

## REDUCTION OF EXPERIMENTAL ERROR IN COCONUT WITH ADJUSTMENT BY AN INTEGRATED INDEX DEVELOPED THROUGH PRINCIPAL COMPONENT ANALYSIS USING VEGETATIVE AND REPRODUCTIVE CHARACTERS

D T Mathes, H A J Gunathilaka and D N S Fernando

*Coconut Research Institute, Lunuwila, Sri Lanka*

### ABSTRACT

Precision in estimation and accuracy of the results from experiments is dependent on the nature of experimental variation. Controlling variation through calibration, other than through local control, is a well defined concept since 1953.

It is important that the choice of a calibrating variate or variates to be, cost effective, less time consuming in recording, and efficient in reducing the experimental error. For experiments in coconut, one and two years pre-experimental yield is used as calibrating variates. Collecting yield data over a period of one or two years to use as a calibrating variate is a long wait before commencement of an experiment. Financially too it is high.

This paper explores the merits of using, an 'Integrated Index' (incorporating the vegetative and reproductive characters) as a calibrating variate evaluated through the technique of 'Principal Component Analysis'. The characters were, 'Trunk Girth ( $X_1$ )', 'Number of Inflorescences ( $X_2$ )', and Number of nuts per bunch ( $X_3$ )', and 'Number of green fronds ( $X_4$ )'. The Integrated Index - 'I' developed through principal component analysis reflects the 'Vigour of the Palm'.

The eigen vectors of the four characters and the eigen values of the four vectors were evaluated through Principal Component Analysis. The standardized vector scores for each palm was taken as the Integrated Index for each palm. The average of these values for a given experimental plot was taken as the Index for each plot and was used as the calibrating variate.

The results suggest that, the overall range in the values for the 'coefficient of variation' when adjusted for 'I' was 7.05 to 10.78% as compared to the range, 8.35 to 21.02% observed without adjustment.

The recording of variables  $X_1$  to  $X_4$  was completed within 3 to 4 days before the

commencement of the trial. Thus recording of the characters was less time consuming and cost effective.

## INTRODUCTION

Success of field experiments on perennial crops is greatly dependent on the kind of experimental variation (or the "Noise" of the experiment). Lower the variation, higher the precision in estimation and the comparison of the treatment differences. However existence of high variation in perennial crop experimentation is well known. In controlling this variation great effort is made towards selecting the material and land. High genetic variability in coconut, however reduces the effectiveness of such measures of controlling variation. Conscious of this situation and working towards controlling such variation, the concept of "Calibration" took its pride (Pearce 1953). Pearce defined calibration as, the use of some prior information pertaining to a given experimental material to control (by statistical means) its variability during the experimental phase.

As demonstrated later by Pearce & Brown(1960), Vernon & Morris (1964), Moore & Pearce(1976) this technique could contribute in a very large measure towards the success of field experiments on perennial crops through the proper choice of a calibrating variate or variates. One of the most important criterion in the choice of a calibrating variate, is that, it should be free of the treatment effects.

Field experiments in coconut, in general occupies a large extent of land and is conducted over a period of time. An experiment with 6 treatments and four replicates needs a land space of 2 hectares with guard rows. The yield recording being done at two monthly intervals. Thus it is very important that the choice of a calibrating variate or variates of the day to be, cost effective, less time consuming in recording and finally efficient in reducing the experimental variation. Peiris and Salgado (1937), tried to improve on the experimental error in coconut through the choice of large plot size (18 palms), which is more than twice the present plot size (8 palms). Abeywardena (1970) and Abeysinghe (1986), used one and two years pre-experimental yield data as calibrating variates with reduced plot size (6 to 8 palms). They were able to achieve coefficients of variation in the range 6 to 14% after adjustment.

Collecting yield data over a period of one or two years to use as a calibrating variate is a long wait before the commencement of an experiment. It could even frustrate the Scientist. Financially too it is high. Nonetheless it was the choice of the day. Other than using the yield as a calibrating variate/s there are other variates; 'vegetative' and 'reproductive' in character that could be measured or counted at the commencement or after imposing the treatments. These variates are described elsewhere.

This paper explores the relative merits of using these vegetative and reproductive characters individually. It is also described here the use of an 'Integrated Index' (incorporating the vegetative and reproductive characters), as a calibrating variate evaluated through the technique of 'Principal Component Analysis'.

## MATERIALS AND METHODS

The data for this study were obtained from two field experiments (Exp-1, Exp-2) carried on adult coconut palms conducted by the Coconut Research Institute, of Sri Lanka in the Intermediate Zone and at the same locality.

### Exp-1 (Rehabilitation trial)

The coconut palms used for this trial were around 40 years old and were about 10 meters in height. The experiment consisted of ten (10) different rehabilitation treatments and a control. A randomized block design with three replicates were used. A plot consisted of 8 effective palms with a guard row. The planting was square with 8 meters between the palms. Total of six harvests during an year was taken as the total yield for a given experimental year. For the purpose of this study total yields for the experimental years were used. The experimental years are defined as  $E_1Y_1$  to  $E_1Y_5$ .

### Exp-2 (Ploughing trial)

The status of the coconut palms used was as similar to that of the Exp-1. The planting density and the plot size too were similar. The experiment consisted of six different ploughing methods and a control. Here too a randomized block design with three replicates were used. The experiment commenced in simultaneously with the Exp-1. For the purpose of this study total yields for the experimental years were used. The experimental years are defined as  $E_2Y_1$  to  $E_2Y_5$ .

### Additional Measurements

At the time of commencement of these two trials, the following vegetative and reproductive characters were measured or counted by sending a climber up in the tree:

- $x_1$  = Trunk girth just below the crown ( mean of two measurements taken one foot apart.
- $x_2$  = Number of open inflorescences
- $x_3$  = Number of nuts per bunch (averaged over all opened inflorescences and mature bunches).
- $x_4$  = Number of green fronds present.

The capacity for the trees to produce different yields is dependent upon these

characters. If we consider the concept of "Plant Vigour" it would be reasonable to expect that more often than not a more vigorous palm will have a large trunk girth ( $x_1$ ), more bunches ( $x_2$ ), more nuts per bunch ( $x_3$ ), and more fronds ( $x_4$ ).

In order to represent the plant vigour an 'Integrated Index-(I)' was developed through the use of Principal Component Analysis of the correlation matrix derived from the four variables  $x_1$  to  $x_4$  and select the Eigen Vector wherein the vector elements, corresponding to the four variables, all show positive signs and are nearly of equal dimensions. In this case the vector elements should approximate to 0.5, 0.5, 0.5, 0.5 if plant vigour is to reflect in equal measure by all four characters. The integrated index for each plant was used as a calibrating variate and compared with the use of variables  $x_1$  to  $x_4$  in reducing the experimental error in field experiments on coconut.

### Integrated Index (I)

The Eigen vectors of the four characters and Eigen values of the four vectors were evaluated through Principal Component Analysis. The standardized vector scores for each palm in respect of the vector would indicate how 'Vigorous' a given palm is. This standardized vector score is taken as the Integrated Index (I) for each palm. The average of these values for a given experimental plot (8 palms) is taken as the Integrated Index for each plot and was used as the calibrating variate. The value  $I_i$  for each palm is given by;

$$I_i = \sum a_j z_{ij}$$

where,  $z_{ij} = (x_{ij} - \bar{x}_j)/s_x$   
and  $\sum a_j = 1$

### Gain in Precision due to Reduction in Error

The precision of an experiment is given by the coefficient of variation (CV). Reduction in the coefficient of variation measures the gain in precision. As stated by Cochran (1957) the gain in precision through the use of, covariance analysis depends on the absolute size of the "Within group correlation coefficient". Measure of this correlation coefficient ( $r$ ) could be considered as "Relative Calibrating Efficiency" (RCE).

## RESULTS

### Principal Component Analysis

The principal component analysis of the correlation matrix of the four vegetative and reproductive characters ( $x_1$  to  $x_4$ ) gave rise to the Eigen vectors and Eigen roots of the two experiments (Tables 1 and 2).

**Table 1. Eigen Vectors and Eigen Values of the Vectors (Exp-1)**

Vector	Vector Elements ( $a_i$ )				Eigen value	Proportion of variance accounted by the vector
	Trunk girth	No. of bunches	No. of nuts per bunch	No. of fronds		
	$a_1$	$a_2$	$a_3$	$a_4$		
$V_1$	0.3032	0.5067	0.5734	0.5680	2.1961	54.9 %
$V_2$	0.8736	-0.4524	-0.1536	0.0923	0.9598	24.0 %
$V_3$	0.3770	0.6794	-0.2128	-0.5925	0.4544	11.4 %
$V_4$	-0.0523	0.2776	-0.7761	0.5638	0.3897	9.7 %

**Table 2. Eigen Vectors and Eigen Values of the Vectors (Exp-2)**

Vector	Vector Elements ( $a_j$ )				Eigen value	Proportion of variance accounted by the vector
	Trunk girth $a_1$	No. of bunches $a_2$	No. of nuts per bunch $a_3$	No. of fronds $a_4$		
$V_1$	0.4123	0.4810	0.5624	0.5313	1.9859	49.6 %
$V_2$	0.7347	-0.5638	-0.2873	0.2445	0.8856	22.1 %
$V_3$	0.5266	0.5052	-0.2009	-0.6535	0.6421	16.1 %
$V_4$	-0.1135	0.4421	-0.7488	0.4805	0.4864	12.2 %

Exp-1

Table 3. *Within group Correlation Coefficients between yield vs  $x_1$  to  $x_4$  and I.*

Experimental Year	Girth	No. of bunches	No. of nuts per bunch	No. of fronds	Integrated Index
	$x_1$	$x_2$	$x_3$	$x_4$	I
$E_1Y_1$	-0.0565	0.2184	0.1649	0.1501	0.1990
$E_1Y_2$	0.2406	0.1873	0.7049***	0.5593**	0.8661***
$E_1Y_3$	0.2439	0.0465	0.4099*	0.2799	0.5679**
$E_1Y_4$	0.2718	0.3135	0.4823*	0.4773*	0.5520**
$E_1Y_5$	0.3065	0.0447	0.4533*	0.2963	0.6362**

\*p = 0.05    \*\*p = 0.01    \*\*\*p = 0.001

Exp-2

Table 4. *Within group Correlation Coefficients between yield vs  $x_1$  to  $x_4$  and I.*

Experimental Year	Girth	No. of bunches	No. of nuts per bunch	No. of fronds	Integrated Index
	$x_1$	$x_2$	$x_3$	$x_4$	I
$E_2Y_1$	0.0409	-0.0979	-0.3188	-0.0094	-0.1222
$E_2Y_2$	0.5617**	0.2739	0.7755***	0.6458**	0.8745***
$E_2Y_3$	0.5929**	0.4456	0.8291***	0.7939***	0.8677***
$E_2Y_4$	0.6321**	0.3077	0.6469**	0.5052*	0.5761**
$E_2Y_5$	0.2156	-0.1206	-0.0800	0.1641	0.0722

\* $p=0.05$

\*\* $p=0.01$

\*\*\* $p=0.001$

Exp-1

Table 5. *Coefficient of variation (%) for the yield without and with adjustment by the Covariates  $x_1$  to  $x_4$  and I*

Exp. Year	Without adjustment	Adjustment with Covariates				
		$x_1$	$x_2$	$x_3$	$x_4$	I
$E_1Y_1$	8.96	9.18	8.97	9.07	9.09	9.01
$E_1Y_2$	21.02	20.93	21.18	15.30	17.88	10.78
$E_1Y_3$	17.06	16.97	17.48	15.96	16.80	10.40
$E_1Y_4$	12.34	12.18	12.02	11.09	11.12	9.56
$E_1Y_5$	9.96	9.73	10.21	9.11	9.76	7.89

Table 6. *Coefficient of Variation (%) for the yield without and with after adjustment by the variates  $x_1$  to  $x_4$  and I*

Exp. year	Without adjustment	Adjustment with Covariates				
		$x_1$	$x_2$	$x_3$	$x_4$	I
$E_2Y_1$	8.35	8.58	8.55	8.14	8.59	8.02
$E_2Y_2$	14.28	12.16	14.13	9.28	11.21	7.13
$E_2Y_3$	15.50	12.85	14.28	8.92	9.71	7.93
$E_2Y_4$	8.99	7.17	8.80	7.56	7.98	7.05
$E_2Y_5$	9.33	9.37	9.53	9.57	9.47	9.07

**Table 7.** *Average yields for the Control Treatment and adjusted & unadjusted CV values over the trial period*

Exp.	Exp. Year	No. of Nuts/palm (control)	CV values (%)		
			Unadj.	Adj.(x <sub>3</sub> )	Adj.(I)
Exp-1	E <sub>1</sub> Y <sub>1</sub>	59.7	8.96	9.17	9.01
	E <sub>1</sub> Y <sub>2</sub>	44.3	21.02	15.30	10.78
	E <sub>1</sub> Y <sub>3</sub>	32.3	17.06	15.96	10.40
	E <sub>1</sub> Y <sub>4</sub>	72.3	12.34	11.09	9.56
	E <sub>1</sub> Y <sub>5</sub>	63.4	9.96	9.11	7.89
Exp-2	E <sub>2</sub> Y <sub>1</sub>	53.4	8.35	8.14	8.02
	E <sub>2</sub> Y <sub>2</sub>	36.3	14.28	9.28	7.13
	E <sub>2</sub> Y <sub>3</sub>	29.8	15.50	8.92	7.93
	E <sub>2</sub> Y <sub>4</sub>	71.6	8.99	7.56	7.05
	E <sub>2</sub> Y <sub>5</sub>	68.6	9.33	9.57	9.07

For the two experiments the first vector ( $V_1$ ) has the Vector Elements all positive and each approximating to 0.5. This vector also explains 54.9 % and 49.6 % of total variation for the two experiments. Thus for all intents and purposes this vector reasonably fulfills the requirement of an integrated expression of the correspondence between these four characters and may be construed as an index of "Plant Vigour".

The integrated index for the  $i$ th palm is;

$$\text{Exp.1 } I_i = (0.3032)(x_{i1} - x_i)/s_x + (0.5067)(x_{i2} - x_2)/s_x + \dots$$

$$\text{Exp.2 } I_i = (0.4123)(x_{i1} - x_i)/s_x + (0.4810)(x_{i2} - x_2)/s_x + \dots$$

The average of  $I_i$  over 8 experimental palms for each plot was taken as a calibrating variate.

### Relative Calibrating Efficiency (RCE).

The within group correlation coefficients between characters,  $x_1$  to  $x_4$ , the Integrated Index (I), with the experimental yields are shown in tables 3 and 4, for the two experiments.

for the year  $E_1Y_1$  neither of the calibrating variates showed any significant within group correlation with the yield. The variates  $x_1$  and  $x_2$  showed no relation with the yield for all the experimental years. The variate  $x_4$  showed a relationship only for the experimental years  $E_1Y_2$  and  $E_1Y_4$ . However the two calibrating variates  $x_3$  and the Integrated Index showed significant correlations from  $E_1Y_2$  onwards. The Index 'I' showed superiority over  $x_3$  in respect of the relative calibrating efficiency. None of the calibrating variates showed any significant relation between the yield for the experimental years  $E_2Y_1$  and  $E_2Y_5$ . The variable  $x_2$  showed no relation for all the years. The variates  $x_1, x_3, x_4$ , and 'I' showed significant relationships for the experimental years  $E_2Y_2, E_2Y_3$ , and  $E_2Y_4$ . The Integrated Index 'I' showed superiority over all the variates in respect of relative calibrating efficiency.

### Absolute Efficiency

The coefficients of variation for the yield without adjustment and with adjustment using all the calibrating variates are shown in Tables 5 and 6.

The yield without adjustment showed a very low Coefficient of Variation (8.96%) for the experimental year  $E_1Y_1$  and with hardly any difference shown for the calibrating variates  $x_1$  to  $x_4$  and 'I'. The range being 8.97 to 9.07%. The yield showed high variabilities of 21.02 and 17.02% for the experimental years  $E_1Y_2$  and  $E_1Y_3$ , respectively. Considerable reduction in variation was observed for using  $x_3$  and 'I' as calibrating variates. The index 'I' showed the lowest variabilities of 10.78 and 10.40%. The experimental year  $E_1Y_4$  showed a CV of 12.34% for the yield without

adjustment and a reduced CV of 9.56% was indicated when adjusted with the index 'I'. Very low CV values were observed for the year  $E_1Y_5$  for the yield without adjustment and for the calibrating variates; the lowest being for 'I'. The range of values was 7.89 to 10.21%.

### Exp-2

Very low coefficients of variation was shown for the experimental years  $E_2Y_1$ ,  $E_2Y_2$  and  $E_2Y_5$  for the yield without adjustment and also when adjusted with the calibrating variates. The range being 7.05 to 9.57%. The values for the integrated index 'I' showed the lowest values.

Experimental years  $E_2Y_2$  and  $E_2Y_5$  showed high CV values (14.28 & 15.50%) for the yield without adjustment. However with adjustment with the integrated index 'I', the result showed values of CV being reduced to as low as 7.13 and 7.93% respectively.

## DISCUSSION

Conducting of an experiment is not without complications. The problems are more when it is a field experiment on a perennial crop where the material used has a large genetic variation. In this respect field experiments on coconut palm is no exception. Various efforts are made by the Experimenter and the Biometrician in controlling variation and evaluating results so as to achieve the objectives of the experiments in it's best form.

The reduction of the experimental error can be ensured very simply by experimenting with uniform experimental material and under uniform conditions. Well, these are billiard table experiments. In field experiments on biological material, this approach to reducing experimental error is wellnigh impossible. Therefore, for the experiment to be of any practical value, it is necessary to experiment within the natural framework subject to what ever variation there is. The philosophy should be "how can an experimenter do better when faced with fearful odds".

The pioneering attempt in controlling experimental error was through local control with adequate replication. This was achieved through intra-block homogeneity and inter block heterogeneity. However with time it was realized that controlling experimental error in this manner in perennial crops, more often than not was so efficient. It was at this time that the concept of "Calibration" came into being. (Pearce 1953).

In case of coconut experiments the reduction of experimental error was achieved through local control assisted by the use of pre-experimental yield data as a

calibrating variate. A certain degree of efficiency was achieved through this methodology. However the collection of yield has to be done over two years or at least one year before the commencement of the experiment. This caused concern over many aspects; cost, time etc. It is also a well gathered information that there is a high degree of natural variation due to changes in the weather pattern. Better the rainfall distribution in a given year high is the crop in the subsequent year and vice versa. This is uncontrollable. The fig (1), which is self explanatory will indicate the kind of the yield variation that is generally observed.

The field experiments on coconut are carried out under such conditions of variation. Further it is also observed different experimental errors in different years. Abeyasinghe (1986) reported that high variability is associated with high yielding years. However the results of the experiments reported in here suggest a different status (Table 7). The table,

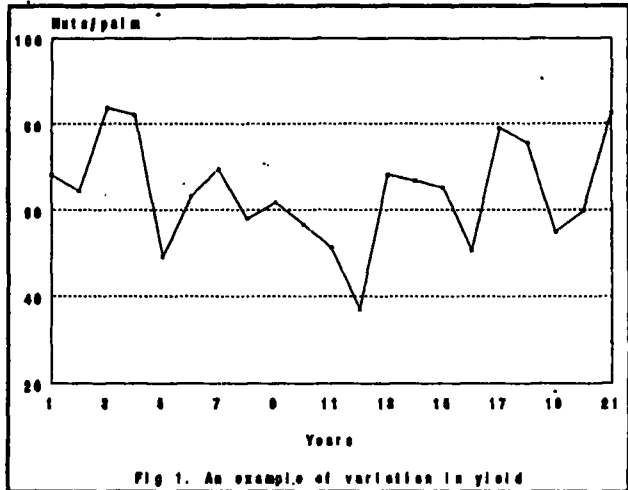


Fig 1. An example of variation in yield

shows the average yields for the control treatment and the adjusted & unadjusted CV values for the two experiments over the trial period. For all intents and purposes this average yield indicates the status quo of the weather pattern of the previous year. For the two experiments the years 1987 and 1988 are two poor crop years. It is evident that high variation (CV) is observed when the yields of the crop year are poor.

These results, indicate that no general claim can be made as to the nature of the variation one would observe at the end of the year, when the results are analysed. It is apparent that every experiment is unique and has to be treated in it's own right. What is more important and the prime objective, is to reduce the error through the choice of a calibrating variate(s) when high variability is observed. Thus prior arming with suitable calibrating variate(s) will be to the best advantage of the Experiment.

In view of these expressed and silent constrains and for other good reasons some vegetative and reproductive characters were measured and counted and used as calibrating variates independently and in association through the use of Principal Component Analysis. The characters were, trunk girth( $x_1$ ), number of open inflorescences( $x_2$ ), number of nuts per bunch( $x_3$ ), and number of green fronds( $x_4$ ). The

'Integrated Index-I' developed through the use of Principal Component Analysis using the characters  $x_1$  to  $x_4$ , reflects the "Vigour of the Palm".

Amongst the characters the number of bunches( $x_2$ ) was a poor calibrating variate to consider as an independent variable. However it was a credible variate to be considered in association with the other variables. This was shown by "Vector Elements" in Principal Component Analysis. The two variables  $x_1$  and  $x_4$  taken independently were marginal in reducing experimental error. Among the four variables,  $x_3$  (no. of nuts per bunch) showed superiority in reducing the experimental error taken independently.

In the Integrated Index developed through the use of Principal Component Analysis, the four variables have shown similar strength in equal measure. This fact is well proved in reducing experimental error when 'I' was used as a calibrating variate. Taking the results of the two experiments together, the overall range in the CV values when adjusted for 'I' was 7.05 to 10.78% as compared to the range 7.56 to 15.96% when adjusted with  $x_3$  as a calibrating variate. The range observed without adjustment was 8.35 to 21.02%.

The results (Tables 5 & 6) indicated that out of 10 experimental situations of analyses in the two experiments, 5 situations showed high variability. When the variation was low without adjustment, the error reduction through the use of 'I' was very low or marginal. It may be due to that, there is a limit below which experimental error cannot be reduced. To maintain a variation (CV) of the range 8 to 10% in field experiments for coconut can be considered as excellent.

In situations when high variations were observed, under unadjusted conditions, the error reduction brought about through the use of 'I' was very commendable. Thus the end result of this study was that the 'Integrated Index-I' developed through the use of Principal Component Analysis could be effectively and efficiently used in reducing experimental error in field experiments in coconut. With little effort the 'Integrated Index-I' could be evaluated with the kind of computers around. The recording of the variables  $x_1$  to  $x_4$  could be done within 3 to 4 days before or on commencement of the trial. Thus recording and the use of 'I' is further cost effective and less time consuming.

## REFERENCES

- Abeyasinghe, Tilak. (1986). Calibration of experiments on perennial crops using covariance analysis: The case of coconut. *Expl. Agric.* 22: 353-361.
- Abeywardana, V. (1970). The efficiency of pre-experimental yield in the calibration of coconut experiments. *Ceylon Coconut Quarterly.* 21: 85-91.

- Cochran, W. G. (1957). Analysis of covariance: its nature and uses. *Biometrics*. 13 : 261-81.
- Moore, C. S. and Pearce, S. C. (1976). Methods of reducing crop variation in apple trials. *J. Hort. Sci.* 51: 533-544.
- Pearce, S. C. (1953). *Tech. Commun. Commonw. Bur. Hort. Plant. N. Crops* No. 23.
- Pearce, S. C. and Brown, A. H. F. (1960). Improving fruit tree experiments by a preliminary study of the tree. *J. Hort. Sci.* 35: 56-65.
- Pieris, W. V. D. and Salgado M. L. M. (1937). Experimental error in field experiments with coconut. *Tropical Agriculture, Trinidad* 89: 75-85.
- Vernon, A. J. and Morris M. G. (1964). Variability in West African peasant cocoa farms; Time studies on agodi cocoa yields. *J. Hort. Sci.* 39: 241-253.