

NOTION D'AGROSYSTEME

PREMIERE PARTIE

NOTION DE BASE ET NOTIONS DERIVEES

METHODE D'ETUDE

Chapitre 3

COMPARAISON DE DEUX OU PLUSIEURS AGROSYSTEMES

par

B. BONZON et J. DEJARDIN

H = 5497-1

~~N AG 1083/2/4~~

→ n'envoie qu'un Nouman

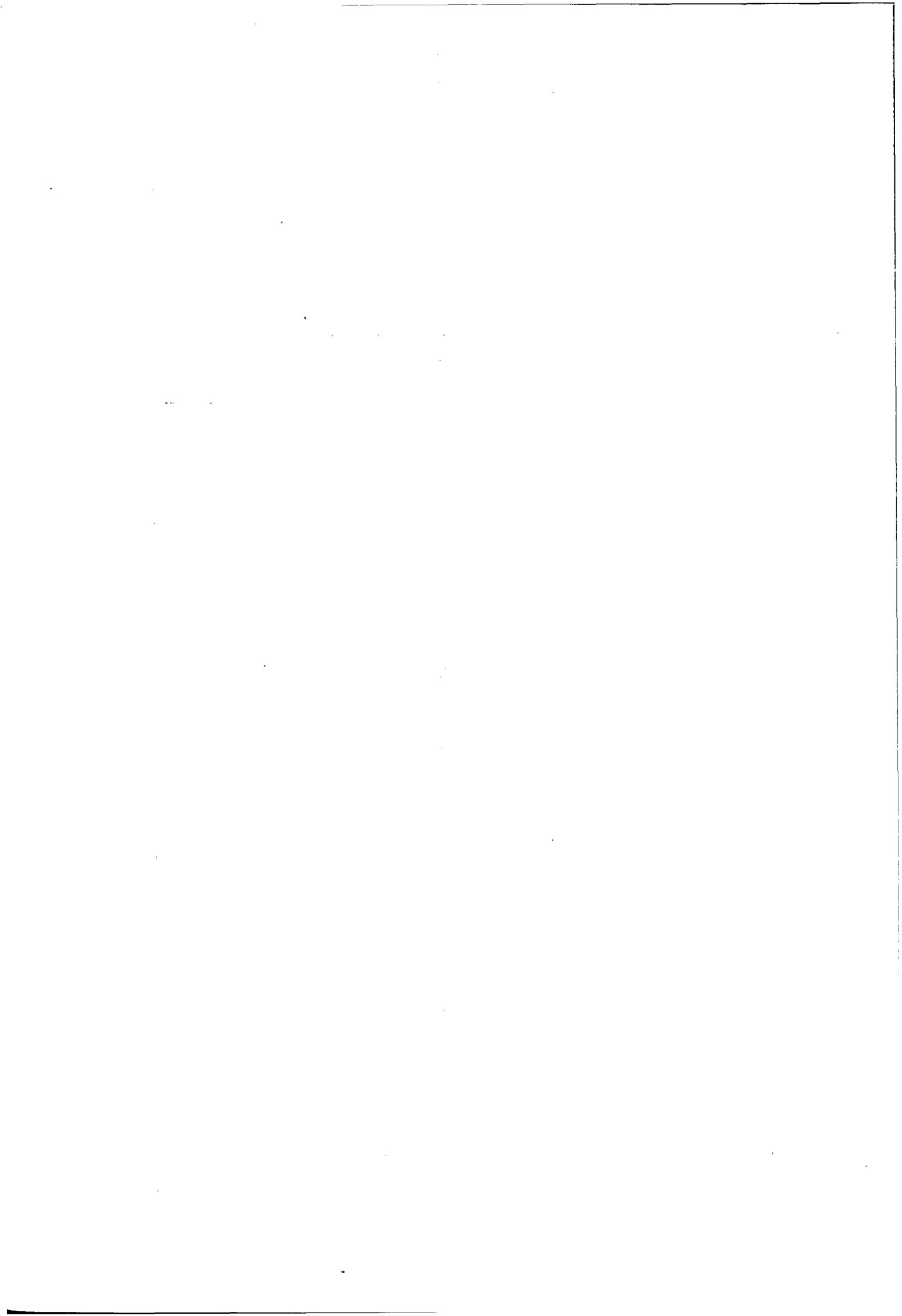
Fonds Documentaire IRD



010025728

Fonds Documentaire IRD

Cote: Bx 25728 Ex: *venezuela*



COMPARAISON DE DEUX OU PLUSIEURS AGROSYSTEMES
ASPECTS STATISTIQUES.

	<u>Pages.</u>
1 - NOTIONS D'AGROSYSTEMES EQUIVALENTS ET D'AGROSYSTEMES SEMBLABLES	2
11 - Notion d'agrosystèmes équivalents.....	2
12 - Notion d'agrosystèmes semblables	3
13 - Aspects géométriques de ces définitions.....	3
131 - Conséquences de l'équivalence et de la similitude au niveau des volumes de dispersion des observations.....	3
132 - Conséquences de l'équivalence et de la similitude au niveau des plans de régression simple.....	4
133 - Extension des conséquences précédentes au niveau des régressions partielles	5
2 - INTERET ET OBJECTIFS GENERAUX DE LA COMPARAISON D'AGROSYSTEMES. ETAPES SUCCESSIVES ET VOLUME DES ANALYSES.....	5
3 - COMPARAISON DES VARIANCES D'UN ELEMENT COMMUN A k_L AGROSYSTEMES	6
31 - Cas où $k_L = 2$	6
32 - Cas où $k_L > 2$	7
33 - Conséquences de la comparaison des variances d'un élément	7
4 - COMPARAISON DES MOYENNES D'UN ELEMENT COMMUN A k_L AGROSYSTEMES	8
41 - Cas où $k_L = 2$	8
411 - Cas où chaque agrosystème est caractérisé à partir d'au moins 30 sites.....	8
412 - Cas où le nombre de sites est inférieur à 30.....	9
42 - Cas où $k_L > 2$	9
5 - COMPARAISON DES COEFFICIENTS DE CORRELATION ENTRE DEUX ELEMENTS COMMUNS A k_L AGROSYSTEMES	10
51 - Estimation et signification du coefficient de corrélations moyen r	11
52 - Homogénéité des k_L coefficients de corrélation	12
521 - Cas où r est significativement différent de zéro	12
522 - Cas où r n'est pas significativement différent de zéro	13
53 - Conditions d'application de la méthode du coefficient de corrélation transgénéralisé pour la comparaison de plusieurs corrélations	13

6 - COMPARAISON DES COEFFICIENTS DE REGRESSION D'UN ELEMENT X SUR UN ELEMENT U COMMUNS A k_L AGROSYSTEMES	14
61 - Estimation d'un coefficient de régression moyen de X sur U....	14
62 - Tests de signification du coefficient de régression moyen b_{xu} et de nonparallélisme des droites de régression.....	15
63 - Estimation et signification du coefficient de régression moyen de U sur X et vérification du parallélisme des droites de régression correspondantes	17
64 - Remarque sur une autre procédure d'estimation du coefficient de corrélation moyen r_{xu}	17
CONCLUSION	18

COMPARAISON DE DEUX OU PLUSIEURS AGROSYSTEMES

ASPECTS STATISTIQUES.

--0--

Au chapitre précédent nous avons passé en revue les principales méthodes permettant de caractériser la partie à p éléments d'un agrosystème A.

Ayant appliqué ces méthodes à k_L agrosystèmes A, ..., D, ..., L, ou du moins à leur partie à p éléments communs U, ..., X, ..., Z, nous désirons à présent les comparer.

A priori nous concevons bien que cette comparaison ait à porter à la fois sur :

- les niveaux moyens de leurs éléments constitutifs ;
- les champs de variation de ces éléments ;
- les intensités des liens qui les unissent ;
- les relations correspondantes.

Si les p caractéristiques de ces k_L agrosystèmes sont distribuées normalement et liées linéairement, lorsqu'elles sont liées significativement - ce qui nous supposons acquis pour la suite de l'étude -, les problèmes à résoudre consistent donc en ceux de la comparaison de moyennes, de variances, de coefficients de corrélation et de coefficients de régression.

Cependant, afin de préciser l'intérêt et les objectifs généraux de ces comparaisons et, surtout, leurs enchainements logiques, nous commencerons par définir deux notions de base : celle d'agrosystèmes équivalents et celle d'agrosystèmes semblables.

Concernant par ailleurs les fondements ou les développements théoriques des méthodes qui seront examinées ensuite, le lecteur voudra bien se référer aux ouvrages de base déjà cités au précédent chapitre.

1. NOTIONS D'AGROSYSTEMES EQUIVALENTS ET D'AGROSYSTEMES SEMBLABLES.

Soient donc k_L agrosystèmes A, \dots, D, \dots, L sur chacun desquels ont été observés, indépendamment les uns des autres, les mêmes p éléments U, \dots, X, \dots, Z .

Ces éléments ont fait l'objet de n_A observations sur A, \dots, n_D sur D, \dots, n_L sur L .

Les matrices regroupant ces informations sont :

$$\hat{A}_{(p, n_A)} = \begin{bmatrix} u_1 & u_2 & \dots & u_i & \dots & u_{n_A} \\ A & A & & A^i & & A^{n_A} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_1 & x_2 & \dots & x_i & \dots & x_{n_A} \\ A & A & & A^i & & A^{n_A} \\ \vdots & \vdots & & \vdots & & \vdots \\ z_1 & z_2 & \dots & z_i & \dots & z_{n_A} \\ A & A & & A^i & & A^{n_A} \end{bmatrix} \quad (1)$$

$$\hat{D}_{(p, n_D)} = \begin{bmatrix} u_1 & u_2 & \dots & u_i & \dots & u_{n_D} \\ D & D & & D^i & & D^{n_D} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_1 & x_2 & & x_i & & x_{n_D} \\ D & D & & D^i & & D^{n_D} \\ \vdots & \vdots & & \vdots & & \vdots \\ z_1 & z_2 & & z_i & & z_{n_D} \\ D & D & & D^i & & D^{n_D} \end{bmatrix} \quad (2)$$

$$\hat{L}_{(p, n_L)} = \begin{bmatrix} u_1 & u_2 & \dots & u_i & \dots & u_{n_L} \\ L & L & & L^i & & L^{n_L} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_1 & x_2 & \dots & x_i & \dots & x_{n_L} \\ L & L & & L^i & & L^{n_L} \\ \vdots & \vdots & & \vdots & & \vdots \\ z_1 & z_2 & \dots & z_i & \dots & z_{n_L} \\ L & L & & L^i & & L^{n_L} \end{bmatrix} \quad (3)$$

1.1. Notion d'agrosystèmes équivalents.

Nous dirons que A, \dots, D, \dots, L sont équivalents si les trois séries de conditions suivantes sont réunies :

1°/ - les k_L valeurs centrales de chaque élément sont statistiquement égales :

$$\begin{aligned} \bar{u} &= \dots = \bar{u} = \dots = \bar{u} = \bar{u} \\ A & \quad \quad \quad D & \quad \quad \quad L \\ \vdots & & & & & \\ \bar{x} &= \dots = \bar{x} = \dots = \bar{x} = \bar{x} \\ A & \quad \quad \quad D & \quad \quad \quad L \\ \vdots & & & & & \\ \bar{z} &= \dots = \bar{z} = \dots = \bar{z} = \bar{z} \\ A & \quad \quad \quad D & \quad \quad \quad L \end{aligned} \quad (4)$$

2°/ - les champs de variation des k_L séries d'observations de chaque élément sont égaux :

$$\begin{aligned} s_{A^u} &= \dots = s_{D^u} = \dots = s_{L^u} = s_u \\ s_{A^x} &= \dots = s_{D^x} = \dots = s_{L^x} = s_x \\ s_{A^z} &= \dots = s_{D^z} = \dots = s_{L^z} = s_z \end{aligned} \quad (5)$$

3°/ - les k_L coefficients de corrélation de chacun des $p(p-1)/2$ couples d'éléments sont statistiquement égaux, par exemple :

$$r_{xu}^{AA} = \dots = r_{xu}^{DD} = \dots = r_{xu}^{ZZ} = r_{xu} \quad (6)$$

1.2. Notion d'agrosystèmes semblables.

Si les deux dernières séries de conditions seulement sont réunies nous dirons que les k_L agrosystèmes sont simplement semblables.

1.3. Aspects géométriques de ces définitions.

1.3.1. Conséquences de l'équivalence et de la similitude au niveau des volumes de dispersion des observations.

Considérons l'espace à p dimensions bâti sur les p caractéristiques observées et les points P_i de coordonnées $u_i, \dots, x_i, \dots, z_i$.

Dans cet espace, l'agrosystème A est représenté par le nuage

des n_A points $P_{A_i} (u_{A_i}, \dots, x_{A_i}, \dots, z_{A_i}), \dots$, l'agrosystème D par celui

des n_D points $P_{D_i} (u_{D_i}, \dots, x_{D_i}, \dots, z_{D_i}), \dots$, l'agrosystème L par celui

des n_L points $P_{L_i} (u_{L_i}, \dots, x_{L_i}, \dots, z_{L_i})$.

Si les k_L agrosystèmes sont équivalents, les n_A points P_{A_i} du premier se dispersent dans le même volume et de la même façon autour du point central $M(\bar{u}, \dots, \bar{x}, \dots, \bar{z})$ que les n_D points P_{D_i} du second et, finalement, que les n_L points P_{L_i} du L^{eme}.

S'ils ne sont que semblables les volumes et les modalités de dispersion des points P_i de chaque agrosystème restent comparables, mais chacun de ces volumes est centré sur un point M distinct des autres : le volume des points P_i est alors centré sur M $(\bar{u}, \dots, \bar{x}, \dots, \bar{z})$, celui des points P_i est centré sur M $(\bar{u}, \dots, \bar{x}, \dots, \bar{z})$, etc. Une translation de vecteur \overrightarrow{MM} permettrait alors de superposer ces deux volumes, etc...

1.3.2. Conséquence de l'équivalence et de la similitude au niveau des plans de régression simple.

L'égalité des variances et celle des coefficients de corrélations simples entraînent, naturellement, l'égalité des coefficients de régression simple de chaque couple de caractéristiques. Soit (X,U) l'un de ces $p \cdot (p-1)/2$ couples. Considérons les k_L droites de régression d'équation :

$$\begin{aligned} x \begin{matrix} (u_i) \\ A_i \\ A \end{matrix} &= \bar{x} + b_{xu} \begin{matrix} (u_i - \bar{u}) \\ A_i \\ A \end{matrix} \\ \vdots & \\ x \begin{matrix} (u_i) \\ D_i \\ D \end{matrix} &= \bar{x} + b_{xu} \begin{matrix} (u_i - \bar{u}) \\ D_i \\ D \end{matrix} \\ \vdots & \\ x \begin{matrix} (u_i) \\ L_i \\ L \end{matrix} &= \bar{x} + b_{xu} \begin{matrix} (u_i - \bar{u}) \\ L_i \\ L \end{matrix} \end{aligned} \quad (7)$$

Si les k_L agrosystèmes sont équivalents ces droites sont confondues. S'ils ne sont que semblables elles sont parallèles.

1.3.3. Extension des conséquences précédentes au niveau des régressions partielles.

Les conséquences de l'équivalence et de la similitude s'étendent, naturellement, aux relations partielles entre trois, ou plusieurs éléments, si ceux-ci apparaissent tous liés significativement entre eux. Dans ces conditions les plans de régressions partielles doivent être confondus si les k_L agrosystèmes sont équivalents, parallèles s'ils ne sont que semblables.

2. INTERET ET OBJECTIFS GENERAUX DE LA COMPARAISON D'AGROSYSTEMES. ETAPES SUCCESSIVES ET VOLUME DES ANALYSES.

. Equivalence et similitude doivent s'observer dans les conditions réelles : par exemple lorsque l'on compare les états de parcelles découpées à l'intérieur d'un champ particulièrement homogène, chacune de ces parcelles comportant plusieurs sites d'observations et pouvant être considérée, par suite, comme le support d'un agrosystème particulier.

Pour sa part, la similitude est requise sur les dispositifs expérimentaux au champ et sa vérification, lorsqu'elle est possible, conforte les résultats des analyses de variance et autorise la covariance.

Cependant, en raison de la diversité potentielle des facteurs de variations aléatoires et de la variabilité de leurs modalités d'intervention, on peut postuler qu'elles ne sont pas la règle générale.

Comparer des agrosystèmes observés dans des situations manifestement hétérogènes pour certaines caractéristiques présente, néanmoins, toujours un grand intérêt: celui de permettre la quantification d'un certain nombre d'inférences sur les causes de la variabilité des objets étudiés : le rendement d'une culture, la composition minérale d'un produit, l'évolution d'une caractéristique sol, etc...

. D'après ce qui précède (conf. paragraphe 11 ci-dessus) comparer k_L agrosystèmes nécessite donc la réalisation :

- 1°. de p séries de comparaisons de k_L moyennes,
- 2°. de p séries de comparaisons de k_L variances,
- 3°. de $p \cdot (p-1)/2$ séries (au plus) de comparaisons de k_L coefficients de corrélation,
- 4°. de $p \cdot (p-1)/2$ séries (au plus) de comparaisons de coefficients de régression.

Cependant, comme les méthodes de comparaison de moyennes supposent l'égalité des variances des populations-parents, nous commencerons par examiner le problème de la comparaison des variances.

3. COMPARAISON DES VARIANCES D'UN ELEMENT COMMUN A k_L AGROSYSTEMES.

Soit donc X l'un des p éléments observés sur les k_L agrosystèmes A, \dots, D, \dots, L et $s_{A^X}^2, \dots, s_{D^X}^2, \dots, s_{L^X}^2$ les k_L variances à comparer.

Pour mémoire, ces variances sont estimées avec $(n_A-1), \dots, (n_D-1), \dots, (n_L-1)$ degrés de liberté.

3.1. Cas où $k_L = 2$.

Si l'on a seulement deux agrosystèmes à comparer, A et D par exemple, on peut utiliser le test F (déjà appliqué au chapitre précédent pour tester la non-linéarité des droites de régression et la signification du rapport de corrélation).

Il suffit alors pour chaque élément (X ici pour l'exemple) de calculer le rapport

$$F_{X(A/D)} = \frac{s_{A^X}^2}{s_{D^X}^2}, \text{ si } s_{A^X}^2 > s_{D^X}^2 \quad (8)$$

ou le rapport

$$F_{X(D/A)} = \frac{s_{D^X}^2}{s_{A^X}^2}, \text{ si } s_{A^X}^2 < s_{D^X}^2 \quad (9)$$

et de comparer la valeur trouvée à celle théorique correspondante au seuil α retenu ($\alpha = 0.05, 0.01$ ou ...) pour (n_A-1) et (n_D-1) degrés de liberté s'il s'agit de $F_{X(A/D)}$, ou pour (n_D-1) et (n_A-1) degrés de liberté, naturellement, s'il s'agit de $F_{X(D/A)}$.

Si F observé $> F_\alpha$ les deux variances seront considérées comme différentes.

On peut également calculer la probabilité α' correspondant à la valeur de F observée et la comparer, au besoin, aux seuils α retenus (cf paragraphe 3 de l'annexe 1).

3.2. Cas où $k_L > 2$.

Si l'on a plus de deux agrosystèmes à comparer le test de BARTLETT semble le mieux adapté à la comparaison de plusieurs variances.

Il consiste :

1°/ - en le calcul de la quantité à $(k_L - 1)$ degrés de liberté

$$\chi^2_{\text{obs}} = \frac{(N - k_L) \cdot L \cdot N \cdot s_X^2 - \sum_{D=A}^L [(n_D - 1) \cdot L \cdot N \cdot s_{DX}^2]}{1 + [1/3 \cdot (k_L - 1)] \cdot \left[\sum_{D=A}^L 1/(n_D - 1) - 1/(N - k_L) \right]} \quad (10)$$

dans laquelle

$$N = \sum_{D=A}^L n_D \quad \text{et} \quad s_X^2 = \left[\sum_{D=A}^L (n_D - 1) s_{DX}^2 \right] / (N - k_L) \quad (11)$$

2°/ - en la comparaison de χ^2_{obs} à χ^2_{α} pour $(k_L - 1)$ degrés de liberté.

Si $\chi^2_{\text{obs}} > \chi^2_{\alpha}$ on conclura à l'hétérogénéité des variances comparées.

Dans le cas contraire les k_L variances de X seront considérées comme homogènes et s_X^2 constituera la meilleure approximation de la variance de X commune aux k_L agrosystèmes.

L'application de ce test ne serait satisfaisante, néanmoins, que si les effectifs n_D sont au moins égaux à 4 et si le nombre de variances à comparer n'est pas trop élevé par rapport aux nombres n_D .

Enfin, une réalisation automatique du test de BARTLETT est possible grâce à la formule approchée donnant la probabilité α' correspondant à la valeur observée de χ^2 (Annexe 1, paragraphe 4).

3.3. Conséquence de la comparaison des variances d'un élément.

La comparaison des k_L variances d'un élément aboutit ainsi à la conclusion,

- soit de leur homogénéité et dans ce cas il est possible de poursuivre les opérations de comparaison ;

- soit de leur hétérogénéité et l'on peut se poser alors la question de savoir s'il ne serait pas possible de stabiliser ces variances en transgénérant la variable. Cette question a déjà été examinée au chapitre précédent : elle se rattache à celle de la transgénération des variables envisagées pour linéariser des régressions (chap. II, parag. 3.3.1. et 3.3.2) ou normaliser des distributions (chap. II, parag. 4). S'il est impossible finalement de trouver une telle fonction on conclura à l'hétérogénéité des agrosystèmes observés pour l'élément étudié, et la comparaison des k_L moyennes $\bar{x}_A, \dots, \bar{x}_D, \dots, \bar{x}_L$ ne sera pas effectuée de même que celle des (p-1) séries de comparaisons de k_L coefficients de corrélation impliquant X.

4. COMPARAISON DES MOYENNES D'UN ELEMENT COMMUN A k_L AGROSYSTEMES.

Le test des variances de X ayant permis, par contre, de conclure qu'elles étaient homogènes on peut donc comparer maintenant les k_L moyennes correspondantes.

Comme pour les variances, le test à utiliser est différent selon le nombre d'agrosystèmes à comparer.

4.1. Cas où $k_L = 2$.

Si l'on a seulement deux agrosystèmes à comparer, A et D par exemple, comparer \bar{x}_A et \bar{x}_D peut consister simplement en le test par rapport à zéro de la variable réduite

$$t_{A \times D} = \frac{(\bar{x}_A - \bar{x}_D)}{s_{(\bar{x}_A - \bar{x}_D)}} \quad (12)$$

Selon l'importance des effectifs n_A et n_D deux cas particuliers se présentent néanmoins concernant l'estimation de $s_{(\bar{x}_A - \bar{x}_D)}$ et le test de $t_{A \times D}$.

4.1.1 - Cas où chaque agrosystème est caractérisé à partir d'au moins 30 sites

Dans ces conditions :

$$1^\circ / s_{(\bar{x}_A - \bar{x}_D)}^2 = (s_{\bar{x}_A}^2 / n_A) + (s_{\bar{x}_D}^2 / n_D) \quad (13)$$

2°/ $t_{(A-D)}$ peut être comparée à t de P(t) pour différentes valeurs α de P(t)

Plus simplement, si l'on prend les seuils $\alpha = 0.05$, $\alpha = 0.01$ et $\alpha = 0.001$, on comparera t à 1.96, 2.58 et 3.29. Si $t_{obs} > 1,96$, par exemple, les deux moyennes seront considérées comme significativement différentes au seuil $\alpha = 0.05$.

Si l'on désire plus de précision, on peut aussi déterminer la probabilité $P(t)$ correspondant à la valeur observée de t , à l'aide de l'approximation rappelée à l'annexe 1 (paragraphe 212) et comparer ensuite, au besoin, cette probabilité à α égal successivement à 0.05, 0.01, 0.001. Si $\alpha < 0.05$, par exemple, les deux moyennes seront considérées comme significativement différentes au seuil α .

4.1.2. Cas où le nombre de sites est inférieur à 30.

Dans ces conditions :

$$1^\circ / \left\{ \frac{s_A^2}{n_A} - \frac{s_D^2}{n_D} \right\} = \left\{ \frac{(n_A + n_D)}{[n_A \cdot n_D \cdot (n_A + n_D - 2)]} \right\} \cdot \left\{ s_A^2(n_A - 1) + s_D^2(n_D - 1) \right\} \quad (14)$$

2°/ $t_{A \times D}$ est comparée à la valeur de t aux différents seuils α égaux à

0.05, 0.01, 0.001 de :

- . $P(t)$ si $n_A + n_D - 2 > 30$ (on compare alors, comme ci-dessus, t observée à 1.96, 2.58 et 3.29),
- . $P(t)$ de STUDENT-FISCHER dans le cas contraire et pour $(n_A + n_D - 2)$ degrés de liberté.

On peut aussi déterminer la probabilité $P(t)$ ou $P'(t)$, correspondant à la valeur observée de t , à l'aide des approximations des valeurs de ces fonctions rappelées à l'annexe 1 (paragraphe 212 et 222) et comparer ensuite, au besoin, cette probabilité à α égal successivement à 0.05, 0.01 et 0.001.

4.2. Cas où $k_L > 2$.

Lorsque l'on a plus de deux agrosystèmes à comparer, la probabilité de tester, à l'aide du test précédent et à un seuil α donné, une différence significative entre deux moyennes est fonction, en partie, des rangs de ces moyennes dans l'ensemble des k_L moyennes classées de la plus petite à la plus grande et de ce nombre k_L . Il en résulte, plus précisément, que le risque de première espèce (celui d'accepter l'hypothèse d'une différence significative au seuil α retenu alors qu'elle est fausse) augmente avec le nombre de comparaisons à réaliser

$$c_{k_L}^2 = k_L \cdot (k_L - 1) / 2 \quad (15)$$

Pour diminuer ce risque, NEWMAN puis KEULS ont mis au point un test qui repose, pratiquement, sur des tables de valeurs critiques d'une constante $q(1-\alpha)$ pour différents seuils de signification α . Pour un seuil donné, ces valeurs critiques dépendent de la différence de rang des moyennes comparées et du degré de liberté de la variance commune aux k_L moyennes comparées.

Si c et c' sont ainsi les rangs de deux moyennes données à comparer ($c' < c$) et si s_X^2 est la variance de X commune aux k_L agrosystèmes avec $(N - k_L)$ degrés de liberté, la plus petite différence significative au seuil α entre deux moyennes sera donnée

$$\text{par } d_{(c-c')} = q(1-\alpha) \cdot \sqrt{s_X^2 / n} \quad (16)$$

si les k_L effectifs $n_A, \dots, n_D, \dots, n_L$ sont tous égaux à n ,

$$\text{par } d_{(c-c')} = q(1-\alpha) \cdot \sqrt{(n_c + n_{c'}) / (2 \cdot n_c \cdot n_{c'}) \cdot s_X^2} \quad (17)$$

si les k_L effectifs sont en partie différents, n_c étant l'effectif correspondant à la moyenne \bar{x}_c de rang c , $n_{c'}$ étant celui correspondant à la moyenne $\bar{x}_{c'}$ de rang c' (d'après C.Y. KRAMER, 1956).

La quantité $q(1-\alpha)$ est prise dans la table donnant ces quantités au seuil α à la ligne $(N - k_L)$ degrés de liberté et dans la colonne $(c - c' + 1)$.

Si

$$(\bar{x}_c - \bar{x}_{c'}) > d_{(c-c')}$$

les deux moyennes seront considérées comme significativement différentes.

5. COMPARAISON DES COEFFICIENTS DE CORRELATION ENTRE DEUX ELEMENTS COMMUNS A k_L AGROSYSTEMES.

Soient X et U deux des éléments communs aux k_L agrosystèmes étudiés.

Nous désirons maintenant comparer les k_L coefficients de corrélation

$$r_{xu, \dots, r_{xu}, \dots, r_{xu}$$

AA DD LL

dont certains peuvent ne pas être significativement différents de zéro, voir de signe opposé à celui des autres.

Plus précisément, faisant l'hypothèse de l'existence d'un lien unique moyen d'intensité ρ commun aux k_L agrosystèmes (nous avons pour cela déjà vérifié l'homogénéité des variances) nous voulons :

- 1°/ - estimer ρ et le seuil de signification de son estimation r ,
- 2°/ - vérifier l'homogénéité des k_L coefficients de corrélation observés, que r , coefficient de corrélation moyen estimé, soit ou non significativement différent de zéro.

5.1. Estimation et signification du coefficient de corrélation moyen r .

Nous avons vu au chapitre précédent (paragraphe 232) que la variable

$$z = L_N (1 + r) / (1 - r) \tag{18}$$

est distribuée presque normalement quelque soit la valeur exacte ρ de r et l'effectif n des couples (x_i, u_i) .

Nous avons vu également que la variance de z était donnée par

$$s_z^2 = 1 / (n-3) \tag{19}$$

Dans ces conditions si à $r_{xu}, \dots, r_{xu}, \dots, r_{xu}$, nous faisons correspondre
AA DD LL

$z_A, \dots, z_D, \dots, z_L$, nous pouvons facilement :

- 1°/ - estimer \bar{z} , moyenne pondérée des coefficients $z_A, \dots, z_D, \dots, z_L$, par

$$\bar{z} = \sum_{D=A}^L (n_D - 3) z_D / \sum_{D=A}^L (n_D - 3) \tag{20}$$

et $s_{\bar{z}}^2$ variance de \bar{z} , par :

$$s_{\bar{z}}^2 = 1 / \sum_{D=A}^L (n_D - 3) \tag{21}$$

- 2° - tester si \bar{z} est significativement différent de zéro en calculant la valeur de la variable réduite

$$t = \frac{\bar{z}}{s_{\bar{z}}} \tag{22}$$

et en comparant cette valeur à celles données à différents seuils par la fonction $P(t)$ ou, plus simplement, à $t_{0.05} = 1,96$, $t_{0.01} = 2,58$, etc..., ou encore, en calculant t , puis la probabilité correspondante \det à l'aide de l'approximation rappelée à l'annexe 1 (paragraphe 212) et en comparant au besoin cette probabilité aux seuils α égaux à 0.05, 0.01 et 0.001;

3°/ - calculer

$$r = (e^{2\bar{z}} - 1) / (e^{2\bar{z}} + 1) \quad (23)$$

estimation du coefficient de corrélation moyen ρ .

5.2. Homogénéité des k_L coefficients de corrélation.

5.2.1. Cas où r est significativement différent de zéro.

Faisons alors l'hypothèse que les k_L coefficients de corrélation transgénéérés $z_A, \dots, z_D, \dots, z_L$, proviennent d'une même population de moyenne exacte z (transgénéérée de ρ inconnu) et de moyenne estimée \bar{z} . Dans ces conditions :

1°/ - les k_L quantités

$$(z_D - \bar{z}) / s_{\bar{z}} = (z_D - \bar{z}) \cdot \sqrt{n_D - 3} \quad (24)$$

doivent être distribuées normalement autour de zéro avec un écart-type égal à l'unité ;

2°/ - la quantité

$$U = \sum_{D=A}^L (n_D - 3) \cdot (z_D - \bar{z})^2 \quad (25)$$

doit être distribuée selon une loi de χ^2 à $(k_L - 1)$ degrés de liberté.

Si l'on constate alors que $U < \chi^2_{0.05}$, on conclura qu'effectivement les k_L coefficient z_D , donc les k_L coefficients r_{XU}^{DD} sont homogènes et que r transformée inverse de \bar{z} est une bonne estimation du lien moyen commun au couple (X, U) dans les k_L agrosystèmes.

Dans le cas contraire ($U \geq \chi^2_{0,05}$) cette estimation est évidemment impossible.

Le test de χ^2 peut être réalisé pour sa part automatiquement, comme indiqué à l'annexe 1 paragraphe 5.

5.2.2. Cas où r n'est pas significativement différent de zéro.

Dans l'hypothèse où les k_L coefficients de corrélation transgénérés proviennent d'une même population dont la moyenne exacte est nulle,

1°/ - les k_L quantités

$$(z_D - \bar{z}) / s_{\bar{z}} = z_D \cdot \sqrt{n_D - 3} \quad (26)$$

sont encore distribuées normalement autour de zéro avec un écart-type égal à 1,

2°/ - la quantité

$$U_0 = \sum_{D=A}^L (n_D - 3) \cdot (z_D - \bar{z})^2 = \sum_{D=A}^L (n_D - 3) \cdot z_D^2 \quad (27)$$

est elle aussi distribuée selon une loi de χ^2 , mais à k_L degrés de liberté cette fois.

Si $U_0 < \chi^2_{0,05}$, on conclura à l'homogénéité des k_L coefficients de corrélation qui seront donc considérés comme équivalents à zéro. En d'autres termes, les caractéristiques X et U seront dites globalement non-liées significativement.

Si au contraire $U_0 \geq \chi^2_{0,05}$, on conclura à l'hétéronégéité des k_L coefficients de corrélation.

5.3. Conditions d'application de la méthode du coefficient de corrélation transgénéré pour la comparaison de plusieurs corrélations.

Du fait du caractère seulement quasi-normal de la distribution de z, la méthode qui vient d'être présentée pour estimer \bar{z} et tester l'homogénéité des k_L coefficients de corrélation ne serait valable que si ce nombre est petit vis à vis de l'importance moyenne des échantillons $n_A, \dots, n_D, \dots, n_L$.

6. COMPARAISON DES COEFFICIENTS DE REGRESSION D'UN ELEMENT X SUR UN ELEMENT U COMMUNS A k_L AGROSYSTEMES.

En principe, avec les conditions dans lesquelles nous nous sommes placés au départ - normalité des distributions des variables et linéarité des régressions -, la comparaison des coefficients de régression est redondante avec l'ensemble des comparaisons des variances et des coefficients de corrélation : si les variances de U et de X sont homogènes et si les coefficients de corrélation le sont aussi, les coefficients de régression b_{xu} et b_{ux} doivent l'être également.

Mais, lorsque la linéarité des régressions n'a pu être testée du fait de l'insuffisance des effectifs ($n_D < 20$) - cas fréquent - et, ou lorsque l'on se trouve dans des conditions limites d'acceptation des résultats des tests d'homogénéité des variances ou des coefficients de corrélation, tester l'homogénéité des k_L coefficients b_{ux} , est utile.

D'autre part, le test mis en oeuvre repose sur l'estimation d'un coefficient de régression moyen dont la valeur peut être utile ou nécessaire à l'interprétation agronomique des résultats, et dont il est possible, par ailleurs, de tester directement la signification.

6.1. Estimation d'un coefficient de régression moyen de X sur U.

Faisons donc l'hypothèse que les k_L coefficients de régression b_{xu} sont égaux :

$$b_{xu} = b_{xu} = \dots = b_{xu} = \dots = b_{xu}$$

AA BB DD LL

Dans ces conditions nous pouvons estimer un coefficient de régression moyen b_{xu} par :

$$b_{xu} = \frac{\sum_{D=A}^L SPE_{xu}}{\sum_{D=A}^L SCE_u} \quad (28)$$

en désignant par

$$SPE_A^{xu}, \dots, SPE_D^{xu}, \dots, SPE_L^{xu}$$

écarts de X et U

les k_L sommes des produits des

et par

$$SCE_A^u, \dots, SCE_D^u, \dots, SCE_L^u$$

les k_L sommes des carrés de U.

En effet, si

$$b_{xu} = \dots = b_{xu} = \dots = b_{xu} \quad (29)$$

AA DD LL

nous devons vérifier que

$$\frac{SPE_{xu}}{SCE_u} = \dots = \frac{SPE_{xu}}{SCE_u} = \dots = \frac{SPE_{xu}}{SCE_u} = \frac{\sum_{D=A} SPE_{xu}}{\sum_{D=A} SCE_u} \quad (30)$$

6.2 - Tests de signification du coefficient de régression moyen b_{xu} et de non-parallélisme des droites de régression.

Considérons alors successivement :

1°/ les k_L droites de régression particulières telles que

$$x(u_i) = \bar{x} + b_{xu} \cdot (u_i - \bar{u}) \quad (31)$$

D D D D

2°/ les k_L droites de régression moyennes définies à partir du coefficient de régression moyen b_{xu} telles que

$$x'(u_i) = \bar{x} + b_{xu} \cdot (u_i - \bar{u}) \quad (32)$$

D D D D

3°/ l'équation des écarts sur l'ensemble des k_L agrosystèmes

$$(x_i - \bar{x}) = (x_i - \bar{x}) + (\bar{x} - \bar{x}) \quad (33)$$

D D D D

dont le terme $(x_i - \bar{x})$ peut se décomposer lui-même en :

$$(x_i - \bar{x}) = \{x_i - x(u_i)\} + \{x'(u_i) - \bar{x}\} + \{x(u_i) - x'(u_i)\} \quad (34)$$

D D D D D D D

4°/ la somme des carrés des écarts correspondante :

$$Q = Q' + Q_A \quad (35)$$

dont les termes ont respectivement $(N-1)$, $(N-k_L)$ et (k_L-1) degrés de liberté et dont le terme Q' peut lui-même se décomposer en trois termes ayant respectivement $(N - 2k_L)$, 1 et $(k_L - 1)$ degrés de liberté :

$$Q' = Q_{xu} + Q_R + Q_{HP} \quad (36)$$

5°/ les carrés moyens de chacun des termes précédents de la somme des carrés des écarts. Ces carrés moyens représentent :

$$s_X^2 = Q/(N - 1) = \left\{ \sum_D \sum_i (x_i - \bar{x})^2 \right\} / (N-1) \quad (37)$$

une estimation de la variance de X indépendante de toute hypothèse sur X ;

$$s_A^2 = Q_A / (k_L - 1) = \left\{ \sum_D \sum_I (\bar{x} - \bar{x})^2 \right\} / (k_L - 1) \quad (38)$$

une estimation de la variance de X s'il n'y a aucune influence significative sur X de la part d'un facteur global "agrosystème" (question qui sera examinée au chapitre suivant) ;

$$s_X'^2 = Q' / (N - k_L) = \left\{ \sum_D \sum_i (x_i - \bar{x})^2 \right\} / (N - k_L) \quad (39)$$

une estimation de la variance de X, abstraction faite d'un éventuel effet "agrosystème", mais non d'une influence éventuelle de la variable U sur la variable X ;

$$s_{X,U}^2 = Q_{X,U} / \nu = \left\{ \sum_D \sum_i \{x_i - x(u_i)\}^2 \right\} / \nu = \sum_D (SCE_{X,DD} - b_{XU}^2 \cdot SCE_{U,DD}) / \nu$$

avec $\nu = (N - 2k_L)$, (40)

une estimation de la variance de X, abstraction faite de l'influence de tout facteur de variation ;

$$s_R^2 = Q_R / 1 = \sum_D \sum_i \{x'(u_i) - \bar{x}\}^2 = b_{XU}^2 \cdot \sum_D SCE_{U,DD} \quad (41)$$

une estimation de la variance de X s'il n'y a pas de liaison linéaire entre X et U sur l'ensemble des k_L agrosystèmes ;

$$s_{NP}^2 = Q_{NP} / (k_L - 1) = \left\{ \sum_D \sum_i \{x(u_i) - x'(u_i)\}^2 \right\} / (k_L - 1) \quad (42)$$

soit encore

$$s_{NP}^2 = \sum_D (b_{XU,DD}^2 - b_{XU}^2) \cdot SCE_{U,DD} / (k_L - 1) \quad (43)$$

une estimation de la variance de X si les k_L droites de régression sont parallèles à la droite de régression moyenne.

Ces différentes estimations conditionnelles de la variance de X nous permettent alors de tester simultanément :

. la signification du coefficient de régression moyen b_{xu} en comparant s_R^2 à $s_{x,u}^2$ l'aide du test F. Si

$$F_R = s_R^2 / s_{x,u}^2 \quad (44)$$

est supérieur ou égal à F_α pour 1 et $(N-2k_L)$ degrés de liberté, alors b_{xu} sera considéré comme significativement différent de zéro au seuil de signification α ;

. le non-parallélisme des droites de régression avec la régression moyenne, en comparant s_{NP}^2 à $s_{x,u}^2$. Si

$$F_{NP} = s_{NP}^2 / s_{x,u}^2 \quad (45)$$

est supérieur ou égal à F_α pour (k_L-1) et $(N-2k_L)$ degrés de liberté, alors les k_L droites de régression de X sur U devront être considérées comme globalement non-parallèles.

6.3. Estimation et signification du coefficient de régression moyen de U sur X.

Vérification du parallélisme des droites de régression correspondantes.

L'analyse des k_L régressions de U sur X se conduit, naturellement, de la même façon que celle des régressions inverses.

Toutefois, elle peut être limitée à la seule estimation du coefficient de régression moyen b_{ux} si :

1°/ - la linéarité des k_L droites de régression de U sur X a été vérifiée en même temps que celle des k_L droites de régression de X sur U, à l'occasion de la caractérisation de chaque agrosystème ;

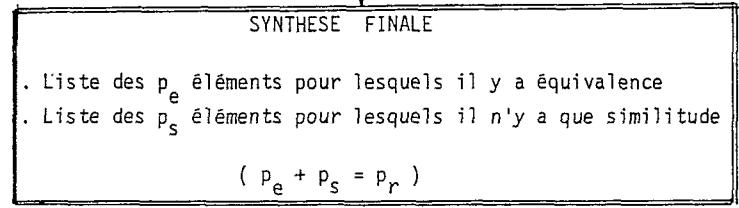
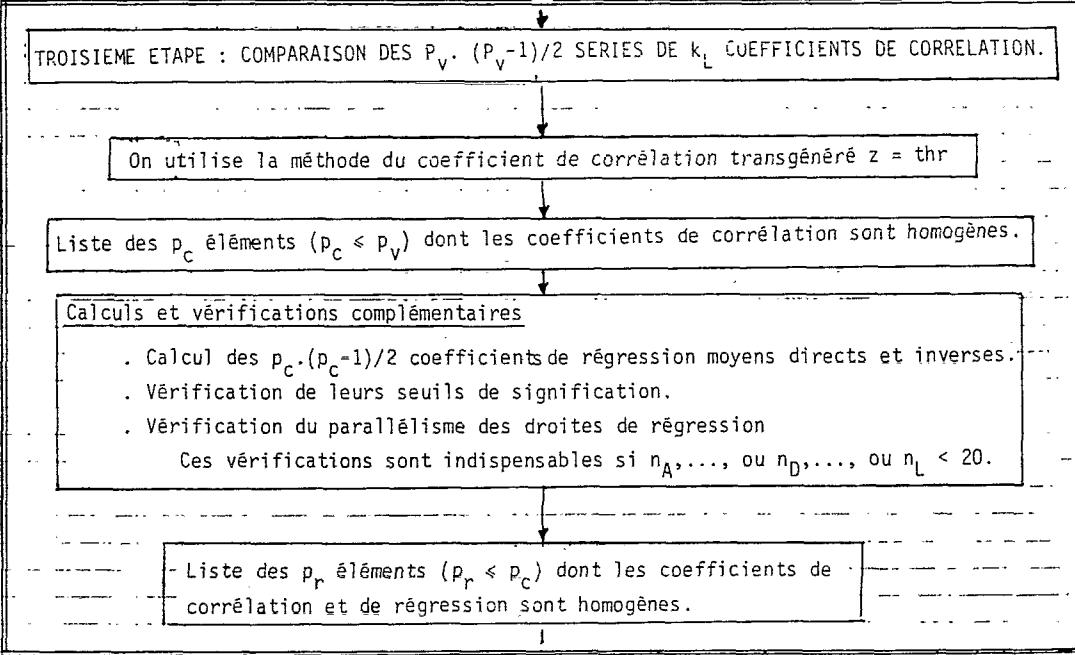
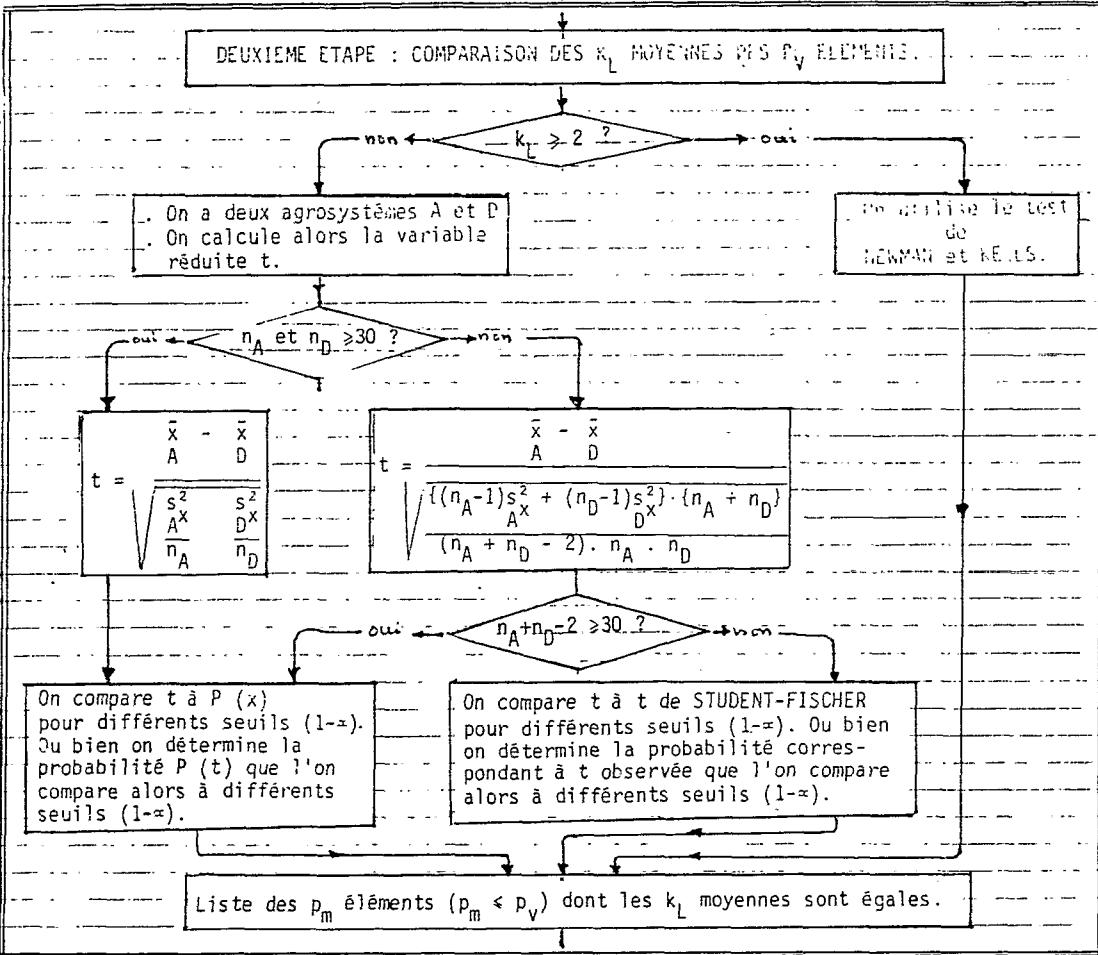
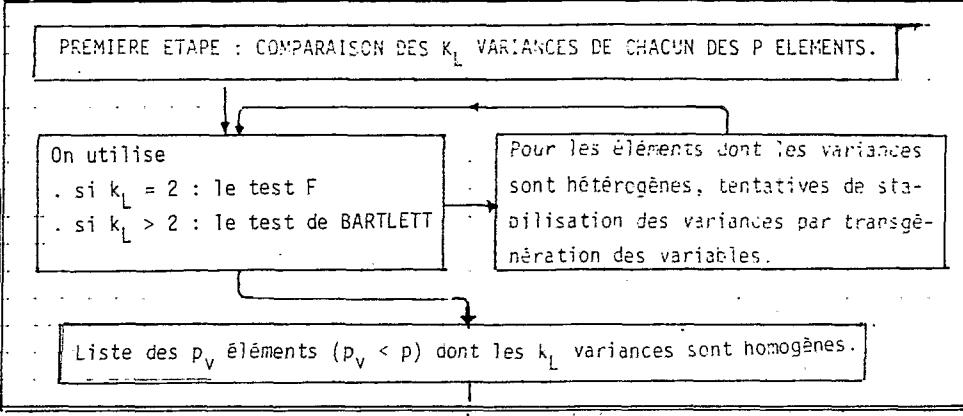
2°/ - l'homogénéité des k_L coefficients de corrélation a été également vérifiée.

6.4. Remarque sur une autre procédure d'estimation du coefficient de corrélation moyen r_{xu} .

Comme on vient de le voir, les tests d'homogénéité des coefficients de corrélation et de régression sont indépendants les uns des autres et sont redondants lorsque l'on a pu vérifier antérieurement la linéarité des régressions au niveau de chaque agrosystème. En fait, même lorsque cette vérification

COMPARAISON DE DEUX OU PLUSIEURS AGROSYSTEMES

Soient k_L agrosystèmes A, ..., D, ..., L, à p éléments communs U, ..., Y, ..., Z, observés chacun sur $n_A, \dots, n_D, \dots, n_L$ sites. Les distributions des éléments sont normales. La linéarité des régressions a été vérifiée chaque fois que $n_D > 20$.



préalable n'a pu être effectuée, on pourrait se limiter aux tests d'homogénéité des coefficients de régression de X sur U et de U sur X et déduire de leur homogénéité celle d'un coefficient de corrélation moyen r_{xu} que l'on estimerait alors à partir de la relation

$$r_{xu} = \sqrt{b_{xu} \cdot b_{ux}} = \frac{\sum_D SPE_{xu}}{\sum_D SCE_x} / \sqrt{\frac{\sum_D SCE_x \cdot \sum_D SCE_u}{\sum_D SCE_x \cdot \sum_D SCE_u}} \quad (46)$$

Cette estimation serait, cependant, moins précise que celle obtenue par la méthode présentée plus haut (paragraphe 5).

Pour cette raison, il serait préférable de tester simultanément coefficients de corrélation et coefficients de régression.

CONCLUSION.

La comparaison d'agrosystèmes peut viser plusieurs objectifs.

D'une façon générale, les résultats des comparaisons des variances, des moyennes, des coefficients de corrélation et de régression doit permettre de préciser les éléments pour lesquels les agrosystèmes sont équivalents, semblables, ou différents.

Du fait du nombre élevé de comparaisons à réaliser il convient toutefois d'être exigeant quant aux seuils d'acceptation de l'homogénéité.

En réalité, il semble qu'il serait judicieux d'établir équivalence et similitude à plusieurs niveaux de probabilité : ceci rendrait probablement mieux compte des différences qui existent aux niveaux des éléments et de leurs relations mutuelles.

L'ensemble des opérations à réaliser peut finalement se présenter comme sur l'organigramme ci-contre.

