

# STUDIES OF BUNCH ANALYSIS

## 2 - BUNCH SAMPLING TO ESTIMATE OIL YIELD

CORLEY, R H V\*

### ABSTRACT

*To estimate oil yield in research trials, fresh fruit bunch (FFB) yield is multiplied by the oil/bunch determined on a sample of bunches, but there is no standard method for sampling bunches. Data from palms in which all bunches were analysed for seven years allowed actual mean oil/bunch (O/B) and oil yields to be calculated, and compared with estimates from samples. Samples of five bunches per palm gave poor precision, with 95% confidence limits of about  $\pm 4\%$  O/B. Increasing the number of bunches to 10 per palm gave confidence limits of about  $\pm 2.5\%$  O/B. Alternatives include analysing more bunches only if the coefficient of variation (CV) for five bunches is large, or measuring fruit/bunch and oil/mesocarp on 10 bunches, and mesocarp/fruit, which is less variable, on only two or three bunches.*

*The mean O/B of five bunches may give acceptable estimates of oil yield per palm. For progeny or treatment mean oil yields, the mean of 50 bunches per progeny gives a CV of about 10%. Precision is increased if a bunch is sampled from every palm, rather than at random from the progeny.*

*When considering mechanisation of bunch analysis, a method in which the bunch components are determined on different bunches has no disadvantages.*

**Keywords:** oil to bunch, progeny trials, precision.

**Date received:** 23 May 2017; **Sent for revision:** 14 June 2017; **Received in final form:** 9 August 2017; **Accepted:** 4 April 2018.

### INTRODUCTION

The aim of most oil palm research trials is to study effects of treatments on oil yield, but oil yield is never measured directly. Fresh fruit bunch (FFB) yield of individual palms is recorded, and a sample of bunches is analysed to estimate oil/bunch (O/B). The product of FFB yield and O/B gives an estimate of oil yield. There is little information on the accuracy or precision of the estimate, but Ooi and Tam (1976) noted that 'sampling errors are too large to allow detection of small but economic differences between progenies'. In this article, I attempt to rectify this.

There are three possible sources of error in the estimate of oil yield: errors in FFB yield, errors in the bunch analysis method, and non-representative samples of bunches for analysis. Over a period of, say, five years' recording in a breeding trial, the weight of all bunches is recorded, so FFB yield of each individual palm for that period is known exactly (though the relative contributions of genotype and environment are unknown). Blaak (1965) and others have shown that mean yield over the first five years is usually well correlated with later yields. Given that early yields are also the most financially valuable, in terms of return on the capital cost of planting, we may assume that FFB yield data give useful estimates of the life-time yields of individual palms.

The Nigerian Institute for Oil Palm Research (NIFOR) method of bunch analysis (Blaak *et al.*, 1963) is almost universally used. Published work

\* Highlands, New Road, Great Barford, Bedford MK44 3LQ, United Kingdom.  
E-mail: herewardc@aol.com

indicates that, though results may be laboratory dependent, the method estimates true O/B with high precision (Rao *et al.*, 1983; Corley, 2018).

FFB yield data are accurate, therefore, and bunch analysis results probably have high precision. It is not usually feasible to analyse all bunches from a trial, so to allow oil yield to be estimated a sample of bunches must be taken for analysis. The sample must be representative if the oil yield estimate is to be reliable, but there is no standard sample size or method for sampling. To evaluate a progeny of 60 palms, Blaak *et al.* (1963) recommended analysing 180 bunches from every progeny in the first year of production, followed by all bunches from selected progenies in the second year (perhaps 600 bunches, from 60 palms), but no justification for these figures was given. The Malaysian Standard for seed production (MS157:2005) requires 50 bunches from a *tenera* progeny, but in practice 30 bunches per progeny is often considered adequate; this would typically be less than 1% of the bunches produced over five years by 60 palms of a progeny. Bunches are usually taken at random, with no effort made to ensure that all palms are represented; with 30 bunches from 60 palms that would not be possible anyway. For individual seed parent palms, the same Malaysian Standard requires at least three bunches for a *dura* seed parent, perhaps a 5% sample of five years' production. Without information on the variation between bunches from a single palm, the precision of O/B and oil yield estimates from such samples is unknown.

Corley (2018) showed that variation between bunches from a single palm is mainly attributable to variation in fruit/bunch, and to a lesser extent oil/mesocarp. In this article, I have looked at methods of bunch sampling, using data from three small plots of 20 commercial palms. All bunches were weighed and put through bunch analysis, so actual oil yield (within the limits of the accuracy of bunch analysis) could be calculated for each individual palm. I have then simulated various bunch sampling schemes, and compared estimated and actual mean O/B and oil yields.

In some bunch analysis laboratories, only good looking and undamaged bunches are accepted for analysis. The argument for this approach is that one is trying to estimate genetic potential, and damaged bunches would fall short of this. With the data available for this study, it was not possible to simulate this approach; only random sampling has been tested.

## MATERIAL AND METHODS

Three plots of 20 palms chosen at random from commercial fields on Ulu Remis Estate, in Johor, Malaysia (approximately 1° 51'N, 103° 28'E), were

recorded over a period of seven years four months, from September 1968 to December 1975. Details of the three plots are given in *Table 1*. As described by Corley (2018), there was a high level of *dura* contamination in the 1958 and 1960 plots; the 1963 plot was 100% *tenera*. For this study, I have combined data from the *teneras* in the 1958 and 1960 plantings; there were thus two groups of *duras* and two groups of *teneras*.

All bunches were weighed, and analysed by the standard NIFOR bunch analysis method (Blaak *et al.*, 1963). The weight of oil in each bunch was calculated, with oil yield for each palm being the total of the individual bunch figures. The mean O/B for all bunches was also calculated for each palm. A few bunches (Plot A: 18 bunches; Plot B: 20; Plot C: 12) were not completely analysed owing to laboratory errors; these were excluded from the yield calculations and all subsequent analysis.

The bunch analysis data were used to study variability in bunch composition, as described by Corley (2018). Repeatabilities were calculated following Falconer (1981). To simulate conventional sampling for bunch analysis, random samples of bunches were taken from each individual palm. The sampling methods tested were as follows:

- random sampling and complete analysis of a pre-determined number of bunches;
- random sampling and complete analysis, with additional samples for F/B and O/WM;
- progressive sampling, with additional bunches sampled if CV of the initial sample is high;
- sampling restricted to a three-year period;
- random sampling, with individual bunch data adjusted according to the month of sampling; and
- random sampling with partial analysis, each bunch component being measured on a different bunch.

Further details of each method are given under Results and Discussion section. Some methods were only tested on one or two groups of palms. With the method of random sampling used, the same bunch was sometimes included more than once in a batch. As a result for about 3% of five-bunch samples and 10% of 10-bunch samples the true number of bunches in the sample was lower than shown.

Where the number of analyses differed between bunch components, sample O/B was calculated from the sample means for the individual components ( $O/B = F/B \times M/F \times O/WM$ , where F/B = fruit/bunch, M/F = mesocarp/fruit and O/WM = oil/fresh mesocarp). To calculate seasonal and age-related trends, the overall mean for each palm was calculated for all components, and the figures for each bunch were then expressed as a percentage of the mean for that palm.

For each sampling method tested, the sampling was repeated 50 times, and correlations between

estimated and actual mean O/B were calculated. The correlations (*r*) obtained with different sampling methods were compared by one-way analysis of variance after transformation to *z* [where  $z = 0.5 \{ \log_e (1+r) - \log_e (1-r) \}$ ], which is approximately normally distributed (Snedecor and Cochran, 1966).

Additional data came from irrigated progeny trials in Southern Thailand; details of these trials were described by Corley and Palat (2013).

## RESULTS AND DISCUSSION

Table 1 shows overall yields of the plots. In the first part of this study (Corley, 2018), means and standard deviations for bunches from individual palms for all bunch components were described; results are summarised in Table 2. Variation between bunches in mesocarp/fruit (M/F) was generally low, reflecting strong genetic control. In all groups, fruit/bunch (F/B) was the most variable bunch component, followed by oil to fresh mesocarp (O/WM).

In the following sections, I have investigated various bunch sampling methods, and determined the precision obtained. Another approach would be to specify the size of difference to be detected, whether for progeny or individual palm selection, and then use the data in Table 2 to determine the sample size required.

## Bunch Sampling

If O/B has a standard deviation of 3.0% (Table 2), a sample of three bunches will have a standard error of 1.73% O/B, and will give an estimate of the mean with 95% confidence limits (CL) of  $\pm 7.4\%$  O/B (CL =  $t_{0.05} SD/\sqrt{n}$ ). A sample of five bunches would have CL of  $\pm 3.7\%$  O/B; that is, for a palm in Plot C with a mean O/B of 24%, the CL would be between 20.3% and 27.7% O/B. A sample of 10 bunches would give an estimate within  $\pm 2.1\%$  of the true mean O/B.

Repeatabilities are shown in Table 3. According to Falconer (1981), with a repeatability of 0.5 there will be little gain in accuracy from taking more than five samples, but with a repeatability of 0.25, seven or eight samples will be worthwhile. Thus for M/F five samples should be adequate, but for other components and for O/B more samples will improve accuracy.

Another way of looking at the data is to consider the correlation between the sample estimates and the actual or 'true' O/B for each palm. True O/B was taken as the mean of all bunches analysed. Table 4 shows that the correlations were significantly higher for five-bunch samples than for three bunches, and higher again for 10-bunch samples. Figure 1 shows examples of a set of three-bunch samples with a correlation between samples and actual oil/bunch

TABLE 1. DETAILS OF THE PLOTS STUDIED

Plot	A				B				C	
	1958				1960				1963	
	<i>Dura</i>		<i>Tenera</i>		<i>Dura</i>		<i>Tenera</i>		<i>Tenera</i>	
Year of planting	Mean	SD*	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Number of palms	13		7		17		3		20	
FFB yield (kg/palm.yr)	145	16.3	183	22.8	157	29.8	171	27.9	159	27.1
Bunch number/palm.yr	6.4	1.18	8.3	0.93	9.3	2.38	8.8	2.40	11.1	3.75
Mean bunch weight (kg)	23.8	3.83	22.1	2.47	17.3	2.56	20.1	2.15	15.2	3.30
Oil/bunch (%)	18.5	1.63	22.7	1.96	17.9	1.57	25.7	0.83	24.1	2.17
Oil yield (kg/palm.yr)	27.3	4.58	41.6	7.15	28.5	6.67	44.4	8.62	38.7	7.41

Note: \*Standard deviations for palm means.

TABLE 2. MEANS, STANDARD DEVIATIONS AND COEFFICIENTS OF VARIATION FOR BUNCH COMPONENTS\*

Component	Plot A, <i>dura</i>			Plot B, <i>dura</i>			Plots A+B, <i>tenera</i>			Plot C, <i>tenera</i>		
	Mean	SD	CV	Mean	SD	CV	Mean	SD	CV	Mean	SD	CV
Fruit/bunch	65.0	6.2	9.6	65.5	7.6	11.7	62.3	6.8	10.9	66.1	7.2	11.0
Mesocarp/fruit	59.5	2.5	4.2	60.9	2.8	4.6	79.0	2.8	3.6	77.9	2.9	3.7
Dry matter/mesocarp	64.2	4.0	6.2	60.7	4.7	7.8	63.5	3.8	6.0	63.1	4.5	7.2
Oil/dry mesocarp	75.0	2.1	2.8	74.3	2.3	3.1	75.7	1.9	2.5	74.4	2.3	3.1
Oil/fresh mesocarp	48.2	3.6	7.5	45.2	4.3	9.5	48.0	3.9	8.2	47.0	4.2	9.0
Oil/bunch	18.5	2.5	13.5	17.9	2.8	15.7	23.6	3.3	14.1	24.1	3.5	14.7
No. of bunches analysed		609			1 164			620			1 635	
Bunches/palm (mean, range)	47	36 - 63		68	41 - 96		62	47 - 89		82	48 - 169	

Note: \*Averages of means, standard deviations (SD) and coefficient of variation (CV) for bunches from individual palms.

TABLE 3. REPEATABILITY OF BUNCH COMPONENTS

Component	Plot A <i>dura</i>	Plot B <i>dura</i>	Plots A+B <i>tenera</i>	Plot C <i>tenera</i>	Mean all palms
Fruit/bunch (F/B)	0.14	0.21	0.07	0.34	0.19
Mesocarp/fruit (M/F)	0.64	0.36	0.74	0.47	0.55
Dry matter/mesocarp (DM/W)	0.41	0.16	0.23	0.12	0.23
Oil/dry mesocarp (O/DM)	0.44	0.28	0.36	0.35	0.36
Oil/fresh mesocarp (O/WM)	0.45	0.19	0.20	0.18	0.25
Oil/bunch (O/B)	0.28	0.23	0.31	0.35	0.29

of 0.798, a five-bunch sample set with a correlation of 0.838, and a 10-bunch sample set with a correlation of 0.904 (these correlations were close to the average across all four groups). The better fit of the 10-bunch samples to the true O/B means is clear.

For some purposes it may be sufficient to rank palms, rather than determining absolute values of O/B. As an example, although the rank correlation for the palms in *Figure 1a* was highly significant [ $r = 0.764$ , 18 degree of freedom (d.f.)], the figure shows that the second best palm, with 27% O/B, ranked only 10th based on three-bunch samples. For reliable ranking, a larger sample of bunches would be needed.

Adopting a 10-bunch sample as standard would double the cost of individual palm selection, so I looked for cheaper alternatives. Given the high variability of F/B (*Table 2*), I investigated the possibility of analysing additional bunches for F/B only. [Corley (2018) found a negative correlation between F/B and M/F within palms, but there is no obvious way to allow for this in sampling]. *Table 4*

shows that the correlations between sample means and true values were improved by increasing the number of F/B analyses to 10 per palm, while keeping numbers for the other components unchanged. With three complete analyses per palm the correlation was significantly higher in all groups if a further seven F/B analyses were added. With five complete analyses, a further five F/B analyses improved the correlation significantly in two groups.

After F/B, O/WM is the next most variable component, so I also tested the value of additional O/WM analyses. In three of the groups, three complete analyses plus seven F/B and seven O/WM gave higher correlations than three complete analyses plus seven F/B only (*Table 4*). Similarly, five complete analyses plus five F/B and five O/WM was significantly better than five complete plus five F/B only in three groups, and was only significantly worse than 10 complete analyses in one group.

With M/F being less variable than other components (*Table 2*) and contributing little to between-bunch variation (Corley, 2018), a cost

TABLE 4. CORRELATIONS BETWEEN SAMPLE MEANS AND TRUE OIL/BUNCH FOR INDIVIDUAL PALMS

No. of analyses			Correlations + true O/B							
Compl.	F/B only	O/WM +F/B	Plot A <i>duras</i>		Plot B <i>duras</i>		Plots A+B <i>teneras</i>		Plot C <i>teneras</i>	
			Mean	Range	Mean	Range	Mean	Range	Mean	Range
3	-	-	0.799a	0.56 - 0.91	0.718a	0.35 - 0.89	0.812a	0.39 - 0.97	0.731a	0.46 - 0.85
5	-	-	0.869bc	0.64 - 0.97	0.801bc	0.50 - 0.90	0.885b	0.52 - 0.98	0.804b	0.58 - 0.92
10	-	-	0.923f	0.78 - 0.97	0.885f	0.70 - 0.95	0.937c	0.83 - 0.99	0.891e	0.78 - 0.95
3	7	-	0.848b	0.71 - 0.97	0.772b	0.46 - 0.91	0.927c	0.80 - 0.99	0.800b	0.44 - 0.92
5	5	-	0.885cd	0.66 - 0.97	0.828c	0.67 - 0.93	0.935c	0.84 - 0.98	0.835c	0.55 - 0.93
3	-	7	0.907e	0.75 - 0.97	0.864de	0.70 - 0.93	0.927c	0.80 - 0.99	0.872d	0.56 - 0.95
5	-	5	0.912ef	0.78 - 0.97	0.877ef	0.76 - 0.94	0.935c	0.84 - 0.98	0.871d	0.61 - 0.94
1	-	9	0.876c	0.62 - 0.96	0.819c	0.62 - 0.93	0.903b	0.71 - 0.98	0.833c	0.49 - 0.94
2	-	8	0.898de	0.68 - 0.96	0.853d	0.66 - 0.94	0.923c	0.79 - 0.99	0.861d	0.48 - 0.94
$r$ for $P = 0.01$			0.684	11 d.f.	0.606	15 d.f.	0.765	8 d.f.	0.561	18 d.f.

Note: For each method, random samples were taken, and the correlation for that set of samples was calculated; sampling was repeated 50 times; the means are for 50 correlations. For the 5 and 10 bunch samples bunches were added to the initial samples. Correlations were transformed to  $z$  for analysis of variance; means (back-transformed to  $r$ ) followed by the same letter do not differ significantly (95% LSD).

O/B - oil/bunch.

F/B - fruit bunch.

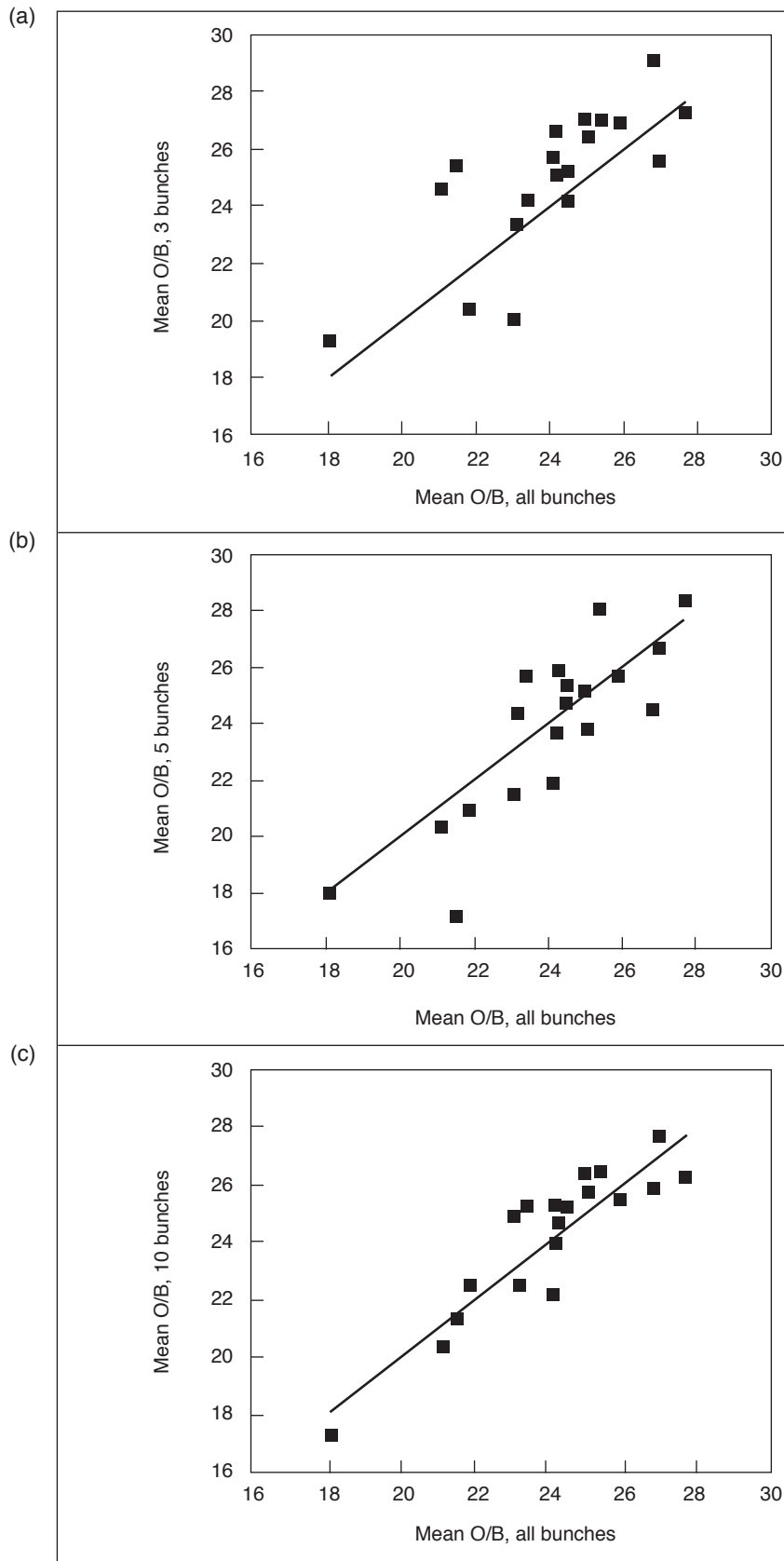
O/WM - oil/fresh mesocarp.

M/F - mesocarp/fruit.

$r$  - correlation coefficient.

$P$  - probability.

LSD - least significant difference.



Note: O/B – oil/bunch.

Figure 1. True oil/bunch and estimates for palms in Plot C.

Lines indicate 1:1 relationship;

(a) - Estimates based on three bunches; correlation with true O/B = 0.798.

(b) - Estimates based on five bunches; correlation with true O/B = 0.838.

(c) - Estimates based on 10 bunches; correlation with true O/B = 0.904.



saving might be made by reducing the number of M/F analyses. Reducing the number of M/F analyses from five to two, with 10 F/B and 10 O/WM analyses, reduced the average correlation significantly in one group, but reducing it to one analysis reduced the correlations in all groups (Table 4). However, in all groups, one complete analysis with nine additional F/B and O/WM analyses was better than three complete analyses only.

The only methods for which the correlation was always significant at  $P = 0.01$  or lower were complete analysis of 10 bunches per palm and five complete analyses plus five additional F/B and O/WM (Table 4). The former method gave a significantly higher average correlation in one group, but the latter method would be cheaper.

### Progressive Sampling

A procedure sometimes recommended is to accept a three-bunch sample if results are consistent [with a low coefficient of variation (CV)], but to sample more bunches if results are very variable. The palms in Plot C did differ in the variability from bunch to bunch, with O/B CV for individual palms ranging from 8.6% to 21.8%. This appears to support the approach, but the overall CV and CV of three-bunch samples were not highly correlated: for 50 sets of samples, the average correlation was 0.267 (range from -0.200 to 0.698, with 18 d.f.), with only 11 of 50 being significant at 5%. A small CV for the sample does not necessarily indicate accuracy either. As examples, a three-bunch sample with CV of 3.4% had a mean of 28.0% O/B, compared to an actual O/B for that palm of 24.1%; conversely, a sample with a CV of 24% had a mean O/B of 21.5%, from a palm with an actual O/B of 21.7%. Over 50 sets of samples, the average correlation between

sample CV and the absolute difference between the estimate and the true O/B was only 0.145 (range -0.411 to 0.732).

Despite the above observations, I simulated several possible progressive sampling schemes in two groups, as shown in Table 5. The first sample of three bunches was accepted if the CV was below 10%; for samples with higher CV (an average of 46% of palms), either two more bunches were taken to give a total of five, or seven more bunches were analysed for F/B only. Table 5 shows that neither of these methods improved the correlation with actual O/B compared to three bunches alone. Taking five bunches only where the CV for three bunches was high was significantly worse than taking five bunches from all palms.

With five bunches as the basic sample, adding analyses of F/B alone where CV was above 10% (62% of palms on average) did not make a significant difference. However, adding five more complete analyses gave significantly higher correlations, and in Plot C was not significantly worse than taking 10 bunches from every palm. This method looks promising, therefore. If 60% of palms have to be resampled, costs would be 20% lower than taking 10 bunches from every palm - but 60% higher than sampling only five bunches from every palm.

### Seasonal and Age-related Trends

Table 6 shows that O/B was low in all groups in 1969, and fairly low in 1970, mainly because of low F/B. As the three groups were different ages, this is unlikely to reflect a simple trend with age, so may indicate that pollination was poor in 1969 for extraneous reasons. I tested the effect of restricting sampling to 1971 - 1973 in Plot C. The improvement in average correlations between sample means and

TABLE 5. CORRELATIONS BETWEEN SAMPLE MEANS AND TRUE OIL/BUNCH FOR INDIVIDUAL PALMS, PROGRESSIVE SAMPLING

Initial sample	Additional analyses for palms + CV > 10%	Plot B <i>duras</i>			Plot C <i>teneras</i>		
		%	Mean	Range	%	Mean	Range
3 bunches	no more analyses	53	0.734a	0.29 - 0.88	55	0.759a	0.44 - 0.95
3 bunches	7 analyses F/B only	47	0.755a	0.48 - 0.89	45	0.772a	0.31 - 0.88
3 bunches	2 complete analyses		0.772a	0.44 - 0.94		0.782a	0.60 - 0.94
5 bunches	no more analyses	35	0.812b	0.60 - 0.96	41	0.822b	0.59 - 0.94
5 bunches	5 analyses F/B only	65	0.832b	0.64 - 0.97	59	0.841b	0.41 - 0.92
5 bunches	5 complete analyses		0.868c	0.64 - 0.98		0.873c	0.73 - 0.93
10 bunches	no more analyses	100	0.889d	0.73 - 0.98	100	0.899c	0.76 - 0.95

Note: For each method, random sampling was repeated 50 times, and the correlation for that set was calculated; the means shown are thus the means of 50 correlations. Correlations were transformed to  $z$  for analysis of variance; means (back-transformed to  $r$ ) followed by the same letter do not differ significantly (95% LSD).

CV - coefficient of variation.

F/B - fruit/bunch.

$r$  - correlation coefficient.

LSD - least significant difference.

TABLE 6. ANNUAL AND MONTHLY MEANS FOR BUNCH COMPONENTS AS PERCENTAGES OF OVERALL MEAN

Year	Plot A, <i>durras</i>				Plot B, <i>durras</i>				Plots A + B, <i>teneras</i>				Plot C, <i>teneras</i>							
	O/B	F/B	M/F	D/WM	O/DM	O/B	F/B	M/F	D/WM	O/DM	O/B	F/B	M/F	D/WM	O/DM	O/B	F/B	M/F	D/WM	O/DM
1969	90.8	93.9	101.8	95.6	99.2	95.4	96.2	101.2	98.1	100.3	94.4	94.0	100.3	98.0	100.3	90.5	92.6	101.5	96.9	99.3
1970	94.5	95.5	101.6	98.0	99.3	99.9	99.0	101.6	99.1	99.8	98.8	99.5	101.4	98.6	99.7	99.6	100.6	101.0	98.0	100.1
1971	104.6	104.5	101.7	98.5	100.0	103.6	103.4	100.6	99.2	100.2	98.6	101.8	99.6	97.4	99.4	102.0	102.8	100.6	98.6	99.8
1972	108.3	105.9	100.9	100.4	101.1	103.7	102.2	100.4	100.4	100.3	105.5	103.7	100.2	100.6	100.9	105.0	103.1	99.9	100.9	101.0
1973	103.4	98.4	100.4	103.9	101.3	101.5	101.8	98.5	100.8	100.0	100.7	98.9	100.7	101.1	100.4	105.4	101.4	99.3	103.4	101.1
1974	102.4	100.9	99.3	102.7	99.7	102.7	101.7	98.6	101.9	100.6	99.9	98.5	99.8	101.6	100.0	102.2	101.0	98.3	102.9	100.1
1975	103.9	103.9	98.6	102.6	98.8	101.3	101.7	98.2	102.0	99.5	103.8	103.1	98.8	102.7	99.3	104.3	102.0	98.8	104.4	99.3

Month	Plot A, <i>durras</i>				Plot B, <i>durras</i>				Plots A + B, <i>teneras</i>				Plot C, <i>teneras</i>							
	O/B	F/B	M/F	D/WM	O/DM	O/B	F/B	M/F	D/WM	O/DM	O/B	F/B	M/F	D/WM	O/DM	O/B	F/B	M/F	D/WM	O/DM
Jan	103.9	104.5	99.6	99.5	100.1	102.3	104.7	97.9	100.0	99.5	99.0	102.2	99.0	97.8	99.2	101.8	103.3	99.6	99.6	99.4
Feb	105.9	102.7	101.7	101.2	100.0	96.8	100.4	99.0	98.1	98.9	102.8	101.6	100.4	99.8	99.8	101.6	104.5	99.1	99.2	98.8
Mar	106.8	104.6	101.3	101.4	99.6	102.7	102.5	99.2	101.3	99.9	103.8	101.9	100.3	100.1	100.1	104.2	104.0	99.8	100.8	99.5
Apr	104.0	102.5	100.3	101.3	99.9	98.6	98.7	100.8	99.9	99.2	103.7	100.0	101.5	99.9	99.9	102.5	101.1	100.0	101.6	99.8
May	104.4	100.0	103.4	100.5	100.3	100.2	98.0	101.5	101.3	99.9	98.9	95.5	102.4	99.7	99.7	96.9	94.0	102.4	100.9	99.8
Jun	90.2	87.9	103.5	101.2	97.3	101.2	98.7	101.5	100.4	101.1	100.0	97.9	102.0	100.1	100.1	98.4	94.7	102.7	100.3	100.7
Jul	96.2	99.3	102.2	96.6	99.8	101.9	100.5	101.4	99.3	100.5	98.6	97.0	100.7	100.4	100.4	95.1	97.3	100.2	97.9	99.5
Aug	96.6	100.7	99.2	97.2	99.7	105.7	103.0	102.0	100.0	100.7	100.0	102.7	98.8	99.4	99.4	101.9	101.5	99.3	101.0	100.2
Sep	99.1	97.9	100.4	100.6	100.0	99.2	96.9	101.4	100.6	100.2	101.4	100.0	100.5	100.8	100.8	100.4	99.7	100.0	100.6	100.0
Oct	99.0	98.4	99.1	101.3	100.5	98.6	98.9	99.5	100.0	100.0	96.4	98.9	98.8	100.4	100.4	100.8	100.3	98.6	101.0	101.1
Nov	100.2	101.4	99.1	99.5	100.1	99.4	101.3	97.9	99.8	100.9	95.8	96.3	98.4	99.7	99.7	105.8	102.2	98.9	103.4	101.5
Dec	101.1	100.5	99.0	101.0	100.8	100.8	101.9	98.0	100.6	99.6	102.7	104.5	99.0	100.6	100.6	106.0	103.6	98.6	102.7	100.8

Note: Data for each individual bunch expressed as percentage of the overall mean for the palm, and averaged for all palms in each plot.

O/B – oil/bunch.

F/B – fruit/bunch.

M/F – mesocarp/fruit.

D/WM – dry matter/fresh mesocarp.

O/OM – oil/dry matter.

true O/B resulting from this age restriction was statistically significant, but very small; for 400 sets of samples mean correlation was 0.861 for the later years only, compared with 0.837 when samples were taken from any year.

There are also month-to-month differences in O/B and its components (Table 6). It is not feasible to deal with this by restricting sampling to certain months, as the yield of individual palms can be strongly seasonal. In general, the higher yielding palms tended to fruit more regularly, but palm 14 in Plot C, for example, had above average FFB and oil yield but over seven years only produced three bunches in March - May, out of a total of 67 bunches. It is unlikely that this palm would have been sampled if bunch analysis had been restricted to those months.

An alternative to restricting sampling dates is to adjust monthly figures to long-term averages, as was done by Dumortier *et al.* (1992). For adjustment, they used monthly means for the three standard crosses included in all their trials. They found that without adjustment monthly mean O/B ranged from 93% to 111% of the annual mean, so in a strongly seasonal climate such as that in the Democratic Republic of Congo, adjustment for this should be worthwhile. In Johor, the range of variation in O/B between months in Plot C was slightly smaller (between 95% and 106% of the annual mean), and using monthly averages over all years for adjustment did not improve the correlations between sample means and true oil/bunch (Table 7).

Based on these results, therefore, there appears to be little advantage in restricting bunch analysis to certain years or months, and no benefit from adjusting for month of sampling. However, in a climate with a more marked dry season than Johor seasonal variation should be allowed for. This could be done by sampling all progenies in a trial in the same months, but for individual palm selection adjustment may be useful.

## Mechanisation

If the amount of bunch analysis is to be increased, mechanisation must be given more consideration. With the current bunch analysis method, all components are determined on the same bunch. Thus the complete analyses in the simulations discussed above were from single bunches; where additional F/B and O/WM analyses were included, both were also done on the same additional bunches. With mechanisation, it may be convenient to do analysis of each component on a separate bunch. Thus, F/B might be determined using a single bunch stripper; this would damage the fruit, but M/F could be estimated on samples of undamaged fruit from other bunches, perhaps using the depericarper developed for seed production (Escobar, 1980). O/M could be measured on mesocarp slices from other bunches, if the within-fruit variation described by Ooi and Tam (1976) can be accounted for. Another option might be to use NMR with intact fruit (Shaarani *et al.*, 2010), eliminating the need for mesocarp sampling and soxhlet extraction.

The possibility of determining the components on different bunches is simulated for the two largest groups in Table 8, and compared with five complete analyses plus five additional F/B and O/WM analyses, or 10 complete analyses. In Plot C, separate analyses for each component gave significantly higher correlations than doing all analyses on the same bunches, while for Plot B *duras* there was no difference between the methods. Thus, we may conclude that analysing the components separately will not reduce the precision of individual palm bunch analysis.

## Progeny Means

For estimating progeny means in breeding trials and treatment means in agronomy trials,

TABLE 7. CORRELATIONS BETWEEN SAMPLE MEANS AND TRUE OIL/BUNCH (O/B) FOR INDIVIDUAL PALMS, PLOT C

Adjustment for monthly differences	Number of bunches	Correlation + true O/B	
		Mean	Range
Unadjusted	3	0.781a	0.531- 0.911
Adjusted for month of sampling	3	0.790a	0.529 - 0.923
Unadjusted	5	0.837b	0.618 - 0.935
Adjusted for month of sampling	5	0.843b	0.636 - 0.947
Unadjusted	10	0.905c	0.695 - 0.970
Adjusted for month of sampling	10	0.907c	0.715 - 0.958

Note: For each method, random sampling was repeated 100 times (50 with samples taken from all years, 50 with samples from 1971-1973 only), and the correlation for that set of samples was calculated; the means shown are thus the means of 100 correlations. Correlations were transformed to *z* for analysis of variance; means (back-transformed to *r*) followed by the same letter do not differ significantly (95% LSD).

O/B – oil to bunch.

LSD – least significant difference.



TABLE 8. COMPARISON OF SINGLE BUNCH ANALYSES WITH INDEPENDENT SAMPLES.

Method	No. of analyses			Correlation + true O/B			
	F/B	M/F	O/WM	Plot B <i>duras</i>		Plot C <i>teneras</i>	
				Mean	range	Mean	range
Complete analyses	10	5	10	0.889	0.79 - 0.96	0.877 a	0.76 - 0.94
Complete analyses	10	10	10	0.892	0.80 - 0.95	0.883 a	0.78 - 0.98
Each analysis on a different bunch	10	5	10	0.878	0.71 - 0.96	0.901 b	0.71 - 0.97

Note: Each sampling method was repeated 50 times, and the correlation with true oil/bunch calculated. Correlations were transformed to  $z$  for analysis of variance; means (back-transformed to  $r$ ) followed by the same letter do not differ significantly (95% SD).

F/B - fruit to bunch.

M/F - mesocarp/fruit.

O/WM - oil/fresh mesocarp.

O/B - oil/bunch.

$r$  - correlation coefficient.

LSD - least significant difference.

the standard deviation between bunches will be larger than that for bunches from a single palm, as it will include genetic and environmental variation between palms. In breeding trials, palms within a plot or treatment are usually full sibs, but in this study the relationship between the palms was unknown. Some may have been half-sibs, but a closer relationship than that is unlikely in mixed commercial material. Thus, these plots can be regarded as typical of plots in an agronomy trial.

For all the bunches harvested from Plot C, the standard deviations (SD) was 4.38% O/B; a random sample of 30 bunches would have 95% CL of  $\pm 1.64\%$  O/B, or  $\pm 7\%$  of the mean of 24.1% O/B. With a 50-bunch sample, the 95% CL would  $\pm 1.24\%$ . The SD for bunches from a single progeny should be smaller because the palms will be sibs, but even for bunches from single palms, where there is no genetic variation between bunches, the average SD in Table 2 was 3.4%. Thus, the CL for progeny means may not be very much smaller than for agronomic treatments.

### Oil Yields

As FFB yield is one component of oil yield, and actual FFB yield is recorded, not estimated, correlations between estimated and actual oil yield for individual palms are higher than for oil/bunch. Thus, while the three-bunch samples in Figure 1 had a correlation of 0.798 with actual O/B, the oil yields calculated from the same samples had a correlation of 0.927 with actual oil yield (Figure 2). For five-bunch samples with a correlation of 0.838 with actual O/B, the correlation between estimated and actual oil yields was 0.947, and for 10-bunch samples, 0.973 for oil yield compared to 0.904 for O/B. To estimate oil yield of individual palms, therefore, a five-bunch sample may be adequate, but where breeders want to look specifically at bunch composition for selection or for linkage studies, five bunches per

palm will probably not give the precision needed, as discussed above.

For a progeny, the most important characteristic will usually be the mean oil yield. A common practice appears to be to sample bunches from each progeny as a whole, disregarding the statistical design of replicated plots. Progeny mean oil yield can then be calculated from O/B  $\times$  FFB yield, but statistical analysis of the data will take no account of possible variation in O/B between plots. If a separate sample of bunches is taken from each plot then the O/B and oil yield can be calculated for each plot, and both can be subjected to the same statistical analysis as FFB yields. In progeny trials in Thailand a separate sample of bunches was analysed from each plot, and Figure 3a shows that the CV of progeny means for O/B decreased as the total number of bunches sampled was increased. With 30 bunches per progeny (10 per plot from three replicates) the standard error of progeny mean O/B in these trials was about 0.97%, giving a 95% LSD between progenies of about 1.6% O/B. By increasing the number of bunches to 50 (c. 17/plot), the standard error was reduced to about 0.77%, and the least significant difference (LSD) to about 1.2% O/B.

The standard error of oil yield will depend on the plot to plot variation in FFB yield, as well as the accuracy of O/B estimates. FFB yields of individual plots are affected by environmental factors (experimental error), and this variation is additional to variation in O/B. Thus, the CV for oil yield will generally be larger than that for O/B, as shown in Figure 3b. With 30 bunches analysed, the CV is about 13%, and the 95% LSD between treatment means is about 14% of the trial mean. By increasing the number of bunches to 50 per treatment, the CV is reduced to 10%, and the LSD to 9% of the mean. Where standard errors and LSD for oil yield are undesirably large, therefore, the researcher should look at the standard errors for FFB and O/B

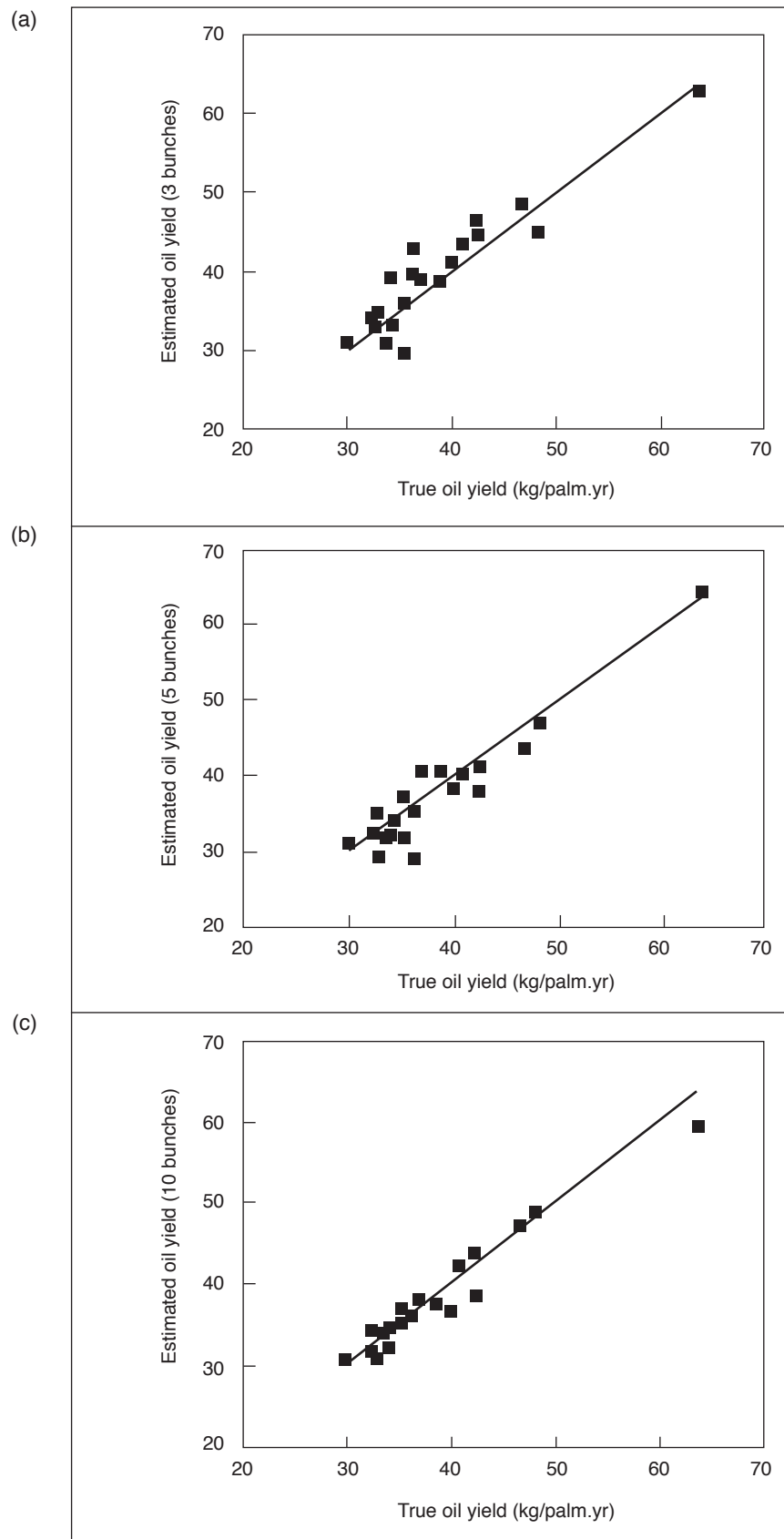


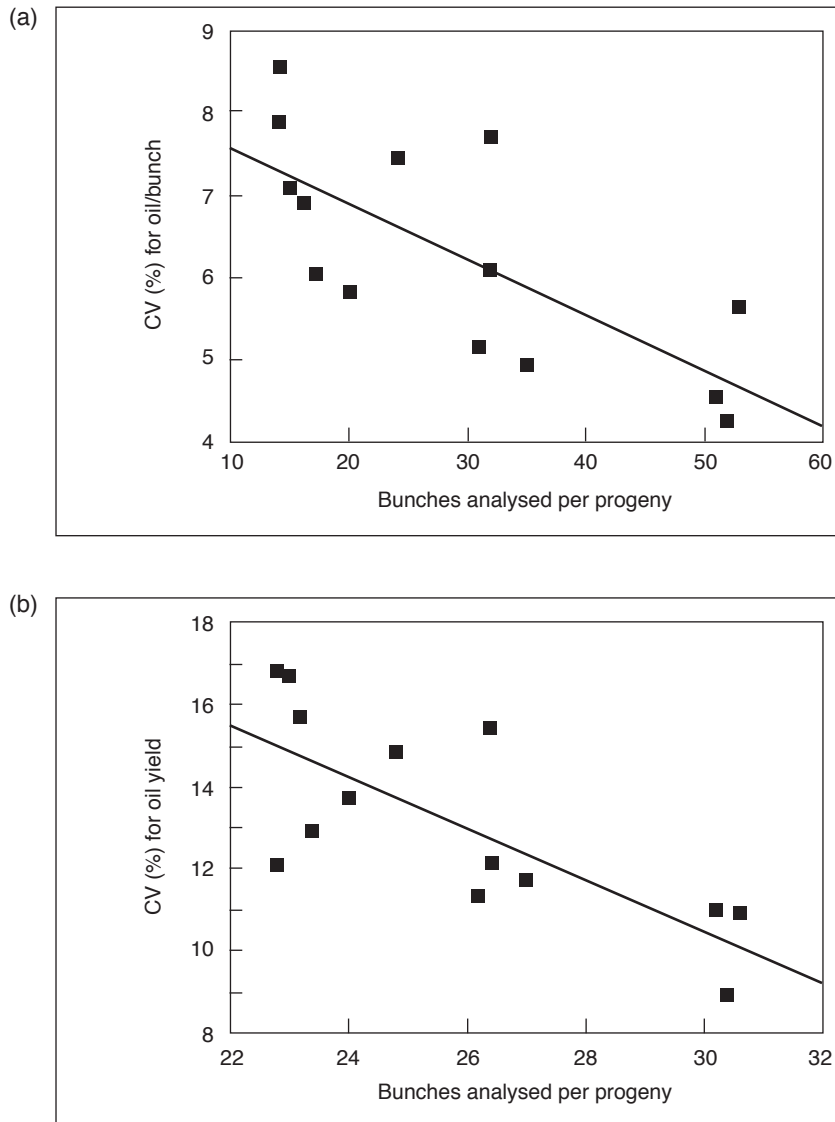
Figure 2. True oil yield (kg/palm.yr) and estimates for palms in Plot C.

Lines indicate 1:1 relationship; and:

(a) - Estimates from three bunches/palm; correlation with true oil yield = 0.927.

(b) - Estimates from bunches/palm; correlation with true oil yield = 0.947.

(c) - Estimates from 10 bunches/palm; correlation with true oil yield = 0.973.



Note: Data provided by Univanich Palm Oil PCL.

Figure 3. Effect of number of bunches sampled on coefficient of variation of progeny means. (a) - Oil/bunch. (b) - Oil yield. The lines are fitted linear regressions.

separately. If the latter is large, more bunch analysis may improve precision.

If bunch analysis costs were lower, it would be feasible to sample one or more bunches from every palm. Plot mean oil yield could then be calculated from the mean of individual palm yields. This was simulated in Plot C. Table 9 shows that the sampling method made little difference to accuracy: means for both O/B and oil yield were within  $\pm 3\%$  of the true values for the plot. However, by taking bunches from each palm rather than at random from the plot, SD were reduced. This was particularly so if only 20 bunches were sampled; in this case, an average of only eight palms out of 20 was included with random sampling, and precision was clearly greater if one bunch was taken from each palm. It made little difference whether a single bunch was

analysed completely, or each of the components was analysed separately as might be done with a mechanised analysis method.

### CONCLUSION

Although oil palm breeding has made much progress since these plots were planted in the 1960s, there has been no attempt to reduce bunch-to-bunch variability within palms, so it is likely that these results remain relevant to present-day breeding programs. It is clear that conventional bunch sampling gives estimates of oil/bunch with large confidence limits. For progeny or treatment means, 10 to 20 bunches per plot with three replications should give reasonably reliable estimates of oil

TABLE 9. ESTIMATES OF MEAN OIL YIELD, PLOT C TENERAS

Sampling method	Palms Sampled*	Oil/bunch (%)		Oil yield (kg/p.yr)	
		Mean	SD	Mean	SD
Random, 20 bunches	8	23.6	1.18	37.6	1.88
1 bunch from each palm	20	24.1	0.75	38.3	1.24
1 separate analysis per palm of F/B, M/F and O/WM	20	24.1	0.76	38.2	1.23
Random, 40 bunches	17	23.7	0.80	37.7	1.28
2 bunches from each palm	20	24.1	0.62	39.2	1.03
2 separate analyses per palm of F/B, M/F and O/WM	20	24.1	0.58	38.2	0.95
True values	-	24.1	-	38.7	-

Note: \*Average number of palms included.

Each sampling method was repeated 50 times; figures are means of these. With random samples, oil yield was estimated from means of O/B and FFB. Where each palm was sampled, oil yield was estimated for each palm, and then averaged.

O/B – oil/bunch.

F/B - fruit/bunch.

M/F - mesocarp/fruit.

O/WM - oil/fresh mesocarp.

SD - standard deviations.

FFB – fresh fruit bunch.

yield, but for individual palm selection on bunch components, 10 bunches per palm gives greater precision than the usual 3 - 5 bunches. However, the cost of bunch analysis in a breeding programme is already high, so doubling it is obviously undesirable. Alternatives are to adopt a progressive sampling scheme where additional bunches are sampled only if the CV of the initial sample is high, or to increase the number of analyses for F/B and O/WM, while possibly reducing the number of M/F analyses to two or three.

Perhaps the most important conclusion is that a cheaper, and preferably mechanised method of bunch analysis is needed. With mechanisation, the different bunch components could be measured on different bunches, without reducing precision.

DM/WM	Dry matter content of fresh mesocarp (%)
DxP	<i>Dura x pisifera</i>
F/B	Fruit/bunch (%)
FFB	Fresh fruit bunch yield
LSD	Least significant difference
M/F	Mesocarp/fruit (%)
NIFOR	Nigerian Institute for Oil Palm Research
O/B	Oil/bunch (%)
O/DM	Oil/dry mesocarp (%)
O/WM	Oil/fresh mesocarp (%)
P	Probability
r	Correlation coefficient
SD	Standard deviations

#### ACKNOWLEDGEMENT

The bunch analysis project was part of the programme of the Oil Palm Genetics Laboratory, jointly funded in the 1960s and early 1970s by Unilever Plantations, Kumpulan Guthrie, Harrisons and Crosfield, and Dunlop Estates. I am grateful to Univanich Palm Oil PCL for permission to include the data from irrigated progeny trials in Figure 3.

Readers wishing to test their own sampling schemes can obtain a spreadsheet with the bunch analysis data from the author at herewardc@aol.com, on condition that any useful results are made public.

#### ABBREVIATION

CL	Confidence limits
CV	Coefficient of variation
d.f.	Degrees of freedom

#### REFERENCES

- BLAAK, G; SPARNAAIJ, L D and MENENDEZ, T (1963). Breeding and inheritance in the oil palm (*Elaeis guineensis* Jacq.). Part II. Methods of bunch quality analysis. *J. W. Afr. Inst. Oil Palm Res.*, 4: 146-155.
- BLAAK, G (1965). Breeding and inheritance in the oil palm (*Elaeis guineensis* Jacq.). Part III. Yield selection and inheritance. *J. W. Afr. Inst. Oil Palm Res.*, 4: 262-283.
- CORLEY, R H V (2018). Studies of bunch analysis: 1 – Variation within and between palms. *J. Oil Palm Res.* DOI: <https://doi.org/10.21894/jopr.2018.0018>.
- CORLEY, R H V and PALAT, T (2013). Maximising lifetime yield for greater economic sustainability. Paper presented at the PIPOC 2013 International Palm Oil Congress - Green Opportunities from the Golden Crop, Kuala Lumpur, 19-21 November 2013.

DUMORTIER, F; VAN AMSTEL, H and CORLEY, R H V (1992). *Oil palm breeding at Binga, Zaire, 1970 - 1990*. Unilever Plantations, London.

ESCOBAR, R (1980). An improved oil palm fruit depulper for single bunch lots. *The Planter*, 56: 540-542.

FALCONER, D S (1981). *Introduction to Quantitative Genetics*. Longman, London & New York.

OOI, S C and TAM, T K (1976). The determination of the within bunch components of oil yield in the oil palm (*Elaeis guineensis* Jacq). I. Pattern of variation. *The Planter*, 52: 467-475.

RAO, V; SOH, A C; CORLEY, R H V; LEE, C H; RAJANAIDU, N; TAN, Y P; CHIN, CW; LIM, K C; TAN, S T; LEE, T P and NGUI, M (1983). A critical reexamination of the method of bunch quality analysis in oil palm breeding. *PORIM Occasional Paper No. 9*: 1-28.

SHAARANI, S M; CARDENAS-BLANCO, A; AMIN, M H G; SOON, N G and HALL, L D (2010). Monitoring development and ripeness of oil palm fruits (*Elaeis guineensis*) by MRI and bulk NMR. *Int. J. Agric. Biol.*, 12: 101-105.

SNEDECOR, G W and COCHRAN, W G, (1966). *Statistical Methods*. Iowa State University Press, Ames, Iowa, USA.