

27 2 53 66

# APPLICATION DE L'ANALYSE MULTIVARIABLE A UN ESSAI COMPARATIF DE PRODUITS INSECTICIDES

par

**P. L. LEFORT<sup>1</sup> R. COUILLOUD<sup>2</sup> et A. ANGELINI<sup>3</sup>**

## RÉSUMÉ

Quatre variables relatives à des observations faites en cours d'expérience sur un essai comparatif de cinq produits insecticides ont été mesurées. Aucune ne permet, individuellement, de conclure à des différences d'efficacité entre les produits. Par contre, leur traitement en analyse de variance généralisée fait apparaître des différences statistiquement significatives. Les distances généralisées de Mahalanobis conduisent à la séparation des cinq produits en trois groupes. Les composantes principales permettent, d'une part, de classer les variables en fonction de leur pouvoir de séparation vis-à-vis des traitements, et fournissent, d'autre part, les éléments d'interprétation de ces différences.

Dans les essais comparatifs de produits insecticides, l'appréciation et la différenciation de la valeur des formules employées se font, en général, d'après l'examen des rendements en coton-graine.

L'étude globale, ou même fractionnée, du seul facteur « rendement » ne représente, cependant, que le bilan de l'activité d'un produit vis-à-vis de différents ravageurs, bilan qui peut être plus ou moins faussé par divers phénomènes de compensation de la part de la plante. Aussi s'attache-t-on, au cours de chaque expérience, à réunir des observations multiples, susceptibles d'apporter le maximum de précision sur le comportement des produits : étude quotidienne des chutes d'organes, appréciation dans ce « shedding » de la part due aux insectes, recensement des

chenilles dans les organes attaqués, etc.

On dispose donc, en fin de campagne, et du rendement final et d'un ensemble de données plus fines. Si le traitement de ces informations, variable par variable, apporte des précisions sur les caractéristiques des produits testés, il semble cependant que, dans un tel cas, l'utilisation de l'analyse multivariante, plus puissante et plus performante, puisse conduire à une meilleure exploitation des mesures effectuées.

Le propos de cet article est de montrer, à partir d'un exemple, ce que peuvent apporter, par rapport aux analyses à une seule variable, les méthodes de traitement simultané des données.

## 1. — L'ESSAI COMPARATIF ÉTUDIÉ

L'efficacité de cinq formules insecticides a été testée dans un essai comparatif en blocs Fisher, comprenant cinq répétitions. Ces formules étaient les suivantes :

- A - Endrine-DDT  
300 + 1 125 g de matière active par hectare.
- B - Monitor\*-DDT  
717 + 1 200 g de matière active par hectare.
- C - Triazophos-DDT  
600 + 1 200 g de matière active par hectare.
- D - Gardona\*-DDT  
700 + 1 200 g de matière active par hectare.
- E - Phosalone-DDT  
625 + 925 g de matière active par hectare.

Les observations faites en cours d'expérience ont été les suivantes :

- x = nombre de boutons floraux tombés.
- y = nombre de boutons floraux tombés et troués par des chenilles.
- z = nombre de capsules tombées et trouées par des chenilles.
- t = nombre de chenilles d'*Heliothis armigera* trouvées dans les boutons floraux et les capsules tombés et troués.

<sup>1</sup> Généticien et <sup>2,3</sup> Entomologistes à la Station principale de Bouaké, B.P. 604, Côte d'Ivoire.

\* Monitor : diméthylphosphoramidithioate.

\* Gardona : tétrachlorvinfos

L'essai a été mis en place au mois d'août. Le choix de la date de semis, et la définition des quatre variables observées correspondaient, *a priori*, au désir

de tester l'efficacité des formules vis-à-vis d'*H. armigera*, ravageur important du cotonnier en Côte d'Ivoire.

## II. — LES MÉTHODES D'ANALYSES

Les méthodes auxquelles nous faisons référence ont été développées essentiellement par RAO (1952, 1965), ANDERSON (1958), WILKS (1962) et KENBALL (1972). Elles constituent trois analyses distinctes : l'analyse de variance généralisée, l'analyse des distances généralisées et l'analyse des composantes principales.

### 1° L'analyse de variance généralisée

C'est une extension de l'analyse de variance à une variable au cas où les observations, au lieu d'être des variables indépendantes aléatoires à une dimension, sont des variables vecteurs indépendantes aléatoires d'ordre  $p$ . Dans la pratique, on effectue les calculs d'analyse de variance sur chacune des  $p$  variables, puis ceux d'analyse de covariance pour chaque couple. On dispose ces analyses en un tableau des variances-covariances généralisé, où chaque source de variation est représentée non plus par une somme de carrés, mais par une matrice somme des produits, carrée, symétrique, et d'ordre  $p$ . La matrice inter-traitements est appelée (B) ou matrice « Between », la matrice erreur est appelée (W) ou matrice « Within ». La première prend en compte la variabilité inter-traitements, la seconde la variabilité intra-traitements. La signification des sources de variation, définies pour trois niveaux de degrés de liberté, est testée grâce à des tables de valeurs limites particulières ; on peut également utiliser les tests classiques du F ou du  $\chi^2$  grâce à certaines formules (RAO, 1965, chap. VIII).

### 2° L'analyse des distances généralisées

Le calcul de l'indice  $D^2$  ou « distance généralisée de MAHALANOBIS », va permettre de situer les traitements les uns par rapport aux autres.

Si  $(\bar{X}_n)$  et  $(\bar{X}_{n'})$  sont les variables vecteurs lignes relatives aux traitements  $n$  et  $n'$ , la quantité  $D^2$  entre ces deux traitements sera :

$$D^2_{n,n'} = 2 (b - 1) (\bar{X}_n - \bar{X}_{n'}) (W)^{-1} (\bar{X}_n - \bar{X}_{n'})'$$

où  $b$  est le nombre de traitements,  $(W)^{-1}$  la matrice inverse de (W), et  $(\bar{X}_n - \bar{X}_{n'})'$  le transposé du vecteur  $(\bar{X}_n - \bar{X}_{n'})$ , soit le vecteur colonne de mêmes éléments.

Différentes méthodes de regroupements permettent de visualiser, par représentation graphique, les relations entre les traitements après calcul deux à deux des indices  $D^2$  :

- Les constellations (RAO, 1952 ; VAN DEN DRIESCH, 1965) ;
- Les dendrogrammes (SOKAL et SNEDECOR, 1963) ;
- La méthode nodale (*id.*).

### 3° L'analyse des composantes principales

Elle permet, une fois schématisées les relations entre traitements, de recombinaison les variables de façon à obtenir des combinaisons linéaires et indépendantes entre elles des caractères, combinaisons pouvant rendre compte d'un traitement ou d'un groupe de traitements. Les phénomènes complexes établis en termes de corrélations parmi les variables considérées, sont alors exprimés comme des fonctions d'un petit nombre de nouvelles variables synthétiques (appelées composantes principales ou facteurs) contenant l'information permettant de les expliquer.

L'opération consiste en pratique à déterminer le changement d'axe (ici une rotation) conduisant à un nouveau système où toute projection de l'ellipsoïde de confiance (généralisant à  $p$  variables l'ellipse de confiance) dans un plan déterminé par deux axes est une ellipse où les corrélations sont nulles, les variables étant toutes indépendantes deux à deux. Ce changement d'axe est obtenu par diagonalisation de la matrice des variances-covariances intertraitements (B), c'est-à-dire en calculant D, matrice diagonale (matrice dont tous les éléments sont nuls, excepté ceux de la première diagonale) équivalente à (B).

Le calcul correspond à la recherche des vecteurs propres de la matrice des corrélations intertraitements, soit la matrice (B) des variances-covariances normée par les écarts-types. Cette matrice est symétrique, et il existe une matrice orthogonale X telle que :

$$(X)' (B) (X) \text{ soit diagonale}$$

les valeurs propres de (B) sont les éléments diagonaux de cette matrice diagonale.

Les calculs ont été faits, dans notre exemple, au moyen d'une machine de bureau de faible capacité. Ceci est possible jusqu'à cinq variables ; au-delà, il faut faire appel aux ordinateurs pour lesquels les programmes d'analyse multivariable commencent à être assez répandus.

## III. — RÉSULTATS

Tableau 1. — Données brutes des observations, moyennes des cinq répétitions, pourcentages du traitement A.

Variables	Traitement	Rép. I	Rep. II	Rép. III	Rép. IV	Rép. V	Moyenne	% du trait. A
x	A	635	1 034	351	644	688	670,4	100,0
	B	920	892	507	962	660	788,2	117,6
	C	641	490	529	678	726	612,8	91,4
	D	552	783	577	1 177	317	683,2	101,9
	E	489	988	478	1 217	807	795,2	118,7
y	A	346	366	83	170	148	222,6	100,0
	B	424	350	129	302	217	284,4	127,3
	C	166	140	159	142	184	158,2	71,1
	D	195	340	166	428	121	250,0	112,3
	E	164	240	220	450	225	259,8	116,7
z	A	541	542	257	309	338	397,4	100,0
	B	582	583	289	510	425	477,8	120,2
	C	450	277	272	108	302	281,8	70,9
	D	461	435	352	479	253	396,0	99,7
	E	245	596	320	574	436	434,2	109,3
t	A	114	140	84	58	86	96,4	100,0
	B	153	153	95	106	132	127,8	132,6
	C	104	80	87	30	91	78,4	81,3
	D	69	100	85	95	81	86,0	89,2
	E	51	85	118	103	133	98,0	101,7

Tableau 2. — Analyses de variances.

Origine de la variation	D.L.	variable x			variable y		
		S.C.	C.M.	F	S.C.	C.M.	F
Totale .....	24	1 384 448			279 434		
Répétitions .....	4	628 106	157 026	3,99%	87 227	21 807	2,40
Traitements .....	4	126 056	31 514	< 1	46 662	11 665	1,28
Erreur .....	16	630 286	39 393		145 545	9 097	
	D.L.	variable z			variable t		
		S.C.	C.M.	F	S.C.	C.M.	F
Totale .....	24	428 508			21 721		
Répétitions .....	4	117 105	29 276	2,28	3 140	785	1,09
Traitements .....	4	105 913	26 480	2,06	7 082	1 770	2,46
Erreur .....	16	205 485	12 843		11 499	719	

Tableau 3. — Analyses de covariances.

Origine de la variation	D.L.	variable xy			variable xz			variable xt		
		S.P.	P.M.	F	S.P.	P.M.	F	S.P.	P.M.	F
Totale .....	24	524 131			562 356			70 245		
Répétitions .....	4	210 117	52 529	3,71*	164 392	41 098	2,23	— 11 041	— 2 760	< 1
Traitements .....	4	67 725	16 931	1,55	103 593	25 898	1,41	23 103	5 776	1,59
Erreur .....	16	246 289	15 393		294 571	18 411		58 183	3 636	

	D.L.	variable yz			variable yt			variable zt		
		S.P.	P.M.	F	S.P.	P.M.	F	S.P.	P.M.	F
Totale .....	24	278 331			46 148			67 618		
Répétitions .....	4	84 443	21 111	2,70	— 2 732	— 683	< 1	6 811	1 703	< 1
Traitements .....	4	68 707	17 177	2,20	14 086	3 521	1,62	23 393	5 848	2,50
Erreur .....	16	125 181	7 824		34 794	2 175		37 414	2 338	

Les effets « traitements » ne sont significatifs ni pour les analyses de variance, ni pour les analyses

de covariance; l'étude des variables une par une ne permet donc, dans aucun cas, de séparer les traitements.

Tableau 4. — Analyse de variance-covariance généralisée.

Origine de la variation	D.L.	Matrices sommes de produits				
Totale .....	24	1 384 448	524 131	562 356	70 245	} = (Sij)
			279 434	278 331	46 148	
				428 508	67 618	
					21 721	
Répétitions .....	4	628 106	210 117	164 392	11 041	} = (Rij)
			87 226	84 443	2 732	
				117 105	6 811	
					3 140	
Traitements .....	4	126 056	67 725	103 593	23 103	} = (Bij)
			46 662	68 707	14 086	
				105 918	23 393	
					7 082	
Erreur .....	16	630 287	246 289	294 571	58 183	} = (Wij)
			145 546	125 181	34 794	
				295 485	37 414	
					11 499	

Ces quatre matrices étant carrées et symétriques, seuls les termes au-dessus de la première diagonale ont été écrits.

Les effets « répétitions » et « traitements » sont testés par les rapports de déterminants de matrice suivants :

$$\text{Effet répétitions } \Lambda_B = \frac{|Wij|}{|Rij| + |Wij|} = 0,059^{**}$$

$$\text{Effet traitements } \Lambda_B = \frac{|Wij|}{|Bij| + |Wij|} = 0,234^*$$

Donc, l'analyse simultanée des quatre variables permet de mettre en évidence un effet traitement significatif avec une probabilité de 95 %.

Tableau 5. — Calcul et niveau de signification des distances D' entre chaque traitement.

D'	A	B	C	D	E
A	0	15,6*	20,5*	15,4*	5,6
B		0	38,0**	54,5**	25,4*
C			0	49,8**	24,2*
D				0	12,4
E					0

Les niveaux de signification sont testés par la quantité :

$$F = \frac{2b - p - 1}{p} \frac{b}{4(b-1)} D^2, \text{ à } p \text{ et } (2b - p - 1) \text{ D.L.}$$

où  $b$  est le nombre de traitements et  $p$  le nombre de variables.

Ce test ne doit en fait être pris que comme une indication, la comparaison des moyennes deux à deux par les  $D^2$  n'étant pas rigoureuse. Il existe pour ces comparaisons une généralisation du  $S$  de SCHÉRRÉ (test des contrastes) valable en toute rigueur.

On peut représenter graphiquement dans l'espace les « distances » entre traitements (fig. 1).

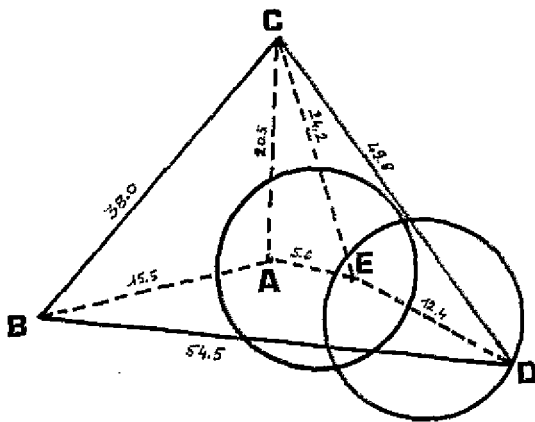


Fig. 1. — Représentation dans l'espace des distances entre traitements. (Les cercles ont pour diamètre la valeur du  $D^2$  au seuil  $P$  0,05.

En définitive :

B, C et D diffèrent entre eux à  $P = 0,01$  ;

A diffère de B, C et D à  $P = 0,05$ , mais ne diffère pas de E ;

E diffère de B et C à  $P = 0,05$ , mais ne diffère ni de A, ni de D.

### Les composantes principales

Ce sont les vecteurs propres de la matrice des corrélations inter-traitements. Cette matrice s'obtient en normant par les écarts-types correspondants les éléments de la matrice (B) des variances-covariances inter-traitements.

On obtient quatre composantes principales qui, classées par valeurs décroissantes des coefficients de leurs éléments, sont les suivantes :

$$CP_1 = z + 0,98 y + 0,95 x + 0,91 t$$

prenant en charge 89,5 % de la variation.

$$CP_2 = t - 0,43 y - 0,35 x - 0,12 z$$

prenant en charge 6,6 % de la variation.

$$CP_3 = x - 0,58 y - 0,43 z - 0,04 t$$

prenant en charge 3,5 % de la variation.

$$CP_4 = -z + 0,86 y + 0,27 t + 0,06 x$$

prenant en charge 0,3 % de la variation.

Les trois premières composantes principales expliquant 99,6 % de la variation totale, nous ne nous attacherons pas à l'interprétation de la quatrième.

Explicitons ces composantes :

1 - Des valeurs fortes de  $CP_1$  correspondront à des traitements caractérisés par ordre d'importance décroissante par :

- de nombreuses capsules tombées et trouées ;
- de nombreux boutons floraux tombés et troués ;
- de nombreux boutons floraux tombés ;
- de nombreux *H. armigera* dans les boutons et les capsules tombés et troués.

C'est une nouvelle variable synthétique qui tient compte additivement et équitablement de chacune des mesures, elle pondère l'abondance d'*H. armigera* et les dégâts.

2 - Des valeurs fortes de  $CP_2$  correspondront à des traitements caractérisés par ordre d'importance décroissante par :

- de nombreux *H. armigera* ;
- peu de boutons floraux tombés et troués ;
- peu de boutons floraux tombés ;
- peu de capsules tombées et trouées.

Cette variable purifie la mesure du nombre d'*H. armigera* en éliminant les mesures des dégâts. Pour des dégâts équivalents, la variable est d'autant plus grande que le nombre d'*H. armigera* est plus grand. On peut ainsi la considérer comme une mesure purifiée d'inefficacité spécifique d'un traitement contre *H. armigera*.

3 - Enfin, la troisième composante peut être considérée comme rendant compte du shedding non parasitaire, ou de façon plus exacte du shedding non dû aux chenilles.

A partir de ces trois composantes, les caractéristiques des traitements seraient parfaitement décrites dans un espace à trois dimensions. La projection sur des plans définis par deux composantes permet d'analyser l'information apportée par chacune (fig. 2, 3 et 4).

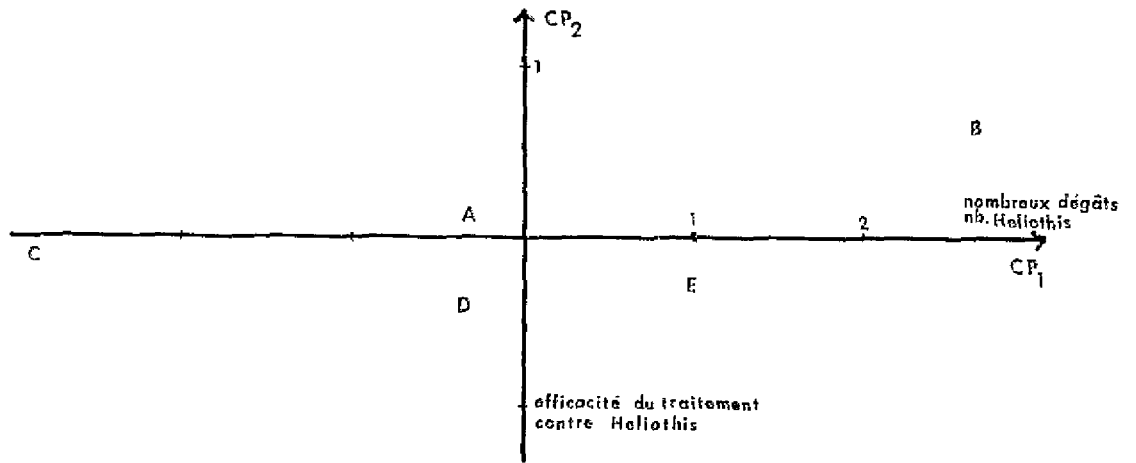


Fig. 2.

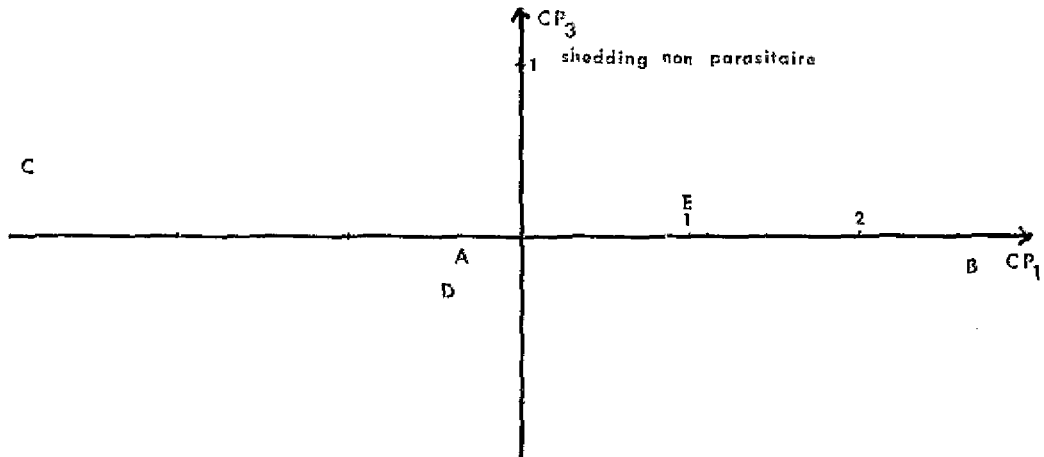


Fig. 3.

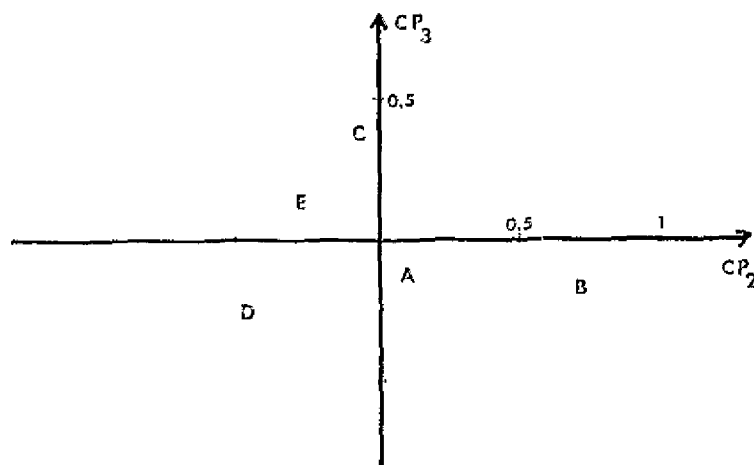


Fig. 4.



## IV. — DISCUSSION

Aucune des quatre séries d'observations réalisées en cours d'expérience ne permet, individuellement, de mettre en évidence des différences dans l'efficacité des cinq traitements comparés. Par contre, leur utilisation simultanée en analyse de variance généralisée, et le calcul des distances  $D^2$  entre traitements deux à deux aboutissent au classement suivant :

C      A      E      D      B  
 \_\_\_\_\_  
 \_\_\_\_\_

La formule C (Triazophos-DDT) est la meilleure, la formule B (Monitor-DDT) la moins bonne. Les trois autres formules sont intermédiaires, A (endrine-DDT) différant de D (Gardona-DDT), alors que E (Phosalone-DDT) ne diffère ni de A ni de D.

L'analyse des composantes principales fournit les éléments d'interprétation des différences rencontrées.

Elle permet, en premier lieu, de classer le pouvoir séparateur de chaque variable vis-à-vis des traitements, soit, par ordre décroissant :

z = nombre de capsules tombées et trouées ;  
 y = nombre de boutons floraux tombés et trousés ;  
 x = nombre de boutons floraux tombés ;  
 t = nombre de chenilles d'*H. armigera* dans les boutons et les capsules tombées et trouées.

On peut justifier, *a posteriori*, cette hiérarchie. En effet, les deux premières séries de données traduisent directement l'incidence du parasitisme dû aux chenilles de la capsule, essentiellement *H. armigera*, dans l'expérience présente. Le nombre de capsules tombées et trouées est, des deux mesures, la plus précise, car les dégâts dus aux chenilles sur les capsules sont manifestes et ne peuvent échapper à l'observation, alors que sur les boutons floraux, les lésions et perforations provoquées par les chenilles sont souvent moins apparentes. Ainsi, un certain nombre d'observations à classer en y, l'ont été en x, diminuant la précision de la variable y.

La médiocrité du pouvoir séparateur de la variable x (nombre total de boutons floraux tombés) est liée au fait que les différences entre traitements sont « tamponnées » par l'existence chez le cotonnier d'un shedding non parasitaire touchant 50 à 60% des jeunes organes fructifères.

Enfin, le mauvais classement de la variable t s'explique par la grande variabilité intertraitement de ce type d'observations, variabilité due à l'hétérogénéité des populations d'insectes et aux modifications dans le temps de l'attractivité des plantes d'une parcelle. Cette attractivité, identique au départ, varie en effet par la suite, en raison directe de l'efficacité d'un traitement donné, puisqu'elle dépend du plus ou moins grand nombre d'organes fructifères demeurant en place sur les plantes d'une parcelle.

La description des traitements par les composantes principales recoupe bien la constellation obtenue à partir des  $D^2$  (fig. 1), en particulier la projection sur le plan défini  $CP_1$  et  $CP_2$  (fig. 2). La première composante, qui prend en part 89,5% de la variation totale, redonne les trois groupes C A D E et B, en séparant toutefois assez nettement le traitement E de A et de D.

Cette analyse permet, enfin, d'expliquer et d'interpréter les différences mises en évidence par l'analyse des distances généralisées :

— C (Triazophos-DDT) est, de façon nette, le meilleur traitement. Il se caractérise surtout par de faibles dégâts sur capsules et boutons floraux, l'incidence, apparemment plus forte que pour les autres traitements, des dégâts non dus aux chenilles ou plus généralement du shedding non parasitaire, étant la conséquence de sa bonne efficacité.

— A (endrine-DDT) et D (Gardona-DDT) ne sont séparés que par la deuxième composante principale, D apparaissant comme plus destructeur d'*H. armigera*.

— E (Phosalone-DDT), bien que se rattachant au groupe composé des deux traitements précédents, supporte plus de dégâts, dégâts pour une part importante non dus aux chenilles : son efficacité contre *H. armigera* semble bonne.

— B (Monitor-DDT) est de loin le traitement le moins bon, il se montre particulièrement inefficace contre *H. armigera*.

En fin de campagne, l'analyse statistique des récoltes de coton-graine donnait le classement suivant :

C      B      A      E      D  
 \_\_\_\_\_

la seule différence significative observée étant entre le traitement D et tous les autres, le groupe C B A E étant homogène.

Ce résultat ne s'oppose aux conclusions précédentes que par la place du traitement B dans le groupe de tête, illustrant par là le biais que peuvent introduire les phénomènes de compensation de la part de la plante, dans les tests de comparaison basés sur la seule production des parcelles.

En définitive, l'essai visant essentiellement à déterminer la formule la plus efficace contre *H. armigera*, l'analyse multivariable des données issues des observations réalisées en cours d'expérience, permet de dire qu'il s'agit de la formule C (Triazophos-DDT). Les traitements A (endrine-DDT), D (Gardona-DDT) et E (Phosalone-DDT) apparaissent, pour ce critère, d'efficacité équivalente, la formule B (Monitor-DDT) étant la moins bonne.

## BIBLIOGRAPHIE

- ANDERSON T.W., 1958. — An introduction to multivariate statistical analysis. *J. Wiley, New York.*
- KENDALL M.G., 1972. — A course in multivariate analysis. *Griffin, London.*
- RAO C.R., 1952. — Advanced statistical methods in biometric researches. *J. Wiley, New York.*
- RAO C.R., 1965. — Linear statistical inference and its applications. *J. Wiley, New York.*
- SOKAL R.R. et P. SNEATH, 1963. — Principles of numerical taxonomy. *W.H. Freeman, San Francisco.*
- VAN DEN DRIESCH R., 1965. — La recherche des constellations de groupe à partir des distances généralisées de MAHALANOBIS. *Biométrie-Praximétrie*, 6, 1, 36-47.
- WILKS S.S., 1962. — Mathematical statistics. *J. Wiley, New York.*

## SUMMARY

*Four variables relating to observations made in the course of experiments were measured in a comparative trial of four insecticides. None, individually, enabled conclusion as to the differences in the efficiency of these products to be drawn. On the other hand, treatment by generalized variate analysis revealed statistically significant differences.*

*The generalized Mahalanobis distances lead to the separation of the five products into three groups. The principle components enable on the one hand the variables to be classified in accordance with their separating power regarding treatments, and on the other, provide the information to enable these differences to be interpreted.*

## RESUMEN

*Se han medido cuatro variables relativas a observaciones realizadas durante la experiencia, en un ensayo comparativo de cinco productos insecticidas. Ninguno permite, individualmente, concluir en cuanto a diferencias de eficacia entre los productos. En cambio, su tratamiento en el análisis de variancia generalizada hace aparecer diferencias estadística-*

*mente significativas. Las distancias generalizadas de Mahalanobis conducen a la separación de los cinco productos en tres grupos. Los componentes principales permiten, por un lado, clasificar las variables en función de su poder de separación con respecto a los tratamientos, y proporcionan, por otro lado, los elementos de interpretación de esas diferencias,*