

CENTRE TECHNIQUE FORESTIER TROPICAL

Département forestier du C.I.R.A.D.

**Etude des répartitions spatiales
Utilisation de la méthode de Diggle**

Juillet 1989
Jacques Gignoux

UTILISATION DE LA METHODE DE DIGGLE

On distingue au sein des méthodes d'étude des répartitions spatiales deux grandes catégories : les méthodes par comptages d'individus dans des grilles de quadrats (méthode de Greig-Smith et de Chessel) et les méthodes par étude des distributions de distances, comme la méthode de Diggle. Cette méthode utilise les distances entre individus et les distances d'un point d'échantillonnage fixé à un individu pour tester l'écart d'une répartition spatiale à une distribution de type aléatoire. Nous allons tout d'abord rappeler les principes de la méthode, puis tenter de déduire un schéma d'interprétation des résultats qu'elle fournit permettant d'aller plus loin que la simple caractérisation d'une distribution régulière, aléatoire ou agrégative.

1. La méthode de Diggle - Construction des tests.

A partir d'une cartographie précise des individus dans le plan, on peut calculer une distribution observée des distances entre chaque individu et son plus proche voisin ; on peut faire de même avec des distances entre un point d'échantillonnage fixé et le plus proche individu. La méthode de Diggle fournit deux tests statistiques permettant de comparer ces distributions observées aux distributions théoriques des distances, aisément calculables dans l'hypothèse d'une répartition spatiale aléatoire des individus. On peut alors, à l'aide de ces tests, caractériser qualitativement et surtout quantitativement la répartition spatiale observée, les deux approches - distances inter-individus et distances point-individu - étant complémentaires.

a) Calcul des distributions théoriques.

Dans l'hypothèse d'une répartition aléatoire des individus, on peut calculer la distribution théorique des distances inter-individus en négligeant les effets de bords, qui compliquent considérablement le calcul ; l'expression calculée sera donc valable "loin des bords" - nous discuterons de ce point ultérieurement.

Pour un point donné, la probabilité d'avoir un voisin à une distance inférieure ou égale à w est $\pi w^2/A$, A représentant la surface de la région étudiée. On a alors la fonction de répartition des distances :

$$G(w) = P(W < w) = 1 - (1 - \pi w^2/A)^{n-1}$$

avec : n = nombre d'individus ;

W = variable aléatoire exprimant la distance d'un individu à son plus proche voisin.

On peut approcher cette fonction, pour n "grand", par :

$$G(w) = 1 - e^{-\mu \pi w^2},$$

avec $\mu = n/A$.

Pour les distances point d'échantillonnage - plus proche individu, le calcul, par un raisonnement similaire, conduit à :

$$F(x) = 1 - e^{-\mu \pi x^2},$$

x représentant la distance d'un point d'échantillonnage au plus proche individu. On peut interpréter $F(x)$ de façon simple : $1 - F(x)$ représente une estimation de l'espace vide, c'est-à-dire de la surface contenant tous les points situés à une distance supérieure ou égale à x de n'importe quel individu (voir Figure 1).

b) Estimation des distributions des distances de la situation observée.

On estime les distributions de distances par les distributions empiriques correspondantes ; pour $G(w)$, on a :

$$\hat{G}_1(w) = n_w/n,$$

avec : w = distance au plus proche voisin ;

n_w = nombre d'individus ayant leur plus proche voisin à une distance inférieure à w .

De manière évidente, on calcule $\hat{F}_1(x)$ selon un procédé similaire.

Pour supprimer les effets de bords, on n'utilise dans cette formule, pour chaque valeur de x ou w , que les individus ou les points qui sont plus proches de leur plus proche voisin que de la bordure de la parcelle (voir annexe).

c) Construction des tests.

Une statistique de test est donnée par :

$$dw = \sup |G(w) - \hat{G}(w)|$$

$$\text{et } dx = \sup |F(x) - \hat{F}(x)|$$

On compare la valeur de dw (ou dx) obtenue pour l'échantillon observé avec celles de 99 échantillons issus de simulations d'une distribution aléatoire. Si le dw observé est très supérieur aux autres (par exemple à au moins 95 des échantillons simulés), on conclut au rejet de l'hypothèse de distribution aléatoire pour une distance égale à w (ou x).

d) Complément pour les distances point - individu.

Upton et Fingleton (1981) proposent une vérification supplémentaire dans le cas d'une distribution régulière : ils suggèrent de représenter graphiquement $\mu \pi x^2 - \hat{F}(x)$ en fonction de x . En effet, si on a une distribution régulière de points, alors il existe une distance x pour laquelle les cercles de rayon x ne se recouvrent pas : ils totalisent alors une surface de $n \pi x^2$, soit une proportion $\mu \pi x^2$ de la surface totale. Si la distribution est régulière, alors on doit avoir $\mu \pi x^2 = \hat{F}(x)$ jusqu'à un certain seuil qui représente la demi distance entre individus, et ensuite $\mu \pi x^2 > \hat{F}(x)$ (voir Figure 1).

e) Problèmes posés.

*** les effets de bords.**

Pour atténuer l'effet de bord qui empêche les distributions théoriques de G et F d'être exactes, Diggle (1983) propose d'utiliser plutôt pour la construction des tests les moyennes $\bar{G}(w)$ et $\bar{F}(x)$ des 99 distributions simulées $\hat{G}_1(w)$ et $\hat{F}_1(x)$.

La solution la plus rigoureuse serait de n'analyser que la partie centrale du relevé, de façon à laisser autour de cette zone centrale une bordure de largeur égale à la distance maximale possible entre deux points de la région centrale : on serait alors bien "loin des bords", comme l'exigent les expressions analytiques de G et F (voir Figure 1). Tous les points de la région étudiée serviraient au calcul des distances, mais seuls ceux de la région centrale seraient pris en compte dans la construction des tests.

Il est évident que cette solution, mathématiquement idéale, est rarement matériellement réalisable ; mais, pour une densité de points élevée dans toute la parcelle, on peut penser que les fortes valeurs de w ou x seront rares, et qu'en conséquence le biais de la distribution théorique observé pour les fortes valeurs de x et w n'aura que peu d'incidence sur les tests. On pourrait alors, dans ce cas, envisager une réduction de la largeur de la bordure. Il serait de toutes façons profitable d'étudier plus précisément l'impact de la largeur de la bordure sur la précision des résultats.

Ripley (1981) propose un compromis pour les parcelles rectangulaires, qui consiste à reproduire 8 fois la parcelle sur chacun de ses côtés, ce qui revient à en supprimer les bords en en faisant un tore, et donc à supposer que l'environnement de la parcelle est de même nature que la parcelle.

*** les points d'échantillonnage fixés pour le calcul de F.**

Il convient de savoir comment ces points doivent être placés et combien sont nécessaires. Si les divers auteurs sont d'accord pour prendre une distribution régulière des points selon une grille rectangulaire, leur avis varie sur le nombre de points à utiliser. Ripley (1981) propose de prendre un nombre de points compris entre $10n$ et $20n$, avec un maximum de 1000. Diggle (1983) serait plutôt favorable à un nombre à peu près égal à n . Upton et Fingleton (1985) tranchent en suggérant de prendre au moins n points, et de préférence au moins 100. Le tout est de trouver le compromis entre le temps de calcul de l'ordinateur et la précision des résultats.

2. L'interprétation des résultats.

a) les résultats obtenus.

Nous disposons au CTFT-Guyane d'un programme "DIGGLE", écrit en langage C, qui permet d'analyser les données du dispositif forêt naturelle de Paracou par la méthode décrite précédemment.

Ce programme utilise les expressions analytiques approchées des distributions théoriques de F et G et non pas les moyennes des distributions simulées. Pour F, le nombre de points d'échantillonnage est de 12×12 , qui est optimal pour étudier la répartition de 100-200 individus (d'après Diggle).

Les sorties du programme (voir Figure 2) donnent :

- des graphiques de dw et dx observés en fonction respectivement de w et x , avec l'enveloppe des 99 dw ou dx simulés ;

- deux tableaux donnant la fonction de répartition des dw et des dx (nombre d'échantillons simulés pour lesquels dw ou dx est supérieur à la valeur tabulée), qu'on peut comparer à la valeur de dw ou dx observé ;

- un graphique de $\hat{F}(x) - \mu \pi x^2$ en fonction de x .

b) Comment interpréter ces résultats ?

*** rejet de l'hypothèse nulle.**

Les tests basés sur dw et dx concluent au rejet de l'hypothèse nulle de distribution aléatoire des individus dans les cas où les statistiques observées dépassent un certain nombre de valeurs obtenues par simulation, ce nombre étant choisi en fonction du niveau de signification qu'on désire garantir (en général 95 ou 99).

*** distinction entre distributions agrégatives et régulières.**

On peut cependant aller plus loin que le seul rejet de l'hypothèse nulle. En effet, on peut distinguer à l'aide des graphiques les distributions régulières des distributions agrégatives. On a la situation suivante :

Distribution agrégative	Distribution régulière
$\hat{G}(w) > G(w)$	$\hat{G}(w) < G(w)$
$\hat{F}(x) < F(x)$	$\hat{F}(x) > F(x)$

*** puissance des tests.**

D'après Diggle (1979), dx serait un meilleur détecteur de tendances agrégatives, alors que dw détecterait mieux les situations régulières.

*** utilisation de $\hat{F}(x) - \mu \pi x^2$.**

Dans le cas de distributions régulières seulement, on peut déduire du graphique une approximation de la demi distance entre arbres, qui est l'abscisse du point où la courbe "décroche".

c) Toujours plus loin avec la méthode de Diggle...

Il nous a paru intéressant de tenter d'interpréter plus précisément les valeurs des diverses distances fournies par les graphiques de dw et dx (abscisse du maximum, des points d'ordonnée nulle, paliers...) : en effet, pour caractériser une répartition spatiale, il peut être nécessaire de connaître la distance moyenne entre arbres, entre agrégats, la densité intra-agrégats, la taille ou la forme de ces agrégats. Nous avons donc créé plusieurs répartitions spatiales type (voir Figure 3) dont les paramètres étaient bien connus, pour explorer les qualités prédictrices de la méthode vis à vis de ces paramètres.

*** cas des distributions régulières.**

Nous avons testé 9 distributions présentant des degrés divers de régularité (voir Figure 3). Nos conclusions sont les suivantes (voir Figure 4 et Tableau 1) :

- la régularité est détectable même pour les distributions agrégatives avec disposition régulière des arbres à l'intérieur des agrégats ;

- on peut prédire de façon très sûre la distance entre arbres d par le point w_0 de la courbe ($w_0 \approx d$);

- de même, il est possible d'estimer l'écart d'une distribution par rapport à une distribution parfaitement régulière en comparant w_{max} à w_0 . Si on considère que les arbres sont situés à l'intérieur de cercles de rayon R disposés suivant

une grille à une distance d fixée les uns des autres, on a alors une très bonne relation entre R/d et w_{max} ($w_{max}=101-4 R/d$; $P<0.05$; $r^2=1.00$). Un examen plus attentif révèle que l'on a en fait $R/d \approx (w_0 - w_{max})/w_0$;

- pour ce qui concerne $F(x)$, il apparaît que l'abscisse du point de décrochement x_D de la courbe $\hat{F}(x) - \mu \pi x^2$ semble moins efficace que w_0 pour estimer la distance moyenne entre arbres ($x_D=1.39+0.37 d$; $P<0.05$; $r^2=0.91$).

* cas des distributions agrégatives.

Nous ne sommes pas entrés dans une nomenclature très détaillée. Nous avons simplement essayé de retrouver les paramètres suivants à l'aide de la méthode de Diggle : taille moyenne des agrégats, distance inter-agrégats et densité intra-agrégats. 19 distributions agrégatives pour lesquelles ces paramètres sont connus ont été créées (voir Figure 3).

Les résultats obtenus sont moins nets que pour les distributions régulières. Il semblerait que le seul paramètre réellement prédictible sans trop d'erreur soit la distance inter-agrégats, qu'on peut déduire de la valeur de x_∞ ($dist=10.8+0.0203 x_\infty^2$; $P<0.05$; $r^2=0.91$; voir Figure 5 et Tableau 2). On observe aussi une relation entre densité et w_{max} ($densité=299-1.8 w_{max}^2$; $P<0.05$; $r^2 = 0.42$), mais qui est trop floue pour permettre des prédictions.

Malgré la diversité des situations de répartitions agrégatives qu'on ne peut caractériser qu'à l'aide de nombreux paramètres, ces résultats permettent tout de même une première description globale de ces distributions. Il sera sans doute nécessaire, pour étudier à fond une distribution agrégative, d'avoir recours à d'autres méthodes ayant une approche plus nettement axée sur la notion d'agrégat (méthode de Chessel et classification automatique).

Il ressort de cette étude que la méthode de Diggle permet non seulement de détecter facilement les distributions régulières des distributions agrégatives et aléatoires, mais aussi d'aller plus loin en quantifiant certaines caractéristiques de ces distributions. Néanmoins, pour une étude complète d'une distribution donnée, l'utilisation de cette méthode en parallèle avec d'autres serait profitable. Enfin, il manquerait encore pour bien maîtriser cette méthode une analyse plus complète du biais introduit par les effets de bord et les remèdes qu'on peut y apporter.

ANNEXE
Calculs effectués par l'ordinateur

Nous prenons ici l'exemple du calcul et du test de dw , celui de dx s'effectuant de manière similaire.

1) Calcul, pour chaque individu, de :

w : distance à son plus proche voisin ;

b : distance à la plus proche bordure.

On obtient alors, pour n individus, $2n$ valeurs dont :

$2k$ valeurs (k valeurs de b et k valeurs de w) telles que $w > b$;

$n-k$ valeurs de w utilisables pour le calcul ;

$n-k$ valeurs de b utilisables pour le calcul.

Des $2k$ valeurs vérifiant $w > b$, on ne garde que les k valeurs de b , qu'on distingue en les notant c .

2) Calcul des effectifs suivants, pour v variant sur toute la gamme des distances possibles (notation : $N(x > v)$ = nombre d'individus pour lesquels on a $x > v$) :

$N(b \leq v)$; $N(c \leq v)$; $N(w \leq v)$

3) Calcul de $G(v)$ à l'aide de son expression analytique (voir 1.b) ; calcul de $\hat{G}(v)$ à l'aide de la formule suivante :

$$\hat{G}(v) = \frac{N(w \leq v) - N(b \leq v)}{n - N(b \leq v) - N(c \leq v)}$$

On peut aussi prendre $\bar{G}(v) = (1/99) \sum \hat{G}(v)$ au lieu de $G(v)$.

4) Calcul de $dw = |G(w) - \hat{G}(w)|$ et recherche du maximum.

BIBLIOGRAPHIE

DIGGLE P.J., 1979, On parameters estimation and Goodness-of-fit testing for spatial point patterns, *Biometrics* 35, 87-101.

DIGGLE P.J., 1983, Statistical analysis of spatial point patterns, *Academic press*, pp 10-23.

RIPLEY B.D., 1981, Spatial statistics, pp 152-158.

UPTON G.J.G. et B. FINGLETON, 1985, Spatial data analysis by example, tome I, *Wiley* pp 79-87.

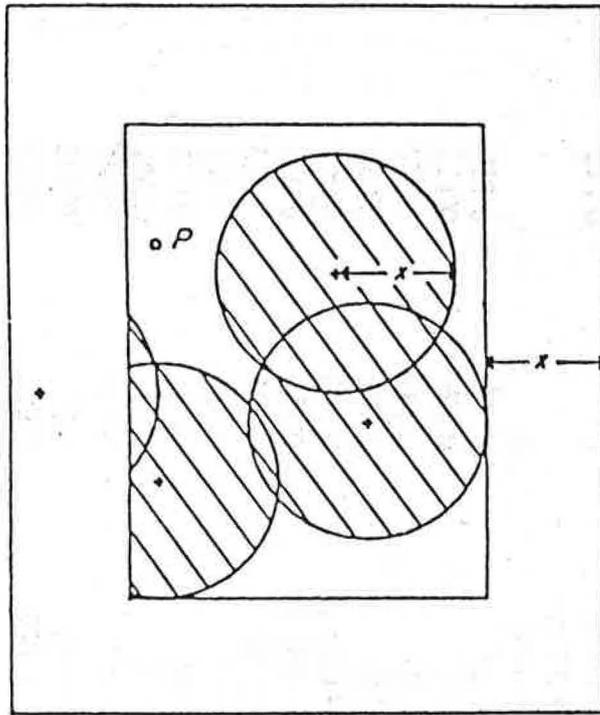


Figure 1: Interprétation de $F(x)$ - Effet de bord.
 A une distance x de la bordure, l'aire hachurée fournit une estimation non biaisée de $F(x)$. Cette estimation suppose que la portion des individus appartenant à la bordure soit connue.
 Si la distribution des individus est régulière, alors $\mu \pi x^2$ est égal à $F(x)$ jusqu'à la valeur-seuil à laquelle les cercles se recouvrent.

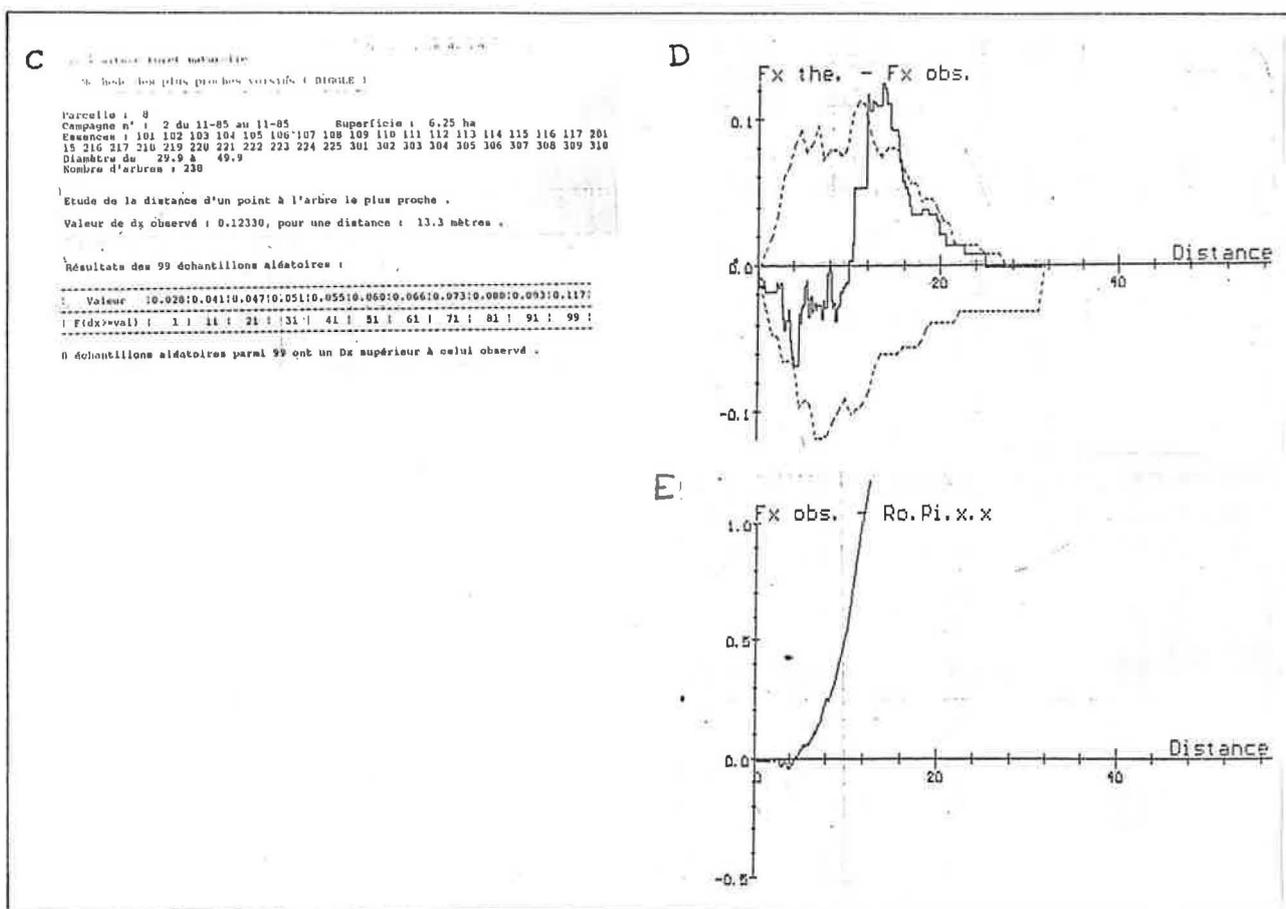
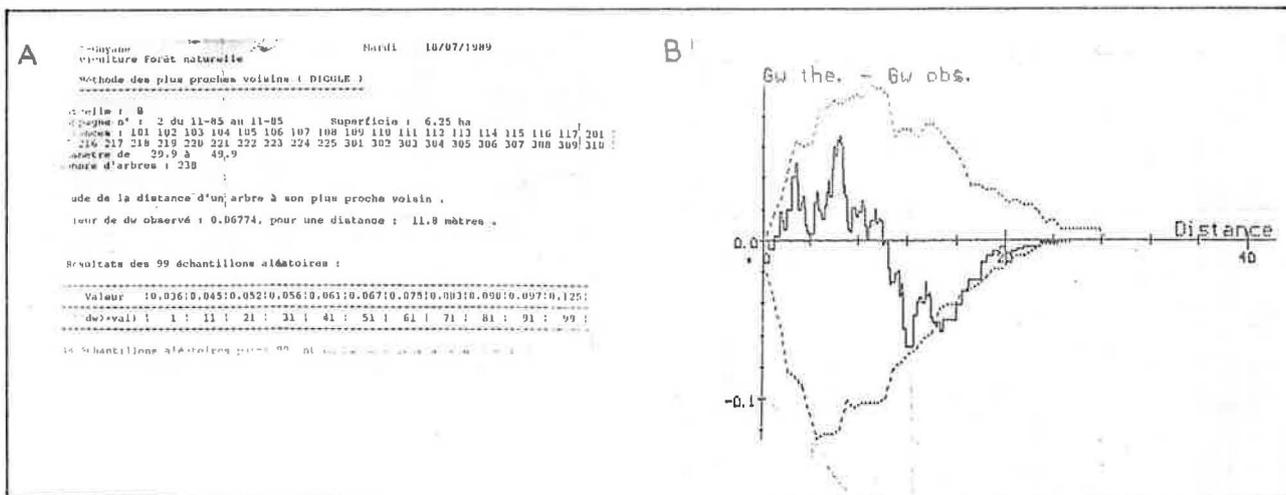


Figure 2 : Sorties du programme "DIGGLE"

- A Tableau des d_w
- B Graphique de d_w avec l'enveloppe des 99 simulations
- C Tableau des d_x
- D Graphique des d_x .
- E Graphique $F(x) - \mu \pi x^2$ en fonction de x

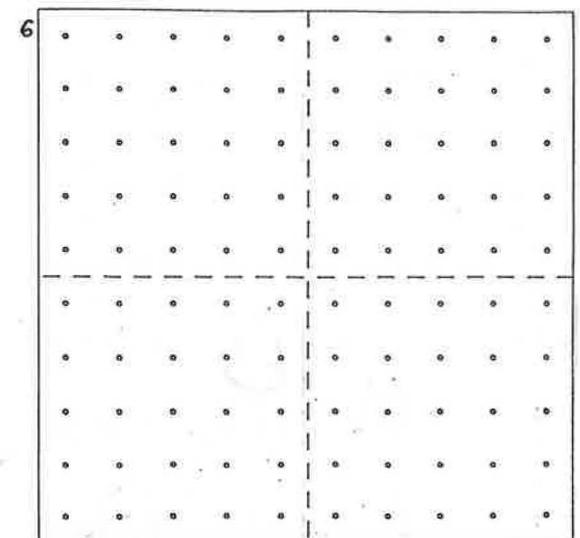
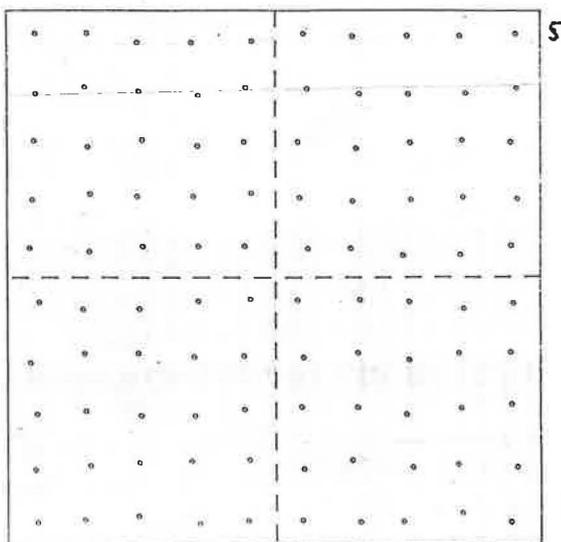
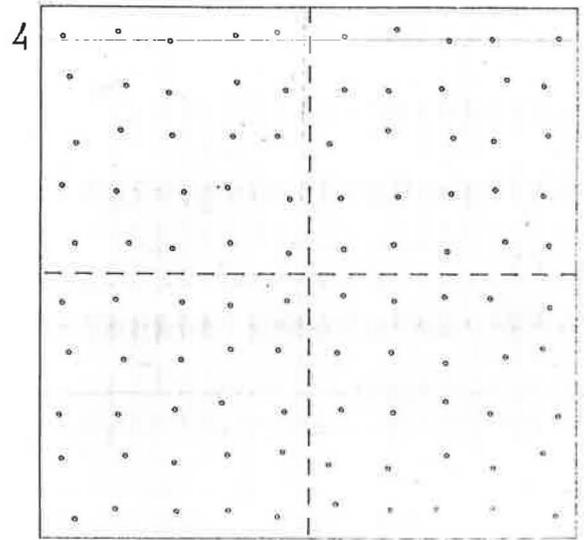
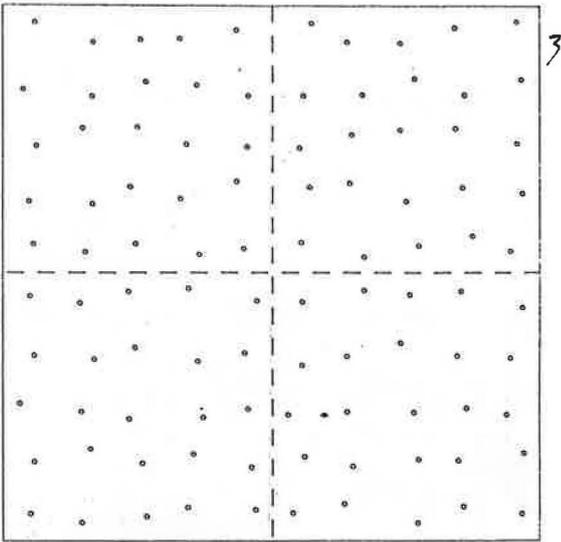
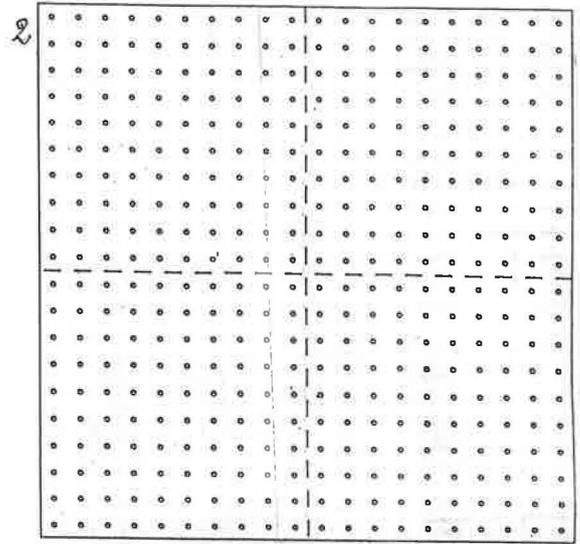
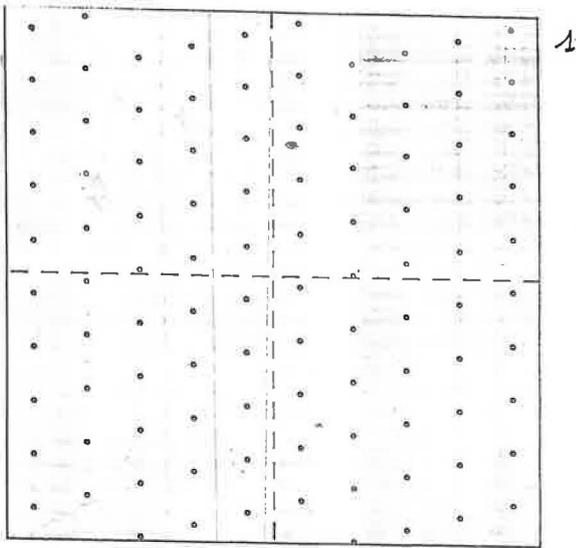


Figure 3: Distributions spatiales - type études

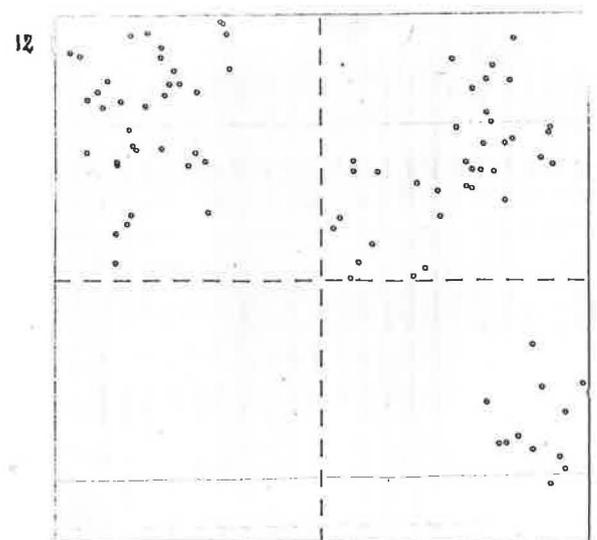
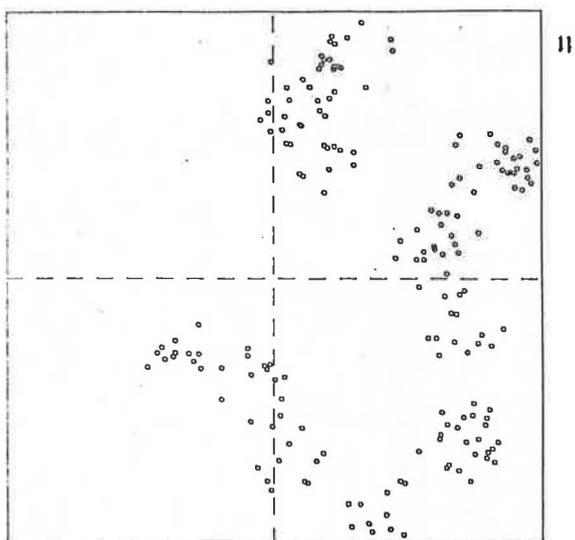
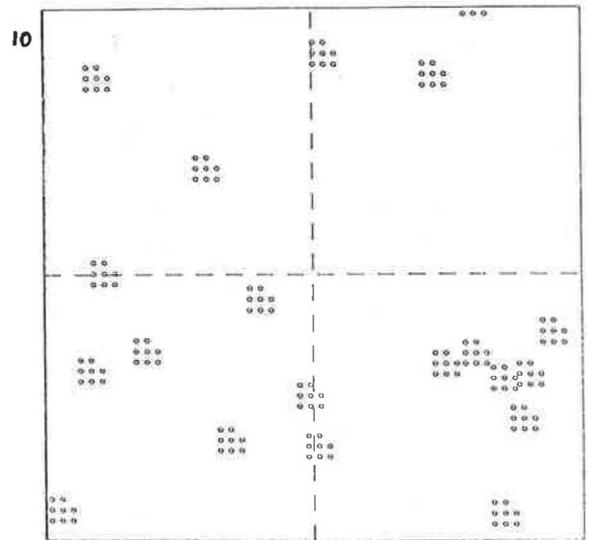
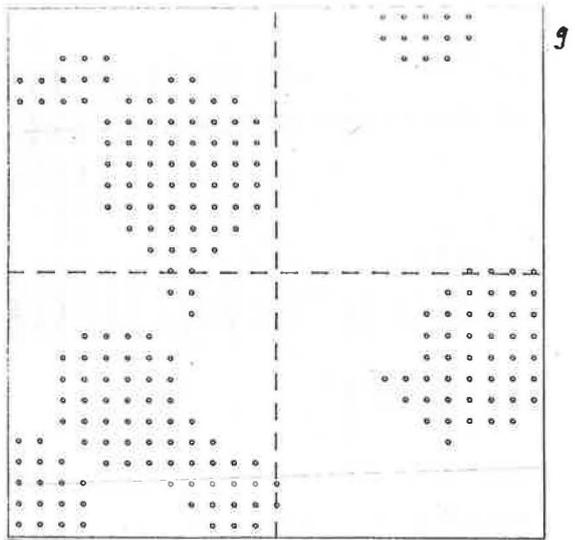
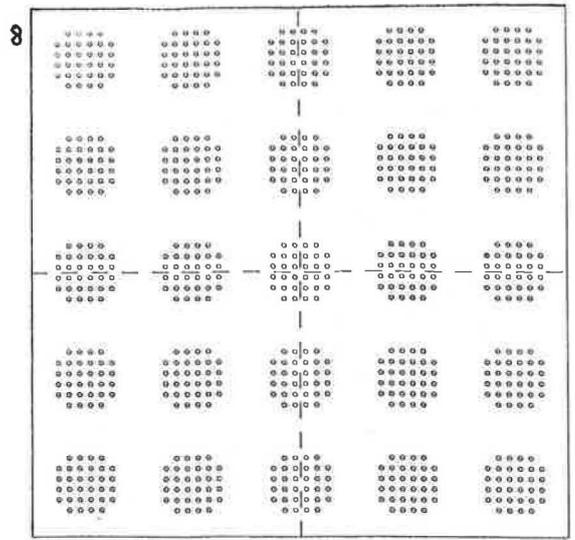
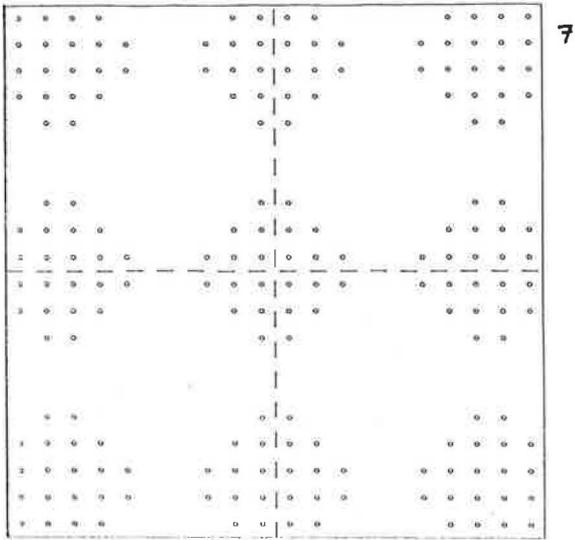


Figure 3 (suite)

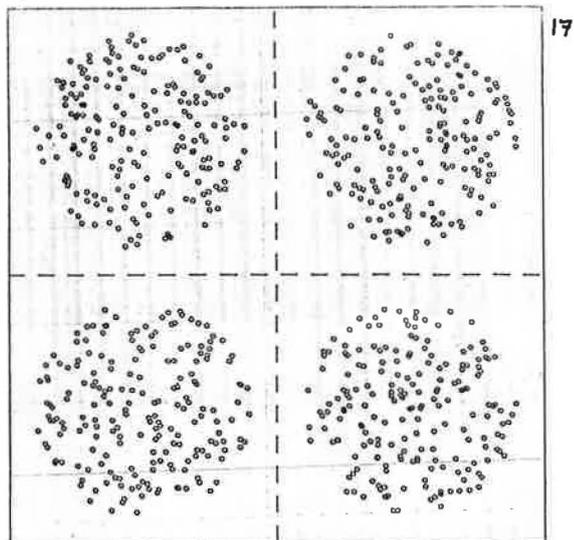
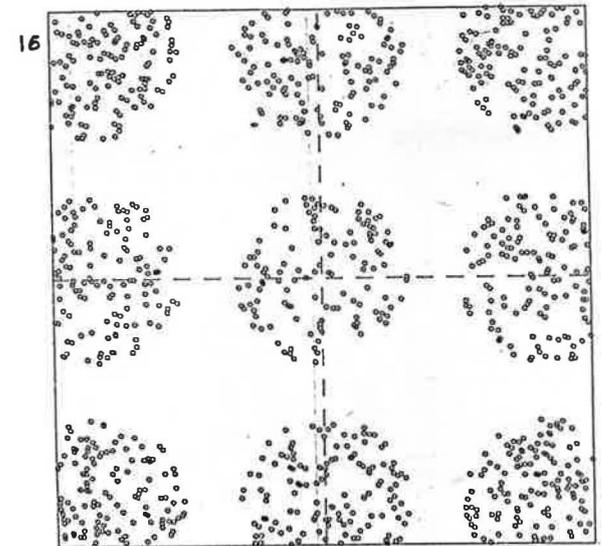
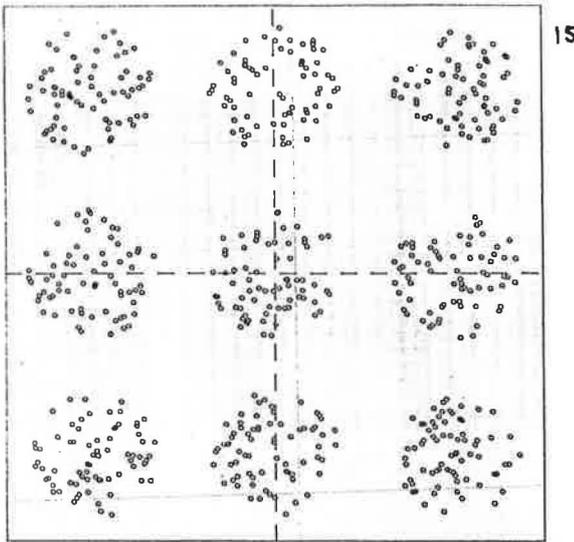
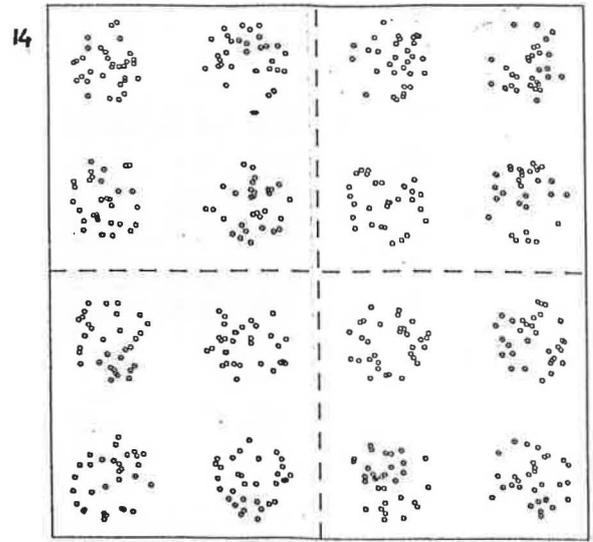
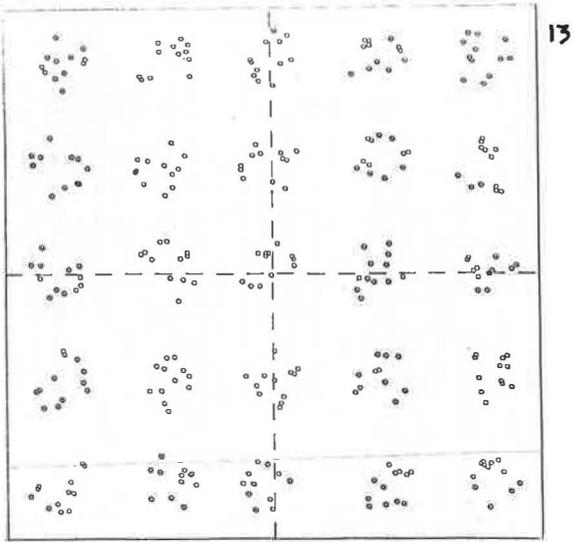


Figure 3 (suite)

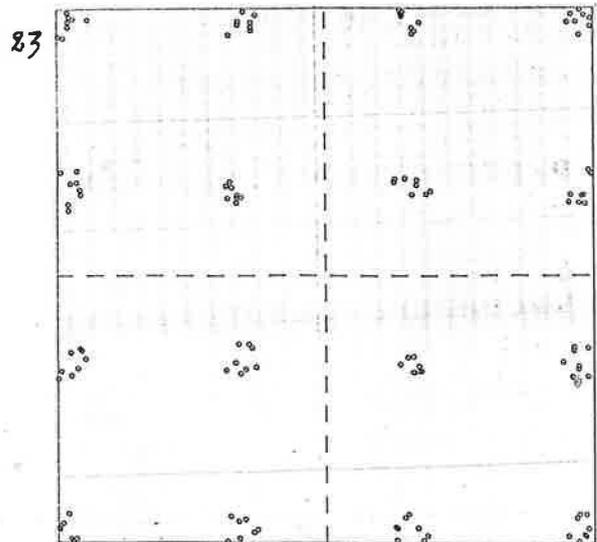
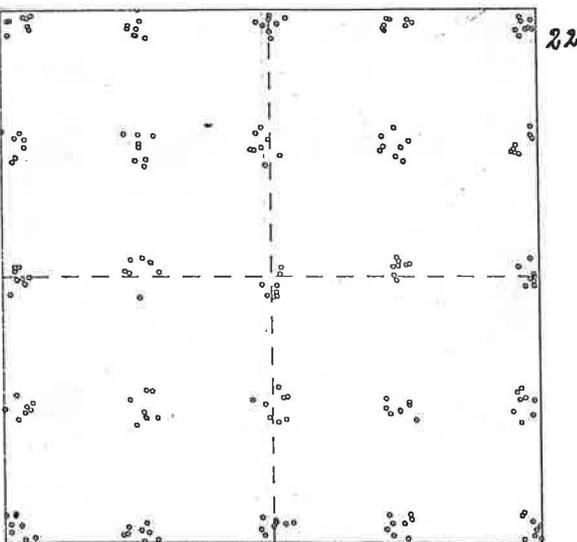
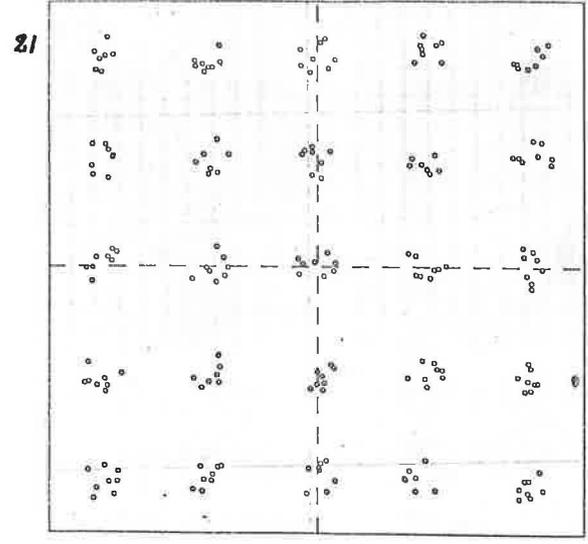
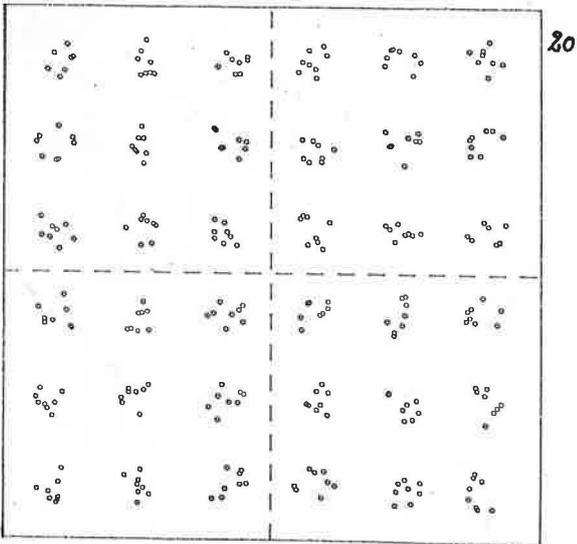
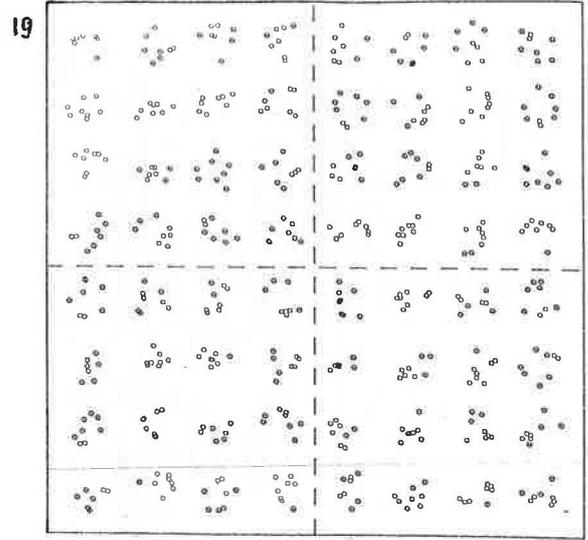
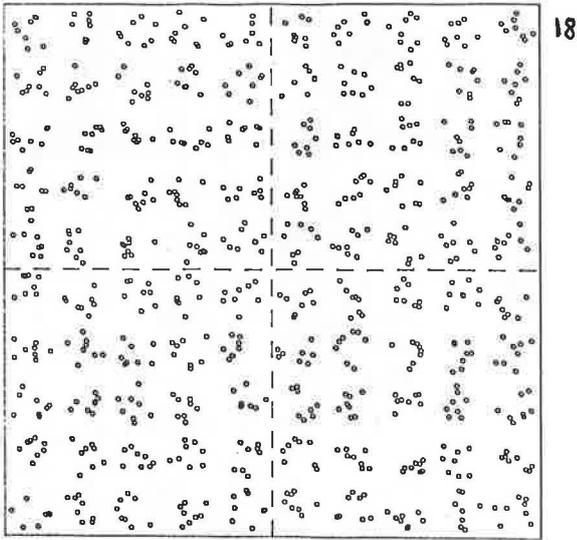


Figure 3 (suite)

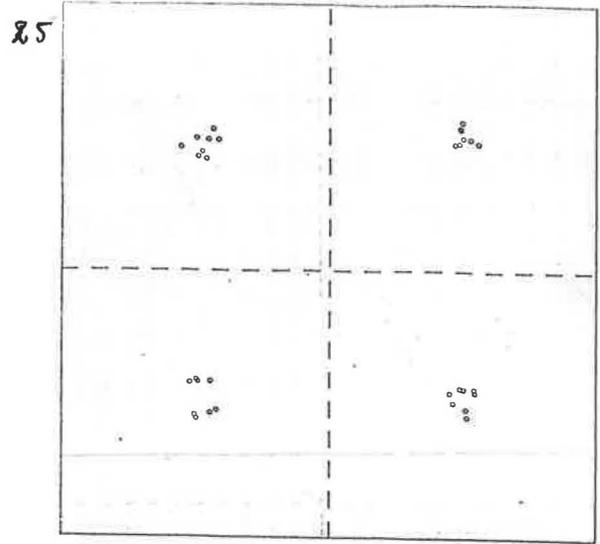
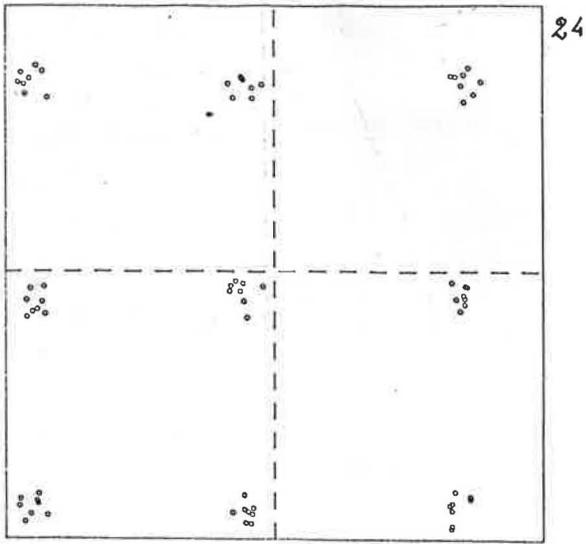
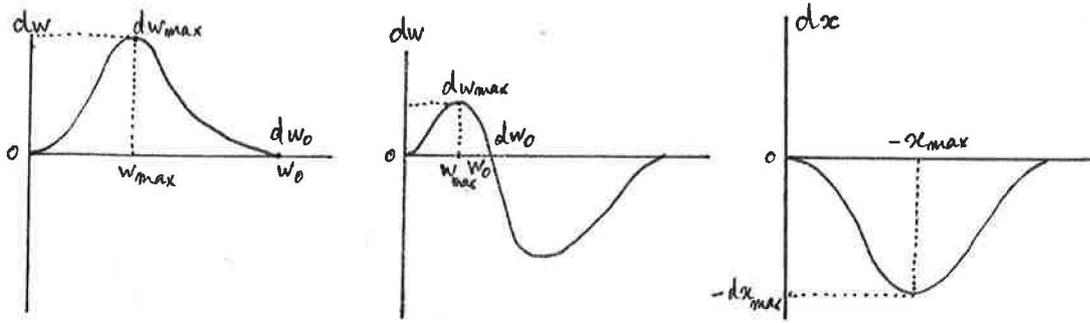


Figure 3 : (fin)

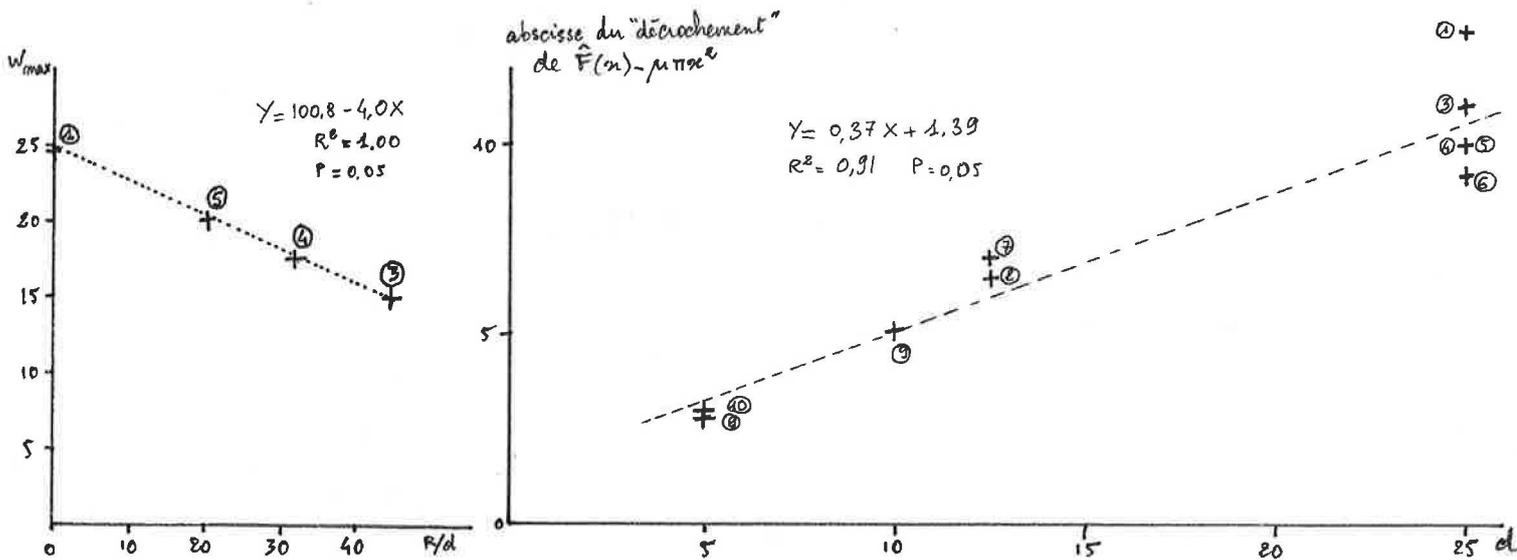
DISTRIBUTIONS REGULIERES



a) Conventions de notation utilisées

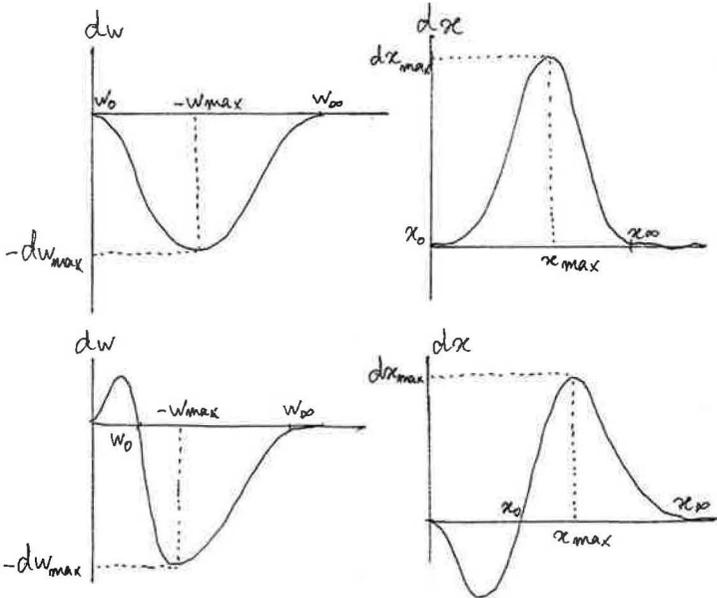
Distribution N° (cf Fig.)	1	2	3	4	5	6	7	8	9	10
distance moyenne entre arbres d	25m	12,5	25	25	25	25	12,5	5	10	5
écart à la régularité R/d (cf texte)	0	0	44%	32%	20%	0	0	0	0	0
w_{max}	24	non testé	15,1	17,5	20,2	non testé	12,5	0 (agrégats)	9,6	5
w_0	25	non testé	25	25	25	non testé	12,5	5	10	5
$-x_{max}$	13,8	6	15,6	16	14	14,8	6	2,5	(agrégats)	(agrégats)
point de "décroche- ment" de la courbe $\hat{F}(x) - \mu \pi x^2$	13	6,5	11	10	10	9,2	7	2,7	5,1	3

b). Tableau 1: Résultats de la méthode de Diggle pour les distributions régulières.



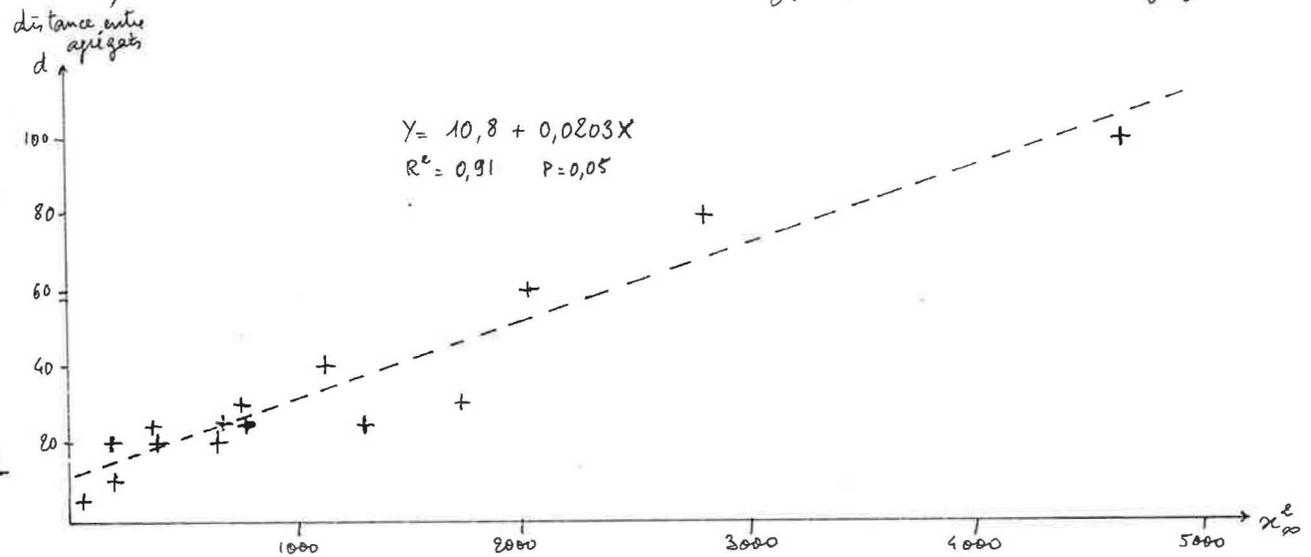
c) Figure 4: à gauche: Relation entre w_{max} et écart à la répartition parfaitement régulière
à droite: Relation entre point de décrochement de la courbe $\hat{F}(x) - \mu \pi x^2$ et distance inter-arbres.

Distribution N° (cf Fig.)	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
distance entre agrégats (m)	30	20	aléatoire	aléatoire	aléatoire	aléatoire	20	25	25	25	25	5	10	20	30	40	60	80	100
rayon des agrégats (m)	35	15	aléatoire (24,3)	7,5	15	35	15	20	30	40	50	10	10	10	10	10	10	10	10
densité (pieds/ha) intra agrégats	64	400	100	400	170	30	170	250	250	250	250	250	250	250	250	250	250	250	250
$-W_{max}$	12,5	5	9,6	5	6,5	8,3	5,5	4,9	3	2,9	4,1	3,9	5,4	5,1	5,5	5	5,6	7,5	5,8
w_{00}	34	17,5	24,5	30	25	38	24	16	15	11	12,7	13	16	22	24	55,5	76	38	62,5
x_{max}	14,8	6,1	14,8	18,6	16,7	25,6	11,6	8,6	8,8	10,5	9,6	6,0	5,5	7,5	12,6	18,1	19	31,2	24,7
x_{00}	41,5	18,5	49	61	57	73	25	19	25,5	35,5	27,5	8,6	13,6	13,2	28	33,5	45	53	68



a) Conventions de notation utilisées

b) Tableau 2: Résultats de la méthode de Diggle pour les distributions agrégatives.



c) Figure 5: Régression distance inter agrégats / x_{00}^e .

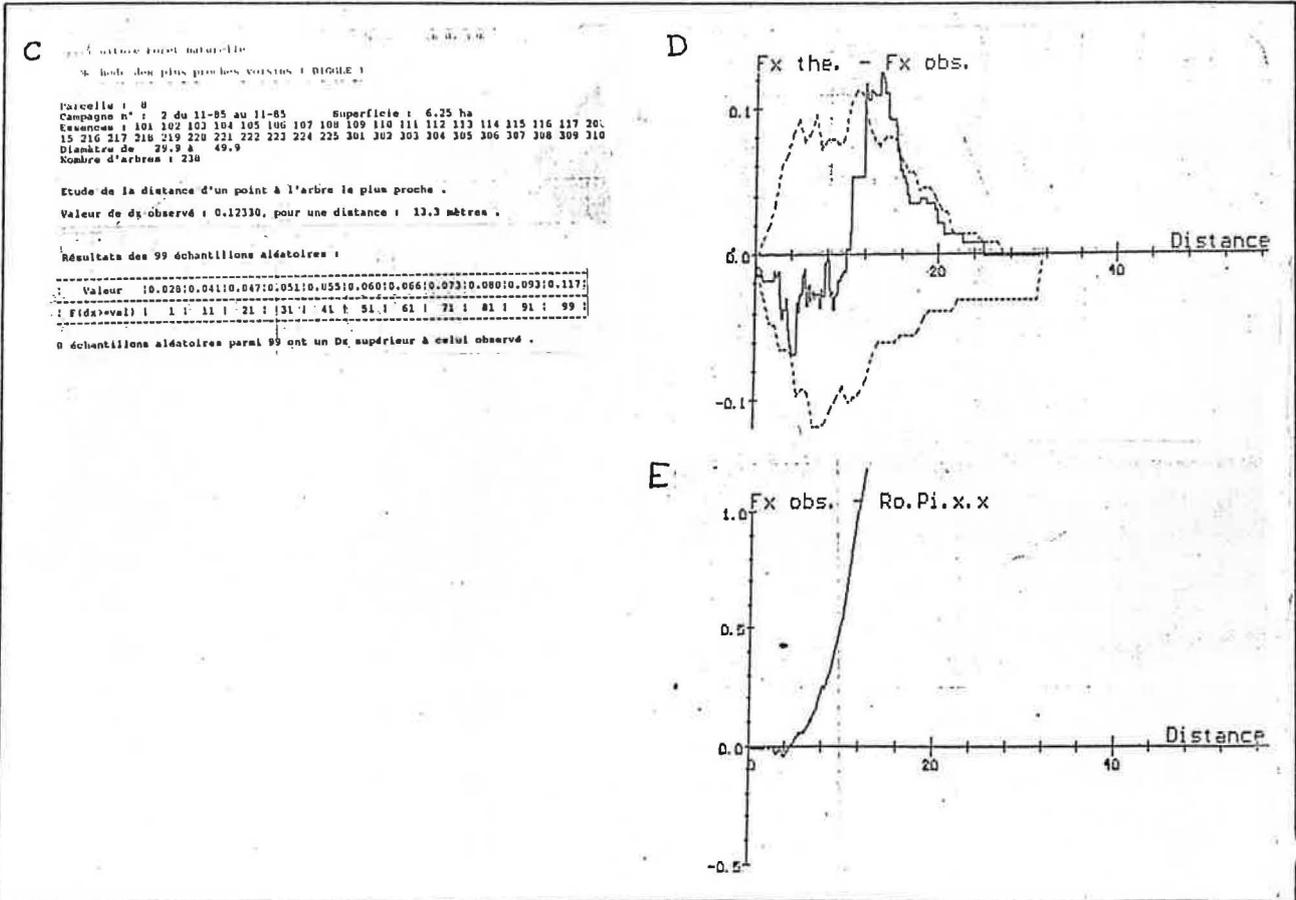
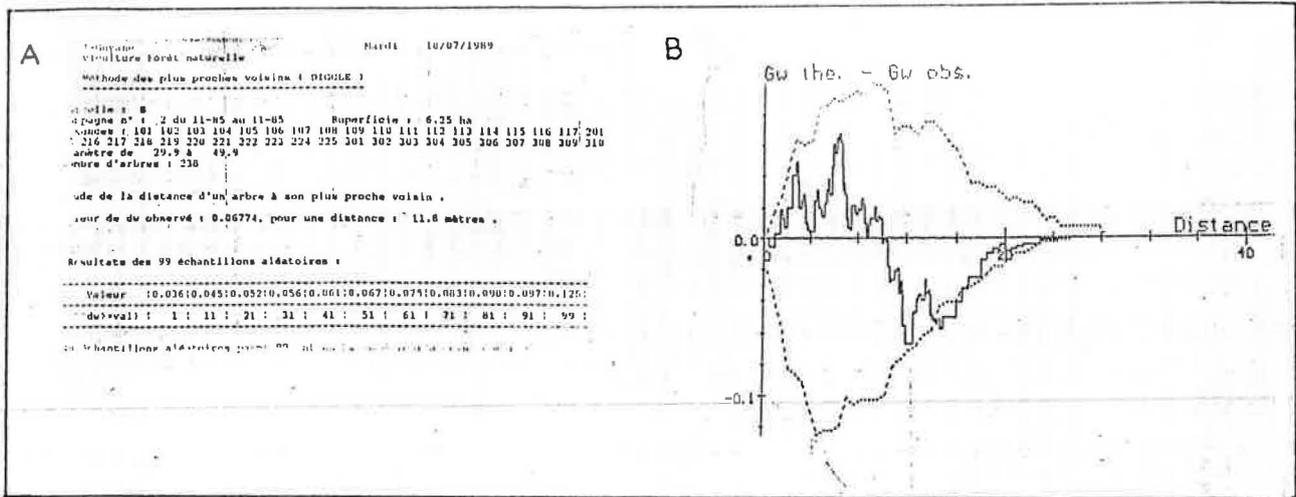


Figure 2 : Sorties du programme "DIGGLE"

- A Tableau des d_w
- B Graphique de d_w avec l'enveloppe des 99 simulations
- C Tableau des d_x
- D Graphique des d_x .
- E Graphique $F(x) - \pi x^2$ en fonction de x