

Use of Alpha-Designs in Oil Palm Breeding Trials

Léon Rob Verdooren

Danone Nutricia Research, Utrecht, The Netherlands

Email address:

l.r.verdooren@hetnet.nl

To cite this article:

Léon Rob Verdooren. Use of Alpha-Designs in Oil Palm Breeding Trials. *American Journal of Theoretical and Applied Statistics*. Vol. 8, No. 4, 2019, pp. 136-143. doi: 10.11648/j.ajtas.20190804.12

Received: June 26, 2019; **Accepted:** August 1, 2019; **Published:** August 14, 2019

Abstract: In Oil Palm Breeding trials the plots have palms at vertices of equilateral triangles with side length of 9 m. The plots consist of $6 \times 6 = 36$ palms, hence a plot is a rectangle of 46.8 x 54m. The number of tested varieties is 20 – 40, the experimental design needed is an incomplete block design, with usually 3 replications; the alpha-designs can give a connected incomplete block design. Current Oil Palm planting materials are DxP hybrid based on crossing selected *dura* palms (female parents) with *pisifera* palms (male parents) to produce *tenera* palms with thin shelled fruits. The crossing scheme of *A dura* and *B pisifera* is an incomplete diallel if the number of crossings C is smaller than $A \times B$. To make a connected crossing scheme the alpha-design can be used. In the analysis of an oil palm breeding trial an additive model of the *dura* and *pisifera* effects is applied to estimate the general combining ability of the parents after removing the fixed replication effect and the random blocks within the replication effects. The analysis can be done with the package SAS or IBM SPSS Statistics with program Mixed; further with R and the R package lme4.

Keywords: Incomplete Block Designs, Alpha-designs, Incomplete Diallel Crossing Scheme,
Use Alpha-designs for a Connected Crossing Scheme,
Use SAS or IBM SPSS Statistics with Program Mixed for Analysis

1. Variety Trials of Field Crops

In 1966 Poland wished to reorganize the variety testing of field crops. The director of the in 1966 new founded Polish “Centre for Research on Varieties of Agricultural Crops (COBORU = Centralny Ośrodek Badania Odmian Roslin Uprawnych)” in Słupia Wielka, Dr. Eugeniusz Bilski, invited J. Hogen Esch, the deputy director of the Netherlands State “Institute for Research on Varieties of Field Crops” in Wageningen (who was a specialist on potato varieties) and the head of the Statistical Department of that Institute, Rob Verdooren, to advise COBORU with the set-up of variety testing and the design of variety trials to cope with the climatic regions in Poland. Early July 1967 we both went to Słupia Wielka. The statistical advisor of COBORU, the Statistical Lecturer of Academia Rolnicza (Agricultural University) in Poznań, Dr. Tadeusz Caliński visited with me some variety trials in the surrounding area. Meanwhile we were discussing the set up of variety testing. My advice to COBORU was to apply Incomplete Block designs in the variety trials, because the number of varieties was too large

to lay down in homogeneous complete blocks, and further to use not more than three replications per experimental site but instead to go for more experimental sites. This was in contrast to the practice of variety-trial designs of the “Bundes-Sortenamt” of West Germany (the West German state variety trial organization). The combination of the results of variety trials in climate zones of Poland turned also out to be large incomplete two-way layouts as well.

But variety trials in incomplete two-way layouts were at that time difficult to analyze by hand. The Best Linear Unbiased Estimators (BLUE) for varietal contrasts were then very cumbersome to calculate, by using matrix inversion for the solution of the normal equations. There were no mainframe computers available. But in The Netherlands at the Agricultural University of Wageningen we had done this BLUE analysis already by hand with an iterative method using mechanical electrical calculators. Back in 1948 W. L. Stevens [14] published an iterative solution to get the BLUE for incomplete two-way layouts, but he did not prove that this iterative procedure was converging and always gave a solution. The Mathematics Professor Dr. Nico H. Kuiper of the Agricultural University of Wageningen produced evidence in

1952 for the incomplete two-way layouts by the projection of vectors [7]. His first staff member, Leo Corsten, obtained his PhD degree of the Agricultural University of Wageningen in 1958, with supervisor Prof. Dr. N. H. Kuiper, in which he elaborated the iterative method also for incomplete three-way classification designs [4]. In the iterative Kuiper-Corsten method for the analysis of incomplete blocks, however, there was a missing link, namely for the calculation of the standard error of varietal contrasts.

During my visit in Poland I gave a lecture "The use of incomplete block designs in agricultural research and its analysis" at the Statistical Department of the Agricultural University (Academia Rolnicza) in Poznań. I demonstrated how to find the BLUE by hand for a small variety-trial example with the iterative Kuiper-Corsten method and mentioned that in the analysis of incomplete block designs with the iterative procedure, there was a missing link namely for the calculation of the standard error of varietal contrasts. Later at the 7th International Biometric Conference in Hannover in August 1970 Dr. Tadeusz Caliński gave in his lecture the solution how the Standard Error of varietal contrasts could be derived with the same Kuiper-Corsten iterative procedure [2].

At the first Working Seminar on Statistical Methods in Variety Testing in 1976 Dr. Eugeniusz Bilski of COBORU invited Prof. Dr. Leo Corsten and Rob Verdooren from Wageningen and Prof. Dr. Desmond Patterson and Mike Talbot from Edinburgh (Scotland). Then Prof. Dr. Desmond Patterson showed the work by him and his PhD student Emlyn Williams about the extension of the Incomplete Block designs with *alpha*-designs. These *alpha*-designs are very useful for Variety Testing because they make it easier to find designs for a large number of varieties and different (even small) sizes of incomplete blocks. COBORU adopted these designs immediately.

In variety testing we have the problem that we often have a large number C of varieties to be compared in just 3 or 4 replications. A Randomized Complete Block Design, with block size $k = C$, can often not be used if $C > 10$. To take care of the heterogeneous growing conditions in an experimental field one can use a Randomized Incomplete Block Design. With a smaller block size $k < C$ we can find homogeneous parts of the experimental field; in that case an *incomplete* block design is used. The well known incomplete block designs, such as balanced incomplete block designs often require too much replications. Initially, one used the incomplete block designs such as lattices (for the case $C = k \times k$) and rectangular lattices (for the case $C = k \times (k+1)$); see [3]. Often the number of tested progenies C does not fit into lattices or rectangular lattices. An extension of the incomplete block designs is given in the studies [9, 10]. They introduced for binary connected incomplete block designs the so-called *alpha*-designs. They start with a rectangular array with column lengths of k (the size of the incomplete blocks) of the sequence of 1, ..., C varieties and shift the columns according to an array. For many combinations of progenies C and block sizes k they give a procedure to construct *alpha*-

designs. (The name of *alpha* comes from the first letter of array in the Greek alphabet). There is now a computer program CycDesign [5] available to generate incomplete block designs as *alpha*-designs and cyclic designs.

In variety testing trials one wants to use *resolvable* incomplete block designs where the design can be divided into r groups (= *replications*) such that each group contains each of the C crosses exactly once. The resolvable incomplete block designs, and particularly the so-called generalized lattice (GL) or *alpha*-designs, have become most suitable for crop variety trials [11, 13]. The program CycDesign can give such resolvable incomplete block designs. All these above-mentioned designs are *connected*. In a connected incomplete block design one can estimate all differences between the varieties.

Now with the Personal Computers and statistical packages as SAS, IBM-SPSS Statistics or R, the Kuiper-Corsten iterative method is not needed any more and we solve the varietal effects with the Least Squares Method using the matrix solution of the normal equations with $b = (X'X)^{-1} X'y$ and the variance of a varietal contrast $p'b$ is found as $\sigma^2 p'(X'X)^{-1} p$.

2. Oil Palm Breeding Trials

In July 1987 I was sent by Harrisons Fleming Advisory Services (HFAS) for two weeks to the Oil Palm Research Station (OPRS) at Dami, on the island of New Britain north of Papua New Guinea. There was a breeding trial with oil palm (*Elaeis guineensis* Jacquin) done with 40 DxP hybrids, crosses of selected *dura* palms, as female parents, with *pisifera* palms, as male parents, to produce *tenera* palms with thin-shelled fruits. Statistical analysis of yield and its component of oil did not show significant differences between the crosses. The experimental design was a randomized complete block design with three replications and 6 x 6 palms per plot. The palms were planted at the corners of equilateral triangles with side lengths of 9 m. The distance between the rows is then equal to the height of this triangle = $9 \times \sin(\pi/3) = 9 \times (1/2) \times \sqrt{3} = 7.80$ m. Hence the length of a plot is $6 \times 7.80 = 46.80$ m and the width is $6 \times 9 = 54$ m. For the yielding capacity of a cross the yield of the $4 \times 4 = 16$ inner palms is recorded.

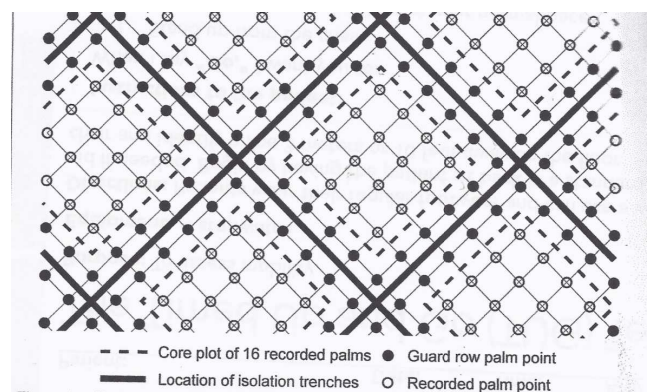


Figure 1. Part of an experimental layout using diamond-shaped plots.



Figure 2. An oil palm breeding trial.

With the English agronomist I walked about two kilometers over the trial to the end of a replication and back again through the other replications over the hilly land. After the visit of the trial I asked him what his impression was about the experimental field. He said the plots were well maintained. I agreed with him, but I explained to him that in the experimental design a block is supposed to have homogeneous growing conditions. In England with their centuries of agricultural history on flat land, a homogeneous part along the ditch may have uniform growing conditions suitable for a block in the field trial. Another parallel part of the land along this first block has also the same growing circumstances but with a different water level. Therefore in English books of Design of Experiments there were always figures with rectangular blocks. Most of the agronomists in tropical countries were taught about Experimental Designs with those English books, and applied, therefore, rectangular blocks even on freshly cleared forests after removing the tall trees by excavating or dynamiting the large trees. Neighboring plots could have different soil structure. In the oil palm trial at Dami the plots on hill tops will have different growing conditions than those on the slopes and in the valleys. These varying conditions preclude the required level of chance of finding significant difference between the crosses. The trial has therefore to be laid down with incomplete blocks.

Back in the office, I made a post-mortem analysis by using incomplete blocks of 8 adjacent plots, hence a replication consists of 5 incomplete blocks. I had my own FORTRAN program with me to analyze incomplete two-way classifications with the Kuiper-Corsten iteration method which could be run at the PC in the office. In the rows of the two-way classification system were the 40 DxP crosses and in the columns the $3 \times 5 = 15$ incomplete blocks. Now the analysis showed statistically highly significant differences between the crosses.

The HFAS company had send a colleague of me to set up a data base of all the oil palm crossings at the OPRS in Dami and further me to solve the problems of the breeding trials with oil palm. They assumed that we would require each two weeks to complete our assignment. My colleague needed these two weeks, but I had finished my report and explaining lecture to the staff about the main goals of the breeding trials

after one week. Because I had nothing else to do I asked the OPRS oil palm breeder about the main goal of the breeding program. He confirmed that this was to find the best *dura* \times *pisifera* combinations to maximize oil yields of the resulting *tenera* palms. I solved this problem in the second week of my stay.

First of all I would like to give some background on oil palm breeding. The main economic product is palm oil extracted from the mesocarp, i.e. the fruit flesh surrounding the pit (or stone) of the oil palm fruit. 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 set fruit, so the *pisifera* is not grown for commercial use. The other homozygote, the *dura*, has a thick shell. The heterozygote of the *dura* \times *pisifera* cross, the *tenera*, has a thin shell. The *tenera* is the fruit form preferred for commercial use, because of its larger proportion of oil-bearing mesocarp.

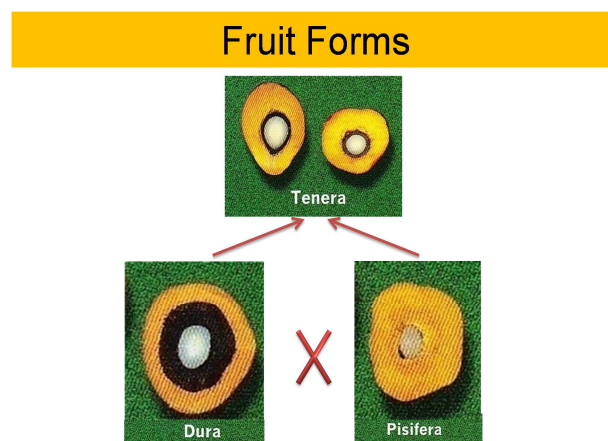
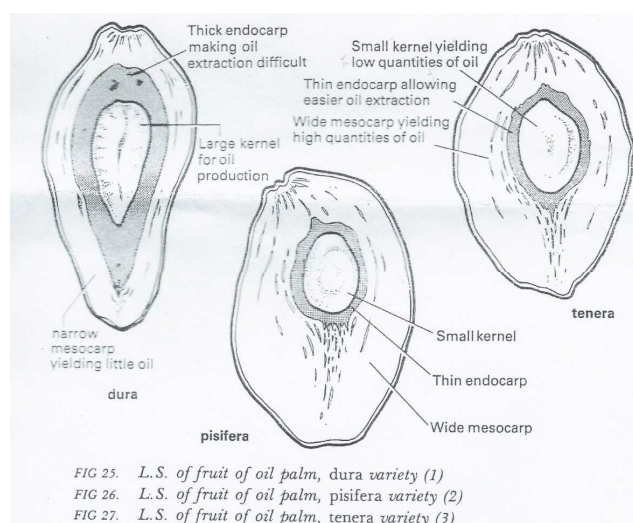


Figure 3. Fruit forms of *tenera*, *dura* (female parent) and *pisifera* (male parent).



(Note: L.S. = Longitudinal Section; *pisifera* fruits have no shell)

Figure 4. Fruit flesh surrounding the pit of the oil palm fruit in *dura*, *pisifera* and *tenera*.

As *pisifera* are predominantly female sterile, i.e. early abortion of fruit bunches, the *dura* is used as the female and the *pisifera* as the male parent (male inflorescences produce fertile pollen) 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. Fortunately, as in other crops, 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*×*pisifera* crosses. The additive model, however, does 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 this SCA effect is usually much smaller than the effect of GCA. A model for the expected yield of the *tenera* offspring ($D_i \times P_j$) is then, $E(y_{ij}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$ where γ_{ij} is the SCA of the cross $D_i \times P_j$.

To exploit both GCA and SCA effects the parents must be crossed; If we have A *dura* and B *pisifera* and all A×B *dura*×*pisifera* are realized, then we have a complete diallel crossing scheme; if there are only made C *dura*×*pisifera* crosses where $C < A \times B$, then we have a partial diallel or incomplete diallel crossing scheme. 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 scheme 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 two-way table given in Table 1.

Table 1. Realized 8 crosses of an incomplete diallel of 4 *dura* and 4 *pisifera*.

		<i>pisifera</i>			
		P ₁	P ₂	P ₃	P ₄
<i>dura</i>	D ₁	*		*	
	D ₂		*		*
	D ₃	*		*	
	D ₄		*		*

From the cross of *dura* D_1 with *pisifera* P_1 , $D_1 \times P_1$, we can make a chain to the cross of *dura* D_3 with P_1 , $D_3 \times P_1$; from $D_3 \times P_1$ we can go to the cross $D_3 \times P_3$, and from this cross $D_3 \times P_3$ we can go to the cross $D_1 \times P_3$, and then we come back

to the cross $D_1 \times P_1$. In this chain we have missed *dura* D_2 and D_4 . Hence this crossing scheme is disconnected. When we have rearranged the two-way table as follows (interchange P_3 with P_2 and also interchange D_3 with D_2), we see directly that there are two disconnected sets of four crosses each, see Table 2.

Table 2. Rearranged two-way table of Table 1 showing directly the disconnected crossing scheme.

		<i>pisifera</i>			
		P ₁	P ₃	P ₂	P ₄
<i>dura</i>	D ₁	*	*		
	D ₃	*	*		
	D ₂			*	*
	D ₄			*	*

The first set contains the 4 connected crosses $D_1 \times P_1$, $D_1 \times P_3$, $D_3 \times P_1$ and $D_3 \times P_3$; the second set contains the 4 connected crosses $D_2 \times P_2$, $D_2 \times P_4$, $D_4 \times P_2$ and $D_4 \times P_4$. In such a disconnected crossing scheme no unbiased estimate can be made for the difference in GCA effect between, for example *dura* D_1 and D_2 or from the difference in GCA 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 the crossing scheme is connected.

A necessary (but not sufficient) condition to have a connected design is that C must be at least equal to the sum of the degrees of freedom of Intercept, *dura* and *pisifera* $C \geq 1 + (A-1) + (B-1) = A + B - 1$. In the example above we have $A = 4$ and $B = 4$, so $C \geq 4+4-1=7$ crosses are sufficient for a connected design. In Table 3 the crossing scheme is connected when the following $C=8$ crosses were made:

Table 3. A connected incomplete diallel with 8 crosses of 4 *dura* and 4 *pisifera*.

		<i>pisifera</i>			
		P ₁	P ₂	P ₃	P ₄
<i>dura</i>	D ₁	*	*		
	D ₂		*	*	
	D ₃			*	*
	D ₄	*			*

Here we have 8 crosses and the crossing scheme is still connected when $C=7$, for example, if the cross $D_4 \times P_1$ was not made.

In the past the author found many large crossing schemes that were not connected, because the oil palm breeder looked only for including previous good *dura* and *pisifera* parents. But the easiest way is to produce random yields with A *dura* and B *pisifera* parents and analyse the two-way classification of the crosses with a Personal Computer with statistical packages as SAS, IBM SPSS-Statistics or R, to find in the Analysis of Variance the Sum of Squares (SS) Type III option (SS of *dura* after correction for *pisifera* and SS of *pisifera* after correction for *dura*). If the degrees of freedom (df) of *dura* = $A-1$ and the df of *pisifera* = $B-1$, then the crossing

scheme is connected, otherwise it is disconnected.

Now we will discuss the construction of a good crossing design when we have *A dura* and *B pisifera* and we want to use *C* crosses in an incomplete diallel connected scheme where $A+B-1 \leq C \leq A \times B$. The choice between several connected mating designs can best be tested on the standard error of the estimator for the difference in the GCA value of all the *dura* pairs and of the *pisifera* pairs. The standard error of the estimator for the difference in the GCA value between two *dura* parents D_i and D_j is $(S_{Dij}) \times \sigma$, or between two *pisifera* parents P_i and P_j is $(S_{Pij}) \times \sigma$, where σ is the residual standard deviation and the value of S_{Dij} and S_{Pij} depends solely on the mating scheme. The value of σ depends on the studied trait (e.g. yield), the variation between the plots in the experimental field and the plot size. For complete crossing schemes (as a complete diallel) with *A dura* and *B pisifera*, where 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_{Dij} is $\sqrt{2/(B \times r)}$; 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_{Pij} is $\sqrt{2/(A \times r)}$. For incomplete 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 based on the coefficients S_{Dij} and S_{Pij} .

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. In such incomplete block designs there must be compared v varieties (or treatments) in blocks of sizes of k plots, where the block size $k < v$. Well known incomplete block designs are lattices where $v = k \times k$ or rectangular lattices where $v = k \times (k+1)$. To extend the possibilities for v unequal to $k \times k$ or $k \times (k+1)$ there are the so-called *alpha*-designs (see [9, 10]). To use such an incomplete block design the role of treatments is played by the *dura* and the role of the incomplete blocks 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*; further-more 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 chosen by the author as an *alpha*-design, with $v = 20$ treatments (= *dura*), $k = 40/10 = 4$ as block size, $b = 10$ blocks (= *pisifera*), $r = 2$ replications where the first replication consists of blocks 1-5 and the second replication consists of blocks 6-10, hence a resolvable design.

Table 4. Three crossing schemes with $C = 40$ crosses of 20 *dura* and 10 *pisifera*.

Design I											Design II											Design III										
<i>pisifera</i>											<i>pisifera</i>											<i>pisifera</i>										
<i>dura</i>	1	2	3	4	5	6	7	8	9	10	<i>dura</i>	1	2	3	4	5	6	7	8	9	10	<i>dura</i>	1	2	3	4	5	6	7	8	9	10
1	*										1	*										1	*					*				
2		*									2		*				*					2		*					*			
3			*								3			*								3			*					*		
4				*							4			*				*				4			*					*		
5					*	*					5			*	*							5				*					*	
6						*					6			*				*				6	*								*	
7					*	*					7			*	*							7		*				*				
8						*					8			*					*			8			*				*			
9						*	*				9			*	*							9			*				*			
10						*	*				10			*	*					*		10			*				*			
11						*	*				11			*	*							11	*						*			
12						*	*				12	*			*							12	*		*						*	
13						*	*				13			*	*			*	*			13			*			*		*		
14					*		*				14		*		*				*			14		*		*			*			
15						*	*				15			*	*			*	*			15			*		*		*			
16				*			*				16		*		*				*			16	*			*			*		*	
17						*	*				17			*	*				*	*		17		*		*			*		*	
18		*				*	*				18			*	*				*	*		18		*		*			*		*	
19						*	*				19			*	*				*	*		19			*		*		*		*	
20	*					*	*			*	20			*	*			*	*		*	20			*		*		*		*	

It can be directly seen that all these three mating designs are connected. In the following table the minimum,

maximum and average of the coefficients S_{Dij} and S_{Pij} for the standard errors of the estimators of the difference between

GCA values of pairs of *dura* and *pisifera* parents, for the three designs are given in Table 5.

Table 5. Coefficients S_{Dij} and S_{Pij} for the standard errors of the estimators of the difference between GCA values.

mating design	S_{Dij} for pairs of <i>dura</i>			S_{Pij} for pairs of <i>pisifera</i>		
	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	1.250	1.125
III	1.125	1.291	1.214	0.949	1.080	1.001

From Table 5 it is clear that design III (the *alpha*-design), which has the smallest average value for S_{Dij} and S_{Pij} for the *dura* and the *pisifera* pairs, and moreover has the smallest range (max – min) for S_{Dij} and S_{Pij} , is the mating design which must be preferred. Hence it is worthwhile to use an *alpha*-design which gives always a connected mating design and be careful to rely too much on “experience”!

3. Analysis of an Oil Palm Breeding Trial

Let us consider the case that we have made $C = 10$ connected *tenera* crosses T_1, T_2, \dots, T_{10} derived from $A = 5$ *dura* mothers and $B = 5$ *pisifera* fathers. In the following two-way table the crossing scheme is given; a dot (•) indicates a cross which has not been made, see Table 6.

Table 6. A connected incomplete diallel of 10 crosses *tenera* from 5 *dura* and 5 *pisifera*.

		<i>pisifera</i>				
		P ₁	P ₂	P ₃	P ₄	P ₅
<i>dura</i>	D ₁	T ₁	•	•	•	T ₁₀
	D ₂	T ₂	T ₃	•	•	•
	D ₃	•	T ₄	T ₅	•	•
	D ₄	•	•	T ₆	T ₇	•
	D ₅	•	•	•	T ₈	T ₉

The palm plot consists of 6 rows of 6 palms, where the palms are laid down at the corners of an equilateral triangle with 9 m sides. For the yielding capacity of a cross the yield of the $4 \times 4 = 16$ inner palms is recorded. Suppose that the experimental field is very heterogeneous, and that we can only find homogeneous growing conditions (blocks) of five adjacent plots.

A resolvable *alpha*-design with block size $k = 5$ and with $r = 4$ replications is used. The index of the *tenera* crosses T_i is given by the program CycDesign in the randomized resolvable *alpha*-design in Table 7.

Table 7. Randomized design of 10 *tenera* crosses in the incomplete blocks of size 5.

Rep 1	Block 1	9	4	1	6	2
	Block 2	5	7	8	10	3
Rep 2	Block 3	5	4	3	8	1
	Block 4	6	2	7	9	10
Rep 3	Block 5	10	4	9	6	3
	Block 6	7	8	1	5	2
Rep 4	Block 7	9	1	3	5	7
	Block 8	10	4	8	6	2

After we have laid out the design of Table 7 in the field,

we gathered after a year the yield y in ton/ha. We made then the following data file of the results from this example (see Table 8) for IBM-SPSS Statistics, where we used consecutive block numbers 1 – 8 for the blocks in the replications.

Table 8. Data file for the yields of the design of Table 7 for IBM-SPSS Statistics.

rep	block	tenera	dura	pisifera	y
1	1	9	5	5	20.10
1	1	4	3	2	17.50
1	1	1	1	1	15.70
1	1	6	4	3	18.10
1	1	2	2	1	14.60
1	2	5	3	3	16.20
1	2	7	4	4	18.70
1	2	8	5	4	21.10
1	2	10	1	5	18.80
1	2	3	2	2	16.70
2	3	5	3	3	17.20
2	3	4	3	2	15.80
2	3	3	2	2	17.90
2	3	8	5	4	18.10
2	3	1	1	1	15.10
2	4	6	4	3	16.70
2	4	2	2	1	15.00
2	4	7	4	4	20.00
2	4	9	5	5	21.70
2	4	10	1	5	18.60
3	5	10	1	5	16.00
3	5	4	3	2	15.60
3	5	9	5	5	21.50
3	5	6	4	3	17.20
3	5	3	2	2	16.00
3	6	7	4	4	18.40
3	6	8	5	4	17.30
3	6	1	1	1	14.70
3	6	5	3	3	16.70
3	6	2	2	1	13.50
4	7	9	5	5	20.00
4	7	1	1	1	16.00
4	7	3	2	2	14.60
4	7	5	3	3	16.70
4	7	7	4	4	17.20
4	8	10	1	5	17.20
4	8	4	3	2	14.90
4	8	8	5	4	19.10
4	8	6	4	3	18.30
4	8	2	2	1	13.80

The IBM-SPSS Statistics syntax to find the GCA values for the *dura* and the *pisifera* and the adjusted *tenera* means (EMMEANS= Estimated Marginal Means) is given in Table 9.

Table 9 IBM-SPSS Statistics syntax file for the analysis of data file Table 8.

```
Title 'Table 8, blocks random with module MIXED'.
MIXED y BY dura pisifera rep
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10)
SCORING(1) SINGULAR(0.000000000001)
HCONVERGE(0,ABSOLUTE)
LCONVERGE(0,ABSOLUTE)
PCONVERGE(0.000001, ABSOLUTE)
/FIXED= rep dura pisifera | SSTYPE(3)
```

```

/METHOD=REML
/PRINT=SOLUTION TESTCOV
/RANDOM=INTERCEPT | SUBJECT(block)
COVTYPE(VC)
/EMMEANS=TABLES(dura) COMPARE ADJ(LSD)
/EMMEANS=TABLES(pisifera) COMPARE ADJ(LSD).

```

Title 'Table 8, tenera, blocks random with module MIXED'.

```

MIXED y BY tenera rep
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10)
SCORING(1) SINGULAR(0.000000000001)
HCONVERGE(0,ABSOLUTE)
LCONVERGE(0,ABSOLUTE)
PCONVERGE(0.000001, ABSOLUTE)
/FIXED= rep tenera | SSTYPE(3)
/METHOD=REML
/PRINT=SOLUTION TESTCOV
/RANDOM=INTERCEPT | SUBJECT(block)
COVTYPE(VC)
/EMMEANS=TABLES(tenera) COMPARE ADJ(LSD).

```

The SAS syntax to find the GCA values for the *dura* and the *pisifera* and the adjusted tenera means (LSMEANS = Least Square Means; this LSMEANS is the same as what SPSS called EMMEANS), is given in Table 9.

Table 9. SAS syntax for the analysis of the yields of the design of Table 7.

```

data table7;
input rep block tenera dura pisifera y;
datalines;

```

1	1	9	5	5	20.10
1	1	4	3	2	17.50
1	1	1	1	1	15.70
1	1	6	4	3	18.10
1	1	2	2	1	14.60
1	2	5	3	3	16.20
1	2	7	4	4	18.70
1	2	8	5	4	21.10
1	2	10	1	5	18.80
1	2	3	2	2	16.70
2	3	5	3	3	17.20
2	3	4	3	2	15.80
2	3	3	2	2	17.90
2	3	8	5	4	18.10
2	3	1	1	1	15.10
2	4	6	4	3	16.70
2	4	2	2	1	15.00
2	4	7	4	4	20.00
2	4	9	5	5	21.70
2	4	10	1	5	18.60

3	5	10	1	5	16.00
3	5	4	3	2	15.60
3	5	9	5	5	21.50
3	5	6	4	3	17.20
3	5	3	2	2	16.00
3	6	7	4	4	18.40
3	6	8	5	4	17.30
3	6	1	1	1	14.70
3	6	5	3	3	16.70
3	6	2	2	1	13.50
4	7	9	5	5	20.00
4	7	1	1	1	16.00
4	7	3	2	2	14.60
4	7	5	3	3	16.70
4	7	7	4	4	17.20
4	8	10	1	5	17.20
4	8	4	3	2	14.90
4	8	8	5	4	19.10
4	8	6	4	3	18.30
4	8	2	2	1	13.80

```
;
```

```
run;
```

```
proc print data = table7;
```

```
run;
```

```

title 'Table 7; proc Mixed, blocks random';
proc mixed method=REML data = table7;
class dura pisifera rep block;
model y = rep dura pisifera / ddfm=satterthwaite;
random block / type =VC;
lsmeans dura / diff=all;
lsmeans pisifera / diff=all;
run;

```

```

title 'Table 7; proc Mixed, tenera, blocks random';
proc mixed method=REML data = table7;
class tenera rep block;
model y = rep tenera / ddfm=satterthwaite;
random block / type =VC;
lsmeans tenera / diff=all;
run;

```

From the output we find that the *dura* is significant (P-value = 0.005) and that the *pisifera* is significant (P-value = 0.001) and the *tenera* is significant (P-value = 0.000).

The LSMEANS (= EMMEANS) are collected in the following EXCEL worksheet in Table 10.

Table 10. Excel Worksheet with the Least Squares means of *tenera*, *dura* and *pisifera* to calculate the GCA.

Table 10			LS mean	LS mean	LS mean	Additive			LS mean		LS mean
tenera	dura	pisifera	tenera	dura	pisifera	mean D+P	SCA	dura	GCA dura	psifera	GCA pisifera
1	1	1	15.375	16.533	15.703	15.0284	0.3466	1	16.533	1	15.703
2	2	1	14.225	16.077	15.703	14.5724	-0.3474	2	16.077	2	17.083
3	2	2	16.3	16.077	17.083	15.9524	0.3476	3	16.422	3	17.138
4	3	2	15.95	16.422	17.083	16.2974	-0.3474	4	17.993	4	17.442
5	3	3	16.7	16.422	17.138	16.3524	0.3476	5	19.013	5	18.672
6	4	3	17.575	17.993	17.138	17.9234	-0.3484				

Table 10			LS mean	LS mean	LS mean	Additive		LS mean	LS mean
tenera	dura	pisifera	tenera	dura	pisifera	mean D+P	SCA	dura	GCA pisifera
7	4	4	18.575	17.993	17.442	18.2274	0.3476	17.2076	17.2076
8	5	4	18.9	19.013	17.442	19.2474	-0.3474	Mean	Mean
9	5	5	20.825	19.013	18.672	20.4774	0.3476		
10	1	5	17.65	16.533	18.672	17.9974	-0.3474		

The first *dura* mean in column 5 (for *dura* 1) is done with Excel-function VLOOKUP(B3,\$I\$3:\$J\$7;2;1).

The first *pisifera* mean in column 5 (for *pisifera* 1) is done with Excel-function VLOOKUP(C3,\$K\$3:\$L\$7;2;1). Then this has been done for all the other *tenera*.

The Additive Mean of D+P (= *dura* + *pisifera*) for *tenera* = LSMEAN GCA *dura* + LSMEAN *pisifera* – Mean GCA; for *tenera* 1 we get as Additive Mean = 16.533 + 15.703 – 17.2076 = 15.0284.

The SCA (Specific Combining Ability) of *tenera* is calculated as LSMEAN(*tenera*) – Additive Mean; the SCA for *tenera* 1 is 15.375 – 15.0284 = 0.3466. Then this has been done for all the other *tenera*.

In a recent handbook about Oil Palm Breeding, see Chapter 12 “Field Experimentation” [12] the use of incomplete block designs for Oil Palm breeding trials is also treated.

For the use of statistical selection procedures to select the best set of *dura* and *pisifera* parents with the indifference zone approach of selection of Bechhofer [1], or the subset selection procedure of Gupta [6, 8].

4. Conclusions

In Oil Palm Breeding trials the breeder must be aware that he must use a connected incomplete crossing scheme of the *dura* (female parents) and the *pisifera* (male parents) to produce the *tenera* hybrids. The use of an alpha-design to produce such a connected incomplete diallel is easily be done by using the *dura* as the treatment and the *pisifera* as the incomplete blocks.

Because the field plots of the *tenera* palms, which are planted at the side of an equilateral triangle with sides of 9 m, are quite large, the breeder must use an incomplete block design. For the trial design of the *tenera* crosses in the field use an incomplete block design given by an alpha-design, because this gives easily a solution for the number of *tenera* with a small block size of the incomplete blocks.

Using the program CycDesign the needed alpha-designs are easily found for the incomplete diallel and later for the experimental design in the field.

The analysis can be done on a Personal Computer with a statistical package which can analyze a Mixed Model. Syntax are given for SAS and IBM-SPSS Statistics. But the analysis can also be done with the free program R with the package lme4.

References

- [1] 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.
- [2] Caliński, T. (1971): On some desirable patterns in block designs, *Biometrics*, 27, 275–292.
- [3] Cochran, W. G. and Cox, G. M. (1957): *Experimental Designs, second edition*. John Wiley & Sons, New York, London, Sydney.
- [4] Corsten, L. C. A. (1958): Vectors, a tool in statistical regression theory. Thesis Landbouwhogeschool Wageningen, Veenman.
- [5] CycDesign (2014): A package for the computer generation of Experimental Designs, <http://www.vsnl.co.uk/software/cycdesign/>.
- [6] Gupta, S. S. (1965): On some multiple decision (selection and ranking) rules. *Technometrics*, 7, 225-245.
- [7] Kuiper, N. H. (1952): Variantie-analyse. *Statistica*, 6, 149-194. *An English translation of this Dutch article is published in 1983 as “Analysis of Variance”, Mededelingen van de Landbouwhogeschool te Wageningen, Nederland*, 83 (10).
- [8] Laan, P. van der, Verdooren, L. R. (1989): Selection of Populations. An Overview and Some Recent Results. *Biometrical Journal*, 31, 383-420.
- [9] Patterson, H. D., Williams, E. R. (1976): A new class of resolvable incomplete block designs, *Biometrika*, 63, 83–92.
- [10] Patterson, H. D., Williams, E. R. and Hunter, E. A. (1978): Block designs for variety trials. *Journal of Agricultural Science*, Cambridge, 90, 395-400.
- [11] Patterson, H. D., Silvey, V. (1980): Statutory and recommended list trials of crop varieties in the United Kingdom (with discussion). *J R Stat Soc A*, 143, 219–252.
- [12] Verdooren, Rob; Soh, Aik Chin and Roberts, Jeremy (2017): Field Experimentation, Chapter 12 in Soh, Aik Chin, Mayes, Sean and Roberts, Jeremy (editors): *Oil Palm Breeding, Genetics and Genomics*. Taylor & Francis Group, CRC Press, Boca Raton.
- [13] Williams, E. R. (1977): Iterative analysis of generalized lattice designs. *Aust J Stat*, 19, 39–42.
- [14] Stevens, W. L. (1948): Statistical analysis of a non-orthogonal tri-factorial experiment. *Biometrika*, 35, 346-367.