

Analyse de la structure spatiale de semis de points hétérogènes : exemples d'application à des peuplements forestiers

Résumé

Les outils classiques d'analyse de la structure spatiale, comme la fonction de Ripley, sont définis uniquement pour des semis de points homogènes. La réalité écologique est souvent plus complexe, et présente différentes sources d'hétérogénéité qui peuvent fausser les résultats de l'analyse de la structure spatiale.

Cette documentation thématique présente une méthode pour analyser simplement la structure spatiale d'un semis hétérogène à l'aide des modules Ripley et Density d'ADS. L'essentiel reprend des travaux développés dans Goreaud & Pélissier (1999), Goreaud (2000) et Pélissier & Goreaud (2001).

Plan

1 — Introduction.....	2
2 — Théorie : l'analyse des semis de points hétérogènes	2
2.1 — Le cadre mathématique d'étude de la structure spatiale.....	2
2.2 — Analyse de la structure spatiale par la méthode de Ripley.....	5
2.3 — Prise en compte de l'hétérogénéité.....	13
3 — Applications à des données d'écologie forestière.....	18
3.1 — Exemple 1 : Dispositif de Paracou (Guyane française).....	19
3.2 — Exemple 2 : Dispositif du Beau Poirier (Nancy, France)	28
3.3 — Exemple 3 : Dispositif d'Uppangala (Inde).....	38
4 — Références.....	48

F. Goreaud et R. Pélissier

1 — Introduction.

La structure spatiale d'un écosystème, c'est-à-dire la manière dont sont organisés dans l'espace les individus qui le composent, joue souvent un rôle essentiel dans son fonctionnement. Pour un peuplement forestier, par exemple, la structure spatiale détermine l'environnement local autour de chaque arbre (en particulier le nombre de voisins) et donc influence les processus naturels comme la croissance et la mortalité (Delvaux 1981, Barot *et al.* 1999), voire détermine le choix d'un scénario d'éclaircies pour les peuplements faisant l'objet d'une gestion sylvicole. La structure spatiale influence également le développement des graines et donc la régénération du peuplement. Inversement, les processus naturels ou les actions anthropiques modifient à leur tour la structure spatiale du peuplement qui se trouve donc dans un cycle de rétroaction schématisé par la Figure 1.

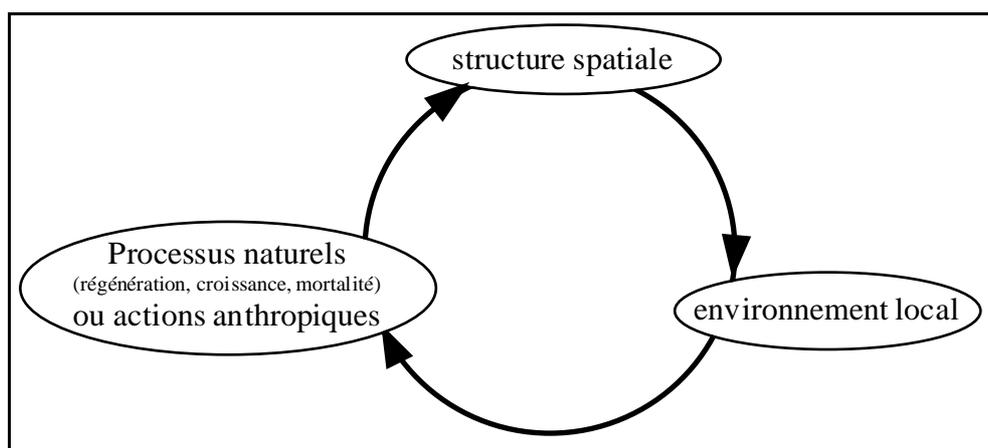


Figure 1 : La structure spatiale dans la dynamique d'un peuplement forestier.

L'objectif du présent document est de montrer, à partir de quelques exemples pris dans le domaine de l'écologie forestière, comment les outils proposés dans ADS permettent d'étudier la structure spatiale d'un peuplement hétérogène.

La première partie présente brièvement, sur un plan théorique, les outils d'analyse de la structure spatiale utilisés : le formalisme des processus ponctuels, la fonction de Ripley et les fonctions associées, ainsi qu'une méthode simple d'étude des peuplements hétérogènes. La seconde partie détaille l'utilisation de ces outils dans le contexte spécifique de l'étude de la structure spatiale de peuplements forestiers hétérogènes, en illustrant l'utilisation des modules univariés d'ADS (Ripley et Density), ainsi que celle des différents utilitaires d'ADS et d'ADE-4 qui permettent de préparer les analyses et de représenter les résultats.

2 — Théorie : l'analyse des semis de points hétérogènes.

2.1 – Le cadre mathématique d'étude de la structure spatiale.

La structure spatiale est un concept complexe, qui se rapporte à la fois à l'utilisation verticale et horizontale de l'espace par les éléments d'un écosystème (Begon *et al.* 1996). Nous nous intéressons ici uniquement à la position des individus dans le plan horizontal, que l'on peut représenter, à un instant t d'observation, par un semis de points (Cressie 1993). Pour un peuplement forestier, chaque point peut par exemple correspondre au centre de la base d'un arbre. Le formalisme des processus ponctuels permet d'étudier la structure spatiale d'un tel semis de points.

2.1.1 – Le formalisme des processus ponctuels.

Un processus ponctuel P est un processus aléatoire, un objet mathématique similaire à une variable aléatoire, dont les réalisations sont des semis de points. Les propriétés du processus définissent des contraintes sur ces réalisations (densité, voisinage, structure...). En d'autres termes, un seul processus ponctuel P peut générer une infinité de semis de points, tous différents, mais qui partagent certaines propriétés communes, et en particulier la structure. Inversement, pour un semis de points donné, il existe une infinité de processus ponctuels dont ce semis pourrait être une réalisation dans le domaine d'étude.

Lorsqu'on étudie la structure d'un semis de points donné, représentant par exemple un peuplement réel, on fait implicitement l'hypothèse qu'il existe un processus sous-jacent P dont notre semis a une forte probabilité d'être une réalisation. On utilise alors les propriétés de ce processus pour décrire la structure du semis. Notons que, contrairement aux applications statistiques classiques, on ne dispose alors que d'une seule réalisation (une seule mesure) du processus, le semis réel. Cette restriction importante impose des hypothèses implicites supplémentaires d'homogénéité, d'isotropie et d'ergodicité pour rendre possible l'analyse de la structure spatiale. Les phénomènes d'hétérogénéité ou d'anisotropie sont donc particulièrement difficiles à étudier.

Un processus est dit homogène (les statisticiens parleraient plutôt de processus stationnaire) s'il est invariant par translation, c'est-à-dire si ses propriétés ne varient pas d'une position à l'autre de l'espace.

Un processus est dit isotrope s'il est invariant par rotation, c'est-à-dire si ses propriétés ne varient pas avec l'orientation de l'espace.

Les hypothèses d'homogénéité et d'isotropie assurent que le semis présente les mêmes propriétés partout dans le domaine d'étude (Figure 2), conditions nécessaires pour pouvoir définir des variables synthétiques simples, représentatives de l'ensemble du semis.

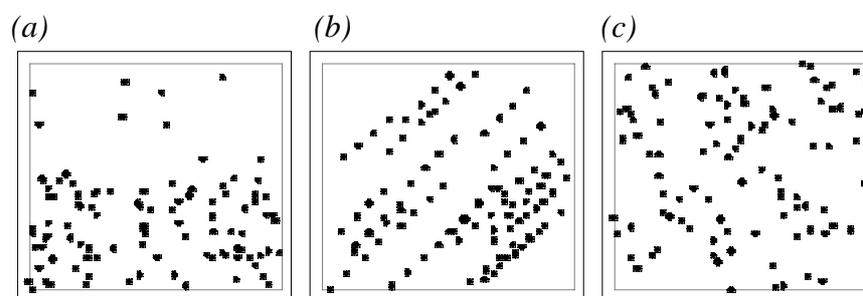


Figure 2 : Exemples (a) d'un semis hétérogène (deux zones de densités différentes), (b) d'un semis anisotrope (une direction privilégiée) et (c) d'un semis homogène isotrope (semis de Poisson).

Pour un semis dans un domaine borné (par exemple une parcelle d'échantillonnage), l'hypothèse d'homogénéité est délicate à définir, puisque par nature les propriétés au bord du domaine, notamment en termes de voisinage, sont différentes de celles du centre. Pour pouvoir utiliser malgré tout l'hypothèse d'homogénéité, nous considérerons par la suite que les semis de points étudiés proviennent de processus non bornés dans l'espace, qui vérifient l'hypothèse d'homogénéité, et sur lesquels seules nos mesures sont bornées, c'est-à-dire restreintes au domaine d'étude. Il sera par conséquent nécessaire de corriger les biais dus aux effets de bord dans les calculs effectués sur des semis réels.

2.1.2 – Les propriétés de premier et second ordres.

La propriété la plus simple d'un processus ponctuel, dite propriété du premier ordre, est sa densité λ . Pour un processus homogène, λ est constante et peut être estimée par le rapport N/D du nombre de points N sur la surface D du domaine d'étude. La probabilité de présence d'un point du semis dans une surface élémentaire dS centrée en (x,y) est alors donnée par :

$$\Pr(\text{un point dans } dS) = \lambda dS$$

La propriété de second ordre d'un processus ponctuel décrit la probabilité de présence conjointe de deux points dans des surfaces élémentaires dS_1 et dS_2 centrées en (x_1,y_1) et (x_2,y_2) . Elle peut être caractérisée par la fonction de densité des paires de points $g((x_1,y_1),(x_2,y_2))$ (*pair correlation function* ; Cressie 1993) qui exprime, d'une certaine façon, les relations de voisinage entre deux points d'un semis. Pour cette raison, la propriété de second ordre d'un processus est usuellement utilisée comme outil de description de la structure. Pour un semis homogène, la probabilité de présence conjointe de deux points dans dS_1 et dS_2 est donnée par :

$$\Pr(\text{un point dans } dS_1 \text{ et un point dans } dS_2) = \lambda^2 g((x_1,y_1),(x_2,y_2)) dS_1 dS_2$$

Dans le cas d'un processus homogène isotrope, on peut montrer que la fonction g ne dépend que de la distance r entre (x_1,y_1) et (x_2,y_2) (voir par exemple Ripley 1977). La fonction $g(r)$ est alors proportionnelle au nombre de couples de points à distance r l'un de l'autre.

2.1.3 – L'hypothèse nulle d'un processus de Poisson.

Un processus de Poisson d'intensité λ est un processus ponctuel homogène isotrope pour lequel la disposition des points est complètement aléatoire à chaque réalisation. La probabilité de présence d'un point à une position donnée est en particulier indépendante de la position des autres points du processus. Ce qui permet d'écrire :

$$\Pr(\text{un point dans } dS_1 \text{ et un point dans } dS_2) = \Pr(\text{un point dans } dS_1) \Pr(\text{un point dans } dS_2) = \lambda^2 dS_1 dS_2$$

et de conclure d'après cette équation que, pour un tel processus, la propriété de second ordre $g((x_1,y_1),(x_2,y_2))$ est constante, égale à 1.

Le processus de Poisson sert de référence : il correspond à une structure complètement aléatoire (*complete spatial randomness* ou CSR) et est usuellement utilisé comme hypothèse nulle pour les tests statistiques concernant la structure spatiale. Une telle structure aléatoire présente toutefois, à toutes les échelles d'observation une forte variabilité intrinsèque du comportement local, avec des zones plus ou moins denses (Figure 3). On peut d'ailleurs montrer que $N(S)$, le nombre de points contenus dans une zone fixée de surface S , suit une loi de Poisson de paramètre λS :

$$\Pr(N(S) = n) = e^{-\lambda S} \frac{(\lambda S)^n}{n!}$$

En particulier pour un processus de Poisson sur un domaine fermé de surface D , le nombre de points des différentes réalisations n'est pas constant, mais suit, lui aussi, une loi de Poisson de paramètre λD .

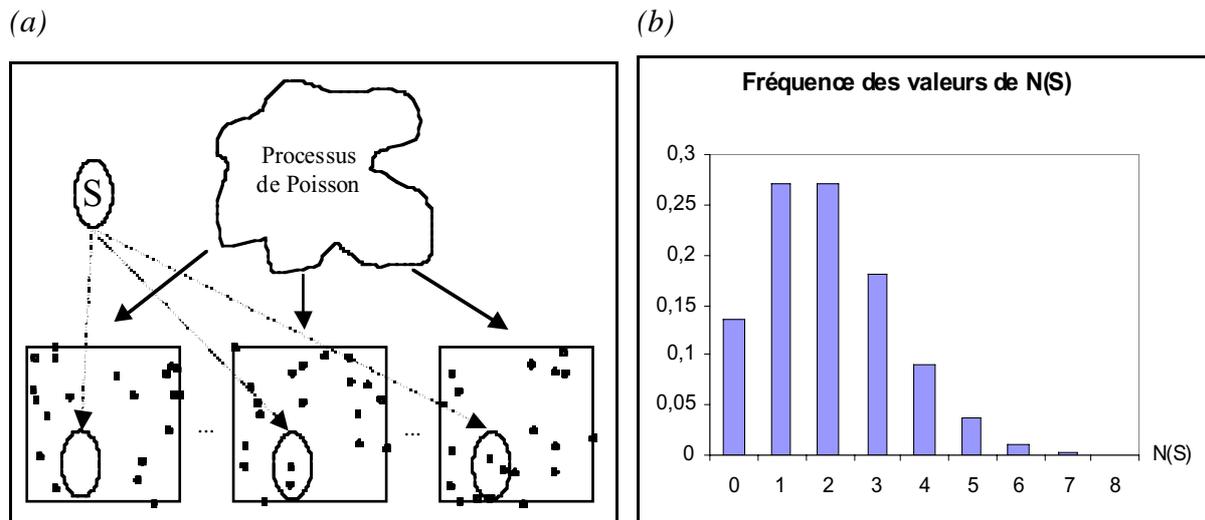


Figure 3 : Le nombre de points contenus dans une surface S quelconque pour différentes réalisations d'un processus de Poisson (a) suit une loi de Poisson de paramètre λS (b).

2.1.4 – L'hypothèse d'ergodicité.

Un processus ponctuel est dit ergodique si la moyenne au sens probabiliste, c'est-à-dire portant sur les différentes réalisations, peut être remplacée par la moyenne spatiale sur le domaine d'étude (Cressie 1993, p.53).

A partir d'une unique réalisation (un semis réel par exemple), il est impossible de calculer des propriétés moyennes au sens probabiliste du terme. L'hypothèse d'ergodicité permet de remplacer ces moyennes statistiques par des moyennes portant sur des régions différentes du domaine d'étude. Elle est donc indispensable à toute étude du processus ponctuel, et en particulier à l'analyse de la structure spatiale. Pour un semis de points donné, l'hypothèse d'ergodicité ne peut être testée à cause de l'unicité de la réalisation. Toutefois, on peut montrer qu'elle est vérifiée pour certaines classes de processus homogènes (Cressie 1993, p.57).

2.2 — Analyse de la structure spatiale par la méthode de Ripley.

De nombreuses méthodes d'analyse de la structure spatiale des semis de points peuvent être trouvées dans la littérature (e.g. Ripley 1981, Diggle 1983, Cressie 1993). On les classe généralement en fonction des données nécessaires à leur mise en œuvre. On distingue en particulier les méthodes *basées sur les quadrats*, pour lesquelles les données sont des nombres d'individus dans des quadrats de positions et de tailles variables (voir par exemple Chessel 1978) ; et les méthodes *basées sur les distances*, pour lesquelles les données sont des distances entre points ou individus, ou encore des positions. Parmi ces dernières, certaines méthodes ne nécessitent que la connaissance des plus proches voisins de chaque point, comme la méthode de Clark et Evans (1954). D'autres nécessitent une carte exhaustive du domaine d'étude. C'est le cas de la méthode d'analyse de voisinage de second ordre de Ripley, qui présente l'avantage de décrire la structure spatiale d'un semis de points à plusieurs échelles simultanément.

2.2.1 – La fonction $K(r)$ et les fonctions associées.

Pour un processus ponctuel homogène et isotrope d'intensité λ , Ripley (1976, 1977) a montré qu'on peut caractériser la propriété de second ordre par une fonction $K(r)$ telle que l'espérance du nombre de voisins à distance r d'un point quelconque du semis soit :

$$N(r) = \lambda K(r)$$

On peut déduire de $N(r)$ une fonction de densité locale de voisinage de second ordre $n(r)$:

$$n(r) = N(r) / \pi r^2$$

où πr^2 correspond à la surface du disque de rayon r .

La fonction de Ripley est en fait une normalisation de $N(r)$ qui permet de comparer les structures spatiales de semis de densités variables :

$$K(r) = N(r) / \lambda$$

Elle est directement reliée à la fonction de densité de paires de points $g(r)$ définie au paragraphe précédent :

$$K(r) = \int_{s=0}^r g(s) 2\pi s ds$$

Plus précisément, on peut écrire :

$$g(r) = P(r) / \lambda ds$$

où $P(r)$ est la probabilité qu'il y ait un point dans une surface ds se trouvant à une distance r d'un point quelconque du semis.

* Pour un processus de Poisson d'intensité λ , $N(r)$, l'espérance du nombre de voisins à distance r d'un point quelconque du semis est $\lambda \pi r^2$, donc $n(r) = \lambda$ et $K(r) = \pi r^2$. $P(r)$ vaut alors λds , d'où $g(r) = 1$.

* Pour un processus agrégé, les points ont en moyenne plus de voisins que pour l'hypothèse nulle, et donc $K(r) > \pi r^2$. Inversement, pour un processus régulier les points ont en moyenne moins de voisins que pour l'hypothèse nulle, et $K(r) < \pi r^2$ (Figure 4).

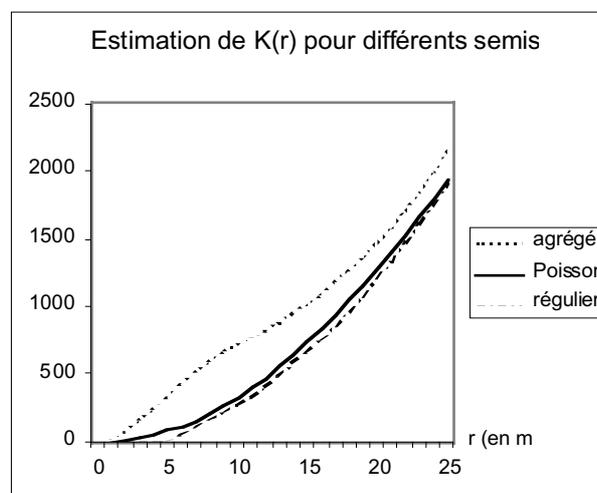


Figure 4 : Estimations de la fonction $K(r)$ pour un semis agrégé, un semis de Poisson, et un semis régulier.

La fonction $K(r)$ n'est cependant pas toujours très facile à interpréter, puisque la courbe obtenue pour l'hypothèse nulle d'un semis de Poisson est une parabole. C'est pourquoi Besag (1977) propose une fonction modifiée¹ $L(r)$:

$$L(r) = \sqrt{K(r)/\pi} - r$$

Cette fonction est plus facile à interpréter que $K(r)$, car pour un processus de Poisson $L(r) = 0$ à toutes distances r , les processus agrégés et réguliers se situant respectivement au-dessus et au-dessous de l'axe des abscisses (Figure 5).

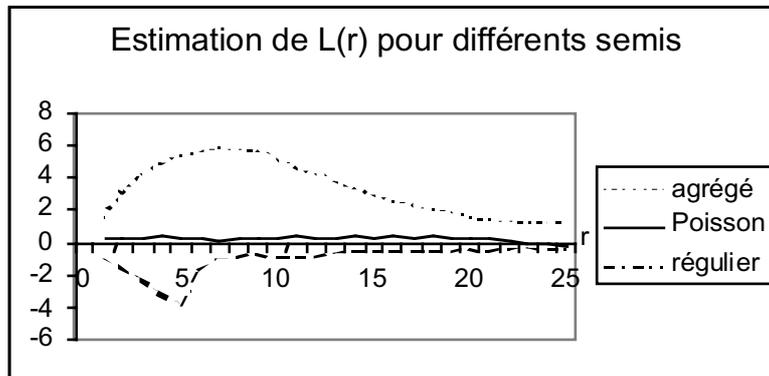


Figure 5 : Un exemple d'estimation de la fonction $L(r)$ pour un semis agrégé, de Poisson et régulier.

2.2.2 – Les estimateurs $\hat{n}(r)$, $\hat{K}(r)$, $\hat{g}(r)$ et $\hat{L}(r)$.

Dans la plupart des cas, on ne connaît pas le processus à l'origine du semis étudié et les différentes fonctions doivent par conséquent être estimées à partir d'une réalisation unique. On approche alors l'espérance du nombre de voisins autour d'un point quelconque, par sa moyenne sur l'ensemble du semis (ce qui n'est possible que sous l'hypothèse d'ergodicité) :

$$\hat{N}(r) = \frac{1}{N} \sum_{i=1}^N \sum_{i \neq j} k_{ij}$$

où N est le nombre de points dans le domaine d'étude de surface D et k_{ij} prend la valeur 1 si la distance entre les points i et j est inférieure à r , et 0 dans le cas contraire.

On peut alors définir les estimateurs :

$$\hat{n}(r) = \hat{N}(r) / \pi r^2$$

$$\hat{K}(r) = \hat{N}(r) / \hat{\lambda}, \text{ où } \hat{\lambda} = N / D$$

et

$$\hat{L}(r) = \sqrt{\hat{K}(r) / \pi} - r$$

Ces estimateurs naïfs sont biaisés, car les points situés près de la bordure ont dans le domaine d'étude moins de voisins qu'un point quelconque du processus : si on considère un

¹ Besag propose en fait une fonction $L(r) = \sqrt{K(r)/\pi}$ et étudie $L(r)-r$ en fonction de r .

processus non borné, certains voisins potentiels sont en dehors du domaine d'étude (Figure 6). C'est un problème très général, connu sous le nom d'effet de bord, qui se rencontre dans beaucoup de méthodes d'analyse de la structure spatiale.

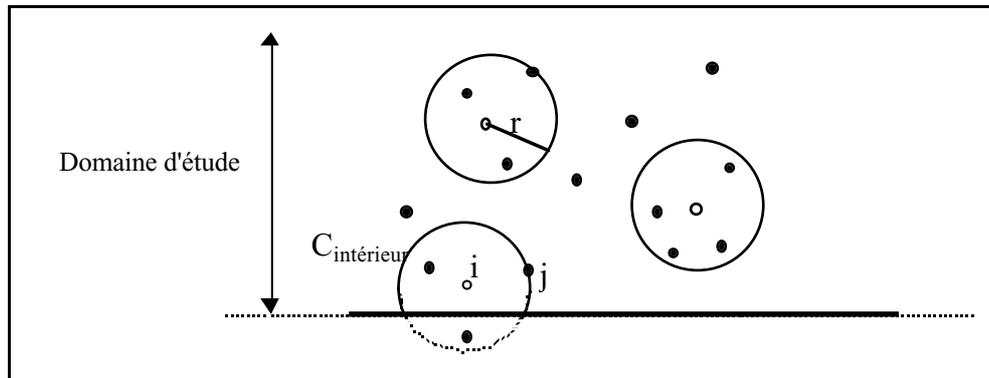


Figure 6 : Illustration du problème des effets de bord.

Il existe de nombreuses méthodes pour prendre en compte les problèmes d'effet de bord. La plus simple consiste à ignorer dans les calculs, les points placés à proximité de la bordure, ce qui revient à définir une zone tampon autour du domaine étudié, mais réduit sérieusement la quantité de données disponibles. Gignoux *et al.* (1999) ont mesuré la perte de puissance que cette méthode pouvait entraîner sur des tests effectués sur certaines fonctions d'analyse de la structure spatiale.

Plusieurs auteurs ont fait le bilan des méthodes de correction des effets de bord (voir par exemple Ripley 1982 ou Biber 1999). En ce qui concerne $\hat{K}(r)$ deux méthodes semblent avoir été principalement utilisées :

- (1) une méthode de correction globale, dite de Ohser & Stoyan, qui repose sur les propriétés géométriques du domaine d'étude (Stoyan *et al.* 1995). Cette méthode a l'avantage d'être assez simple et donc rapide à calculer ;
- (2) une méthode de correction locale, proposée par Ripley (1977), qui porte sur la contribution de chaque point situé près de la bordure du domaine d'étude. Cette correction locale présente l'avantage d'être également utilisable pour calculer des indices individuels relatifs à chacun des points du semis, comme celui proposé par Getis & Franklin (1987). Certains travaux semblent montrer d'autre part, que cette classe d'estimateurs est plus robuste que les estimateurs de type Ohser & Stoyan (Kiêu & Mora 1999).

Pour les travaux présentés dans ce document, nous avons choisi d'utiliser la méthode de correction des effets de bord proposée par Ripley (1977). Elle consiste à remplacer, pour les points situés près des bords du domaine d'étude, le coefficient k_{ij} par l'inverse de la proportion du périmètre du cercle C_{ij} (centré sur i et passant par j) qui se trouve dans le domaine d'étude. Ce qui correspond à une estimation du nombre de points à même distance qui seraient en dehors du domaine d'étude :

$$k_{ij} = \frac{\text{périmètre total}}{\text{périmètre dans le domaine d'étude}} = \frac{2\pi r}{C_{\text{intérieur}}} \geq 1$$

Ripley (1977) montre que l'estimateur $\hat{K}(r)$ qui prend en compte cette correction des effets de bord est non biaisé. La formule est très générale et peut être utilisée pour de nombreuses autres fonctions d'analyse de la structure spatiale.

La Figure 7 compare l'estimation de $K(r)$ obtenue pour un semis de Poisson avec et sans correction de l'effet de bord. Le biais introduit lorsque les effets de bord ne sont pas corrigés

est relativement important aux grandes distances, si bien qu'il est fortement souhaitable d'effectuer la correction pour pouvoir interpréter les courbes correctement.

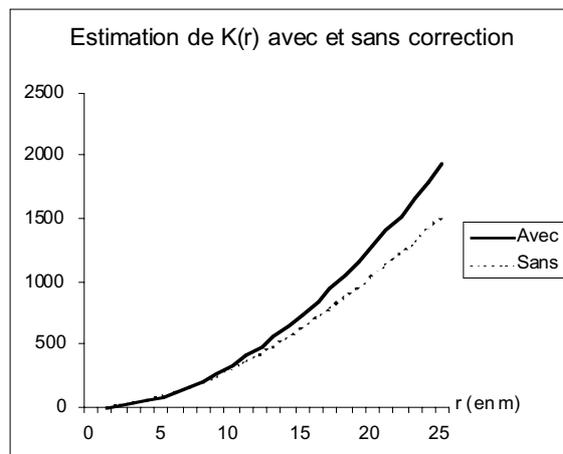


Figure 7 : Comparaison des estimations de la fonction $K(r)$ calculée, pour un même semis de Poisson de 200 points sur un domaine carré de 100 x 100 m, avec et sans correction des effets de bord.

En théorie, cette méthode permet d'obtenir un estimateur non biaisé de $K(r)$ quelle que soit la forme du domaine d'étude. Cependant, les formules explicites qu'on peut trouver dans la littérature ne s'appliquent généralement qu'à des domaines d'étude de formes simples, circulaires ou rectangulaires, et dans ce dernier cas seulement pour des distances r inférieures à la moitié de la largeur du rectangle (voir par exemple Diggle 1983). Certains programmes prennent en compte des domaines de forme polygonale (Rowlingson & Diggle 1993), mais restent souvent peu documentés. Nous avons donc été amenés à calculer et expliciter les formules de correction des effets de bord, d'une part pour des distances plus grandes dans le cas rectangulaire, et d'autre part dans le cas de domaines d'étude de forme complexe, approchés par des polygones (Goreaud & Pélissier 1999). Ces formules permettent de calculer un estimateur non biaisé de $K(r)$ jusqu'à des valeurs de r égales à la moitié de la longueur de la zone étudiée, ce qui peut s'avérer utile lorsqu'on travaille sur un transect par exemple. Il est toutefois prudent de limiter l'interprétation à des distances moindres, par exemple de l'ordre du quart de la longueur du domaine d'étude, car le nombre de couples de points utilisés dans le calcul diminue rapidement lorsque r augmente, si bien que les valeurs de $\hat{K}(r)$ n'ont plus beaucoup de sens aux grandes distances.

De plus, il est également prudent d'estimer $K(r)$ pour des intervalles de distance supérieurs à la précision du positionnement des individus. Pour un peuplement forestier, par exemple, la précision sur le positionnement des individus est souvent de l'ordre de 0,1 à 0,5 m. Il est alors conseillé de choisir un intervalle de distance d'au moins 1 ou 2 m pour estimer $K(r)$.

Notons que l'estimateur $\hat{L}(r)$ présente l'avantage d'avoir une variance beaucoup plus stable que $\hat{K}(r)$. Par contre, il n'est plus non-biaisé à cause de la non-linéarité de la transformation.

Dans ADS, ces estimateurs sont calculés pour des valeurs de r équidistantes d'un pas dr , tel que $r = tdr$ avec $t = 1, \dots, t_{\max}$. La surface comprise entre les cercles tdr et $(t-1)dr$ correspond à $S(r) = \pi(tdr)^2 - \pi((t-1)dr)^2$. La probabilité de rencontrer un point du semis dans cette surface est estimée par $\hat{P}(r) = \hat{N}(tdr) - \hat{N}((t-1)dr)$, à partir de laquelle on peut estimer $g(r)$:

$$\hat{g}(r) = \hat{P}(r) / \hat{\lambda}S(r) = [\hat{K}(tdr) - \hat{K}((t-1)dr)] / S(r)$$

2.2.3 – Construction d'un intervalle de confiance.

Les différents estimateurs des fonctions de voisinage de second ordre sont des variables aléatoires avec une certaine variance, si bien que pour un même processus de Poisson dans un domaine d'étude borné, on observe généralement de légères fluctuations des valeurs estimées (Figure 8).

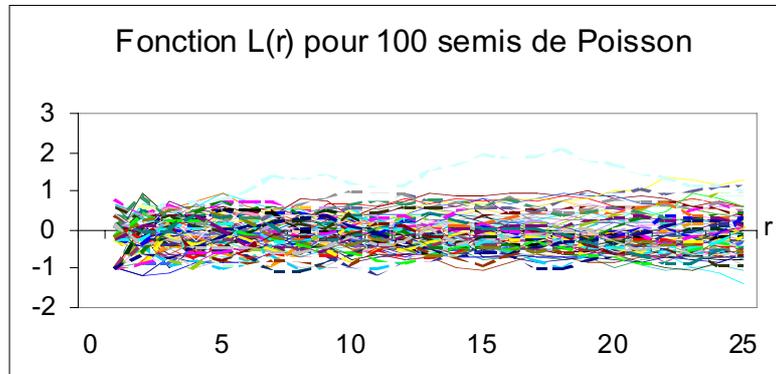


Figure 8 : Fonctions $L(r)$ obtenues pour 100 réalisations d'un processus de Poisson de 100 points sur un domaine carré de 100 x 100 m.

Pour tester l'hypothèse nulle correspondant à un processus de Poisson à partir de données réelles, il est donc nécessaire de tenir compte de cet aléa. Comme la loi des estimateurs des fonctions de voisinage de second ordre n'est pas connue, on utilise généralement la méthode de Monte Carlo pour estimer leur variation (Besag & Diggle 1977). Pour cela, on simule un grand nombre (N_{sim}) de réalisations d'un processus de Poisson de même densité que le semis réel étudié. Notamment à cause de la correction des effets de bord, et particulièrement pour des domaines d'étude de forme complexe, il est nécessaire de simuler un processus de Poisson sur un même domaine d'étude et avec un même nombre de points que le semis étudié. On calcule alors pour chaque réalisation du processus les valeurs des différents estimateurs aux distances r considérées. En toute rigueur, les N_{sim} simulations devraient être indépendantes pour chaque distance r . Mais en pratique, on peut considérer que le biais introduit en utilisant les mêmes simulations pour toutes les distances est négligeable (Goreaud 2000).

A r fixé, on peut construire un test statistique de l'hypothèse nulle en comparant la déviation absolue par rapport à la valeur théorique attendue des valeurs estimées, pour le semis réel et pour les différentes réalisations du processus de Poisson (Barot *et al.* 1999). Pour les différents estimateurs, cette déviation est définie par :

$$\delta n(r) = |\hat{n}(r) - \lambda| ; \delta K(r) = |\hat{K}(r) - \pi r^2| ; \delta g(r) = |\hat{g}(r) - 1| \text{ et } \delta L(r) = |\hat{L}(r)|.$$

Pour cette valeur de r , l'hypothèse nulle ne pourra pas être rejetée lorsque la déviation absolue calculée pour le semis réel sera inférieure ou égale à la déviation absolue calculée pour une réalisation du processus de Poisson. La proportion des N_{sim} simulations qui conduisent au non-rejet de l'hypothèse nulle donne le seuil de signification du test (P -value).

Toujours pour r fixé, on peut construire un intervalle de confiance local pour un risque de première espèce α donné, c'est-à-dire tel que la valeur de la fonction estimée pour un semis de Poisson sorte de cet intervalle avec une probabilité α . Les bornes de cet intervalle peuvent être estimées de façon robuste à partir des valeurs triées, obtenues pour les N_{sim} simulations, par les statistiques d'ordre $N_{sim} \alpha/2$ et $N_{sim} (1-\alpha/2)$.

Pour une distance r fixée, la comparaison des fonctions estimées pour le semis réel et des bornes de l'intervalle de confiance est un test de l'hypothèse nulle : si, pour cette valeur de r , la fonction estimée pour le semis réel est en dehors de l'intervalle de confiance, alors on peut conclure (avec un risque α) que la structure du semis à cette distance est significativement régulière ou agrégée, selon le signe de la déviation. Si par contre la fonction estimée pour le semis réel est à l'intérieur de l'intervalle, on ne peut rejeter l'hypothèse d'une structure aléatoire.

Il faut noter que la précision de l'estimation de cet intervalle de confiance augmente avec le nombre de simulations (Figure 9). De nombreuses publications, en particulier les plus anciennes, utilisent un intervalle de confiance calculé à partir d'un petit nombre de simulations : 19 simulations pour un risque de 5% ou 99 pour un risque de 1%. On considère de façon générale que le test est peu fiable lorsque $\alpha N_{simu} < 5$ (Martens *et al.* 1997). Grâce à la puissance des ordinateurs actuels, il est possible, et souhaitable, d'utiliser un nombre de simulations plus élevé. Dans ce document nous avons utilisé 10000 simulations.

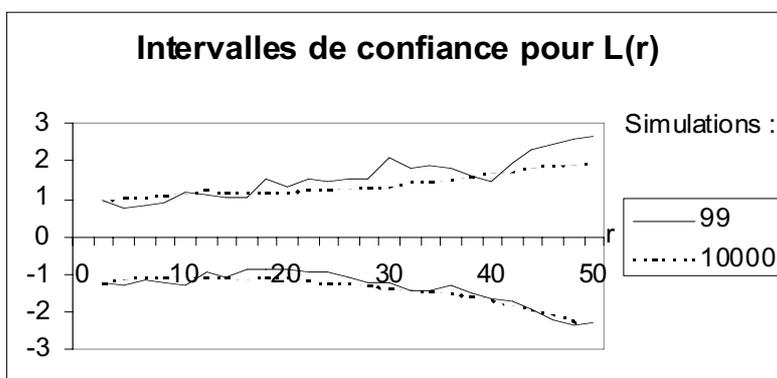


Figure 9 : Exemple d'un intervalle de confiance à 1% calculé pour la fonction $L(r)$ avec 99 et 10000 simulations, pour un semis de 100 points sur un domaine carré de 100 x 100 m.

La largeur de l'intervalle de confiance, et donc sa capacité à discriminer des hypothèses, augmente avec le nombre de points N et diminue avec la surface D du domaine d'étude (Figure 10). Ripley (1979) précise que, pour un domaine rectangulaire et une large gamme de distances² r , les valeurs des bornes de l'intervalle de confiance peuvent être approchées par $\pm 1,68\sqrt{D/N}$ pour un risque $\alpha = 1\%$.

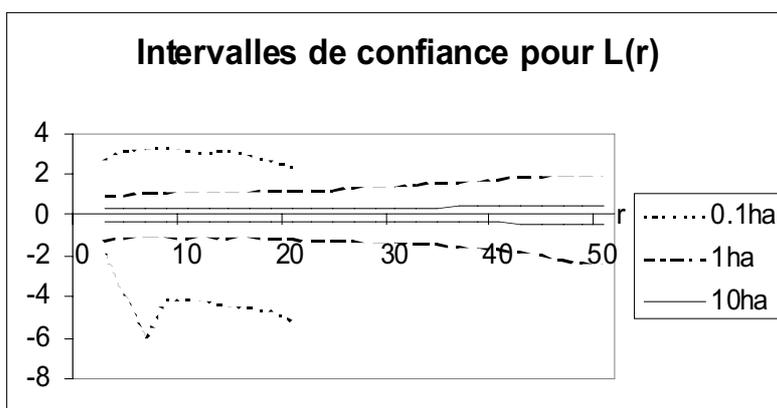


Figure 10 : Intervalles de confiance locaux à 1% calculés pour un semis de 100 points sur un domaine carré de 0,1 ha, 1 ha et 10 ha.

² En fait pour des distances r inférieures au quart de la largeur du domaine d'étude et supérieures à la distance moyenne entre arbres.

Il est important de bien noter que cet intervalle de confiance est local : il n'a de sens que pour une valeur donnée de r . Cependant, dans la pratique on a souvent besoin d'analyser la structure spatiale d'un semis de points sur l'ensemble des valeurs calculées aux différentes distances r (par exemple de 1 à 25 m). En toute rigueur, il faudrait construire pour cela un intervalle de confiance *global*, constitué par exemple, de deux fonctions $L_-(r)$ et $L_+(r)$ telles que α soit le risque qu'une courbe $L(r)$, calculée sur un semis de Poisson donné, sorte de l'intervalle $[L_-(r); L_+(r)]$ au moins une fois à une distance r quelconque. Diggle (1983, p.77) propose un test global sur une mesure intégrale de la distance entre les valeurs mesurées de $K(r)$ et les valeurs moyennes obtenues par simulation de Monte Carlo pour l'hypothèse testée. Mais un intervalle de confiance global reste difficile à construire en pratique et l'intervalle de confiance local est souvent utilisé pour analyser la structure spatiale sur l'ensemble des distances r considérées. Cette erreur entraîne une sous-estimation du risque de première espèce, car le risque global concerne toutes les courbes qui sortent de l'intervalle de confiance à quelque distance que ce soit, alors que le risque local ne concerne que les courbes qui en sortent à une distance r fixée, qui sont moins nombreuses. On peut estimer ce risque global par le pourcentage de courbes calculées sur des semis de Poisson, qui sortent d'un intervalle de confiance local, calculé préalablement pour un risque α fixé. Pour un semis de 100 points sur un domaine carré de 100 x 100 m et 10000 simulations, on obtient un risque de première espèce global d'environ 8.8% pour $\alpha = 1\%$, 36% pour $\alpha = 5\%$, et 56% pour $\alpha = 10\%$.

On peut donc considérer qu'un intervalle de confiance local calculé avec un risque $\alpha = 1\%$ donne une bonne approximation d'un intervalle de confiance global pour un risque de première espèce de l'ordre de 10%. Par abus de langage, on parlera alors d'agrégation ou de régularité significative lorsque la fonction $\hat{L}(r)$ sortira de l'intervalle de confiance, bien qu'il ne s'agisse pas dans ce cas d'un test statistique (Figure 11).

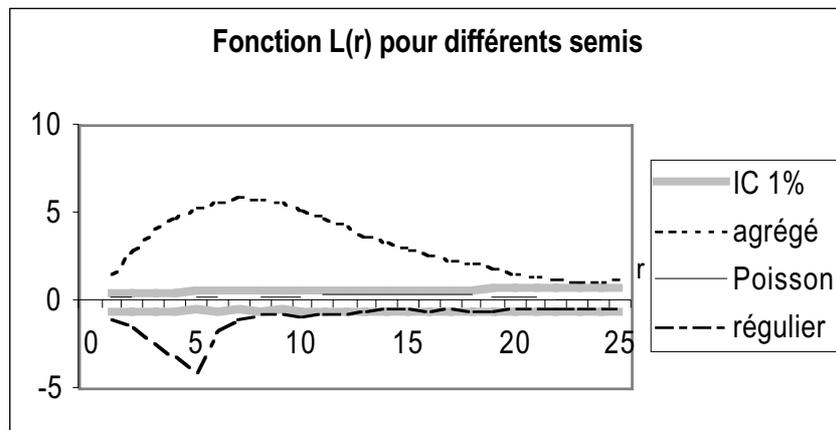


Figure 11 : Positions relatives des fonctions $L(r)$ calculées pour trois semis caractéristiques (agrégé, de Poisson et régulier) par rapport à l'intervalle de confiance de l'hypothèse nulle d'une répartition aléatoire du même nombre de points dans le même domaine d'étude.

Notons enfin que, comme le risque de première espèce global est mal connu et certainement assez élevé, il faut rester prudent dans les interprétations, en particulier lorsque la fonction estimée ne sort que légèrement de l'intervalle de confiance et en un seul point. Après la première phase d'analyse des données, il est toujours possible de faire un test rigoureux, pour une distance r fixée, en utilisant un deuxième jeu de données. Dans ce cas, l'utilisation de l'intervalle de confiance local à la distance r est pleinement justifiée.

2.3 — Prise en compte de l'hétérogénéité.

L'hypothèse d'homogénéité qui préside à l'utilisation des fonctions de voisinage de second ordre n'est pas toujours satisfaisante, notamment lorsqu'on s'intéresse à des systèmes biologiques qui présentent souvent des sources variées d'hétérogénéité. Ripley (1981) définit un semis de points hétérogène comme "... *a spatial point pattern that vary in a systematic way from place to place ...*". Pour pouvoir tester cette hypothèse rigoureusement, il est donc nécessaire de disposer de plusieurs réalisations du processus étudié. En pratique, lorsqu'on ne dispose que d'une seule réalisation du processus sous-jacent, la notion d'hétérogénéité dépend fortement de l'échelle d'observation choisie (Goreaud 2000).

Sous l'hypothèse d'ergodicité, les propriétés de premier et second ordres d'un semis de points homogène sont définies comme des espérances (c'est-à-dire des moyennes) sur le domaine étudié. Pour un semis de points hétérogène, ces définitions ne sont plus pertinentes, car les propriétés du processus peuvent varier selon la position. On peut par contre définir des valeurs locales de ces propriétés, et analyser la manière dont varient localement ces propriétés à l'intérieur du domaine d'étude (Getis & Franklin 1987; Pélissier & Goreaud 2001). C'est la méthode que nous avons utilisé dans ce document, où nous considérons deux types de fonctions locales dérivées du formalisme des processus ponctuels : la densité locale basée sur les propriétés locales de premier ordre du semis de points et la densité locale de voisinage de second ordre.

2.3.1 – La densité locale permet de déceler certaines hétérogénéités.

La densité correspond simplement au nombre de points par unité de surface. On peut définir, pour une position quelconque A de coordonnées (x,y) , la fonction locale $N_A(r)$ correspondant au nombre de points contenus dans le disque de rayon r centré en A . La densité locale, c'est-à-dire le nombre de points par unité de surface calculé dans un disque de rayon r centré en A , se déduit simplement par :

$$n_A(r) = N_A(r) / \pi r^2$$

où πr^2 correspond à la surface du disque de rayon r .

Pour un processus homogène d'intensité λ , l'espérance de la fonction de densité locale, calculée sur les différentes réalisations du processus, est $E[n_A(r)] = \lambda$ quelle que soit la position (x,y) de A et quelle que soit la distance r considérée. La figure 12a présente quatre réalisations d'un processus de Neyman-Scott homogène d'intensité $\lambda = 0,02$ (10 agrégats de 10 m de rayon et comportant 20 arbres sur un domaine d'étude carré de 100 x 100 m). A partir de 1000 réalisations de ce processus, Goreaud (2000) montre que la densité locale estimée dans un disque de rayon $r = 5$ m situé au centre du domaine d'étude est 0,020079, très proche de la valeur théorique $\lambda = 0,02$. L'histogramme des valeurs locales estimées en 1000 positions aléatoires, toujours sur ces 1000 réalisations, montre une distribution unimodale centrée sur la valeur théorique 0,02 (Figure 12b).

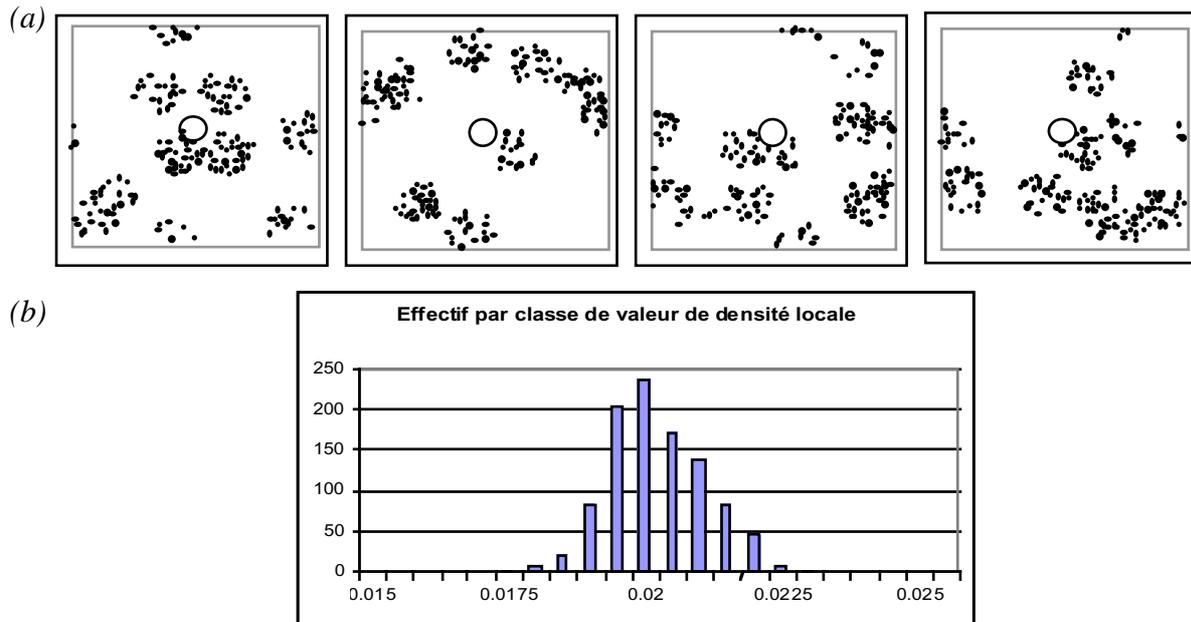


Figure 12 : (a) Exemple de 4 réalisations d'un processus de Neyman-Scott homogène (10 agrégats de rayon 10 m et comportant 20 arbres chacun sur un domaine d'étude de 100 x 100 m). (b) Histogramme des valeurs locales de la densité, estimées en 1000 positions aléatoires du domaine d'étude, avec une fenêtre d'échantillonnage circulaire de rayon 5 m, et pour 1000 réalisations du processus de Neyman-Scott.

Inversement, pour un processus ponctuel qui présente une hétérogénéité de la densité locale, les valeurs estimées en différentes positions du domaine d'étude sont différentes, et l'histogramme des valeurs locales s'éloigne généralement d'une distribution unimodale.

Lorsque plusieurs réalisations d'un processus sont disponibles, l'histogramme des valeurs de la densité locale, estimées en plusieurs positions du domaine d'étude, permet donc généralement (sauf processus hétérogènes très particuliers) de détecter des hétérogénéités de la densité. Dans certains cas de processus homogènes, la loi des estimateurs de la densité locale est connue et permet de tester de façon formelle, à r fixé, l'hypothèse d'homogénéité de la densité. Ainsi, pour un processus de Poisson homogène d'intensité λ , $N_A(r)$ suit une distribution de Poisson de paramètre $\lambda\pi r^2$.

Lorsqu'on ne dispose que d'une seule réalisation du processus (le semis de points étudié), la densité locale, qui est une moyenne sur un grand nombre de réalisations, n'a plus de sens. Considérons par exemple une seule réalisation du processus de Neyman-Scott de la Figure 12. Les valeurs de la densité locale varient fortement d'une position à l'autre, traduisant simplement le fait que tout semis de points, même s'il est la réalisation d'un processus homogène, présente des irrégularités dans la position des points (Figure 13).

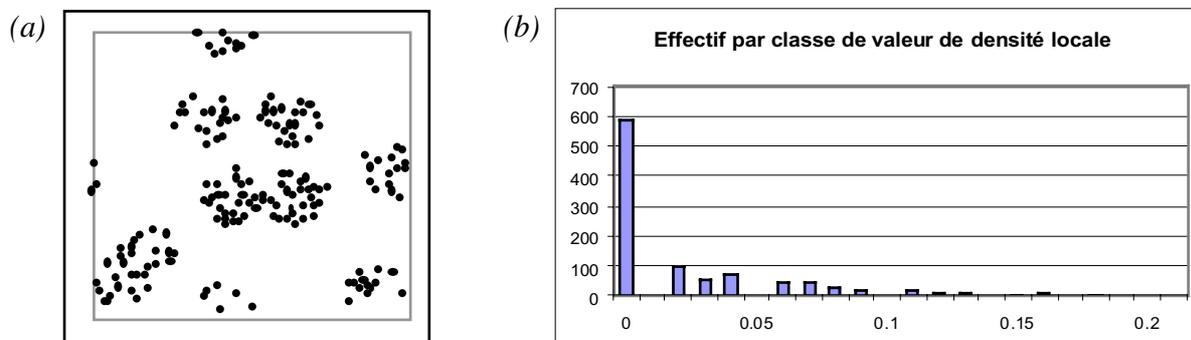


Figure 13 : (a) Une réalisation d'un processus de Neyman-Scott. (b) Histogramme des valeurs de la densité locale estimées en 1000 positions aléatoires du domaine d'étude, avec une fenêtre d'échantillonnage circulaire de rayon 5 m, pour cette seule réalisation.

La notion d'hétérogénéité dépend alors fortement de l'échelle d'observation. Le semis de points de la Figure 14 est simulé par la superposition d'un processus de Poisson homogène et du processus de Neyman-Scott précédent. Sur l'ensemble du domaine d'étude D3, c'est-à-dire à une échelle d'observation de 200 m, le semis est homogène avec une structure agrégée, et peut être modélisé par la superposition des deux processus homogènes simples. Dans le domaine d'étude D2, c'est-à-dire à une échelle d'observation de 50 m, le semis de points est perçu comme hétérogène, car il présente une forte différence de densité entre la partie gauche et la partie droite. Il pourrait être modélisé par un processus hétérogène introduisant un gradient de densité de la gauche vers la droite. Pour les domaines d'études plus petits D1 et D1', correspondant à une échelle d'observation de 10 m, le semis apparaît homogène et pourrait être modélisé dans les deux cas par un processus de Poisson, d'intensité plus forte en D1 qu'en D1'.

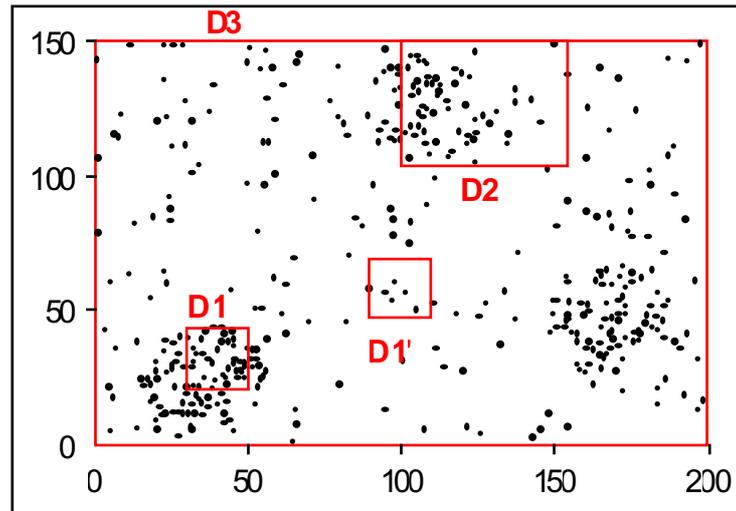


Figure 14 : Exemple d'un semis simulé qui peut être considéré comme homogène ou hétérogène selon l'échelle d'observation.

On voit donc apparaître une ambivalence entre structure et hétérogénéité liée à l'échelle d'observation. Par exemple, la zone de forte densité du domaine D2 crée une hétérogénéité à cette échelle. Mais elle peut être considérée comme un élément de structure à l'échelle D3 de l'ensemble du domaine. Ces considérations permettent, dans certains cas, d'expliquer le comportement des fonctions $L(r)$ calculées sur des semis hétérogènes (Pélissier & Goreaud 2001).

Pour un semis de points unique, si l'on précise le choix des échelles (Figure 15), la densité locale peut toutefois permettre de caractériser d'éventuelles hétérogénéités. Il est d'abord nécessaire de préciser E1, l'échelle de la structure spatiale que l'on souhaite étudier. Par définition, les irrégularités dans la position des points du semis observées à des échelles inférieures à E1, seront considérées comme des éléments de structure (taille des agrégats, distance minimale entre points pour une régularité, etc.), et les irrégularités observées à des échelles supérieures à E1 comme des hétérogénéités. La taille E3 du domaine d'étude est déterminante pour le choix de E1 : E3 limite la taille maximale à laquelle on peut étudier la structure et les résultats seront d'autant meilleurs que E3 est supérieure à E1, c'est-à-dire que la structure étudiée aura de chances de se reproduire dans E3. L'échelle E2, comprise entre E1 et E3, précise l'échelle d'échantillonnage des valeurs de la densité locale. E2 doit être suffisamment supérieure à E1 pour que les valeurs de densité locale soient relativement indépendantes de la structure étudiée. Le choix de E2 est également contraint par l'échelle d'observation E3, car il est nécessaire de disposer d'un nombre suffisant d'échantillons pour étudier les variations de densité locale, et qu'on peut souhaiter que ces échantillons restent relativement indépendants, c'est-à-dire ne se superposent pas trop. En pratique, le système d'échantillonnage le plus efficace consiste à répartir les échantillons sur une grille systématique, idéalement en respectant la relation $E1 \ll E2 \ll E3$, ce qui n'est pas toujours facilement réalisable.

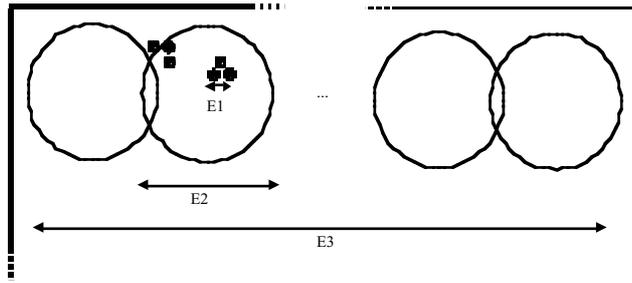


Figure 15 : Contraintes sur le choix des échelles E1, E2 et E3.

Pour estimer, dans un domaine d'étude donné, la densité locale à la position A de coordonnées (x,y) , représentant par exemple un nœud d'une grille d'échantillonnage, on définit l'estimateur :

$$\hat{N}_A(r) = \sum_{i=1}^N k_{Ai}$$

où N est le nombre de points dans le domaine d'étude ; et k_{Ai} prend la valeur 1 si la distance entre la position A et un point i du semis est inférieure à r , 0 sinon. On déduit alors un estimateur de la densité locale par : $\hat{n}_A(r) = \hat{N}_A(r) / \pi r^2$.

Ces estimateurs comportent un biais lié aux effets de bord, que l'on corrige, comme pour l'estimation de $K(r)$, par la méthode de Ripley (1977) : lorsque A est situé près de la bordure du domaine d'étude, k_{Ai} est remplacé par l'inverse de la proportion du périmètre du cercle C_{Ai} centré en A et passant par i , qui se trouve dans le domaine d'étude.

Considérons à titre d'exemple, un semis virtuel hétérogène dans un domaine carré de 100×100 m ($E3 = 100$ m), dont la moitié supérieure correspond à une réalisation d'un semis de Poisson d'intensité $\lambda = 0,05$ et la partie inférieure à une réalisation d'un semis de Poisson d'intensité $\lambda = 0,01$ (Figure 16a). Supposons que la structure que l'on souhaite étudier définisse le choix de $E1 = 10$ m. On pourra choisir $E2 = 25$ m, c'est-à-dire des disques d'échantillonnage de 12,5 m de rayon, de manière à respecter $E1 \ll E2 \ll E3$. Si l'on souhaite que ces échantillons soient indépendants, ils doivent être répartis sur une grille systématique de $25 \text{ m} \times 25 \text{ m}$, soit 16 positions d'échantillonnage, ce qui est peu. Dans la mesure où il n'est de toute façon pas possible de tester formellement l'hypothèse d'homogénéité de la densité à partir d'une unique réalisation du processus, on préférera augmenter le nombre d'échantillons en utilisant, par exemple, une grille de 5×5 m, même si, dans ce cas, les échantillons ne sont plus complètement indépendants. L'histogramme de la Figure 16b montre clairement une distribution bimodale qui correspond à la superposition des deux distributions de Poisson, et met par conséquent en évidence l'hétérogénéité de la densité du semis étudié.

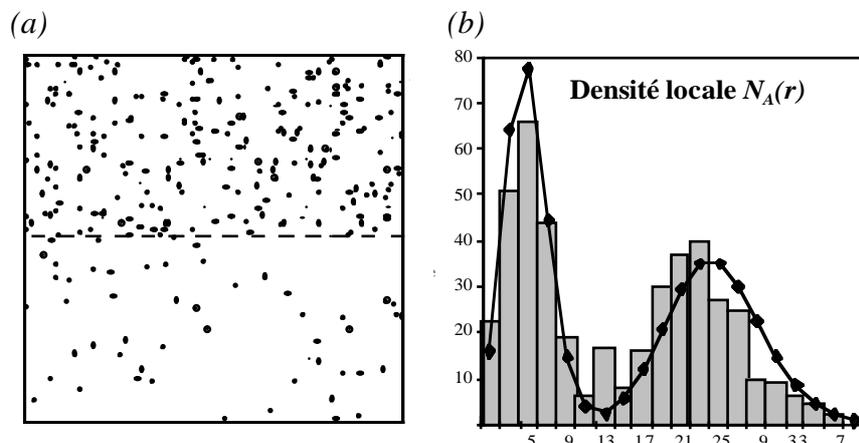


Figure 16 : (a) Exemple de semis virtuel hétérogène correspondant à la juxtaposition de 2 semis de Poisson d'intensités différentes sur un domaine d'étude de 100×100 m. (b) Histogramme des valeurs de la densité locale estimées aux nœuds d'une grille de 5×5 m avec une fenêtre d'échantillonnage circulaire de rayon 12,5 m.

La fonction $L(r)$ estimée pour ce semis virtuel (Figure 17) montre alors une divergence positive aux grandes valeurs de r par rapport à l'intervalle de confiance de l'hypothèse nulle d'un semis de Poisson homogène : la zone dense est perçue comme un gros agrégat dont la taille est supérieure à l'échelle d'observation.

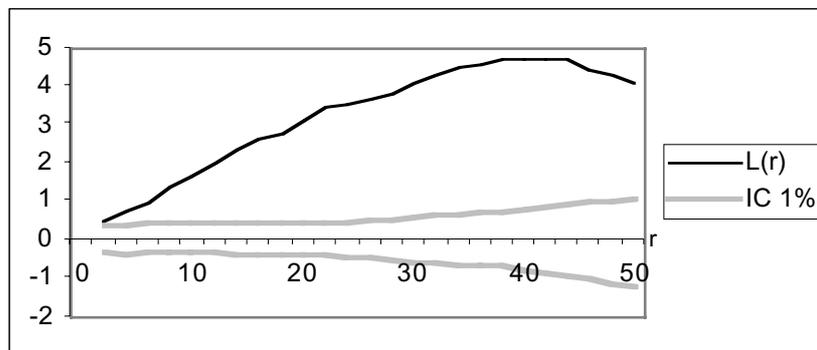


Figure 17 : Fonction $L(r)$ estimée pour le semis hétérogène de la Figure 16 avec un intervalle de confiance à 1% de l'hypothèse nulle d'un semis de Poisson homogène.

Bien que l'analyse de la distribution des valeurs de la densité locale ne fournisse pas de test formel de l'hypothèse d'homogénéité du processus sous-jacent, nous verrons à partir d'exemples au chapitre suivant, comment celle-ci peut-être utilisée pour définir des zones homogènes à l'intérieur desquelles la structure spatiale pourra être décrite correctement par la fonction $L(r)$.

2.3.2 – La densité locale de voisinage de second ordre.

Même lorsque la densité locale du semis étudié est homogène, la structure spatiale peut présenter des hétérogénéités. L'étude des valeurs locales de la propriété de second ordre permet parfois de mettre en évidence de telles hétérogénéités.

Par analogie avec les fonctions de Ripley, on peut définir pour chaque point i d'un semis dans un domaine d'étude donné, une fonction locale (ou individuelle) $N_i(r)$, correspondant au nombre de voisins situés dans un disque de rayon r centré sur i . La densité locale de voisinage de second ordre, c'est-à-dire le nombre de voisins par unité de surface, calculé dans un disque de rayon r centré sur i , se déduit de cette fonction par :

$$n_i(r) = N_i(r) / \pi r^2$$

où πr^2 correspond à la surface du disque de rayon r .

Pour un processus de Poisson d'intensité λ , l'espérance du nombre de voisins à distance r d'un point i du semis est $\lambda \pi r^2$, donc $E[n_i(r)] = \lambda$. Une valeur de $n_i(r) > \lambda$ indique donc une agrégation locale autour du point i , et une valeur de $n_i(r) < \lambda$ indique une régularité locale autour du point i .

Pour un semis de points réel, on estime, sous l'hypothèse d'ergodicité, le nombre de voisins contenus dans un disque de rayon r centré sur i , par :

$$\hat{N}_i(r) = \sum_{j=1, i \neq j}^N k_{ij}$$

où k_{ij} prend la valeur 1 si la distance entre les points i et j du semis est inférieure à r , et 0 sinon.

En transposant les relations du paragraphe 2.2.1, on peut alors définir les estimateurs individuels : $\hat{n}_i(r)$, $\hat{K}_i(r)$, $\hat{g}_i(r)$ et $\hat{L}_i(r)$ (Getis & Franklin 1987). Comme pour la fonction de Ripley, le biais introduit par les effets de bord est corrigé en remplaçant, pour les points i situés près de la bordure du domaine d'étude, le coefficient k_{ij} par l'inverse de la proportion du périmètre du cercle C_{ij} , centré sur i et passant par j , qui se trouve dans le domaine d'étude.

Comme nous l'avons vu au paragraphe précédent, la structure spatiale d'un semis peut apparaître comme une hétérogénéité selon l'échelle d'observation à laquelle on se place, et on peut toujours trouver des différences locales dans la position relative des points les uns par rapport aux autres. Getis & Franklin (1987) proposent d'utiliser une cartographie des valeurs individuelles de la fonction $L_i(r)$ à r fixé, pour mettre ces variations en évidence (Figure 18).

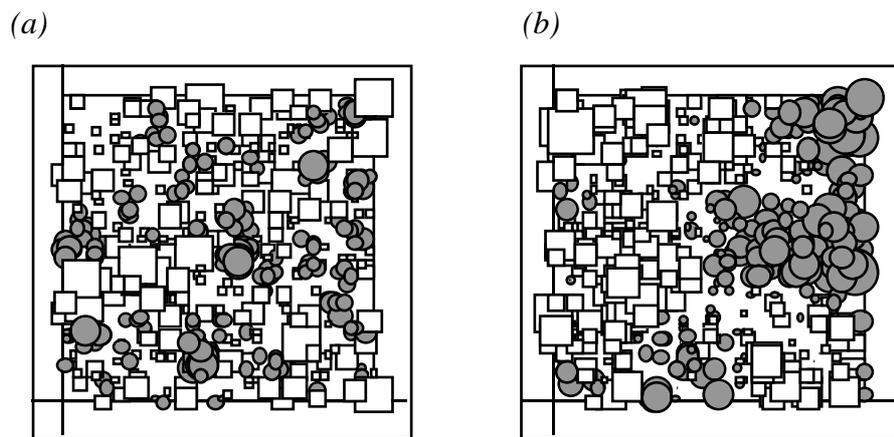


Figure 18 : Représentation des valeurs estimées de la fonction $L_i(r)$ à $r = 5$ m (a) et $r = 20$ m (b) pour un semis de Poisson de 100 points sur un domaine carré de 100×100 m. Les cercles (valeurs positives) et les carrés (valeurs négatives) sont de taille proportionnelle à la valeur estimée de $L_i(r)$ au point i .

A certaines échelles, ces différences locales peuvent faire apparaître une zonation géographique à l'intérieur du domaine d'étude. Le semis présente alors une hétérogénéité des propriétés locales de second ordre qui peut conduire à une mauvaise interprétation de la structure moyenne du semis décrite par la fonction de Ripley (Pélissier & Goreaud 2001). On peut généraliser cette approche à l'ensemble des fonctions individuelles d'analyse de voisinage de second ordre introduites précédemment.

Ces fonctions étant liées à un point particulier du semis, elles ne peuvent être considérées comme des fonctions aléatoires du processus sous-jacent. Par conséquent, il n'existe pas de test formel pour ces fonctions. Nous verrons cependant à partir d'exemples au chapitre suivant comment les valeurs individuelles peuvent être utilisées pour définir des zones homogènes, à l'intérieur desquelles la structure spatiale pourra être décrite correctement par la fonction $L(r)$.

3 — Applications à des données d'écologie forestière.

Le formalisme des processus ponctuels s'étend à l'étude des processus non-stationnaires, c'est-à-dire qui ne respectent pas l'hypothèse préalable d'homogénéité. On peut trouver dans la littérature des outils spécifiques permettant d'analyser certains de ces processus (voir par exemple Diggle 1983), mais la théorie mathématique sous-jacente, plus complexe que pour les processus homogènes, reste souvent difficile à appliquer à des cas réels pour lesquels on ne connaît pas *a priori* le processus à l'origine des structures observées (Dessart 1996, Batista & Maguire 1998).

L'objectif de ce paragraphe est d'illustrer, à partir d'exemples pris en écologie forestière, une approche qui permet d'utiliser les outils d'analyse des processus homogènes introduits dans les pages précédentes (et donc les modules d'ADS), pour étudier la structure spatiale dans certains cas relativement simples de peuplements forestiers hétérogènes. Les irrégularités à grande distance dans la position des points d'un semis sont interprétées comme l'expression d'une hétérogénéité qui permet de définir, à l'intérieur d'un domaine d'étude hétérogène, des sous-régions homogènes à l'intérieur la structure spatiale peut être décrite correctement par les fonctions de Ripley (Pélissier & Goreaud 2001).

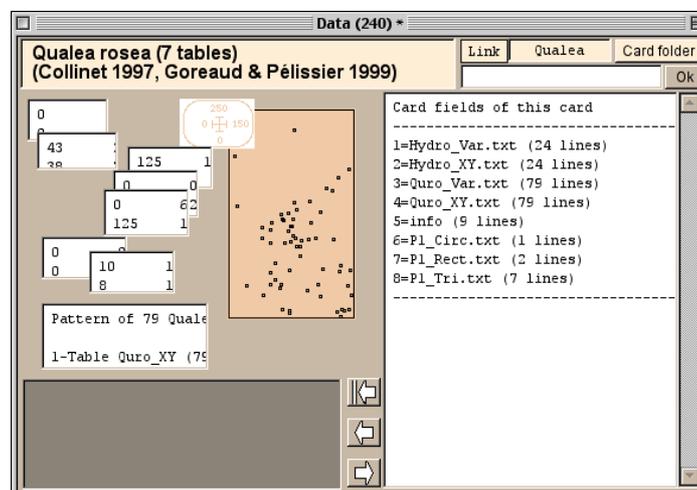
3.1 – Exemple 1 : Dispositif de Paracou (Guyane française).

Les formations végétales sont généralement sensibles aux différentes conditions locales de l'environnement (comme le sol, la topographie, l'ensoleillement, l'humidité, etc.), dont les variations peuvent entraîner une hétérogénéité de la structure spatiale de l'écosystème. On parle dans ce cas d'hétérogénéité d'origine exogène, c'est-à-dire due à des facteurs extérieurs au système étudié (Pélissier & Goreaud 2001). Ainsi, on constate souvent que la composition et la structure d'un peuplement forestier varient, d'un endroit à l'autre, en fonction de ces facteurs. Collinet (1997) a montré sur le dispositif expérimental de Paracou en Guyane française, que la densité de la plupart des espèces était influencée par le fonctionnement hydrodynamique des sols. A l'échelle d'une parcelle d'étude, ces irrégularités peuvent entraîner une hétérogénéité de la structure spatiale.

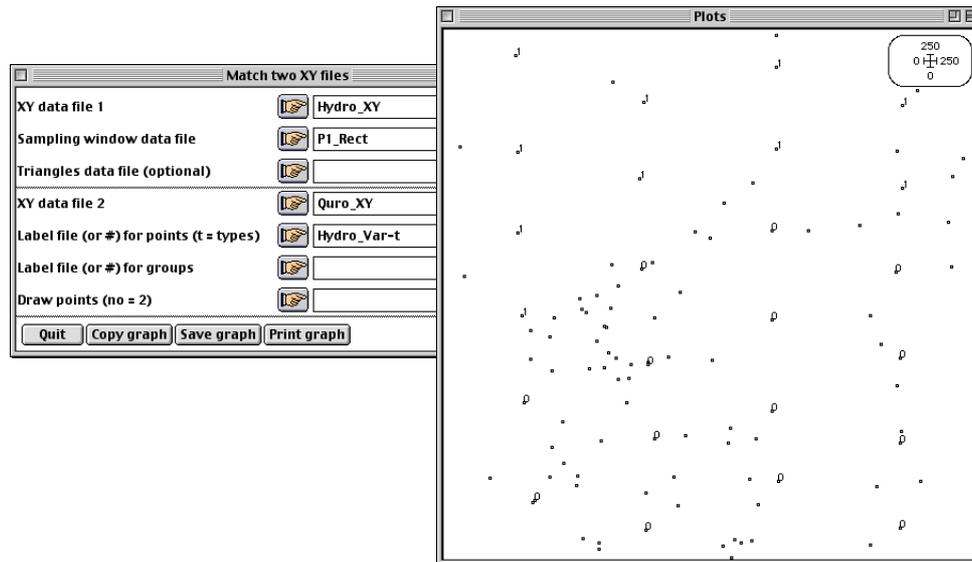
Ce premier exemple présente le cas de l'espèce *Qualea rosea* (Vochysiaceae) sur la parcelle 1 du dispositif de Paracou, qui montre une distribution spatiale fortement hétérogène parce qu'elle fuit les sols de bas-fond qui connaissent des périodes prolongées de saturation en eau des horizons de surface (Collinet 1997). Cette première application montre comment l'hétérogénéité du semis invalide toute interprétation correcte de la structure spatiale, analysée sur l'ensemble de la parcelle par les fonctions de Ripley. On montre ensuite comment il est possible, à partir de données pédologiques, de définir à l'intérieur de cette parcelle une sous-région homogène, correspondant aux sols hors bas-fond, sur laquelle la méthode de Ripley permet d'interpréter correctement la structure spatiale du semis.

3.1.1 – Les données de la parcelle 1 de Paracou.

Les données de cette placette sont consignées dans la carte *Qualea* de la pile de données. Les tables *Quro_XY* et *P1_Rect* contiennent respectivement les coordonnées des 79 points du semis et les coordonnées de l'origine et du coin opposé du domaine d'étude rectangulaire. Les tables *Hydro_XY* et *Hydro_Var* contiennent respectivement les coordonnées de 24 points de sondages pédologiques (Barthès 1991) et un code binaire indiquant la présence (1) ou non (0) de traces de saturation en eau dans les horizons de surface (hydromorphie de surface) :



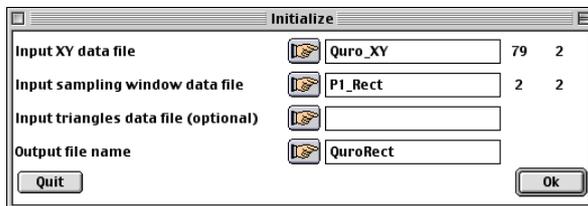
L'option Plots : Match two XY files permet de représenter le semis et les positions des sondages (le fichier Hydro_Var est préalablement converti au format texte) :



Le fond de carte est sauvegardé sous le nom Quro_Map.

3.1.2 – Analyse de la structure spatiale sur l'ensemble de la parcelle.

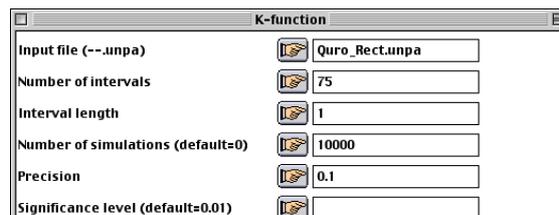
On prépare l'analyse de la structure spatiale du semis en liant les fichiers Quro_XY et P1_Rect par l'option Ripley : Initialize :



 ew TEXT file QuroRect.unpa contains the parameters:
 ----> XY data file: Quro_XY [79][2]
 ----> Shape of the sampling window: 1
 1 = rectangular
 2 = circular
 3 = irregular within a rectangular window
 4 = irregular within a circular window
 ----> Sampling window data file: P1_Rect [2][2]

Le listing signale la création du fichier de paramètres QuroRect.unpa qui servira d'entrée pour les options de calcul.

L'option Ripley : K-function permet de faire l'analyse de la structure spatiale que l'on choisit de réaliser sur 75 pas de distance de 1 m. La précision des coordonnées simulées pour le calcul d'un intervalle de confiance local est fixée à 0,1 m et le nombre de simulations de Monte Carlo à 10000 pour un risque de première espèce $\alpha = 0,01$:



Le listing renvoie une description des fichiers de données définis par QuroRect.unpa et rappelle les paramètres du calcul :

```
-----
Univariate second-order neighbourhood functions (Ripley 1977)
---XY data file: Quro_XY
It contains 79 points
---Sampling window data file: P1_Rect
Xmin: 0.0000e+00 Ymin: 0.0000e+00
Xmax: 2.5000e+02 Ymax: 2.5000e+02
Number of points within the rectangular sampling window: 79
---Input parameters
Area of the study region: 6.2500e+04
Number of points within the study region: 79 (density = 1.2640e-03)
rmax = 7.5000e+01
dr = 1.0000e+00
Number of Monte Carlo simulations = 10000
Precision of the simulated coordinates = 1.0000e-01
```

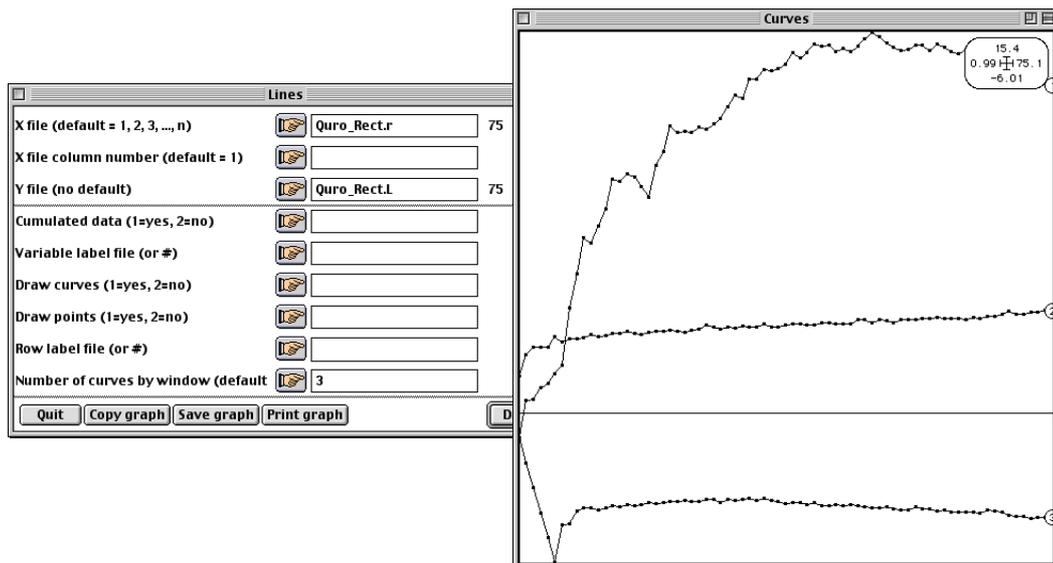
Le tableau suivant donne, pour chaque pas de distance r , le seuil de signification du test local de l'hypothèse nulle d'une répartition aléatoire des points du semis :

```
---Local significance level of the null hypothesis
P-values of |Obs| <= |Simu|
-----
|  r  |  g(r)  |  n(r)  |  K(r)  |  L(r)  |
-----
| 1.0000 | 9.999e-01 | 9.999e-01 | 9.999e-01 | 9.999e-01 |
| 2.0000 | 3.238e-01 | 9.999e-01 | 7.165e-01 | 7.165e-01 |
...
| 74.0000 | 5.693e-01 | 0.000e+00 | 0.000e+00 | 0.000e+00 |
| 75.0000 | 6.111e-01 | 0.000e+00 | 0.000e+00 | 0.000e+00 |
-----
```

Les fichiers de résultats contiennent les valeurs des différents estimateurs et les bornes de l'intervalle de confiance local à 1% :

```
---File QuroRect.r contains the distance steps
t has 75 rows and 1 column
---File QuroRect.g contains the pair density function
t has 75 rows and 3 columns
d 1: g(r)
d 2,3: 99% confidence interval
---File QuroRect.n contains the local neighbour density function
t has 75 rows and 3 columns
d 1: n(r)
d 2,3: 99% confidence interval
---File QuroRect.K contains the Ripley's function
t has 75 rows and 3 columns
d 1: K(r)
d 2,3: 99% confidence interval
---File QuroRect.L contains the modified Ripley's function
t has 75 rows and 3 columns
d 1: L(r)
d 2,3: 99% confidence interval
-----
```

Les résultats sont représentés graphiquement par l'option Curves : Lines en groupant la fonction estimée pour le semis réel (1) et les bornes de l'intervalle de confiance (2 et 3) :



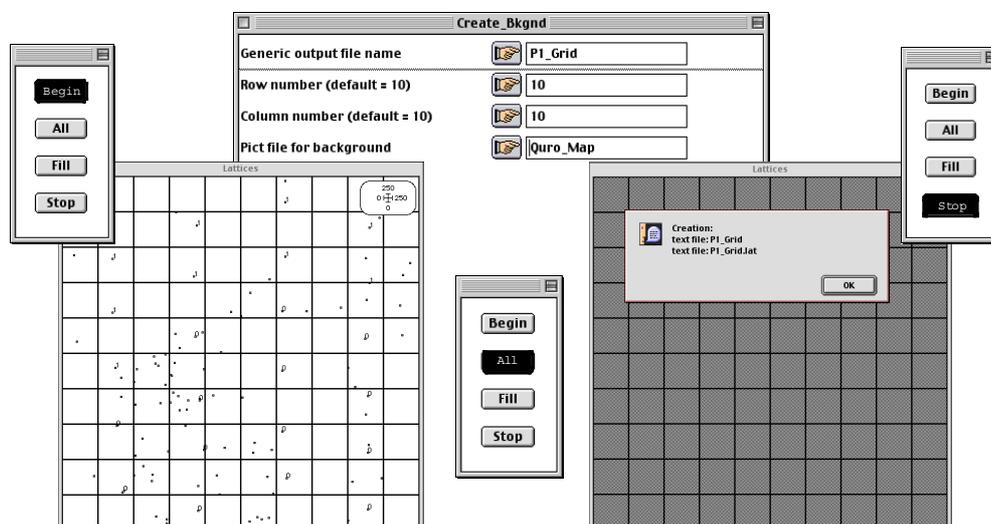
La divergence de la courbe $\hat{L}(r)$ aux grandes distances témoigne de l'hétérogénéité du semis (présence d'un unique agrégat de taille supérieure à l'échelle du domaine d'étude) qui empêche d'interpréter correctement la structure fine du semis (Pélissier & Goreaud 2001). Cette hétérogénéité provient de la forte variation de la densité du semis à l'échelle d'observation de la parcelle.

3.1.3 – Définition d'un domaine d'étude homogène.

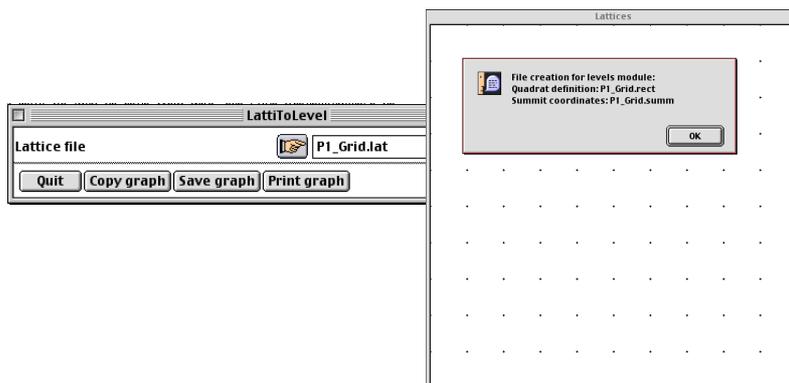
La relation avérée entre densité de *Qualea rosea* et un facteur pédologique cartographié nous ramène à un cas classique d'échantillonnage stratifié, où les valeurs 0 et 1 de la variable qualitative exogène permettent de définir des sous-régions homogènes. On utilisera la ligne de contour correspondant à la valeur 0,5 de la variable Hydro_Var pour délimiter ces sous-régions.

Le module Levels d'ADE-4 permet d'interpoler les valeurs d'une variable estimée aux nœuds d'une grille systématique par la méthode Lowess (local weighted scatterplot smoothing ; Cleveland 1979). Quelques étapes préalables sont cependant nécessaires :

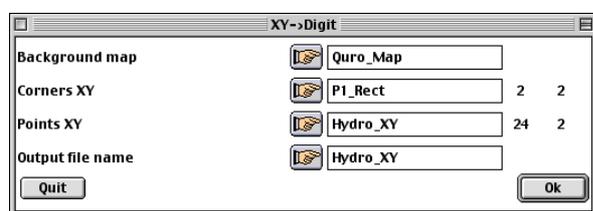
1) On prépare à partir du fond de carte Quro_Map, une grille d'échantillonnage de 10 x 10 m par l'option Lattices : Create Bkgnd :



Le programme signale la création de deux fichiers à partir desquels l'option Lattices : LattiToLevel génère deux fichiers de paramètres qui serviront d'entrée dans Levels :



2) Les coordonnées des positions des sondages dans Hydro_XY sont en mètres, alors que la grille d'échantillonnage que l'on vient de créer est définie en coordonnées écran (pixels). Il convient donc de ramener ces coordonnées à l'échelle physique du fond de carte Quro_Map par l'option ADSUUtil : XY->Digit³ :



Scaling actual coordinates into pixel coordinates
----Input background map: Quro_Map
----Input corner file: P1_Rect
Xmin: 0.0000e+00 Ymin: 0.0000e+00
Xmax: 2.5000e+02 Ymax: 2.5000e+02
----Input XY file in actual coordinates: Hydro_XY
It contains 24 points

----File Hydro_XY_Pt contains point coordinates in pixels
It has 24 rows and 2 columns

3) L'option MapUtil : 2D lowess error donne l'erreur d'estimation en fonction du nombre de voisins pris en compte dans la régression locale utilisée par le programme Levels :

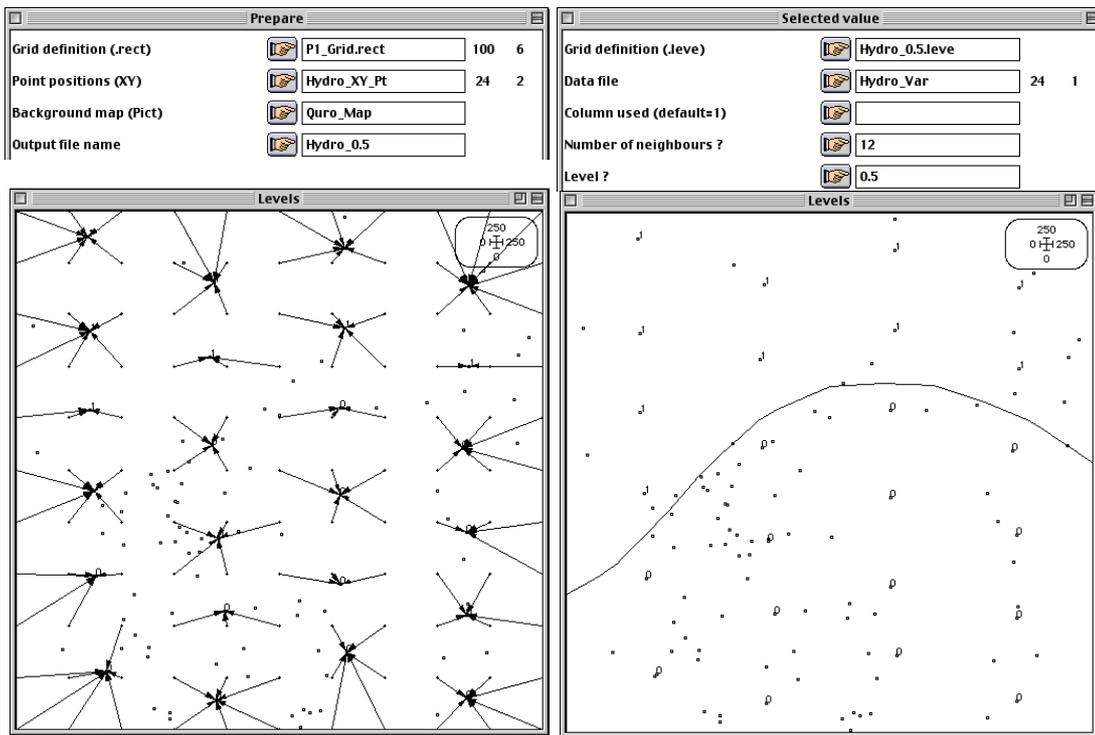


owesserror: total error of bidimensional local regression
Y coordinates file: Hydro_XY
-- Number of rows: 24, columns: 2
ata file: Hydro_Var
-- Number of rows: 24, columns: 1
umber of neighbors: 5a20
3.8030E-01
3.6480E-01
3.5391E-01
3.6447E-01
3.4709E-01
3.4438E-01
3.4482E-01
3.5154E-01
3.5712E-01
3.5538E-01
3.6132E-01
3.6042E-01
3.6029E-01
3.6420E-01
3.6612E-01
3.6740E-01
output file: HydroVar.2DError
-- Number of rows: 16, columns: 1
his file contains, for each variable (in columns),
nd for each number of neighbors (in rows) the sum
f estimation error for all sampling points

Comme les sondages pédologiques sont réalisés selon une grille systématique, chaque position a respectivement 2, 6, 8, 12, etc. voisins situés à la même distance. On choisit donc d'utiliser les 12 plus proches voisins pour prédire la variable pédologique.

³ Sous Windows, les fonds de carte sauvegardés à partir des modules graphiques sont au format .WMF, alors que les fonds de carte appelés doivent être au format .BMP. Il est donc nécessaire de coller les fonds de carte créés par les modules graphiques dans un logiciel de dessin (p.e. MSPaint) et de les enregistrer en bitmap.

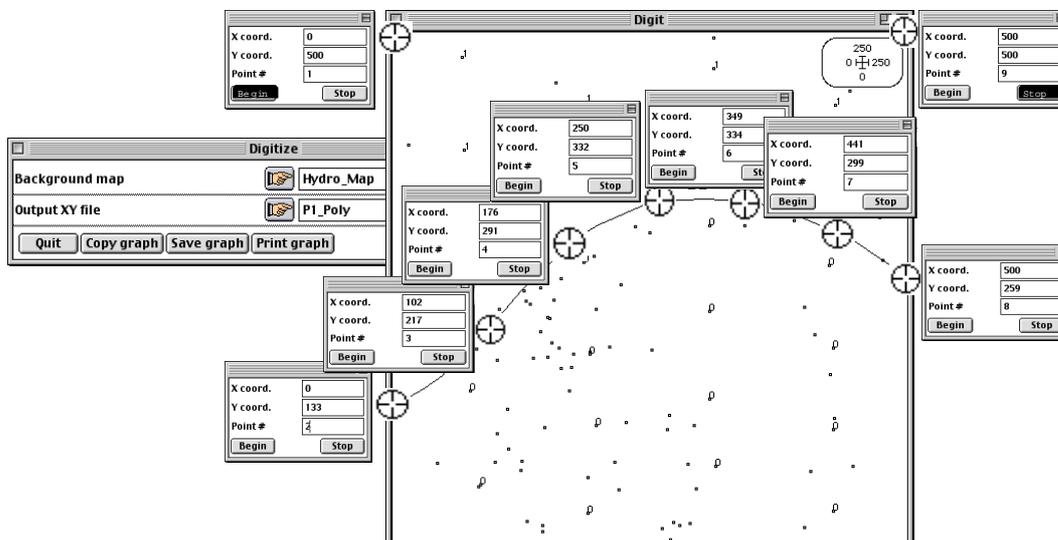
4) On dispose maintenant de tous les fichiers nécessaires au tracé de la ligne de contour 0,5, qu'on initialise par Levels : Prepare et qu'on trace définitivement par Levels : Selected value en utilisant les 12 plus proches voisins :



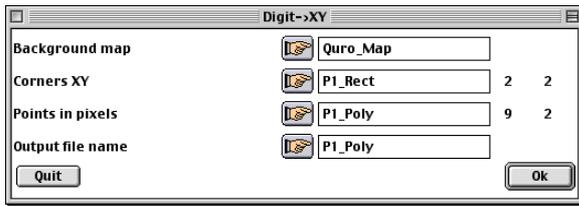
La figure obtenue est sauvegardée sous le nom Hydro_Map. Elle montre clairement la segmentation du domaine initial en deux sous-régions homogènes, l'une de très faible densité sur sol hydromorphe (en haut), l'autre de densité plus élevée sur sol non-hydromorphe (en bas).

La carte précédente va servir de référence pour définir un nouveau domaine d'étude, correspondant à la zone hors bas-fond sur laquelle on souhaite raffiner l'analyse. Ce nouveau domaine de forme complexe est approché par un polygone que l'on définit en excluant la zone de bas-fond du domaine rectangulaire initial (Goreaud & Pélissier 1999). Les étapes suivantes sont nécessaires :

1) Le module Digit : Digitize permet de saisir, à partir du fond de carte Hydro_Map, les coordonnées des sommets du polygone qui couvre la zone de bas-fond :



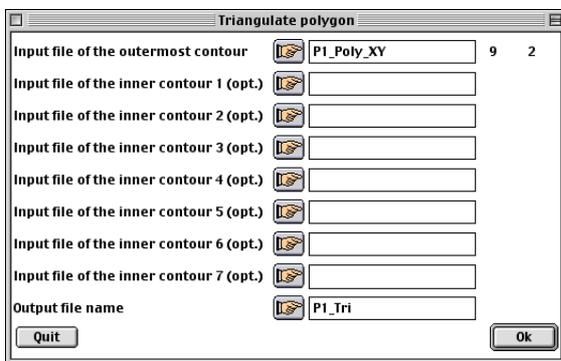
2) Les coordonnées des sommets dans P1_Poly sont en pixels. On opère la conversion en coordonnées métriques par ADSUtil : Digit->XY :



caling pixel coordinates into actual coordinates
 ---Input background map: Quro_Map
 ---Input corner file: P1_Rect
 min: 0.0000e+00 Y min: 0.0000e+00
 max: 2.5000e+02 Y max: 2.5000e+02
 ---Input XY file in pixels : P1_Poly
 t contains 9 points

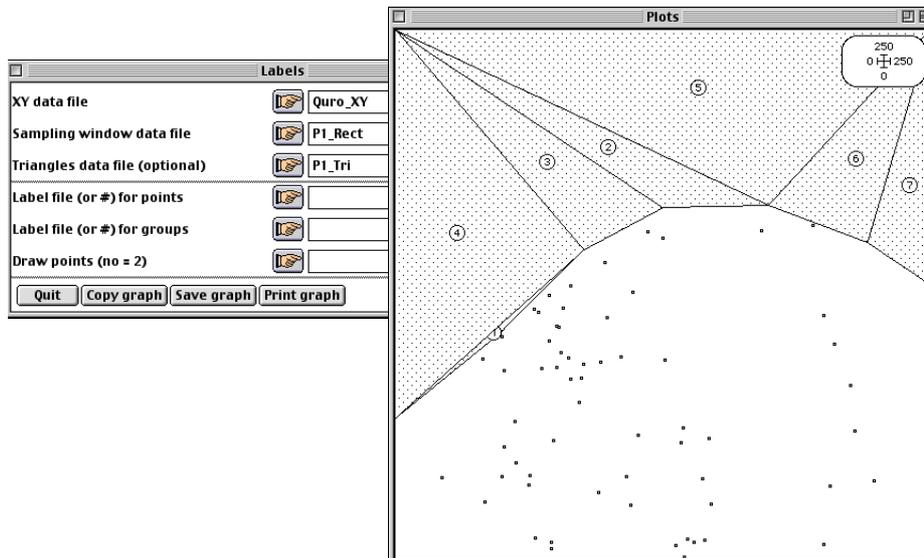
---File P1_Poly_XY contains actual point coordinates
 t has 9 rows and 2 columns

3) Le polygone défini par P1_Poly_XY est décomposé en triangles contigus par ADSUtil : Triangulate polygon :



ast polygon triangulation (Seidel 1991, Narkhede & Manocha 995)
 ---Input XY data file of the outermost contour: P1_Poly_XY
 t contains 9 points
 ---file P1_Tri contains coordinates of the triangle vertices
 t has 7 rows and 6 columns

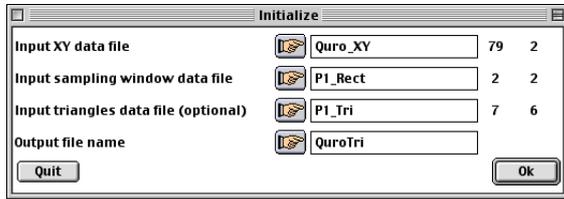
Le fichier P1_Tri contient les coordonnées des sommets des 7 triangles qui composent le polygone à exclure de l'analyse. L'option Plots : Labels permet de les représenter :



On obtient ainsi une zone homogène sur laquelle on peut effectuer une analyse de la structure spatiale avec les outils classiques.

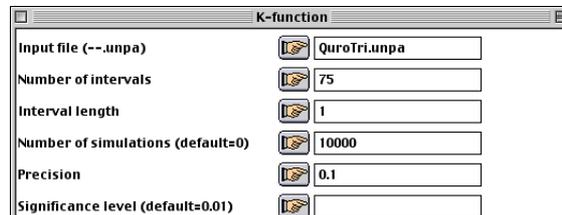
3.1.4 – Analyse de la structure spatiale sur sol non-hydromorphe.

On prépare l'analyse de la structure spatiale du semis hors bas-fond en liant les fichiers Quro_XY, P1_Rect et P1_Tri par l'option Ripley : Initialize :



```
-----
new TEXT file QuroTri.unpa contains the parameters:
---> XY data file: Quro_XY [79][2]
---> Shape of the sampling window: 3
      1 = rectangular
      2 = circular
      3 = irregular within a rectangular window
      4 = irregular within a circular window
---> Sampling window data file: P1_Rect [2][2]
---> Triangle data file: P1_Tri [7][6]
-----
```

Le listing signale la création du fichier de paramètres QuroTri.unpa qui servira d'entrée pour l'option Ripley : K-fonction :



Le programme signale que les points qui se trouvent à l'intérieur des triangles sont éliminés :

```
-----
nivariate second-order neighbourhood functions (Ripley 1977)
---XY data file: Quro_XY
t contains 79 points
---Sampling window data file: P1.Rect
min: 0.0000e+00 Ymin: 0.0000e+00
max: 2.5000e+02 Ymax: 2.5000e+02
umber of points within the rectangular sampling window: 79
---Input triangle data file: P1_Tri
t contains 7 triangles
umber of triangles within the rectangular sampling window: 7
umber of points within the triangles: 17
---Input parameters
rea of the study region: 3.4767e+04
riangles within the study region: 7
umber of points within the study region: 62 (density = 1.7833e-03)
max = 1.2500e+02
r = 1.0000e+00
umber of Monte Carlo simulations = 10000
recision of the simulated coordinates = 1.0000e-01
```

```
---Local significance level of the null hypothesis
-values of |Obs| <= |Simu|
```

r	g(r)	n(r)	K(r)	L(r)
1.0000	9.999e-01	9.999e-01	9.999e-01	9.999e-01
2.0000	7.130e-01	6.736e-01	6.736e-01	6.736e-01
...				
74.0000	6.982e-01	5.599e-03	5.599e-03	5.599e-03
75.0000	9.430e-01	5.799e-03	5.799e-03	5.799e-03

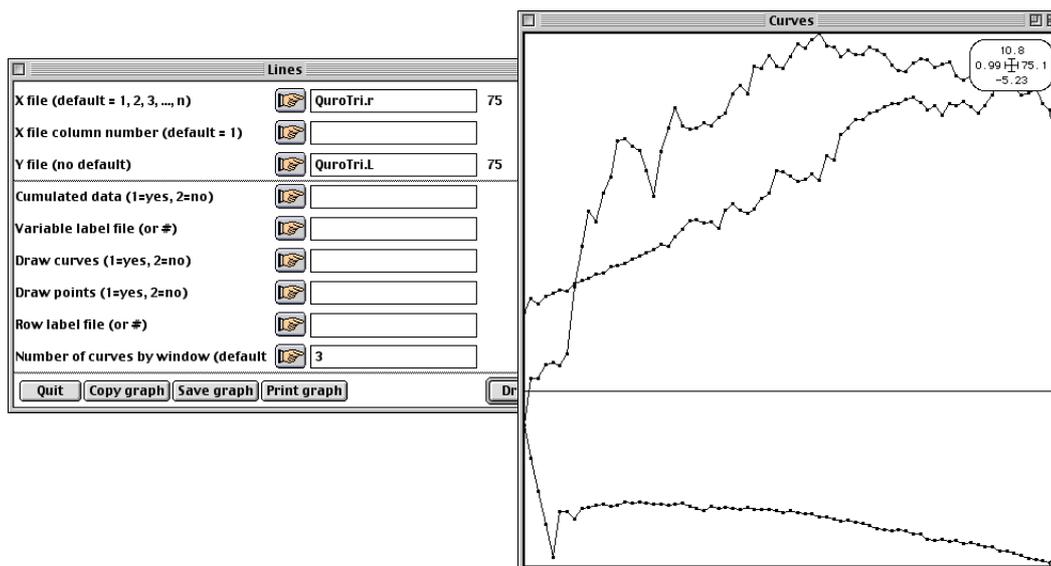
Les fichiers de résultats contiennent les valeurs des différents estimateurs et les bornes de l'intervalle de confiance à 1% :

```

----File QuroTri.r contains the distance steps
It has 75 rows and 1 column
----File QuroTri.g contains the pair density function
It has 75 rows and 3 columns
Col 1: g(r)
Col 2,3: 99% confidence interval
----File QuroTri.n contains the local neighbour density function
It has 75 rows and 3 columns
Col 1: n(r)
Col 2,3: 99% confidence interval
----File QuroTri.K contains the Ripley's function
It has 75 rows and 3 columns
Col 1: K(r)
Col 2,3: 99% confidence interval
----File QuroTri.L contains the modified Ripley's function
It has 75 rows and 3 columns
Col 1: L(r)
Col 2,3: 99% confidence interval
-----

```

Les résultats peuvent être représentés par l'option Curves : Lines en groupant la fonction estimée (1) et les bornes de l'intervalle de confiance (2 et 3) :



La courbe de la fonction $\hat{L}(r)$ revient dans l'intervalle de confiance aux grandes distances, ce qui confirme qu'à cette échelle, le semis de points peut être considéré comme homogène et donc sa structure interprétée correctement. Elle révèle l'existence d'agrégats significatifs de taille variant de 10 à 60 m de rayon, pour l'espèce *Qualea rosea* sur les sols non-hydromorphes. Ce type de structure est lié d'une part aux exigences édaphiques de l'espèce et d'autre part à une dissémination anémochore d'un rayon limité autour des pieds mères (Collinet 1997).

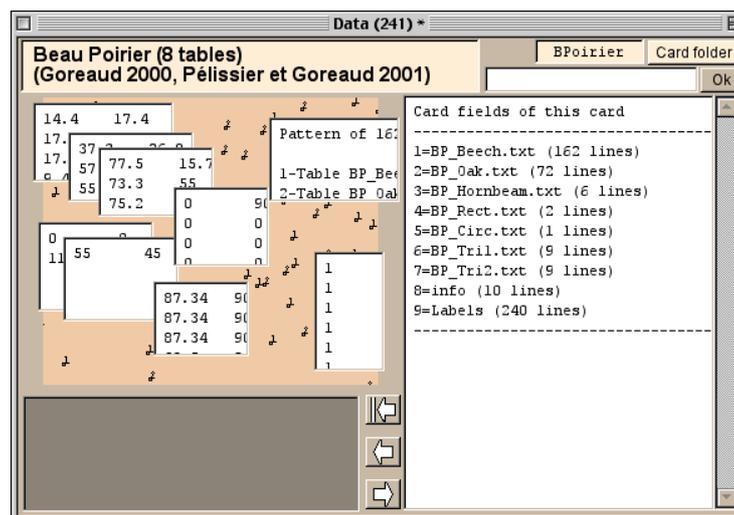
3.2 – Exemple 2 : Dispositif du Beau Poirier (Nancy, France).

Dans certains cas, un peuplement forestier peut présenter une hétérogénéité de structure sans que les facteurs à l'origine de cette hétérogénéité soient connus ou mesurés. La méthode présentée dans l'exemple précédent ne peut alors être appliquée directement. Nous proposons alors d'utiliser les propriétés locales du semis (par exemple, la densité, la hauteur moyenne ou la composition spécifique) pour définir des zones homogènes. Ainsi, si on note $f(x,y)$ une telle propriété du semis à la position (x,y) , une carte des valeurs de $f(x,y)$ peut jouer le même rôle que la carte d'un facteur exogène pour définir des zones homogènes (Pélissier & Goreaud 2001).

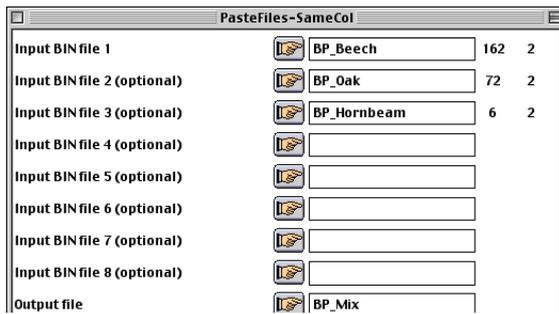
Cet exemple reprend le cas de la placette du Beau Poirier étudiée par Goreaud (2000). Cette placette de 1 ha en forêt de Haye, France, abrite un peuplement mixte de hêtres (*Fagus sylvatica*), chênes (*Quercus petraea*) et charmes (*Carpinus betullus*). Elle a été installée en 1882, est a été soumise à des interventions sylvicoles dont le but est de préserver, par des éclaircies sélectives, le chêne qui tend à être naturellement éliminé par le hêtre. Cet exemple montre que le peuplement actuel présente une hétérogénéité de composition et de structure qui empêche d'interpréter correctement la structure spatiale du semis sur l'ensemble de la parcelle à partir des fonctions de Ripley. Cette hétérogénéité, probablement liée à l'action conjuguée de facteurs pédologiques et des interventions sylvicoles, mais pour lesquels on ne dispose d'aucune donnée quantifiée, peut-être appréhendée à partir des propriétés locales du peuplement. On montre en particulier comment la densité locale permet de définir des sous-régions homogènes, à l'intérieur desquelles la structure spatiale peut être analysée correctement par les fonctions de Ripley.

3.2.1 – Les données de la placette du Beau Poirier.

Les données de cette placette sont consignées dans la carte BPOirier de la pile de données. Les fichiers BP_Beech, BP_Oak et BP_Hornbeam contiennent respectivement les coordonnées de 162 hêtres, 72 chênes et 6 charmes. Le fichier BP_Rect contient les coordonnées de l'origine et du coin opposé du domaine d'étude rectangulaire :

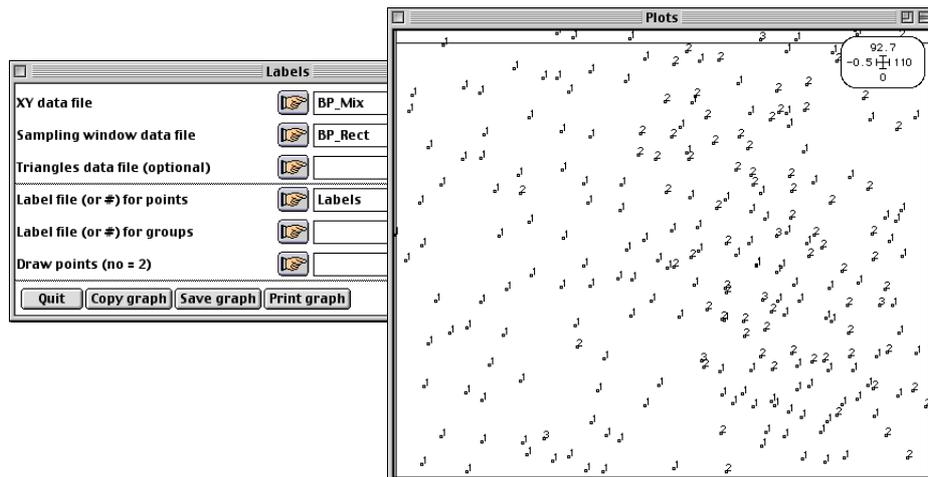


Comme on s'intéresse ici à la structure du mélange, on opère une concaténation des fichiers des trois espèces par FilesUtil : PasteFiles-SameCol :



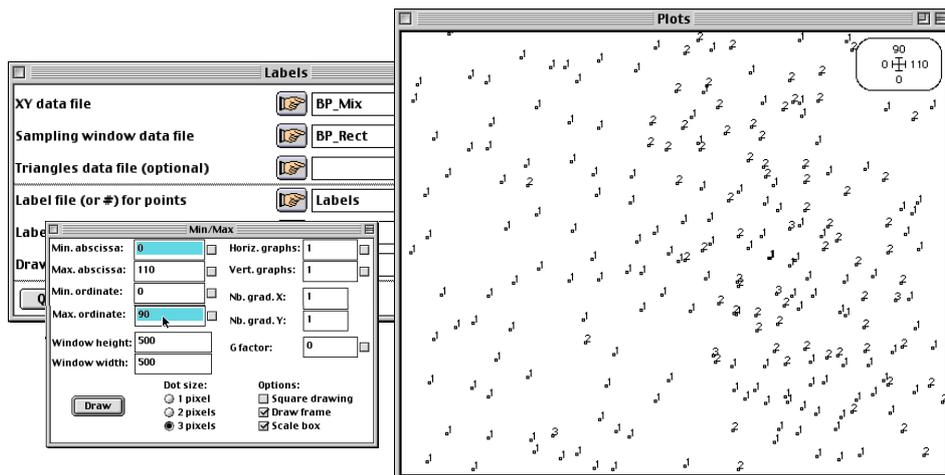
 astelig : Paste files with same column number
 nput file: BP_Beech
 -- Number of rows: 162, columns: 2
 nput file: BP_Oak
 -- Number of rows: 72, columns: 2
 nput file: BP_Hornbeam
 -- Number of rows: 6, columns: 2
 utput file: BP_Mix
 -- Number of rows: 240, columns: 2

Le fichier BP_Mix contient les coordonnées des 240 points du semis mélangé que l'on peut représenter par Plots : Labels (1=hêtre; 2=chêne; 3=charme) :



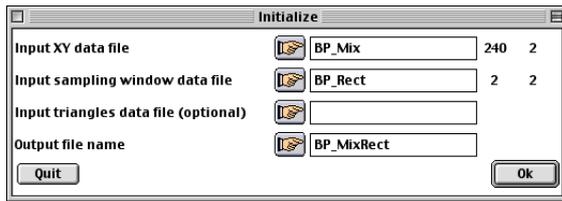
On notera que quelques points sont situés à l'extérieur du domaine d'étude rectangulaire.

On réduit les dimensions de l'image dans la fenêtre Min. & Max. afin de ne pas représenter les points situés à l'extérieur de la parcelle et on sauvegarde cette nouvelle figure sous le nom BP_Map :



3.2.2 – Analyse de la structure spatiale sur l'ensemble de la parcelle.

On prépare l'analyse de la structure spatiale du semis en liant les fichiers BP_Mix et BP_Rect par Ripley : Initialize :



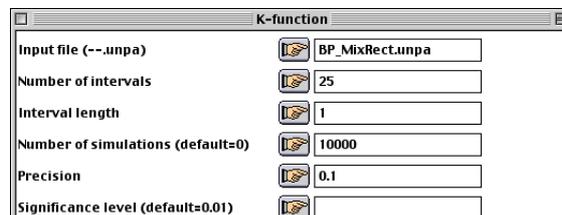
```

-----
ew TEXT file BP_MixRect.unpa contains the parameters:
--> XY data file: BP_Mix [240][2]
--> Shape of the sampling window: 1
    1 = rectangular
    2 = circular
    3 = irregular within a rectangular window
    4 = irregular within a circular window
--> Sampling window data file: BP_Rect [2][2]
-----

```

Le listing signale la création du fichier de paramètres BP_MixRect.unpa qui servira d'entrée pour les options de calcul.

L'analyse de la structure spatiale par l'option Ripley : K-function est réalisée sur 25 pas de distance de 1 m. La précision des coordonnées simulées pour le calcul de l'intervalle de confiance local est fixée à 0,1 m et le nombre de simulations de Monte Carlo à 10000 pour un risque de première espèce $\alpha = 0,01$:



Le listing renvoie une description des fichiers de données définis par BP_MixRect.unpa et indique que les points situés à l'extérieur du domaine d'étude rectangulaire ne sont pas pris en compte dans l'analyse :

```

-----
nivariate second-order neighbourhood functions (Ripley 1977)
--XY data file: BP_Mix
t contains 240 points
---Sampling window data file: BP_Rect
min: 0.0000e+00 Ymin: 0.0000e+00
max: 1.1000e+02 Ymax: 9.0000e+01
umber of points within the rectangular sampling window: 231
---Input parameters
rea of the study region: 9.9000e+03
umber of points within the study region: 231 (density = 2.3333e-02)
max = 2.5000e+01
r = 1.0000e+00
umber of Monte Carlo simulations = 10000
recision of the simulated coordinates = 1.0000e-01
-----

```

Le tableau suivant donne, pour chaque pas de distance r , le seuil de signification du test local de l'hypothèse nulle d'une répartition aléatoire des points du semis :

```

---Local significance level of the null hypothesis
-values of |Obs| <= |Simu|
-----
  r | g(r) | n(r) | K(r) | L(r) |
-----
1.0000 | 3.320e-02 | 2.460e-02 | 2.460e-02 | 1.240e-02 |
2.0000 | 5.799e-03 | 8.999e-04 | 7.999e-04 | 0.000e+00 |

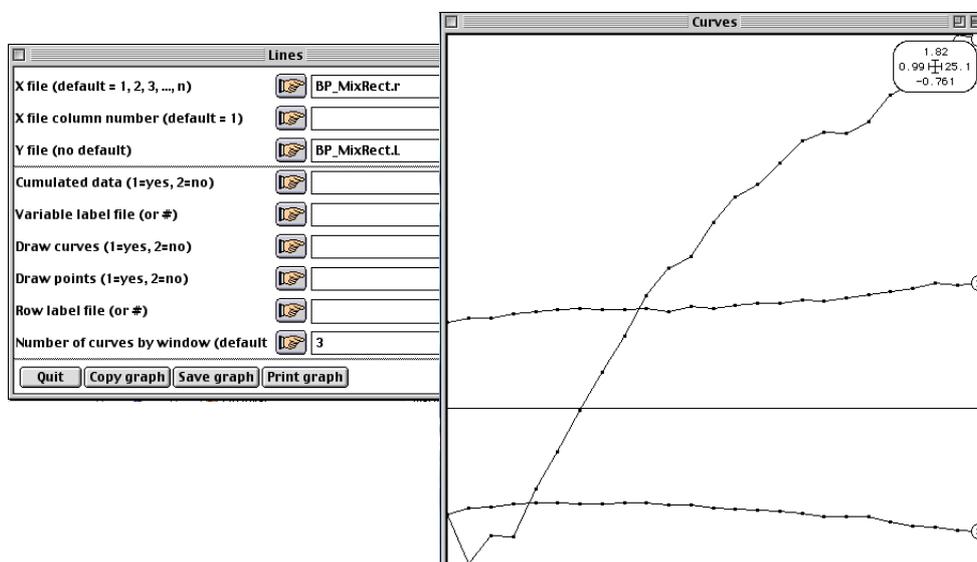
24.0000 | 9.999e-05 | 0.000e+00 | 0.000e+00 | 0.000e+00 |
25.0000 | 4.120e-01 | 0.000e+00 | 0.000e+00 | 0.000e+00 |
-----

```

Les fichiers de résultats contiennent les valeurs des différents estimateurs et les bornes de l'intervalle de confiance local à 1% :

----File BP_MixRect.r contains the distance steps
 It has 25 rows and 1 column
 ----File BP_MixRect.g contains the pair density function
 It has 25 rows and 3 columns
 Col 1: $g(r)$
 Col 2,3: 99% confidence interval
 ----File BP_MixRect.n contains the local neighbour density function
 It has 25 rows and 3 columns
 Col 1: $n(r)$
 Col 2,3: 99% confidence interval
 ----File BP_MixRect.K contains the Ripley's function
 It has 25 rows and 3 columns
 Col 1: $K(r)$
 Col 2,3: 99% confidence interval
 ----File BP_MixRect.L contains the modified Ripley's function
 It has 25 rows and 3 columns
 Col 1: $L(r)$
 Col 2,3: 99% confidence interval

La fonction $L(r)$ peut être représentée graphiquement par l'option Curves : Lines en groupant la fonction estimée pour le semis observé (1) et les bornes de l'intervalle de confiance (2 et 3) :



La courbe de $\hat{L}(r)$ montre une divergence aux grandes distances que l'on interprète comme le signe d'une hétérogénéité liée à la différence de densité entre les deux moitiés de la parcelle. A cause de cette hétérogénéité, on ne peut pas utiliser ce premier calcul pour décrire la structure spatiale fine du semis (Pélissier & Goreaud 2001).

3.2.3 – Analyse des propriétés endogènes du semis.

L'hétérogénéité de la densité observée à l'échelle de la parcelle incite à utiliser la variable densité locale pour définir des sous-régions homogènes. L'option Density : 1st-order density permet de calculer la densité locale aux nœuds d'une grille systématique de 10 x 10 m. On effectue le calcul sur 25 pas de distance de 0,5 m :



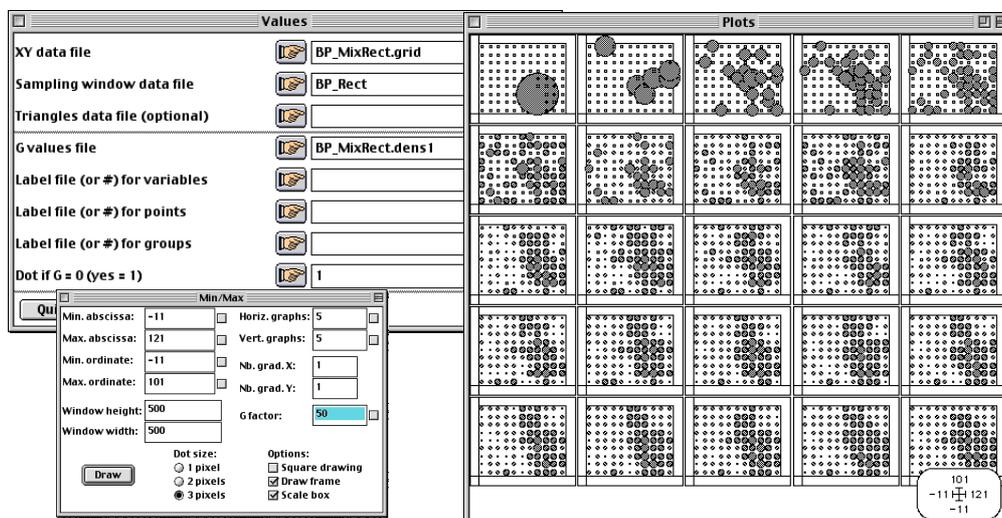
Le listing renvoie les caractéristiques du semis et indique que les points situés à l'extérieur du domaine d'étude rectangulaire ne sont pas pris en compte :

```
-----
first-order local density functions (Pelissier & Goreaud 2001)
---XY data file: BP_Mix
t contains 240 points
---Sampling window data file: BP_Rect
min: 0.0000e+00 Y min: 0.0000e+00
max: 1.1000e+02 Y max: 9.0000e+01
umber of points within the rectangular sampling window: 231
---Input parameters
rea of the study region: 9.9000e+03
umber of points within the study region: 231 (density = 2.3333e-02)
nter-node distance on X: 1.0000e+01
on Y: 1.0000e+01
umber of sampling points within the study region: 99
max= 1.2500e+01
r= 5.0000e-01
```

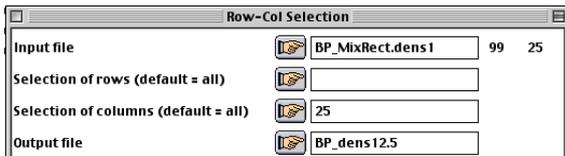
Le fichier BP_MixRect.grid contient les coordonnées des positions d'échantillonnage, le fichier BP_MixRect.dens1 contient les valeurs de la densité locale, estimées à chaque position d'échantillonnage (en lignes) et aux différents pas de distance (en colonnes) :

```
-----
---File BP_MixRect.grid contains the coordinates of the sampling points A
t has 99 rows and 2 columns
---File BP_MixRect.count1 contains the counting function NA(r)
t has 99 rows and 25 columns
---File BP_MixRect.dens1 contains the local density function nA(r)
t has 99 rows and 25 columns
-----
```

L'option Plots : Values permet d'obtenir une carte de ces valeurs. Pour une meilleure lisibilité de la figure, le facteur G est réglé à 50 dans la fenêtre Min. & Max. :

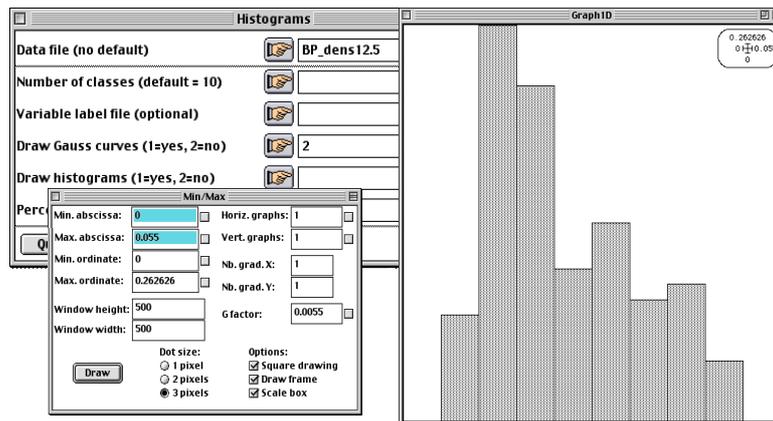


La figure fait apparaître une zone de plus forte densité dont les contours paraissent assez stables au-delà de $r = 5$ m (fenêtre 10). Compte tenu de l'échelle d'observation définie par la taille du domaine d'étude (50 m) et de l'échelle présumée des structures intéressantes (< 20 m), on choisit d'utiliser la carte des valeurs de la densité locale à $r = 12,5$ m (échelle = 25 m) pour délimiter des sous-régions homogènes en densité. Pour des raisons pratiques, on extrait la colonne 25 du fichier BP_MixRect.dens1 par FilesUtil : Row-Col selection :



 selectrc: Selection of rows & columns in a file
 Input file: BP_MixRect.dens1
 --- Number of rows: 99, columns: 25
 Output file: BP_dens12.5
 --- Number of rows: 99, columns: 1

L'histogramme des valeurs de la densité locale à $r = 12,5$ m pour les 99 positions d'échantillonnage est réalisé par l'option Graph1D : Histograms. Les paramètres de la représentation sont réglés de manière à obtenir des classes de densité de largeur 0,005 :



La distribution présente une forme grossièrement bimodale qui détermine deux classes de densité $[0; 0.025]$ et $[0.025; 0.055]$.

3.2.4 – Définition des sous-régions homogènes.

L'analyse la densité locale calculée à $r = 12,5$ m conduit à utiliser la ligne de contour 0,025 pour délimiter des sous-régions homogènes en densité à l'intérieur du domaine rectangulaire initial. On se retrouve donc, mais avec une variable continue, dans une situation analogue à celle du paragraphe 3.1.3, dont on pourra suivre les différentes étapes pour tracer cette ligne de contour par Levels :

1) Préparer, à partir du fond de carte BP_Map, une grille d'échantillonnage de 10 x 10 m (9 cellules en ligne et 11 cellules en colonne) en utilisant les options Lattices : Create Bkgnd (création des fichiers BP_Grid et BP_Grid.lat) et Lattices : LattiToLevel (création des fichiers BP_Grid.rect et BP_Grid.summ).

2) Ramener les coordonnées des positions d'échantillonnage de la densité locale du fichier BP_MixRect.grid à l'échelle écran de BP_Map en utilisant l'option ADSUtil : XY->Digit (création du fichier BP_MixRect.grid_Pt).

3) Calculer l'erreur d'estimation en fonction du nombre de voisins pris en compte dans la régression locale par l'option MapUtil : 2D lowess error :



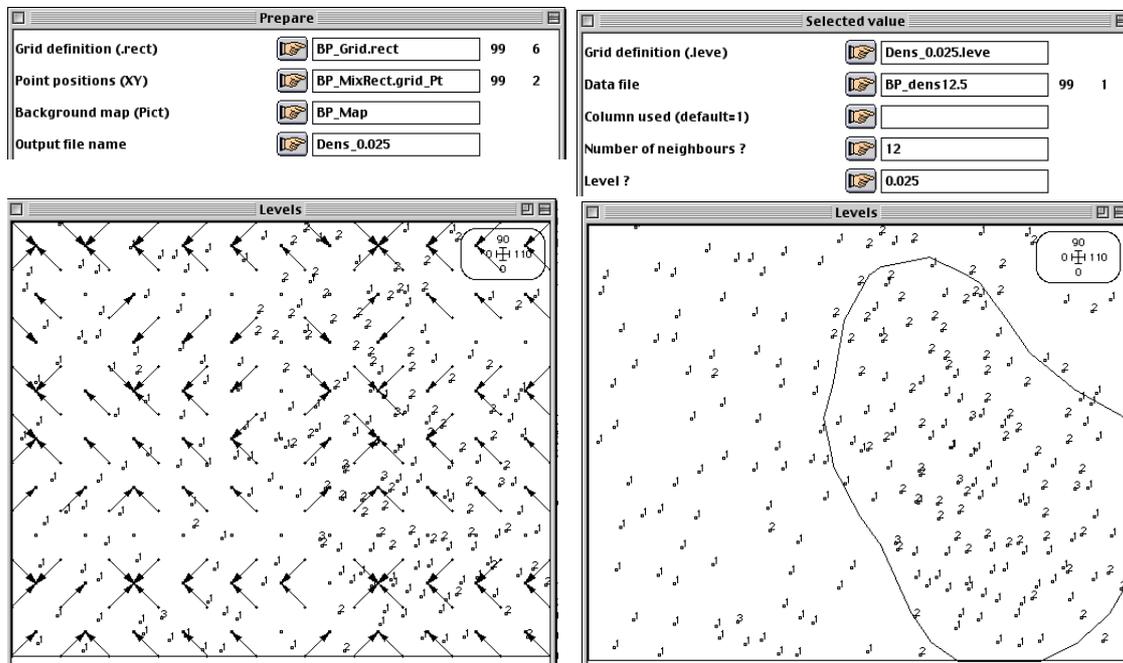
```

-----
owesserror: total error of bidimensional local regression      1.8387E-01
Y coordinates file: BP_MixRect.grid                        1.8759E-01
-- Number of rows: 99, columns: 2                         1.8797E-01
ata file: BP_dens12.5                                    1.8656E-01
-- Number of rows: 99, columns: 1                         1.9935E-01
umber of neighbors: 5a20                                  1.9852E-01
1.9685E-01                                                2.0334E-01
1.9685E-01                                                2.1775E-01
1.8883E-01
1.8937E-01
1.8937E-01
1.8550E-01
1.8534E-01
1.8452E-01
-----
output file: DensVar.2DError
-- Number of rows: 16, columns: 1
his file contains, for each variable (in columns),
nd for each number of neighbors (in rows) the sum
f estimation error for all sampling points
-----

```

Comme les valeurs de la densité locale sont estimées sur une grille systématique, chaque position d'échantillonnage a respectivement 4, 8, 12, 20, etc. voisins situés à la même distance. On choisit donc d'utiliser les 12 plus proches voisins pour prédire la densité locale.

4) Tracer la ligne de contour 0,025 en utilisant les options Levels : Prepare et Levels : Selected value :



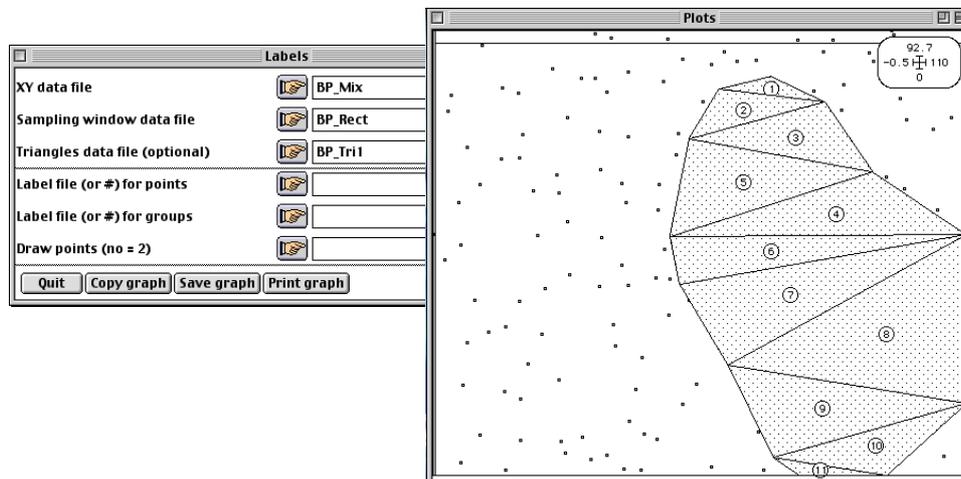
La figure ainsi obtenue montre que l'hétérogénéité de la densité correspond assez précisément à l'hétérogénéité du mélange spécifique : la parcelle est divisée en une zone 1 de forte densité abritant un mélange de hêtres (espèce 1) et chênes (espèce 2) ; et une zone 2 de faible densité, largement dominée par le hêtre. Goreaud (2000) montre que des valeurs locales de la hauteur ou de la circonférence des hêtres présentent une hétérogénéité similaire, ce qui lui permet de conclure à une différence de fertilité locale du sol entre les deux zones.

Le nouveau fond de carte, sauvegardé sous le nom Dens_Map, va servir de référence pour définir deux nouveaux domaines d'étude, l'un correspondant à la zone 1 de forte densité, l'autre à la zone 2 de faible densité. Ces deux domaines, de formes complexes, sont approchés par des polygones que l'on définit par exclusion de triangles du domaine rectangulaire initial (Goreaud & Pélissier 1999). On pourra encore une fois suivre les étapes détaillées au paragraphe 3.1.3 pour :

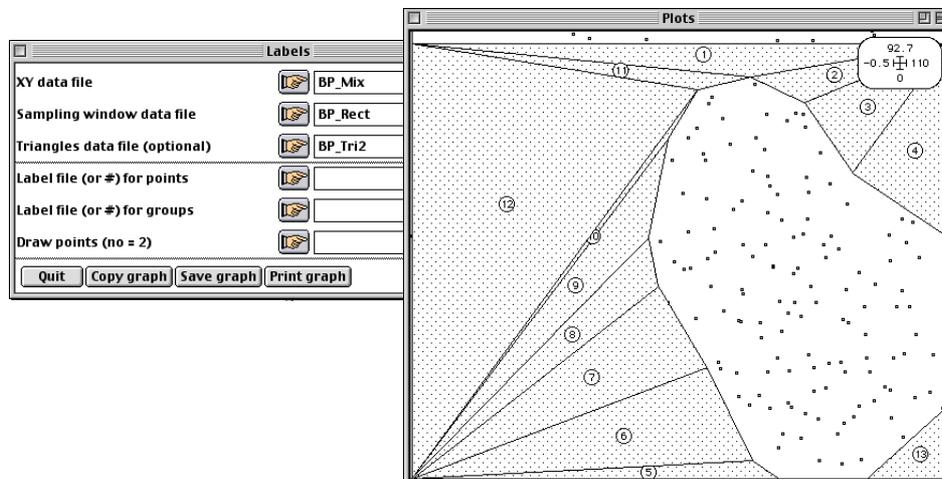
1) Saisir les coordonnées des sommets du polygone qui couvre la zone 1 par Digit : Digitize (création du fichier BP_Poly1).

2) Convertir les coordonnées écran de BP_Poly1 en coordonnées métriques par ADSUtil : Digit->XY (création du fichier BP_Poly1_XY).

3) Décomposer le polygone défini par BP_Poly1_XY en triangles contigus par ADSUtil : Triangulate polygon (création du fichier BP_Tri1). L'option Plots : Labels permet de les représenter :



On procède de manière analogue pour créer deux fichiers distincts correspondant aux deux polygones qui couvrent la zone 2. On opère ensuite une concaténation de ces fichiers en un seul par FilesUtil : PasteFiles-SameCol (création du fichier BP_Tri2) contenant 13 triangles que l'on représente par Plots : Labels :

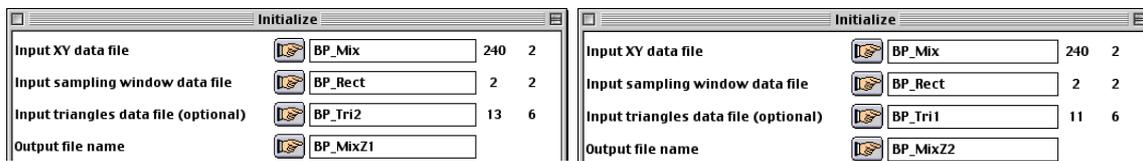


On a ainsi délimité deux sous-régions, homogènes en densité, à l'intérieur desquelles on pourra alternativement analyser la structure spatiale à l'aide des fonctions de Ripley.

Dans cet exemple on effectue simplement un partage de l'espace en deux zones homogènes complémentaires. Dans certains cas on peut toutefois observer un gradient de transition entre deux zones de densité différentes. Pélissier & Goreaud (2001) proposent alors de définir une zone tampon entre ces zones.

3.2.5 – Analyse de la structure spatiale dans les sous-régions homogènes.

On prépare l'analyse de la zone 1 en liant les fichiers BP_Mix, BP_Rect, BP_Tri1 et l'analyse de la zone 2 en liant les fichiers BP_Mix, BP_Rect, BP_Tri2 par Ripley : Initialize :



```

-----
ew TEXT file BP_MixZ1.unpa contains the parameters:
--> XY data file: BP_Mix [240][2]
--> Shape of the sampling window: 3
    1 = rectangular
    2 = circular
    3 = irregular within a rectangular window
    4 = irregular within a circular window
--> Sampling window data file: BP_Rect [2][2]
--> Triangle data file: BP_Tri2 [13][6]
-----

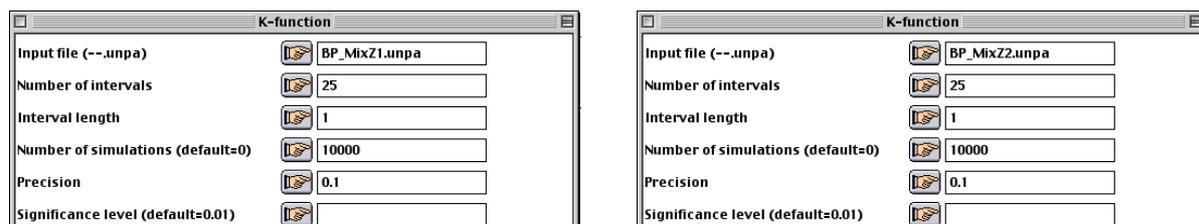
```

```

-----
ew TEXT file BP_MixZ2.unpa contains the parameters:
--> XY data file: BP_Mix [240][2]
--> Shape of the sampling window: 3
    1 = rectangular
    2 = circular
    3 = irregular within a rectangular window
    4 = irregular within a circular window
--> Sampling window data file: BP_Rect [2][2]
--> Triangle data file: BP_Tri1 [11][6]
-----

```

Les fichiers BP_MixZ1.unpa et BP_MixZ2.unpa servent d'entrée pour l'option de calcul Ripley : K-function, qu'on lance en utilisant les mêmes paramètres que pour l'analyse préliminaire :



Les listings signalent que les points situés à l'extérieur du domaine rectangulaire initial et ceux situés à l'intérieur des triangles ne sont pas pris en compte dans les analyses :

```

-----
Univariate second-order neighbourhood functions (Ripley 1977)
---XY data file: BP_Mix
It contains 240 points
---Sampling window data file: BP_Rect
Xmin: 0.0000e+00 Ymin: 0.0000e+00
Xmax: 1.1000e+02 Ymax: 9.0000e+01
Number of points within the rectangular sampling window: 231
---Input triangle data file: BP_Tri2
It contains 13 triangles
Number of triangles within the rectangular sampling window: 13
Number of points within the triangles: 99
---Input parameters
Area of the study region: 3.6190e+03
Triangles within the study region: 13
Number of points within the study region: 132
(density = 3.6474e-02)
rmax = 2.5000e+01
dr = 1.0000e+00
Number of Monte Carlo simulations = 10000
Precision of the simulated coordinates = 1.0000e-01
-----

```

```

-----
Univariate second-order neighbourhood functions (Ripley 1977)
---XY data file: BP_Mix
It contains 240 points
---Sampling window data file: BP_Rect
Xmin: 0.0000e+00 Ymin: 0.0000e+00
Xmax: 1.1000e+02 Ymax: 9.0000e+01
Number of points within the rectangular sampling window: 231
---Input triangle data file: BP_Tri1
It contains 11 triangles
Number of triangles within the rectangular sampling window: 11
Number of points within the triangles: 132
---Input parameters
Area of the study region: 6.2810e+03
Triangles within the study region: 11
Number of points within the study region: 99
(density = 1.5762e-02)
rmax = 2.5000e+01
dr = 1.0000e+00
Number of Monte Carlo simulations = 10000
Precision of the simulated coordinates = 1.0000e-01
-----

```

----Local significance level of the null hypothesis
P-values of $|Obs| \leq |Simu|$

r	g(r)	n(r)	K(r)	L(r)
1.0000	4.720e-02	4.720e-02	2.970e-02	7.299e-03
2.0000	1.760e-02	3.200e-03	3.200e-03	4.000e-04
...
24.0000	8.439e-02	2.296e-01	2.296e-01	2.343e-01
25.0000	7.818e-01	2.905e-01	2.905e-01	2.956e-01

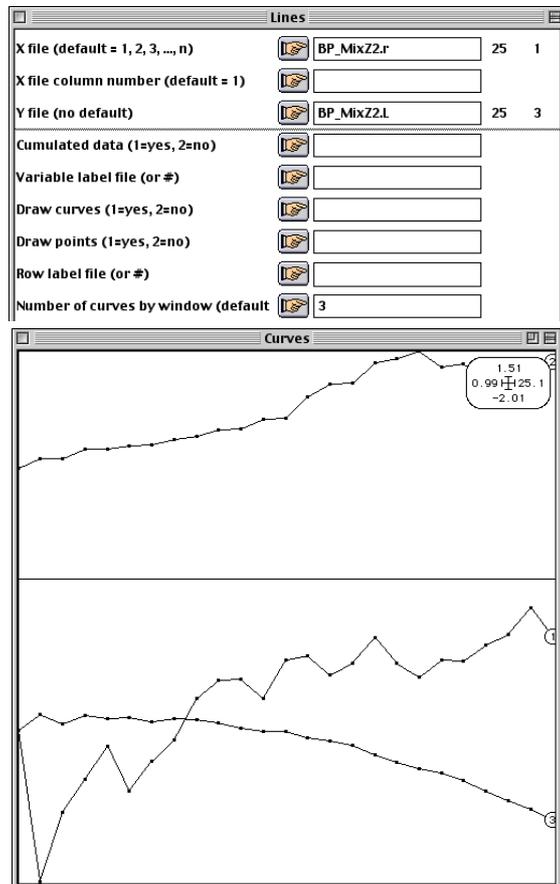
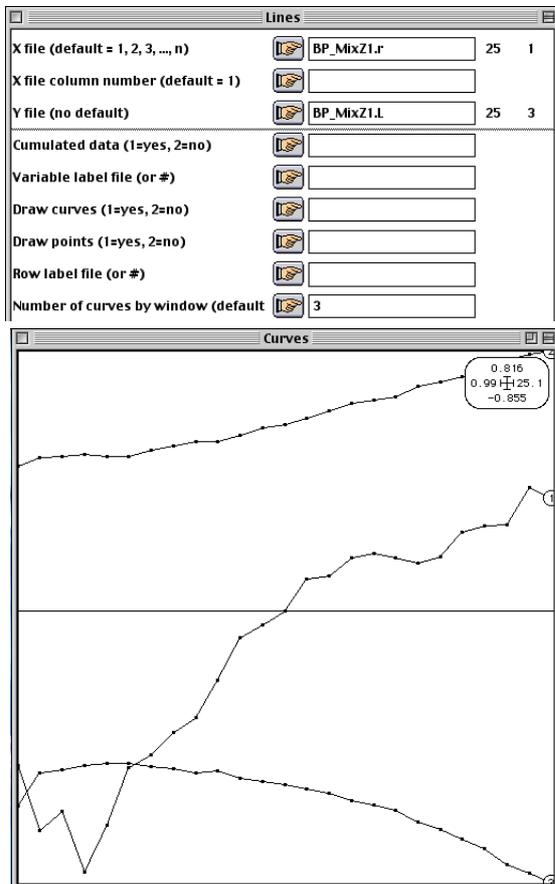
----File BP_MixZ1.r contains the distance steps
It has 25 rows and 1 column
----File BP_MixZ1.g contains the pair density function
It has 25 rows and 3 columns
Col 1: g(r)
Col 2,3: 99% confidence interval
----File BP_MixZ1.n contains the local neighbour density function
It has 25 rows and 3 columns
Col 1: n(r)
Col 2,3: 99% confidence interval
----File BP_MixZ1.K contains the Ripley's function
It has 25 rows and 3 columns
Col 1: K(r)
Col 2,3: 99% confidence interval
----File BP_MixZ1.L contains the modified Ripley's function
It has 25 rows and 3 columns
Col 1: L(r)
Col 2,3: 99% confidence interval

----Local significance level of the null hypothesis
P-values of $|Obs| \leq |Simu|$

r	g(r)	n(r)	K(r)	L(r)
1.0000	1.803e-01	1.803e-01	1.803e-01	9.239e-02
2.0000	9.699e-03	3.600e-03	3.600e-03	0.000e+00
...
24.0000	2.908e-01	7.641e-01	7.641e-01	7.629e-01
25.0000	1.929e-01	5.491e-01	5.491e-01	5.479e-01

----File BP_MixZ2.r contains the distance steps
It has 25 rows and 1 column
----File BP_MixZ2.g contains the pair density function
It has 25 rows and 3 columns
Col 1: g(r)
Col 2,3: 99% confidence interval
----File BP_MixZ2.n contains the local neighbour density function
It has 25 rows and 3 columns
Col 1: n(r)
Col 2,3: 99% confidence interval
----File BP_MixZ2.K contains the Ripley's function
It has 25 rows and 3 columns
Col 1: K(r)
Col 2,3: 99% confidence interval
----File BP_MixZ2.L contains the modified Ripley's function
It has 25 rows and 3 columns
Col 1: L(r)
Col 2,3: 99% confidence interval

On peut représenter les courbes de $L(r)$ par **Curves : Lines** en groupant la fonction estimée sur le semis réel (1) et les bornes de l'intervalle de confiance (2 et 3) :



Les courbes reviennent dans l'intervalle de confiance à 1% aux grandes distances, ce qui confirme que les deux sous-régions peuvent être considérées comme homogènes (Pélissier & Goreaud 2001). Les analyses de la structure spatiale mettent en évidence une régularité des deux semis, mais avec une intensité moindre dans la zone 1. Goreaud (2000) interprète cette différence par l'intervention du sylviculteur qui cherche généralement à favoriser le chêne face au hêtre. Dans la zone 2, les conditions de fertilité très favorables au hêtre, auraient conduit le forestier à pratiquer de fortes éclaircies pour tenter de conserver le chêne, créant ainsi une structure plus régulière que dans la zone 1, où les éclaircies auraient été plus modérées.

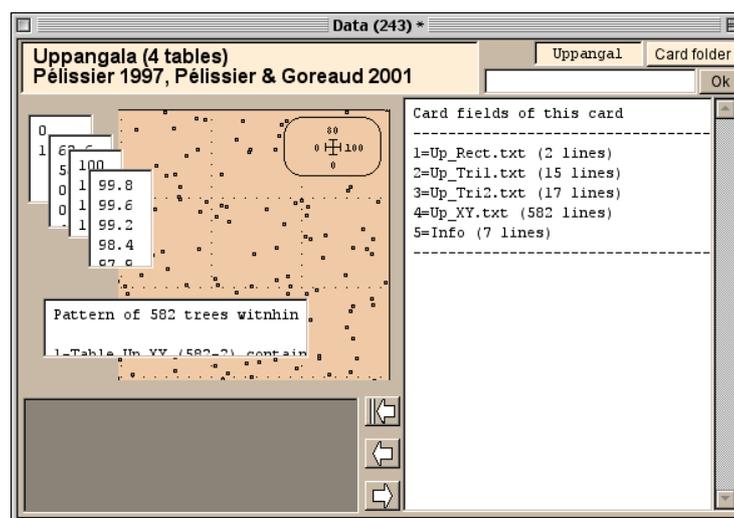
3.3 – Exemple 3 : Dispositif d'Uppangala (Inde).

Dans certains cas, bien que les propriétés de premier ordre du semis de points soient homogènes sur le domaine d'étude, la structure fine du semis peut apparaître hétérogène (Getis & Franklin 1987) et conduire à une interprétation incorrecte de la structure spatiale (Pélissier & Goreaud 2001).

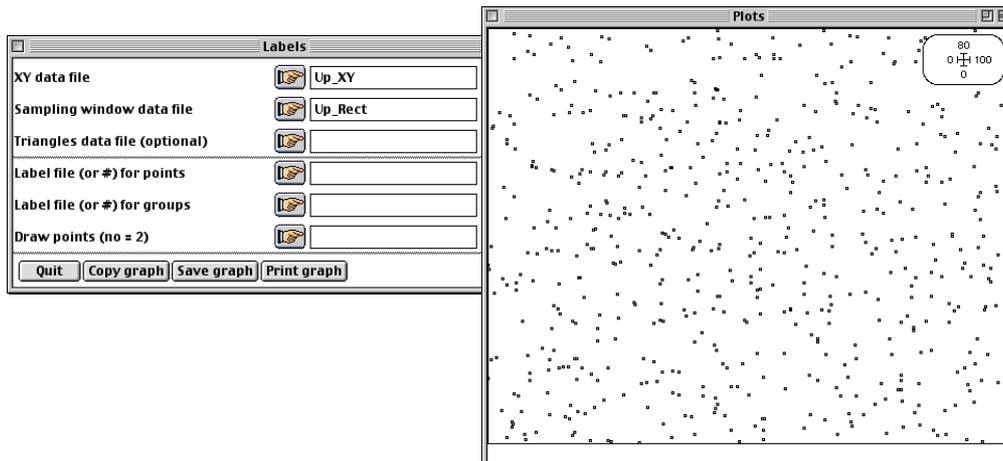
Ce chapitre reprend l'exemple d'une parcelle de 0.8 ha étudiée par Pélissier (1997) et abritant un peuplement plurispécifique de forêt dense humide sempervirente de la station d'Uppangala dans les Ghâts occidentaux de l'Inde. Cet exemple montre comment, malgré une homogénéité de la densité locale (de premier ordre), l'examen des variations de la densité locale de voisinage de second ordre révèle une hétérogénéité géographique à l'intérieur de la parcelle. La prise en compte de cette hétérogénéité conduit à définir des sous-régions dont la structure spatiale interne peut être analysée par les fonctions de Ripley et permettre une interprétation plus précise de l'origine des structures observées.

3.3.1 – Les données de la parcelle d'Uppangala.

Les données de cette parcelle sont consignées dans la carte Uppangal de la pile de données. Les fichiers Up_XY et Up_Rect contiennent respectivement les coordonnées de 582 individus de diamètre supérieur à 10 cm et les coordonnées de l'origine et du coin opposé du domaine d'étude rectangulaire de 100 x 80 m :



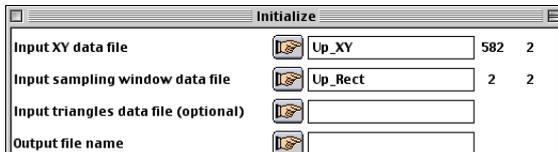
L'option Plots : Labels permet de représenter le semis à partir de ces fichiers :



Le fond de carte est sauvegardé sous le nom Up_Map.

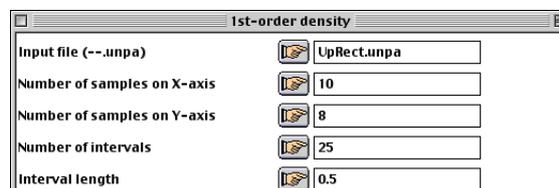
3.3.2 – Analyse de la structure spatiale sur l'ensemble de la parcelle.

La carte du semis ne laisse pas apparaître d'hétérogénéité majeure des propriétés de premier ordre. On peut s'en assurer en étudiant la distribution spatiale de la variable densité locale, dont on prépare l'analyse en liant les fichiers Up_XY et Up_Rect par l'option Density : Initialize :



```
-----
New TEXT file UpRect.unpa contains the parameters:
----> XY data file: Up_XY [582][2]
----> Shape of the sampling window: 1
      1 = rectangular
      2 = circular
      3 = irregular within a rectangular window
      4 = irregular within a circular window
----> Sampling window data file: Up_Rect [2][2]
-----
```

Le listing signale la création du fichier de paramètres UpRect.unpa qui sert d'entrée pour l'option de calcul Density : 1st-order density. La densité locale est estimée aux nœuds d'une grille systématique de 10 x 10 m, pour 25 pas de distance de 0,5 m :



Le listing renvoie une description des fichiers de données définis par UpRect.unpa et rappelle les paramètres du calcul :

```

-----
First-order local density functions (Pélissier & Goreaud 2001)
----XY data file: Up_XY
It contains 582 points
----Sampling window data file: Up_Rect
Xmin: 0.0000e+00 Ymin: 0.0000e+00
Xmax: 1.0000e+02 Ymax: 8.0000e+01
Number of points within the rectangular sampling window: 582
----Input parameters
Area of the study region: 8.0000e+03
Number of points within the study region: 582 (density = 7.2750e-02)
Inter-node distance on X: 1.0000e+01
                    on Y: 1.0000e+01
Number of sampling points within the study region: 80
rmax= 1.2500e+01
dr= 5.0000e-01

```

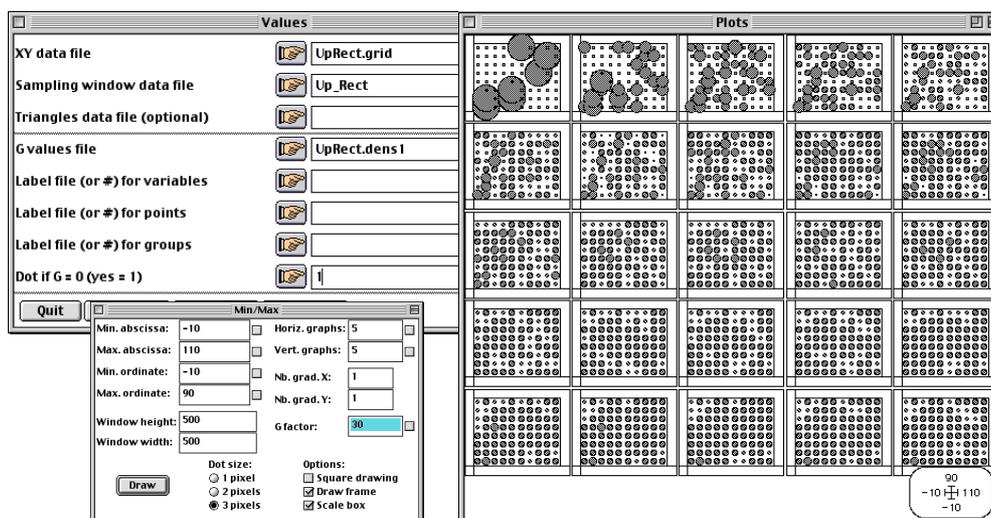
Le fichier UpRect.grid contient les coordonnées des positions d'échantillonnage et le fichier UpRect.dens1 les valeurs de la densité locale, estimées à chaque position d'échantillonnage (en lignes) et aux différents pas de distance (en colonnes) :

```

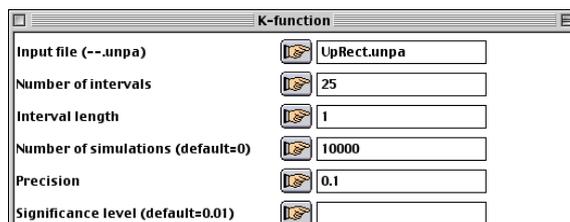
----File UpRect.grid contains the coordinates of the sampling points A
It has 80 rows and 2 columns
----File UpRect.count1 contains the counting function NA(r)
It has 80 rows and 25 columns
----File UpRect.dens1 contains the local density function nA(r)
It has 80 rows and 25 columns
-----

```

L'option **Plots : Values** permet d'obtenir une carte des valeurs de la densité locale. Pour une meilleure lisibilité de la figure, le facteur G est réglé à 30 dans la fenêtre **Min. & Max.** :



La figure confirme l'homogénéité de la densité de la parcelle. L'option **Ripley : K-fonction** est lancée à partir du fichier UpRect.unpa. On choisit d'effectuer le calcul sur 25 pas de distance de 1 m, avec une précision des coordonnées simulées pour le calcul d'un intervalle de confiance local fixée à 0,1 m, 10000 simulations de Monte Carlo et un risque de première espèce $\alpha = 0,01$:



Le listing renvoie une description des fichiers de données définis par UpRect.unpa et rappelle les paramètres du calcul :

```
-----
nivariate second-order neighbourhood functions (Ripley 1977)
---XY data file: Up_XY
t contains 582 points
---Sampling window data file: Up_Rect
min: 0.0000e+00 Ymin: 0.0000e+00
max: 1.0000e+02 Ymax: 8.0000e+01
umber of points within the rectangular sampling window: 582
---Input parameters
rea of the study region: 8.0000e+03
umber of points within the study region: 582 (density = 7.2750e-02)
max = 2.5000e+01
r = 1.0000e+00
umber of Monte Carlo simulations = 10000
recision of the simulated coordinates = 1.0000e-01
```

Le tableau suivant donne, pour chaque pas de distance r , le seuil de signification du test local de l'hypothèse nulle d'une répartition aléatoire des points du semis :

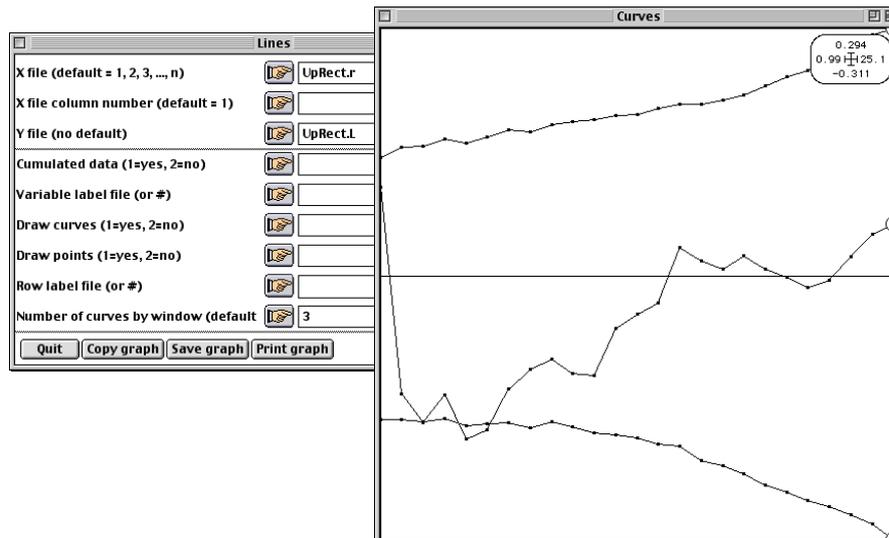
```
---Local significance level of the null hypothesis
-values of |Obs| <= |Simu|
-----
  r |   g(r) |   n(r) |   K(r) |   L(r) |
-----
1.0000 | 6.479e-02 | 6.479e-02 | 6.479e-02 | 8.799e-02 |
2.0000 | 4.000e-04 | 3.080e-02 | 3.080e-02 | 2.690e-02 |

24.0000 | 2.733e-01 | 6.648e-01 | 6.648e-01 | 6.651e-01 |
25.0000 | 5.937e-01 | 6.121e-01 | 6.121e-01 | 6.128e-01 |
-----
```

Les fichiers de résultats contiennent les valeurs des différents estimateurs et les bornes de l'intervalle de confiance à 1% :

```
---File UpRect.r contains the distance steps
t has 25 rows and 1 column
---File UpRect.g contains the pair density function
t has 25 rows and 3 columns
d 1: g(r)
d 2,3: 99% confidence interval
---File UpRect.n contains the local neighbour density function
t has 25 rows and 3 columns
d 1: n(r)
d 2,3: 99% confidence interval
---File UpRect.K contains the Ripley's function
t has 25 rows and 3 columns
d 1: K(r)
d 2,3: 99% confidence interval
---File UpRect.L contains the modified Ripley's function
t has 25 rows and 3 columns
d 1: L(r)
d 2,3: 99% confidence interval
-----
```

On représente graphiquement la fonction $L(r)$ par Curve : Lines en groupant la fonction estimée (1) et les bornes de l'intervalle de confiance local (2 et 3) :



La courbe $\hat{L}(r)$ revient dans l'intervalle de confiance aux grandes distances, et ne permet donc pas de mettre en évidence une hétérogénéité de la structure spatiale. Cependant, la présence d'un pic important d'agrégation à 1 m, immédiatement suivi d'une régularité significative entre 2 et 6 m soulèvent des difficultés d'interprétation. Bien que certains modèles de réaction/diffusion permettent parfois d'expliquer la combinaison de phénomènes de facilitation à courte distance et d'inhibition à moyenne distance (Couteron 1997), il paraît plus probable ici que les deux types d'interaction ne s'adressent pas aux mêmes individus. Pour vérifier cette hypothèse, on peut étudier la distribution des valeurs la densité locale de voisinage de second ordre calculée par l'option Ripley : Ki-values⁴ :



```

-----
nivariate individual second-order neighbourhood values (Getis & Franklin 1987)
---XY data file: Up_XY
t contains 582 points
---Sampling window data file: Up_Rect
min: 0.0000e+00 Ymin: 0.0000e+00
max: 1.0000e+02 Ymax: 8.0000e+01
umber of points within the rectangular sampling window: 582
---Input parameters
rea of the study region: 8.0000e+03
umber of points within the study region: 582 (density = 7.2750e-02)
max = 2.5000e+01
r = 1.0000e+00

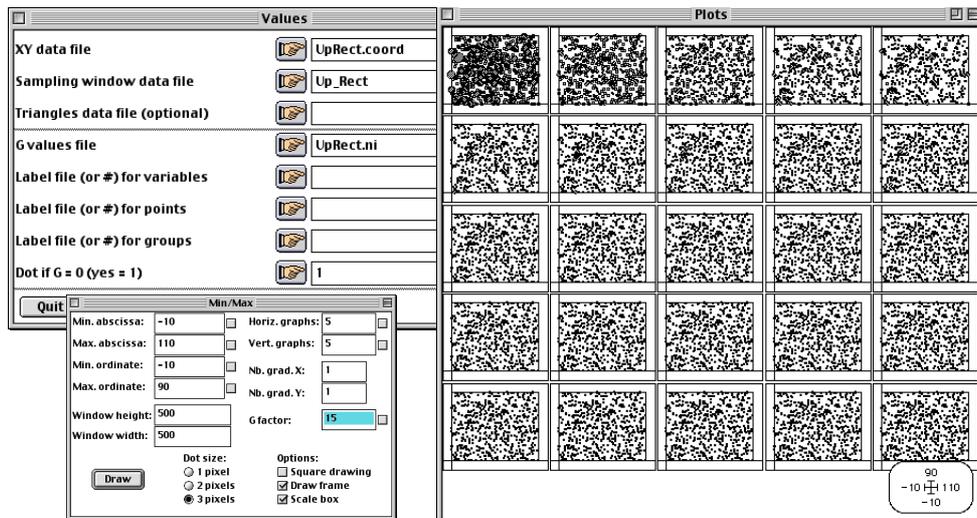
---File UpRect.coord contains the point coordinates
t has 582 row and 2 columns
---File UpRect.gi contains individual pair density values
t has 582 row and 25 columns
---File UpRect.ni contains individual local neighbour density values
t has 582 row and 25 columns
---File UpRect.Ki contains individual Ripley's values
t has 582 row and 25 columns
---File UpRect.Li contains modified individual Ripley's values
t has 582 row and 25 columns
-----

```

Le fichier UpRect.ni contient les valeurs de la densité locale de voisinage de second ordre pour chacun des points du semis (en lignes) aux différents pas de distance (en colonnes).

⁴ Le même calcul pourrait être réalisé par l'option Density : 2nd-order density.

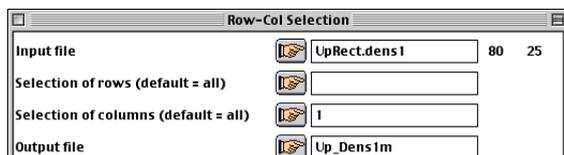
L'option Plots : Values permet de visualiser la distribution spatiale de ces valeurs. Le facteur G est réglé à 15 dans la fenêtre Min. & Max. pour améliorer la lisibilité de la figure :



La figure montre un regroupement, aux petites distances, des valeurs élevées de $n_i(r)$ dans la partie gauche de la parcelle.

3.3.3 – Définition de sous-régions homogènes.

A partir des résultats précédents, on choisit d'utiliser les valeurs de la densité locale de voisinage de second ordre calculée à $r = 1$ m pour définir des sous-régions homogènes à l'intérieur du domaine rectangulaire initial. Pour des raisons pratiques, la colonne 1 est extraite du fichier UpRect.dens1 par FilesUtil : Row-Col selection :



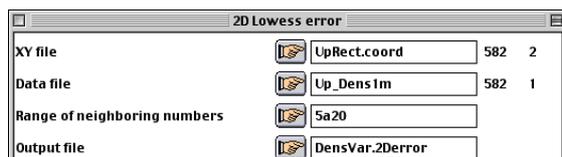
```
-----
electrc: Selection of rows & columns in a file
nput file: UpRect.ni
-- Number of rows: 582, columns: 25
utput file: Up_Dens1m
-- Number of rows: 582, columns: 1
-----
```

Pour un processus de Poisson homogène, la valeur attendue de la densité locale de voisinage de second ordre est $\lambda = 582/(100*80) = 0,073$, quel que soit r . On choisit d'utiliser cette ligne de contour pour délimiter des sous-régions homogènes à l'intérieur du domaine rectangulaire initial. On pourra suivre les étapes détaillées au paragraphe 3.1.3 pour tracer la ligne de contour par le module Levels :

1) Préparer, à partir du fond de carte Up_Map, une grille d'échantillonnage de 10 x 10 m (8 cellules en ligne et 10 cellules en colonne) en utilisant les options Lattices : Create Bkgnd (création des fichiers Up_Grid et Up_Grid.lat) et Lattices : LattiToLevel (création des fichiers Up_Grid.rect et Up_Grid.summ).

2) Ramener les coordonnées des points du semis à l'échelle écran de la carte Up_Map en utilisant l'option ADSUtil : XY->Digit (Création du fichier UpRect.coord_Pt).

3) Calculer l'erreur d'estimation en fonction du nombre de voisins pris en compte dans la régression locale par l'option MapUtil : 2D lowess error :



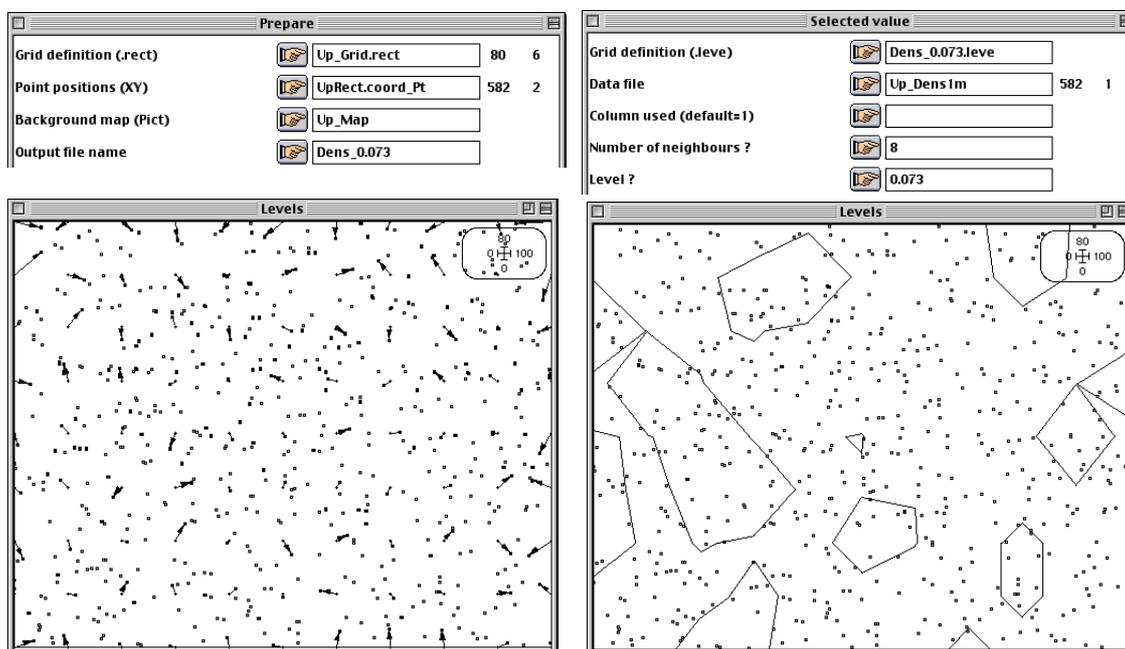
```

-----
Lowess error: total error of bidimensional local regression
Y coordinates file: UpRect.coord
-- Number of rows: 582, columns: 2
Data file: Up_Dens1m
-- Number of rows: 582, columns: 1
Number of neighbors: 5a25
1.1169E+00
8.7606E-01
7.8863E-01
7.4980E-01
7.6581E-01
7.8423E-01
7.9088E-01
8.0946E-01
8.2032E-01
8.4110E-01
+8.5134E-01
+8.6693E-01
+8.7485E-01
+8.7959E-01
+8.8738E-01
+8.9119E-01
+8.9258E-01
+8.9506E-01
+9.0128E-01
+9.0396E-01
+9.0622E-01
Output file: UpRect.2Derror
--- Number of rows: 21, columns: 1
This file contains, for each variable (in columns),
and for each number of neighbors (in rows) the sum
of estimation error for all sampling points
-----

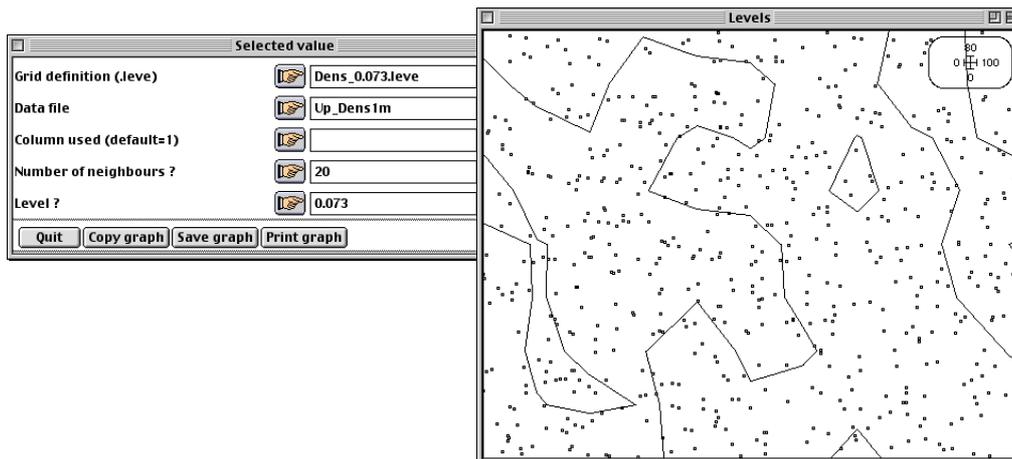
```

On choisit d'utiliser les 8 plus proches voisins pour prédire les valeurs de la densité locale de voisinage de second ordre.

4) Tracer la ligne de contour 0,073 en utilisant les options Levels: Prepare et Levels Selected value :



La figure fait apparaître plusieurs îlots qui tendent à se regrouper lorsqu'on augmente le nombre de voisins pris en compte dans la régression. A partir de 20 voisins, une sous-région de forme stable apparaît dans la partie gauche de la parcelle :



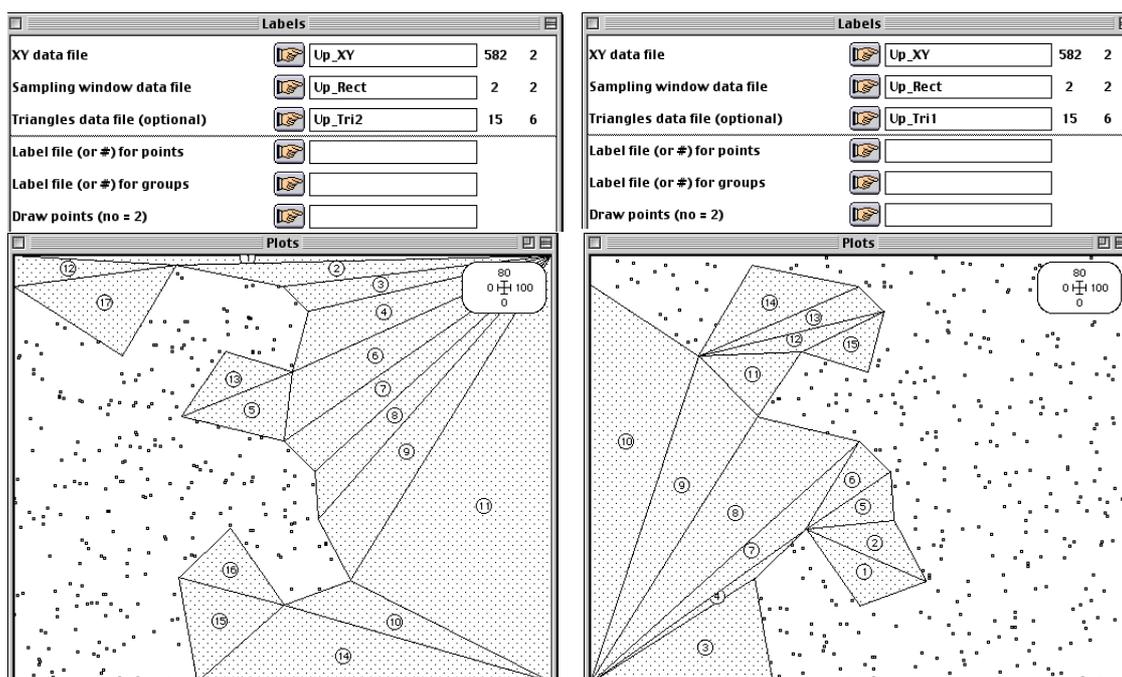
Pour des raisons de simplicité de définition des sous-régions homogènes, on préférera cette figure à la précédente. Le nouveau fond de carte est sauvegardé sous le nom Dens_Map.

Cette carte va servir de référence pour définir deux nouveaux domaines d'étude, l'un correspondant à la partie gauche de la parcelle (zone 1) où se trouvent regroupés les individus ayant plus de voisins dans une distance < 1 m que ne le laisserait supposer le hasard, l'autre (zone 2) correspondant à la partie droite de la parcelle. Ces deux domaines, de formes complexes, sont approchés par des polygones que l'on définit en excluant alternativement les zones 1 et 2 du domaine rectangulaire initial (Goreaud & Pélissier 1999). On pourra encore une fois suivre les étapes détaillées au paragraphe 3.1.3, pour :

1) Saisir les coordonnées des sommets du polygone qui couvre la zone 1, puis les coordonnées des sommets du polygone qui couvre la zone 2 par Digit : Digitize (création des fichiers Up_Poly1 et Up_Poly2).

2) Convertir les coordonnées écran de Up_Poly1 et Up_Poly2 en coordonnées métriques par ADSUtil : Digit->XY (création des fichiers Up_Poly1_XY et Up_Poly2_XY).

3) Décomposer les polygones définis par Up_Poly1_XY et Up_Poly2_XY en triangles contigus par ADSUtil : Triangulate polygon (création des fichiers Up_Tri1 et Up_Tri2). L'option Plots : Labels permet de représenter les deux sous-régions :



On a ainsi délimité deux sous-régions à l'intérieur desquelles on pourra alternativement analyser la structure spatiale à l'aide des fonctions de Ripley.

3.3.4 – Analyse de la structure spatiale dans les sous-régions.

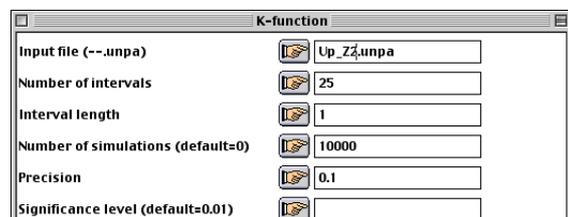
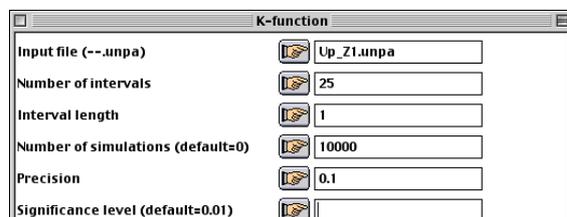
On prépare l'analyse de la zone 1 en liant les fichiers Up_XY, Up_Rect, Up_Tri2 et l'analyse de la zone 2 en liant les fichiers Up_XY, Up_Rect, Up_Tri1 par Ripley : Initialize :



```
-----
ew TEXT file Up_Z1.unpa contains the parameters:
--> XY data file: Up_XY [582][2]
--> Shape of the sampling window: 3
    1 = rectangular
    2 = circular
    3 = irregular within a rectangular window
    4 = irregular within a circular window
--> Sampling window data file: Up_Rect [2][2]
--> Triangle data file: Up_Tri2 [17][6]
-----
```

```
-----
ew TEXT file Up_Z2.unpa contains the parameters:
--> XY data file: Up_XY [582][2]
--> Shape of the sampling window: 3
    1 = rectangular
    2 = circular
    3 = irregular within a rectangular window
    4 = irregular within a circular window
--> Sampling window data file: Up_Rect [2][2]
--> Triangle data file: Up_Tri1 [15][6]
-----
```

L'analyse de la structure spatiale de chacune des zones est lancée par Ripley : K-fonction à partir des fichiers Up_Z1.unpa et Up_Z2.unpa en utilisant les mêmes paramètres que pour l'analyse préliminaire :



```
-----
Univariate second-order neighbourhood functions (Ripley 1977)
---XY data file: Up_XY
It contains 582 points
---Sampling window data file: Up_Rect
Xmin: 0.0000e+00 Ymin: 0.0000e+00
Xmax: 1.0000e+02 Ymax: 8.0000e+01
Number of points within the rectangular sampling window: 582
---Input triangle data file: Up_Tri2
It contains 17 triangles
Number of triangles within the rectangular sampling window: 17
Number of points within the triangles: 334
---Input parameters
Area of the study region: 3.2157e+03
Triangles within the study region: 17
Number of points within the study region: 248
(density = 7.7122e-02)
rmax = 2.5000e+01
dr = 1.0000e+00
Number of Monte Carlo simulations = 10000
Precision of the simulated coordinates = 1.0000e-01
-----
```

```
-----
Univariate second-order neighbourhood functions (Ripley 1977)
---XY data file: Up_XY
t contains 582 points
---Sampling window data file: Up_Rect
min: 0.0000e+00 Ymin: 0.0000e+00
max: 1.0000e+02 Ymax: 8.0000e+01
umber of points within the rectangular sampling window: 582
---Input triangle data file: Up_Tri1
t contains 15 triangles
umber of triangles within the rectangular sampling window: 15
umber of points within the triangles: 248
---Input parameters
rea of the study region: 4.7803e+03
riangles within the study region: 15
umber of points within the study region: 334
density = 6.9871e-02)
rmax = 2.5000e+01
r = 1.0000e+00
umber of Monte Carlo simulations = 10000
recision of the simulated coordinates = 1.0000e-01
-----
```

---Local significance level of the null hypothesis
P-values of $|Obs| \leq |Simu|$

r	g(r)	n(r)	K(r)	L(r)
1.0000	0.000e+00	0.000e+00	0.000e+00	2.000e-04
2.0000	1.190e-02	9.432e-01	9.432e-01	9.433e-01
...				
24.0000	6.038e-01	6.587e-01	6.587e-01	6.594e-01
25.0000	9.898e-01	6.888e-01	6.888e-01	6.894e-01

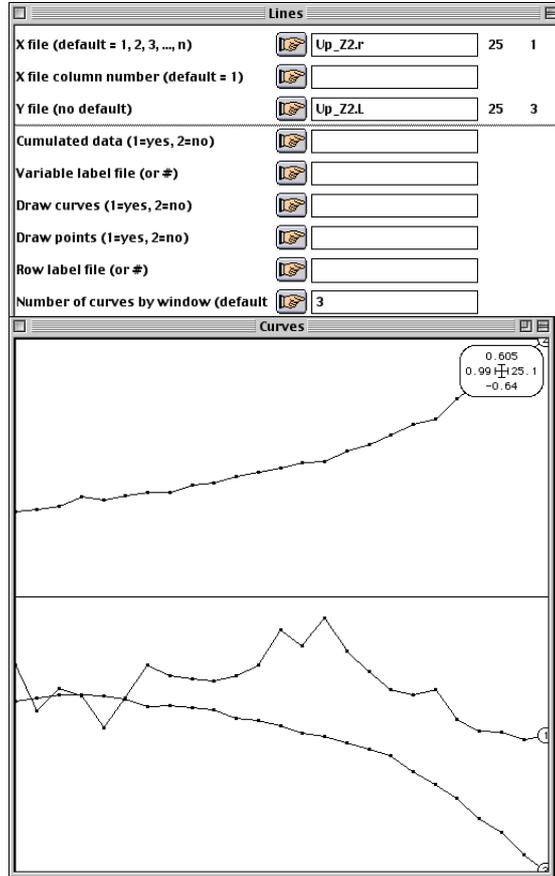
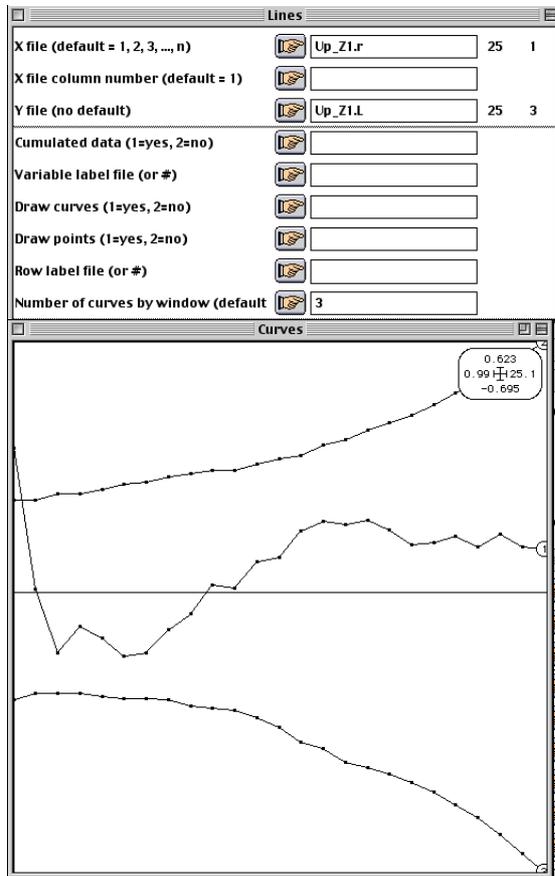
---File Up_Z1.r contains the distance steps
It has 25 rows and 1 column
---File Up_Z1.g contains the pair density function
It has 25 rows and 3 columns
Col 1: g(r)
Col 2,3: 99% confidence interval
---File Up_Z1.n contains the local neighbour density function
It has 25 rows and 3 columns
Col 1: n(r)
Col 2,3: 99% confidence interval
---File Up_Z1.K contains the Ripley's function
It has 25 rows and 3 columns
Col 1: K(r)
Col 2,3: 99% confidence interval
---File Up_Z1.L contains the modified Ripley's function
It has 25 rows and 3 columns
Col 1: L(r)
Col 2,3: 99% confidence interval

---Local significance level of the null hypothesis
-values of $|Obs| \leq |Simu|$

r	g(r)	n(r)	K(r)	L(r)
1.0000	8.379e-02	7.979e-02	8.379e-02	6.689e-02
2.0000	1.880e-02	3.500e-03	3.500e-03	2.300e-03
...				
24.0000	5.275e-01	1.330e-01	1.330e-01	1.311e-01
25.0000	9.724e-01	1.693e-01	1.693e-01	1.675e-01

---File Up_Z2.r contains the distance steps
t has 25 rows and 1 column
---File Up_Z2.g contains the pair density function
t has 25 rows and 3 columns
col 1: g(r)
col 2,3: 99% confidence interval
---File Up_Z2.n contains the local neighbour density function
t has 25 rows and 3 columns
col 1: n(r)
col 2,3: 99% confidence interval
---File Up_Z2.K contains the Ripley's function
t has 25 rows and 3 columns
col 1: K(r)
col 2,3: 99% confidence interval
---File Up_Z2.L contains the modified Ripley's function
t has 25 rows and 3 columns
col 1: L(r)
col 2,3: 99% confidence interval

On représente les courbes de $L(r)$ par Curves : Lines en groupant la fonction estimée pour le semis réel (1) et les bornes de l'intervalle de confiance local (2 et 3) :



Les analyses mettent en évidence une nette séparation des pics d'agrégation et de régularité de l'analyse préliminaire, ce qui montre que la segmentation du domaine d'étude initial en deux sous-régions permet de prendre en compte l'hétérogénéité de la structure de second ordre du semis de points. Ainsi la zone 1 montre un pic d'agrégation à faible distance et la zone 2 une nette régularité avec deux pics significatifs à 2 et 5 m. Pélissier (1997, 1998) interprète cette hétérogénéité comme le résultat de fonctionnements dynamiques différents : dans la zone 1, le renouvellement du peuplement est dominé par une régénération des jeunes au pied des adultes ; dans la zone 2, le renouvellement est dominé par le développement de jeunes à la périphérie de la couronne des adultes. Plusieurs hypothèses pourraient permettre d'expliquer ce phénomène : 1) une différence de composition floristique entre les deux zones ; 2) la présence de mycorhizes facilitant localement l'installation des jeunes au pied des parents ; 3) une disponibilité plus grande de la lumière sous la couronne des grands arbres de la zone 1 due au fait que cette partie de la parcelle se trouve dans une pente orientée à l'ouest.

4 — Références.

- Barot, S., Gignoux, J. & Menaut, J.-C. 1999. Demography of a savanna palm tree: predictions from comprehensive spatial pattern analyses. *Ecology*, 80: 1987-2005.
- Barthès, B. 1991. *Caractérisation pédologique des parcelles du dispositif "forêt naturelle" du CTFT à Paracou*. Orstom, Cayenne: 1-14.
- Batista, J.L.F. & Maguire, D.A. 1998. Modeling the spatial structure of tropical forests. *Forest Ecology and Management*, 110: 293-314.
- Begon, M., Harper, J.L. & Townsend, C.R. 1996. *Ecology: individuals, populations and communities* (3rd edition). Blackwell Science, Oxford: 1-1068.
- Besag, J. 1977. Contribution to the discussion of Dr Ripley's paper. *Journal of the Royal Statistical Society*, B39: 193-195.
- Besag, J. & Diggle, P.J. 1977. Simple Monte Carlo tests for spatial pattern. *Applied Statistics*, 26: 327-333.
- Biber, P. 1999. Ein verfahren zum ausgleich von rabdeffekten bei der berechnung von konkurrenzindizes. Deutscher verband forstlicher forschungsanstalten. Jahrestagung 1999, Volpriehausen, 19-21 mai 1999: 189-202.
- Chessel, D. 1978. Description non paramétrique de la dispersion spatiale des individus d'une espèce. In *Biométrie et écologie*, Legay, J.-M. & Tomassone, R. (Eds.), Société Française de Biométrie, Paris: 45-133.
- Collinet, F. 1997. *Essai de regroupement des principales espèces structurantes d'une forêt dense humide d'après l'analyse de leur répartition spatiale (Forêt de Paracou, Guyane)*. Thèse de Doctorat, Université Claude Bernard, Lyon: 1-313.
- Couteron, P. 1998. *Relations spatiales entre individus et structure d'ensemble dans des peuplements ligneux soudano-sahéliens au nord-ouest du Burkina-Faso*. Thèse de Doctorat, Université Paul Sabatier, Toulouse: 1-223.
- Clark, P.J. & Evans, F.C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35: 445-453.
- Cleveland, W.S. 1979. Robust locally weighted regression and smoothing scatterplot. *Journal of the American Statistical Association*, 74: 829-836.
- Cressie, N.A.C. 1993. *Statistics for spatial data*. Wiley, New-York: 1-900.
- Delvaux, J. 1981. Différenciation sociale. *Journal de Foresterie Suisse*, 132: 733-749.

- Dessart, H. 1996. *Estimation de l'intensité locale d'un processus ponctuel de Cox : application à l'analyse spatiale d'un inventaire forestier*. Thèse de Doctorat, Université des Sciences et Techniques du Languedoc, Montpellier: 1-336.
- Diggle, P.J. 1983. *Statistical analysis of spatial point patterns*. Academic Press, London: 1-148.
- Getis, A. & Franklin, J. 1987. Second order neighborhood analysis of mapped point patterns. *Ecology*, 68: 473-477.
- Gignoux, J., Duby, C. & Barot, S. 1999. Comparing the performances of Diggle's tests of spatial randomness for small samples with and without edge effect correction: application to ecological data. *Biometrics*, 55: 156-164.
- Goreaud, F. 2000. *Apports de l'analyse de la structure spatiale en forêt tempérée à l'étude et la modélisation des peuplements complexes*. Thèse de Doctorat, ENGREF, Nancy: 1-528.
- Goreaud, F. & Pélissier, R. 1999. On explicit formulas of edge effect correction for Ripley's K-function. *Journal of Vegetation Science*, 10: 433-438.
- Kiêu, K. & Mora, M. 1999. Estimating the reduced moments of a random measure. *Advances in Applied Probability*, 31: 48-62.
- Martens, S.N., Breshears, D.D., Meyer, C.W. & Barnes, F.J. 1997. Scales of above-ground and below-ground competition in a semi-arid woodland detected from spatial pattern. *Journal of Vegetation Science*, 8: 655-664.
- Pélissier, R. 1997. *Hétérogénéité spatiale et dynamique d'une forêt dense humide dans les Ghâts occidentaux de l'Inde*. Publications du Département d'Ecologie, n° 37, Institut Français de Pondichéry, Inde: 1-148.
- Pélissier, R. 1998. Tree spatial patterns in three contrasting plots of a southern Indian tropical moist evergreen forest. *Journal of Tropical Ecology*, 14: 1-16.
- Pélissier, R. & Goreaud, F. 2001. A practical approach to studying the spatial structure in simple cases of heterogeneous vegetation stands. *Journal of Vegetation Science*, 12: in press.
- Ripley, B.D. 1976. The second order analysis of stationary point process. *Journal of Applied Probability*, 13: 255-266.
- Ripley, B.D. 1977. Modelling spatial patterns. *Journal of the Royal Statistical Society*, B39: 172-212.
- Ripley, B.D. 1979. Test of "randomness" for spatial point pattern. *Journal of the Royal Statistical Society*, B41: 368-374.
- Ripley, B.D. 1981. *Spatial statistics*. Wiley, New-York: 1-250.
- Ripley, B.D. 1982. Edge effects in spatial stochastic processes. In *Statistics in theory and practice*, Ranney, B. (Ed.), Swedish University of Agricultural Sciences, Umea, Sweden: 247-262.
- Rowlingson, B.S. & Diggle, P.J. 1993. Splanx: spatial point pattern analysis code in S-Plus. *Computers and Geosciences*, 19: 627-655.
- Stoyan, D., Kendall, W.S. & Mecke, J. 1995. *Stochastic geometry and its applications*. Wiley, New-York: 1-436.