IV Conference internationale sur les recherches cacaoyères
8-18 janvier 1972. St Augustine. Trinidad

POSSIBILITES D'AMELIORATION DE L'EXPERIMENTATION SUR CACAOYERS

R. LOTODE
I.F.C.C. Cameroun

SUMMARY

The failures often experienced in cacao experimentation have led to a thorough examination of certain points in the experimental designs and techniques of analysing the results.

It is a question of :—

— reducing to the greatest extent the variable factors other than those of which it is desired to measure the effects.
— determining the optimum plot size which can depend on various circumstances.
— bringing into the analysis of variance those factors whose effect can be measured.
— being certain of the validity of the analysis of variance through a proper transformation of the data.

For these studies, the existing records of the Cocoa Research Station at Nkoemvone (Cameroun) from 1950 to 1960 have been used.

Work was carried out on :—

— the optimum plot sie and the precision of any trial according to the design used.
— the correlation between the yield of neighbouring plots in order to measure the effect on a plot of its neighbours.
— the relation between trunk diameter at a given age, or better, the increased in diameter between two ages and the final cumulative yield.

The results follow:—

— regarding plot size, two situations must be considered:—
  1) Where guard rows are not used, single tree plots can be used.
  2) Where guard rows are necessary a plot size of 12 - 25 trees is recommended. 25 trees would be chosen when the area is not limited, staff and money are adequate. If funds are limiting, the number of replications may be reduced, but at the expense of the precision of the experiment. If the area is limited, 12 tree plots would be used. In choosing an experimental design, the standard method should be used.
— The residual error can be reduced in the analysis by using additional measurements such as: the number and date of loss of tree in each plot, measurements of the trunk at 20 cm from the ground taken at one or two years of age before this factor is influenced by the treatments. if this is done, the analysis of these measurements allows a rapid sorting out of the material in the experiment.

RESUMÉ

Les échecs souvent enregistrés dans les expérimentations sur cacaoyers ont conduit à approfondir certains points des dispositifs expérimentaux et des techniques d'analyse des résultats.

Il s'agit :

— de réduire au maximum les facteurs de variation autres que ceux dont on veut mesurer les effets;
— de déterminer une taille de parcelle élémentaire optimum, qui pourra dépendre des circonstances;
— de faire intervenir dans l'analyse de variance des facteurs dont l'action pourra être mesurée'
— de s'assurer de la validité de l'analyse de variance par une transformation adéquate des données.

Pour ces études, ont été utilisées les fiches établies à la Station de recherches sur le cacaoyer de Nkoemvone (Cameroun), de 1950 à 1960, sur chacun des cacaoyers d'une importante collection.

Les travaux ont porté sur :

— la taille optimum des parcelles élémentaires et la précision d'un essai suivant le dispositif adopté;
— la corrélation entre les productions de parcelles contiguës afin de juger de l'intérêt de l'adjonction d'une parcelle témoin a chaque parcelle traitée;
— la liaison entre les diamètres des troncs a un âge donne ou mieux l'accroissement de diamètre entre deux âges donnés, et la production cumulée ultérieure.

Les résultats positifs sont les suivants:

— En ce qui concerne la taille des parcelles élémentaires utiles, il faut distinguer deux cas :
  ■ lorsque les lignes de bordures sont inutiles, on doit adopter la randomisation totale pied par pied. Elle peut s'effectuer, si un contrôle de l'hétérogénéité du champ d'essai est possible, par une stratification en blocs à l'intérieur de chacun d'eux.
  ■ lorsque les lignes de bordures sont nécessaires, une fourchette de 12-25 arbres est recommandée : on opte pour 25 arbres lorsque la surface n'est par limitée, que le personnel et les crédits soient suffisants ou insurrisants, mais dans ce dernier cas le nombre le répétitions étant limité, la précision de l'essai s'en trouve réduite. Si la surface est limitée, on opte pour 12 arbres quand le personnel et les crédits sont suffisants et on se place au milieu de la fourchette lorsque le personnel et les crédits sont limités.

— En ce qui concern le choix du dispositif expérimental, il faut adopter le schéma classique, sans adjonction d'une parcelle témoin contiguë. Le résidu aléatoire pourra être diminué en faisant intervenir dans l'analyse des facteurs dont l'action pourra être mesurée, tels que: nombre et date d'apparition des manquants pour chaque parcelle élémentaire; mensurations du tronc à 20cm du collet effectuées après un et deux ans de plantation, quand cette variable n'est pas influencée par les traitements. Si elle l'est, l'analyse de ces mensurations permet un tri rapide du matériel en expérimentation./—

170

Les nombreux échecs enregistrés dans l'expérimentation sur cacaoyers ont amené a approfondir certains problèmes concernant les dispositifs expérimentaux et les techniques d'analyse des résultats.

Il s'agit:

— de réduire au maximum les facteurs de variation autres que ceux dont on veut mesurer les effets. Nous heutons là à l'hétérogénéité du sol dans le détail, de l'ombrage maintenu ou même installé, du matériel végétal. La plus grande homogénéité possible doit être, bien entendu, recherchée dans tous les domaines autres que ceux étudiés;

— de déterminer une taille de parcelle élémentaire optimum qui pourra dépendre de certaines circonstances (présence ou absence de lignes de bordures autour des parcelles élémentaires, nécessité dans certains cas de l'effet de masse, disponibilités en surface, personnel et crédit), mais qui donnera dans chaque cas le maximum de précision;

— de faire intervenir dans l'analyse de variance des facteurs dont l'action pourra être mesurée et, si elle est significative, viendra réduire le résidu aléatoire et donc augmenter la précision;

— de s'assurer de la validité de l'analyse de variance par une transformation adéquate des données.

A La Station de Nkoemvone (200 km au Sud de Yaoundé), les chercheurs chargés entre autre des problèmes de sélection, avaient dans une première étape, constitué une collection importante de cacaoyers qui ont été suivis individuellement pendant dix ans (1950/1960). Le fichier résultant de cette étude a constitué un document précieux qui a permis:

— une étude sur la taille optimum des parcelles élémentaires et la précision d'un essai suivant le dispositif adopté;

— une étude de la corrélation entre productions de parcelle contiguës afin de voir si l'adjonction d'une parcelle témoin à chaque parcelle traitée (permettant la technique de convariance) est bénéfique.

En outre:

— le problème des manquants a été abordé;

— les données d'un essai comparatif d'hybrides à la Station de Nkoemvone ont montré la liaison qui existe entre les diamètres des troncs à un âge donné et la production cumulée ultérieure, permettant, dans certains cas, la réduction du résidu aléatoire.

DETERMINATION DE LA TAILLE DES PARCELLES ELEMENTAIRES

Les données de base sont les cumuls des productions individuelles dix ans après la plantation. Notons que la collection est installée sous forêt secondaire aménagée.

L'étude porte sur les données de cinq champs (parcelles 21 à 25 du répertoire de Nkoemvone). Dans chacun d'eux, des parcelles élémentaires comprenant: 1, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 64, 96 arbres ont été délimités. Les parcelles ont été consitutées de la façon la plus compacte possible. Elles n'ont pas une

	Champ I-Moy. gle :4.925 (576 arbres)			Champ II-Moy. gle:5.705 (420 arbres)			Champ III-Moy. gle:6.285 (332 arbres)		
Taille de parcelles	Variance données brutes (1)	Variance données trans-formées	d.d.l.	Variance données brutes (1)	Variance données trans-formées	d.d.l.	Variance données brutes (1)	Variance données trans-formées	d.d.l.
1	13.452	0,18091	575	27.679	0,16187	419	23.423	0,18084	331
4	8.009	0.09294	143	17.413	0,08164	104	13.275	0,06105	75
8	5.857	0,06401	71	12.527	0,05604	51	8.682	0,04245	37
12	4.416	0,05326	47	10.266	0,04617	34	8.778	0,04045	24
16	4.163	0,04790	35	11.371	0,04931	25	7.646	0,03662	17
20	3.626	0,04346	27	8.944	0,03843	20	7.520	0,03303	13
24	3.644	0,04497	23	12.232	0,04253	16	7.558	0,03322	11
28	3.858	0,04144	19	10.080	0,03744	14	6.727	0,02514	9
32	2.705	0,02823	17	5.625	0,02991	12	7.780	0,02854	8
36	3.169	0,03248	15	7.414	0,02805	10	6.323	0,02160	7
40	2.519	0,02668	13	4.450	0,02072	9	7.176	0,02986	6
64	2.631	0,03829	8	5.595	0,02384	5	5.143	0,02144	3
96	2.271	0,02199	5	2.029	0,01233	3	7.042	0,02831	2
3.5	3.5	3.5	3.5	3.5	3.5	3.5			

- Données exprimées en grammes de fèves fraîches.

(1) Les variances des données brutes sont experimées en milliers de grammes.

ddl = Degrés de liberté.-

forme parfaitement régulière du fait de la présence de manaquants.

Les variances des moyennes parcellaires brutes et transformées en logarithme (base 10) sont données dans le tableau joint.

A la lecture du tableau, on constate:

données brutes:

Les variances pour des parcelles de même dimension varient nettement d'un champ à l'autre: le rapport des variances extrêmes se place entre 2, 1 et 3, 4 suivant la taille parcellaire. Elles sont significativement différentes:

données transformées:

Les variances pour des parcelles de même dimension, varient peu d'un champ a l'autre. Si on ne considère que celles calculées avec un nombre de degrés de liberté suffisant, le rapport des valeurs extrêmes varie suivant la dimension des parcelles de 1,33 à 1,65. Elles ne sont plus significativement différentes.

Ceci montre une fois de plus, s'il en était encore besoin, l'intérêt de la transformation logarithmique.

Quel que soit le niveau de production des blocs, ou, quel que soit le nombre d'années intervenant dans le cumul (a partir d'un certain minimum toutefois, les deux premières années de production montrant une variabilité telle qu'on ne peut valablement les analyser) la variance du résidu aléatoire sur la moyenne parcellaire (pour une taille donnée) devient constante. Cela va nous permettre d'étudier d'une façon générale la précision d'un essai de type donné.

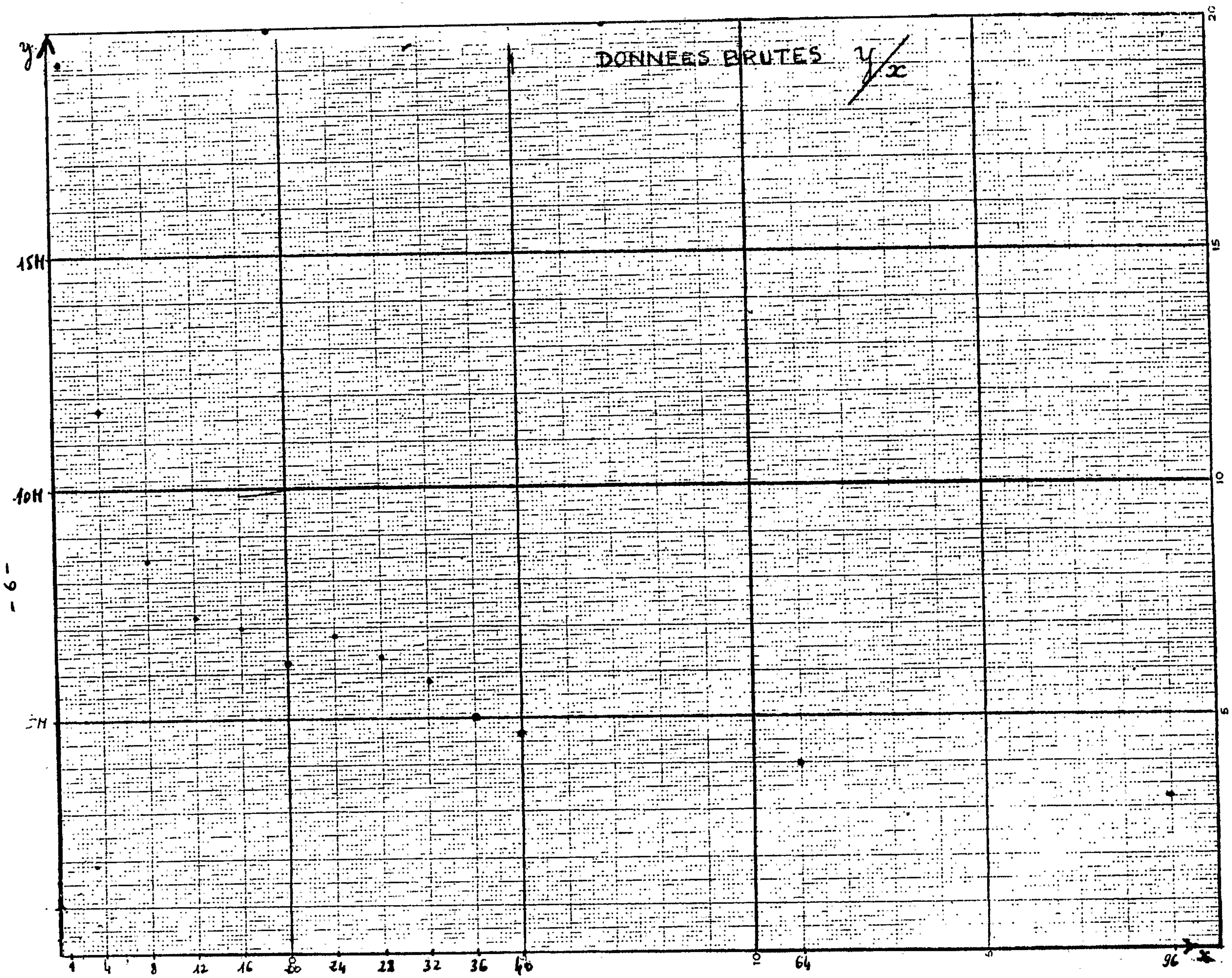
ETUDE DE L'EVOLUTION DES VARIANCES DES DONNEES BRUTES EN FONCTION DE LA TAILLE DES PARCELLES

Plaçons sur un graphique en abcisse les tailles de parcelles (x), en ordonnée la variance de la moyenne parcellaire correspondante (y). Les points correspondants suggerent une relation hyperbolique nette de la forme $y = \frac{a}{x} + b$.

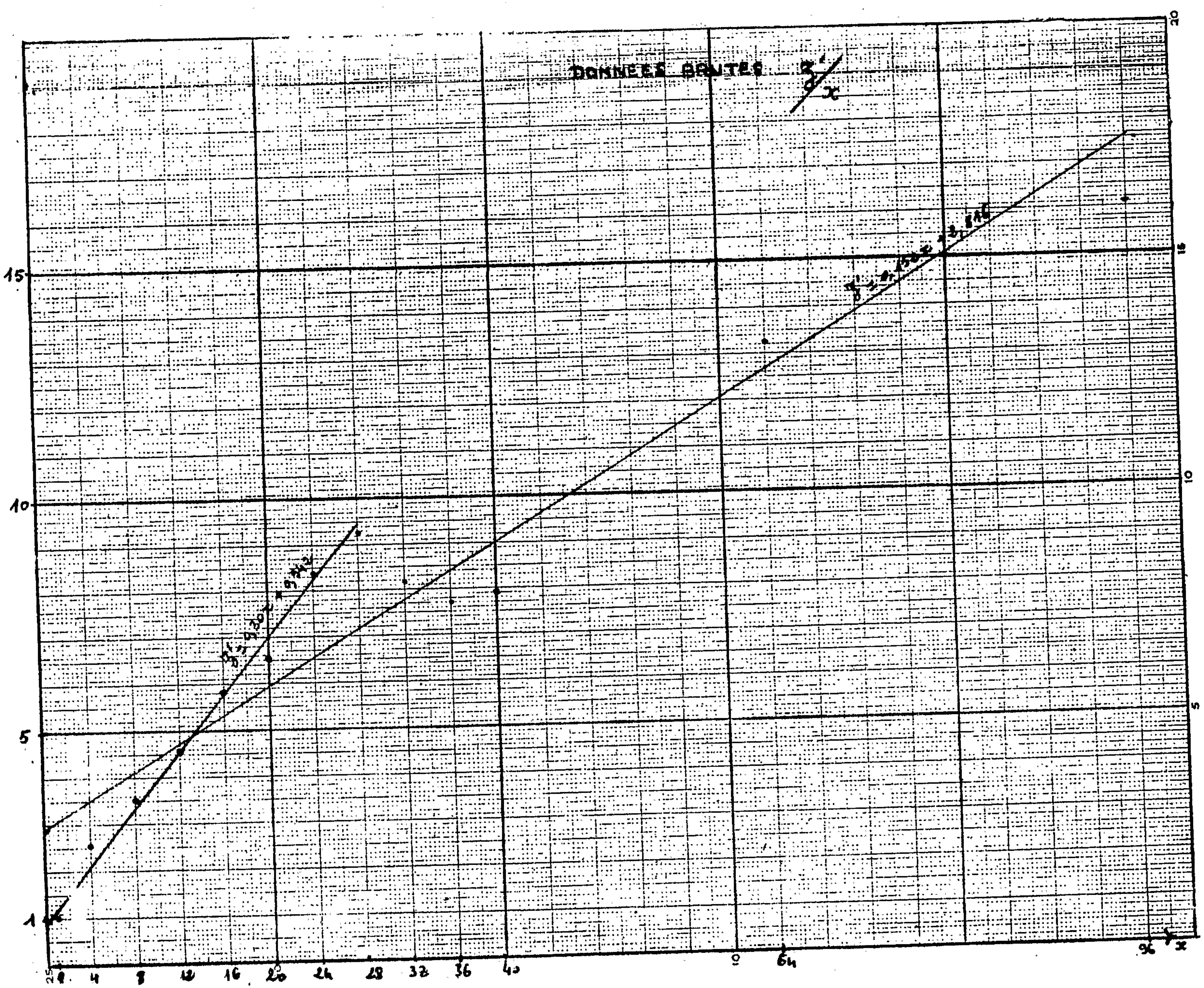
Pour la vérifier, adoptons en ordonnée la variable $z = y.x$. Si la relation précédente est vérifiée z doit être une fonction linéaire de x. ($z = a + bx$).

L'étude de la régression nous permettra alors de définir les coefficients a et b. Pour simplifier les calculs.

Champ IV – Moy.gle: 4.410 (377 arbres)			Champ V – Moy.gle: 6.590 (303 arbres)			Ensemble des champs: (2.008 arbres) Moy.gle don. brutes: 5.477 Moy.gle don. transf.: 3,57971.			
Variance données brutes(1)	Variance données transformées	d.d.l.	Variance données brutes(1)	Variance données transformées	d.d.l.	Variance donnés drutes	Variance données transformées	d.d.l.	C.V.
13.047	0,18211	376	17.342	0.13661	302	19.156	0,17558	2.007	11,7 %
7.448	0,08686	86	9.604	0,06014	72	11.539	0,08389	484	8.09 %
5.366	0,06219	42	8.144	0,05010	35	8.465	0,05996	240	6,85 %
4.499	0,05439	28	6.554	0,04463	23	7.183	0,05181	160	6,36 %
4.832	0,05296	20	6.039	0,03483	17	6.975	0,04772	118	6,10 %
3.583	0,03216	16	6.124	0,03325	13	6.194	0,03997	93	5,59 %
3.728	0,03828	13	6.780	0,04191	11	6.789	0,04229	78	5,74 %
3.799	0,04151	11	5.246	0,02867	9	6.292	0,03845	66	5,48 %
3.194	0,04189	9	4.534	0,02903	8	4.826	0,03311	58	5,08 %
4.222	0,04918	8	4.891	0,02976	7	4.039	0,03497	51	5,22 %
3.211	0,02655	7	5.179	0,03723	6	3.714	0,02823	45	4,69 %
2.326	0,01536	4	4.976	0,02588	3	3.996	0,02822	27	4,69 %
2.648	0,02405	2	4.557	0,02325	2	3.227	0,02079	18	4,03 %



655


DONNEES BRUTES
15
10
5
1
4
8
12
16
20
24
28
32
36
40
0
64
96
x
20
15
10
5

656

prenons comme unité en ordonnée, la valeur de la variance des données individuelles: 19.156 x 10^3. Nous obtenons le tableau suivant:

x	z	z'
1	19.156.431	1
4	46.154.168	2,409
8	67.717.768	3,535
12	86.200.620	4,500
16	111.592.512	5,825
20	123.888.800	6,467
24	162.928.512	8,505
28	176.189.972	9,197
32	154.424.384	8,061
36	145.402.672	7,590
40	148.553.040	7,755
64	255.760.896	13,351
96	309.760.032	16,170

Les points sont sensiblement alignés comme le montre le graphique (z'/x).

Effectivement:

coefficient de corrélation : r = 0,95 (11 ddl)

pente de la régression : b = 0,152

droite de régression : z' = 0,152 x + 2,816

$$y = 0{,}152 + \frac{2{,}816}{x}$$

L'ajustement à une hyperbole est bon et, comme l'ont montré H. MARTICOU et R.A. MULLER (Essai de mise au point d'une méthode d'expérimentation adaptée aux conditions de la cacaoyère camerounaise traditionnelle. Café-Cacao-Thé, No3 — Juillet/Septembre 1964, page 185), la variance de la moyenne parcellaire ne tend pas vers 0 quand la taille augmente indéfiniment, mais vers une valeur correspondant à 15% environ de la variance des données individuelles. Si la répartition des arbres à l'intérieur des plantations était réalisée strictement au hazard, c'est-à-dire si l'hypothèse de l'indépendence de la production d'arbres voisins était exacte, la variance y suivrait la loi générale théorique $y = \frac{b}{x}$ (b étant la variance des productions individuelles). Le coefficient a, introduit, chiffre le degré d'association des arbres de production comparable.

Notons que la valeur de a calculée pour l'ensemble des points (0,152) est très voisine de la valeur calculée par H. MARTICOU et R.A. MULLER (0,16) alors que les plantations étudiées sont différentes à beaucoup d'égards. C'est assez remarquable.

Si l'on reprend la formule obtenue: z = 0, 152 x + 2,816, on constate que l'ajustement n'est pas bon pour les premiers points. Si l'on examine le graphique, on s'aperçoit que la courbe passant au plus près des points montre un infléchissement vers l'axe des abcisses, c'est-à-dire, semble tendre vers une asymptote. Notons également que les dernières variances calculées avec un nombre réduit de degrés de liberté sont peu précises. Les huit premiers points sont presque parfaitement alignés:

r = 0.998 (6 ddl)

b = 0,30

z' = 0,30 x + 0,942

$$y = 0{,}30 + \frac{0{,}942}{x}$$

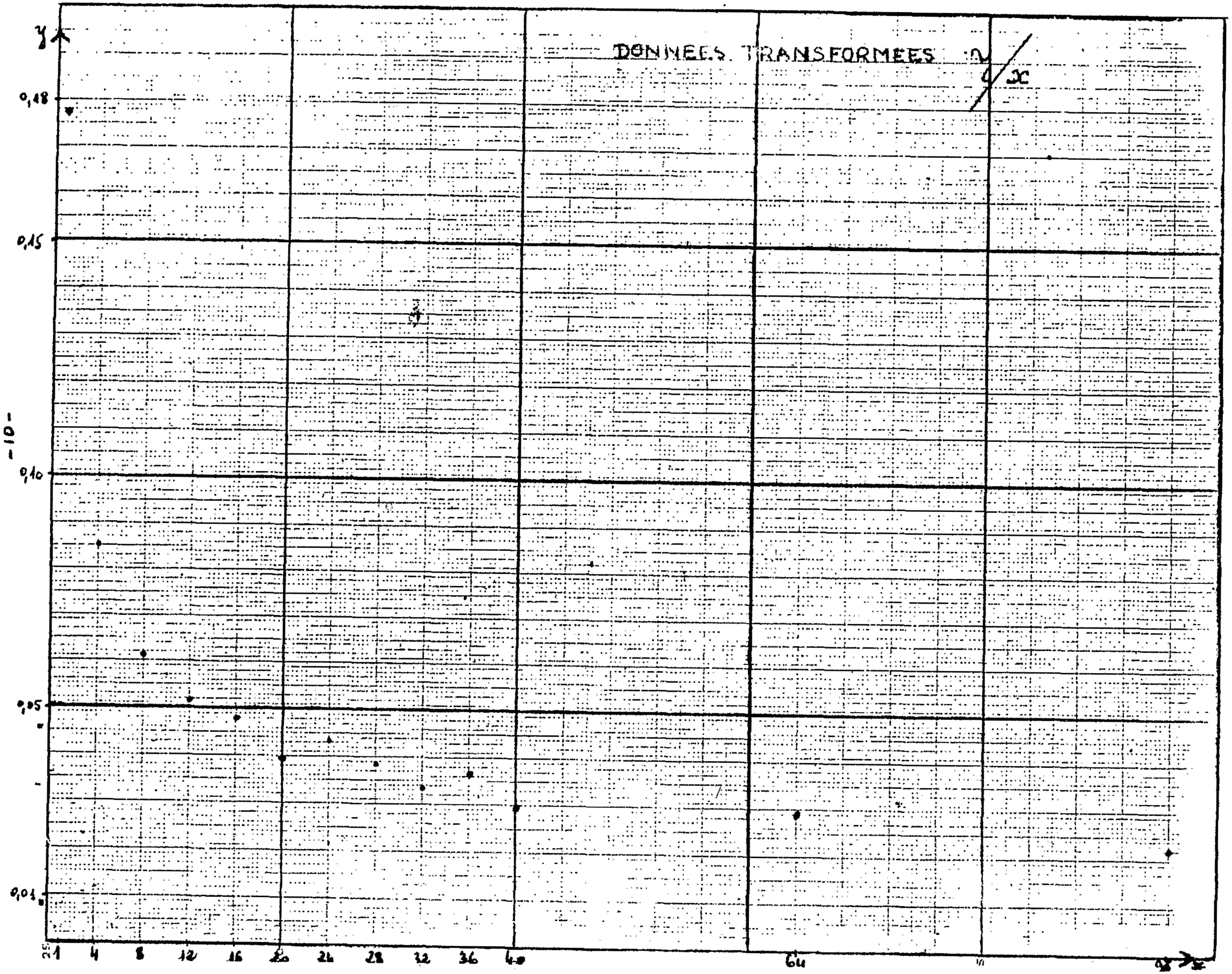
L'ajustement devient excellent. La valeur supérieure de b (0,30), dans de cas, indique l'influence importante de la liaison entre arbres de productions comparables jusqu'aux parcelles de 28 arbres environ. Elle diminue ensuite ce qui indique que l'influence de l'association des arbres à production comparable diminue quand la taille des parcelles augmente; c'est logique : les taches de fertilité indentique sont petites (hétérogénéité dans le détail du sol), de même l'influence d'un arbre d'ombrage particulier s'étend sur une zone de faible étendue.

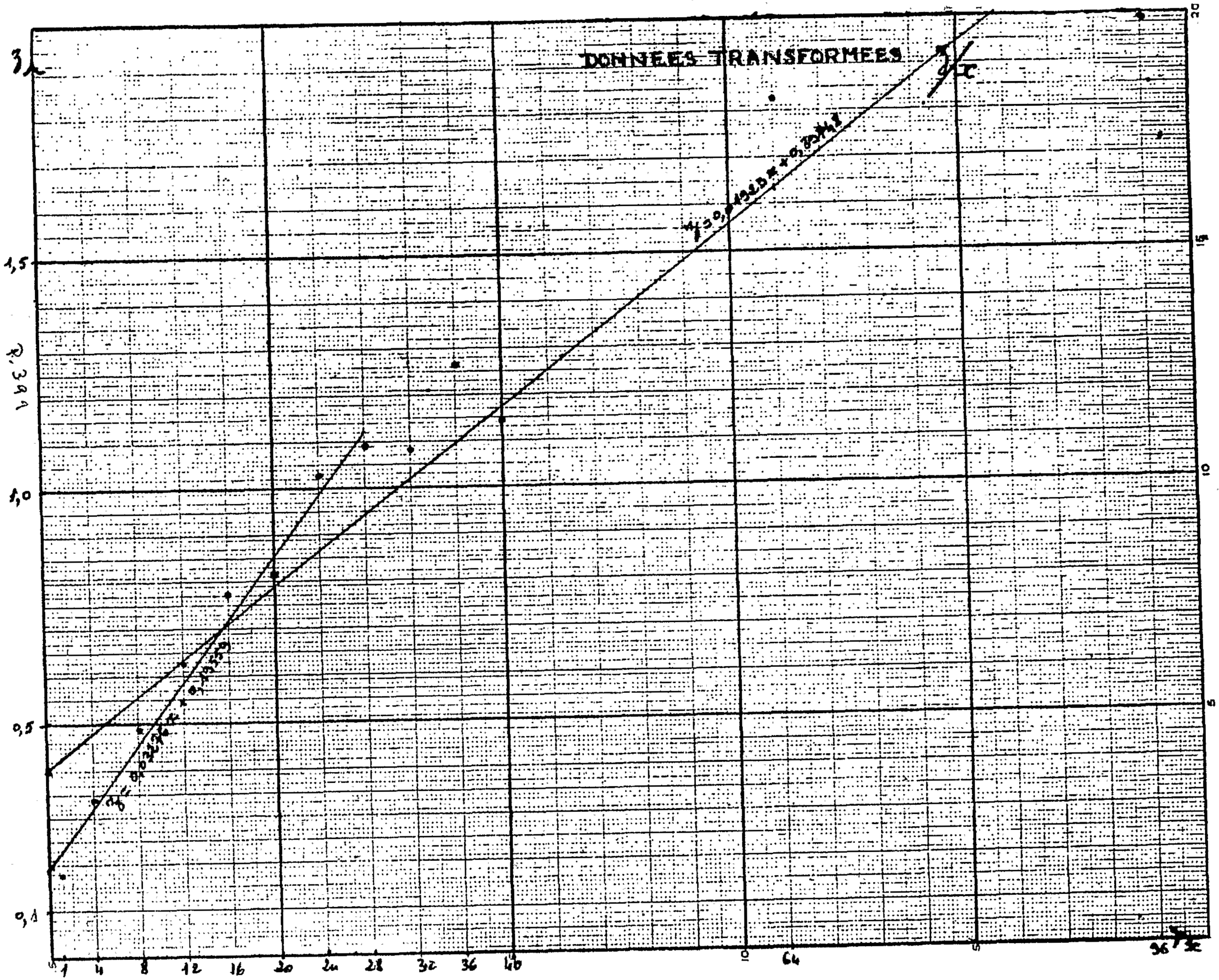
ÉTUDE DE L'EVOLUTION DES VARIANCES DES DONNÉES TRANSFORMÉES EN FONCTION DE LA TAILLE DES PARCELLES

Plaçons sur un graphique en abcisse les tailles de parcelles (x), en ordonnée la variance de la moyenne parcellaire transformée correspondante (y). Les points correspondants suggèrent là aussi une relation hyperbolique nette de la forme $y = \frac{a}{x} + b$.

Comme précédemment, étudions la régression linéaire devant exister entre z = y. x et x (voir graphique z/x données transformées). :

x	z
1	
4	0,17558
8	0,33556
12	0,47968
16	0,62172
20	0,76352
24	0,79940
28	1,01496


DONNEES TRANSFORMEES
y
0,18
0,15
0,10
0,05
0,01
1
4
8
12
16
20
24
28
32
36
40
64
x


DONNEES TRANSFORMEES
1,5
1,0
0,5
0,1
1
4
8
12
16
20
24
28
32
36
40
64
96

1,07674	32
1,05952	36
1,25892	40
1,12920	64
1,80608	96
1,99584	

r = 0,96 (ddl)

pente de la régression = 0,01929

z = 0,01929 x + 0,39748

$$y = 0{,}01929 + \frac{0{,}39748}{x}$$

L'ajustement a une hyperbole est correct et dans ce cas aussi, bien entendu, la variance ne tend pas vers 0 quand x augmente mais vers une valeur 0,01929 représentant 10% de la variance des données individuelles. Le degré d'association semble inférieur après transformation; ceci est dû tassement consécutif des distributions, a l'importance relative moins grande donnée aux valeurs extrêmes.

Plus nettement que sur les données brutes, on constate un infléchissement constant de la courbe ajustée au mieux aux points, vers l'axe des abcisses. Le degré d'association des arbres à production comparable voit son influence sur la variance diminuer avec l'accroissement de la taille parcellaire.

L'ajustement à une droite n'est véritablement excellent que pour les 8 premiers points : r = 0.99 (6 ddl)

pente de la régression = 0,03276

z = 0.03276 x + 0,19559

$$y = 0{,}03276 + \frac{0{,}19559}{x}$$

Les valeurs ajustées sont alors très voisines des valeurs calculées pour ces huit points:

	Valeur ajustée	Valeur calculée
x = 1	y = 0,22835	0,17558
x = 4	y = 0,08166	0,08389
x = 8	y = 0,05721	0,05996
x = 12	y = 0,04906	0,05181
x = 16	y = 0,04499	0,04772
x = 20	y = 0,04254	0,03997
x = 24	y = 0,04091	0,04229
x = 28	y = 0,03975	0,03845

Pratiquement, *à partir de 20/25 arbres, l'augmentation de la taille de la parcelle élémentaire n'amene qu'une diminution négligeable de la variance du résidu aléatoire et on n'a absolument aucun intérêt à aller au-delà.*

ÉTUDE DE LA PRECISION D'UN ESSAI SUIVANT LE DISPOSITIF ADOPTE

La précision de la moyenne générale sur l'ensemble des champs en observation ser d'autant plus grande que la variance de cette moyenne sera plus petite.

Supposons un ensemble comprenant N arbres. Posons-nous le problème suivant : quelle doit être la taille de la parcelle élémentaire pour que la précision sur la moyenne générale soit la plus grande étant donné les renseignements précédents?

1/ — *Cas où l'essai ne nécessite pas une ligne de bordure autour des parcelles élémentaires* (essais de variétés, d'hybrides, de clones).

On prendra en considération, pour les variances, les valeurs ajustées.

Parcelle élémentaire (Nbre d'arbres)	Nbre de parcelles élémentaires dans un champ e N arbres	Variance de la moyenne générale
1	N	$\frac{0,22835}{N}$
4	N/4	$\frac{0,08166}{N/4} = \frac{0,32664}{N}$
8	N/8	$\frac{0,05721}{N/8} = \frac{0,45768}{N}$
12	N/12	$\frac{0,04906}{N/12} = \frac{0,58872}{N}$
16	N/16	$\frac{0,04499}{N/16} = \frac{0,71984}{N}$
20	N/20	$\frac{0,04254}{N/20} = \frac{0,85082}{N}$
24	N/24	$\frac{0,04091}{N/24} = \frac{0,98184}{N}$
28	N/28	$\frac{0,03975}{N/28} = \frac{1,1130}{N}$

On constate que d'une manière très nette, la randomisation totale pied par pied est supérieure à tout autre système.

A titre d'exemple, comparons deux dispositifs installés dans un champ de N arbres:

— la randomisation totale;

— les blocs randomisés complets avec parcelle élémentaire de 20 arbres (système adopté à Nkoemvone pour les premiers essais comparatifs d'hybrides).

Soit d la plus petite différence significative entre deux moyennes/hybrides consécutives après classement par ordre de grandeur dans le 1er cas et d' dans le second cas.

$d = t \times \sqrt{2 \times \text{variance du résidual}}$ sur la moyenne (t de Student)

$$d = t \times \sqrt{2 \times \frac{0,22835}{N}}$$

$$d' = t \times \sqrt{2 \times \frac{0,85080}{N}}$$

$$\frac{d'}{d} = \sqrt{\frac{0,85080}{0,22835}} = 1,93.$$

En prenant les variances calculées au lieu de celles ajustées, on trouve $\frac{d'}{d} = 2,13$.

Dans le 2ème dispositif, la plus petite différence significative entre moyennes transformées doit être en gros, 2 fois supérieure à la différence nécessaire dans le 1er dispositif.

A quoi correspond pratiquement cette différence?

Soit M'1 et M'2 deux moyennes consécutives après classement dans le 2ème dispositif; M1 et M2 dans le 1er dispositif. Le rapport précédent devient:

$$\frac{\log M'2 - \log M1}{\log M2 - \log M1} \# 2$$

$$\log \frac{M'2}{M'1} \# \frac{\log \frac{M2}{M1}}{2} \# \log \left(\frac{M2}{M1}\right)^2$$

$$\frac{M'2}{M'1} \# \left(\frac{M2}{M1}\right)^2.$$

(Notons que les "M" sont les moyennes géométriques obtenues par retransformation simple, présentant donc un biais par rapport aux moyennes arithmétiques).

Si le rapport significatif dans le 1er système est 1,2 dans le deuxième, il sera 1,44 — De 1,3, il passera à 1,69. Les différences sont très importantes et la randomisation totale est à adopter quand les moyens le permettent bien entendu, surtout en personnel qualifié. Les essais de variétés, hybrides, clones devraient être établis suivant ce dispositif dans toutes les stations de recherches.

Si on estime pouvoir contrôler une partie de l'hétérogénéité du champ d'essai par une stratification en bloc, on adoptera le dispositif "randomisation totale par bloc" ce qui permettra:

— de réduire le résidu aléatoire;

— de continuer à bénéficier de l'avantage de la randomisation totale (variance de la moyenne/traitement minimum).

2/ — *Cas où l'essai nécesite une ligne de bordure autour des parcelles élémentaires* (essais d'engrais, de densité, d'ombrage, etc...)

Parcelle élémentaire utile	Parcelle élémentaire totale	Nbre de parcelles dans un champ de N arbres	Variance de la moyenne générale
1	9	N/ 9	2,055/N
4	16	N/16	1,307/N
6	20	N/20	1,307N
8	24	N/24	1,373/N
9	25	N/25	1,362/N
12	30	N/30	1,472/N
16	36	N/36	1,620/N
20	42	N/42	1,787/N
24	48	N/48	1,964/N
25	49	N/49	1,989/N
28	54	N/54	2,147/N

Théoriquement, la parcelle élémentaire de 6 arbres donne la précision maximum par unité de surface d'essai. Mais dans ce genre d'essai interviennent d'autres considérations:

a) *Effet de masse*

Dans un essai d'engrais par exemple, une modification des équilibres dans le sol par l'application d'engrais, ne peut se concevoir que si on la réalise sur une surface d'une certaine taille. Dans un essai d'ombrage, une modification d'un microclimat ne peut se concevoir que sur une surface encore nettement plus grande : l'effet de masse est ici très important.

Dans un essai d'engrais, une parcelle élémentaire de 20 arbres (4 x 5) pour 6 utiles (2 x 3) semble insuffisante pour créer "l'ambiance sol" recherchée. D'autre part, le rapport arbres utiles/arbres totaux est défavorable. Il est à prendre en considération lorsque le coût de l'essai est un élément du choix. Un parcelle élémentaire de 30 arbres (5 x6) pour 12 utiles (3 x 4) semble être une limite inférieure raisonnable. La présence d'essences variées d'arbres d'ombrage agissant sur de petites étendues par modification physico-chimique du sol, et constitution de microclimat très localisé, augmente le dégré d'association des arbres a production comparable. Dans un essai établi apres déforestation totale, ce degré d'association diminue certainement mais dans quelle proportion? L'hétérogénéite du sol existe toujours, due:

— à la microtopographie;
— à la profondeur très variable des concressions latéritiques;
— à l'influence variable des essences forestières qui persiste, après abattage, pendant quelques années.

Toutefois, le coefficient a devenant plus faible, la taille optimum théorique augment et c'est une raison de plus d'adopter 12 comme limite inférieure. D'autre part, nous avons vu qu'on n'avait aucun intérêt à aller au-delà de 20 à 25 arbres utiles.

Voici donc une fourchette 12-25 arbres utiles à l'intérieur de laquelle on doit se placer.

b) *Disponibilités en surface, en personnel qualifié*

1/— *On n'est pas limité en surface*

a) S'il n'y a pas de problèmes de personnel, on se placera en haut de la fourchette et on adoptera des parcelles élémentaires utiles de 25 arbres. Le nombre de répétitions sera fonction de la précision désirée.

b) S'il y a un problème de personnel, on adoptera toujours 25 arbres mais le nombre de répétitions sera alors fonction de ce problème et la précision de l'essai sera réduite.

2/ — *On est limité en surface*

a) S'il n'y a pas de problème de personnel, on se placera en bas de la fouchette et on adoptera des parcelles élémentaires utiles de 12 arbres donnant le maximum de précision par unité de surface d'essai (à l'intérieur de la fourchette). Le nombre de répétitions sera déterminé par la surface disponible et une certaine précision en résultera.

b) Si, en plus, il y a un problème de personnel, on se placera au milieu de la fourchette de façon à:
— occuper la surface disponible;
— avoir un nombre de répétitions en rapport avec le personnel disponible.

EXEMPLE — PRECISION DE DEUX DISPOSITIFS

a) *Dispositif 3^3en parcelles élémentaires de 25 arbres utiles avec 4 répétitions.*

La formule donnée plus haut nous donne une valeur approximative de la variance de la moyenne parcellaire (données transformées): 0,04058. Il y a 12 répétitions effectives pour chaque effet principal.

Soit M1 et M2, deux moyennes de traitement consécutives après classement. Elles seront significativement différentes si :

$$\frac{\log M2 - \log M1}{\sqrt{2 \times \frac{0,04058}{12}}} > 2$$

$$\log \frac{M2}{M1} > 0,16446$$

$$\frac{M2}{M1} > 1,46$$

(Notons que les "M" sont les moyennes géométriques obtenues par retransformation simple, présentant donc un biais par rapport aux moyennes arithmétiques).

Le rapport significatif est très élevé dans les conditions adoptées à Nkoemvone. Pour l'étude de l'interaction, le rapport des moyennes devra être supérieur à 1,93 ! On peut espérer qu'après déforestation totale, la variance de la moyenne parcellaire sera nettement réduite et par conséquent la plus petite différence significative, mais on ne peut dire dans quelle proportion car les données obtenues avec cette technique cultural manquent.

b) *Dispositif 3^3 en parcelles élémentaires de 12 arbres utiles avec 8 répétitions.* (Soit une augmentation e 22,5% de la surface de l'essai par rapport au précédent).

En prenant la variance ajustée de la moyenne parcellaire : 0,04906, des calculs similaires conduisent aux rapports significatifs suivants:

pour l'effet principal :

$$\frac{M2}{M1} > 1{,}34$$

pour l'interaction :

$$\frac{M2}{M1} > 1{,}67$$

Quoique inférieurs aux précédents, ils sont toujours relativement élevés.

Ce premier point étant acquis, il s'agit maintenant d'essayer de réduire le résidu aléatoire en étudiant la possibilité de mesurer l'action de certains facteurs in tervenant dans celui-ci.

1) *ETUDE DE LA CORRELATION ENTRE PRODUCTIONS DE PARCELLES CONTIGUES DE 12 ARBRES*, en vue de déterminer, si elle existe, l'intérêt de l'adjonction d'une parcelle témoin à chaque traitée pour mesurer l'influence de facteurs divers ayant une action localisée en taches (fertilité, microtopographie, arbres d'ombrage a développement variable, etc. . .).

Elle a été entreprise sur les parcelles 21 à 25 de Nkoemvone.

79 couples indépéndants de parcelles contiguës de 12 arbres ont été constitués (couples 1–2, 3–4, etc..) L'étude de la corrélation entre moyennes de production (transformées en logarithme) des parcelles contigues conduit aux résultats suivants (notons que la transformation des données normalise la distribution et permet cette étude):

$r = 0{,}344$ pour 77 degrés de liberté, ce qui est très hautement significatif (r au seuil 0,01 et pour 77 ddl #0,29);

pente de la régression, $b = 0{,}392$.

La même étude entreprise sur les données brutes montre une liaison plus importante ($r = 0{,}53$, $b = 0{,}57$). Mais nous ne sommes pas là dans les conditions de validité nécessaires pour l'analyse. La distribution des moyennes parcellaires brutes, (calculées pourtant sur 12 données) est assez nettement dissymétrique et éloignée de la normale.

Appelons x les données de l'ensemble formé par un élément de chaque couple, y les données de l'ensemble complémentaire. Nous aurons, les données étant transformées, $s_x^2 \# s_y^2$.

Coefficient de corrélation $r = \frac{sxy}{sxsy}$

Puisque sx # sy, $r \# \frac{sxy}{s^2}$. Cette dernière expression représente la pente de la régression b. On ^{sx}a donc r # b. On le constate effectivement dans notre cas : 0,344 et 0,392.

Comparons les deux dispositifs suivants disposés sur la même surface, les parcelles élémentaires ayant la même taille (12 arbres):

— adjonction à chaque parcelle traitée d'une parcelle témoin contigue afin d'entreprendre ensuite une analyse de convariance (chaque traitement est appliqué dans n parcelles);

— dispositif classique comportant par conséquent un nombre double de répétitions (2 n).

1er cas. On a déterminé le coefficient de corrélation entre moyennes de parcelles contigüees $r = \frac{sxy}{sxsy} = 0{,}344$

(sxy = covariance x/y - sx et sy = écarts-type de x et y.

Dans l'analyse de convariance, la somme des carrés des écarts résiduelle est diminuée de la quantité :

$$Q = \frac{(Sxy\ \text{résiduel})^2}{Sx2\ \text{résiduel}}.$$

(S représentant une somme de carrés d'écarts ou de produits d'écarts).

$$Q = r^2\ S_y^2\ \text{résiduel}.$$

La somme des carrés des écarts résiduels devient

S_y^2 résiduel – r^2 . S_y^2 résiduel = $(1\text{-}r^2)\ S_y^2$ résiduel et le carré moyen résiduel : $(1\text{-}r^2)\ s_y^2$ résiduel.

La variance de la moyenne sera approximativement $\frac{(1\text{-}r^2)\ s_y^2\ \text{résiduel}}{n}$ (une correction doit, en fait, intervenir pour tenir compte de l'imprécision supplémentaire due a l'utilisation du coefficient de régression b mais on peut la négliger dans ce calcul approximatif).

2eme cas. Le carré moyen résiduel sera ici calculé avec un nombre de degrés de liberté supérieur mais il doit etre peu différent du s_y^2 résiduel précédent. Les deux dispositfs étant supposés installés sur deux champs identiques, la variance du residu aléatoire sera constante quelque soit le dispositif.

La variance de la moyenne est alors :

$$\frac{s_y^2 \text{ résiduel}}{2n}$$

L'efficacité relative du 2ème dispositif par rapport au premier sera le rapport des variances des moyennes:

$$\frac{s_y^2 \text{ res. } (1-r^2)}{n} \Big/ \frac{s_y^2 \text{ res.}}{2n}$$
$$= 2\,(1-r^2) = 2-2\,r^2.$$

Cela montre que l'avantage reviendra au deuxième dispositif si $r<\frac{\sqrt{2}}{2}$, soit $r<0{,}7$ et inversement. Il est imponsable que cette valeur puisse etre atteinte en champ et on peut donc dire que l'avantage reviendra toujours, dans la pratique, au deuxième dispositif.

Dans notre cas, l'efficacité relative, calculée compte non tenu des nombres de degrés de libertá, est 1,76.

Revenons sur le travail effectué par H.MARTICOU et R.A MULLER (pages 198 et 199 de l'étude précédente). Ils proposent d'adjoindre à chaque parcelle traitée une parcelle témoin et d'utiliser le modèle suivant:

$$P\,1 = Po.\,k\,1.\,I.\,E = Po.\,\frac{T1}{To}.\,I.\,E.$$

où Po = production de la parcelle à traiter l'année 0
P1 = production de la parcelle traitée l'année 1
To = production de la parcelle témoin l'année 0
T1 = production de la parcelle témoin l'année 1
I = un facteur multiplicatif caractérisant l'effet du traitement.
E = une variable aléatoire de moyenne 1.

En passant aux logarithmes,

$$\log P1 - \log Po = \log T1 - \log To + \log I + \log E$$

(log E devenant une variable aléatoire normale, de moyenne nulle).

Ceci revient à déterminer si entre et l'année 1 et l'année 0, l'accroissement de production dans la parcelle traitée est supérieur à l'accroissement de production dans la parcelle témoin adjacente, les données étant exprimées en logarithmes. C'est un modèle de convariance pour lequel on suppose que la pente (b) est égale à l'unité. On a vu au début du chapitre qu'il n'en était rien puisque aec nos données r # b # 0,35. Une valeur proche de 1 ne peut être envisagée et par conséquent le deuxième dispositif est décidément a recommander.

D'autre part, le phénomène d'alternance dans les productions annuelles d'un arbre nécessite la prise en considération du cumul d'un certain nombre d'année et non l'étude d'une seule production annuelle.

2) *PROBLEME DES MANQUANTS DÛ A DES PERTES ACCIDENTELLES*

Un manquant dans une parcelle élémentaire va provouer une *perturbation* sur la donnée parcellaire que l' on doit mesurer et faire intervenir dans l'analyse si la mortalité ne peut dépendre du traitement (dans le cas contraire, l'analyse portera sur la production ramentée au nombre total initial d'àrbres). La meilleure façon de procéder est d'employer l'analyse de covariance suivante. On travaille sur la moyenne parcellaire calculée par plant survivant et on ajuste cette donnée par le nombre de manquants par parcelle. On peut améliorer la technique en pondérant chaque manquant par le temps écoulé depuis la mort de l'arbre; par exemple, en attribuant une valeur 5 à un arbre manquant depuis 5 ans, une valeur 2 à un arbre manquant depuis 2 ans, etc. . . La convariance va permettre de calculer sur l'ensemble de l'essai l'influence moyenne d'un manquant, de 2 manquants, etc. . . sur le rendement parcellaire rapporté au nombre d'arbres survivants et de faire au mieux la correction. (Notons que l'influence des manquants variera suivant qu'ils sont disposés en tache ou dispersés dans la parcelle. On ne pourra en tenir compte).

Ce procédé est jugé par PEARCE comme le meilleur (Field Experimentation with Fruit Trees and Other Perennial Plants, East Malling, maidstone, Kent — Page 74 et la suite).

Le problème des manquants ne se pose pas de la même façon dans le dispositif "randomisation totale". L'influence d'un manquant se répartira sur plusieurs hybrides, 8 au mieux. La moyenne d'un traitement sera donc moins perturbée que dans le dispositif en bloc. Toutefois, on peut résoudre le problème de façon similaire. On utilisera une pseudo-variate de cette façon:

Pour chaque arbre, on notera le nombre de manquants contigus (i) et pour chacun de ceux-ci, le nombre d'années écoulées depuis la mort de l'arbre (ai). La variable indépendante liée à la production de l'arbre sera E ai. On pourra d'ailleurs chercher la meilleure liaison possible entre les deux variables en cherchant une expression pouvant être plus représentative de l'effet de compétition que E ai. Cette expression sera celle diminuant au maximum le résidu, aléatoire dans l'analyse de covariance. Cette analyse permettra de chiffrer l'influence moyenne sur la production d'un arbre de la présence d'un manquant contigu, 2 manquants, 3, etc. . . Le fait que les manquants soient disposés en tache ou dispersés n'a plus ici d'importance.

3) RELATION ENTRE MENSURATION A UN AGE DONNE ET PRODUCTION CUMULEE ULTERIEURE

Dans l'essai comparatif d'hybrides mis en place en 1964 à Nkoemvone, parcelle 90 (5 répétitions – parcelle élémentaire de 20 arbres — 40 familles – 4.000 arbres), des mesures de diamètre du tronc à 20 cm du collet ont été effectuées à 11 et 23 mois de plantations.

Deux analyses de covariance ont été tentées en prenant comme variable dépendante le cumul moyen des productions par parcelle élémentaire, calculée par arbre vivant après 5 ans de plantations (transformé en logarithme) et comme variable indépendante la moyenne parcellaire des mensurations à 11 mois et à 23 mois.

1ère analyse : (diamètres à 11 mois)

Source de variation	Degrés de liberté	Sx^2	Sxy	S_y^2	$\frac{(Sxy)^2}{Sx2}$	S_y^2 réduit	D.L.	C.M.
Famille	39	4,9596	2,58310	9,92490		8,65922	39	
Bloc	4	0,4347	0,61660	3,08800		2,42395	4	
Résidual.	156	5,5108	3,83294	8,52386	2,66593	5,85793	155	0,03779
Fam. + rés.al.	195	10,4704	6,41604	18,44876	3,93161	14,51715		
Blocs + rés.al.	160	5,9455	4,44954	11,61186	3,32998	8,28188		

La pente de la régression (b = 0,69) est significativement différente de 0 (F = 17,3). Le coefficient de corrélation linéaire est : 0,56. Le carré moyen du résidu est réduit de 0,05464 à 0,037795, soit de 31 % de sa valeur.

2eme analyse : (diamètres à 23 mois)

Source de variation	Degrés se liberté	Sx^2	Sxy	S_y^2	$\frac{(Sxy)^2}{Sy^2}$	S_y^2 réduit	D.L.	C.M.
Famille	39	14,0613	6,91459	9,92490		6,55770	39	0,16815
Bloc	4	5,8366	3,97441	3,08800		0,67430	4	0,16857
Résidual.	156	23,0308	9,90897	8,52386	4,26332	4,26054	155	0,02749
Fam. + res.al.	195	37,0921	16,82356	18,44876	7,63052	10,8124	194	
Blocs + res.al.	160	28,8674	13,88338	11,61186	6,67702	4,93484	159	

La pente de la régression (b = 0,43) est significativement différente de 0 (F = 38). Le coefficient de corrélation linéaire est: 0,71. Le carré moyen du résidu est réduit de 0,05464 à 0,02749, soit de 50 % de sa valeur.

La liaison est plus intéressante à 23 mois. Il semble donc que c'est cette variable indépendante que l'on doit utiliser dans les analyses. (Une analyse effectuée avec l'accroissement du diamètre de 11 à 23 mois a donné des résultats du même ordre). Les calculs repris après six ans de plantation ont apporté des conclusions semblables. Il semble toutefois que l'accroissement du diamètre du tronc entre 11 et 23 mois pourrait être une variable indépendante plus intéressante que le diamètre à 23 mois. Quoiqu'il en soit, les deux analyses doivent être effectuées afin de prendre en considération celle amenant le maximum de réduction du carré moyen résiduel.

L'analyse de covariance ne peut être effectuée que si la variable auxilliaire n'est pas influencée par les traitements. Or, ici, la variable "diamètre moyen" est influencée par :

— la fertilité moyenne parcellaire;
— le facteur étudié : l'hybride.

En effet, l'analyse de variance des diamètres moyens montre que les différences sont significatives entre hybrides. Remener dans ce cas toutes les moyennes de diamètre/hybride à la moyenne générale nous conduirait à ne pas tenir compte de la valeur intrinsèque de l'hybride qui se manifeste justement par des croissances variables dans le jeune âge.

Dans les essais où les parcelles élémentaires ont même constitution génétique, le diamètre moyen parcellaire sera le reflet de la fertilité moyenne parcellaire uniquement. L'analyse de covariance indiquée sera alors très bénéfique et nous devons la pratiquer chaque fois que l'on en a la possibilité. Dans un essai de fertilisation minérale commençant à l'entrée en production des arbres, les mensurations effectuées avant le premier traitement à 11 et 23 mois par exemple, pourront être utilisées comme indiqué. C'est indipendable si on veut avoir des différences significatives raisonnables.

Les rapports significatifs seront, en effet, abaissés très approximativement aux valeurs suivantes (toujours sous forêt amémagée) :

1er dispositif: $(3^3) \times 4$ (parcelle élémentaire de 25 a.)
effet principal : 1,3
interaction : 1,58.

2ème dispositif: $(3^3) \times 8$ (parcelle élémentaire de 12 a.)
effet principal : 1,23
interaction : 1,44.

Dans les essais comparatifs d'hybrides, de familles, de variétés, la liaison indiquée pourra permettre de faire un premier choix rapide en analysant les mensurations sans attendre l'entrée en production des parcelles.

En conclusion, les résultats positifs obtenus peuvent se résumer ainsi :

— *Taille des parcelles élémentaires utiles*
— *lignes de bordure inutiles :* on doit adopter la randomisation totale pied par pied. Elle pourra s'effectuer, si un contrôle de l'hétérogénéité du champ d'essai est possible par stratification en blocs, à l'intérieur de chacun d'eux;

— *lignes de bordure nécessaires :* une fourchette 12-25 arbres est à recommander :
— si la surface n'est pas limitée:
. personnel et crédit suffisant . . . . . .25 arbres;
. personnel et crédit limité . . . . .25 arbres (mais le nombre de répétitions étant limité par ce facteur, la précision de l'essai sera déterminée par celcu-ci);
— si la surface est limitée :
. personnel et crédit suffisant . . . . . .12 arbres;
. personnel et crédit limité.
On se placera au milieu de la fouchette de façon à occuper la surface disponible et avoir un nombre de répétitions en rapport avec le personnel disponible.

— *Le schéma challique sera adopté* mais le résidu aléatoire pourra être nettement diminué par l'intervention dans l'analyse de facteurs dont l'action pourra être mesurée :
. nombre et date d'apparition des manquants par parcelle élémentaire (dans tous les cas);
. mensurations du tronc à 20 cm du collet effeectuées apres un an et deux ans de plantation quand cette variable n'est par influencée par les traitements. Si elle l'est (essai d'hybrides par exemple), l'analyse de ces mensurations permet un trirapide du matériel en expérimentation./–