1943; Arbez *et al.*, 1974; Namkoong, 1976; Lin, 1978; Burdon, 1979; Vangen, 1979; Shelbourne and Low, 1980; Cotterill and Jackson, 1981; Burdon, 1982; Christophe and Birot, 1983; Smith, 1983; Falkenhagen, 1986; Volker *et al.*, 1990) that review the basic theory, look at modifications of the index, discuss limitations and consider the application of the theory. However, the researcher noted the absence of studies considering the evaluation of individuals identified as being genetically superior by different indices, as was the case in this study.

It should be noted that the commercially realistic economic weightings assigned to the traits in this population, and used in the selection indices, would have had a great influence on the results obtained for this study. If different sets of relative economic weights were used, it would have produced very different outcomes of index selection. The results obtained in this study must therefore be viewed in the light of the economic weightings used in the construction of the indices.

4.2.6   Research Aim 6

To select individual trees, using various selection options, for production and deployment purposes.

The researcher's goal was to find the most desirable individuals in the population to be used for production purposes, such as clonal forestry. Therefore, various selection options, namely, (1) selection of individuals with phenotypic values in the top 10% for all traits in the population, (2) selection of individuals with phenotypic values in the top 20% and, (3) selection for a 'commercial situation' for the traits under investigation, were used to select superior individuals for use in production and deployment. Consequently, the number of individuals with phenotypic trait values within a particular percentage was ascertained. The individuals with phenotypic values in the top percentages of the population were recognized as superior individuals and were therefore regarded as potential clones for clonal forestry. The "commercial selection" option was the only option successful in identifying an individual that met the required criteria for the four traits in the population of 475 individuals. Furthermore, the results obtained suggested that a greater number of individuals were suitable for production purposes when: (1) fewer traits were selected for in the index and, (2) a larger population size was considered.

With regard to the number of traits included in an index, Stern (1964) stated that the demand to include as many traits as possible in the index is countermanded by the fact that the selection intensity for other traits is lowered by every new trait considered. If the selection intensity is lowered, this causes a decrease in the expected genetic gains (Falconer and Mackay, 1996). Therefore, the breeder has to restrict the number of traits in the index by choosing only the most important traits. Depending on the genetic correlations between traits, Harwood *et al.* (2005) noted that it is generally not feasible to incorporate more than three to four traits in a tree improvement programme, as including more traits does not facilitate progress towards the breeding objective.

From the results obtained, the researcher noted that the position at which a tree is ranked, on the basis of its index score does not indicate whether the individual may still meet all the selection option's trait criteria. For example, tree 616 was ranked 74th in the index M1234 (selecting for the four traits under investigation) and yet, displayed above average values for the four traits in the population. Under normal circumstances this tree would probably not be selected due to its low index ranking within the population. On review of *Dataset*, the researcher noted that this individual belonged to a family in the population with poor performing relatives. Hence, a reason for this individual's low ranking in the index was assumed to be due to the poor family weighting. Furthermore, in this investigation, when selecting individuals with the particular phenotypic values necessary for production purposes, equal weightings were given for each trait, in this case 25%, as opposed to the economic weightings stipulated in the selection index. Thus, the index may be beneficial if the researcher wanted to improve the breeding population but, as shown by the results, not so reliable when selecting individuals for production and deployment purposes.

It must be noted that an individual's index value (BLP) is determined by individual performance as well as information from various types of relatives for more than one trait. When using an index, the breeder is trying to find the individual with the highest breeding value in order to produce the best progeny. The index value of a tree should therefore represent the best possible guide to the overall breeding value for multiple traits of the tree (Hettasch *et al.*, 2006). However, if an individual with particularly good phenotypic values came from a family of poor performance, this would negatively influence the individual's index ranking. In the case of vegetative propagation however, the breeder only has to consider the phenotypic value of the individual as it is the individual's entire genotype that will be mass produced. Thus index selection or BLP may not be advantageous when trying to identify particular individuals for clonal production.

In this investigation, the researcher aimed to identify a small group of phenotypically superior individuals that could potentially be used for clonal forestry production. Having examined three selection options, it may be concluded that, the results suggested breeders should consider large populations and only a few important traits in order to obtain a greater number of individuals suitable for mass propagation in clonal forestry.

4.2.7   Research Aim 7

To determine the effect of population size on the number of individuals that could be selected for a clonal trial.

In order to determine the effect of population size on the number of individuals suitable for clonal forestry, a large hypothetical population was generated using Excel®. This was accomplished using between family and within family standard deviation values from *Dataset*. In the hypothetical population of 1000 individuals, 12 individuals with phenotypic values in the criteria specified for a 'commercial situation', suitable for production purposes were identified. This result was far higher than for the real population of 475 individuals, further indicating that a larger population will provide a greater number of individuals appropriate for use in production and deployment.

It is interesting to note that from the observed individual rankings (BLP) in the hypothetical population, Tree 31 was noted to be ranked first in the index. However, this tree did not meet the percentage requirements for the commercial selection option since its corrected splitting value was documented as below the average of the population. This was taken as further evidence by the researcher that index selection or BLP may not be the best method of identifying particular individuals for clonal production.

4.3 SUGGESTIONS FOR FUTURE RESEARCH

This investigation successfully addressed the aims by: (1) calculating individual breeding values (BLP) and ranking individuals accordingly, (2) predicting the breeding population's response to selection, according to three strategies, for the four traits under investigation, and (3) selecting superior individuals for use in commercial clonal forestry. However, future research could consider the effect of different sets of economic weightings on index rankings in a population and the influence that population structure has on the optimal genetic gains obtained.

4.3.1   Economic weightings

A major difference between forest tree breeding and breeding for most crops and animals is the much longer time between investment in a breeding programme and returns via an improved harvest (Libby *et al.*, 1969). The economic value of a trait is certainly problematic; it alters with shifting market conditions and technological changes. Hence, it is difficult for tree breeders to predict accurate estimates of economic weightings to several traits combined in an index, especially considering the long generation interval. In some instances, it may be necessary for tree breeders to collaborate with forest economists in order to alleviate the problems associated with 'accurate' economic weightings.

Cotterill and Jackson (1985) outlined three methods for estimating economic weights for use in selection indices. Each method produced different sets of weights and the consequences of these different weights were examined in terms of expected genetic gains and the phenotypic values of individuals retained following index selection. Furthermore, a study conducted by Dean *et al.* (1983) calculated the genetic gains for various traits from mass selection on indices using various combinations of economic weightings. It was concluded by Dean *et al.* (1983) that altering the relative economic weights in the selection index achieved very different combinations of expected gains in the traits.

In an investigation such as this, it would be beneficial for the researcher to change the economic weightings used in the indices in order to observe the effect that various economic weightings would have on the outcome of index rankings and consequent genetic gains.

4.3.2 Population structure

Particular parameters affect genetic gains achieved in breeding and production strategies. Hence, there are various strategy choices and scenarios that need to be considered, such as, the number of families in the population as well as family sizes (the number of individuals per family).
The effect of population structure on expected genetic gains in breeding and production strategies has been investigated by Verryn *et al.*, (2000) and, Verryn and Snedden (2000). Both studies concluded that an optimal population structure will have an impact on predicted genetic gain.

Consequently, future research could study a range of population compositions to give an impression of the optimal population structure required for desirable genetic gains. The hypothetical population in this investigation could be used to generate various populations based on:
- Varying number of trees per family, and/or
- Varying numbers of families in the total population.

4.4 CONCLUDING REMARKS

In planning for the future, tree breeders are building up commercial forests that have desirable genetic qualities. It is critically important in any genetic improvement programme that an appropriate objective is chosen, as this objective drives the direction of genetic change achieved through breeding. Hence, in formulating a breeding strategy, tree breeders are required to have vision with the aim of progressing quickly and cost-effectively towards the objective of developing genetically improved material for production and deployment purposes.

This investigation successfully examined the application of BLP for breeding and clonal production purposes in a *Eucalyptus grandis* population. The investigation explored the use of BLP to: (1) predict the breeding population's expected response to selection, across three selection strategies,

for the four traits under investigation, and (2) select superior individuals for use in commercial clonal forestry production and deployment.

A number of questions that require further investigation were revealed by this research. These include:

- To what extent will the index rankings of individuals differ when different sets of relative economic weightings are used in the index?
- Does population structure influence the expected genetic gains in breeding and production strategies?

REFERENCES

Arbez, M., Baradat, P.H., Mauge, J.P., Millier, C. and Badia, J. 1974. Some problems related to use of selection indices in forest tree breeding. *Proceedings, Joint IUFRO Meeting of working parties in population and ecological genetics, breeding theory and progeny testing.* S.02.04.1-3, Session II. Stockholm, Sweden.

Becker, W.A. 1975. *Manual of Quantitative Genetics.* Washington State University Press. Pullman, WA. 170pp.

Borralho, N.M.G. 1995. The impact of individual tree mixed models (BLUP) in tree breeding strategies. *Proceedings, CRC-IUFRO Conference, Eucalypt Plantations: Improving fibre yield and quality*, 19-24 Feb, Hobart, Tasmania. 141-145.

Burdon, R.D. 1979. Generalisation of multi-trait selection indices using information from several sites. New Zealand Journal of Forestry Science. **9**:145-152.

Burdon, R.D. 1982. Selection indices using information from multiple sources for the single-trait case. Silvae Genetica. **31**: 81-85.

Christophe, C. and Birot, Y. 1983. Genetic structures and expected genetic gains from multitrait selection in wild populations of Douglas fir an Sitka spruce II. Practical application of index selection on several populations. Silvae Genetica. **32**: 173-181.

Costa E Silva, J., Borralho, N.M.G. and Wellendorf, H. 2000a. Genetic parameter estimates for diameter growth, pilodyn penetration and spiral grain in *Picea abies* (L.) KARST. Silvae Genetica. **49**: 29-36.

Costa E Silva, J., Wellendorf, H. and Borralho, N.M.G. 2000b. Prediction of breeding values and expected genetic gains in diameter growth, wood density and spiral grain from parental selection in *Picea abies* (L.) KARST. Silvae Genetica. **49**: 102-109.

Cotterill, P.P. 1985. On index selection II. Simple indices which require no genetic parameters or special expertise to construct. Silvae Genetica. **34**: 64-69.

Cotterill, P.P. and Dean, C.A. 1990. *Successful tree breeding with index selection.* CSIRO, Australia Publications.

Cotterill, P.P. and Jackson, N. 1981. Short Note: Index selection with restrictions in tree breeding. Silvae Genetica. **30**: 106-108.

Cotterill, P.P. and Jackson, N. 1985. On index selection I. Methods of determining economic weight. Silvae Genetica. **34**: 56-63.

Crossa, J. 1990. Statistical analyses of multilocation trials. Advances in Agronomy. **44**: 55-85.

Dean, C.A., Cotterill, P.P. and Cameron, J.N. 1983. Genetic parameters and gains expected from multiple trait selection of Radiata Pine in Eastern Victoria. Australian Forest Research. **13**: 271-278.

Dean, C.A., Cotterill, P.P. and Eisemann, R.L. 1986. Genetic parameters and gains expected from selection in *Pinus caribaea* var. hondurensis in Northern Queensland, Australia. Silvae Genetica. **35**: 229-235.

Denison, N.P and Kietzka, J.E. 1993. The use and importance of hybrid intensive forestry in South Africa. South African Forestry Journal. **165**: 55-60.

Department of Water Affairs and Forestry. 2006. South African Forestry Facts. Republic of South Africa. Available at: http://www.forestry.gov.za.

Department of Water Affairs and Forestry. 2005. Briefing document on forestry in the department. Republic of South Africa. Available at: http://www.dwaf.gov.za.

Doran, J.C., Baker, G.R., Chudleigh, P. and Simpson, S. 2000. Using clones to establish tea tree plantations. RIRDC Research paper series No. 73. RIRDC, Canberra. Available at: http://www.rirdc.gov.au/pub.

Elgin, J.H., Hill, Jr.R.R. and Zeiders, K.E. 1970. Comparison of four methods of multiple trait selection for five traits in *Alfalfa*. Crop Science. **10**: 190-193.

Elston, R.C. 1963. A weight-free index for the purpose of ranking or selection with respect to several traits at a time. Biometrics. **19**: 85-97.

Falconer, D.S. and Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics* (4th ed.). Longman Group Ltd., England.

Falkenhagen, E.R. 1985. Genotype x Environment Interactions in South Africa Pine progeny trials: Implications in tree breeding. South African Forestry Journal. **135**: 53-60.

Falkenhagen, E.R. 1986. Understanding multiple trait index selection and its problems. South African Forestry Journal. **137**: 26-32.

Ferreira, M. and Santos, P.E.T. 1995. Eucalypts wood traits for species/provenances/plus trees and clones planted in Brazil – A review applied to genetic improvement for pulp production. *Proceedings, IUFRO Conference, Eucalypt Plantations: Improving fibre yield and quality,* Theme 4, Hobart, Tasmania. 258-160.

Fins, L., Friedman, S.T. and Brotschol, J.V. 1992. *Handbook of Quantitative Forest Genetics.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Franklin, E.C. 1979. Model relating levels of genetic variance to stand development for four North American Conifers. Silvae Genetica. **28**: 207-212.

Harris, D.L. 1964. Expected and predicted progress from index selection involving estimates of population parameters. Biometrics. **20**: 46-72.

Harvey, W.R. 1990. *User's guide for LSMLMW and MIXMDL PC-2 Version.* Dr. W.R. Harvey, Ohio, U.S.A.

Harvey, W.R. and Townsend, A.M. 1985. Selection on a combination of individual, family, and stand merit in provenance tests. Forest Science. **31**: 813-821.

Harwood, C.E., Bird, R., Butcher, T., Bush, D.J., Jackson, T., Johnson, I., Stackpole, D. and Underdown, M. 2005. *Update of hardwood breeding strategies – ALRTIG.* Rural Industries Research and Development Corporation. RIRDC Publication No 05/023, Canberra.

Hazel, L.N. 1943. The genetic basis for constructing selection indexes. Genetics. **28**: 476-490.

Hazel, L.N. and Lush, J.L. 1943. The efficiency of three methods of selection. The Journal of Heredity. **33**: 393-399.

Heidhues, T. 1961. Relative accuracy of selection indices based on estimated genotypic and phynotypic parameters. Biometrics. **17**: 502-503.

Henderson, C.R. 1963. Selection index and expected genetic advance. In: *Statistical genetics and plant breeding.* (Ed. Hanson, W.D. and Robinson, H.F.). NAS NRC Publication 982. 141-163.

Henderson, C.R. 1984. *Applications of linear models in animal breeding.* University of Guelph, Guelph.

Hettasch, M.H., Snedden, C.L., Lunt, K.A., Pierce, B.T. and Verryn, S.D. 2004. *Practical Data Analysis Tools for Tree Breeders Manual.* Environmentek, CSIR. Pretoria, South Africa.

Hettasch, M.H., Eatwell, K.A., Fossey, A., Pierce, B.T., Snedden, C.L., Steyn, D.J., Venter, H.M. and Verryn, S.D. 2006. *Tree Breeding Course Manual.* Natural Resources and the Environment, CSIR. Pretoria, South Africa.

Hill, J., Becker, H.C. and Tigerstedt, P.M.A. 1998. *Quantitative and ecological aspects of plant breeding.* Chapman & Hall, London.

Huber, D.A., White, T.L. and Hodge, G.R. 1994. Variance component estimation techniques compared for two mating designs with forest genetic architecture through computer simulation. Theoretical and Applied Genetics. **88**: 236-242.

Ikemori, Y.K. 1990. Genetic variation in characteristics of *E. grandis* (HILL) MAIDEN raised from micro-propagation, macro-propagation and seed. Ph.D. Dissertation. Green College, Oxford. 123pp.

Jacquard, A. 1983. Heritability one word three concepts. Biometrics. **39**: 465-477.

Kang, H. and Nienstaedt, H. 1987. Managing long-term tree breeding stock. Silvae Genetica. **36**: 30-39.

Lambeth, C.C. 1980. Juvenile-Mature correlations in *Pinaceae* and implications for early selection. Forest Science. **26**: 571-580.

Lessells, C.M. and Boag, P.T. 1987. Unrepeatable Repeatabilities: A Common Mistake. The Auk. **104**: 116-121.

Libby, W.J., Stettler, R.F. and Seitz, F.W. 1969. Forest genetics and forest-tree breeding. Annual Review of Genetics. **3**: 469-494.

Lin, C.Y. 1978. Index selection for genetic improvement of quantitative characters. Theoretical and Applied Genetics. **52**: 49-56.

Lindgren, D. 2002. Advantages of clonal propagation. In: *Proceedings of workshop on high quality birch – clonal propagation and wood properties.* (Ed. Welander, M. and Zhu, L.H.). August 27-28, 2001.

Lynch, M. and Walsh, B. 1998. *Genetics and analysis of quantitative traits.* Sinauer Associates, Inc., Sunderland, U.S.A.

Magnussen, S. 1990. Selection index: economic weights for maximum simultaneous genetic gain. Theoretical and Applied Genetics. **79**: 289-293.

Magnussen, S. 1991. A probability distribution model for age-age correlations and its application in early selection. Canadian Journal for Forestry Research. **21**: 1550-1558.

Malan, F.S. 1979. The control of end splitting in saw logs: A short literature review. South African Forestry Journal. **109**: 14-18.

Mayo, O. 1980. *The Theory of Plant Breeding.* Oxford University Press, Oxford.

Mayo, O. 1987. *The Theory of Plant Breeding.* (2nd ed.). Oxford University Press, Oxford.

McRae, J.B., Stelzer, H.E., Foster, G.S. and Caldwell, T. 1993. Genetic test results from a tree improvement program to develop clones of loblolly pine for reforestation. Reforestation, Nurseries and Genetic Resources. Available at: http://www.rngr.net/Publications/sftic/.

Molony, J. 1999. An overview of the South African pulp and paper industry. Tappi Journal. **82**: 64-68.

Namkoong, G. 1969. Problems of Multiple Trait Breeding. *FAO Second World Consultation on Forestry Tree Breeding. Breeding for High Yielding characters,* Vol. 1. Washington, August. 775-781.

Namkoong, G. 1976. A Multiple-Index selection strategy. Silvae Genetica. **25**: 199-201.

Namkoong, G. 1979. Introduction to quantitative genetics in forestry. U.S. Department of Agriculture, Technical Bull. No. 1588.

Namkoong, G., Kang, H.C. and Brouard, J.S. 1988. *Tree Breeding: principles and strategies.* Springer-Verlag, New York.

Namkoong, G. and Kang, H.C. 1990. Quantitative genetics of forest trees. Plant Breeding Reviews. **8**: 139-188.

Osorio, L.F. 1999. Estimation of genetic parameters, optimal test designs and prediction of the genetic merit of clonal and seedling material of *Eucalyptus grandis*. Dissertation. University of Florida. 128pp.

Pait, J. 2005. Clonal forestry: out of the lab, finally. *Proceedings, 28th Southern Forestry Tree Improvement Conference*, Raleigh, NC. Reforestation, Nurseries and Genetic Resources. Available at: http://www.rngr.net/Publications/sftic/.

Pot, D., Chantre, G., Rozenberg, P., Rodrigues, J.C., Jones, G.L., Pereira, H., Hannrup, B., Cahalan, C. and Plomion, C. 2002. Genetic control of pulp and timber properties in maritime pine (*Pinus pinaster Ait.*). Annals of Forest Science. **59**: 563-575.

Potts, B.M. 2004. Tree breeding, practices: genetic improvement of Eucalypts. Elsevier Ltd. 104–114.

Romagosa, I. and Fox, P.N. 1993. Genotype x environment interaction and adaptation. In: *Plant Breeding: Principles and prospects.* (Ed. Hayward, M.D., Bosemark, N.O. and Romagosa, I.). Chapman & Hall, London. 373-390.

Santos, P.E.T., Geraldi, I.O. and Garcia, J.N. 2004. Estimates of the genetic parameters of wood traits for sawn timber production in *Eucalyptus grandis*. Genetics and Molecular Biology. **27**: 567-573.

Shelbourne, C.J.A. and Low, C.B. 1980. Multi-trait index selection and associated genetic gains of *Pinus radiata* progenies at five sites. New Zealand Journal of Forestry Science. **10**: 307-324.

Smith, C. 1983. Effects of changes in economic weights on the efficiency of index selection. Journal of Animal Science. **56**: 1057-1064.

Snedden, C.L. and Verryn, S.D. 1999. An investigation into the occurrence and nature of genotype by environment interaction in *Pinus patula*. South African Forestry Journal. **186**: 67-75.

Snedden, C.L. and Verryn, S.D. 2003. *Matgen 6.1: A user's guide*. Environmentek, CSIR, Pretoria, South Africa.

Snustad, D.P. and Simmons, M.J. 2000. *Principles of genetics* (2nd ed.). John Wiley & Sons, Inc., U.S.A.

Squillace, A.E. 1974. Average genetic correlations among offspring from open pollinated forest trees. Silvae Genetica. **23**: 149-156.

Stern, K. 1964. Population genetics as a basis for selection. Unasylva. **18**: (2-3), 21-29.

Stonecypher, R.W. 1970. Multiple-trait breeding. Unasylva. **24**: (2-3), 48-51.

Stonecypher, R.W. and Arbez, M. 1976. Methods of Selection. Proceedings, Joint IUFRO *Meeting on Advanced Generation Breeding*. Bordeaux, June 14-18, Session II. 31-46.

Tuskan, G.A. 1997. Clonal forestry, heterosis and advanced-generation breeding. Reforestation, Nurseries and Genetic Resources. Available at: http://www.rngr.net/Publications/sftic/.

Van Wyk, G. 1983. Improving forestry yields through genetics. *SAIF Symposium. Forestry Quo Vadis?* Pietermaritzburg, South Africa. 73-87.

Van Wyk, G. and Falkenhagen, E.R. 1984. Genotype X Environment interaction in South African breeding material. *Proceedings, IUFRO Symposium on Site and Productivity of Fast Growing Plantations.* Pretoria, South Africa. South African Forestry Research Institute, Vol. **1**: 215-231.

Van Wyk, G. and Verryn, S.D. 2000. The basic principles of tree breeding in South Africa. *South African Forestry Handbook* (4th ed.). The Southern Africa Institute of Forestry, Vol. **1**: 61-68.

Vangen, O. 1979. Studies on a two trait selection experiment in pigs. Acta Agriculturæ Scandinavica. **29**: 305-319.

Verryn, S.D. 1993. Research and development report compiling estimates of genetic parameters in open pollination and diallel trials. *CSIR internal report.* FOR-DEA 601. 9pp.

Verryn, S.D. and Geerthsen, J.M.P. 2003. *Matgen 6.1 (Software).* Environmentek, CSIR, Pretoria, South Africa.

Verryn, S.D. and Snedden, C.L. 2000. Optimising the expected genetic gains of various breeding and selection strategies. *Proceedings, Forest Genetics for the next millennium.* IUFRO working party 2.08.01. ICFR, Durban. 240-243.

Verryn, S.D., Snedden, C.L. and Parfitt, R.C. 2000. Program for the deterministic modeling of genetic gains of tree breeding and seed and production strategies. Southern African Forestry Journal. **189**: 3-8.

Verryn, S.D. and Turner, P. 1999. The prediction and selection of *E. grandis* solid wood: Phase one. *CSIR Internal Report.* Environmentek, CSIR, Pretoria, South Africa.

Volker, P.W., Dean, C.A., Tibbits, W.N. and Ravenwood, I.C. 1990. Genetic parameters and gains expected from selection in *Eucalyptus globulus* in Tasmania. Silvae Genetica. **39**: 18-21.

Williams, J.S. 1962. The evaluation of a selection index. Biometrics. **18**: 375-393.

White, T.L. 1987. A conceptual framework for tree improvement programs. New Forests. **4**: 325-342.

White, T.L. 1995. Prototype breeding strategies for clonal forestry: a case study of Eucalyptus grandis. *Proceedings, IUFRO Conference, Eucalypt Plantations: Improving fibre yield and quality,* Theme 4, Hobart, Tasmania.

White, T.L. 1996. Genetic parameter estimates and breeding value predictions: Issues and Implications in tree improvement programs. *Proceeding, QFRI: IUFRO Conference, Tree Improvement for sustainable tropical forestry.* Caloundra, Australia. 110-117.

White, T.L. and Hodge, G.R. 1989. *Predicting breeding values with applications in forest tree improvement.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Wricke, G. and Eberhard Weber, W. 1986. *Quantitative genetics and selection in plant breeding.* Walter de Gruyter & Co., Berlin.

Wright, J.W. 1976. *Introduction of Forest Genetics.* Academic Press, Inc., New York.

Wright, J.W. 1995. Operational gains and constraints with clonal *Eucalyptus grandis* in Columbia. Contributed paper. *Proceedings, CRC/IUFRO Conference, Eucalypt Plantations: Improving fibre yield and quality.* Hobart, Australia. pp 308-310.

Wu, R., Zeng, Z., McKeand, S. and O'Malley, D. 2000. The case for molecular mapping in forest tree breeding. In: *Plant Breeding Reviews, volume 19.* (Ed. Janick, J). John Wiley & Sons, Inc., U.S.A. 41-68.

Young, S.S.Y. 1961. A further examination of the relative efficiency of three methods of selection for genetic gains under less restricted conditions. Genetical Research. **2**: 106-121.

Zobel, B.J. and Talbert, J. 1984. *Applied Forest Tree Improvement.* John Wiley & Sons, New York.

APPENDIX A

EXTRACT FROM MATGEN® (2003) OUTPUT, USING INDEX M1234 AS AN EXAMPLE

```
Matgen version 6.1 for Windows(TM) by S.D.Verryn & J.M.P.Geerthsen
Copyright (2003)
                                              Pg: 1   05/12/06 17:08:17


          BLP for DBH + Stem + Splitting + Density (4 Traits)
BLP Combined Individual and Family Mean Selection for CSPLIT CDBH CSTEM CDEN
Data file: Use .DBF
Input stored in file:  .TXT
Output file: M1234.PRN

GENETIC INPUT SUMMARY

Coefficient of Relationship  0.30

Genetic (co)variances (*cv)
              Trait    CSPLIT     CDBH      CSTEM     CDEN
              CSPLIT    0.2268   -0.0633   -0.1031    0.0346
              CDBH     -0.0633    0.4288    0.3478    0.0852
              CSTEM    -0.1031    0.3478    0.3713    0.0986
              CDEN      0.0346    0.0852    0.0986    0.5340

Within family (co)variances
              Trait    CSPLIT     CDBH      CSTEM     CDEN
              CSPLIT    0.9966   -0.0121   -0.1042    0.0107
              CDBH     -0.0121    0.7890    0.6341    0.0002
              CSTEM    -0.1042    0.6341    0.8025    0.0205
              CDEN      0.0107    0.0002    0.0205    0.8221

Economic weights
 CSPLIT    0.250
 CDBH      0.500
 CSTEM     0.180
 CDEN      0.070

SUMMARY STATISTICS

 Trait      Mean
 CSPLIT    -0.0237
 CDBH       0.2953
 CSTEM      0.2641
 CDEN       0.0000

Total number of observations:  475

LISTING CONDITIONS

All families listed.
All individuals of each family listed.
Observations regarded as missing in calculations if: source->split*source->dbh*source->stem*source->den = 0
```

Rank	Family	Index	TREE	REP	PLOT	SPLIT	CSPLIT	DBH	CDBH	STEM	CSTEM	DEN	CDEN
1	28820	0.76	773	3	270	32.039	1.505	520	2.047	7	1.028	0.678	1.287
2	28817	0.753	575	3	200	9.495	0.547	506	1.882	8	1.825	0.686	1.416
3	28820	0.75	427	2	150	30.464	1.358	451	1.316	6	0.313	0.798	3.14
4	28820	0.748	425	2	150	-11.469	-0.424	506	1.967	7	1.11	0.699	1.546
5	28876	0.74	270	2	96	46.592	2.043	570	2.723	7	1.11	0.565	-0.611
6	28952	0.723	355	2	126	12.296	0.586	522	2.156	7	1.11	0.797	3.124
7	28855	0.642	653	3	227	7.811	0.475	527	2.13	8	1.825	0.654	0.901
8	28817	0.634	173	1	62	-86.715	-3.62	512	1.873	7	1.321	0.638	0.637
9	28876	0.628	19	1	7	17.662	0.815	428	0.88	6	0.524	0.784	2.988
10	28820	0.624	34	1	12	0.163	0.071	496	1.684	8	2.119	0.64	0.669
11	28817	0.61	573	3	200	-4.374	-0.042	480	1.574	6	0.23	0.655	0.917
12	28877	0.61	443	2	156	-4.248	-0.117	508	1.99	8	1.907	0.608	0.081
13	28891	0.603	137	1	50	-60.755	-2.517	533	2.121	8	2.119	0.705	1.716
14	28892	0.601	143	1	52	29.466	1.316	535	2.145	6	0.524	0.585	-0.216
15	28876	0.598	21	1	7	21.502	0.978	490	1.613	6	0.524	0.648	0.798
16	28817	0.59	172	1	62	36.921	1.633	521	1.979	7	1.321	0.557	-0.667
17	28876	0.578	627	3	218	28.965	1.374	474	1.504	6	0.23	0.651	0.853
18	28855	0.532	57	1	20	-0.321	0.05	546	2.275	7	1.321	0.564	-0.554
19	28892	0.532	145	1	52	5.238	0.287	542	2.228	8	2.119	0.516	-1.327
20	28877	0.528	701	3	244	10.406	0.586	478	1.551	7	1.028	0.619	0.337
21	28820	0.527	426	2	150	13.514	0.638	388	0.571	6	0.313	0.765	2.609
22	28817	0.527	171	1	62	11.523	0.554	463	1.294	7	1.321	0.627	0.46
23	28822	0.519	748	3	260	2.665	0.257	571	2.65	8	1.825	0.536	-0.999
24	28889	0.507	212	1	76	18.356	0.844	534	2.133	7	1.321	0.574	-0.393
25	28839	0.499	231	1	83	-14.153	-0.537	475	1.436	7	1.321	0.693	1.523
26	28889	0.498	214	1	76	18.914	0.868	530	2.086	7	1.321	0.575	-0.377
27	28819	0.495	48	1	17	21.502	0.978	490	1.613	7	1.321	0.644	0.734
28	28817	0.477	574	3	200	22.338	1.093	423	0.901	7	1.028	0.653	0.885
29	28877	0.47	2	1	1	-1.794	-0.012	442	1.045	5	-0.273	0.675	1.233
30	28877	0.469	700	3	244	-0.477	0.123	470	1.456	7	1.028	0.596	-0.033
31	28892	0.466	764	3	-267	102.253	-4.201	489	1.681	7	1.028	0.595	-0.049
32	28817	0.456	341	2	121	-4.39	-0.123	407	0.796	7	1.11	0.667	1.031