Guidelines for testing and selecting parent palms in oil palm:
Practical aspects and statistical methods

C.J. Breure[1] and L.R. Verdooren[2]

SUMMARY

Practical aspects of implementing a three-step selection procedure in oil palm are described; these steps are: (i) phenotypic selection of *dura* and *pisifera* parents, (ii) further selection on the basis of General Combining Ability (GCA) values obtained from a progeny test and (iii) testing crosses between elite families to exploit both GCA and Specific Combining Ability (SCA).

A detailed account is given of the statistical method to estimate the GCA and SCA values of the parents from the *tenera* crosses. Parents can only be compared if the crosses are connected; this aspect is clarified. Incomplete block designs as for example alpha-designs are most suitable to compare the GCA values of the parents.

To compare sources of planting material, reliability can be enhanced by increasing the number of progenies per source rather than the number of replications.

Crossing work can be speeded by assigning the parents according to the sequence of emergence of inflorescences. The large number of crosses involved usually dictates the use of incomplete block designs, the choice and the statistical analysis of which are reviewed.

Among the statistical selection procedures studied, the Subset Selection Procedure of Gupta is most flexible and permits the elimination of inferior progenies.

The ultimate objective is to select parent palms for high yield of oil and kernels per ha. Selection for maintaining a high Harvest Index (HI), the proportion of dry matter used for the production of oil and kernels, is advocated. This can best be achieved through indirect selection for high GCA values of Leaf Area Ratio (LAR), (i.e. the ratio of new total leaf area produced to new Vegetative Dry Matter (VDM)), low GCA values for VDM, low values of height increment and high values of magnesium content in the leaves. A further aim is to achieve optimal Leaf Area Index (LAI), (i.e. the total leaf area per unit ground area), quickly after planting by the selection of parameters derived from a logistic growth function fitted through mean leaf area against palm age. Leaves are best measured 6, 42, 66 and 90 months after planting; precision is enhanced by including measurements 12 months later.

Palm height with time also fits a logistic growth function, so actual height must be measured to compare progenies.

Recording techniques of palm characteristics are described, illustrated by technical drawings, and a recording schedule is proposed.

Content

Relevant aspects

[1] PT Tania Selatam, JI. BasukiRahmat 788, Pelembang, Indonesia. 2. Department of Mathematics, Wageningen Agricultural University, Dreijenlaan 4, Wageningen. The Netherlands

1. INTRODUCTION

Selection in oil palm aims to increase the production of oil and kernels per ha. A selection strategy to develop elite planting material, involving large-scale testing of parent palms, has been formulated by Breure & Bos (1992). But information on the practical aspects to implement such a program is scarce. The main economic product is palm oil extracted from the mesocarp. The shell thickness is therefore an important characteristic as this determines the proportion of the fruit available for the oil bearing mesocarp.

Shell thickness is determined by a single gene. One homozygote, the *pisifera*, is shell-less; many *pisifera* palms fail to fruit, so the *pisifera* is not grown for commercial use. The other homozygote, the *dura*, has a thick shell. The heterozygote of the *dura* x *pisifera* cross, the *tenera*, has a thin shell. The *tenera* is the fruit form preferred for commercial use, because more of the pericarp consists of mesocarp than in the *dura*. As *pisifera* usually produce bunches with predominantly sterile fruits, the *dura* is used as the female and the *pisifera* as the male parent of *tenera* planting material. The search is thus for *dura* and *pisifera* parents which transmit high bunch yield and oil-and-kernel extraction per hectare to their *tenera* offspring.

Breure & Bos (1992) proposed to select *dura* and *pisifera* parents in three steps:

   i. preliminary selection of *dura* and *pisifera* palms on phenotypic characters, i.e. those characters measured on the parent palms,
   ii. further selection among these on the basis of General Combining Ability (GCA) values, i.e. the additive genotypic effects of the parents, obtained from a progeny test, and

iii. testing of families, derived from intercrossing palms selected in step (ii) in all combinations, with the main objective of exploiting both GCA and Specific Combining Ability (SCA), i.e. the contribution of the interaction effect of parents on the performance of the offspring.

Although the SCA effect is usually much less than the GCA effect, the ultimate aim is to search for specific crosses between *dura* and *pisifera* parents.

Within these elite crosses one can find outstanding *tenera*. Clones derived from these *tenera* may yield higher than the source family. Testing clones may therefore be the final stage in upgrading planting material.

The three-step selection of *dura* and *pisifera* parents is explained in section 2.1. Section 2.2 outlines the procedure of estimating the contribution of GCA and SCA effects on the performance of the offspring.

Section 2.3 shows how a proper choice of the mating design can enhance the precision in comparing GCA values of pairs of parents.

Usually, a large number of parents is involved in such testing trials; progress in crossing work depends on the availability of female inflorescences on the *dura* parents and males on the *pisifera*. An efficient method of implementing the crossing program is described in section 2.4. The lay-out of progeny experiments should be such that progenies are as much as possible arranged in blocks of uniform soil conditions. However, it is usually difficult to find sufficiently large blocks to accommodate the total set of progenies because of diversity in drainage and other physical and chemical soil characteristics. In that case progenies must be arranged in incomplete blocks, i.e. by subdividing replications into smaller homogeneous blocks containing only part of the progenies, as is illustrated in section 2.5.

The statistical methods to identify the set of elite parents in step 2 for further testing in step 3 is described in section 2.6. The way secondary characters are taken into account in the selection strategy is described in section 2.7. The general breeding strategy is described in section 2.8. Section 2.9 describes the size and shape of plots for experiments to select parent palms and also those to compare sources of planting material.

During the final stage of recording, differences in the progenies's ability to compete for light may create an important diversity in mutual shading among progenies; the stronger competitors then tend to gain yield at the expense of their weaker neighboring progenies. The way the effect of light competition on selection efficiency can be minimized by the arrangement of progenies in the field is described in section 2.10.

Plantation companies often like to test planting material from distinct origins at the onset of large-scale planting programs. Section 3.1 describes the trial design for evaluating different sources of planting material; its statistical analysis is given in section 3.2.

The actual technique of recording yield and growth has often received little attention. These include yield, oil and kernel extraction, secondary selection characters derived from records to determine the area and weight of the leaves, leaf production, trunk diameter and height increment.

Frequent and accurate measurements are needed to obtain meaningful parameters. Section 4.1 covers the recording of the components of oil and kernel yield and growth illustrated with technical drawings; in particular a novel technique to measure vertical stem growth is described.

Section 4.2 covers measurements to estimate growth parameters. Section 4.3 outlines the period of recording of the various parameters. Section 4.4 illustrates the components of growth and section 4.5 the calculation of growth parameters.

2. Relevant aspects of parent selection

2.1 Steps in selection

The first step is to select *dura* female parents in *dura* x *dura* crosses, which segregate into 100% *dura*; and to select *pisifera* male parents in *tenera* x *tenera* (25% *pisifera*) or in *tenera* x *pisifera* crosses (50% *pisifera*). Breure et al. (1987) describe the history of phenotypic parent selection from bunch yield and physical bunch analysis in early commercial plantings to the present method whereby all components of oil extraction rate, growth and leaf-Mg status are also taken into account. The method of growth recording was developed in Malaysia (Hardon et al., 1969; Corley et al., 1971), and an early account on how these measurements were applied in breeding work elsewhere was given by Breure et al. (1982).

The standard procedure is to select first *dura* x *dura* families. Individual *dura* female parents are then selected within these selected families based on more detailed recording, in terms of bunch analysis, growth and sometimes leaf-Mg level. Family selection in the source of *pisifera* male parents is handicapped by the occurrence of sterile *pisifera*. Yield and bunch analysis records are therefore restricted to *tenera* palms only. Once elite families are identified, *pisifera* selection can only be based on growth and magnesium level.

The genotypic performance of palms selected for phenotype (step 1 in the selection procedure) is still masked by all sorts of errors. Indeed, the correlation between phenotypic characters, in particular yield, of the *dura* parents and their *tenera* offspring is usually poor, while as mentioned before, *pisifera* cannot be selected for yield per se.

Fortunately, as in other crops, in oil palm the performance of the offspring can be quite accurately estimated by adding the genotypic effects of the female and male parents (Breure & Bos, 1992). Hence the expected yield of the *tenera* offspring, E(y), can then be described as a constant + the genotypic effect of the *dura* female parent + the genotypic effect of the *pisifera* male parent. These additive effects of the parents are in quantitative genetics termed General Combining Ability (GCA) values. Reliability of selection can therefore be greatly improved by selecting parents according to GCA values estimated from results of *dura* x *pisifera* crosses (step 2 in the selection procedure).

The additive model does, however, still not fully predict the performance of the *tenera* offspring; crosses may perform better or worse than estimated by adding GCA values of the parents. This deviation is due to the effect of Specific Combining Ability (SCA); but remember this SCA effect is usually much smaller than the effect of GCA.

To exploit both GCA and SCA effects the parents must be crossed in all combinations. Elite families selected among these crosses can be reproduced by crossing specific parents (step 3).

Le Guen et al. (1991) reported a considerable increase in yield and extraction rate of *tenera* clones above the mean of the palms of the *dura* x *pisifera* family from which the *tenera* was selected. This is not surprising as *tenera* can be considered as relatively heterozygous while the *dura* and *pisifera* parents are relatively homozygous genotypes. Recombination of favorable alleles in response to the cross can therefore generate an excess of heterozygous and thus superior genotypes. Cloning these elite *tenera* palms yield thus palm populations consisting of homogeneous and heterozygous genotypes. The final phase in improving planting material therefore appears to clone outstanding *tenera* selected within elite families of step 3.

The female parents tested in step 2 and step 3 are reproduced as *dura* selfings. Selfings of all parents tested are planted at the same time as the step 2 parent testing program is established. Female parents for seed production are then selected in selfings of *dura* which are selected on the basis of the outcome of the test crosses.

It has been shown in maize (Hallauer & Miranda, 1981; p. 281-283) that there is a sufficiently large correlation between performance per se of the inbred lines and their GCA value calculated from all crosses obtained from these lines. For this reason the *dura* selfings are also recorded in detail to support results obtained from the *dura* x *pisifera* test crosses. The aim is to clone all *pisifera* tested in step 2 and step 3; as a safeguard, in case cloning fails, *pisifera* are crossed with an elite *tenera* in the same family. Either clones or *pisifera* selected in elite *tenera* x *pisifera* crosses are selected for seed production.

2.2 Estimating genetic effects of the parents

As the content of this section may not be common knowledge for breeders, we will first describe some basic aspects of the statistical methods used for the interested reader. Later on this will be illustrated with data in Example 2 in section 2.2.1 and Appendix 2 and in Example 3 in Appendix 3.

2.2.1 General combining ability (GCA)

The genotypic effects of the *dura* female and the *pisifera* male parents can be estimated from the performance of their *tenera* offspring provided that these effects are additive. The expected yield of the *tenera* offspring of the cross D_ixP_j, $E(y_{ij})$, can then be written as the sum of a general constant, μ, the genotypic effect $_i$ of the *dura* mother D_i and the genotypic effect $_j$ of the *pisifera* father P_j :

$$E(y_{ij}) = \mu + a_i + b_j .$$

In quantitative genetics these additive effects of the parents are called General Combining Ability (GCA) values.

For a set of C crosses, derived from A *dura* and B *pisifera*, where $C \leq A*B$, the parameters μ, $\alpha 1$, ..., αA, $\beta 1$, ..., βB can be estimated using the Least Squares Method. Assume that there are n_{ij} plots available of a certain *tenera* cross $D_i x P_j$; if no cross has been made then $n_{ij} = 0$. Let us consider first the case that we have used a completely randomized design (CRD), i.e. the plots are allotted at random to the progenies. (In section 2.5 we will consider the use of an incomplete block design to compare the *tenera* offsprings.)

In the following example we have made C=16 crosses between A=5 *dura* and B=4 *pisifera*. In the experimental field there were 20 plots available and a completely randomized design (CRD) was used. The number of asterisk (*) in the table below shows how many plots are used by a certain cross; hence two asterisk means two plots etc. So $n_{11}=2$, $n_{12}=1$, $n_{13}=0$ (no cross of *dura* 1 x *pisifera* 3), $n_{14}=1$, etc.

EXAMPLE 1

pisifera

	1	2	3	4
dura 1	**	*		*
dura 2	*		*	**
dura 3	*	*	*	*
dura 4		**	**	
dura 5	*	*	*	*

The actual yield y_{ijk} of the k-th plot of a *tenera* offspring of the cross $D_i x P_j$ is a random sample of the population of all possible observations from this cross with population mean or expectation $E(y_{ijk})$ and variance σ^2; hence, the statistical model is $y_{ijk} = E(y_{ijk}) + e_{ijk}$, where e_{ijk} is the effect of the environment or error on this k-th plot. These error-terms e_{ijk} are such that the expectation $E(e_{ijk}) = 0$ and the variance $Var(e_{ijk}) = \sigma^2$; these errors are uncorrelated with one another because we have randomized the plots over the crosses. When one uses a randomization procedure to allot the plots of a field to the crosses, as with a completely randomized design (CRD), then the plot-errors can be assumed to be uncorrelated.

With such a model for the yields, this Least Squares Method searches estimates m, α_i, and β_j for the parameters μ, α_i and β_j respectively, such that the sum of the squared deviations between the observation and the estimate of their expected value for k=1,...,n_{ij}, i=1,...,A, and j=1,...,B, $\sum_i \sum_j \sum_k [y_{ijk} - (m+a_i+b_j)]^2$ is minimal.

Good statistical packages such as SAS, SPSS, SYSTAT, BMDP and GENSTAT can provide these Least Squares estimates for the parameters. For the Normal Equations and their solution see Appendix 1.

From the Least Squares estimates for the parameters one can calculate the Least Squares Mean for a *dura* D_i, LSM(D_i), as $m+\alpha_i + \beta b_j /B$ and the Least Squares Mean for a *pisifera* P_j, LSM(Pj),

as m + $\sum \alpha_i$ /A +β_j . We can rank all the *dura* and the *pisifera* according to their General Combining Ability on their Least Squares Means LSM (D_i) and LSM(P_j), provided that the crossing scheme is connected. The term connected crossing scheme will be explained below.

If we make all possible crosses between A *dura* and B *pisifera*, hence we have C=A*B crosses, the crossing scheme is called a complete diallel scheme; if the number of crosses C is less than A*B the crossing scheme is called an incomplete diallel scheme.

In a complete diallel scheme with C=A*B crosses, where each cross has the same number of plots, n_{ij} = n for i=1,...,A and j=1,...,B , the Least Squares Mean for a *dura* or *pisifera* is just the average of the observations.

In this case LSM(D_i)=$y_{i..}$/(n*B), and LSM(P_j) = $y_{.j.}$ /(n*A).A complete diallel is always connected.

For an incomplete diallel scheme with C<A*B crosses one must use a good statistical package to get the Least Squares Means for the *dura* and the *pisifera* parents. The difficulty with an incomplete diallel scheme is that it can be disconnected and not all statistical packages notice this. Furthermore, a good statistical package provides the estimate for the common variance σ^2 as the Mean Square Error (or Mean Square Residual) from the Analysis Of Variance (ANOVA) table.

In order to compare the entire set of the A *dura* and the B *pisifera* on the basis of the GCA values the parents must be crossed according to a so-called connected crossing scheme. A crossing scheme is called connected if for each *dura* pair (D_h,D_i) of the A *dura*, there is a chain of *dura* from *dura* D_h to *dura* D_i, in which each of the adjacent links of the chain occur together with the same *pisifera*. Otherwise the crossing design is called disconnected. In the same vein, the crossing scheme is connected if for each *pisifera* pair (P_k,P_j) of the B *pisifera*, there is a chain of *pisifera* from *pisifera* P_k to *pisifera* P_j, in which each of the adjacent links of the chain occur together with the same *dura*. Another way to check whether the crossing scheme is connected, is to form a two-way table of the crosses with the A *dura* as rows and the B *pisifera* as columns. The crossing scheme is connected if we cannot split the table in separate tables by interchanging rows and columns.

Let us elucidate this by a little example with C=8 crosses made from A=4 *dura* and B=4 *pisifera*. Let the realized crosses be indicated by an asterisk (*) in the table.

From the cross of *dura* D_1 with *pisifera* P_1, D_1xP_1, we can make a chain to the cross of *dura* D_3 with P_1, D_3xP_1; from D_3xP_1 we can go to the cross D_3xP_3, and from this cross D_3xP_3we can go to the cross D_1xP_3, and then we come back to the cross D_1xP_1. In this chain we have missed *dura* D_2 and D_4. Hence this crossing scheme is disconnected.

Pisifera		P1	P2	P3	P4
Dura	D1	*		*	
	D2		*		*
	D3	*		*	
	D4		*		*

When we have rearranged the table as follows (interchange P_3 with P_2 and also interchange D_3 with D_2),

Pisifera		P1	P2	P3	P4
Dura	D1	*	*		
	D2	*	*		
	D3			*	*
	D4			*	*

we see directly that there are two disconnected sets of four crosses each. The first set contains the connected crosses D_1xP_1, D_1xP_3, D_3xP_1 and D_3xP_3; the second set contains the connected crosses D_2xP_2, D_2xP_4, D_4xP_2 and D_4xP_4. In such a disconnected crossing scheme no unbiased estimate can be made for the difference in effect between, for example, *dura* D_1 and D_2 or for the difference in effect between *pisifera* P_3 and P_4.

A more practical method of checking whether a crossing scheme is connected is to draw a chain from one cross to another following a horizontal or vertical direction only. If all the crosses are connected by one continuous chain then the crossing scheme is connected. In the above mentioned example the crossing scheme is connected if e.g. the following 8 crosses were made:

Pisifera		P1	P2	P3	P4
Dura	D1	*	*		
	D2		*	*	
	D3			*	*
	D4	*			*

A necessary condition to have a connected design is that the number of crosses C must be at least A+B-1. In the example above we have A=4 and B=4, so 4+4-1=7 crosses sufficient for a connected design. But we have 8 crosses and the crossing scheme is still connected if, for example, the cross D_4xP_1 was not made. But we must realize that this condition $C \geq A+B-1$ is not sufficient. We must always check for connectedness by making a continuous chain through the crosses of the crossing scheme.

EXAMPLE 2

Assume that C=9 progenies (2 plots each), from A=5 *dura* and B=3 *pisifera*, are tested in a completely randomized design. We assume an additive model for the genetic effects of the *dura* and *pisifera* parents. Yield records (kg/plot) were as follows:

		Pisifera							
		P1		P2		P3		Total	
Dura	D_1	44	48					92	
	D_2	45	42	45	43			175	
	D_3	33	36	35	32	36	38	210	
	D_4			44	42	46	48	180	
	D_5					53	55	108	
Total		248		241		276		765	

This crossing design is connected because there is one continuous chain which connects all the crosses. For the analysis of this Example 2 see Appendix 2.

A solution of the Normal Equations gives:

m=54; a_1= -4.8333333; a_2= -6.9 ;

a_3= -16.7666667; a_4= -7.2333333; a_5= 0;

b_1= -3.1666667; b_2= -3.5333333; b_3= 0.

The estimate for the common variance $\square^2$ is 3.00303 based on 11 degrees of freedom.

A breeder is not interested in testing the hypothesis that all *dura* effects (or all *pisifera* effects) are the same, but is much more interested in how much pairs of *dura* or *pisifera* can be different in GCA values.

The difference between two *dura* effects, for example D_1 and D_2 is $\alpha_1 - \alpha_2$ and is estimated by $\alpha_1 - \alpha_2 = 2.06667$ with an estimated standard error of 1.61327.

The difference between two *pisifera*-effects, for example $P_2 - P_3$ is $\beta_2 - \beta_3$ is estimated by $\beta_2 - \beta_3 = -0.26667$ with an estimated standard error of 1.183813.

It is often reasonable to assume that the error-terms (and hence the yields) are Normally distributed, so we can construct for example a 95%-confidence interval for differences between the General Combining Abilities of the parents. Let us illustrate this for a 95% confidence interval for such differences. The 5% two-sided significance point for a t-distribution with 11 degrees of freedom is 2.201. Hence the 95%-confidence limits for $\alpha_1 - \alpha_2$ are $2.06667 \pm 1.61327*2.201 = 2.06667 \pm 3.55081$ and the 95%-confidence interval is $-1.48414 < \alpha_1 - \alpha_2 < 5.61748$.

In the same way the 95% confidence limits for $\beta_2 - \beta_3$ are calculated as $-.26667 \pm 1.183813*2.201 = -0.26667 \pm 2.60557$ and the 95%-confidence interval is $-2.87224 < \beta_2 - \beta_3 < 2.33890$.

Note that if a 95% confidence interval for the difference of two parental effects contains zero, then this means that the null hypothesis "These two parental effects are equal" is not rejected with a significance level 5%. If the 95% confidence interval for the difference of two parental effects does not contain zero, this null-hypothesis of equal parental effects is rejected with a significance level of 5%.

To rank the parents according to their GCA values, we can for example use the Least Squares Mean (LSM). The Least Squares Mean of for example *dura* D_1 is estimated by

$$m + a_1 + (p_1 + p_2 + p_3)/3 = 54 + (-4.83333) + [(-3.16667) + (-3.53333) + 0]/3 = 46.93333, \text{etc.}$$

Dura	LSM	rank	*Pisifera*	LSM	rank
D1	46.9333	2	P1	43.6867	2
D2	44.8667	3	P2	43.32	3
D3	35. 000	5	P3	46.8533	1
D4	44.5333	4			
D5	51.7667	1			

We can get the same ranking of parents according to their GCA values, if we use a solution of the Normal Equations for these parental effects.

Normal Equations *Dura*	solution	rank	Normal Equations *Pisifera*	solution	rank
D1	-4.83333	2	P1	-3.16667	2
D2	-6.90000	3	P2	-3.53333	3
D3	-16.7667	5	P3	0	1
D4	-7.23333	4			
D5	0.00000	1			

2.2.2 Specific combining ability (SCA)

Sometimes the additive model of the genetic effects of the parents do not fully explain the performance of their offspring. This is attributable to an interaction effect of the genetic effects of the parents. In other words, besides the additive genetic effects (General Combining Ability) of the parents there is also a specific interaction effect due to the specific combination of the parents. This specific interaction effect is called in quantitative genetics Specific Combining Ability (SCA). For this interaction model the expected yield of the *tenera* offspring of the crossing D_ixP_j, $E(_{yij})$, can then be written as the sum of a general constant, μ *, the GCA effect α_i* of the *dura* mother D_i, the GCA effect β_j* of the *pisifera* father and the SCA effect $(\alpha\beta)_{ij}$* of the realized cross:

$$E(Y_{ij}) = \mu^* + \alpha_i + \beta_j + (\alpha\beta)_{ij} \ 8 = \mu_{ij}$$

When we have a set of C crosses, derived from A *dura* and B *pisifera*, where $C \leq A*B$, the C parameters μ_{ij} can be estimated using the Least Squares Method.

Assume that there are n_{ij} plots available for a certain *tenera* cross D_ixP_j; in the case that there has no cross been made then $n_{ij} = 0$. We consider here the case that we have used a completely randomized design (CRD). In section 2.5 we will consider the case that we have used an incomplete block design.

The actual yield y_{ijk} of the k-th plot of a *tenera* offspring of the cross D_ixP_j is $y_{ijk} = E(y_{ijk}) + e_{ijk}$, where e_{ijk} is the effect of the environment or error on this k-th plot. These errors e_{ijk} are such that the expectation $E(e_{ijk}) = 0$ and the variance $Var(e_{ijk}) = \sigma^2$; these errors are uncorrelated with one another. When one uses a randomization procedure to allot the plots of a field to the crosses, such as in a completely randomized design (CRD), then the plot-errors can be assumed to be uncorrelated.

The Least Squares Method searches estimates m_{ij} for the parameters μ_{ij} such that the sum of the squared deviations between the observation and the estimate of their expected value for k=1,...,n_{ij}, i=1,...,A, and j=1,...,B, $\sum_i\sum_j\sum_k [y_{ij} - m_{ij}]^2$ is minimal.

The Least Squares estimates m_{ij} for the parameters ij are found as solutions of the Normal Equations, which are in this case very easy.

Let us denote the sum of the observations of the n_{ij} plots of the cross $D_i x P_j$ by $y_{ij.}$,

hence $\sum_k y_{ijk} = y_{ij.}$. The Normal Equations are then:

$n_{ij} * m_{ij} = y_{ij.}$ (4)

for i=1,...,A and j=1,...,B . There are only C Normal Equations present, because if a certain offspring $D_i x P_j$ has not been realized, then $n_{ij} = 0$ for such a progeny and we have no observations of this progeny. The parameter estimates are then $m_{ij} = y_{ij.} / n_{ij}$, the progeny means of the crosses $D_i x P_j$. To estimate the Specific Combining Abilities of these progenies we must now calculate the estimates m for μ, a_i for α_i and b_j for β_j for the parameters according to an additive model

$E(y_{ijk}) = \mu + \alpha_i + \beta_j$

As has been explained in section 2.2.1. The estimate for the Specific Combining Ability $(\alpha\beta)_{ij}*$ is

$(ab)_{ij}* = m_{ij} - (m + a_i + b_j)$.

See Appendix 3 for the analysis and Example 3.

R EMARK

For many characteristics of oil palms the Specific Combining Ability is not so large. Hence for a first screening of parents the additive model to estimate the General Combining Abilities is a good tool.

2.3 Evaluating mating designs

Assume that we want to make C crosses derived from A *dura* and B *pisifera*. If C=A*B then we have only one mating design, a complete *diallel* crossing scheme. In an incomplete *diallel* where the number of crosses C is less than A*B there are many possible mating designs. As we have already explained in section 2.2.1, the entire set of A *dura* and B *pisifera* can only be compared on the basis of the GCA values if the crossing scheme is connected. A necessary condition for a connected scheme is that C≥ A+B-1, but this condition is not sufficient. For a given number of crosses C in an incomplete *diallel* scheme, where A+B-1 ≤ C <A*B, the problem is now to find a good connected mating design. The choice between several connected mating designs can best be based on the standard error of the estimator for the difference in the GCA value of all the *dura* pairs and the *pisifera* pairs. The standard error of the estimator for the difference in the GCA value between two *dura* parents or *pisifera* parents P_i and P_j is $S_{ij}*\sigma$, where σ is the residual standard deviation and the value of S_{ij} depends solely on the mating scheme. The value of *s* depends on the studied trait, the variation between the plots in the experimental field and the plot size.

As we have already explained in section 2.2.1 (see Example 2), the GCA values are estimated by solving the Normal Equations M * p=t, where M is the matrix of the coefficients in the Normal

Equations, p is the column of parameters from the linear additive model of the *dura* and *pisifera* genetic effects, and t is the column of the totals in the right-hand side of the Normal Equations. A solution of these Normal Equations can be written as $p = M^- * t$, where M^- is a generalized inversion of M. In other words it fulfills the condition that $M * M^- * M = M$.

A difference of the GCA values between two parents P_i and P_j will be estimated as the difference between the estimates of the parameters of these parents P_i and P_j and this is the same for each solution of the Normal Equations. This difference is a linear combination of the parameter-estimates and can be written in matrix notation as $c' * p$; the variance of $c' * p$ is given by $(c' * M^- * c) * \sigma^2$. The standard error of this estimator $c' * p$ is the square root of the variance of $c' * p$,

hence $S_{ij} \sqrt{(c' * M^- * c)}$

From a mating design follows the matrix M and hence the Sij depends solely on the mating design. For complete crossing schemes (as a complete *diallel*) with A *dura* and B *pisifera* (each cross occurs on r plots) the standard error of the estimator of the difference between the GCA values of the *dura* parents is the same for all pairs of *dura* and S_{ij} is $\sqrt{2/(b * r)}$; also the standard error of the estimator of the difference between the GCA values of the *pisifera* parents is the same for all pairs of *pisifera* and S_{ij} is $\sqrt{2/(A * r)}$.

For incomplete mating designs we can get the same standard error for the estimator of the difference between the GCA values of the *dura* parents if these *dura* are balanced over the *pisifera*. This means that each *pisifera* has the same number of k (<A) *dura* and each pair of *dura* occurs the same number of times together with a *pisifera*. In this case we have an incomplete balanced mating design.

For incomplete unbalanced mating designs the standard error of the estimator of the differences in GCA values varies across the parents. The quality of such mating designs can be measured by the average and range of the standard errors of the estimator of the differences between the GCA values of a pair of *dura* parents or of a pair of *pisifera* parents. As shown above, such quality evaluation can solely be based on S_{ij} values.

To find a good mating design one can search for balanced or partially balanced incomplete mating designs. For such incomplete mating designs one can use the incomplete block designs (see section 2.5). In such incomplete block designs there must be compared v treatments in blocks of sizes of k plots, where the block size k <v. Well known incomplete block designs are lattices where v = k*k or rectangular lattices where v = k*(k+1). (See Cochran & Cox, 1957). To extend the possibilities for v unequal to k*k or k*(k+1) there are the so called alpha-designs (see Patterson, Williams and Hunter, 1978). To use such an incomplete block design the role of the treatments is played by the *dura* and the role of the incomplete blocks is played by the *pisifera*. So we must look for incomplete block designs with A treatments and B blocks. The block size k is then chosen as C/B, where C is the number of crosses used. If there is no incomplete block design which fits the requirements, we can always start from a smaller design and add some extra treatments (=*dura*) to the blocks (=*pisifera*).

As an example we give here some mating designs involving C=40 crosses among A=20 *dura* and B=10 *pisifera*. In these designs each *dura* must be crossed with two *pisifera*; furthermore each *pisifera* must be crossed with four *dura*.

Two designs (I and II) were solely chosen intuitively on the basis of symmetry by two experienced oil palm breeders and the last design (III) is an alpha-design. A realized cross is indicated by an asterisk (*).

Design I

Pisifera	1	2	3	4	5	6	7	8	9	10
Dura 1	*								*	
2	*								*	
3	*	*								
4		*				*				
5		*	*							
6			*			*				
7			*	*						
8				*		*				
9				*	*					
10					*	*				
11					*	*				
12					*	*				
13						*	*			
14			*				*			
15							*	*		
16		*					*			
17								*	*	
18	*							*		
19									*	*
20	*									*

Design II

Pisifera	1	2	3	4	5	6	7	8	9	10
Dura 1	*								*	
2	*				*					
3	*	*								
4		*				*				
5		*	*							
6			*				*			
7			*	*						
8				*				*		
9				*	*					
10					*				*	
11					*	*				
12	*					*				
13						*	*			
14		*					*			
15							*	*		
16		*						*		
17								*	*	
18			*						*	
19									*	*
20					*					*

For each design one can calculate beforehand the average of the S_{ij}-values of the standard errors of the estimator of the differences between the GCA values of the *dura* and *pisifera* parents as well as their range.

In the following table the minimum, maximum and average of the standard errors of the estimator of the difference between GCA values of pairs of *dura* and *pisifera* parents ($S_{ij}*\sigma$), divided by σ, for the mating Designs I, II (constructed by the experienced oil palm breeders) and III (based on an alpha-design) are given.

Design III

Pisifera

Dura	1	2	3	4	5	6	7	8	9	10
1	*					*				
2		*					*			
3			*					*		
4				*					*	
5					*					*
6	*									*
7		*				*				
8			*				*			
9				*				*		
10					*				*	
11	*									*
12		*							*	
13			*			*				
14				*			*			
15					*			*		
16	*							*		
17		*							*	
18			*							*
19				*		*				
20					*		*			

Mating design	Pairs of *dura*			Pairs of *pisifera*		
	min	max	average	min	max	average
I	1.000	2.236	1.561	0.765	2.072	1.417
II	1.000	1.483	1.313	0.841	1.25	1.125
III	1.125	1.291	1.214	0.949	1.08	1.001

From the table it is clear that design III (the alpha-design), which has the smallest average value for Sij for the *dura* and the *pisifera* pairs, and moreover has the smallest range (max - min) for Sij, is the mating design which must be preferred. Hence it is worthwhile to use an alpha-design for a mating design and be careful to rely too much on "experience"!

2.4 Implementing a crossing program

Oil palm produces male and female inflorescences in separate leaf axils on the same palm, one in each axil. A succession of several inflorescences of one sex is followed by a succession of the other.

Preparations for making *dura* x *pisifera* crosses involve collecting pollen from male inflorescences of *pisifera* palms and isolating (bagging) female inflorescences on *dura* palms. For large-scale crossing designs it is necessary to complete several hundreds of crosses. An example of a connected crossing design involving 225 *dura* and 50 *pisifera* with 9 *dura* crossed onto one *pisifera* is given in <u>Table 1</u>.

As germinated seeds should be planted at about the same time, it is crucial to complete a crossing program quickly to avoid differences in the speed of germination rate due to the age of the seeds.

Pollen can easily be stored in ampoules sealed under vacuum for a year, so pollen collection can proceed regardless of whether palms bear anthesizing female inflorescences.

Progress therefore mainly depends on the rate of emergence of female inflorescences on the *dura*, particularly when female parents are assigned to a specific *pisifera* before actual crossing work starts. Speed of completing the crossing program can be increased if numbers are assigned to the female parents according to the emergence of inflorescences. For example, in Table 1 the first 9 *dura* which produce female inflorescences are crossed with *pisifera* 1 and numbered 1, 26, 51, 76, 101, 126, 151, 176 and 201; *pisifera* 2 is crossed onto the second set of nine *dura* with females inflorescences which are numbered 2, 27, 52, 77, 102, 127, 152, 177 and 202.

The first 25 *pisifera* are in this way crossed onto the total of 225 *dura*, the numbers of which are assigned as female inflorescences become available. *Pisifera* 26 to 50, however, must be crossed with the specific 225 *dura* for which the numbers were already assigned when crossed onto *pisifera* 1 to 25.

Once pollen collection has started, initial progress depends thus mainly on the rate of availability of female inflorescences of different *dura*. On the other hand, availability of male inflorescences from *pisifera* 26 to 50 tend to become the constraint for completing the final stage of the crossing program. For that reason, a spare set of *pisifera* should be reserved to ensure that the program can be finalized in time. In the same vein, nominating an additional set of *dura* will speed up completion of the first part of the crossing program.

2.5 Progeny trials in incomplete blocks
It is often difficult to find a sufficiently large homogeneous area to test the number C of *tenera* crosses or progenies from *dura* mothers and *pisifera* fathers in a Completely Randomized Design. The experimental field has often only homogeneous parts which are so large that they can only contain a part of the crosses; in such a homogeneous part (block) of the experimental field the crosses can be compared under the same conditions. To take care of the heterogeneous growing conditions in an experimental field one can use a Randomized Incomplete Block design. If all the progenies are present in a block with size k, it is called a complete block; block size k = C. But often the block (homogeneous part of the experimental field) is not large enough to contain all the progenies; block size k<C. In that case an incomplete block design is used. The well known incomplete block designs, such as balanced incomplete block designs (BIBD), partially balanced incomplete block designs with two associate classes (PBIBD with 2 associate classes), lattices (for the case of C=k*k) and rectangular lattices (for the case of C=k*(k+1)) can be found in the book of Cochran and Cox (1957). Often the number of tested progenies C does not fit with the above mentioned incomplete block designs as given in the book of Cochran and Cox (1957). An extension of the incomplete block designs is given by Patterson, Williams and Hunter (1978). They introduced the so-called alpha-designs. For many combinations of progenies C and block sizes k they give a procedure to construct these alpha-designs.

Table 1. Crossing design for 225 *dura* and 50 *pisifera* with nine *dura* per *pisifera*

Pisifera	Dura								
1	1	26	51	76	101	126	151	176	201
2	2	27	52	77	102	127	152	177	202
3	3	28	53	78	103	128	153	178	203
4	4	29	54	79	104	129	154	179	204
5	5	30	55	80	105	130	155	180	205
6	6	31	56	81	106	131	156	181	206
7	7	32	57	82	107	132	157	182	207
8	8	33	58	83	108	133	158	183	208
9	9	34	59	84	109	134	159	184	209
10	10	35	60	85	110	135	160	185	210
11	11	36	61	86	111	136	161	186	211
12	12	37	62	87	112	137	162	187	212
13	13	38	63	88	113	138	163	188	213
14	14	39	64	89	114	139	164	189	214
15	15	40	65	90	115	140	165	190	215
16	16	41	66	91	116	141	166	191	216
17	17	42	67	92	117	142	167	192	217
18	18	43	68	93	118	143	168	193	218
19	19	44	69	94	119	144	169	194	219
20	20	45	70	95	120	145	170	195	220
21	21	46	71	96	121	146	171	196	221
22	22	47	72	97	122	147	172	197	222
23	23	48	73	98	123	148	173	198	223
24	24	49	74	99	124	149	174	199	224
25	25	50	75	100	125	150	175	200	225
26	1	27	54	83	112	138	170	196	223
27	2	28	55	84	113	139	171	197	224
28	3	29	56	85	114	140	172	198	225
29	4	30	57	86	115	141	173	199	201
30	5	31	58	87	116	142	174	200	202
31	6	32	59	88	117	143	175	176	203
32	7	33	60	89	118	144	151	177	204
33	8	34	61	90	119	145	152	178	205
34	9	35	62	91	120	146	153	179	206
35	10	36	63	92	121	147	154	180	207
36	11	37	64	93	122	148	155	181	208
37	12	38	65	94	123	149	156	182	209
38	13	39	66	95	124	150	157	183	210
39	14	40	67	96	125	126	158	184	211
40	15	41	68	97	101	127	159	185	212
41	16	42	69	98	102	128	160	186	213

42	17	43	70	99	103	129	161	187	214
43	18	44	71	100	104	130	162	188	215
44	19	45	72	76	105	131	163	189	216
45	20	46	73	77	106	132	164	190	217
46	21	47	74	78	107	133	165	191	218
47	22	48	75	79	108	134	166	192	219
48	23	49	51	80	109	135	167	193	220
49	24	50	52	81	110	136	168	194	221
50	25	26	53	82	111	137	169	195	222

All these above mentioned designs are connected. A block design is called connected if for each pair (A_i, A_j) of the C progenies A1,...,AC, there is a chain, $A_i = A_{(1)},...,A_{(m)} = A_j$, in which each two adjacent links of the chain occur together in the same block. The block design is otherwise called disconnected. In a connected block design one can estimate all differences between the progenies.

But later on we also want to estimate, from the yield of the C *tenera* progenies, the General Combining Abilities of the A *dura* mothers and the B *pisifera* fathers. Therefore we must use a connected crossing design for the *dura* and the *pisifera*.

The model for an incomplete block design with C progenies and NB incomplete blocks is such that the expected yield $E_{(ygh)}$ of a *tenera* offspring T_g (g=1,...,C) of a *dura* mother D_i (i=1,...,A) and a *pisifera* father P_j (j=1,...,B), which is allotted to a plot in an incomplete block Bl_h (h=1,...,NB), can be described as the sum of a general constant ϕ, an effect τ_g of the *tenera* T_g and an effect δ_h of the block Bl_h, hence

$$E(y_{gh}) = \phi + \tau_g + \delta_h$$

for g=1,...,C and h=1,...,NB .

The yield y_{gh} of the progeny T_g in the block Bl_h can be described as $y_{gh} = E(y_{gh}) + e_{gh}$, where e_{gh} is the environmental effect or plot error with expectation $E(e_{gh}) = 0$ and variance $Var(e_{gh}) = \sigma^2$, these errors are uncorrelated. Because we have allotted the plots of a block at random to the progenies, which must be tested in this block according to the design, this assumption of uncorrelated errors is reasonable.

The model described for y_{gh} is an additive model of the *tenera* effects and the block effects. In section 2.2.1 we have already described how the parameters of an additive model can be estimated with the Least Squares Method. To estimate these parameters we must solve the so-called Normal Equations. See Appendix 4 for the analysis of an incomplete block design.

Good statistical packages such as SAS, SPSS, SYSTAT, BMDP or GENSTAT can give us a solution of the Normal Equations; an estimate for the variance σ^2 follows from the ANOVA table as Mean Square Residual (or Error).

To calculate the estimates of the General Combining Abilities of the A *dura* and the B *pisifera*, which are used to derive the C crosses, we construct a two-way table with A rows for the *dura* and with B columns for the *pisifera*. In a cell $D_i x P_j$ we insert the Least Squares Mean for $D_i x P_j$, according to the number of plots of this cross in the experimental field. This two-way table is

then analyzed according to an additive model for the *dura* and *pisifera* effects, as has been described in section 2.2.1. The Least Squares estimates for μ, for the GCA *dura* effect α_i and for the GCA *pisifera* effect β_j are m, a_i and b_j respectively. The estimate for the expected value of a cross $D_i x P_j$ according to the additive model of the GCA values is then

$m + a_i + b_j$.

This procedure gives sufficiently accurate General Combining Abilities estimates to rank the *dura* and the *pisifera*. Note that this procedure is an approximate procedure to estimate the General Combining Abilities, using a randomized incomplete block design to compare the C crosses. A more elaborate analysis would need a three-way table analysis according to blocks, *dura* and *pisifera*. But the two-step analysis gives results in a very good approximation; in the case of a completely randomized design or a randomized complete block design the two-step procedure gives us the correct estimates.

Furthermore, with a large crossing trial in an incomplete block design one can get difficulties with the size of the classifications to analyze it directly as a three-way classification with a statistical package. In the case of disconnected crossing schemes, one can always analyze the connected parts of the crossing schemes separately. In each connected part insert the Least Squares Means of the *tenera* offsprings. The ranking of the *dura* and *pisifera* parents belonging to the connected parts of this crossing scheme can then be given.

To calculate the estimates of the Specific Combining Abilities we must calculate the difference between the Least Squares Mean of a *tenera* Tg and the estimate of the expected value of this cross according to the additive model of GCA values $m + a_i + b_j$, hence

SCA (T_g) = LSM (T_g) - $(m + a_i + b_j)$.

If we may assume that the errors (and hence the yields) are normally distributed, then we can test whether an additive model for GCA values is reasonable, otherwise stated that the SCA values are equal.

 To test the null-hypothesis "The SCA values are equal" we use the test-statistic

$$F = \frac{SS(SCA) \, / \, df(SCA)}{SS(res) \, / \, df(res)}$$

The sum of squares for the SCA values, SS(SCA), can be calculated as the sum of the squared SCA values. The degrees of freedom of this SS(SCA) is df (SCA)=C-(A+B-1), where C=number of the *tenera* crossings in the connected crossing scheme, A=number of *dura* and B=number of *pisifera*. If F >F(α %) the null-hypothesis is rejected at significance level α %, where F(α %) is the right-sided α %-point of the F-distribution with df (SCA) and df (res) degrees of freedom.

EXAMPLE 4

Let us consider the case that we have made C=10 connected *tenera* crosses derived from A=5 *dura* mothers and B=5 *pisifera* fathers. In the following table the crossing scheme is given; a dot (.) indicates a cross which has not been made.

Pisifera

P1 P2 P3 P4 P5

Dura D1 T1 .   .   .   T10

D2 T2 T3 .   .   .

D3 .   T4 T5 .   .

D4 .   .   T6 T7 .

D5 .   .   .   T8 T9

Suppose that the experimental field is very heterogeneous, and that we can only find homogeneous parts (blocks) of maximal four plots. We want to have each progeny tested on four plots. An alpha-design with block size 4, 3 and 3 with four replications has been used. Hence there were a total of NB=12 blocks, where the blocks 1, 2 and 3 form one super-block or replication; further blocks {4, 5 and 6}, blocks {7, 8 and 9} and blocks {10, 11 and 12} form other replications.

The Normal Equations are as follows.

The observations per plot were as follows:

Progeny	Block											
	1	2	3	4	5	6	7	8	9	10	11	12
T1	6.83	.	.	6.56	.	.	7.01	.	.	.	6.01	.
T2	.	6.69	.	.	5.95	.	.	5.75	.	6.28	.	.
T3	.	.	6.19	.	.	6.56	.	.	6.42	.	.	6.02
T4	.	6.66	.	.	.	6.58	.	.	7.2	.	6.59	.
T5	.	.	6.71	7.1	.	.	6.4	.	.	6.33	.	.
T6	7.45	.	.	.	7.38	.	.	7.39	.	.	.	6.04
T7	.	.	6.57	.	6.22	.	.	.	7.79	.	6.39	.
T8	8.01	.	.	.	.	6.25	4.41	.	.	6.9	.	.
T9	.	6.29	.	6.11	.	.	.	6.91	.	.	.	5.11
T10	7.61	.	.	.	.	7.44	.	5.81	.	.	7.7	.

The first Normal Equation (1) is:

$40*f + \square_g 4*t_g + 4*d_1 + 3*d_2 + 3*d_3 + 3*d_4 + 3*d_5 + 4*d_6 + 3*d_7 + 4*d_8 + 3*d_9 + 3*d_{10} + 4*d_{11} + 3*d_{12} = 263.62$

the first equation of (2) is:

$4*f + 4*t_1 + d_1 + d_4 + d_7 + d_{11} = 26.4$, etc.

the first equation of (3) is:

$4*f + t_1 + t_6 + t_8 + t_{10} + 4*d_1 = 29.9$, etc.

A solution of these Normal Equations is

$f = 6.0686$, $t_1 = -0.3985$, $t_2 = -0.7802$, $t_3 = -0.6241$, $t_4 = -0.4676$, $t_5 = -0.1808$, $t_6 = 0.2356$, $t_7 = -0.3186$, $t_8 = -0.6645$, $t_9 = -0.6471$, $t_{10} = 0.0000$, $d_1 = 1.6133$, $d_2 = 1.1097$, $d_3 = 0.7959$, $d_4 = 0.9302$, $d_5 = 0.7358$, $d_6 = 1.0780$, $d_7 = 0.2860$, $d_8 = 0.6944$, $d_9 = 1.5382$, $d_1 = 0.9766$, $d_{11} = 0.9001$, $d_{12} = 0.0000$.

Hence the Least Squares Means for the *tenera* offsprings are:

$LSM(T_1) = 6.5583$, $LSM(T_2) = 6.1765$, $LSM(T_3) = 6.3328$, $LSM(T_4) = 6.4892$, $LSM(T_5) = 6.7760$, $LSM(T_6) = 7.1923$, $LSM(T_7) = 6.6382$, $LSM(T_8) = 6.2922$, $LSM(T_9) = 6.3096$, $LSM(T_{10}) = 6.9567$.

The residual sum of squares is SS(res) =

$$\sum_g \sum_h y_{gh}^2 - [f*263.62 + t_1*26.41 + \ldots + d_1*29.9 + \ldots] = 1757.5708 - 17448.4275 = 9.1433$$

with degrees of freedom, df (res) = 40-(10+12-1) = 19 and hence. $s^2 = 9.1433/19 = 0.4812$.

To estimate the General Combining Abilities of the *dura* and *pisifera* parents we make a two-way table for the *dura* D_i and *pisifera* P_j. In each cell of a realized cross the Least Squares Mean is inserted as many times as there are plots for that cross in the Incomplete Block Design; in this example this is four times. In the table we indicate the value only once.

Pisifera

	P1	P2	P3	P4	P5
Dura D1	6.558	.	.	.	6.957
D2	6.177	6.333	.	.	.
D3	.	6.489	6.776	.	.
D4	.	.	7.192	6.638	.
D5	.	.	.	6.292	6.31

The Least Squares estimates for the additive model of *dura* and *pisifera* effects are:

m=6.3588, a_1=0.5487, a_2=0.0685,

a_3=0.1266, a_4=0.4444, a_5=0.0000,

b_1=-0.3000, b_2= -0.0454, b_3=0.3398,

b_4=-0.1158, b_5= 0.0000 .

The estimate of the Specific Combining Ability of $T_1 = D_1 x P_1$ is $LSM(T_1)-(m+a_1+b_1)$

$= 6.5583 - [6.3588 + 0.5487 +(-0.3000)] = -0.0492$.

In the following table the estimates of the Specific Combining Abilities (SCA) are given; each SCA value must be repeated four times but in the table only a SCA value is given once.

| | | Pisifera | | | | |
		P1	P2	P3	P4	P5
Dura	D1	-0.0492	·	·		0.0492
	D2	0.0492	-0.0492	·	·	·
	D3	·	0.0492	-0.0492	·	·
	D4	·	·	0.0492	-0.0492	·
	D5	·	·	·	0.0492	-0.0492

The sum of squares for the Specific Combining Abilities (SCA) is the sum of all the squares of the SCA values (note that each entry in the table must be replicated four times), $SS(SCA)= 4*10*(0.0492)^2 = 0.0968$. The degrees of freedom for this SS(SCA) is $df\,(SCA) =10-(5+5-1)=1$.

The test-statistic to test the null-hypothesis "The SCA values are equal" is

$$F = \frac{SS(SCA)\,/\,df(SCA)}{SS(res)\,/\,df(res)}$$

.

$$F = \frac{0.0968/1}{9.1433/19} = 0.096$$

The 5% right sided significance point of the F-distribution with 1 and 19 degrees of freedom is $F(5\%)=4.38$, hence the null hypothesis is not rejected because $F= 0.096 < F(5\%) = 4.38$.

2.6 Statistical selection procedures

2.6.1 Introduction

Suppose that a breeder is comparing a number k $(k \geq 2)$ of potential oil palm progenies. A progeny is characterized by the expected yield m per plot of constant size. The goal of the breeder is to select one or more good progenies, or, formulated in a more accurate way, he wants

to select ultimately the best progeny, where the best progeny is defined as the progeny with the largest expected yield per plot. The statistical approach of searching for the best progeny is termed Statistical Selection. There are two basic approaches developed for Statistical Selection in the literature. One approach has been developed by Bechhofer (1954). The second approach has been thoroughly investigated by Gupta (1956, 1965). For a review see Van der Laan and Verdooren (1989, 1990). In sections 2.6.2 and 2.6.3 the theoretical background of the two approaches will be outlined for the interested reader. In section 2.6.5 these approaches are illustrated by a practical example with oil palm.

2.6.2 Indifference zone approach of selection

In this section we shall describe Bechhofer's Indifference Zone approach. Assume k (fixed and $k \geq 2$) varieties denoted by V_1, V_2, ..., V_k are given. The experimental design can either be a completely randomized design with n plots for every variety or a randomized complete block design with n blocks each of block size k and the plots in a block randomly assigned to the k varieties. From the observations X_{ij} of the k varieties, we calculate the k sample means

$$\overline{X}_1, \overline{X}_2, ..., \overline{X}_k$$

where

$$\overline{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij} \, (i = 1,2...,k)$$

and these sample means are based on an equal number *n* of independent and normally distributed random observations with expectation μ_i and common variance σ^2. For simplicity we first assume that the common variance σ^2 is known. The parameters μ_i are ranked and indicated by $\mu_{[1]} \mu_{[2]} \leq ... \mu_{[k]}$. In the same way the ranked sample means are denoted by

$$\overline{x}_{[1]} \leq \overline{x}_{[2]} \, \square \, ... \leq \overline{x}_{[k]}.$$

The variety associated with $\mu_{[i]}$ will be denoted by $V_{(i)}$. Then we define the variety $V_{(k)}$ (with associated response $x_{(k)}$), corresponding to $\mu_{[k]}$, as the best variety. If there is more than one contender because there are ties, it is assumed that one of these is appropriately tagged. The goal is to select the variety associated with $\mu_{[k]}$, thus the best variety. We define

$$\delta_{k,k-1} = \mu_{[k]} - \mu_{[k-1]}$$

Then a selection procedure R based on the Indifference Zone approach and using the statistics

$$\overline{X}_1, \overline{X}_2, ..., \overline{X}_k$$, is as follows.

Select V_i if and only if

$$\overline{X}_i = \overline{X}_{[k]}.$$

In this context a correct selection (CS) means that the best variety is selected. The following probability condition for CS given the selection procedure R must be fulfilled:

$P(CS \mid R) \geq P^*$ if $\delta_{k,k}^{-1} \geq \delta^*$

with $k^{-1} < P^* < 1$. Thus the probability of selecting the best variety is at least P^*, provided the best variety is at least δ^* away from the second best variety. This minimal probability P^* can only be guaranteed if the required common sample size n is large enough. The minimum of the $P(CS \mid R)$ is attained for the so-called Least Favourable Configuration (LFC) given by $\mu_{[1]} = \mu_{[2]} = ... = \mu_{[k-1]} = \mu_{[k]} - \delta^*$. One can prove that the probability of correct selection for the LFC is equal to

$$PLFC(CS) = \int_{-\infty}^{\infty} \phi^{k-1}(x + \tau) d\phi(x)$$

where $\phi,(\bullet)$ is the standard Normal cumulative distribution function and

$$\tau = \frac{\delta^* \sqrt{n}}{\sigma}$$

Tables for τ and thus for *n* have been constructed. For a minimal sample size n to find the best variety for a given δ^* and P^* we have to choose

$$n \geq \left(\frac{\tau\sigma}{\delta^*}\right)^2$$

rounded to the nearest (larger) integer, where the quantity τ can be found for various values of P^* and k, in for instance Gibbons et al. (1977) tables A1 and A2, Bechhofer (1954), Gupta (1963), Gupta et al. (1973) and Butler & Butler (1987). The conclusion deduced from the statistical selection procedure R can also be formulated as follows. With the chosen minimal n it can be guaranteed with minimal probability P^* that the selected variety is less than δ^* away from the best variety.

The Indifference Zone approach is important in designing an experiment. It provides a value for the common sample size n needed to meet certain probability requirements. In practice k, the number of varieties, should not be large, otherwise the total number of plots (kn) will be too large.

Using the Indifference Zone approach in the situation of an unknown common variance σ^2 a two-stage procedure is necessary, which is less attractive in practice. For the Indifference Zone

procedure the first stage is necessary in order to get an estimation of σ^2. In the second stage the estimate s^2 is used to obtain the required size of the second sample.

2.6.3 Subset selection approach of selection

The Subset Selection approach of Gupta aims to select a subset of the k varieties considered in section 2.6.2, in order to include the best variety with a certain confidence. The size of the subset is not fixed beforehand (this means the number of selected varieties) and depends among other things on the sample means and the variance σ^2 and on the common sample size n. Obviously, we wish a selection rule which makes the expected number of varieties in the subset as small as possible. It is not necessary to determine the common sample size n at the start of the experiment. The experimental design can be a completely randomized design with n plots for each variety or a randomized complete block design with n blocks each of block size k and the plots in a block randomly assigned to the k varieties.

The selection rule R will be based again on the sample means

$$\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k,$$

where

$$\overline{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij} (i = 1, 2, \ldots, k)$$

is based on an equal number n of independent and normally distributed random observations with expectation μ_i and common known variance σ^2. The rule R can be described as follows. Select V_i in the subset if and only if

$$\overline{X}_i \geq \frac{\max}{1 \leq j \leq k} \overline{X}_j - \tau \frac{\sigma}{\sqrt{n}}$$

where $\tau > 0$ must be determined such that the probability requirement of a Correct Selection (CS) with this selection rule R

$$P(CS) \mid R) \geq P*$$

is met for all possible values of the parameters i. In this context a correct selection CS means that the best variety belongs to the selected subset. It can be proved that the least favorable configuration (LFC) is the limit situation, where $\mu_{[1]}, \ldots, \mu_{[k1]}$ are all equal to $\mu_{[k]}$. It can be proved that

$$PLCF(CS) = \int_{x=-\infty}^{x=\infty} \phi^{k-1}(X+\tau)d\phi(x)$$

Values of τ can be found in the tables mentioned in 2.6.2. The size of the subset reflects the confidence in choosing the best variety. A large subset would mean that either the varieties are close together or the sample sizes are small, or both. In the normal means μ_i situation with the common variance σ^2 unknown, a single-stage procedure can be used for the Subset Selection:

Select variety V_i if and only if

$$\overline{X}_i \geq \overline{X}_{[k]} - S \frac{\tau}{\sqrt{n}}$$

where s^2 is the unbiased estimator of σ^2 based on v degrees of freedom and $\tau = \tau(k,v,P^*)$. Values of the constant τ or h with $\tau = h\sqrt{2}$ can be found in the references mentioned in 2.6.2, with the exception of Butler & Butler (1987) which gives only values of $\tau(k,\infty, P^*)$. The constant h satisfies

$$\int_{w=0}^{w=\infty} \int_{z=-\infty}^{z=\infty} \phi^{k-1}(z+hw\sqrt{2}d\phi(z)d(w) = P^*$$

where $G(w)$ is the distribution function of $w = s/\sigma$ and P^* is the desired confidence. Values of h are given in e.g. table A4 of Gibbons et al. (1977) and in Bechhofer and Dunnett (1988).

2.6.4 Comparison of the two approaches

The Subset Selection approach has certain advantages in practice. We mention the possibility to use the Subset Selection method as screening procedure. Even when the ultimate goal of the breeder is to choose the best, the Subset Selection approach can be applied to eliminate inferior varieties. This is in practice an interesting feature, especially when the number of potential varieties is large, as is usually the case in oil palm parental testing programs.

The Indifference Zone approach aims to indicate the best variety, whereas the Subset Selection approach selects in general more than one variety, so providing less precise information.

However, one has to pay for more precise information in the form of structuring the problem in more detail. Using the Indifference Zone approach we must define δ k,k-1 which is a measure for the distance between the best variety and the second best variety and we must give δ *, which is in practice sometimes embarrassing. The Indifference Zone approach is very useful at the experimental design stage in order to determine the required common sample size n. The designing aspect is an integral accessory of the Indifference Zone methodology. Determination of the required sample size is the central point, rather than the analysis of obtained samples.

The Subset Selection approach can be used without planning the sample size in advance. This enables the breeder to analyze the data when the experiment has already been done and the sample size is not adequate for the Indifference Zone approach. In this sense one can say that the Subset Selection approach is more flexible. Regardless the value of the common sample size n,

the Subset Selection approach can be applied. However, the size of the subset increases as n decreases. That is the toll one has to pay for small sample sizes. This Subset Selection approach is recommended as a method for the oil palm progeny tests.

2.6.5 Some applications

In Papua New Guinea, oil palm cultivation started on a commercial scale in 1968. In 1976, about 12,000 ha were planted. To guide the oil palm cultivation the Dami Oil Palm Research Station was founded at Kimbe, West New Britain, Papua New Guinea. At this station a *dura* x *pisifera* progeny trial was established in 1968. In this experiment nine ex-AVROS *pisifera* with four selected Deli *dura* palms were crossed to get 15 families. These fifteen families were arranged in five randomized complete blocks with sixteen (4 x 4) palms per plot with a 9 m triangular spacing.

For this example we have taken only ten families which remain in four complete blocks, the other families were discarded in several blocks due to diseases. The average fresh fruit bunch yield y (in kg/palm) over the years 1972-1977 of the four inner palms per plot was analyzed. Further, samples of leaf 17 were taken from all inner palms in 1973, bulked per plot, and analyzed at Banting Oil Palm Research Station (O.P.R.S.) in Malaysia.

The Magnesium content x (in %) in leaf 17 was determined. Breure (1987), arrived at the conclusion that this % Mg has a good correlation (r= 0.70) with the yield of oil for the first five years of production (1972-1976); hence the % Mg can be used to indicate good families for oil yield. Because the % Mg determination has been done with the same procedure in Banting O.P.R.S. for a long period, the standard deviation % of the % Mg determination can be stated as known and to be 0.0186. The average % Mg for the 10 families over the four blocks was as follows.

family V_i	$\bar{x}_i$ % Mg	rank number
I = 1	0.212	[5]
2	0.222	[7]
3	0.242	[8]
4	0.204	[3]
5	0.210	[4]
6	0.186	[2]
7	0.218	[6]
8	0.244	[9]
9	0.162	[1]
10	0.248	[10]

Following Bechhofer's procedure the Least Favourable Configuration (LFC) is given by

$$\mu_{[1]} = \mu_{[2]} = \ldots = \mu_{[9]} = \mu_{[10]} - \delta^*.$$

When the minimum probability of correct selection P* and the common sample size n are given, δ * can be determined. From table A1 of Gibbons et al. (1977), we find the following values for τ with k = 10 populations for various values of P*:

P*	τ
0.75	2.2637
0.90	2.9829
0.95	3.4182
0.99	4.2456

From the formula

$$\delta* = \frac{\sigma}{\sqrt{n}} \tau$$

we find in this case the values of

$$\delta* = \frac{0.0186}{\sqrt{4}} \tau$$

for the values of P* as:

P*: 0.75 0.90 0.95 0.99

δ *: 0.021 0.028 0.032 0.039

Otherwise we can also determine the number n of complete blocks to determine a δ * = 0.02 or 0.01 for the different values of P* from

$$n \geq \left[\frac{\sigma \tau}{\delta*} \right]^2$$

rounded to the nearest (larger) integer. In our case we must calculate

$$n \geq \left[\frac{0.0186}{\delta*} \tau \right]^2$$

and the results for n are:

	P:	0.75	0.90	0.95	0.99
$\delta* = 0.02$	n:	5	8	11	16
$\delta* = 0.01$	n:	18	31	41	63

When we apply Gupta's Subset Selection procedure to find the subset which contains the best family with a minimum probability of correct selection P*, we must take those families V_i for which

$$\overline{X_i} \geq \overline{X}_{[10]} - \frac{\delta}{\sqrt{n}}\tau$$

In our case we find

$$\overline{X_i} \geq 0.248 - \frac{0.0186}{\sqrt{4}}\tau$$

The results are:

P*	Selection Level	Subset contains families V i
0.75	0.227	(10, 8, 3)
0.90	0.220	(10, 8, 3, 2)
0.95	0.216	(10, 8, 3, 2)
0.99	0.209	(10, 8, 3, 2)

Conclusion for % Mg

If we want to indicate the families which give the best Magnesium content in leaf 17 across the first five years of production (and hence a good oil yield), we can take the families 10, 8, 3 and 2 with a minimal probability of correct selection P* of 0.90. When we increase P* to 0.99 the subset must be extended with the families 7, 1 and 5.

Now we want to use Gupta's Subset Selection procedure to find the minimum subset of families which contains the best family with a minimum probability of correct selection P* for the average fresh fruit bunch yield y. The Analysis of Variance procedure gives as an estimate of the variance σ^2 a Mean Square Error of 6308.488 based on $v = 27$ degrees of freedom. For the yield y_{ij} of family i in block j we use the following model:

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

(i = 1, 2, ..., 10 and j = 1, 2, 3, 4).

The least squares mean $\mu + \alpha_i + \beta$ for the family V_i is estimated by the family V_i mean y_i. The results were as follows:

family V_i	y_i kg/palm	rank number
1	999.97	[10]
2	890.5	[6]
3	927.24	[9]
4	888.6	[5]
5	697.9	[1]
6	745.83	[2]
7	863.19	[4]
8	916.06	[8]
9	915.62	[7]
10	854.94	[3]

The Subset Selection procedure of Gupta selects family V_i if

$$\overline{Y}_i \geq \overline{Y}_{[10]} - \sqrt{\frac{6308.488}{4}}\, \tau$$

From table A4 of Gibbons et al. (1977), which gives h-values, we derive the following values of $\tau = h\sqrt{2}$ for $k = 10$, $v = 27$, and for $P^* = 0.95$ and $P^* = 0.99$ (we must interpolate with $1/v$ for $v = 25$ and $v = 30$ to find the value of h):

$P^* = 0.95$, $\tau = 2.55\sqrt{2} = 3.606$

$P^* = 0.99$, $\tau = 3.27\sqrt{2} = 4.624$

Hence

P*	selection level	subset contains families V_i.
0.95	856.76	(1, 3, 8, 9, 2, 4, 7)
0.99	816.34	(1, 3, 8, 9, 2, 4, 7, 10)

Conclusion for fresh fruit bunch yield

We then conclude that the most promising families for the fresh fruit bunch yield are the families 1, 3, 8, 9, 2, 4 and 7. The probability that this selection procedure selects in the subset the best family of the 10 tested families is at least $P^* = 0.95$. When we increase P^* to 0.99 then family 10

must also be included in this subset. The subsets can be reduced by using more replications in forthcoming experiments.

2.6.6 Selection trials with incomplete block designs

In case progenies are compared in an incomplete block design, one can use for the Subset Selection Procedure of Gupta the following approximation. Instead of the selection rule in section 2.6.3: "Select variety V_i if and only if

$$\overline{X}_i \geq \overline{X}_{[k]} - \frac{S*\tau}{\sqrt{n}}$$

"we use now the selection rule : "Select variety V_i if and only if

$LSM(V_i) \geq LSM(V[k]) - S*\tau$"

where $S = 1/1\sqrt{2}$ times the average standard error of the differences between pairs of progenies, LSM(Vi) is the Least Squares Mean of variety Vi and LSM(V[k]) is the Least Squares Mean of the variety with the largest LSM (see section 2.2 or section 2.5 for the definition of Least Squares Mean (LSM)). A more accurate (but more elaborate) method can be found in Dourleijn (1993, 1995).

2.7 Characters for selection

2.7.1 Objective

Remember that the aim is to select parent palms for the production of *tenera* offspring with high yield of oil and kernels per hectare. Basically, this means developing *tenera* which optimize the transfer of available resources of the physiological environment (solar radiation, temperature, humidity, soil water and nutrients, etc.) into oil and kernels. Among these resources, interest in oil palm breeding focuses mainly on optimizing the supply of assimilate, produced in the process of photosynthesis, and its transfer into economic products. An understanding of the carbon budget, as reported by Breure (1987) for the oil palm, is therefore crucial to develop the desired *tenera* ideotype, i.e. a biological model which is expected to perform in a predictable manner (Donald, 1968).

2.7.2 Partition of carbohydrates

Briefly, carbohydrates produced in the process of photosynthesis are, as a first priority, used for the maintenance of existing biomass (maintenance respiration). Priority is then given to the production of vegetative dry matter. Once requirements for maintenance respiration and vegetative growth are satisfied, carbohydrates are allocated to bunch production. Oil-and-kernel production thus strongly benefits from increasing photosynthetic production above a certain threshold level and reducing carbohydrate requirements for maintenance and vegetative growth. Maintenance requirements can be calculated from the protein and mineral content of plant tissue

and their metabolic activity (van Kraalingen et al., 1989). But this method is not suitable to quantify, and thus to select for, maintenance respiration in breeding work. Therefore, only selection for reduced vegetative requirements is considered in this present report.

2.7.3 Harvest Index

The greatest scope for increasing yield appears to be by selecting for high Harvest Index (HI), i.e. the proportion of dry matter used for the production of oil-and-kernels (Hardon et al, 1972; Breure & Corley, 1983). Little progress can be achieved by direct selection for HI because of its low heritability (Breure & Corley, 1983; Breure & Bos, 1992).

HI can be increased, according to the carbon budget, in two ways: (i) by reducing carbohydrate requirements for Vegetative Dry Matter production (VDM) and (ii) by increasing photosynthetic production above a certain threshold level. Both options are achieved if the reduction of VDM is restricted as much as possible to non-photosynthetic tissue, that is, by increasing Leaf Area Ratio (LAR), defined as the ratio of the new total leaf area produced to new VDM.

Breeding for speed of photosynthesis without an accompanying increase in VDM can also be achieved through increasing magnesium content in the leaves. It is, however, not obvious that magnesium per se increases photosynthetic production.

Magnesium content in oil palm leaves could be accompanied with other decisive photosynthetic components, such as those factors involved during the biosynthetic process of photosynthesis. This mechanism comprises different biochemical reactions during photosynthesis, which are regulated by enzymes during CO_2 fixation. Shibles (1993) found in annual crops that the photochemical process in C3 plants, like oil palm, is not the bottleneck in terms of photosynthetic efficiency. It is therefore possible that besides rising magnesium content, a parallel rise in relevant enzymes also occur into the mesophylic cells; it may be actually this change which is measured as leaf-Mg level. Nevertheless, Peaslee & Moss (1966) demonstrated that magnesium concentration in the leaves is closely related to photosynthetic production, and Breure (1986) showed that parent selection for magnesium content in the leaves increases HI.

More recently Breure & Bos (1992) confirmed in a multiple regression analysis that HI benefits from selection for high GCA values of LAR and leaf-Mg. They also found that reducing Leaf Production (LPR), the main component of VDM, positively affects HI. This is not surprising as reducing LPR will not affect light interception (and thus photosynthetic production), because a rather constant number of crown leaves is maintained through regular leaf pruning for harvesting the bunches. It does, however, diminish the proportion of carbohydrates allocated to vegetative growth to the benefit of bunch yield. In the same vein, reducing vertical trunk growth is expected to increase HI as the trunk is not contributing to photosynthesis; but this could not be substantiated in the study of Breure & Bos (1992). They argued that the gain in diminishing carbohydrate requirements is probably outweighed by the associated reduction in competitive ability for light and thus photosynthetic production when progenies are planted in a mixture. Thus in a more uniform population, the advantage of slow height increment would probably be greater. In progeny tests the benefit of slow height increment may better show up when progenies are grouped in the field according to the vigour established in the nursery, as is

explained in section 2.10. HI can thus be increased by selection for low GCA values for VDM, with LPR as the main component, low values of height increment, and high values of LAR and Leaf-Mg.

2.7.4 Crown expansion

The canopy may take up to 6 years to close (Squire & Corley, 1987). At that stage photosynthetic production per hectare, and thus bunch yield, has reached its maximum, and Breure (1985) showed that selection for quick canopy closure indeed benefits early yield. The effect of crown expansion on the production and partitioning of carbohydrates is a crucial aspect of oil palm breeding.

Before canopy closure the rate of increase in the area of individual leaves, or when taking into account the number of green leaves in the crown, the Leaf Area Index (LAI), i.e. the total leaf area per unit ground area, is directly related to photosynthetic production per hectare, and thus carbohydrates allocated to bunch yield. Crown expansion is influenced by husbandry practices and, as shown by Breure (1985), also by genotype.

Once the canopy is practically closed, photosynthetic production per unit area has reached its maximum, and both bunch production and vegetative growth by and large stabilize (Breure, 1988). The yield pattern in the period following canopy closure seems to depend, among other things, on the rate at which the crown continues to expand (Breure, 1988). Prolonged expansion does not further contribute to photosynthetic production per unit area, as the canopy is closed already, but does increase carbohydrate requirements for vegetative growth and maintenance respiration at the expense of those allocated to bunch production. The result is a decrease in HI with age.

The search is thus for ideotypes which stabilize at LAI-value which maximizes yield of oil and kernels per hectare. In terms of the carbon budget this means when the gain in light interception does not become outweighed by accompanying losses in carbohydrates for vegetative requirements and maintenance respiration. This optimal LAI should be reached as soon as possible after field planting in order to maximize the proportion of incident solar radiation intercepted by the oil palm canopy. This, albeit over-simplified model, implies that the trend in crown expansion affects yield during the entire economic life of a planting. Expansion of mean leaf area per palm with time fits a logistic growth curve (Breure, 1985). Basically a logistic growth function is of the following form:

$f(t) = A/(1 + B*e^{Ct})$, where A, B and C are positive constants and t is the time of growth. C describes the rate of growth and A is the asymptotic maximum of $f(t)$ which is approached when t runs to infinity. At the start (t=0), $f(t) = A/(1 + B)$. The inflexion point of the logistic growth function is at $t = -(1/C)*ln(1/B) = (1/C)*lnB$ when the function has reached a value of $A/2$.

The parameters of the logistic growth function can be estimated by the Least Squares Method. However this procedure is now much more difficult than in the case of a linear model, because we have here a non-linear regression problem. The Normal Equations can now only be solved iteratively. Statistical computer packages such as SAS, SPSS, SYSTAT, BMDP and GENSTAT

have a module for the Least Squares Method for non-linear regression problems. Also the computer package CADEMO (Computer Aided Design of Experiments and Modeling) has a module "Growth Curves" where the parameters can be estimated; furthermore in this module for a new experiment an optimal design is advised to estimate the logistic growth function (or other growth functions). Breure (1985) describes the mean leaf area (in m^2) as a function of the time (in months) after field planting as a logistic growth function but he used in his publication the following function:

$f(t) = L_m / [(1 + (L_m - L_i)/L_i)) e^{-kt}]$, with

$A = L_m$, $B = (L_m - L_i)/Li$ and $C = k$ or

$L_m = A$, $Li = A/(1+B)$ and $k = C$,

where k = the relative rate of growth of the mean leaf area,

L_m = asymptotic maximum leaf area,

L_i = leaf area at field planting.

For selection purposes this is conveniently expressed as the time to reach 95% of the maximum leaf area ($t_{0.95}$), hence $f(t_{0.95}) = 0.95*L_m$, as follows:

$$t_{0.95} = -\left(\frac{1}{k}\right) * \ln \frac{\dfrac{L_m}{0.95 L_m} - 1}{\dfrac{L_m}{L_i} - 1}$$

For an example of a calculation see section 4.5.6. Thus, in order to minimize the period of sub-optimal LAI, selection should aim at low $t_{0.95}$ and thus high k-values. Final leaf area (L_m) is one of the components of LAI at maturity and is thus an important characteristic in breeding for optimal LAI. Practically, leaf area can first be measured when new leaves have emerged, about 6 months after field planting. Yield recording stops after 90 months in the field, so this is the final age at which measurements can be made.

The logistic growth function is determined by three parameters (A, B and C); hence, three measurements on different leaves is the absolute minimum to determine the logistic growth function. More measurements to estimate the parameters of the growth function will clearly improve the precision, but resources are limited. For practical reasons we propose to measure four times for step 2 of parent palm selection which is meant for a first screening. For step 3 of parent palm selection five measurements is recommended.

From the results of 52 progenies planted at Dami Oil Palm Research Station, New Britain, Papua New Guinea, 11 annual leaf area measurements in m^2 averaged over 6 replications of 4 palm plots (24 palms), were done at month 12, 24, 36, 48, 60, 72, 84, 96, 108, 120 and 132 months from planting (See Table 2).

Using the Least Squares Method the parameters of the fitted logistic growth curve per progeny are given in Table 3. The average of these parameters, together with the standard error of the average and the 95%-confidence interval for the parameter, is given in Table 4.

Using the logistic growth curve with these average parameters we can derive a design for the optimal time of measurements from 6 to maximal 90 months after planting for n = 4 (step 2 of parent palm selection) and for n = 5 (step 3 of parent palm selection or for a sample of palms in step 2 to compare groups of families derived from one source of parents).

The design for optimal time of measurements is such that the determinant of the asymptotic covariance matrix of the parameter estimators is a minimum (D-optimality). Unfortunately this optimum design depends on the parameters of the logistic growth function and it gives therefore only a local optimum for the given parameters (not an optimum for all possible parameters). The theory of the D-optimality design for a logistic growth function is described by Rasch (1992) and the Growth Curve analysis module of the computer package CADEMO (Computer Aided Design of Experiments and Modeling) has been used to find the determinant of the asymptotic covariance matrix of the parameter estimators. For the calculation of the determinant of the asymptotic covariance matrix we used the average values of the parameters A, B and C as given in Table 4. The results are presented in Table 5. For the proposed time measurements we will choose the design which is feasible, in terms of the restricted period of recording, and has a determinant equal to or near the least determinant of the asymptotic covariance matrix for the specified number of time measurements.

Table 2. Mean leaf area in m^2 of 6 replications of 4 palm plots (24 palms), measured on 12 months interval

Pro-geny	Months from planting										
	12	24	36	48	60	72	84	96	108	120	132
1	1.83	3.54	5.35	6.04	7.52	7.86	8.3	9.15	11.37	11.21	10.43
2	1.68	2.5	5.3	5.48	7.25	7.69	8.78	9.37	11.56	12.21	11.38
3	1.76	3.64	5.37	5.94	7.1	8.04	8.69	9.6	11.88	11.76	11.16
4	1.74	3.81	5.48	6.43	8.29	7.71	8.41	9.36	11.48	11.59	10.68
5	1.79	3.9	5.54	5.72	7.7	8.44	8.07	9.22	11.98	11.67	11.16
6	1.8	3.64	5.5	5.61	7.46	8.06	8.89	9.52	12.05	11.63	11.16
7	1.8	3.99	5.54	5.9	7.8	8.07	9.12	9.5	11.34	11.63	10.88
8	1.7	3.8	5.54	6.28	8.13	7.69	8.57	9.32	12.01	11.8	10.74
9	1.76	3.8	5.36	5.86	7.64	7.98	8.93	9.21	12.13	11.96	11.23
10	1.75	3.78	5.33	6.06	7.96	8.12	8.87	9.4	12.52	11.72	11.03
11	1.88	3.85	5.63	6.35	7.86	8.33	8.9	9.17	11.71	11.74	11.32
12	1.88	4.01	5.43	6.11	7.91	8.12	9.13	9.48	11.56	11.92	11.35
13	1.76	4.01	5.82	6.22	7.23	8.32	8.89	9.71	11.93	11.91	11.16
14	1.93	4.01	5.37	6.18	7.91	8.24	8.92	9.87	11.63	11.9	11.34
15	1.84	3.74	5.5	5.85	8.01	8.26	9.1	9.64	12.58	11.86	11.23
16	1.85	3.93	5.51	6.2	7.78	8.54	9.16	9.86	11.92	12.03	11.08

17	1.7	3.74	5.3	6.07	7.91	8.35	8.83	9.84	12.59	11.94	11.68
18	1.86	4.29	5.36	6.47	8.22	8.64	8.86	9.52	11.79	12.17	10.81
19	1.91	4.15	5.54	6.19	8.46	7.99	9.06	9.54	11.77	12.15	11.35
20	1.93	4.14	5.57	6.4	8.29	8.38	9.34	9.63	11.64	12.24	11.03
21	1.91	4.09	5.61	5.99	8.05	8.15	9.06	10.04	12.37	12.35	11.49
22	1.85	3.94	5.56	6.07	8.08	8.86	9.12	10.17	12.75	11.7	11.4
23	1.74	4.18	5.59	6.43	8.25	8.65	9.78	10.23	11.67	12.08	11.24
24	1.82	4.03	5.28	5.9	8.13	8.18	8.99	10.27	13.24	12.25	11.84
25	1.73	4.06	5.71	6.24	8.12	8.37	8.7	9.71	13.34	12.32	11.71
26	1.7	3.93	5.25	6.03	8.41	8.74	8.98	9.74	13.1	12.26	11.9
27	1.83	3.68	5.38	6.06	8.07	8.8	9.24	10.41	13.36	12.11	11.12
28	1.72	4.04	5.75	6.05	8.29	8.33	9.12	9.83	12.87	12.5	11.66
29	2.18	4.45	5.85	5.95	8.37	8.49	9.05	10.27	12.44	12.88	11.44
30	1.98	4.24	5.85	6.65	8.2	8.37	9.77	10.19	11.72	12.86	11.73
31	1.9	4.22	5.62	6.25	7.92	8.68	9.63	10.47	12.66	12.5	11.79
32	1.95	4.14	5.8	6.8	8.5	8.35	9.01	10.08	12.53	12.59	12.14
33	1.85	3.87	5.55	5.98	7.88	8.99	9.5	10.44	13.29	12.63	11.99
34	1.71	4.02	5.54	6.33	8.49	8.92	9.51	10.69	12.41	13.17	11.48
35	1.93	4.32	5.78	6.28	8.29	8.9	9.59	10.33	13.05	12.49	11.57
36	1.95	4.26	6.01	6.65	8.38	9.03	9.27	10.38	12.7	12.59	11.58
37	1.91	4.07	5.98	6.81	8.66	8.84	9.68	9.99	13.11	12.72	11.65
38	1.86	4.04	5.79	6.83	8.15	9.29	9.86	10.45	12.81	12.39	12.34
39	1.81	4.21	5.94	7.2	8.83	8.9	9.3	10.24	13.1	12.47	11.87
40	1.89	4.45	5.48	6.65	8.75	8.94	9.64	10.28	13.19	13.16	12.02
41	1.8	4.21	5.72	6.8	9.27	9.15	9.84	10.06	13.35	12.64	12.13
42	1.97	4.23	5.88	6.47	8.48	8.87	9.86	10.67	13.11	13.03	12.54
43	1.97	4.43	5.62	6.68	8.74	9.16	10.01	10.89	12.68	12.92	12.31
44	2.19	4.59	5.7	6.69	8.69	8.76	10.08	10.75	13.58	12.8	11.85
45	2.2	4.84	5.55	6.94	8.74	9	9.22	10.89	13.44	13.03	11.98
46	2	4.28	5.9	6.6	8.85	9.15	10.22	10.99	13.41	12.69	11.79
47	2.05	4.59	6.08	6.82	8.99	9.13	10.22	11.18	12.71	13.13	12.15
48	2.13	4.41	5.73	6.77	8.97	9.36	10.42	10.89	13.27	13.22	12.6
49	1.94	4.54	6.04	6.61	8.94	9.36	10.17	11.03	13.76	13.57	12.45
50	2.07	4.78	5.95	7.01	8.95	9.67	10.38	10.82	13.32	13.71	12.17
51	1.94	4.4	6.01	6.93	9.06	9.49	10.8	11.2	13.8	14.18	12.98
52	1.83	4.63	6.22	7.29	9.83	10.13	10.7	11.22	14.11	13.87	13.03

Results show that it is crucial to include one measurement before 18 months in the field. For n=4, the set of measurements at 6, 12, 54, and 90 months after planting gives the best design.

The design is improved if a second measurement is included in the first 18 months after planting; we found the best design then at 6, 12, 42, 66 and 90 months after planting.

Note, that for n=4 the design was only slightly less precise when measurements were done at 6, 42, 66 and 90 months, that is, those which are all included in the best design for n=5 measurements. Also that, the last three times of measurements correspond approximately to the end of the 1st, 3rd and 5th year of production which is a convenient timing for calculating growth parameters (see section 4.3).

This set of four measurements (6, 42, 66, and 90 months after planting) is therefore adopted for a step 2 progeny test; an additional measurement at 12 months is recommended for all palms of a step 3 test or for a sample of palms of a step 2 test when the sole objective is to evaluate leaf expansion of different sources of planting material.

2.7.5 Incidence of crown disease

Another way to increase yield through improving crown expansion is to screen genotypes for crown disease, a disorder which appears as bending of the opening spear leaves during the early years after field planting (Breure & Soebagjo, 1991). Crown disease is the most frequently occurring disorder in oil palm. Breure & Soebagjo (1991) showed that losses of oil yield of susceptible material can amount to 4.5% during the first six years of production.

Table 3. Estimates of the parameters of the fitted logistic growth function of leaf area (in m²) with time t after planting (in months) of the progenies from Table 2 in Appendix 5

Progeny	A	B	C
1	11.42287	5.300911	0.036079
2	12.72226	7.103731	0.034777
3	12.52239	5.746017	0.033383
4	11.54769	5.10298	0.036718
5	12.42299	5.191469	0.032389
6	12.39136	5.793531	0.03438
7	11.80467	5.294897	0.036213
8	11.94054	5.262365	0.035377
9	12.60181	5.620537	0.033292
10	12.20454	5.762228	0.036165
11	12.14406	5.070951	0.03451
12	12.34377	5.22397	0.034056
13	12.45161	5.198982	0.03348
14	12.35624	5.288212	0.034475
15	12.42251	5.870623	0.036102
16	12.16422	5.614522	0.036965

17	12.79858	6.034535	0.034962
18	11.8659	5.167217	0.037459
19	12.42581	5.023228	0.033866
20	12.01259	5.196565	0.037141
21	12.97529	5.426776	0.033164
22	12.3429	5.948479	0.038233
23	11.98	5.819282	0.040487
24	13.39402	6.077927	0.03369
25	13.29485	5.510197	0.032629
26	13.00823	6.139615	0.035692
27	12.5234	6.725684	0.039735
28	13.08518	5.633522	0.033998
29	13.23258	4.963921	0.032006
30	12.88902	5.08963	0.034194
31	13.0898	5.728571	0.035203
32	13.37472	5.038803	0.03203
33	13.38446	6.444905	0.035984
34	12.81209	6.30977	0.038793
35	12.87719	5.595573	0.036527
36	12.72905	5.26557	0.036533
37	12.76994	5.515285	0.037613
38	13.00721	5.868044	0.037653
39	12.68981	5.375179	0.038031
40	13.38379	5.637055	0.035478
41	12.83206	5.930891	0.039808
42	13.81779	5.667616	0.034015
43	13.22683	5.735545	0.037238
44	13.32115	5.404799	0.035876
45	13.5809	5.027142	0.033522
46	12.95249	6.035112	0.040041
47	13.17762	5.403957	0.037711
48	13.60597	5.800105	0.037148
49	13.8824	5.794405	0.03622
50	13.49246	5.464826	0.037497
51	14.34248	6.131313	0.036614
52	13.83745	6.140957	0.040785

Breure & Soebagjo (1991) showed that the screening of progenies for crown disease is the most effective control; parental GCA values for the incidence of crown disease is therefore an important aspect in selection. They recorded crown disease both by scoring the severity on newly

emerged leaves and the percentage of affected palms. The former is quite labor intensive, and for parent selection, recording can be restricted to the percentage of affected palms (rate of incidence).

2.7.6 Tolerance of light competition

Yield recording is usually restricted to the first five years of production; harvesting starts, in favourable environments, about 30 months after field planting. Yield components are fixed as early as two years before harvest (Breure & Menendez, 1990; Breure & Corley, 1992), so that yield for virtually the entire recording period is determined under conditions of low competition for light. The ranking for yield of the progenies may change, however, when components are determined under a closed canopy, that is, when inter palm competition for light has reached its maximum. This drawback in selection efficiency can be partly circumvented by selection for characters associated with tolerance to light competition.

Table 4. Average of the 52 estimates of the parameters of the logistic growth function for leaf area (in m²) and time t after planting (in months) from Table 3 (see Appendix 6); the standard error of the averages and the 95% confidence limits of the parameters

Parameter		Standard Error	95 % confidence limits	
			Lower	Upper
A	12.79768	0.087293	12.6224354	12.9729304
B	5.625345	0.062917	5.4990339	5.7516557
C	0.035883	0.000307	0.0352667	0.0371166

In this way it is more likely that high-yielding progenies, identified during the conventional period of yield recording (low light competition), continue to perform well at mature light competition. Remember that, according to the carbon budget, bunch production is more sensitive to light competition than vegetative growth. There is, however, evidence of genotypic differences in adopting vegetative requirements to available light (Corley & Donough, 1992). These genotypes are characterized by a high plasticity of vegetative growth, and are thus expected to maintain a high HI under a mature canopy.

One option is direct selection for trends in HI, but this is not practical. One of the components of HI is bunch yield which tends to show pronounced annual fluctuations. Meaningful HI values require therefore two or three year yield records which, in the limited period of recording, does not yield a sufficient number of values to estimate trends with age. Vegetative dry matter production is the other characteristic determining HI; this component, on the other hand, is less prone to annual fluctuations than bunch yield.

Among the components of vegetative growth, leaf production is particularly sensitive to light competition (Breure, 1982; Corley & Donough, 1992). This component also responds quickly to changes in the amount of light, as shown by Breure (1994). Genotypes which maintain a high HI

at maturity are therefore expected to show a strong decrease in leaf production with the increase of LAI with age. Corley (personal communication) indeed found highly significant differences between clones in the slope of the regression of leaf production on LAI, indicating that selection for trend of rate of leaf production with age may be feasible.

Table 5. Designs for leaf area measurements (time after planting) based on the logistic growth function and the determinant of the asymptotic covariance matrix of the parameter estimators.

A smaller determinant means a better design. Parameters used are A=12.798, B=5.625 and C=0.03588 (Table 4)

Number of Measurements	Time in months after planting							Determinant/σ^6
4	6	12	54	90				$0.1411*10^{-02}$
4	6	18	54	90				$0.1458*10^{-02}$
4	6	18	66	90				$0.1752*10^{-02}$
4	6	42	66	90				$0.1494*10^{-02}$
4	6	54	84	90				$0.1602*10^{-02}$
4	12	36	60	84				$0.2592*10^{-02}$
4	12	42	66	90				$0.1607*10^{-02}$
4	18	42	66	90				$0.1859*10^{-02}$
4	30	54	78	90				$0.3328*10^{-02}$
5	6	12	18	54	90			$0.09603*10^{-02}$
5	6	12	42	66	90			$0.08180*10^{-02}$
5	6	12	42	78	90			$0.10109*10^{-02}$
5	6	12	54	78	90			$0.08517*10^{-02}$
5	6	18	30	54	90			$0.09759*10^{-02}$
5	6	18	42	66	90			$0.08599*10^{-02}$
5	6	18	54	78	90			$0.08766*10^{-02}$
5	42	54	66	78	90			$0.48473*10^{-02}$
6	12	18	42	66		90		$0.06096*10^{-02}$
7	12	24	36	48		72	84	$0.06968*10^{-02}$

2.7.7 Height

Height records are relevant in the selection procedure for two reasons. Firstly, vertical stem growth is one of the parameters to estimate dry matter incorporated in the trunk which is one of the components of VDM. Selection for slow height increment therefore positively affects HI.

Secondly, as fruit bunches must be cut for harvesting, height at the level of the bunches affects the cost of harvesting. At a certain height, usually about 12 m above ground level, harvesting is not economically feasible anymore, and replanting is necessary.

The stem first forms a wide base without internodal elongation; vertical stem increment thereafter increase, until it reaches a by and large constant value. Conventionally, trunk height is measured from ground level to a reference point on a standard leaf in the crown, usually leaf 25 or leaf 41, counted from the youngest fully opened leaf (leaf 1); height increment is then calculated as the difference between sequential measurements.

This method for determining height increment is not preferred. The area around the palm trunk is usually not flat (holes left from removal of the tree crop, slopes etc.), palms may be leaning, and it also proves difficult to define the reference point on the petiole when the lower leaves are still attached to the stem. Moreover, rate of leaf production differs between palms and also declines with age after reaching a maximum in the second year after planting (Breure, 1987). Vertical increment for a certain interval, determined from measurements between a standard leaf in the crown, does therefore not correspond to the actual increment at the level of the growing point.

Reliability can be improved by measuring height to the insertion of the leaf base of known opening date; this corresponds to the level of the growing point. A further improvement can be made by measuring from a reference point on the stem instead of from ground level.

Note, however, that the bases of leaves produced during about the first 18 months after field planting cannot be used as a reference because these are concealed under the expanding trunk base (Breure & Powell, 1987). Only leaves which open about 30 months after planting remain clearly visible on the stem. Stem increment to estimate trunk dry matter production are thus best measured on the stem.

Regarding the height at which fruit bunches can still be conveniently harvested, however, it is clearly not feasible to actually measure the height of palms at the time of reaching the critical height for replanting. So the search is for height measurements during the early life of the palms which closely corresponds to mature height.

From 65 palms representing different progenies from three distinct seed sources of planting material, planted in North Sumatra (Indonesia), the latest fully opened leaf was marked from the start of bunch production, six months later, and then at four successive annual periods. Table 6 presents the height measured from the insertion of the first marked leaf base, that is the leaf which opened at the stage of harvesting (see Fig. 12 for the method of recording). These records per palm fit a logistic growth curve, $f(t) = A/(1 + B * e^{Ct})$, as previously described for leaf area (see section 2.7.4). The characteristic parameters A, B and C, and also the time to reach the

inflexion point of the growth function, $t_{0.50} = (1/C)\ln B$, vary considerably across palms (see Table 7); in other words, there are quick and slow starters. This indicates that increment values obtained between sequential measurements during a restricted period cannot be used to reliably compare palms for actual height at a certain age.

Indeed, correlations between height increments presented in Table 7 see for the periods 6 to 30 months after the start of harvesting and those of 30 to 54 months was very low and none were significant (correlation coefficients of 0.23 (11 d.f., P-value 0.453), 0.28 (16 d.f., P-value 0.258) and 0.33 (14 d.f., P-value 0.207) respectively for the three palm sources).

The correlations are higher and significant for height increment during 6 to 30 months and actual height at 54 months (correlation coefficients of 0.72 (11 d.f., P-value 0.005), 0.60 (16 d.f., P-value 0.009) and 0.46 (14 d.f., P-value 0.071) respectively for the three palm sources).

Mean height values to the base of leaf 25 (in cm) of 24 palms per plot of three different seed sources from the same experiment (see Table 8) show a similar relationship. The correlation coefficient of height increment, measured to the base of leaf 25, from year 1 to 5 after the start of harvesting and the actual height at 11 years was 0.58 (43 d.f., P-value 0.0001). It is therefore concluded that the actual height to a reference point in the crown at the end of the recording period should be measured to compare progenies for height at maturity.

Table 6. Height (cm) to the insertion of the first fully opened leaf, measured at various periods (months after the start of production) for three seed sources: D=Dami, GM= Gunung Melayu, MRS=Marihat Research Station, planted in North Sumatra (Indonesia) in October 1979

Date Time Palm	Nov-82 6	Nov-83 18	Nov-84 30	Nov-85 42	Nov-86 54 month
DM					
d23	19.5	67.4	151.5	238.5	326
d24	.	35.6	88.6	158.5	241
d25	20.4	106.3	188.5	281.8	.
d26	28	71	130	202	283
d27	.	60.6	124	194	276.3
d33	21.3	68.7	136.2	189.3	262
d34	37	98	173.6	258	355.8
d35	.	52.7	114	205.5	266.7
d36	20.3	89.8	166.7	236.5	313.3
d37	30	67.3	151.5	227	307.6
d42	12.9	59.5	116.7	192.4	264.6
d43	18	79.3	153.5	239	340.8
d44	38.5	105.3	172.8	235.2	326
d45	27	66.4	170	255	.
d46	25.7	79	160.7	230.3	284.2
d53	32	133.5	189.8	282	.

d54	18.2	84	155.4	227.6	318.1
d55	.	31.6	95.7	154	228.5
d56	22	88	170	253.3	322
d57	30	89	177.6	261.2	334
GM					
g22	.	23.7	83	148.5	229.8
g23	18.4	78	135.7	181	243.2
g25	34	120.2	206	290.7	.
g26	21.7	72.3	132.3	199	254.2
g27	25	66.6	133.6	181.5	241.5
g32	22.8	70.7	121.5	205.7	278.8
g33	16.6	64	124	184.5	268.2
g34	23.2	85	135.7	231	319
g35	23.9	57.5	115	183.3	248.8
g36	19.6	86	146	228.8	296
g37	15	49.6	102.3	152.7	236.7
g42	24.2	87.8	172	248.5	326.5
g43	36	81.4	155.3	234.2	324
g44	31.7	85	147.2	223.7	309.8
g45	19.2	80.2	145	209.5	218
g46	27	100.2	171.7	247	314
g53	.	39.5	101	191	275
g54	12	58.3	124.3	197.2	266.8
g55	18.7	68	116.5	175.2	234.7
g56	18.7	66	126.3	208	260.5
g57	24.2	84.4	155.2	243.4	310.3
MRS					
m22	30.2	77.3	144	204.1	276.9
m23	27	76.6	133.2	202.4	271.9
m24	11	53.9	130.3	200.5	273.7
m25	.	34.4	117.8	185.4	282.2
m26	14.7	80.5	133	203.4	261.8
m27	31.5	97.3	190	282.8	.
m32	23	67.5	137.7	201.1	261.7
m33	17.5	57	115.6	175	264.1
m34	.	30.5	87.4	156.5	218.8
m35	18.5	73.5	140.5	228.6	321
m36	20.2	68.4	132	231	328.8
m37	19	67.2	123	190.5	260.3
m42	17.6	75	124.5	190	253.4
m43	26.5	86	152.2	241.7	323.8
m44	29	75	135.1	217	300.5
m45	.	55.3	97.6	191.8	275.2
m46	29	89.4	130	229	321.6
m47	.	47.5	111.5	176.3	248.2

m52	.	48	107	171	242.3
m53	22.4	81	149.9	221.7	292
m54	10	96.7	165.4	267.5	.
m55	.	58.9	152.7	241.5	324
m56	24.6	71	150.4	241.5	336
m57	25	64.5	147.7	218.4	301.6

Note: A missing value is indicated by a dot (.)

Table 7. Parameters A, B, C and t0.5 = (1/C)lnB of the logistic growth curve fitted through height and height increment values (cm) between various periods (months after the start of production) for three different seed sources: DM = Dami, GM = Gunung Melayu, MRS = Marihat Research Station.

	palm	A	B	C	$t_{0.5}$	Months 6-30	Months 18-42	Months 30-54
DM	23	391.9	23.3	0.087	36.1	132	171.1	174.5
	26	397.9	16.6	0.068	41	102	131	153
	33	321.1	15.2	0.077	35.5	114.9	120.6	125.8
	34	481.4	14.3	0.068	39.1	136.6	160	182.2
	36	359.9	15.4	0.084	32.6	146.4	146.7	146.6
	37	378	18.2	0.08	36.1	121.5	159.7	156.1
	42	328.4	23.1	0.084	37.4	103.8	132.9	147.9
	43	445.7	20.6	0.077	39.2	135.5	159.7	187.3
	44	440.1	11.1	0.063	38.1	134.3	129.9	153.2
	46	310.8	16.2	0.094	29.8	135	151.3	123.5
	54	398.2	17	0.077	36.8	137.2	143.6	162.7
	56	361.7	17.4	0.09	31.7	148	165.3	152
	57	381.7	16.1	0.086	32.2	147.6	172.2	156.4
Mean		**384.4**	**17.3**	**0.08**	**35.8**	**130.4**	**149.5**	**155.5**
GM	23	279.7	12.5	0.079	31.9	117.3	103	107.5
	26	294.8	15.3	0.084	32.6	110.6	171.7	121.9
	27	282.6	13.4	0.079	32.8	108.6	114.9	107.9
	32	370.1	18.5	0.075	39.1	98.7	135	157.3
	33	364.7	18.9	0.073	40.5	107.4	120.5	144.2
	34	441.8	18.4	0.072	40.7	112.5	146	183.3
	35	320.8	18.2	0.077	37.9	91.1	125.8	133.8
	36	350.5	16.2	0.082	33.9	126.4	142.8	150
	37	377.6	22.7	0.067	46.6	87.3	103.1	134.4

	42	376	16.2	0.085	32.8	147.8	160.7	154.5
	43	443	15.4	0.069	39.7	119.3	152.8	168.7
	44	429.9	14.8	0.067	40.1	115.5	138.7	162.6
	45	227.6	17.3	0.118	24.2	125.8	129.3	143
	46	357.5	13.1	0.083	31.2	144.7	146.8	142.3
	54	314.8	23.7	0.089	35.4	112.3	138.9	142.5
	55	291.1	14.7	0.075	35.7	97.8	107.2	118.2
	56	299.6	20.2	0.091	33.1	107.6	142	134.2
	57	362	17	0.085	33.3	131	159	155.1
Mean		**343.6**	**17**	**0.081**	**35.6**	**114.5**	**135.5**	**142.3**
MRS	22	351	13.2	0.071	36.2	113.8	124.8	132.9
	23	350.9	14.3	0.072	37.1	106.2	125.8	138.7
	24	315.9	26.2	0.094	34.9	119.3	146.6	143.4
	26	303	15.6	0.084	32.7	118.3	122.9	128.8
	32	302.4	16.3	0.085	32.9	114.7	133.6	124
	33	396	21.4	0.069	44.3	98.1	118	108.5
	35	418.3	21.5	0.079	39.1	122	155.1	180.5
	36	446.2	25	0.079	41	118.8	162.6	196.8
	37	330.1	17.3	0.077	37.2	104	123.3	137.3
	42	311.5	15.1	0.077	35.4	106.9	115	128.9
	43	410	16.7	0.076	36.9	125.7	155.7	171.6
	44	416.5	17.2	0.07	40.5	106.1	142	165.4
	46	510	18.1	0.064	45.5	101	139.6	191.6
	53	344.7	15.6	0.081	33.7	127.5	140.7	142.1
	56	429.7	21.7	0.08	38.3	125.8	170.5	185.6
	57	370.4	19.3	0.081	36.4	122.7	153.9	153.9
Mean		**375.4**	**18.4**	**0.078**	**37.6**	**114.4**	**139.4**	**151.9**

Table 8. Mean height values (cm) per plot (24 palms) to the base of leaf 25 at different years after the start of production for three different seed sources: DM = Dami, GM = Gunung Melayu, MRS = Marihat Research Station

Plot		1 year	5 years	8 years	9 years	11 years
High density						
DM	1	90	372	608	677	801
	2	91	383	624	694	815
	3	80	348	592	654	797
	4	84	376	622	686	831
	5	77	355	599	661	802
GM	1	73	319	561	628	751
	2	80	342	585	655	793
	3	71	300	547	611	752
	4	66	314	567	625	776
	5	72	305	561	614	761
MRS	1	84	347	590	654	780
	2	87	345	582	649	774
	3	71	333	596	662	810
	4	72	345	597	661	814
	5	86	364	611	671	823
Medium density						
DM	1	83	346	574	636	756
	2	89	388	606	677	796
	3	80	350	584	649	783
	4	74	350	575	629	770
	5	79	346	588	641	781
GM	1	81	330	555	617	731
	2	80	331	549	613	732
	3	82	351	550	616	740
	4	76	332	570	632	761
	5	70	326	557	608	753
MRS	1	72	328	561	627	751
	2	87	368	581	647	772
	3	85	352	575	646	787
	4	80	345	571	632	786

	5	74	337	582	646	802

Low density

DM	1	88	362	586	653	761
	2	85	354	564	621	723
	3	83	370	581	639	760
	4	69	344	551	612	738
	5	84	353	578	629	753
GM	1	81	310	535	603	715
	2	77	338	538	592	717
	3	71	338	535	593	717
	4	76	326	530	587	712
	5	65	307	526	582	715
MRS	1	79	324	567	635	763
	2	73	332	555	619	758
	3	89	363	579	636	774
	4	78	313	538	596	725
	5	78	330	568	622	758

Note: high density = 160 palms per ha; medium density = 143 palms per ha; low density = 128 palms per ha.

2.8 Selection strategy

2.8.1 Selection on phenotype

The strategy is thus, by adopting the three-step procedure of parent selection, to develop *dura* x *pisifera* planting material with a slow height increment and which maintains a high Harvest Index (HI). The latter is achieved, as shown before, by additional selection for high Leaf Area Ratio (LAR) and Leaf-Mg, and low Vegetative Dry Matter Production (VDM), with Leaf Production (LPR) as the main component. Another desired feature is a quick expansion of the crown, see section 2.7.4, through selection for low $t_{0.95}$ and high k, until optimal Leaf Area Index (LAI) is reached as determined by Lm, the asymptotic maximum leaf area, (see Breure, 1985).

The scope of selection progress for these characters depends on the genetic variance in the source population, while the efficiency, in terms of the number of generations of breeding, depends on the heritability, defined as the square of the additive genotypic value and phenotypic value (Falconer, 1981). Although the genotypic variance of the two main objectives of selection, i.e. oil-and-kernel yield and Harvest Index (HI) is usually fairly high, progress in selection is usually hampered by their low heritability. However, various authors have shown that selection for bunch composition is usually higher than for bunch yield; this was confirmed by Breure & Bos (1992).

Characters for indirect selection (HT, LAR, Leaf-Mg and VDM), on the other hand, usually combine a relatively high genetic variance with a fair level of heritability. Good progress can thus be expected from phenotypic selection for the components of oil extraction rate and these secondary characters, in addition to oil-and-kernel yield and HI.

2.8.2 Selection for GCA values

In contrast to phenotypic characteristics, parental General Combining Ability (GCA) values, estimated by progeny testing the parents, reliably predict the performance of the *tenera* offspring (Breure & Bos, 1992). GCA values for oil-and-kernel yield and HI, as well as secondary characters, are therefore more suitable to develop planting material which maximizes photosynthetic production per hectare and efficiently transfers photosynthates produced per unit area into economic yield of the oil palm. In other words, the desired ideotype can be efficiently shaped through GCA values of the parents.

2.9 Plot size and shape

Conventionally, progenies are arranged in plots of 4 x 4= 16 palms. Recently, Breure & Konimor (1992) presented evidence that the expected response to selection of 16-palm plots is greater than with plots of a smaller number of palms. The plot is the experimental unit to compare progenies, so the average yield per plot is needed to estimate General Combining Ability (GCA) values. Carrying bunches to a central point for weighing is not convenient, mainly because the stacked fronds in the interrow do not allow for the easy movement of workers. Recording yield per palm is therefore preferred. The possibility of identifying outstanding *tenera* for clonal propagation is an additional advantage. Individual palm recording also permits more careful checking harvesting standards and loose fruit collection.

Experiments to compare sources of planting material are usually laid down in commercial estates. It is therefore more convenient to weigh the bunches per plot in bulk; moreover, data processing is simplified as plot yields are immediately available. This method of bulk recording becomes practical when all palms of one plot are bordering the same interrow path for transporting the bunches to the road (harvesters x path).

An inspection path is usually established in the centre of the field, parallel to the collection roads, to improve supervision. For convenience of carrying bunches, the distance between collection roads is about 250 m. There are thus 27 palms between roads, assuming the usual distance of 9 m between palms. The layout of harvesters x paths and inspection paths makes it convenient to establish the progenies in two half-rows (27 palms per plot), as shown in Fig. 1.
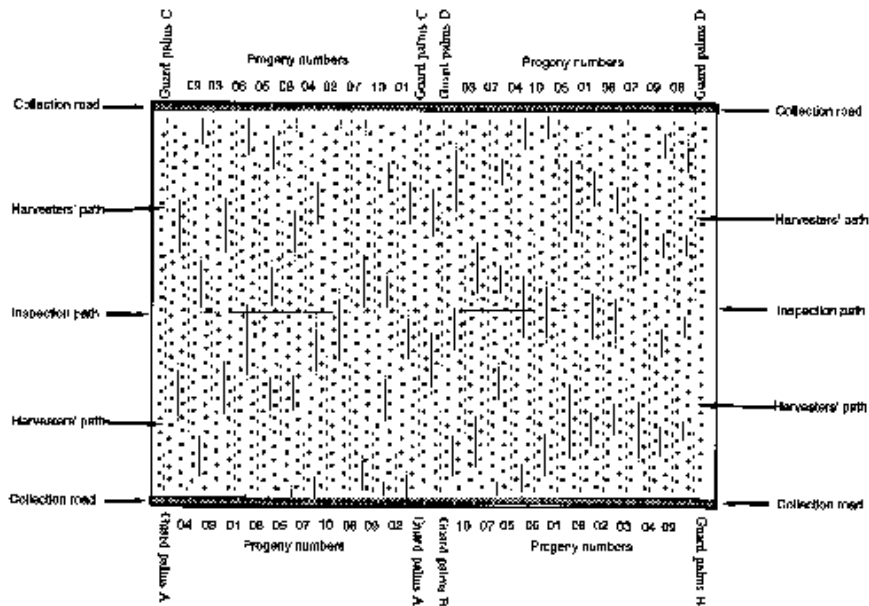
Fig. 1. Layout of a comparison trial for four seed
sources (one replication).

2.10 Arrangement of progenies in the field

Progenies may differ considerably in vigour, and thus mutual shading, which may create a bias in evaluating their yield potential, particularly if derived from distinct sources of parent palms. The effect of mutual shading can be diminished by grouping the progenies according to sources of planting material and, within these sources, to the height and mean leaf area established in the nursery.

3. Comparison of sources of planting material

3.1 Arrangements of progenies to compare sources of planting material

The following layout in Fig. 1 presents one replication of a trial to compare four sources of planting material, A, B, C and D, each represented by 10 progenies.

Progenies are arranged in plots of 27 palms consisting of two rows between the inspection path and the collection road. Sources of planting material are surrounded by a double guard row with a mixture of progenies from the bordering seed source. The fruit bunches per plot of two half-rows are assembled at the collection road and weighed in bulk. A signboard indicating the number of the progeny (1 to 10) is placed at the side of the collection road.

3.2 Statistical analysis for comparing sources of planting material

There is much scope for improving trials to compare planting material. Sources of planting material are often compared by using a mixture of progenies planted in unreplicated commercial blocks; in the worst case sources are not even planted at the same time, and are also often compared under different environmental conditions. Results are more reliable, if sets of identified crosses representing each source of planting material are compared in a replicated randomized block design; with S sources and with P progenies per source; each replication is in this way a complete block accommodating S*P progenies. However the sources are on main-plots and the progenies are randomized over the sub-plots in a main-plot (see Fig. 1). ANOVA Table in this chapter presents the Analysis of Variance (ANOVA) table of such source comparison experiment with R replications:

Before showing how to enhance the power of the tests, a basic aspect of statistical analysis is first briefly reviewed. A significant treatment effect (in this case sources of planting material) means that the F-value in the ANOVA-table should be above a critical F-value; this critical value is mainly determined by the Degrees of Freedom (DF) of the denominator Mean Square in the test-statistic. As can be seen in tables of the F-distribution, F-values required for significance are very high if the number of DF of the denominator is below 7, then values decrease sharply and level off when DF becomes 10. In practice therefore the design of experiments should ensure that the number of DF of the denominator is at least 10.

Because we have a design where per replication main-plots have been used to compare sources of planting material, and within a main-plot the progenies of a source are allotted at random to the sub-plots, we have an Analysis of Variance according to a so-called mixed model. There are three variance components, namely Var(RepxSource) for the main-plot error, Var(Prog (Source)) for the variation caused by taking a random sample of all possible progenies from a source, and Var(Error) for the sub-plot error, (see for example Verdooren (1988) or Searle, Casella & McCullogh (1992)).

The expected values of the Mean Squares (EMS) in the ANOVA table are as follows:

EMS(Rep)= Var (Error) + P* Var (Rep x Source) + Q(Rep),

where Q(Rep) is a quadratic function of the fixed replication effects;

EMS(So)= Var (Error) + R*Var (Prog(Source)) + P*Var (Rep x Source) + Q(Source),

where Q(Source) is a quadratic function of the fixed source-effects;

EMS(RepxSo) = Var(Error) + P* Var(RepxSource) ;EMS(Prog(So)) = Var(Error) + R*Var(Prog(Source)) ; EMS(E)= Var(Error) .

Notice that sometimes one uses an ANOVA table where replicate effects are taken to be random instead of fixed. In that case in the EMS(Rep) the term Q(Rep) is replaced by P*S*Var(Rep). But this change has no effect in the following test.

From these EMS it follows that to test the hypothesis H0: " Var(Prog(Source)) = 0" we must use MS(E) as the denominator in the test statistic, MS(Prog(So))/MS(E), because under H0 EMS(Prog(So)) is equal to EMS(E). This test-statistic has under H0 an F distribution with S(P-1) and S(P-1)(R-1) degrees of freedom. To test the hypothesis H0: "The sources are alike in effect, hence Q(Source)=0 " we must use a combination of Mean Squares as

MS(RepxSo) +MS(Prog(So)) - MS(E)

= MS(Comb) as the denominator in the test statistic, MS(So)/MS(Comb), because under H0 EMS(So) is equal to EMS(RepxSo) + EMS(Prog(So)) - EMS(E).

This test-statistic has an approximate F-distribution with S-1 and C degrees of freedom, where C is to the nearest integer rounded value of

$$[MS(Comb)]^2 /\{ [MS(RepxSo)]^2/(R-1)(S-1) + [MS(Pro(So))]^2 /S(P-1) + [MS(E)]^2 /S(P-1)(R-1)\}.$$

The test of comparing progenies-within-source (with test-statistic MS(Pro(So))/MS(E) becomes more powerful, in terms of DF, with increasing the number of replications (R).

The main objective is, however, to test whether sources of planting material differ significantly. The power of the test of comparing sources (with test-statisitc MS(So)/MS(Comb)) is with a given number of sources enhanced by increasing the number of progenies within sources (P). Thus, as may be expected, an adequate number of identified progenies is crucial for a meaningful comparison of sources of planting material.

ANOVA Table

Source variation	Degrees of freedom	Mean Square	F-test
Replications (Rep)	R-1	MS(Rep)	
Between sources (So)	S-1	MS(So)	MS(So)/MS(Comb)
Main-plot error (RepxSo)	(R-1)(S-1)	MS(RepxSo)	
Between progenies within sources (Prog(So))	S(P-1)	MS(Prog(So))	MS(Prog(So))/MS(E)
Error (E)	S(P-1)(R-1)	MS(E)	
Corrected total	SPR-1		

where R= number of replications, S= number of sources of planting material, and P= number of progenies per source; a combined Mean Square is MS(Comb), defined as MS(Comb)=MS(RepxSo) + MS(Prog(So)) - MS(E).

4. Recording of traits for selection

4.1 Technique of recording

The actual technique of recording yield and growth has often received little attention. There are numerous examples of well designed breeding programs which are recorded with awkward equipment, resulting in unreliable and labour-intensive yield and growth recording. The following sections outline some relevant aspects of recording technique.

4.1.1 Bunch yield

Bunches are either weighed per palm (for example, in the source population of parent palms) or assembled per experimental plot, and weighed in bulk (testing of sources of planting material). Bunches are weighed per palm by using a spring balance attached to a tripod. The tripod is made of thin pieces of board, which are loosely tied together at the top. By extending one leg, the tripod can be used to lift a weight of about 150 kg by one worker, so that total plot yield can be recorded. Both methods are depicted in Fig. 2.

4.1.2 Bunch analysis

In a special building, bunches are analyzed for the components of oil and kernel extraction, i.e. the ratio of fruit-to-bunch, mesocarp-to-fruit, oil-to-mesocarp and kernel-to-fruit, as follows.

Bunches are weighed at their arrival. After weighing, the spikelets are removed and weighed. A random sample of spikelets is then taken and weighed. The fertile and parthenocarpic fruit with oil-bearing mesocarp are removed and weighed to give the ratio of fruit to empty spikelets of the sample. As the weight of the total spikelets of the bunch is known, the fruit-to bunch-ratio can be calculated.

A sample of normal fertile fruits is also taken shortly after arrival of the bunch. The fruit sample is weighed and the mesocarp is scraped off; the nuts are weighed and the weight of the mesocarp is calculated by difference. This gives the weight of the mesocarp at the moment of taking the fruit sample, that is the weight unaffected by drying of the fruit before scraping and during scraping of the mesocarp (the importance of determining the accurate weight of the mesocarp will be explained below).

The mesocarp-to-fruit ratio can be calculated from the weight of the fruit sample and the nuts. Nuts are air dried for about 3 days to facilitate cracking. The shell is removed and the kernels are weighed, so the kernel-to-fruit ratio can be calculated.

All of the scraped mesocarp is dried in an oven at 105° C to constant weight to determine the moisture content of the fresh mesocarp. As mentioned before, the fresh mesocarp is calculated by difference, so it is crucial that all mesocarp is carefully collected during scraping; any mesocarp spilled is calculated as moisture losses during drying in the oven, and thus overestimates the moisture content of the mesocarp.

Figure 3 presents a table, designed to eliminate losses during scraping. It shows how the mesocarp is directly collected in an aluminum tray fitted in a drawer under a hole in the top of the table. The oil content of a sample of oven-dried mesocarp is determined by direct extraction with large Soxhlet extractors. As the moisture content of the sample is determined earlier in the analysis, the ratio of oil-to fresh mesocarp content can be calculated.
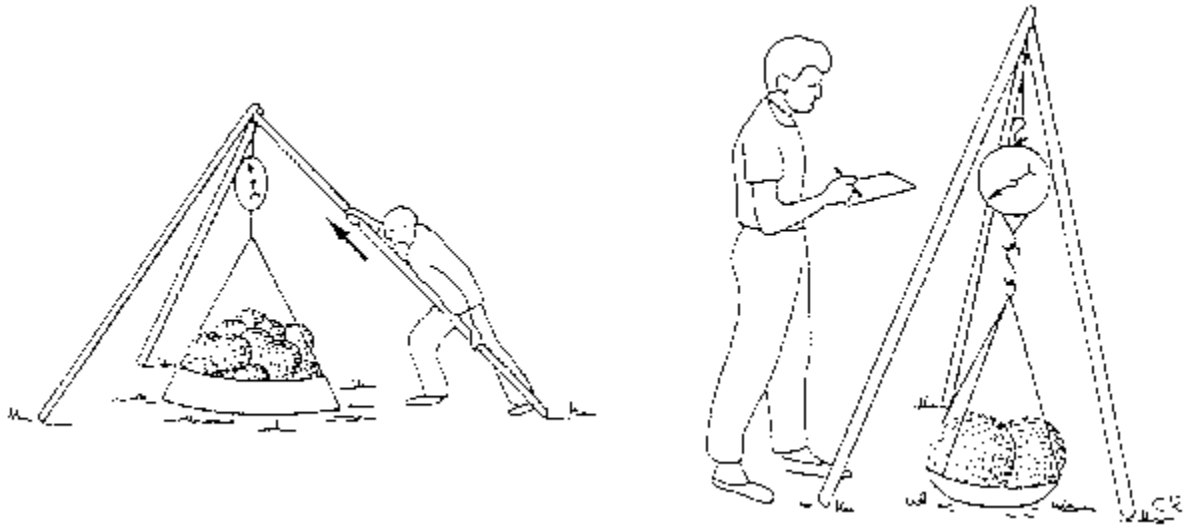
Fig. 2. Weighing fruit bunches per lot (left) and per individual palm (right).

Research centers show a wide variety of bunch analysis procedures, in terms of sampling methods, the weight of the sample and its method of preparation (for a review see Hartley, 1988). Bunch analysis is the most costly part of the breeding procedure. The ultimate aim is to determine the components of oil and kernel extraction of individual palms or families with sufficient precision at minimal costs. More research is still needed to achieve this objective.
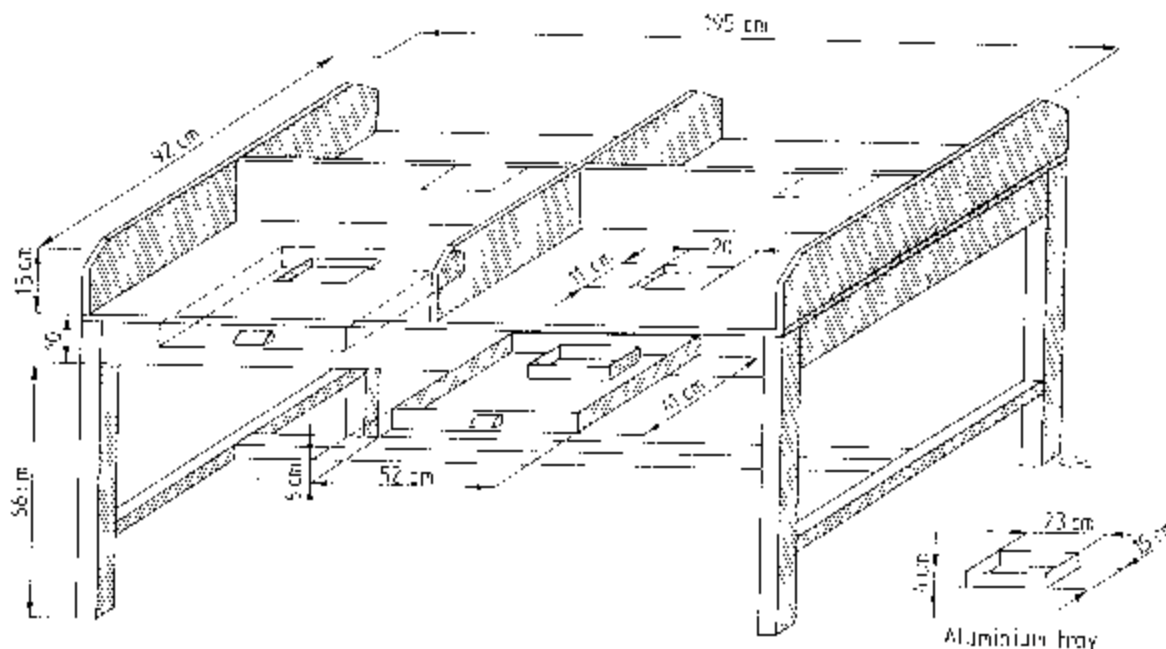
Fig. 3. Scraper table to remove mesocarp in a laboratory for bunch and fruit analysis**.**

4.1.3 Leaf measurements

An oil palm leaf is attached to the stem with the petiole. The leaf bears leaflets on each side of the leaf stalk. The latter may be divided into two zones: the rachis bearing the leaflets, and the petiole, which is much shorter than the rachis and bears only short lateral spines. This is clearly illustrated in Fig. 4. For a more specific description see Hartley (1988).

Leaf marking

Leaves assigned for measuring are marked at the date of opening, either by the number of the month (Fig. 5) or by a paint dot, the color of which corresponds to the date of marking; the latter method is quicker and preferred for progeny testing. As leaf size does not change anymore after opening, measurements can then be done when convenient.
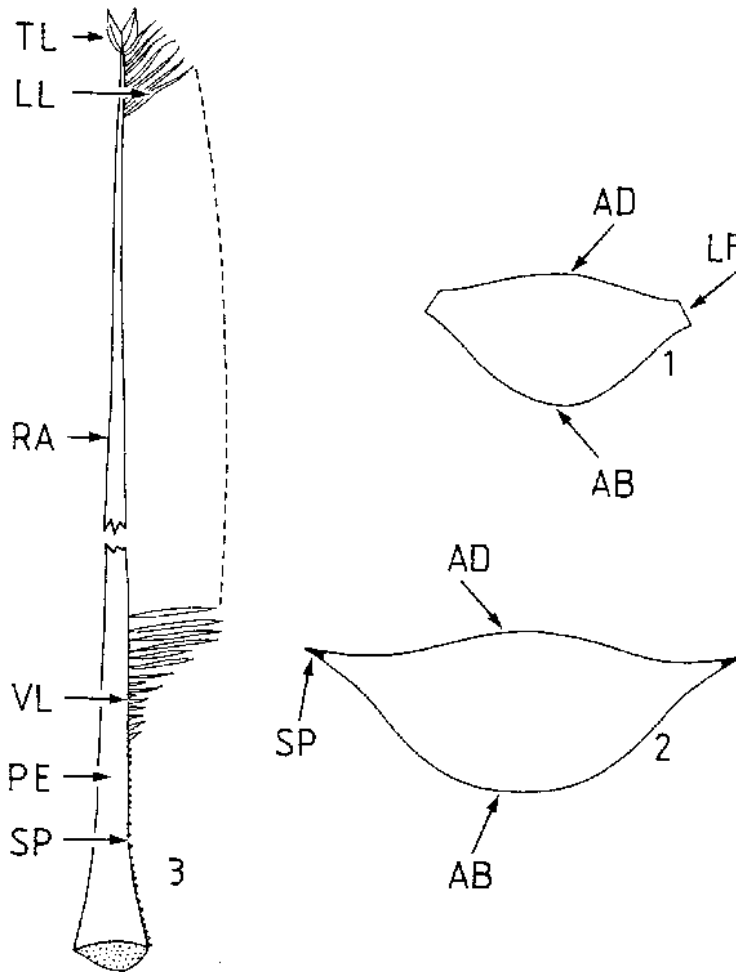
Fig. 4. The oil palm leaf: 1, cross-section of rachis; 2, cross- section of petiole; 3, diagram of oil palm leaf; AD, adaxial face; AB, abaxial face; LF, lateral face; SP, spine; RA, rachis; PE, petiole; TL, terminal pair of ovate leaflets; LL, liner leaflets; VL, leaflets with vestigial laminae

4.1.4 Trunk measurements

Height

Height increment for a given period is measured between the insertion of leaf bases of known opening date. Fig. 13 shows how the level of the lower reference leaf base is obtained by means of a fluid leveller because it is usually not directly above the leaf base marked at a later date; a movable bar is then brought to the insertion of the higher marked leaf base. The height difference between the two bases of the leaves marked at different ages (level of the growing point) is directly read from the attached sliding tape.
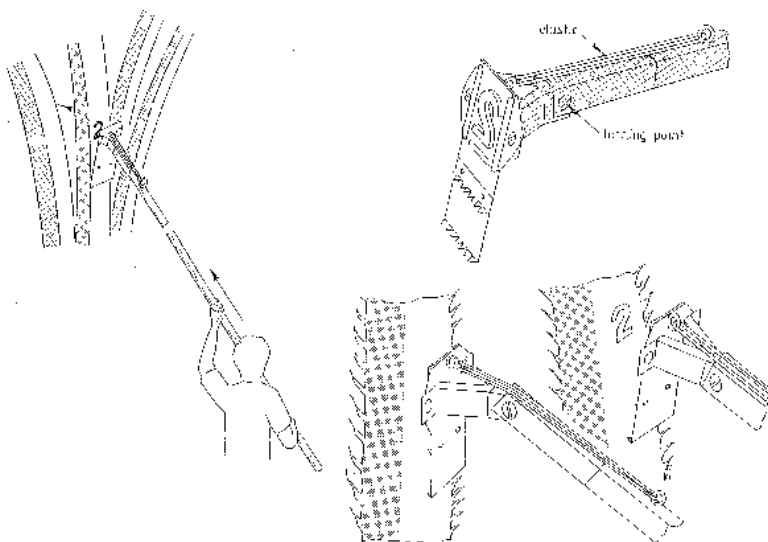
Fig. 5. Marking the latest fully leaf with the number of the month

Actual height is measured to the base of leaf 25 at the end of yield recording height to compare progenies for mature height. This measurement is indicative for the economic life.

Trunk diameter

The diameter is measured at about 150 cm above the ground, that is, when the trunk has already decreased to and largely constant value. Fig. 13 shows how the widest distance between opposite sides of the trunk can be obtained between every 4th spiral. This proves to be a convenient guide for unskilled workers to prune obstructing leaf bases in order to expose the stem. Fig. 14 also depicts the equipment and technique of measuring the stem diameter.
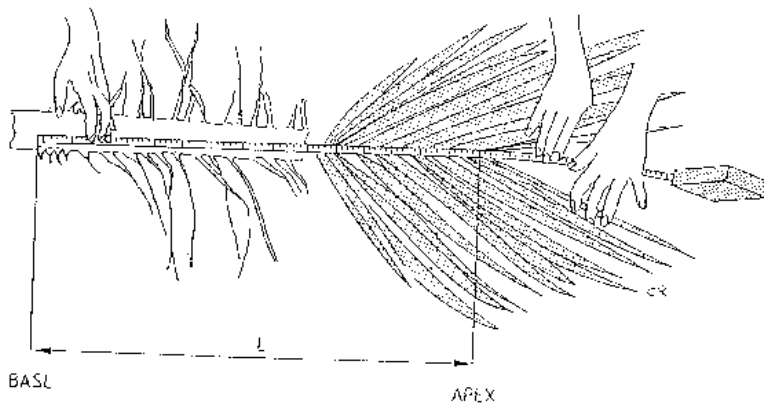
Fig. 6. Measuring the length of the rachis from the base to the apex

4.2 Measurements to estimate growth parameters

Vegetative Dry Matter production (VDM)

VDM is restricted to dry matter incorporated in trunk growth and leaf production (above-ground dry matter production). Trunk dry matter production is estimated from measurements of trunk increment, trunk diameter and an estimate of dry matter per unit trunk volume (kg/dm3); the latter depends on palm age (years) as follows:

0.0076*years after planting + 0.083 (Corley et al., 1971).

Leaf dry matter production is the product of leaf production and mean weight of the leaves. The Rachis length. The rachis is measured from the start of the rudimentary leaflets to the split of the terminal leaflets, the apex, (Fig. 6).

Petiole cross-section area

The width and the depth of the petiole are measured at the initiation of rudimentary leaflets. As the centre of the rachis is usually below the edge, an extension is fitted to a conventional pair of calipers as shown in Fig. 7.
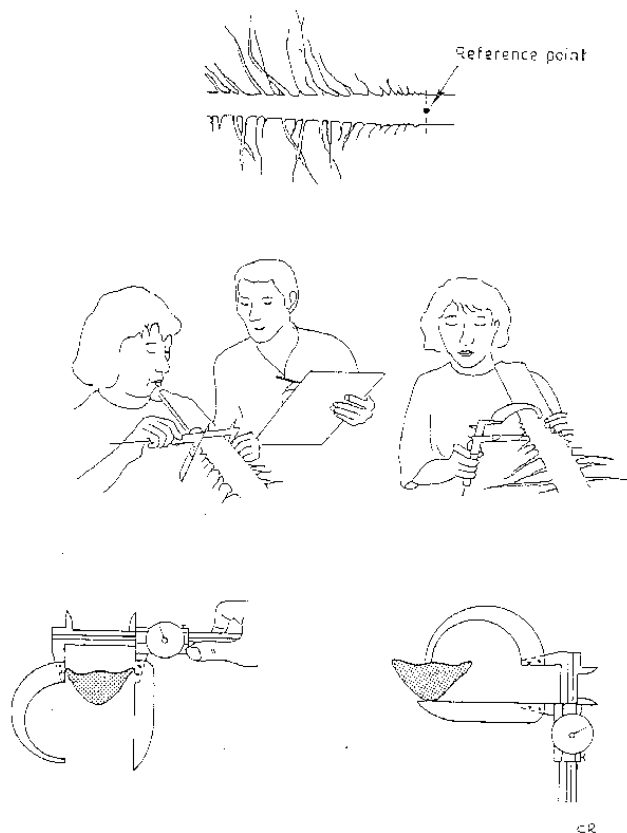
Fig. 7. Measuring the width and depth of the petiole at the point of insertion of the leaflets.

Fig. 7 also shows how the technician reads the measurements directly over the shoulder of the worker holding the calipers.

Length x width of leaflets

From both sides of the rachis a set of 10 leaflets are cut immediately under the region 2/5 from the apex (this region is the area with the larger leaflets). From each of these two sets the three longer leaflets are sampled (Fig. 8) and the length and width are measured (Fig. 9).
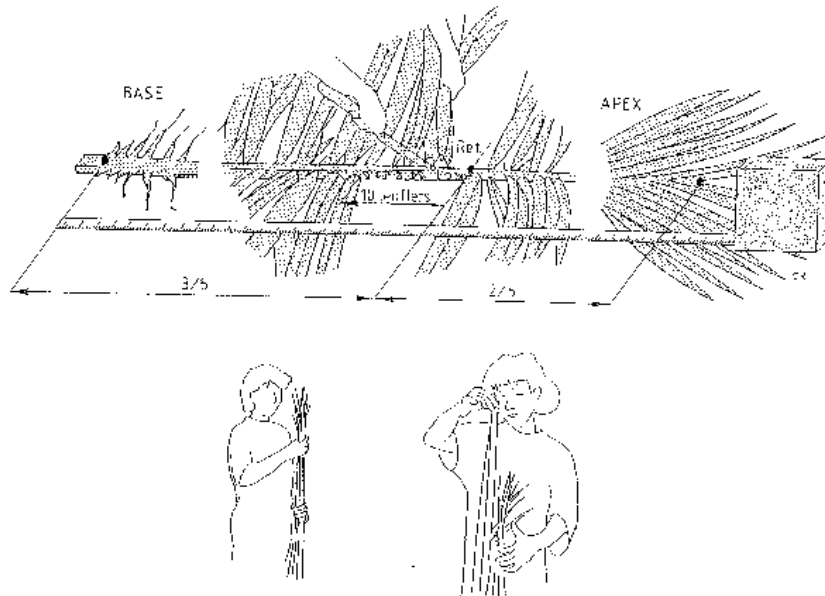
Fig. 8. Sampling of 3 leaflets at each the right and left rachis among 10 leaflets cut at 2/5 from the apex.
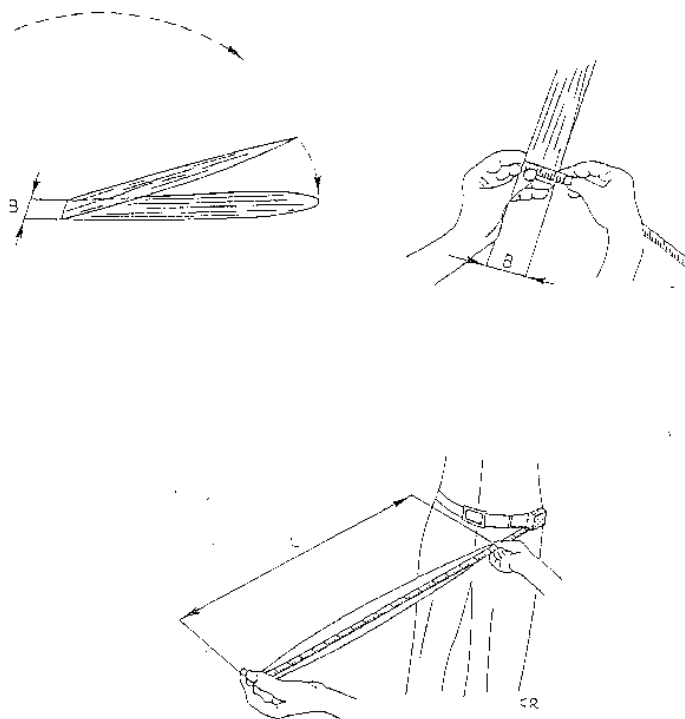
Fig. 9. Measuring the length and width of leaflets.

Counting leaflets

Only the number of leaflets on one side of the rachis are counted, including rudimentary leaflets at the base and fused leaflets at the apex, using a hand-counter (Fig. 10).

Leaf counting

Leaves open at regular intervals, usually in mature palms, about two leaves per month. In order to record the number of leaves which open at a certain time interval, the youngest fully openend leaf (leaf 1) is marked at the start and at the end of the period. As the marked leaf becomes older, a higher rank number is assigned. Since the difference in ranking (age) of sequential leaves on each spiral is eight leaves, the position of the first marked leaf in relation to the latest can easily be obtained from a specially designed diagram (Fig. 11). This gives the order of the leaves in the crown; the rate of leaf production is then obtained by the difference between the order of the leaves.
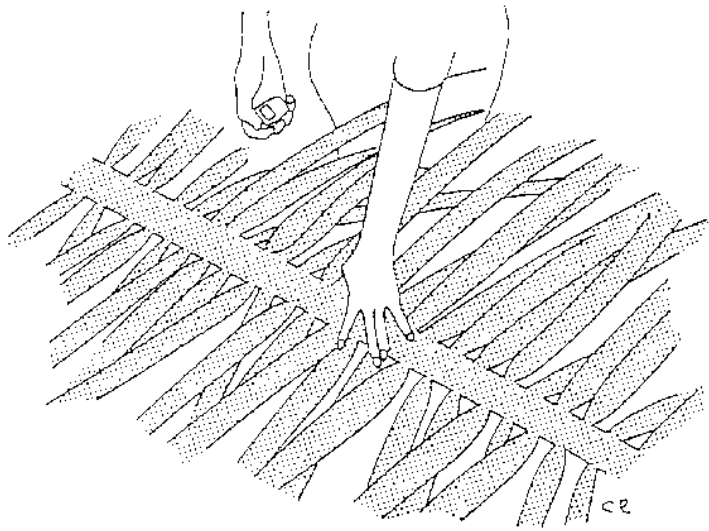
Fig. 10. Counting leaflets with a hand counter

In order to number the spirals, the palm should first be classified whether its direction of the spiral is left- or right-handed (Fig. 12). The side at which inflorescences emerge from the leaf axils can be used for a final classification: inflorescences of right-handed palms emerge from the right side of the leaf axil and those of left-handed from the left. The spiral with the youngest opened leaf is assigned spiral 1. Fig. 12 also shows how spirals can then be conveniently numbered by following the direction of the spiral downwards.

For example, the leaf was marked at opening in November 1992 and again one year later. Assume that leaf production for this period was recorded in December 1993. The leaf marked in 1992 is at that date leaf 4 in spiral 7 (leaf 27) and that marked in 1993 leaf 1 in spiral 4 (leaf 2). Leaf production for November 1992 to November 1993 is then 27-2=25 leaves.

Frond Number	Parastichy Number								Number of Fronds produced since marking
	1	2	3	4	5	6	7	8	
	Present position of previously marked leaf								
1									0
2				①					1
3							1		2
4		1							3
5					1				4
6								1	5
7			1						6
8						1			7
9	2								8
10				2					9
11							2		10
12		2							11
13					2				12
14								2	13
15			2						14
16						2			15
17	3								16
18				3					17
19							3		18
20		3							19
21					3				20
22								3	21
23			3						22
24						3			23
25	4								24
26				4					25
27							④		26
28		4							27
29					4				28
30								4	29
31			4						30
32						4			31
33	5								32
34				5					33
35							5		34
36		5							35
37					5				36
38								5	37
39			5						38
40						5			39
41	6								40
42				6					41
43							6		42
44		6							43
45					6				44
46								6	45
47			6						46
48						6			47

Fig. 11. Diagram for determining the number of leaves produced in a given period, showing, as an example, leaves marked at an interval of one year
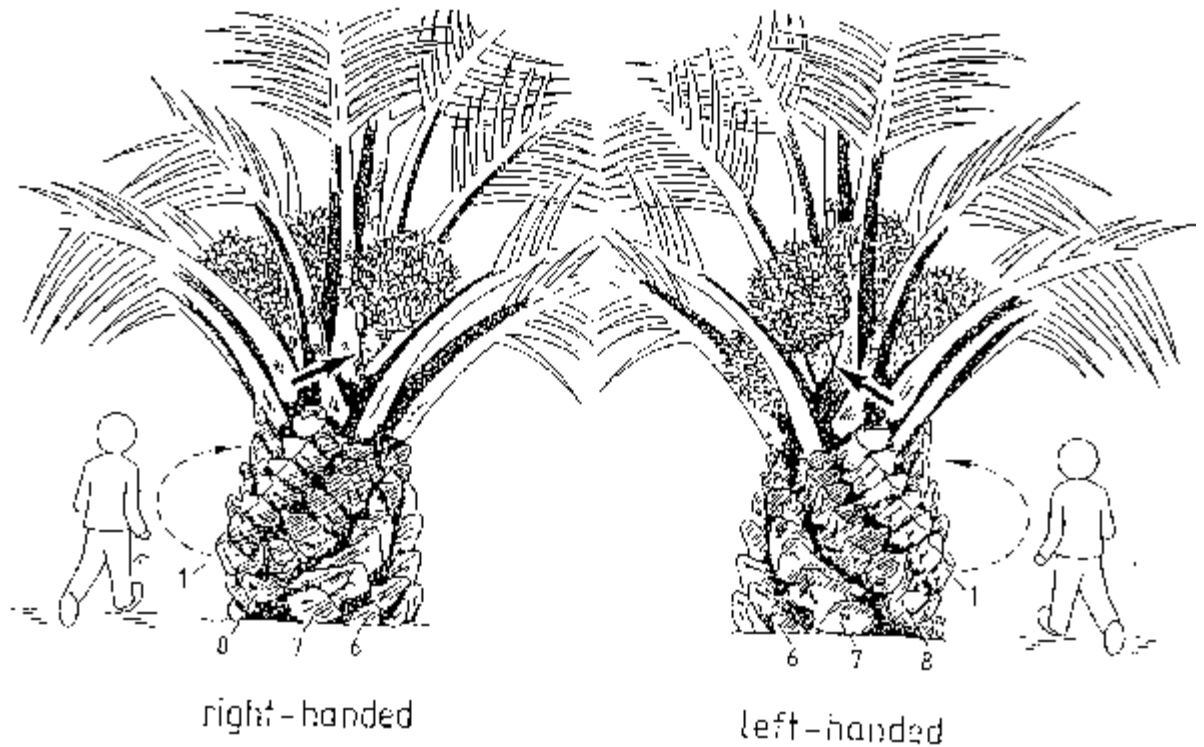
Fig. 12. Direction of numbering the eight spirals on palms showing "right-handed" and "left-handed" phyllotaxis

4.1.4 Trunk measurements

Height

Height increment for a given period is measured between the insertion of leaf bases of known opening date. Fig. 13 shows how the level of the lower reference leaf base is obtained by means of a fluid leveller because it is usually not directly above the leaf base marked at a later date; a movable bar is then brought to the insertion of the higher marked leaf base. The height difference between the two bases of the leaves marked at different ages (level of the growing point) is directly read from the attached sliding tape.
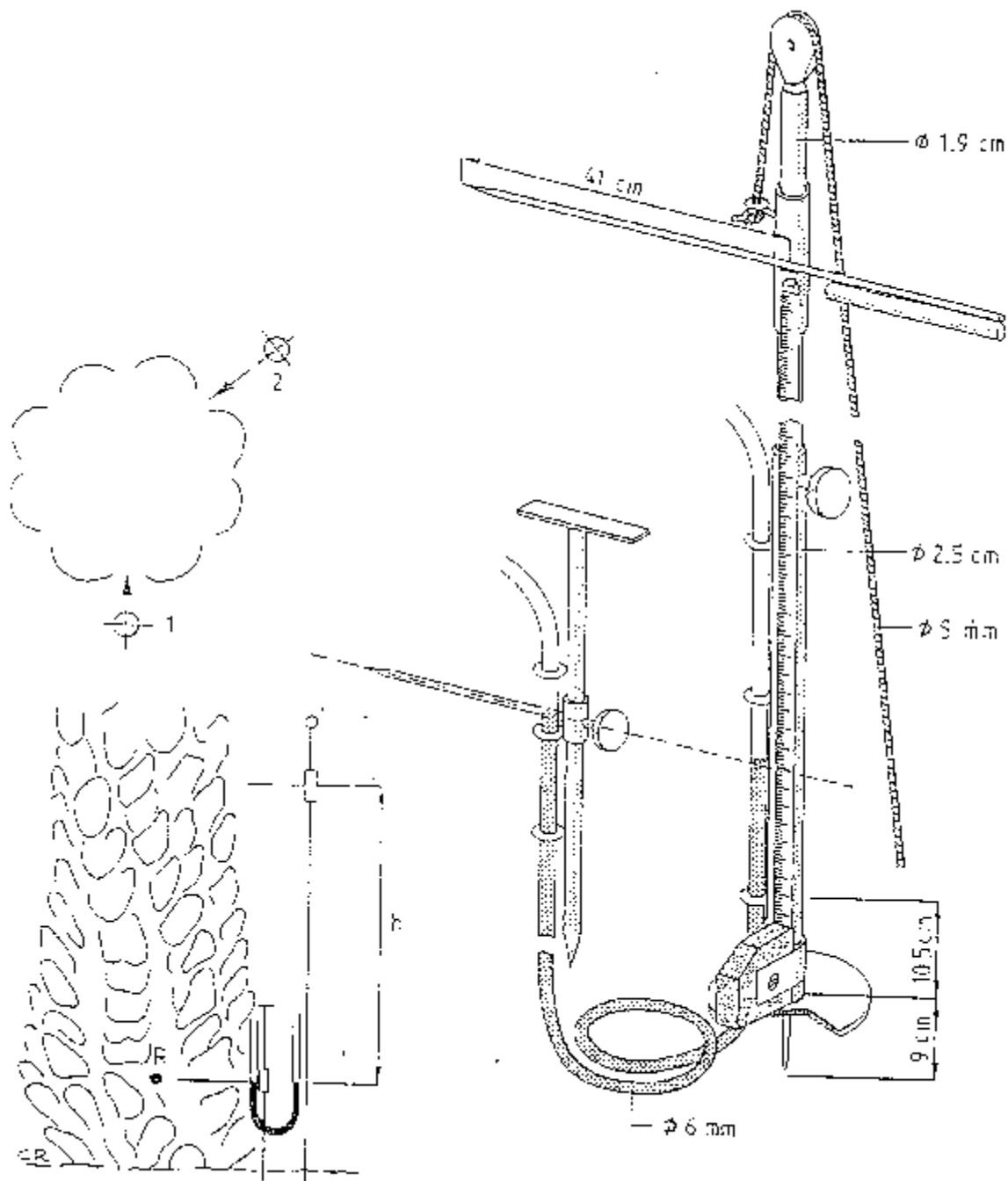
Fig. 13. Measuring height between leaf bases marked at the time of the leaf 1 stage, corresponding to the growing point.

Actual height is measured to the base of leaf 25 at the end of yield recording height to compare progenies for mature height. This measurement is indicative for the economic life.

Trunk diameter

The diameter is measured at about 150 cm above the ground, that is, when the trunk has already decreased to and largely constant value. Fig. 13 shows how the widest distance between opposite sides of the trunk can be obtained between every 4th spiral. This proves to be a convenient guide for unskilled workers to prune obstructing leaf bases in order to expose the stem. Fig. 14 also depicts the equipment and technique of measuring the stem diameter.
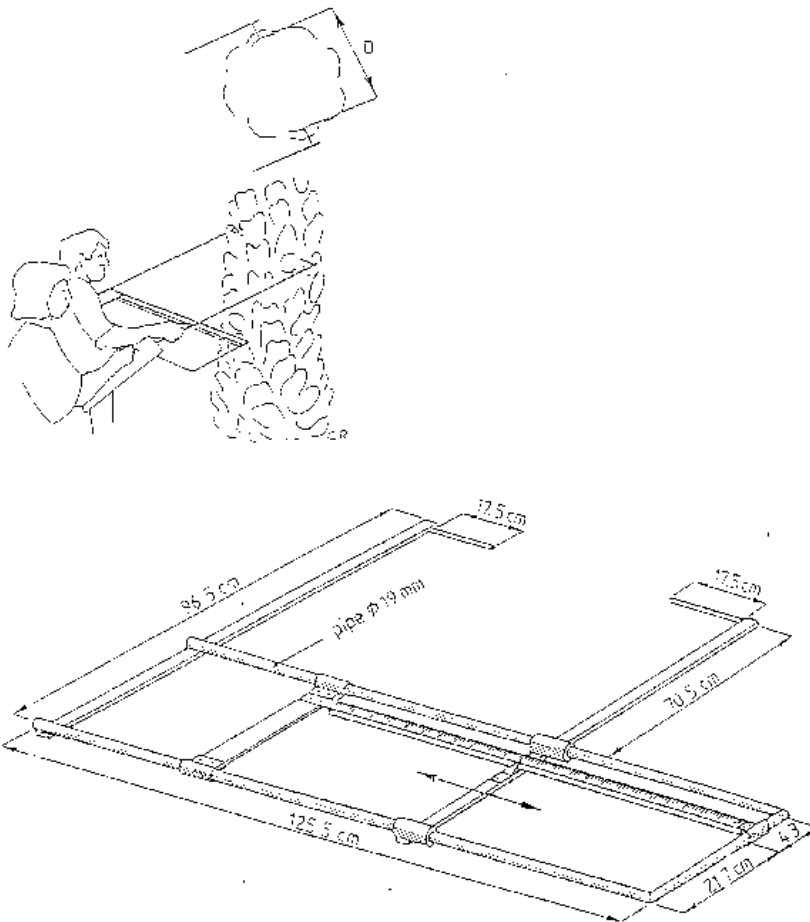
Fig. 14. Measuring the diameter of the trunk

4.2 Measurements to estimate growth parameters

Vegetative Dry Matter production (VDM)

VDM is restricted to dry matter incorporated in in trunk growth and leaf production (above-ground dry matter production). Trunk dry matter production is estimated from measurements of trunk increment, trunk diameter and an estimate of dry matter per unit trunk volume (kg/dm^3); the latter depends on palm age (years) as follows:

0.0076*years after planting + 0.083 (Corley et al., 1971).

Leaf dry matter production is the product of leaf production and mean weight of the leaves. The weight of an individual leaf is estimated as 0.1023*P + 0.2062 , where P is the mean petiole width * depth in cm^2 or petiole cross-section. (Corley et al., 1971).

Leaf area (LA)

The area of a single leaf is estimated, using the method of Hardon et al. (1969) as

c*(n * lw), where

c = a correction factor which varies slightly with palm age (0.51 to 0.57 for palms of 1-2 to 8-11 years, respectively);

n= number of leaflets and

lw= the mean of length * mid-width of three leaflets sampled from each side of the rachis.

Leaf area ratio (LAR)

LAR is defined as N*A/V , where

N= leaf production,

A= mean area per leaf and

V= VDM.

Parameters of the logistic growth curve for leaf area at different years from planting

The logistic growth function f(t) = A/(1 + B* e^{-Ct}) must be fitted to the data by the Least Squares Method to estimate the parameters A, B and C.

Rewriting the logistic growth function as follows:

$$f(t) = \frac{L_m}{(1 + \frac{L_m - L_i}{L_i}) e^{-kt}}$$

with

A= Lm, B= (Lm - Li)/ Li and C= k,

where k = the relative rate of growth of the mean leaf area,

Lm = asymptotic maximum leaf area,

Li = leaf area at field planting.

For selection purposes this is conveniently expressed as the time to reach 95% of the maximum leaf area ($t_{0.95}$) as follows:

$$t_{0.95} = -\left(\frac{1}{k}\right) * \ln\left\{\frac{\left(\frac{L_m}{0.95 L_m}\right) - 1}{\left(\frac{L_m}{L_i}\right) - 1}\right\}$$

Bunch Dry Matter Production (Y)

This is calculated as 53% of the weight at harvesting.

Components of oil and kernel extraction

The percentage oil-to-bunch is calculated as the product of the ratios fruit-to-bunch, mesocarp-to-fruit and oil-to mesocarp. The percentage kernel extraction is the product of fruit-to-bunch and kernel-to-fruit.

Bunch Index (BI)

BI is the ratio of dry weight of fruit bunches to total above-ground dry matter production per palm.

BI=Y/(V + Y), where

Y=dry weight of bunches,

V=VDM.

Harvest Index (HI)

HI is the ratio of oil and kernel yield to total above-ground dry matter production.

HI= (ratio of oil-and-kernel to bunch) * bunch yield/(V + Y).

4.3 Timing of measurements

4.3.1 Bunch yield

Number of bunches and the total weight are determined at each harvesting round of 7 to 10 days. This gives the number of bunches and their mean weight per palm or per plot.

Yield in favourable environments increases rapidly during the first 3 years and then stabilizes (Breure, 1988). As early yield greatly depends on the spread of the leaf area, another two years at mature yield level are recorded to obtain a fair estimate of the yield potential. A total of five years of yield recording is thus needed.

4.3.2 Bunch analysis

Oil content of the mesocarp increases sharply with age during the early period of bunch production. Conventionally, oil analysis starts therefore when the oil content by and large stabilizes, about 2 to 3 years after the start of production. The rate of increase differs among progenies (Corley, personal communication), so precision is enhanced if the analysis of bunches starts shortly after the start of production. Early information on extraction rate is also desirable in case new sources of planting material are tested.

Until more information on the number of samples per progeny is available, a tentative schedule is to analyze 32 bunches/progeny/year during the first three years of production and 64/bunches/progeny/year in the fourth and fifth year.Because of the wide diversity in bunch composition among palms in a progeny, the progeny sample must include as many palms as feasible. It is equally important to analyze each month the same number of samples per progeny to eliminate known seasonal variations in oil content.

4.3.3 Vegetative Growth

Clearly, growth recording should be completed by the end of the five-year period of yield recording. The start of production depends on environmental conditions, age of seedlings etc. It is therefore more convenient to follow initially a recording schedule according to the time after field planting, but change this later on to the time after harvesting.

4.3.4 Leaf measurements

Measurements are required for the following objectives: (i) To establish the logistic growth curve of leaf area against age. The leaf area fits the logistic growth function $f(t)=A/(1+B*e^{Ct})$ as explained in section 2.7.4. A set of measurements at 6, 42, 66 and 90 months after field planting is recommended for a step 2 progeny test (first screening). For a step 3 progeny test an additional measurement is recommended 12 months after field planting.

(ii) To obtain a reliable estimate of Leaf Area Ratio (LAR).

(iii) To obtain the petiole cross-section in order to estimate dry leaf weight.

The timing will be described in the following sections.

4.3.5 Leaf marking

The first measurements are done on leaves marked at 6 months from planting, that is when leaf size starts to increase after a period of so-called "transplanting shock" due to root disturbance during the movement of seedlings to the field. This leaf is assigned L0.

The second leaf, marked one year after the start of bunch production (L1), serves as a reference point for height measurements. The latest fully opened leaves are marked at the end of the third and fifth year of production, and assigned L2 and L3, respectively. Note that these leaves refer to the date of opening while leaf 25 and leaf 41 (see section 2.7.6 and section 4.1.3) are actually assigned according to their ranking in the crown.

4.3.6 Production of leaves

Leaf production is determined between L1 and L2; and between L2 and L3.

4.3.7 Height measurements

Height increment is measured from L1 (reference point) to the insertion of L2 (stem increment for a period of two years). Also to leaf 25 at the end of the fifth year of production.

4.3.8 Trunk diameter

Measurements are done only once at the end of the fifth year of production, when the lower leaves are about 150 cm above the ground.

4.3.9 Leaf magnesium level

To assess magnesium status of individual palms, samples are collected in 6 successive months and bulked (selection of parent palms in the source population). Samples to determine progeny means are collected for each palm once and bulked per plot. Samples thus obtained are analyzed for magnesium content. Sampling should be done when symptoms of magnesium deficiency are most pronounced. This is usually at the end of the second year of production when lower leaves are still exposed to light and palms are bearing the first heavy crop (stress due to high fruiting activity).

4.3.10 Crown disease

Breure & Soebagjo (1991) observed the first symptoms of crown disease on newly emerged leaves at 8 months from planting. Severity reached a peak at 12 months; thereafter, severity gradually diminished until it by and large stabilized between 22 and 35 months. Note, that severity, in terms of the degree of bending of the leaves, was recorded on the nine youngest leaves. The bend is permanent, so symptoms persist for at least a year when scoring on the nine youngest leaves is done.

Incidence can therefore conveniently be scored at 18, 30 and 42 months after planting. Once crown disease has been observed, the palm is marked to avoid double counting in another round. At 42 months the percentage of affected palms, recorded during the three rounds, can be established for each progeny.

4.4 Components of growth

Table 9 presents an example of mean measurements of one progeny assembled at various periods during the first 90 months after field planting; it is assumed that bunch production starts 30 months after planting. The following sections illustrate how the relevant growth parameters are estimated.

4.4.1 Bunch yield (kg /palm)

Bunch yield for year 1 to 5 of production are 65, 140, 230, 185 and 190. The dry weight is estimated as

Year 1: $0.53 * 65 = 34.75$

Year 2: $0.53 * 140 = 74.20$

Year 3: $0.53 * 230 = 121.90$

Year 4: $0.53 * 185 = 98.05$

Year 5: $0.53 * 190 = 100.70$

4.4.2 Oil and kernel extraction ratio

From the bunch analysis results it is calculated that the average extraction of the fresh fruit bunches is 27.3% mesocarp oil and 2.7% kernels.

4.4.3 Leaf measurements

Leaf weight (kg)

LO : $0.1023 * 6.78 + 0.2062 = 0.90$

LO*: $0.1023 * 9.26 + 0.2062 = 1.15$

L1 : $0.1023 * 17.15 + 0.2062 = 1.96$

L2 : $0.1023 * 23.60 + 0.2062 = 2.62$

L3 : $0.1023 * 34.15 + 0.2062 = 3.70$

Leaf area (m^2)

L0 : 0.51 * 151 * 159 = 1.22

L0*: 0.51 * 180 * 207 = 1.90

L1 : 0.53 * 275 * 420 = 6.12

L2 : 0.55 * 340 * 515 = 9.63

L3 : 0.56 * 375 * 547 = 11.49

Table 9. Example of a recording schedule with mean records obtained from one progeny

Month from plant	Bunch yield	pcs(cm)²	Leaflet number	lxw(cm)	frond prod	height(cm)	trunk diam.(cm)
0							
6Lo[1]		6.78	151	159			
12Lo*		9.26	180	207			
18							
24							
30	start						
36							
42 L1	65	17.15	275	420			
48							
54	140						
60							
66 L2	230	23.6	340	515	56.2	152	
72							
78	185						
84							
90 L3	190	34.15	375	547	49.2		55
96							

Reference code of the youngest leaf marked at that date

4.4.4 Stem measurements

Height increment (cm/year)

The distance from the reference point (L1) to the insertion of leaf L2 is 152 cm.

The annual stem increment is thus 152/2 = 76 cm.

Trunk diameter (cm)

The diameter of the exposed stem is 55.2 cm.

4.4.5 Production of leaves

Leaf production between those marked at opening 42 and 66 months from planting (L1 and L2, respectively) is 56.2. Annual leaf production for this period is thus 56.2/2 =28.1.

In the same way the production between months 66 and 90 amounts to 49.2/2 =24.6.

4.5 Calculation of growth parameters

4.5.1 Bunch dry matter production (kg/palm/year)

Month 42 to month 66 (years 2 and 3 of production):(74.20 + 121.90)/2 = 99.38.

4.5.2 Vegetative dry matter production (kg/palm/year)

Leaf dry matter production

Months 42 (L1) to 66 (L2) =

(1.96 + 2.62)/2 * 56.2/2 = 64.35

Months 66 (L2) to 90 (L3) =

(2.62 + 3.70)/2 * 49.2/2 = 77.74.

Trunk dry matter production

Height increment and trunk diameter are assumed to be by and large stabilised at the time of measuring. For months 42 to 66 and for months 66 to 90 trunk dry matter is estimated as follows:

Volume increase:

$(55.2/2)^2 * 76 = 181.79$ dm^3

Weight per volume

$= 0.0076 * (90 - 42)/12 + 0.083$

$= 0.11 \text{ kg/dm}^3$

Trunk dry matter production:

$181.79 * 0.11 = 20.0.$

VDM

Months 42 to 66:

64.35 (leaf DM) + 20.0 (trunk DM) = 84.35

Months 66 to 90:

77.74 (leaf DM) + 20.0 (trunk DM) = 97.74.

4.5.3 Bunch Index

Months 42 to 66:

$98.05/(84.35 + 98.05) = 0.538$

Months 66 to 90:

$99.37/(97.74 + 99.37) = 0.504.$

4.5.4 Harvest Index

Months 42 to 66:

$(185.0 * 0.30)/(84.35 + 98.05) = 0.304$

Months 66 to 90:

$(187.5 * 0.30)/(97.74 + 99.37) = 0.285.$

4.5.5 Leaf area ratio (m^2 /kg)

Months 42 to 66:

$\{28.1 * (6.12 + 9.63)/2\}/84.35 = 2.62$

Months 66 to 90:

{24.6 * (9.63 + 11.49)/2}/97.74 = 2.66.

4.5.6 Parameters of the logistic growth curve

The curve is fitted through estimated leaf area (y) at the following months after planting (t); the data pairs (t,y) were as follows:

(6, 1.22); (12, 1.90); (42, 6.12); (66, 9.63) and (90, 11.49).

The logistic growth function $f(t) = A/(1 + B* e^{-Ct})$ has been fitted to these data by the Least Squares Method and gives A= 12.180772, B= 11.286396, C= 0.057470.

Rewriting the logistic growth function as follows:

$$f(t) = \frac{L_m}{(1 + \frac{L_m - L_i}{L_i})e^{-kt}}$$

with A= L_m, B= $(L_m - L_i)/ L_i$ and C= k,

where k = the relative rate of growth of the mean leaf area,

L_m = asymptotic maximum leaf area,

L_i = leaf area at field planting.

For selection purposes this is conveniently expressed as the time to reach 95% of the maximum leaf area ($t_{0.95}$) as follows:

$$t_{0.95} = -(\frac{1}{k})*\ln \frac{\frac{L_m}{0.95L_m} - 1}{\frac{L_m}{L_i} - 1}$$

This gives the following parameters for selection:

L_m = A =12.180772, L_i = A/(1+B) = 0.991403, k= C =0.057470.

Now 0.95*L_m = 11.571733 and hence

$$t_{0.95} = -\left(\frac{1}{0.057470}\right) * \ln\left\{\frac{\left(\frac{12.180772}{11.571733}\right) - 1}{\left(\frac{12.180772}{0.991403}\right) - 1}\right\}$$

= -17.400383 * (-5.368037) = 93.4059 months after planting.

References

Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist., 25, 16-39.

Bechhofer, R.E. and Dunnett, C.W. (1988). Percentage points of multivariate Student t distributions. In: Selected tables in mathematical statistics, vol. 11. Providence, American Mathematical Society, Rhode Island.

Breure, C.J. (1982). Factors affecting yield and growth of oil palm tenera in West New Britain. Oléagineux, 37, 213-227.

Breure, C.J. (1985). Relevant factors associated with crown expansion in oil palm (*Elaeis guineensis* Jacq.). Euphytica, 34, 161-175.

Breure, C.J. (1986). Parent selection for yield and bunch index in the oil palm in West New Britain. Euphytica, 35, 65-72.

Breure, C.J. (1987). Factors associated with the allocation of carbohydrates to bunch dry matter production in oil palm (*Elaeis guineensis* Jacq.). Ph. D. Dissertation, Wageningen Agricultural University, The Netherlands, pp. 259.

Breure, C.J. (1988). The effect of different planting densies on yield trends in oil palm. Experimental Agriculture, 24, 37-52.

Breure, C.J. (1994). Development of leaves in oil palm (*Elaeis guineensis*) and the determination of leaf opening rate. Experimental Agriculture, 30, 467-472.

Breure, C.J. and Bos, I. (1992). Development of elite families in oil palm (*Elaeis guineensis* Jacq.). Euphytica , 64, 99-112.

Breure, C.J. and Corley, R.H.V. (1983). Selection of oil palms for high density planting. Euphytica, 32, 177-186.

Breure, C.J. and Corley, R.H.V. (1992). Fruiting activity, growth and yield of oil palm. II. Observations in untreated populations. Experimental Agriculture, 28, 11-121.

Breure, C.J. and Konimor, J. (1992). Parent selection for oil palm clonal seed gardens. In: Proceedings of the 1990 ISPOB International Workshop on Yield Potential in the Oil Palm, 122-144. Malaysia, Kuala Lumpur, International Society of Oil Palm Breeders and Palm Oil Research Institute.

Breure, C.J., Konimor, J. and Rosenquist, E.A.R. (1982). Oil palm selection and seed production at Dami Oil Palm Research Station, Papua New Guinea. Oil Palm News, 26 , 6-22.

Breure, C.J. and Menendez, T. (1990). The determination of bunch yield components in the development of inflorescences in oil palm (*Elaeis guineensis* Jacq.). Experimental Agriculture , 26 , 99-115.

Breure, C. and Powell, M.S. (1987). The one-shot method of establishing growth parameters in oil palm. Proceedings of 1987 International Oil Palm Conference, Kuala Lumpur, Malaysia, 17, 203-209.

Breure, C.J., Rosenquist, E.A.R., Konimor, J. and Powell, M.S. (1987). Oil Palm introductions to Papua New Guinea and formulation of selection methods at the Dami Oil Palm Research Station. Proceedings of the international workshop on oil palm germplasm and utilisation. PORIM, Malaysia, 10 , 189-197.

Breure, C.J. and Soebagjo, F.X. (1991). Factors associated with occurence of crown disease in oil palm (*Elaeis guineensis* Jacq.) and its effect on growth and yield. Euphytica, 54, 55-64.

Butler, K.L. and Butler, D.G. (1987). Tables for selecting the best population. Queensland, Biometrical Bulletin 2, Department of Primary Industries, Queensland Government, Brisbane.

Cochran, W.G. and Cox, G.M. (1957). Experimental Designs: second edition. John Wiley & Sons, New York, London, Sydney.

Corley, R.H.V. and Donough, C.R. (1992). Potential yield of oil palm clones - The importance of planting density. In: Proceedings of the 1990 ISPOB International Workshop on Yield Potential in the Oil Palm, 58-71. Malaysia, Kuala Lumpur: International Society of Oil Palm Breeders and Palm Oil Research Institute.

Corley, R.H.V., Hardon, J.J. and Tan, G.Y. (1971). Analysis of growth of the oil palm (*Elaeis guineensis* Jacq.). I. Esimation of growth parameters and application in breeding. Euphytica, 20, 307-315.

Donald, C.M. (1968). The breeding of crop ideotypes. Euphytica, 17, 385-404.

Dourleijn, C.J. (1993). On statistical selection in plant breeding. Ph.D. Dissertation, Wageningen Agricultural University, The Netherlands, pp. 191.

Dourleijn, C.J. (1995). Subset selection in plant breeding practice. Euphytica, 81, 207-216.

Falconer, D.S. (1981). Introduction to quantitative genetics. 2nd Edition, Longman, London, pp. 340.

Gibbons, J.D., Olkin, I. and Sobel, M. (1977). Selecting and ordering populations: A new statistical methodology. John Wiley & Sons, New York.

Gupta, S.S. (1956). On a decision rule for a problem in ranking means. Ph. D. Dissertation (and Mimeograph Series No. 150), Institute of Statistics, University of North Carolina, Chapel Hill.

Gupta, S.S. (1963). Probability integrals of multivariate normal and multivariate t. Ann. Math. Statist., 34, 792-828.

Gupta, S.S. (1965). On some multiple decision (selection and ranking) rules. Technometrics, 7, 225-245.

Gupta, S.S., Nagel, K. and Panchapakesan, S. (1973). On the order statistics from equally correlated normal random variables. Biometrika, 60, 403-413.

Hallauer, A.R. and Miranda, J.B. (1981). Quantitative genetics in Maize breeding. Iowa State University Press, Ames, Iowa.

Hardon, J.J., Corley, R.H.V. and Ooi, S.C. (1972). Analysis of growth in oil palm. II. Estimation of genetic variances of growth parameters and yield of fruit bunches. Euphytica, 21, 257-264.

Hardon, J.J., Williams, C.N. and Watson, I. (1969). Leaf area and yield in the oil palm in Malaysia. Experimental Agriculture, 5, 25-52.

Hartley, C.W.S. (1988). The oil palm (*Elaeis guineensis* Jacq.), 3rd edition. Longman Group UK Limited, Harlow, England.

Kraalingen, D.W.G. van, Breure, C.J. and Spitters, C.J.T. (1989). Simulation of oil palm growth and yield. Agriculture and Forest Meteorology, 46, 227-244.

Laan, P. van der and Verdooren, L.R. (1989). Selection of Populations: An Overview and Some Recent Results. Biometrical Journal, 31, 383-420.

Laan, P. van der and Verdooren, L.R. (1990). A review with some applications of statistical selection procedures for selecting the best variety. Euphytica, 51, 67-75.

Le Guen, V., Samaritaan, G., Zarin Othman, A., Chin, C.W., Konan, K. and Durand-Gasselin, T. (1991). Oil production in young oil palm clones. Oléagineux, 46, 347-359.

Patterson, H.D., Williams, E.H., and Hunter, E.A. (1978). Block designs for variety trials. Journal of Agricultural Science, Cambridge, 90, 395-400.

Peaslee, D.E. and Moss, D.N. (1966). Photosynthesis in K- and Mg- deficient maize (Zea Mays, L.) leaves. Soil Sci. Soc. Amer. Proc., 30, 220-223.

Rasch. D. (1992). Logistic Growth Models and Related Problems. In: Handbook of the logistic distribution, 427-447, (Ed. N. Balakrishnan). Marcel Dekker, Inc., New York.

Searle, S.R., Casella, G. and McCullogh, C.E. (1992). Variance Components. John Wiley & Sons Inc., New York.

Shibles, R. (1993). Do the photochemical reactions limit photosynthesis? In: Manual Course of Crop Physiology. Ames, Iowa.

Squire, G.R. and Corley, R.H.V. (1987). Oil Palm. In: Tree Crop Physiology, 141-167(Eds M.R. Sethuraj and A.S. Raghavendra). Elsevier Science Publisher B.V. , Amsterdam.

Verdooren, L.R. (1988). Statistical inference on variance components. Ph. D. Dissertation, Wageningen Agricultural University, The Netherlands, pp. 218.

APPENDIX 1

Least Squares estimates for General Combining Ability (GCA)

The actual yield y_{ijk} of the k-th plot of a *tenera* offspring of the cross $D_i x P_j$ is a random sample of the population of all possible observations from this cross with population mean or expectation $E(y_{ijk})$ and variance s^2; hence, the statistical model is $y_{ijk} = E(y_{ijk}) + e_{ijk}$, where e_{ijk} is the effect of the environment or error on this k-th plot. These error-terms e_{ijk} are such that the expectation

$E(e_{ijk}) = 0$ and the variance$Var(e_{ijk}) = s^2$;

these errors are uncorrelated with one another because we have randomized the plots over the crosses. When one uses a randomization procedure to allot the plots of a field to the crosses, as with a completely randomized design (CRD), then the plot-errors can be assumed to be uncorrelated.

With such a model for the yields, the Least Squares Method searches estimates m, a_i, and b_j for the parameters m, a_i and b_j respectively, such that the sum of the squared deviations between the observation and the estimate of their expected value for k=1,...,n_{ij}, i=1,...,A, and j=1,...,B,

$S_{iS jS k} [y_{ijk} - (m + a_i + b_j)]^2$ is minimal.

In statistical text books these Least Squares estimates for the parameters are found as solutions of the so called Normal Equations.

$S_{iS jS k} y_{ijk} = y...$, $S_{jS k} y_{ijk} = y_i..$, $S_{iS k} y_{ijk} = y.j.$,

$S_{iS_j} n_{ij} = n.. , S_j n_{ij} = n_i. , S_i n_{ij} = n.j$.

The Normal Equations are then:

$n.. * m + S_i n_i. * a_i + S_j n.j * b_j = y... $ (1)

$n_i. * m + n_i. * a_i + S_j n_{ij} * b_j = y_i..$

for i=1,...,A (2)

$n.j * m + S_i n_{ij} * a_i + n.j * b_j = y.j.$

for j=1,...,B (3)

Note that these equations are not independent. Equation (1) is equal to the sum of the equations of (2); also equation (1) is equal to the sum of the equations of (3). Hence there are two linear dependencies between the Normal Equations. From this follows that the Normal Equations are not uniquely solvable. For a solution of the Normal Equations one can choose freely one value for an a_i and one value for a b_j . The statistical package SAS chooses $a_A = 0$ and $b_B = 0$. The statistical package SPSS chooses in the ANOVA procedure a solution such that $S_i n_i. * a_i = 0$ and $S_j n.j * b_j = 0$; however SPSS chooses in the MANOVA procedure a solution such that $S_i a_i = 0$ and $S_j b_j = 0$. Nevertheless whichever solution of the Normal Equations is chosen, the differences between the *dura*-parameters a_i-a_h are estimated by a_i-a_h for i, h=1,...,A and the differences between the *pisifera*-parameter b_j-b_k are estimated by b_j-b_k for j,k=1,...,B; these estimates are always the same, irrespective of the solution one has chosen. The differences between the *dura*-parameters and the *pisifera*-parameters are therefore called estimable.

The so-called Least Squares Mean for a *dura* D_i, LSM(D_i), is $m+a_i+S_j b_j /B$ and this is the estimate for $m + a_i + S_j b_j /B$; also the Least Squares Mean for a *pisifera* P_j, LSM(P_j), is $m + S_i a_i /A + b_j$ and this is the estimate for $m + S_i a_i /A + b_j$. These Least Squares Means are also uniquely estimated, irrespective of the solution one has chosen from the Normal Equations. These Least Squares Means are also estimable. Note that the difference in effect between two *dura*, $a_i - a_h$, is estimated as LSM(D_i) - LSM(D_h) = $a_i - a_h$; analogously the difference in effect between two *pisifera*, $b_j - b_k$, is estimated as LSM(P_j) - LSM(P_k) = $b_j - b_k$. We can therefore rank all the *dura* and the *pisifera* according to their General Combining Ability estimates of the parameters a_i and b_j or their Least Squares Means LSM(D_i) and LSM(P_j).

To estimate the variance s^2 we must first calculate the sum of squares of the residuals according to this additive model, SS(res-A). This is :

$SS(res-A) = S_{iS_jS_k} y_{ijk}^2 - [m*y... + S_i a_i * y_i.. + S_j b_j * y.j.]$.

This SS(res-A) has df(res-A) = n.. - (A + B -1) degrees of freedom, where n.. is the total number of plots and A= number of *dura* and B= number of *pisifera* in the connected crossing design. The estimate of the variance s^2 is

s^2 = SS(res-A)/df(res-A)

= SS(res-A)/[n.. -(A+B-1)].

When we may assume that the errors (and hence the yields) are Normally distributed, we can construct for example 95% confidence intervals for the difference between the GCA values of the *dura* or *pisifera*.

The above mentioned procedure is illustrated in the following Example 2 (see Appendix 2), where C=9 progenies derived from A=5 *dura* and B=3 *pisifera* are tested in a completely randomized design with two plots per progeny. It is assumed that the genetic effects of the *dura* and *pisifera* parents are additive.

Appendix 2

Analysis of Example 2

EXAMPLE 2.

Assume that C=9 progenies (2 plots each), from A=5 *dura* and B=3 *pisifera*, are tested in a completely randomized design. We assume an additive model for the genetic effects of the *dura* and *pisifera* parents. Yield records (kg/plot) were as follows:

| | | *Pisifera* | | | | | | |
		P1		P2		P3		Total
Dura	D1	44	48					92
	D2	45	42	45	43			175
	D3	33	36	35	32	36	38	210
	D4			44	42	46	48	180
	D5					53	55	108
Total		248		241		276		765

This crossing design is connected because there is one continuous chain which connects all the crosses. The Normal Equations are:

18*m+2*a_1+4*a_2+6*a_3+4*a_4+2*a_5+6*b_1+6*b_2+6*b_3= 765 (1)

2*m+2*a_1+0*a_2+0*a_3+0*a_4+0*a_5+2*b_1+0*b_2+0*b_3= 92

4*m+0*a_1+4*a_2+0*a_3+0*a_4+0*a_5+2*b_1+2*b_2+0*b_3= 175

$6*m+0*a_1+0*a_2+6*a_3+0*a_4+0*a_5+2*b_1+2*b_2+2*b_3= 210$ (2)

$4*m+0*a_1+0*a_2+0*a_3+4*a_4+0*a_5+0*b_1+2*b_2+2*b_3= 180$

$2*m+0*a_1+0*a_2+0*a_3+0*a_4+2*a_5+0*b_1+0*b_2+2*b_3= 108$

$6*m+2*a_1+2*a_2+2*a_3+0*a_4+0*a_5+6*b_1+0*b_2+0*b_3= 248$

$6*m+0*a_1+2*a_2+2*a_3+2*a_4+0*a_5+0*b_1+6*b_2+0*b_3= 241$ (3)

$6*m+0*a_1+0*a_2+2*a_3+2*a_4+2*a_5+0*b_1+0*b_2+6*b_3= 276$

In matrix notation these Normal Equations can be written as M * p= t, where M is the 9x9 matrix of the coefficients in the Normal Equations, p is the column or 9x1matrix of parameters and t is the column or 9x1 matrix of the totals in the right-hand side of the Normal Equations. Hence

$$
\mathbf{M}=
\begin{vmatrix}
18 & 2 & 4 & 6 & 4 & 2 & 6 & 6 & 6 \\
2 & 2 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\
4 & 0 & 4 & 0 & 0 & 0 & 2 & 2 & 0 \\
6 & 0 & 0 & 6 & 0 & 0 & 2 & 2 & 2 \\
4 & 0 & 0 & 0 & 4 & 0 & 0 & 2 & 2 \\
2 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 2 \\
6 & 2 & 2 & 2 & 0 & 0 & 6 & 0 & 0 \\
6 & 0 & 2 & 2 & 2 & 0 & 0 & 6 & 0 \\
6 & 0 & 0 & 2 & 2 & 2 & 0 & 0 & 6
\end{vmatrix}
\quad
\mathbf{p}=
\begin{vmatrix}
m \\ a1 \\ a2 \\ a3 \\ a4 \\ a5 \\ b1 \\ b2 \\ b3
\end{vmatrix}
\quad
\mathbf{t}=
\begin{vmatrix}
765 \\ 92 \\ 175 \\ 210 \\ 180 \\ 108 \\ 248 \\ 241 \\ 276
\end{vmatrix}
$$

A solution of these Normal Equations with $a_5=0$ and $b_3=0$ (because we have two linear dependencies between the Normal Equations) is given by p = M⁻* t, where M⁻ is a generalized inverse of M, with the property M* M⁻*M=M,

$$
\mathbf{M^-}=(1/30)
\begin{vmatrix}
15 & -15 & -15 & -15 & -15 & 0 & 0 & 0 & 0 \\
-15 & 50 & 30 & 25 & 20 & 0 & -20 & -10 & 0 \\
-15 & 30 & 36 & 24 & 21 & 0 & -15 & -12 & 0 \\
-15 & 25 & 24 & 26 & 19 & 0 & -10 & -8 & 0 \\
-15 & 20 & 21 & 19 & 26 & 0 & -5 & -7 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -20 & -15 & -10 & -5 & 0 & 20 & 10 & 0 \\
0 & -10 & -12 & -8 & -7 & 0 & 10 & 14 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{vmatrix}
\quad \mathbf{p}=(1/30)
\begin{vmatrix}
1620 \\ -145 \\ -207 \\ -503 \\ -217 \\ 0 \\ -95 \\ -106 \\ 0
\end{vmatrix}
=
\begin{vmatrix}
54 \\ -4.83 \\ -6.9 \\ -16.8 \\ -7.23 \\ 0 \\ -3.17 \\ -3.53 \\ 0
\end{vmatrix}
$$

An estimable linear combination of the parameters

$c_1*m + c_2*a_1 + c_3*a_2 + c_4*a_3 + c_5*a_4 + c_6*a_5 + c_7*b_1 + c_8*b_2 + c_9*b_3$

is unbiasedly estimated with minimum variance by c'*p where

$c' = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9)$ and the variance of c'*p is given by $c'*M^-*c*s^2$.

The common variance s^2 is estimated by $s^2 = SS(res-A)/df(res-A)$, where the residual sum of squares for the additive model is

SS(res-A) is $y'*y - p'*t =$

$S_{iS jS k}\, y_{ijk}^2 - p'*t = S_{iS jS k}\, y_{ijk}^2 - [m*y... + S_i\, a_i * y_{i..} + S_j\, b_j * y_{.j.}]$

and df(res-A) is the residual SS(res-A) degrees of freedom $= n.. -(A+B-1)$.

In our Example 2 we have SS(res-A) $= 33231-33197.96667 = 33.03333$ with $18-(5+3-1)= 11$ degrees of freedom, hence $s^2 = 33.03333/11 = 3.00303$.

The difference between two *dura*-effects, for example D_1 and D_2 is the linear combination of the parameters a_1-a_2 and which is estimated by $a_1-a_2 = C'*p = (0,1,-1,0,0,0,0,0,0)*p = 2.06667$ with an estimated variance $c'*M^-*c*s^2 = [(50 + 36 - 2*30)/30]*3.00303 = 2.602626$ and an estimated standard error

$\sqrt{(2.602626)} = 1.61327$.

The difference between two *pisifera*-effects, for example P_2 and P_3 is b_2-b_3 and which is estimated by $b_2-b_3 = c'*p = (0,0,0,0,0,0,0,1,-1)*p = -0.26667$ with an estimated variance $c'*M^-*c*s^2 = [(14 + 0 - 2*0)/30]*3.00303 = 1.401414$ and an estimated standard error

$\sqrt{(1.4014)} = 1.183813$

Appendix 3

Specific Combining Ability (SCA) analysis

Sometimes the additive model of genetic effects of the parents does not fully explain the performance of their offspring. This is attributable to an interaction effect of the genetic effects of the parents. In other words, besides the additive genetic effects (General Combining Ability) of the parents there is also a specific interaction effect due to the specific combination of the parents. This specific interaction effect is called in quantitative genetics Specific Combining

Ability (SCA). For this interaction model the expected yield of the *tenera* offspring of the crossing $D_i x P_j$, $E(y_{ij})$, can then be written as the sum of a general constant, m^*, the GCA effect a_i^* of the *dura* mother D_i, the GCA effect b_j^* of the *pisifera* father and the SCA effect $(a\ b\ )_{ij}^*$ of the realized cross:

$$E(y_{ij}) = m^* + a_i^* + b_j^* + (a\ b\ )_{ij}^* = m_{ij}.$$

When we have a set of C crosses, derived from A *dura* and B *pisifera*, where

C£ A*B, the C parameters m_{ij} can be estimated using the Least Squares Method.

Assume that there are n_{ij} plots available for a certain *tenera* cross $D_i x P_j$; in the case that there no cross has been made then $n_{ij} = 0$. We consider here the case that we have used a completely randomized design (CRD). In section 2.5 we will consider the case that we have used an (in)complete block design.

The actual yield y_{ijk} of the k-th plot of a *tenera* offspring of the cross $D_i x P_j$ is $y_{ijk} = E(y_{ijk}) + e_{ijk}$, where e_{ijk} is the effect of the environment or error on this k-th plot. These errors e_{ijk} are such that the expectation $E(e_{ijk}) = 0$ and the variance $Var\ (e_{ijk}) = s^2$; these errors are uncorrelated with one another. When one uses a randomization procedure to allot the plots of a field to the crosses, such as in a completely randomized design (CRD), then the plot-errors can be assumed to be uncorrelated.

The Least Squares Method searches estimates m_{ij} for the parameters mij such that the sum of the squared deviations between the observation and the estimate of their expected value for $k=1,...,n_{ij}$, $i=1,...,A$, and $j=1,...,B$,

$$S_{iS\ jS\ k}\ [\ y_{ijk} - m_{ij}\ ]^2 \text{ is minimal.}$$

The Least Squares estimates mij for the parameters m_{ij} are found as solutions of the Normal Equations, which are in this case very easy. Let us denote the sum of the observations of the n_{ij} plots of the cross $D_i x P_j$ by y_{ij}. , hence

$S_k\ y_{ijk} = y_{ij}.$. The Normal Equations are then: $n_{ij}^*\ m_{ij} = y_{ij}.$ (4)

for $i=1,...,A$ and $j=1,...,B$. There are only C Normal Equations present, because if a certain offspring $D_i x P_j$ has not been realized, then $n_{ij} = 0$ for such a progeny and we have no observations of this progeny. The parameter estimates are then $m_{ij} = y_{ij}.\ /\ n_{ij}$, the progeny-means of the crosses $D_i x P_j$. To estimate the Specific Combining Abilities of these progenies we must now calculate the estimates m for m , a_i for a_i and b_j for b_j for the parameters according to an additive model

$$E(y_{ijk}) = m + a_i + b_j, \text{ as has been explained in section 2.2.1 .}$$

The estimate for the Specific Combining Ability $(a\ b\ )_{ij}^*$ is $(ab)_{ij}^* = m_{ij} - (m + a_i + b_j)$.

To estimate the variance s^2 we must calculate the sum of squares of the residual according to this interaction model, SS(res-I), as follows:

$$SS(res-I) = S_{iS \, jS \, k} \, y_{ijk}^2 - S_{iS \, j} \, (y_{ij.}^2)/n_{ij} \, .$$

This SS(res-I) is based on df(res-I)= n.. - C degrees of freedom, where n.. is the total number of plots and C is the number of realized crosses in the CRD. The estimate s^2 for s^2 is SS(res-I)/(n.. - C) . Assuming that the errors (and hence the yields) are Normally distributed, we can test the null-hypothesis "The Specific Combining Abilities are equal", otherwise stated; an additive model for the GCA values is reasonable. For this test we need to calculate the sum of squares of the residual according to the additive model, SS(res-A), as has been explained in section 2.2.1 .

This $SS(res-A) = S_{iS \, jS \, k} \, y_{ijk}^2 - [\, m*y...+S_i \, a_i * y_{i..} +S_j \, b_j * y_{.j.} \,]$

and has df (res-A)=n.. - (A + B -1) degrees of freedom, where n.. is the total number of plots and A= number of dura and B= number of *pisifera* in the connected crossing design. The test-statisitic is

$$F = \frac{\dfrac{SS(res-A) - SS(res-I)}{df(res-A) - df(res-I)}}{\dfrac{SS(res-I)}{df(res-i)}}$$

and if the null-hypothesis of equal Specific Combining Abilities is true, then this test-statistic has an F-distribution with degrees of freedom {df(res-A)-df(res-I)} and df(res-I). In a table of the F-distribution one can find the right-significance point with significance level a % , F(a), and if F >F(a) one can reject the null-hypothesis of equal Specific Combining Abilities.

Another way of calculating SS(res-A)-SS(res-I), which is the sum of squares according to the SCA values, SS(SCA), is to calculate the sum of the squared SCA values. The degrees of freedom df(SCA) for this SS(SCA), is

df(SCA) = C-(A+B-1).

This procedure is illustrated in the following Example 3, where C=9 progenies derived from A=5 *dura* and B=3 *pisifera* are tested in a completely randomized design with two plots per progeny. The observations are the same as considered in Example 2, but now we will also consider the possibility of Specific Combining Ability of the parents.

EXAMPLE 3.

Assume C=9 progenies (2 plots each), from A=5 *dura* and B=3 *pisifera*, were tested in a completely randomized design. We assume an interaction model for the genetic effects of the *dura* and *pisifera* parents. The yields in kg per plot were as follows:

		P1	P2	P3				Total

Pisifera

		P1	P2	P3	Total

Dura	D1	44 48			92
	D2	45 42	45 43		175
	D3	33 36	35 32	36 38	210
	D4		44 42	46 48	180
	D5			53 55	108

Total	248	241	276	765

The estimate for m $_{11}$ is $m_{11} = (44+48)/2 = 46$, for m 21 is $m_{21} = (45+42)/2 = 43.5$, etc.

The Specific Combining Ability (SCA) for D_1xP_1 is estimated as $m_{11}-(m+a_1+b_1)= 46-[54+(-4.83333)+(-3.16667)]=0.00000$; the SCA for D_2xP_1 is estimated as $m_{21} -(m+a_2+b_1)=43.5-[54+(-6.9)+(-3.16667)]=-0.43333$, etc. See Example 2 for the estimates of m, a_i and b_j. The table of Specific Combining Abilities estimates is then (every entry must be given twice, because we have two plots for each cross):

pisifera

	P1	P2	P3

dura 1	0		
2	-0.4333	0.43333	
3	0.43333	-0.2	-0.2333
4		-0.2333	0.23333
5		0	

The residual sum of squares according to this interaction model is

SS(res-I)= 33231 - [46*(44+48) + 43.5*(45 +42)+...+54*(53+55)]=31.50 with

df(res-I)= 18-9 =9 and hence

$s^2 = 31.50/9 = 3.50$.

We assume now that the errors (and hence the yields) are Normally distributed.

For the calculation of the test statistic F to test the null-hypothesis of equal Specific Combining Abilities we get from example 1 the SS(res-A) = 33.03333 with df(res-A)= 11.

Hence F=[{33.03333-31.50}/(11-9)]/[31.50/9] = 0.219 and the right-sided 5% significance value of the F-distribution with (11-9=2) and 9 degrees of freedom is F(5%)= 4.26 and because F=0.219 <F(5%)=4.26 we cannot reject the null-hypothesis. Hence an additive model for the genetical effects of the parents is reasonable and we can use the results of Example 2 (see <u>Appendix 2</u>) to estimate the General Combining Abilities of the *dura* and *pisifera* parents.

Note that we can also calculate SS(res-I)-SS(res-A)= 33.03333 -31.50 = 1.53333 as SS(SCA) = sum of the squared SCA values. Because there are two plots for each progeny, we have as sum of the squared SCA values of the above given table:

SS(SCA)= $2*[2* 0.00000^2 + 3*0.43333^2 + 0.20000^2 + 3*0.23333^2]$

= 1.53331 . The df(SCA)

= C-(A+B-1) = 9-(5+3-1)=2.

Appendix 4

Analysis of an incomplete block design

The model for an incomplete block design with C progenies and NB incomplete blocks is such that the expected yield $E(y_{gh})$ of a *tenera* offspring T_g(g=1,...,C) of a dura mother D_i (i=1,...,A) and a *pisifera* father P_j (j=1,...,B), which is allotted to a plot in an incomplete block Bl_h (h=1,...,NB), can be described as the sum of a general constant f , an effect t_g of the *tenera* T_g and an effect d_h of the block Bl_h, hence

$E(y_{gh}) = f + t_g + d_h$

for g=1,...,C and h=1,...,NB .

The yield y_{gh} of the progeny T_g in the block Bl_h can be described as $y_{gh} = E(y_{gh}) + e_{gh}$, where e_{gh} is the environmental effect or plot error with expectation $E(e_{gh}) = 0$ and variance $Var(e_{gh}) = s^2$, these errors are uncorrelated. Because we have allotted the plots of a block at random to the progenies, which must be tested in this block according to the design, this assumption of uncorrelated errors is reasonable.

The model described for y_{gh} is an additive model of the *tenera* effects and the block effects. In section 2.2.1 we have already described how the parameters of an additive model can be estimated with the Least Squares Method. To estimate these parameters we must solve the so-called Normal Equations. Let the Least Squares estimates be denoted by f for f , t_g for t_g and d_h for d_h. Let further n_{gh} be 1 if progeny T_g is present in block Bl_h and n_{gh} be 0 if progeny T_g is not present in block Bl_h.

The Normal Equations are then:

$n.. * f + S_g \, n_{g.} * t_g + S_h \, n._h * d_h = y..$ (1)

$n_{g.} * f + n_{g.} * t_g + S_h \, n_{gh} * d_h = y_{g.}$

for g=1,..., (2)

$n._h * f + S_g \, n_{gh} * t_g + n._h * d_h = y._h$

for h=1,...,NB (3)

Note that these equations are not independent. Equation (1) is equal to the sum of the equations of (2); also, equation (1) is equal to the sum of the equations of (3). Hence there are two linear dependencies between the Normal Equations. We will choose as a solution of the Normal Equations that solution where we take $t_C=0$ and $d_{NB}=0$.

The Least Squares Mean of an offspring T_g is defined as

$f + t_g + S_h \, d_h$ /NB, and this is the same for every solution of the Normal Equations.

The estimate for the variance s^2 is $s^2 = $ SS(res)/df(res) , where the residual sum of squares SS(res) is calculated as, SS(res) = $S_{gSh} \, y_{gh}^2$ -[$f*y.. + S_g \, t_g * y_{g.} + S_h \, d_h * y._h$], with degrees of freedom

df(res) = n.. - [C + NB -1].