

Nicolas Picard
DEA d'Écologie générale - Université Paris 6

le 28 août 1996

Répartition spatiale des essences forestières :
Étude de distances entre espèces
basées sur des dissimilarités entre
nuages de points

Stage réalisé au service Biométrie du CIRAD-Forêt
Campus Baillarguet - BP 5035
34032 Montpellier Cedex 1

sous la direction de Louis Houde

Répartition spatiale des essences forestières : étude de
distances entre espèces basées sur des dissimilarités entre
nuages de points

Nicolas Picard

DEA d'Écologie Générale - Université Paris 6

Stage réalisé au service Biométrie du CIRAD-Forêt

Campus Baillarguet - BP 5035

34032 Montpellier Cedex 1

sous la direction de Louis Houde

28 août 1996

Remerciements

Je tiens à remercier tous ceux qui m'ont appuyé pendant ce stage, à commencer par mon directeur de stage Louis Houde, ainsi que Jean-Claude Bergonzini, Avner Bar-Hen et Alain Franc pour le suivi qu'ils ont apporté à mon travail, Guillaume Cornu pour ses dépannages informatiques, Vincent Favrichon et Sylvie Gourlet-Fleury pour les indications qu'ils m'ont données.

I would like also to thank very much Dr Klinka for the great support he gave me in Vancouver, and all his research team in UBC for their nice welcome. In particular I would like to thank Marcus Rudyk who helped me counting thousands of trees, and occasionally bears.

Résumé

La répartition spatiale des arbres reflète les mécanismes biologiques qui influent la croissance des arbres (exigences édaphiques, compétition, etc.). On cherche donc à définir une distance entre deux espèces en s'appuyant sur la répartition spatiale des arbres, ce qui revient à définir une distance entre deux nuages de points.

Plusieurs indices de dissimilarité entre nuages de points sont définis. Huit d'entre eux sont étudiés plus précisément, en simulant des répartitions spatiales à l'aide de processus ponctuels, et en regardant le comportement de chacun des indices sur ces simulations. Les indices ne répondent pas tous à la même question ; parmi ceux qui répondent à la même question, certains sont plus performants que d'autres.

Les indices sont ensuite mis en application dans deux cas concrets : une forêt de la côte nord ouest du Pacifique, près de Vancouver (Canada), et la forêt de Paracou en Guyane française. Ces deux forêts diffèrent par la complexité des interactions entre espèces. Les indices permettent de caractériser assez précisément les relations entre espèces dans la forêt de Colombie Britannique, mais sont beaucoup moins concluants pour la forêt guyanaise.

TABLE DES MATIERES

| | | |
|----------|--|-----------|
| 1 | Introduction | 4 |
| 2 | Les répartitions spatiales, reflet de situations biologiques | 5 |
| 2.1 | Pour une espèce | 5 |
| 2.1.1 | Description d'un nuage de points | 5 |
| 2.1.2 | Facteurs écologiques pouvant influencer la répartition d'une espèce | 5 |
| 2.2 | Pour deux espèces | 6 |
| 2.2.1 | Description de la relation entre deux répartitions | 6 |
| 2.2.2 | Facteurs écologiques pouvant influencer le lien les répartitions de deux espèces | 8 |
| 2.3 | La modélisation par les processus ponctuels | 9 |
| 2.3.1 | Modélisation de la répartition de deux espèces qui n'interagissent pas dans un milieu homogène | 9 |
| 2.3.2 | Modélisation de la répartition de deux espèces qui interagissent dans un milieu homogène | 9 |
| 2.3.3 | Modélisation de la répartition de deux espèces qui n'interagissent pas directement dans un milieu inhomogène | 10 |
| 2.4 | Et pour trois (ou plus) espèces ? | 10 |
| 3 | Etude d'indices de dissimilarité | 11 |
| 3.1 | Démarche | 11 |
| 3.2 | Présentation des indices | 11 |
| 3.3 | Résultats des simulations | 13 |
| 3.3.1 | Indices mesurant un écart à l'indépendance | 13 |
| 3.3.2 | Indices mesurant un écart à l'association | 14 |
| 3.3.3 | Indices mesurant un écart à l'étiquetage aléatoire | 14 |
| 3.3.4 | Conclusion | 14 |
| 4 | Applications : étude de répartitions bivariés dans des peuplements forestiers de la côte Nord Ouest du Pacifique et de Guyane | 16 |
| 4.1 | Etude de la forêt de Vancouver | 16 |
| 4.1.1 | Description des peuplements | 16 |
| 4.1.2 | Echantillonnage | 17 |
| 4.1.3 | Etude des répartitions univariées | 17 |
| 4.1.4 | Etude des répartitions bivariées | 18 |
| 4.1.5 | Etude de la répartition trivariée (parcelle 3) | 21 |
| 4.1.6 | Lien avec les caractéristiques écologiques des espèces | 22 |
| 4.2 | Etude de la forêt de Paracou | 23 |
| 4.2.1 | Description du site | 23 |
| 4.2.2 | Choix des espèces | 23 |
| 4.2.3 | Tests de l'indépendance | 24 |
| 4.2.4 | Indices de dissimilarité | 27 |
| 4.2.5 | Conclusion | 28 |
| 5 | Conclusion | 29 |

| | | |
|----------|--|-----------|
| A | Indices de dissimilarité | 30 |
| B | Quelques rappels sur la théorie des processus ponctuels | 31 |
| B.1 | Définitions | 31 |
| B.2 | Propriétés | 31 |
| B.2.1 | Processus simple | 31 |
| B.2.2 | Stationnarité | 31 |
| B.2.3 | Isotropie | 32 |
| B.2.4 | Indépendance | 32 |
| B.3 | Caractéristiques du premier ordre | 32 |
| B.3.1 | Mesure de premier moment | 32 |
| B.3.2 | Intensité | 32 |
| B.4 | Caractéristiques du second ordre | 32 |
| B.4.1 | Mesure de second moment | 32 |
| B.4.2 | Densité produit de second ordre | 33 |
| B.4.3 | Mesure réduite de second moment | 33 |
| B.5 | Processus indépendants et étiquetage aléatoire | 34 |
| B.5.1 | Indépendance | 34 |
| B.5.2 | Dépendance | 34 |
| B.5.3 | Etiquetage aléatoire | 34 |
| B.6 | Processus de Poisson univarié | 34 |
| B.7 | Processus ponctuels bivariés | 35 |
| B.7.1 | Processus de Poisson bivariés | 35 |
| B.7.2 | Processus parents-enfants de Dixon [20] | 36 |
| B.7.3 | Processus d'inhibition mutuelle [17] | 36 |
| B.7.4 | Processus de Poisson isolé [29] | 36 |
| B.7.5 | Processus de Cox bivarié [19] | 36 |
| B.7.6 | Processus ponctuel bivarié éclairci [19] | 37 |
| B.7.7 | Processus de Markov bivarié [14, p.102] | 38 |
| C | Calculs pour les indices de dissimilarité | 41 |
| C.1 | Méthode par présence / absence dans des quadrats | 41 |
| C.2 | Méthode de Kershaw | 42 |
| C.3 | Méthode de Pielou | 43 |
| C.3.1 | Superposition de deux processus de Poisson homogènes indépendants | 44 |
| C.3.2 | Superposition de deux réseaux réguliers indépendants | 44 |
| D | Tableaux de contingence $2 \times 2 \times 2$ | 45 |
| D.1 | Mesure de l'interaction par le rapport θ | 45 |
| D.2 | Absence d'interaction à trois facteurs | 46 |
| D.3 | Indépendance conditionnelle simple | 46 |
| D.4 | Indépendance partielle simple | 47 |
| D.5 | Indépendance totale | 47 |
| D.6 | Indépendance marginale simple | 48 |
| E | Présentation des treize processus ponctuels bivariés de références | 49 |
| E.1 | Superposition de deux processus de Poisson homogènes indépendants (processus PP) | 49 |
| E.2 | Superposition d'un processus de Poisson homogène et d'un processus de Neyman-Scott indépendants (processus PA) | 49 |
| E.3 | Superposition d'un processus de Poisson homogène et d'un processus d'inhibition séquentiel (processus PR) | 50 |
| E.4 | Superposition de deux processus de Neyman-Scott indépendants (processus AA) | 51 |
| E.5 | Superposition d'un processus de Neyman-Scott et d'un processus d'inhibition séquentiel indépendants (processus AR) | 51 |

| | | |
|----------|---|-----------|
| E.6 | Superposition de deux processus d'inhibition séquentiels (processus RR) | 52 |
| E.7 | Processus de Cox équilibré dégénéré (processus ST) | 52 |
| E.8 | Processus de Cox équilibré (processus CE) | 53 |
| E.9 | Processus de Cox lié dégénéré (processus MV) | 53 |
| E.10 | Processus de Cox lié (processus CL) | 54 |
| E.11 | Processus des paires liées (processus Pa) | 54 |
| E.12 | Processus d'inhibition mutuelle (processus In) | 55 |
| E.13 | Processus de Cox composé (processus Ce) | 55 |
| E.14 | Caractéristiques des processus | 56 |
| F | Présentation des indices | 57 |
| F.1 | Méthodes par quadrats | 57 |
| F.1.1 | Méthode par présence / absence dans des quadrats [26, 57, 64] | 57 |
| F.1.2 | Méthode de Kershaw [38] | 60 |
| F.2 | Méthodes par distances | 62 |
| F.2.1 | Méthode de Diggle & Cox [17, 64, p.250] | 62 |
| F.2.2 | Méthode de Goodall [9, 17, 22, 64] | 64 |
| F.2.3 | Méthode des plus proches voisins | 66 |
| F.2.4 | Méthode du Sup de l'Inf [5] | 68 |
| F.2.5 | Méthode de Pielou [20, 45, 54, 64, p.242] | 69 |
| F.2.6 | Méthodes de Ripley | 71 |
| F.3 | Autres méthodes | 72 |
| G | Résultats pour la forêt de Vancouver | 73 |
| G.1 | Etude des répartitions univariées : tests de la répartition aléatoire | 73 |
| G.2 | Etude des répartitions bivariées | 75 |
| G.2.1 | Résultats des tests de l'indépendance | 75 |
| G.2.2 | Valeurs des indices de dissimilarité | 77 |

1. INTRODUCTION

Deux espèces d'arbres poussent-elles ensemble ? Cette question sous-entend plusieurs questions : les deux espèces ont-elles les mêmes exigences écologiques ? les deux espèces entrent-elles en compétition, et si oui, laquelle exclue l'autre ? etc. Pour quantifier la réponse à cette question, on pourrait imaginer de définir une distance entre les deux espèces, qui reflète ces divers phénomènes biologiques.

L'objet du présent travail est de tâcher de définir une distance entre deux espèces, en s'appuyant sur la répartition spatiale des arbres. Si l'on ne retient comme information que la position spatiale des individus, une population d'arbres se ramène à un nuage de points. Le problème se pose alors comme étant de définir une distance entre deux nuages de points. Le problème ainsi posé simplifie énormément le problème plus général de définir une distance entre deux espèces qui résume les phénomènes biologiques ayant conduit ces deux espèces à pousser ensemble ou non.

La première partie de ce rapport discute justement dans quelle mesure la répartition spatiale des arbres intègre les effets des mécanismes biologiques. La deuxième partie présente plusieurs indices de distances entre nuages de points, et tâche de dégager les caractéristiques principales de chacun. La troisième partie enfin présente deux cas concrets de calculs de distances entre deux espèces.

2. LES RÉPARTITIONS SPATIALES, REFLET DE SITUATIONS BIOLOGIQUES

L'idée sous-jacente à l'étude de la répartition spatiale d'une espèce est que celle-ci reflète les facteurs écologiques qui agissent sur la croissance des individus [21, 49, 52].

2.1. Pour une espèce

La répartition spatiale d'une espèce d'arbres est définie par l'emplacement au sol de ses individus ; sa représentation est donc un nuage de points. Avant de voir en quoi les facteurs écologiques peuvent influencer la répartition spatiale d'une espèce, il faut savoir caractériser un nuage de points.

2.1.1. Description d'un nuage de points

Un nuage de points est d'abord caractérisé par sa densité, c'est-à-dire le nombre moyen de points par unité de surface. La densité ne renseigne cependant pas sur la façon dont sont répartis les points dans l'espace. On distingue alors habituellement trois types de répartition : la répartition aléatoire, les répartition régulières, les répartition agrégées [7].

La répartition aléatoire est obtenue en plaçant les points au hasard, indépendamment les uns des autres. Conditionnellement à la présence d'un point en (x, y) , la probabilité de trouver un autre point dans le voisinage de (x, y) est alors égale à la probabilité de trouver un point dans le voisinage de (x, y) .

Les répartition agrégées et régulières ne sont définies que par écart à la répartition aléatoire, et regroupent divers types de répartition : une répartition aléatoire est en fait entièrement caractérisée par sa densité, alors que d'autres paramètres sont nécessaires pour caractériser une répartition agrégée ou régulière. Intuitivement, pour une répartition agrégée (respectivement régulière), conditionnellement à la présence d'un point en (x, y) , la probabilité de trouver un autre point dans le voisinage de (x, y) est supérieure (respectivement inférieure) à la probabilité de trouver un point dans le voisinage de (x, y) .

La distinction entre répartition aléatoire, agrégée ou régulière dépend en fait de l'échelle d'observation. Une répartition peut être aléatoire à grande échelle (c'est-à-dire sur de petites distances) et agrégée à petite échelle (fig.2.1b), ou agrégée à grande échelle et régulière à petite échelle (fig.2.1a) par exemple.

Ces deux exemples montrent qu'il peut y avoir emboîtement de plusieurs échelles d'organisation, la répartition devant être caractérisée à chaque niveau.

2.1.2. Facteurs écologiques pouvant influencer la répartition d'une espèce

On distingue d'une part les facteurs physiques, d'autre part les facteurs biotiques. Les *facteurs physiques* (climat et microclimat, sol, topographie...) définissent généralement des motifs à petite échelle de type agrégatif (plaques, gradients d'intensité...) [13, 25, 53, 62]. Chaque espèce a en effet des exigences écologiques qui définissent un milieu auquel elle est adaptée et des milieux où elle peut potentiellement croître selon sa plus ou moins grande tolérance aux écarts à l'optimum. Dans un milieu hétérogène, l'espèce n'occupera donc pas tout l'espace, et l'échelle d'agrégation de la répartition reflètera l'échelle de variation du milieu. Si le milieu est homogène, il n'a pas d'incidence sur la répartition spatiale de l'espèce.

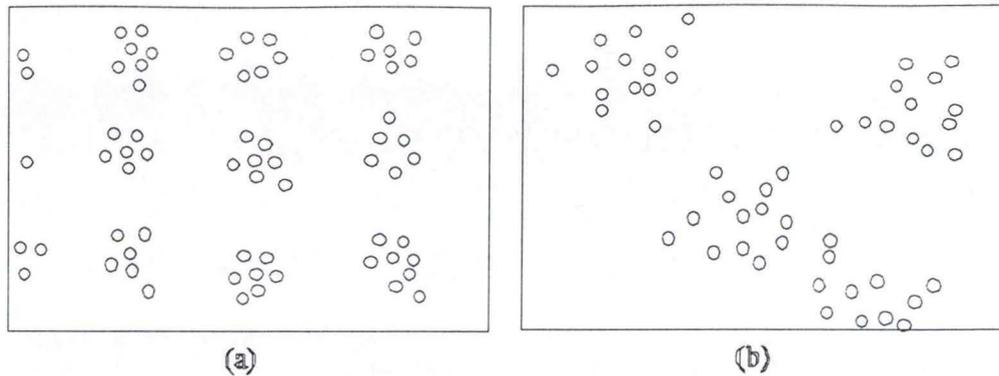


Figure 2.1: Répartitions présentant plusieurs échelles d'organisation

Les *facteurs biotiques* définissent généralement des structures perceptibles à plus grande échelle. La compétition interspécifique entre arbres donnera des motifs de type agrégatif, la compétition intraspécifique donnera au contraire une répartition régulière [55]. L'interaction avec les animaux (herbivorie, dispersion des graines par les animaux, etc.) joue également, ainsi que le mode de reproduction de la plante : une espèce à reproduction végétative, une espèce ayant des graines à faible capacité de dispersion, une espèce dont les graines sont enterrées par les animaux dans des cachettes auront généralement une répartition agrégative ; une espèce anémochore aura tendance au contraire à avoir une répartition plus aléatoire [2, 27].

La répartition observée résulte finalement de l'action conjointe de ces facteurs. Par exemple dans un milieu hétérogène en l'absence d'espèces compétitrices, une espèce suffisamment tolérante aux variations du milieu aura une répartition aléatoire ; dans le même milieu en présence d'espèces compétitrices, la même espèce n'occupera plus que les zones qui lui sont optimales, et présentera donc une répartition agrégée.

La répartition spatiale résulte également de l'intégration dans le temps de l'action de ces différents facteurs. Par exemple chez une espèce comme *Eperua falcata* [2] dont les graines ont une faible capacité de dispersion, les jeunes semis ont une répartition agrégée autour des arbres semenciers ; un mécanisme de mise à distance des semis des arbres semenciers, tel que les semis éloignés de l'arbre semencier ont une probabilité de survie plus grande que ceux situés sous l'arbre, fera que les semis âgés auront finalement une répartition plus régulière.

D'autres facteurs encore [27] peuvent être pris en compte si l'on s'intéresse à des échelles de temps (causes historiques) ou d'espace plus petites (climat régional ; perturbations accidentelles comme les incendies, les tempêtes, les inondations).

2.2. Pour deux espèces

Quand on considère deux espèces simultanément, le problème n'est plus de décrire la répartition de chacune des deux espèces, mais de décrire le lien entre les deux répartitions, et de voir quels facteurs écologiques ont pu influencer ce lien.

2.2.1. Description de la relation entre deux répartitions

Plusieurs points de vue peuvent être adoptés pour décrire la relation entre deux nuages de points :

1. Une approche *géométrique* consiste à chercher à quantifier le mélange entre les deux nuages de points. Un cas extrême correspond à celui de la figure 2.2a, où les deux nuages de points ne sont pas mélangés. A l'autre extrême, on observerait une situation comme celle de la figure 2.2b, où les deux nuages de points sont si proches qu'à la limite on ne les distingue plus : il

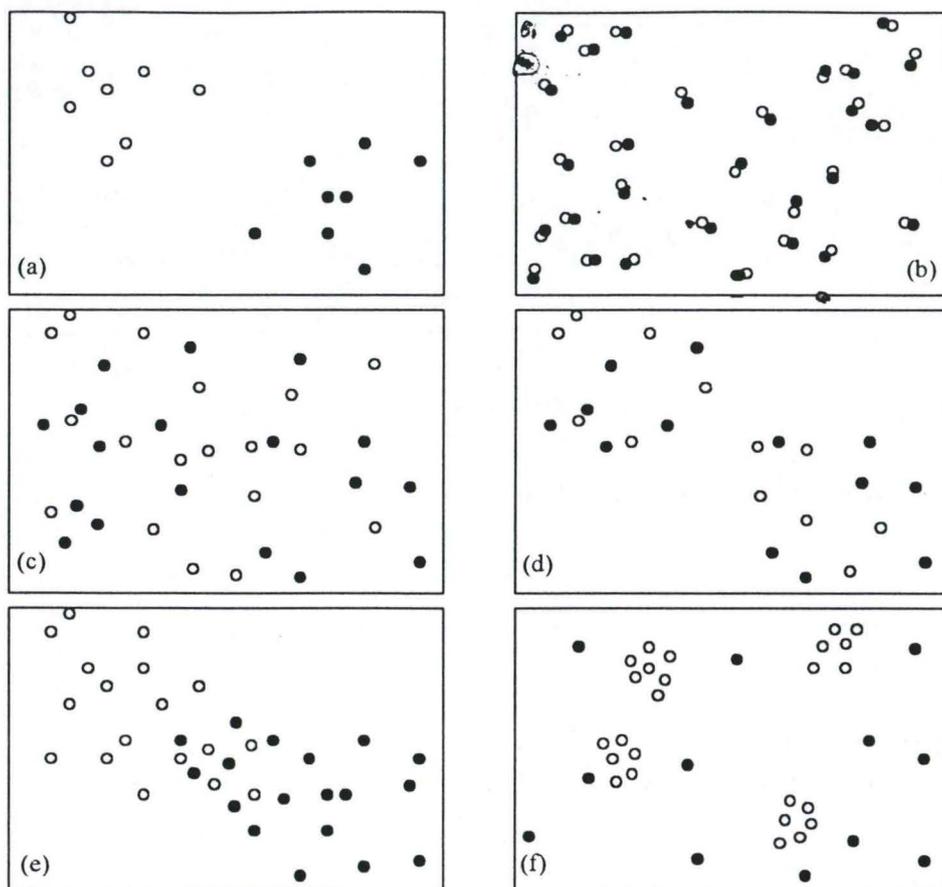


Figure 2.2: Exemples de répartitions bivariées

n'y a alors même plus lieu de parler de mélange de deux nuages de points, car il n'y a plus qu'un seul nuage de points. Un mélange homogène des deux nuages de points correspond aux figures 2.2c et d. Des situations intermédiaires seraient celles des figures 2.2e et f où le mélange est partiel, ou complet et inhomogène

2. Un autre point de vue, celui de l'*étiquetage aléatoire*, consiste à regarder dans quelle mesure la répartition bivariée observée s'écarte de la répartition bivariée que l'on observerait en laissant les individus à leur place mais en leur affectant leur type au hasard. Les figures 2.2c et d correspondent ainsi à un étiquetage aléatoire, tandis que que la figure 2.2f montre un lien entre la densité de la répartition globale (sans distinction de type) et le type noir ou blanc.

Ce point de vue prend tout son sens pour des situations dans lesquelles la localisation d'individus indistinguables survient en premier, puis un type leur est affecté dans un second temps par un autre mécanisme. Un exemple typique est celui de la contamination par une maladie, conduisant à distinguer des arbres infectés et des arbres non infectés.

3. Un troisième point de vue consiste à regarder s'il y a *indépendance* ou non des deux mécanismes qui ont donné naissance aux deux répartitions marginales. Ce troisième point de vue, contrairement aux deux précédents, s'attache donc à mettre en évidence un lien entre les deux nuages de points, indépendamment de la nature même de ces nuages de points. Lorsque la présence d'un point noir en (x, y) augmente (respectivement diminue) la probabilité de

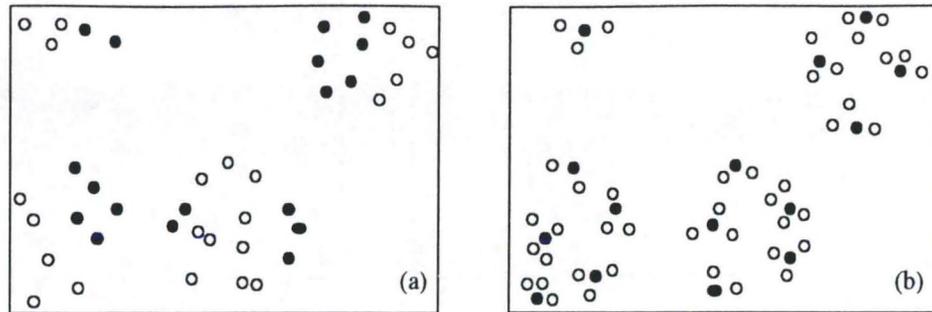


Figure 2.3: Exemple de répartitions bivariées

trouver un point blanc dans le voisinage de (x, y) et réciproquement, on parle d'association positive (respectivement négative) des deux répartitions.

La nature du lien entre deux répartitions dépend en fait là aussi de l'échelle d'observation. La figure 2.3a montre ainsi un cas d'association négative à grande échelle et positive à petite échelle, tandis que la figure 2.3b montre un cas d'association positive à la fois à grande et petite échelle.

2.2.2. Facteurs écologiques pouvant influencer le lien les répartitions de deux espèces

On distingue d'une part les interactions directes entre individus, d'autre part la "coïncidence écologique" [11].

Interactions directes entre individus des deux espèces

Les individus d'une espèce modifient localement l'environnement de sorte que cela a un effet direct sur l'occurrence de l'autre espèce. Une influence positive se traduira par une association positive des deux espèces, une influence négative (essentiellement la compétition) par une association négative. Si deux espèces interagissent directement, leurs répartitions ne seront pas indépendantes.

Coïncidence écologique

Il y a coïncidence écologique lorsque deux espèces répondent de façon indépendante à un même facteur écologique, qu'il soit physique ou biotique. Deux espèces qui répondent de la même façon et indépendamment l'une de l'autre à un même facteur (espèces convergentes) seront associées positivement ; deux espèces qui répondent de façons différentes et indépendamment l'une de l'autre à un même facteur (espèces divergentes) tendront à être au contraire négativement associées.

Le milieu physique est le facteur le plus courant qui entre dans cette catégorie [54]. Les figures 2.2a et e peuvent ainsi s'interpréter par un milieu hétérogène sur lequel se trouvent des espèces ayant des exigences écologiques opposées. Si les deux espèces ont des exigences semblables, mais que l'une tolère des écarts à l'optimum supérieurs, alors on s'attend à voir l'espèce la moins tolérante confinée à des zones restreintes tandis que l'espèce plus tolérante serait plus répandue, comme sur la figure 2.2f. Si les deux espèces ont des exigences identiques, alors elles partagent le même territoire, comme sur les figures 2.2c (milieu homogène) et d (milieu hétérogène).

Les facteurs de coïncidence écologique ont une incidence ambiguë sur l'indépendance des répartitions des deux espèces. Conditionnellement à la connaissance de ces facteurs, les deux répartitions sont indépendantes puisque chaque espèce répond indépendamment l'une de l'autre aux facteurs : chaque espèce aurait ainsi la même répartition en l'absence de l'autre espèce que la répartition qu'elle a effectivement en sa présence. Mais dans l'absolu, les répartitions des deux espèces apparaissent associées, positivement ou négativement. Il n'est en fait pas possible de savoir, à partir

de la simple observation d'une répartition bivariée, si elle est due à une interaction directe (par exemple compétition) ou à une coïncidence écologique (par exemple réponses différentes dans un milieu hétérogène).

2.3. La modélisation par les processus ponctuels

Après avoir discuté des répartitions spatiales et des facteurs écologiques qui les influencent, le problème se pose de formuler cela mathématiquement. La théorie des processus ponctuels fournit un cadre mathématique approprié. De façon non formelle, un processus ponctuel est un modèle stochastique qui contrôle les localisations d'événements dans un ensemble de \mathbb{R}^n . Quelques définitions et résultats sur les processus ponctuels sont rappelés en annexe B.

Un processus ponctuel définit donc le mécanisme qui génère le nuage de points, et permet de générer plusieurs nuages de points par un même mécanisme. Une répartition est une *réalisation* d'un processus ponctuel.

Treize processus ponctuels modélisant des situations biologiques ont été utilisés pour les simulations. Ils sont présentés en annexe E. Sauf indications contraires, les valeurs des paramètres utilisées pour les simulations sont celles indiquées en annexe. Parmi ces processus, seul le processus de Poisson homogène, qui modélise la répartition aléatoire, est incontournable et constitue un processus de référence. Les autres situations biologiques auraient pu être modélisées par d'autres processus que ceux qui ont été choisis, de façon finalement arbitraire. De plus ces processus sont excessivement simples : le lien entre la réalité des interactions biologiques et la modélisation par des processus ponctuel est en fait plus complexe. La modélisation de situations réelles passe souvent par la combinaison de plusieurs mécanismes, et aboutit à un modèle final qui ne correspond à aucun processus ponctuel simple.

2.3.1. Modélisation de la répartition de deux espèces qui n'interagissent pas dans un milieu homogène

Si deux espèces n'interagissent pas directement dans un habitat homogène, les répartitions spatiales de leurs individus se réalisent indépendamment l'une de l'autre. Les processus ponctuels bivariés modélisant de telles situations résultent donc de la superposition de deux processus marginaux indépendants, et homogènes dans la mesure où l'habitat l'est. Selon que les répartitions marginales sont aléatoires, régulières ou agrégées, six situations sont possibles :

| | | | | |
|----------|-----------|-----------------------|---------------------|-----------------------|
| espèce 1 | aléatoire | aléatoire × aléatoire | | |
| | agrégée | agrégée × aléatoire | agrégée × agrégée | |
| | régulière | régulière × aléatoire | régulière × agrégée | régulière × régulière |
| | | aléatoire | agrégée | régulière |
| | | | espèce 2 | |

La répartition aléatoire a été modélisée par un processus de Poisson homogène, la répartition agrégée par un processus de Neyman-Scott, et la répartition régulière par un processus d'inhibition séquentielle (voir annexe E).

2.3.2. Modélisation de la répartition de deux espèces qui interagissent dans un milieu homogène

Deux situations sont possibles, selon que les deux espèces sont en relation de compétition ou de commensalisme. Ces deux situations ont été modélisées par respectivement le processus d'inhibition séquentielle et par le processus des paires liées (voir annexe E). Ce sont des processus homogènes, dont les processus marginaux ne sont pas indépendants : il y a respectivement répulsion et attraction à grande échelle entre les processus marginaux.

2.3.3. Modélisation de la répartition de deux espèces qui n'interagissent pas directement dans un milieu inhomogène

Lorsque l'habitat est inhomogène, deux espèces convergentes du point de vue écologique vont apparaître positivement associées alors que deux espèces divergentes vont apparaître négativement associées. Conditionnellement à l'habitat, les répartitions des individus des deux espèces se font de manière indépendante, mais si l'habitat n'est pas connu, les deux répartitions apparaissent au contraire dépendantes.

La famille des processus ponctuels permettant de modéliser ces situations est celle des processus de Cox (voir annexe B). De façon non formelle, un processus de Cox est un processus de Poisson inhomogène dont l'intensité est elle-même aléatoire. Un processus de Cox peut être généré en tirant une intensité et en construisant le processus de Poisson inhomogène correspondant ; cette démarche peut s'interpréter en disant que la densité est la réponse de la forêt aux hétérogénéités du milieu. Une autre façon de procéder, strictement équivalente à la première du point de vue mathématique, consiste à générer un processus de Poisson homogène, puis à appliquer un champ de survie (voir processus éclairci en annexe). Cette démarche peut s'interpréter en disant qu'il y a mortalité pendant la phase juvénile des arbres, et que la mortalité est fonction du milieu.

Les processus de Cox liés sont tels que les processus marginaux ont des intensités proportionnelles ; ils modélisent les espèces convergentes. Les processus de Cox équilibrés sont tels que la somme des intensités des processus marginaux est constante ; ils modélisent les espèces divergentes. Deux processus de Cox liés ont été utilisés, l'un montrant de l'association positive à petite échelle, l'autre à moyenne échelle ; de même deux processus de Cox équilibrés ont été utilisés, l'un montrant de l'association négative à petite échelle et l'autre à moyenne échelle (voir annexe E).

Par ailleurs, un autre processus de Cox a été retenu, qui résulte de la composition d'un processus de Cox équilibré et d'un processus de Cox lié. Il en résulte un processus bivarié montrant de l'attraction à petite échelle et de la répulsion à grande échelle.

2.4. Et pour trois (ou plus) espèces ?

Le lien entre trois répartitions ne se ramène pas aux trois liens entre les répartitions prises deux à deux. Considérons par exemple le processus ponctuel trivarié suivant. Chacun des trois processus marginaux n'admet que deux réalisations possibles : une répartition à faible densité ou une répartition à forte densité. Le processus trivarié est alors défini de la façon suivante :

- le processus marginal 1 génère sa répartition à faible densité avec la probabilité $\frac{1}{2}$ et sa répartition à forte densité avec la probabilité $\frac{1}{2}$.
- le processus marginal 2 génère sa répartition à faible densité avec la probabilité $\frac{1}{2}$ et sa répartition à forte densité avec la probabilité $\frac{1}{2}$, indépendamment du processus marginal 1.
- le processus marginal 3 est fonction des deux premiers de la façon suivante

| Densité de la répartition marginale 3 | Densité de la répartition marginale 2 | | |
|---------------------------------------|---------------------------------------|--------|--------|
| | forte | faible | |
| Densité de la répartition marginale 1 | forte | forte | faible |
| | faible | faible | forte |

- On vérifie facilement que deux quelconques des processus marginaux de ce processus trivarié sont indépendants. Mais les trois processus marginaux ne sont bien sûr pas indépendants puisque n'importe lequel d'entre eux est fonction des deux autres.

Plus le nombre d'espèces augmente, plus la richesse des liens possibles entre leurs répartitions augmente. On s'intéresse par la suite essentiellement au lien entre deux espèces (un exemple d'analyse de l'interaction entre trois espèces sera simplement traité en partie III). D'autres méthodes, un autre vocabulaire devraient être employées pour étudier le lien entre davantage d'espèces.

3. ETUDE D'INDICES DE DISSIMILARITÉ

3.1. Démarche

Après avoir discuté des façons de décrire une répartition spatiale, nous en venons maintenant au problème de définir un indice de dissimilarité entre deux répartitions. Diverses méthodes ont été proposées. Huit d'entre elles ont été retenues et sont étudiées dans cette partie. Les indices qui sont utilisés ici ne sont pas à proprement parler des indices de dissimilarité, dans le sens mathématique du terme (voir annexe A). Néanmoins, comme il est très facile de passer des indices tels qu'ils sont définis ici à des indices de dissimilarité, nous conservons abusivement cette appellation.

Dans le cadre de la théorie des processus ponctuels, une répartition bivariée n'est qu'une réalisation d'un processus aléatoire. Un indice de dissimilarité défini sur cette répartition bivariée peut donc être considéré comme une *statistique*, dont la distribution dépend de la nature du processus ponctuel sous-jacent. Le calcul de cette distribution étant compliqué, même pour des processus ponctuels simples, on aura recours à des simulations pour l'estimer.

Les distributions d'un indice estimées pour les treize processus ponctuels décrits précédemment constituent une signature de son comportement. Chaque indice de dissimilarité se comporte en effet d'une façon particulière, en fonction de la situation biologique, et les processus ponctuels permettent de faire le lien entre la situation biologique et le comportement de l'indice.

Il faut garder à l'esprit cependant que les treize processus ponctuels ont été choisis arbitrairement, et la vision du comportement de l'indice que l'on obtient est donc très partielle. Seul le processus bivarié obtenu par la superposition de deux processus de Poisson homogènes indépendants, qui constitue le processus de référence, permet de pousser l'analyse plus loin que la simple comparaison entre indices.

3.2. Présentation des indices

Les huit indices utilisés sont présentés rapidement dans le tableau ci-dessous, et plus en détail en annexe F. De plus d'autres méthodes, qui constituent le plus souvent des variantes ou des prolongements de ces huit méthodes, sont également évoquées en annexe.

Présentation des huit indices :

- 1 : ce que mesure l'indice
- 2 : information à collecter pour calculer l'indice
- 3 : principe

Indice des quadrats [26, 57, 64]

- 1 : l'écart à l'indépendance des répartitions marginales.
- 2 : présence / absence des espèces dans des quadrats.
- 3 : construction d'un tableau de contingence croisant l'absence ou la présence des deux espèces dans les quadrats. L'indice est défini comme le coefficient de corrélation de points sur ce tableau. Lorsqu'il y a indépendance, le coefficient de corrélation de points est nul.

Indice de Kershaw [38]

- 1 : l'écart à l'indépendance des répartitions marginales.
- 2 : nombre d'individus de chaque espèce sur une grille de quadrats contigus.
- 3 : estimation de $\text{Cov}(N_1(Q), N_2(Q))$, où $N_i(Q)$ est le nombre d'individus de l'espèce i présents dans le quadrat Q . Lorsqu'il y a indépendance, la covariance est nulle.

Indice de Diggle & Cox [17, 64, p.250]

- 1 : l'écart à l'indépendance des répartitions marginales.
- 2 : distance $X_i^{(k)}$ d'un point arbitraire au k^{eme} plus proche arbre de l'espèce i .
- 3 : mesure de la corrélation entre $X_1^{(k)}$ et $X_2^{(k)}$ à l'aide du coefficient de corrélation des rangs τ de Kendall. Lorsqu'il y a indépendance, la corrélation est nulle.

Indice de Goodall [9, 17, 22, 64]

- 1 : l'écart à l'indépendance des répartitions marginales.
- 2 : distance X_i d'un point arbitraire au plus proche arbre de l'espèce i , et distance Y_{ji} d'un arbre arbitraire de l'espèce j à son plus proche voisin de l'espèce i .
- 3 : comparaison de la distribution de X_i et de Y_{ji} à l'aide de la statistique U de Wilcoxon. Lorsqu'il y a indépendance, X_i et Y_{ji} sont identiquement distribués.

Indice des plus proches voisins

- 1 : l'écart à l'association.
- 2 : distance Y_{ij} d'un arbre arbitraire de l'espèce i à son plus proche voisin de l'espèce j .
- 3 : plus l'association entre les deux répartitions est forte, plus Y_{ij} s'approche de zéro. L'indice est défini comme une moyenne sur Y_{ij} .

Indice du Sup de l'Inf [5]

- 1 : l'écart à l'association.
- 2 : distance Y_{ij} d'un arbre arbitraire de l'espèce i à son plus proche voisin de l'espèce j .
- 3 : plus l'association entre les deux répartitions est forte, plus Y_{ij} s'approche de zéro. L'indice est défini comme le maximum des Y_{ij} .

Indice de Pielou [20, 45, 54, 64, p.242]

- 1 : l'écart à l'association.
- 2 : essence du plus proche voisin d'un arbre de base.
- 3 : construction d'un tableau de contingence croisant l'essence de l'arbre de base et celle de son plus proche voisin. L'indice est défini comme le coefficient de corrélation de points sur ce tableau. Plus l'association entre les deux répartitions est forte, plus le coefficient de corrélation de points s'approche de -1.

Indices de Ripley [16, 19, 47, 50]

- 1 : l'écart à l'indépendance des répartitions marginales (méthode 1), ou l'écart à l'étiquetage aléatoire (méthode 2).
- 2 : cartographie de l'emplacement des arbres.
- 3 : calcul de la fonction K de Ripley, qui renseigne sur le nombre d'arbres se trouvant à une distance inférieure ou égale à r d'un arbre arbitraire. La fonction K peut être calculée de façon théorique sous l'hypothèse d'indépendance (méthode 1) ou sous l'hypothèse d'étiquetage aléatoire (méthode 2). L'indice mesure l'écart entre la courbe théorique et la courbe calculée.

Ce tableau met en relief que les divers indices ne répondent pas aux mêmes questions :

- Certaines indices mesurent un écart à l'indépendance entre les répartitions marginales : ces indices ont tous une origine statistique. La méthode des quadrats, la méthode 1 de Ripley et les méthodes de Goodall, Diggle & Cox et Kershaw entrent dans cette catégorie. Les quatre premières correspondent à un test de l'indépendance des répartitions marginales, tandis que la méthode de Kershaw a été développée à l'origine pour repérer les échelles de structure.
- D'autres indices mesurent un écart à l'association entre les répartitions marginales. Parmi eux, l'indice de Pielou a une origine statistique, tandis que les méthodes du plus proche voisin et du Sup de l'Inf correspondent à une approche géométrique.
- La méthode 2 de Ripley enfin mesure un écart à l'étiquetage aléatoire.

Les indices peuvent être également classés selon la nature de l'information à collecter pour les calculer. L'annexe F présente les huit indices suivant ce classement.

3.3. Résultats des simulations

Les huit indices de dissimilarité ont été calculés pour des réalisations des treize processus ponctuels définis à la partie I. Chaque processus ponctuel a en fait été simulé 20 fois, ce qui donne une fourchette de valeurs de chaque indice pour chaque processus. De plus, pour trois des processus ponctuels, les indices ont été calculés en fonction de l'intensité des processus, et en fonction de la différence d'intensité entre les deux processus marginaux. Les résultats des calculs sont présentés sur les figures 3.1 à 3.29.

Ces simulations renseignent sur le comportement des indices et permet dans une certaine mesure la comparaison entre les indices.

3.3.1. Indices mesurant un écart à l'indépendance

Les cinq indices mesurant un écart à l'indépendance des processus marginaux placent comme on s'y attend au même niveau tous les processus avec indépendance des processus marginaux (c'est-à-dire les six premiers processus sur les figures).

La **méthode de Goodall** (fig. 3.1 à 3.3) donne les plus mauvais résultats, car les amplitudes de variation de l'indice sont telles que tous les types de processus se chevauchent. Cette méthode ne distingue relativement bien que les processus homogènes avec interactions perceptibles à *grande échelle* (Pa et In). Au contraire, le processus ST, dont la structure est perceptible à petite échelle, ne se détache pas des processus avec indépendance des processus marginaux. La méthode de Goodall est la seule à ne pas distinguer ce processus avec répulsion si grossier, alors que c'est en général ce processus qui ressort le mieux par les autres méthodes.

L'amplitude des variations de l'indice de Goodall peut s'expliquer par le fait que la méthode repose sur un échantillonnage partiel. On retrouve ce phénomène chez les autres méthodes avec échantillonnage partiel, notamment chez la méthode de Diggle & Cox et celle des plus proches voisins. Une façon de diminuer les amplitudes de variation consiste à augmenter le nombre de points d'échantillonnage (fig. F.6), ou à tenir compte des k plus proches voisins comme cela a été fait pour la méthode de Diggle & Cox.

La **méthode de Diggle & Cox** (fig. 3.4 à 3.6) présente le même défaut que la méthode de Goodall quand on ne tient compte que du plus proche voisin, mais s'améliore grandement quand on prend en compte les quatre plus proches voisins (fig. 3.7 à 3.9). Il y a alors nette séparation de trois blocs de processus : les processus avec indépendance des processus marginaux (les six premiers sur la figure), les processus avec répulsion des processus marginaux (ST, CE, In), et les processus avec attraction des processus marginaux (MV, CL, Pa). A l'intérieur de chaque bloc cependant, les processus sont indistingables, c'est-à-dire que la méthode ne distingue pas les échelles de structure.

Si au lieu de prendre les quatre plus proches voisins on prend les plus proches voisins dans quatre secteurs du plan, les résultats sont moins bons.

La **méthode par présence / absence dans des quadrats** (fig. 3.10 à 3.12) présente un spectre très semblable à celui de la méthode de Diggle & Cox avec prise en compte des quatre plus proches voisins.

La **méthode de Kershaw** (fig. 3.13), à l'opposé de la méthode de Goodall, détache le mieux les processus dont la structure est perceptible à petite échelle (ST, CL). Cette méthode a l'inconvénient de présenter des "accidents", comme on le voit pour le processus AA, ce qui est lié aux variations considérables que peut présenter la courbe covariance versus taille de quadrat d'une réalisation d'un processus à une autre. Cet indice présente de plus l'inconvénient d'être sensible à l'intensité du processus (fig. 3.14) et à l'écart en intensité entre les processus marginaux (fig. 3.15).

La **méthode 1 de Ripley** (fig. 3.16 à 3.18), tout comme celle de Kershaw, détache le mieux les processus dont la structure est perceptible à petite échelle (ST, MV). Cette méthode se distingue des autres par les faibles amplitudes de variation de l'indice.

3.3.2. Indices mesurant un écart à l'association

La différence de comportement la plus frappante entre les indices mesurant un écart à l'indépendance et les trois indices mesurant un écart à l'association est que ces derniers ne mettent pas sur le même niveau les six processus avec indépendance des processus marginaux.

La **méthode de Pielou** (fig. 3.19 à 3.21) n'est sensible qu'à ce qui se passe à grande échelle. Ainsi les processus MV et CL ne se distinguent pas du processus PP, puisqu'à l'intérieur des patches que dessinent ces deux processus la répartition est aléatoire. Les six processus avec indépendance des processus marginaux sont bien séparés les uns des autres. En particulier le processus RR donne une valeur faible de l'indice, ce qu'expliquent Upton & Fingleton [64, p.244].

La **méthode du Sup de l'Inf** (fig. 3.22 à 3.24) sépare les six processus avec indépendance des processus marginaux de façon moins nette que celle de Pielou. De plus cette méthode présente une grande sensibilité à l'intensité, du moins en ce qui concerne le processus ST.

La **méthode des plus proches voisins** (fig. 3.25 à 3.27), tout comme celle du Sup de l'Inf, sépare moins bien les six processus avec indépendance des processus marginaux que la méthode de Pielou, et présente une grande sensibilité à l'intensité et à l'écart en intensité pour le processus ST. En revanche elle sépare nettement le bloc des trois processus avec attraction des processus marginaux (MV, CL, Pa) du bloc des trois processus avec répulsion des processus marginaux (ST, CE, In).

Le fait d'échantillonner 10% des arbres (fig. 3.28) au lieu de les prendre tous en compte ne change pas la disposition générale des processus, mais élargit le domaine de variation de l'indice. Si on prend en compte les quatre plus proches voisins ou les quatre plus proches voisins dans quatre secteurs du plan tout en faisant un échantillonnage partiel, il n'y a pas d'amélioration notable.

3.3.3. Indices mesurant un écart à l'étiquetage aléatoire

La **méthode 2 de Ripley** (fig. 3.29) présente un spectre écrasé à cause du processus ST qui ressort exagérément.

3.3.4. Conclusion

La plupart des méthodes séparent bien le bloc des processus avec attraction des processus marginaux (MV, CL, Pa) du bloc des processus avec répulsion des processus marginaux (ST, CE, In). En revanche à l'intérieur d'un bloc, les indices séparent généralement mal les trois processus : l'échelle à laquelle est perceptible la dépendance (Pa et In correspondent à une grande échelle, CL et CE à une échelle moyenne, MV et ST à une petite échelle) a donc un rôle moindre que la nature de la dépendance, bien que certaines méthodes (particulièrement celle de Pielou) soient plus sensibles à ce qui se passe à grande échelle, et d'autres comme celles de Ripley, soient plus sensibles à ce qui se passe à petite échelle.

Il est alors intéressant de voir comment se comportent les indices vis-à-vis du processus Ce, qui présente une attraction à petite échelle et une répulsion à grande échelle. Seule la méthode de Pielou détache fortement ce processus comme un processus avec répulsion des processus marginaux, en accord avec le fait que cet indice n'est sensible qu'à ce qui se passe à grande échelle. Les autres indices présentent généralement une large fourchette de valeurs pour ce processus. Les indices mesurant un écart à l'indépendance mettent ce processus du côté des processus avec attraction des processus marginaux, tandis que les indices mesurant un écart à l'association le mettent plutôt du côté des processus avec répulsion des processus marginaux.

Par ailleurs les indices basés sur un échantillonnage partiel présentent toujours une amplitude de variation plus forte que ceux basés sur un échantillonnage total. On peut parler de "multiplicité statistique" pour les indices qui s'appuient sur la collecte de l'information en un grand nombre de points. Augmenter le nombre de points d'échantillonnage revient à augmenter la multiplicité statistique. On peut par ailleurs parler de "multiplicité d'échelles" pour les indices qui combinent l'information prise à différentes échelles. Prendre en compte les k plus proches voisins au lieu du plus proche voisin revient à augmenter la multiplicité d'échelles. Les méthodes les plus robustes

sont peut-être celles qui, comme les méthodes de Ripley, combinent la multiplicité statistique à la multiplicité d'échelles.

Figure 3.1

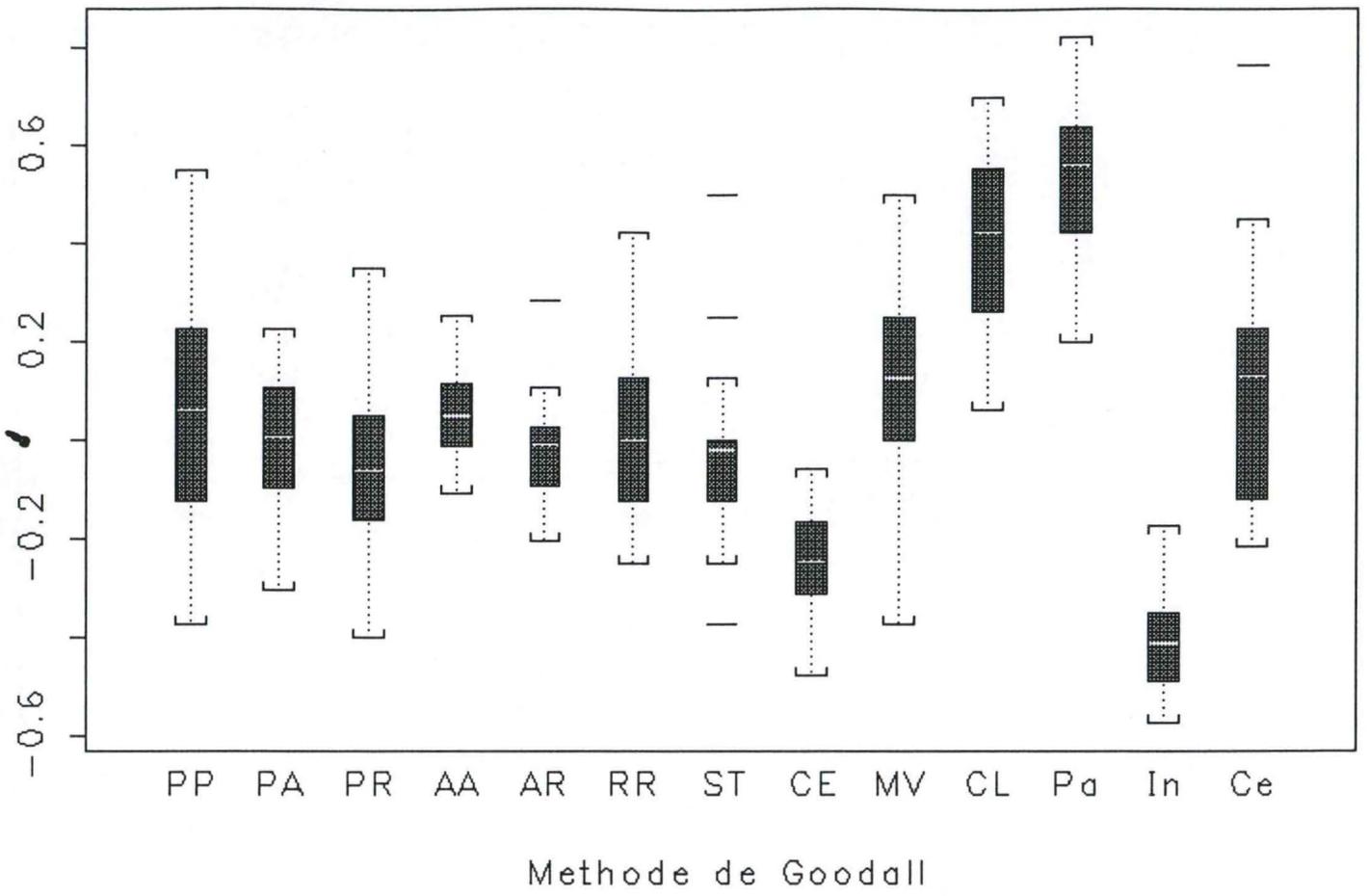


Figure 3.2

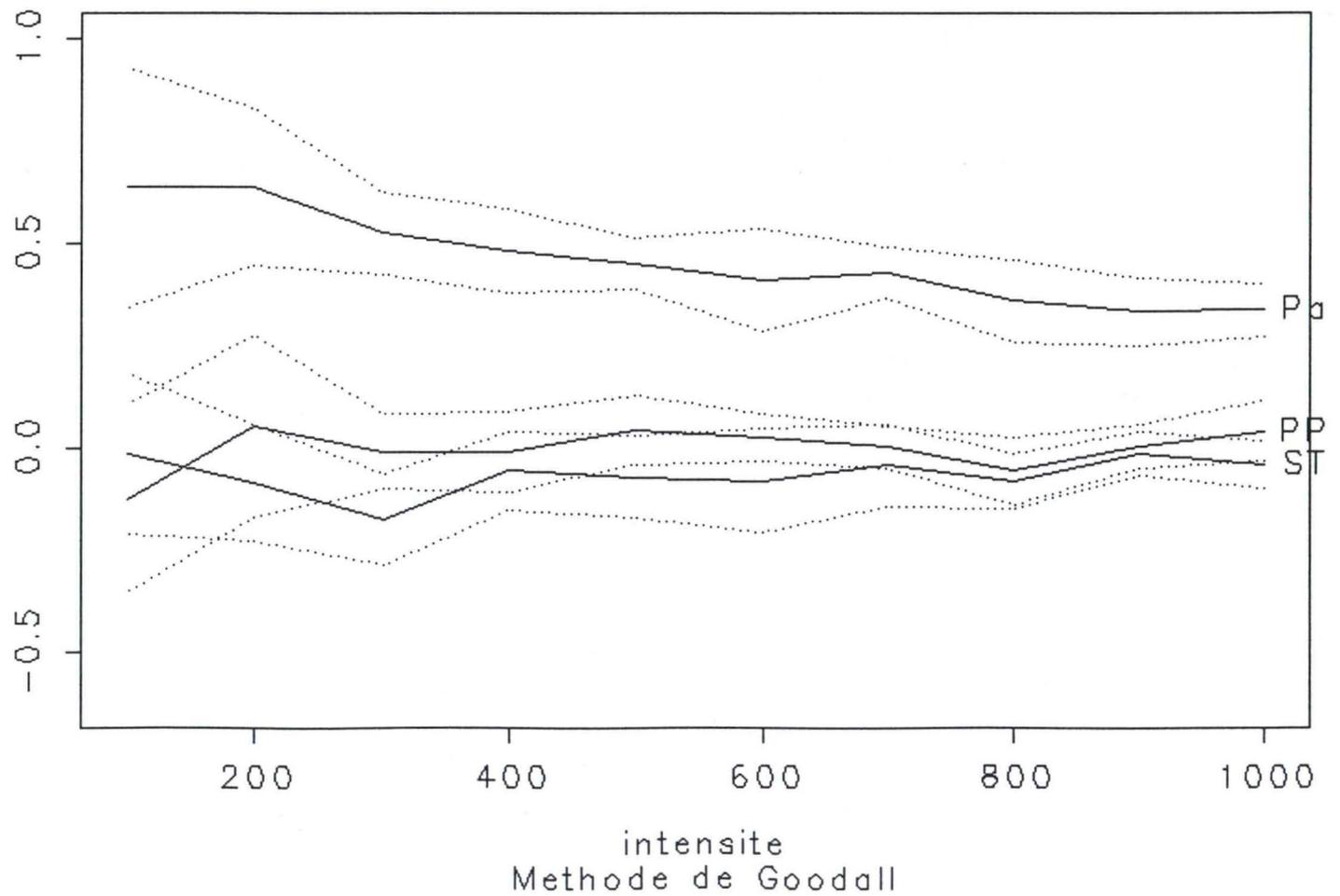
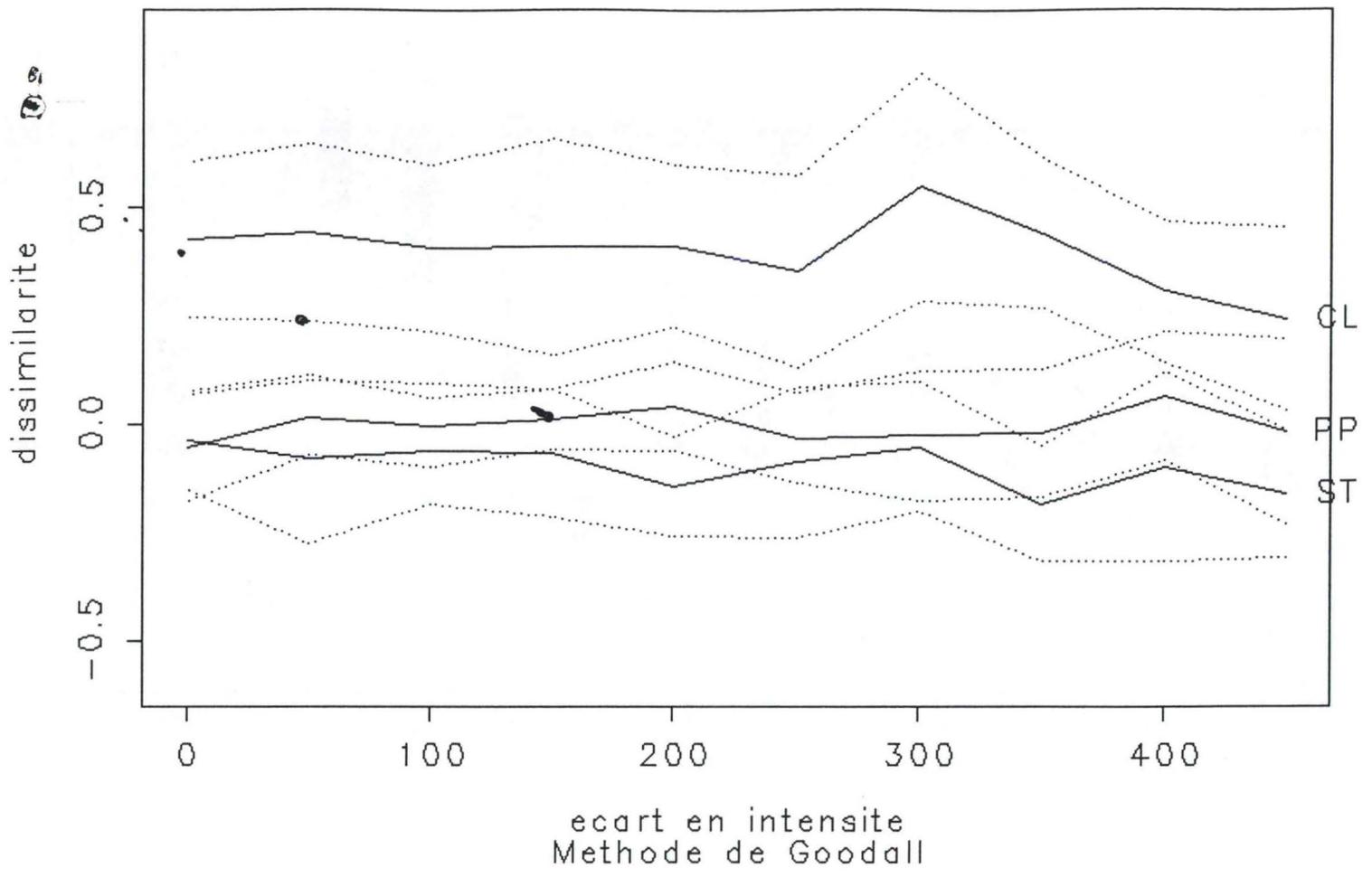
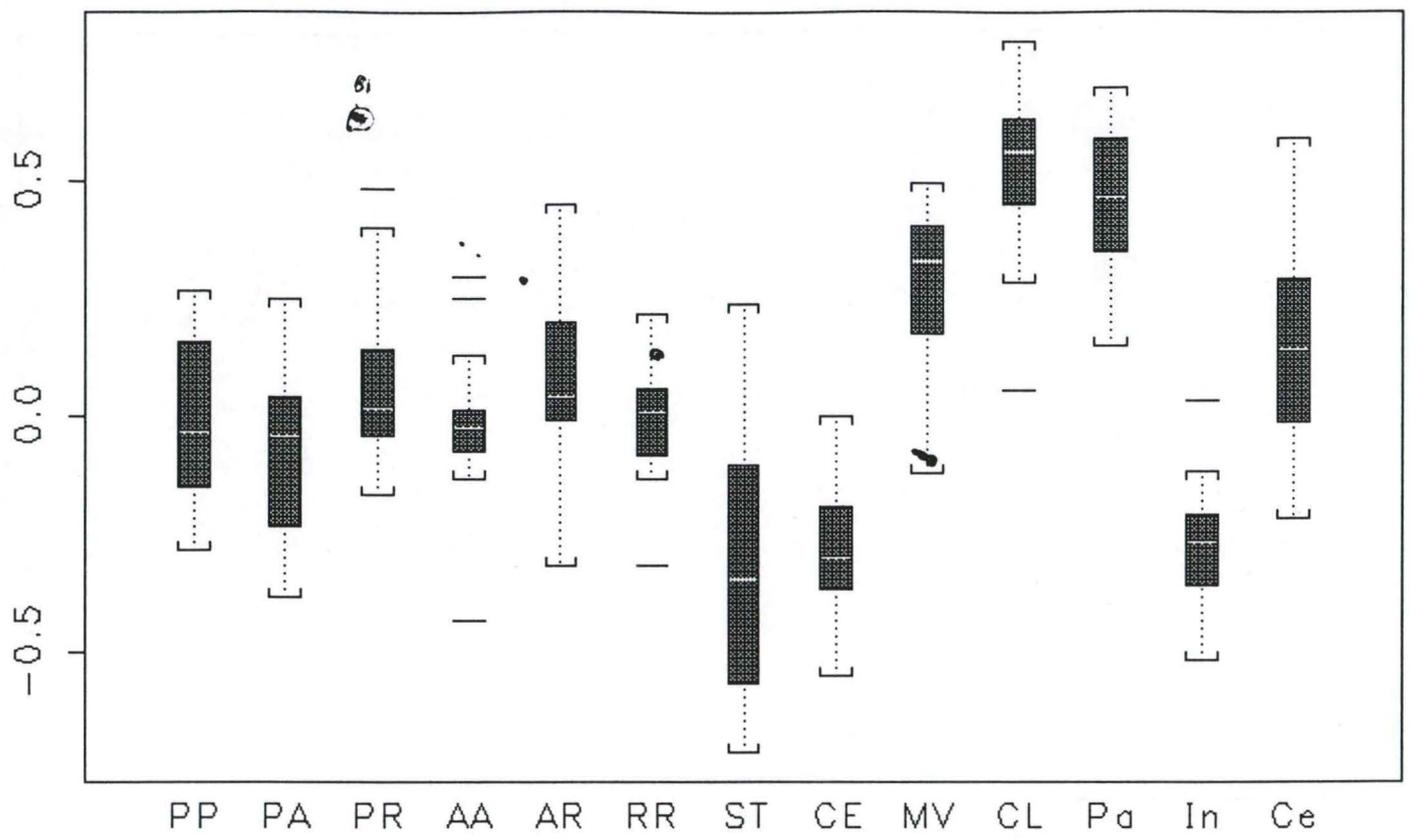


Figure 3.3



— moyenne sur 10 repetitions
..... ecart type

Figure 3.4



Methode de Diggle & Cox

Figure 3.5

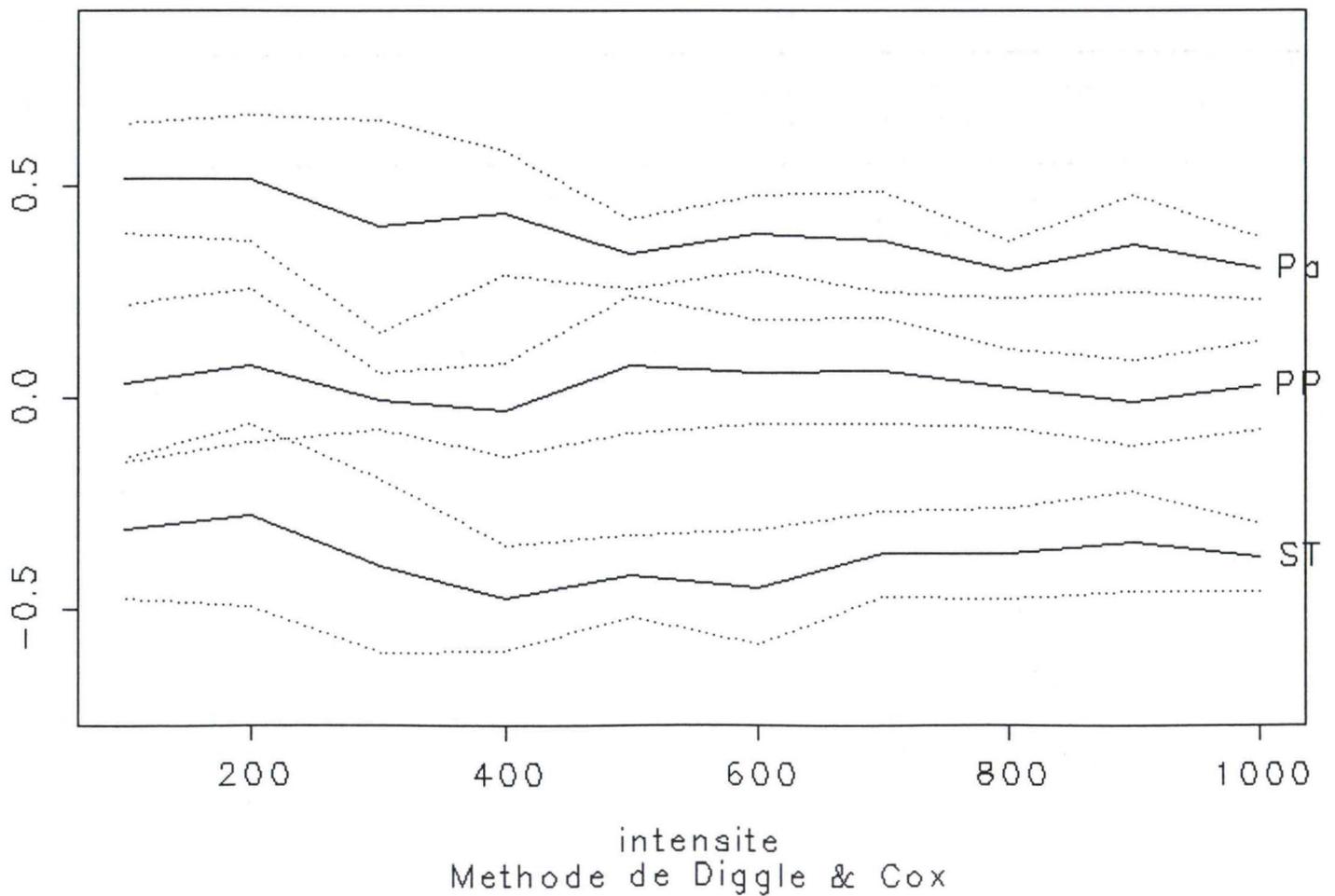
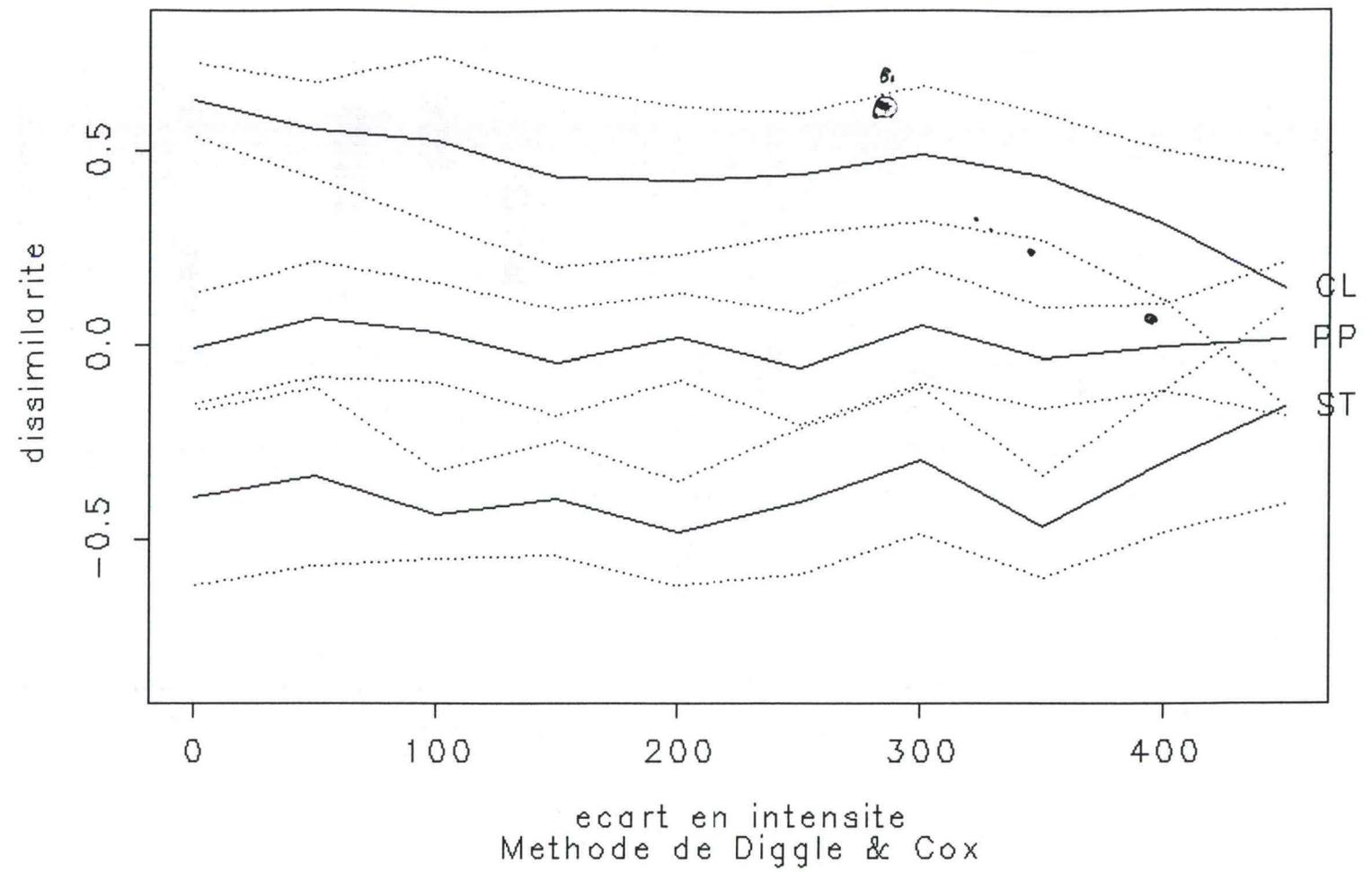
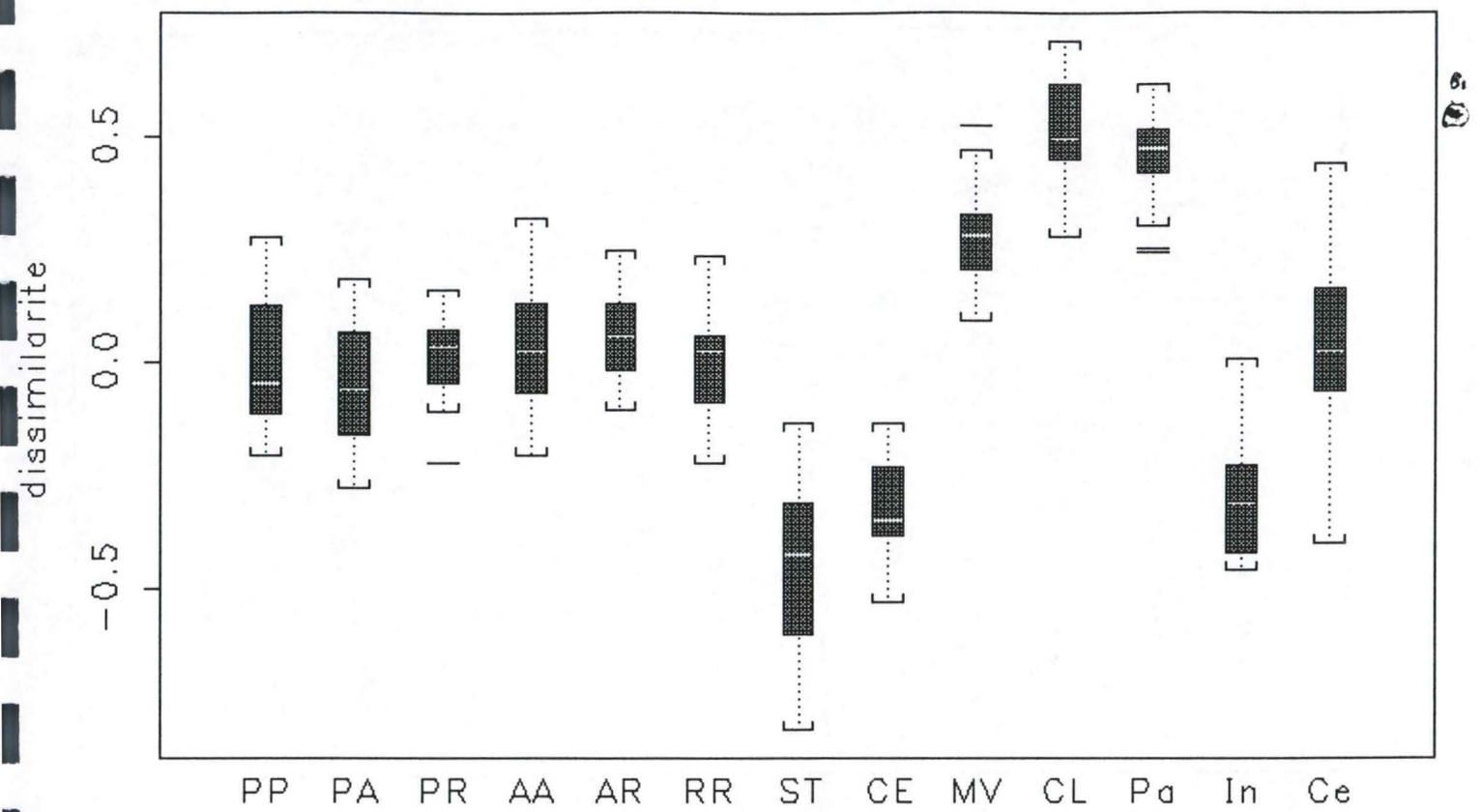


Figure 3.6



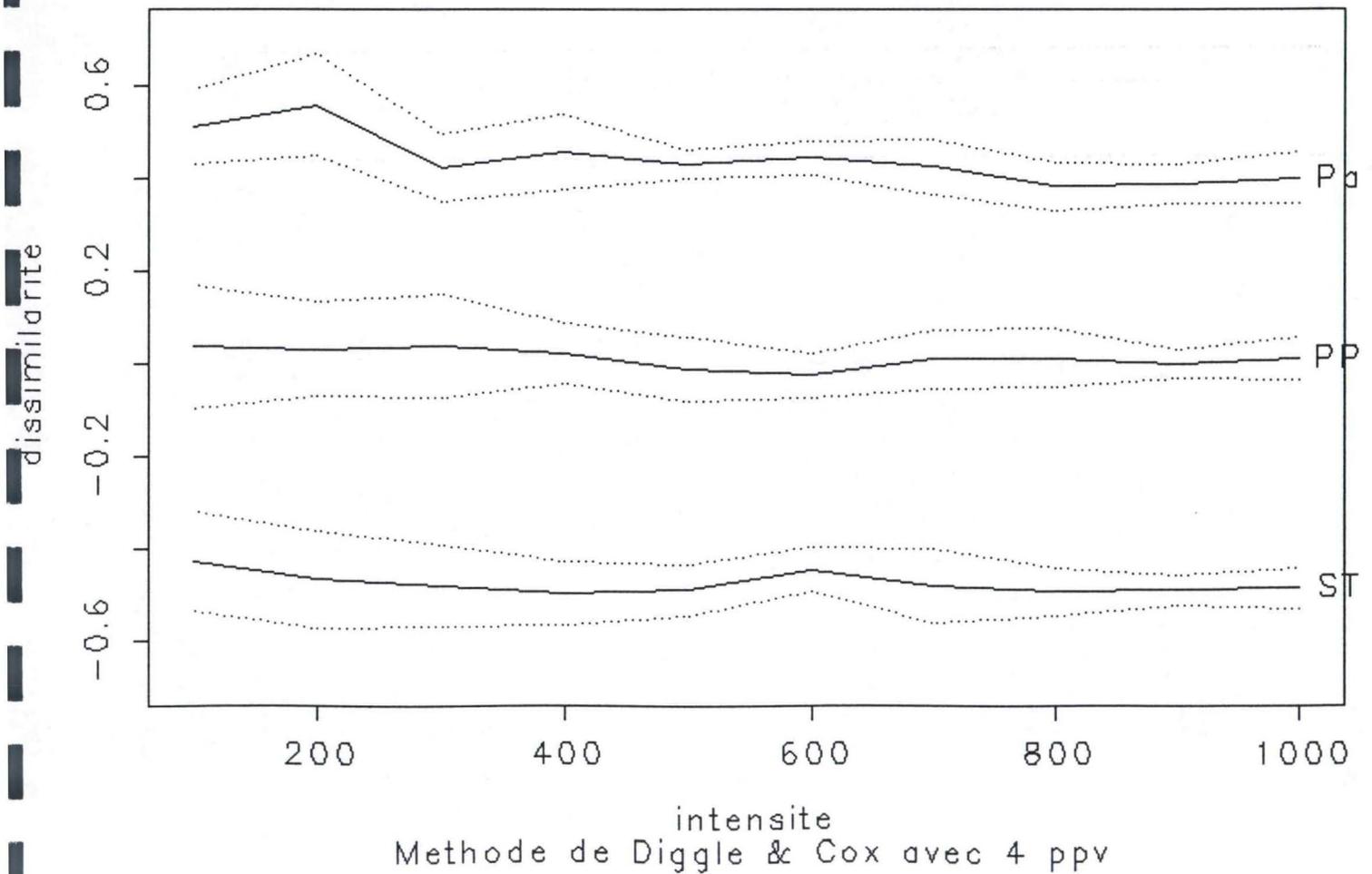
— moyenne sur 10 repetitions
..... ecart type

Figure 3.7



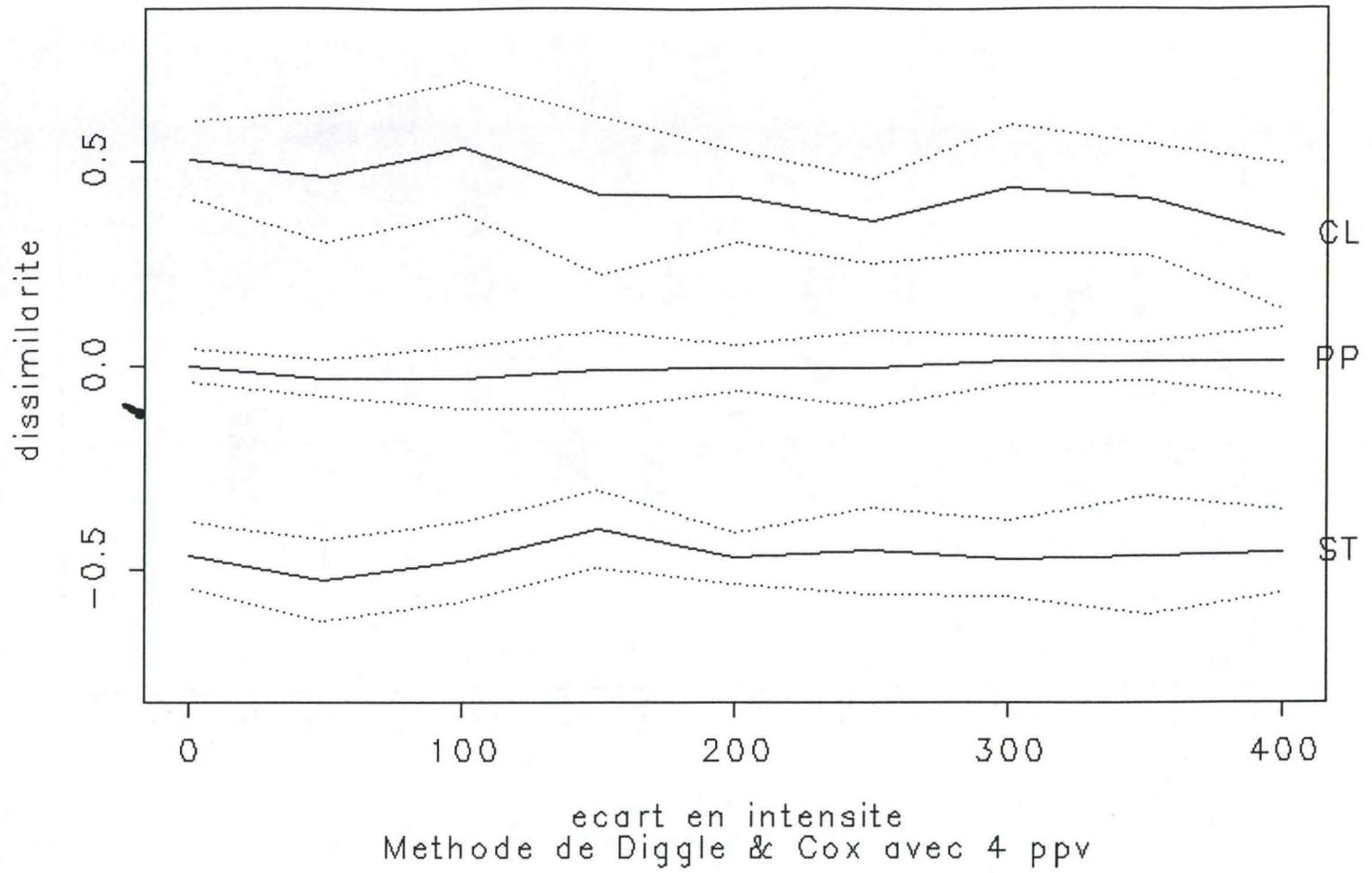
Methode de Diggle & Cox avec 4 ppv

Figure 3.8



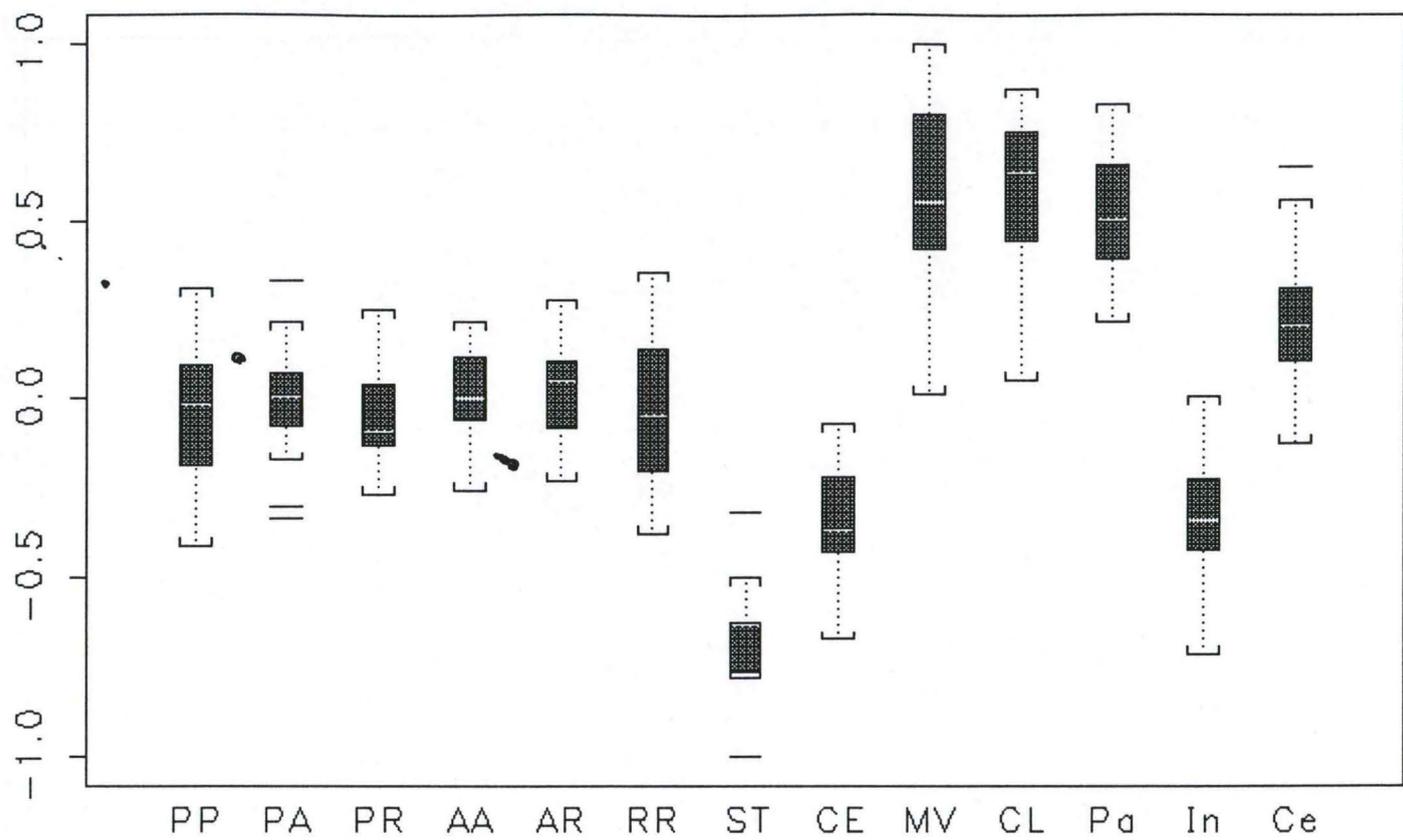
Methode de Diggle & Cox avec 4 ppv

Figure 3.9



— moyenne sur 10 repetitions
..... ecart type

Figure 3.10



Methode des quadrats

Figure 3.11

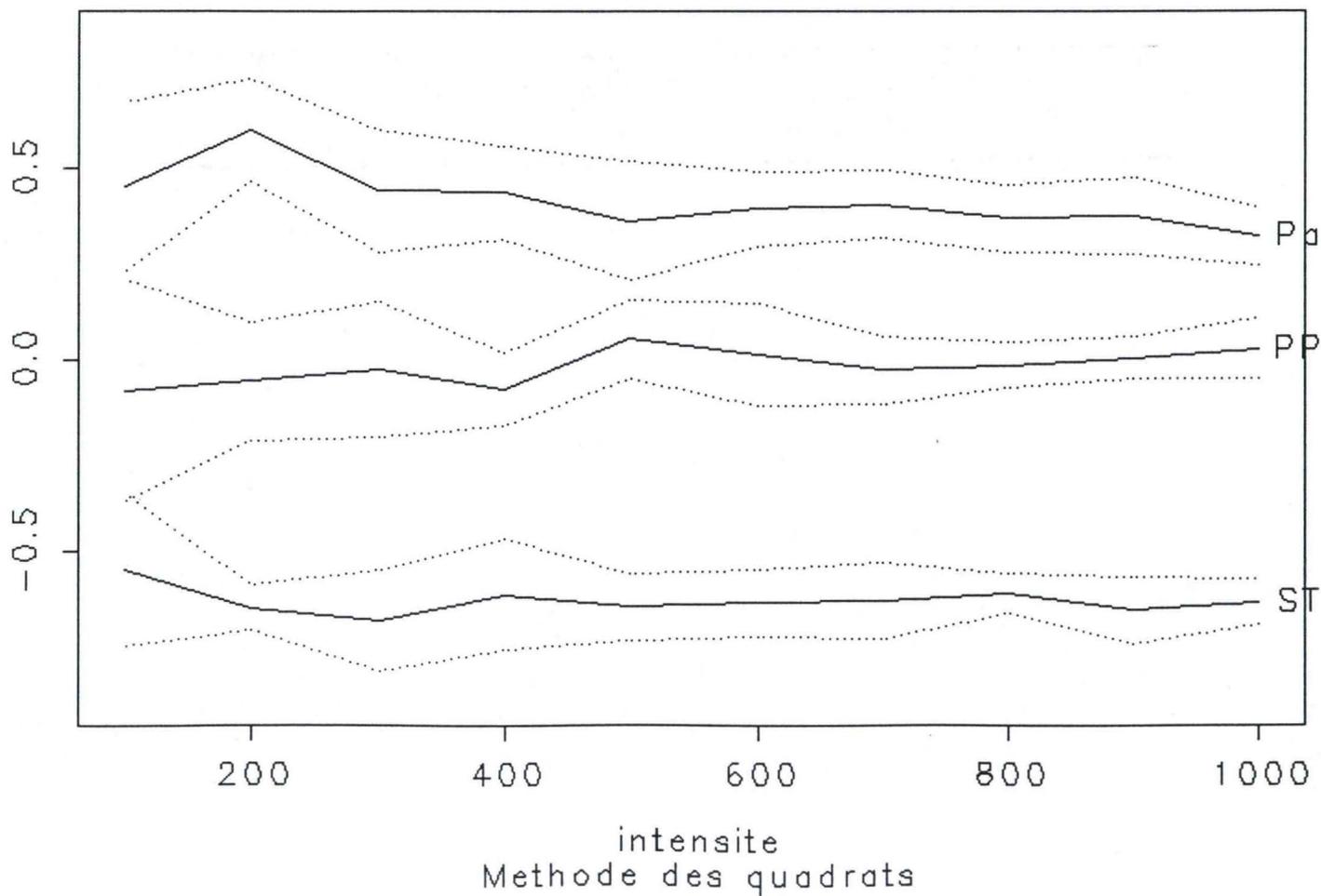
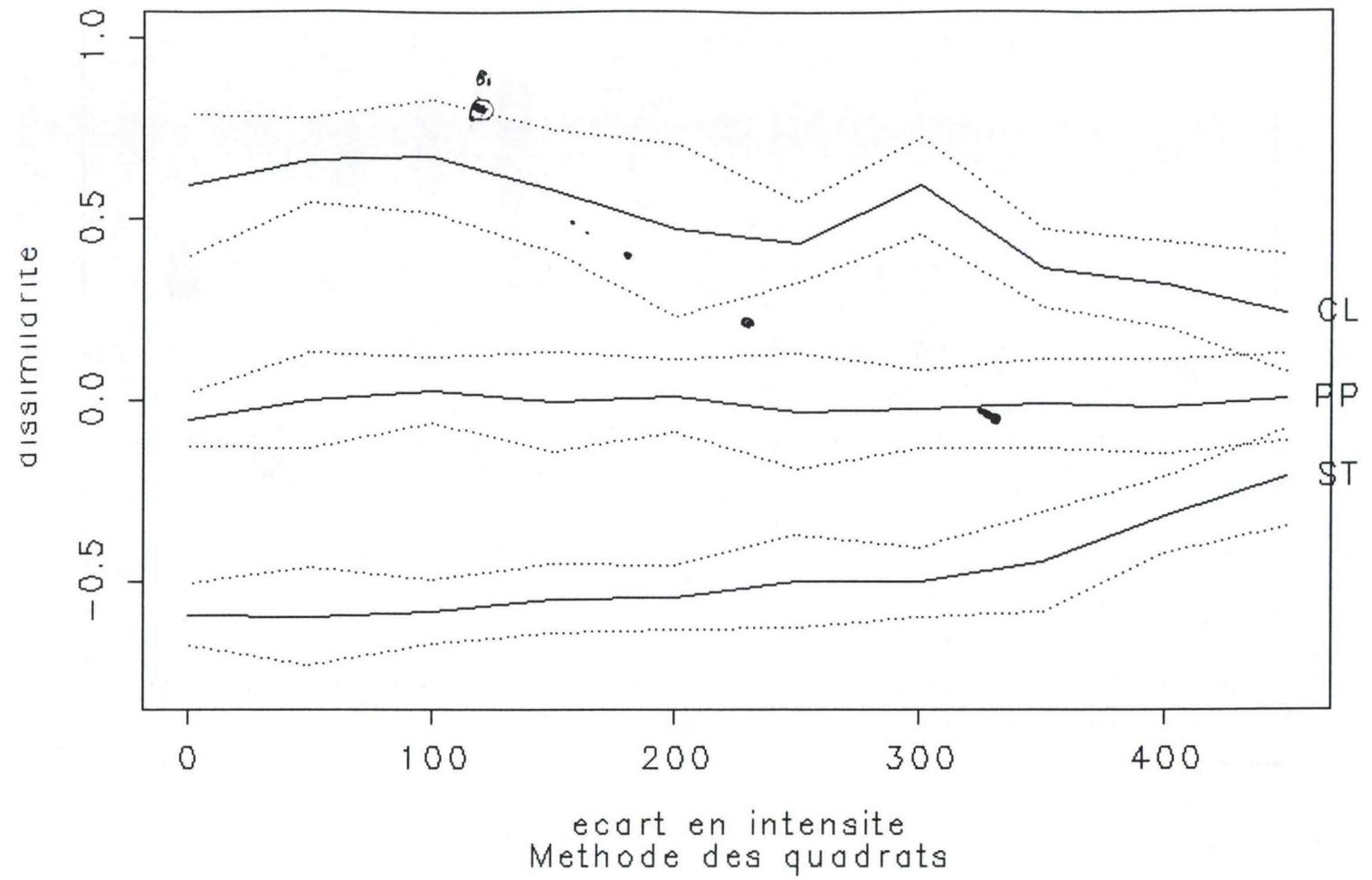
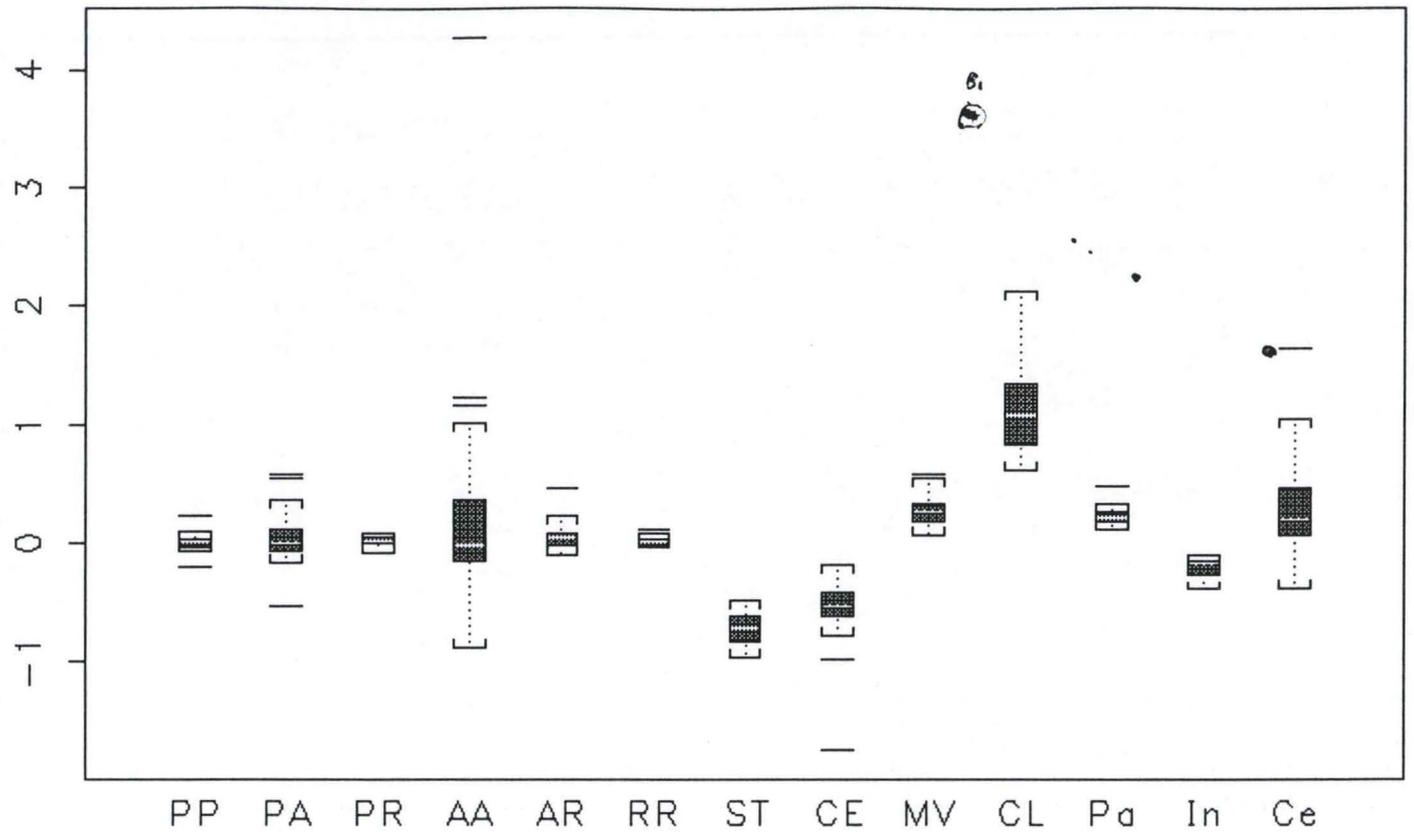


Figure 3.12



— moyenne sur 10 repetitions
..... ecart type

Figure 3.13



Methode de Kershaw

Figure 3.14

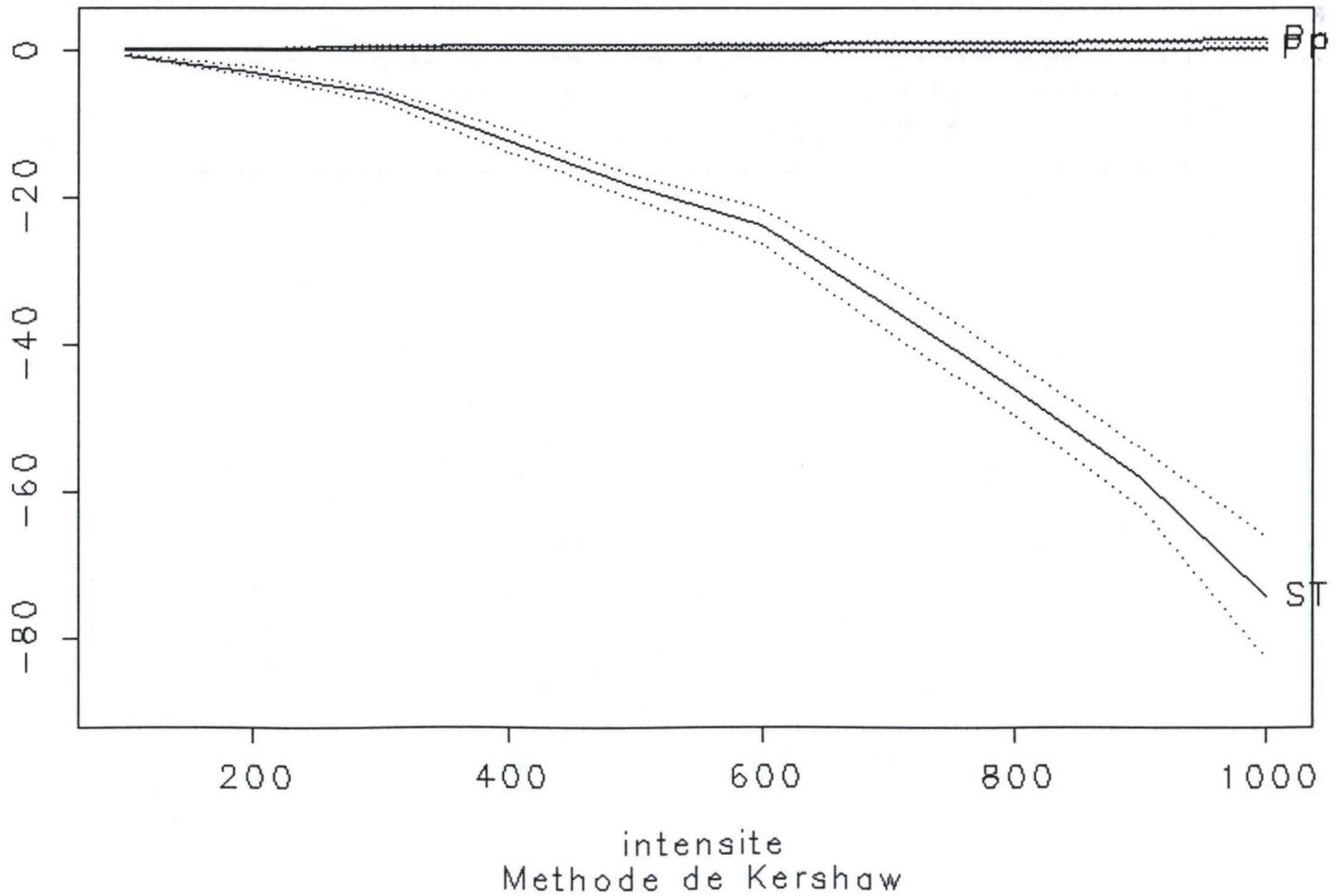
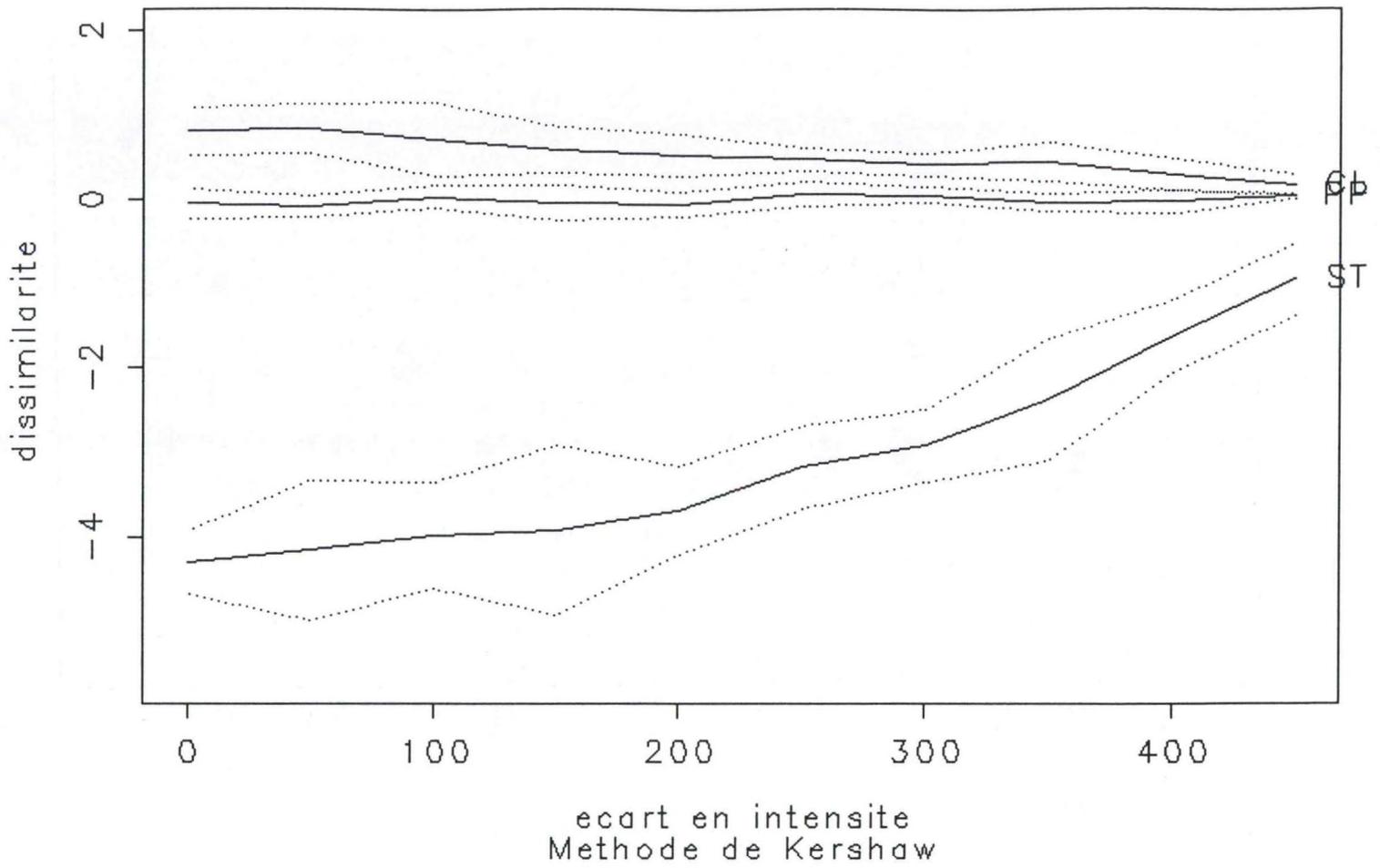
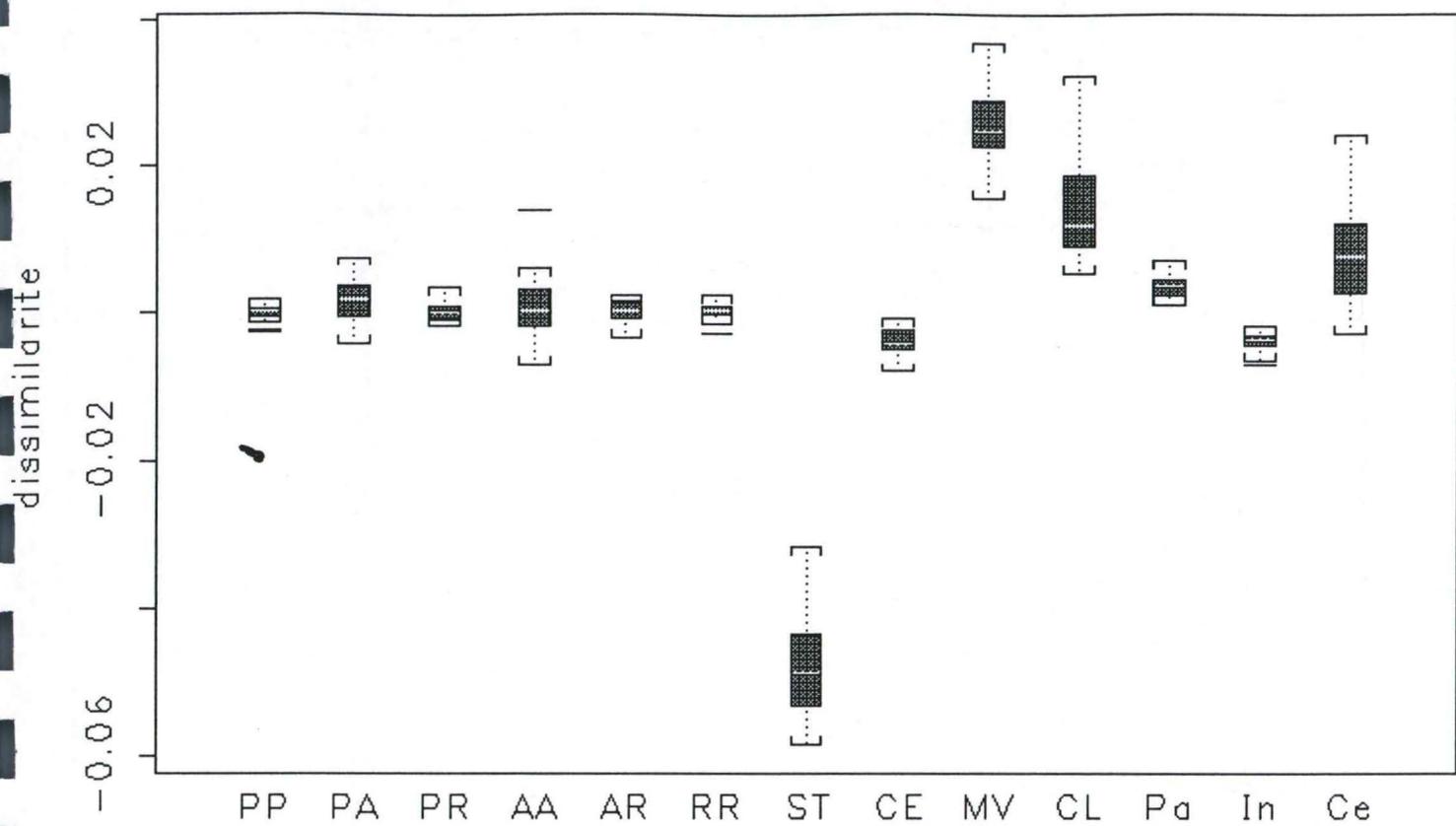


Figure 3.15



— moyenne sur 10 repetitions
..... ecart type

Figure 3.16



Methode 1 de Ripley

Figure 3.17

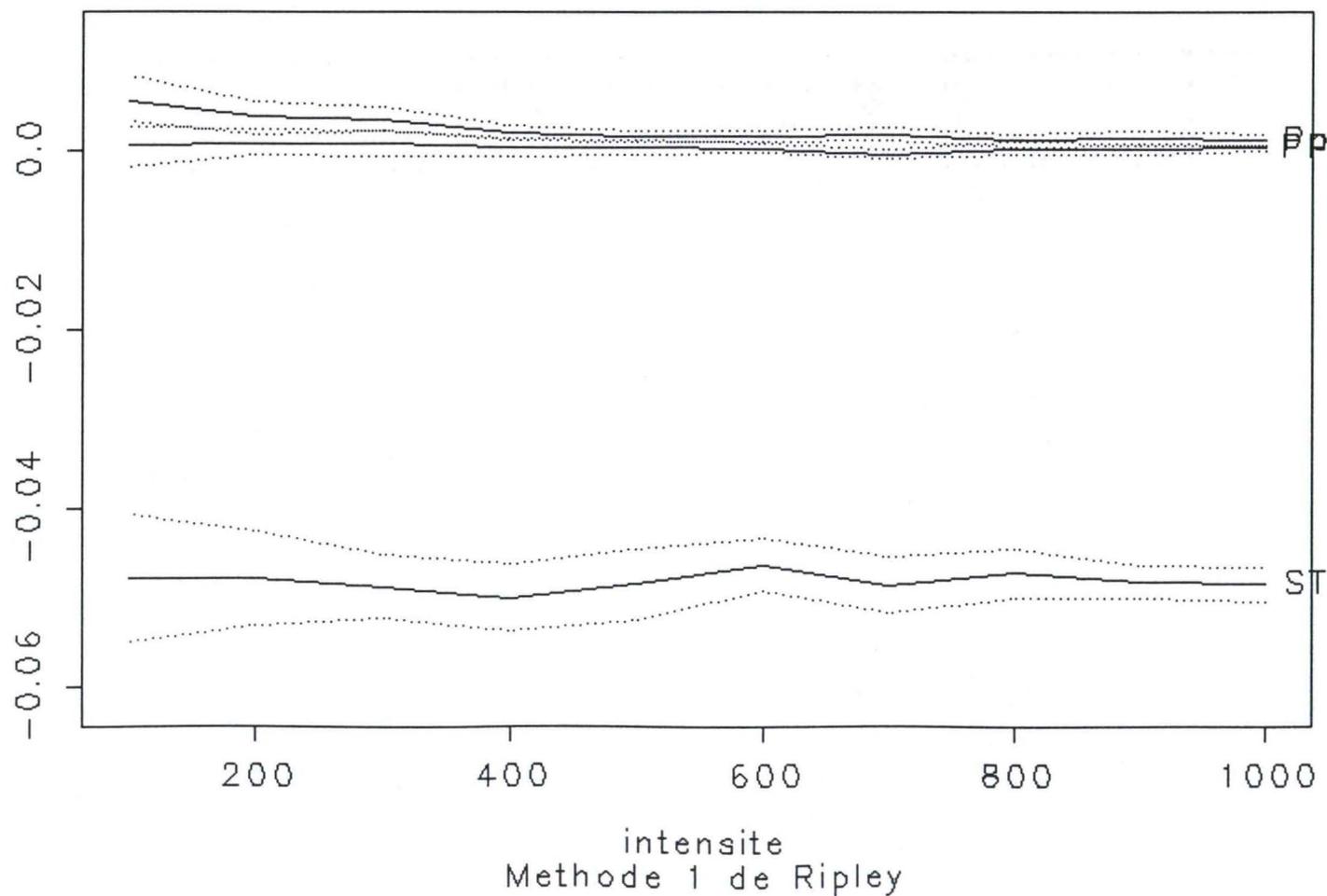
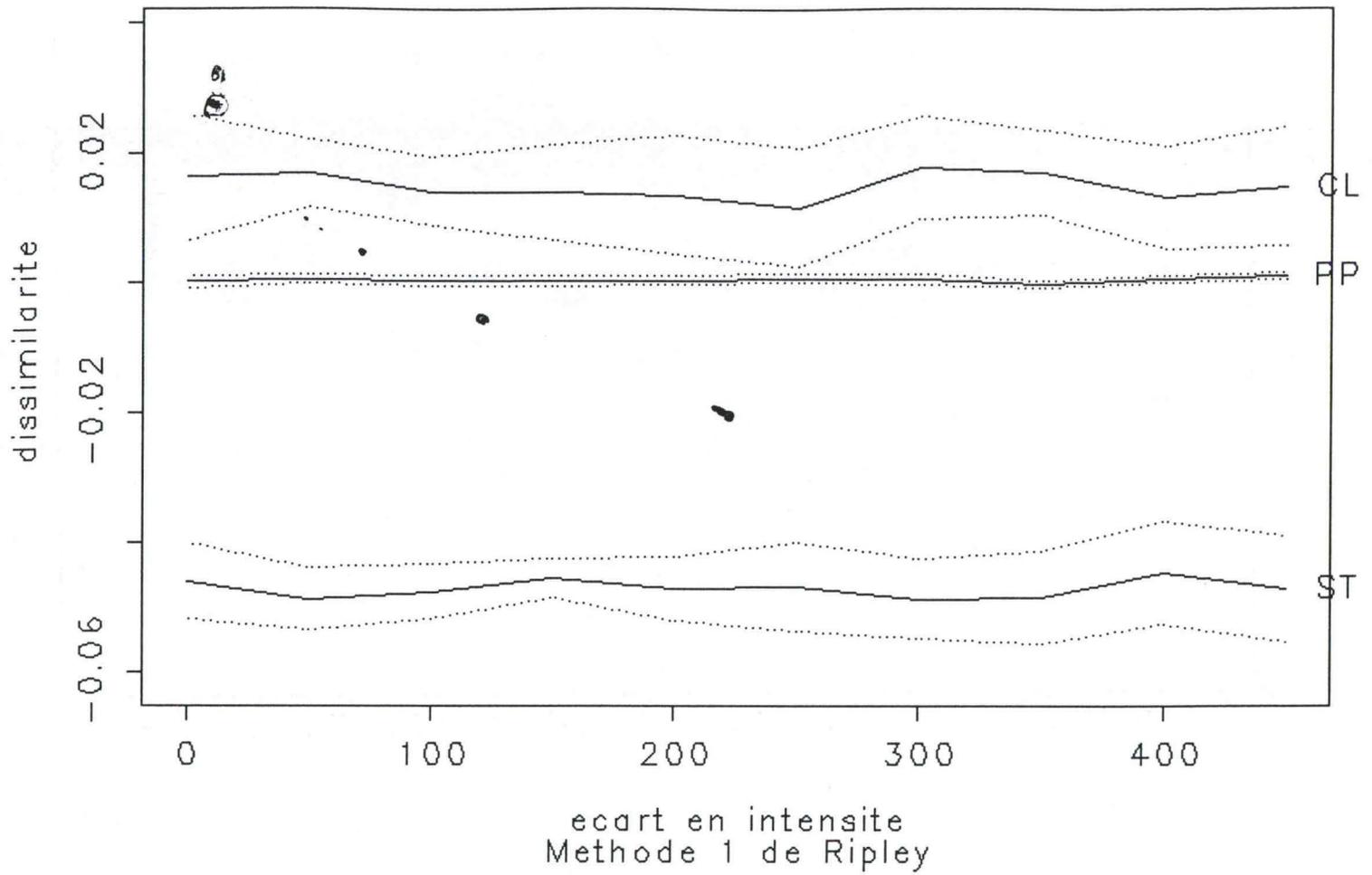
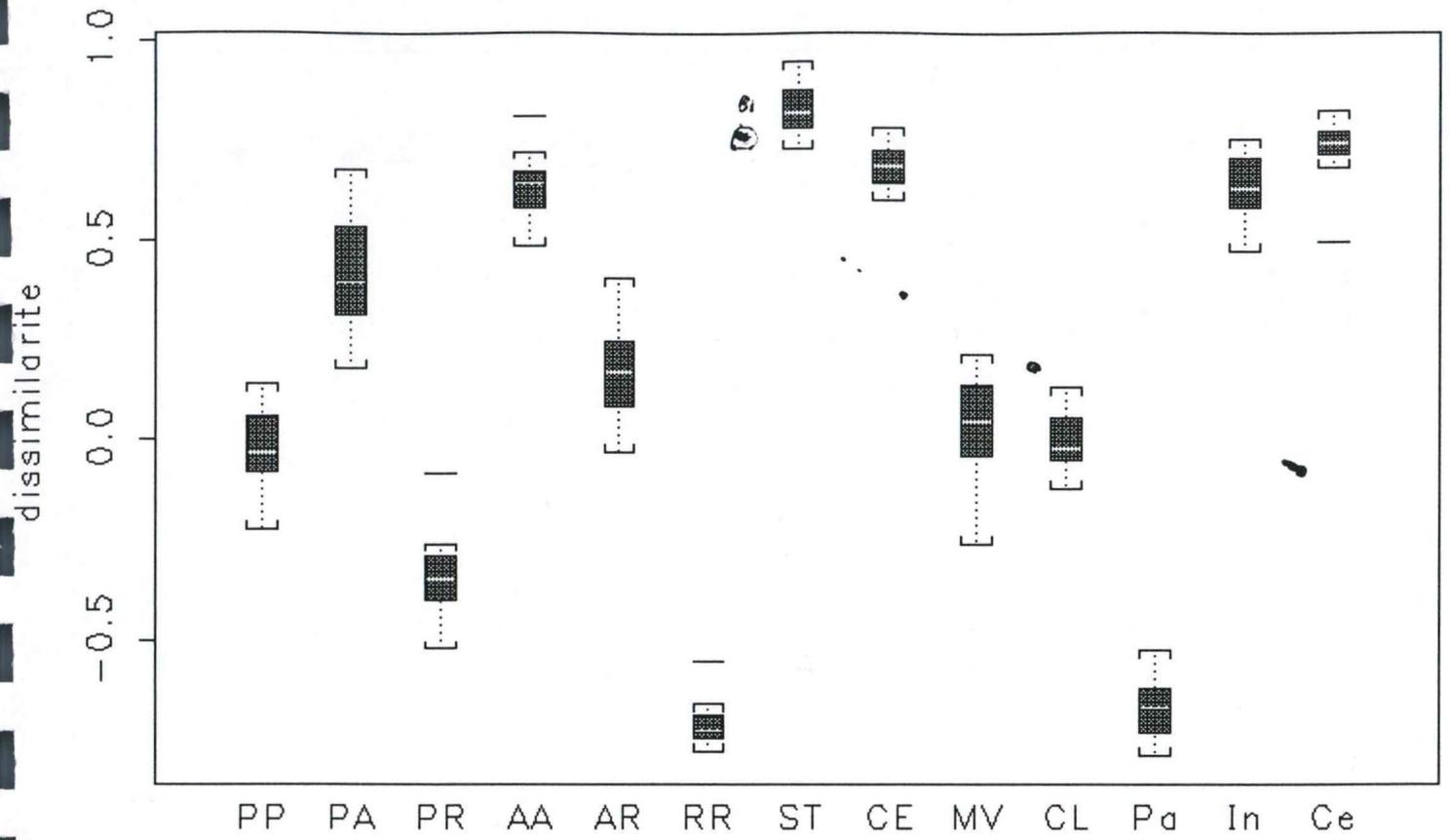


Figure 3.18



— moyenne sur 10 repetitions
..... ecart type

Figure 3.19



Methode de Pielou

Figure 3.20

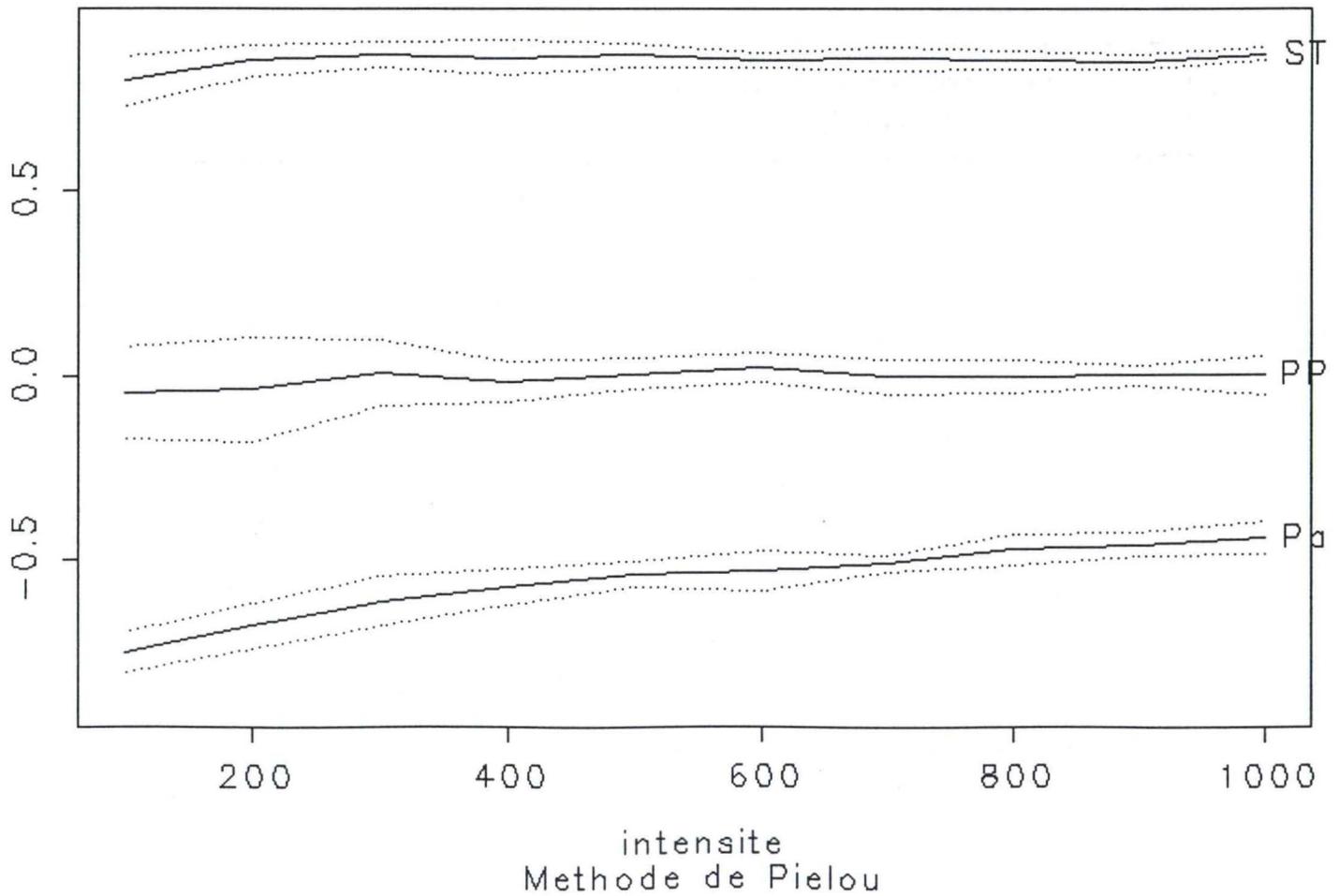
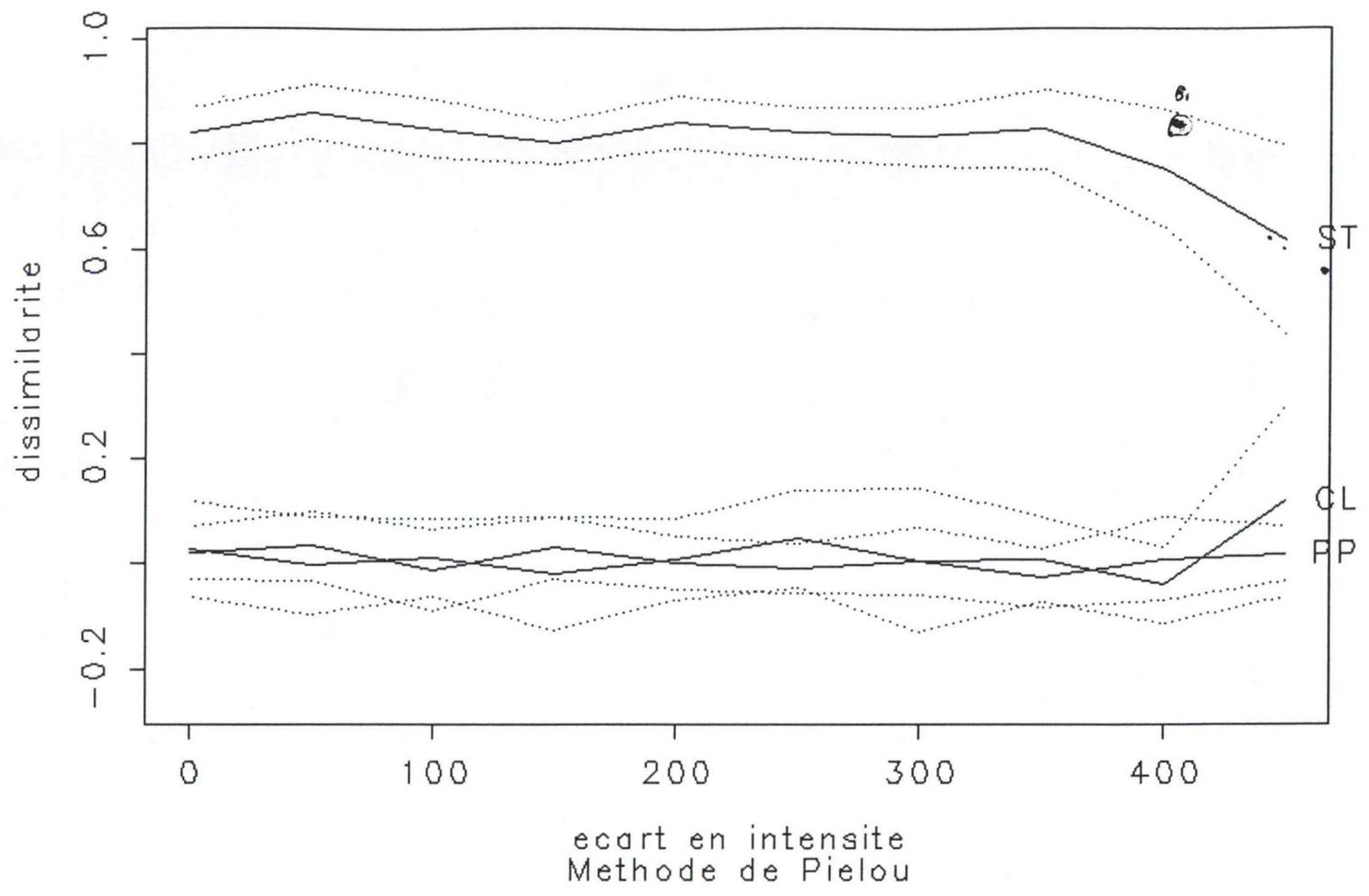
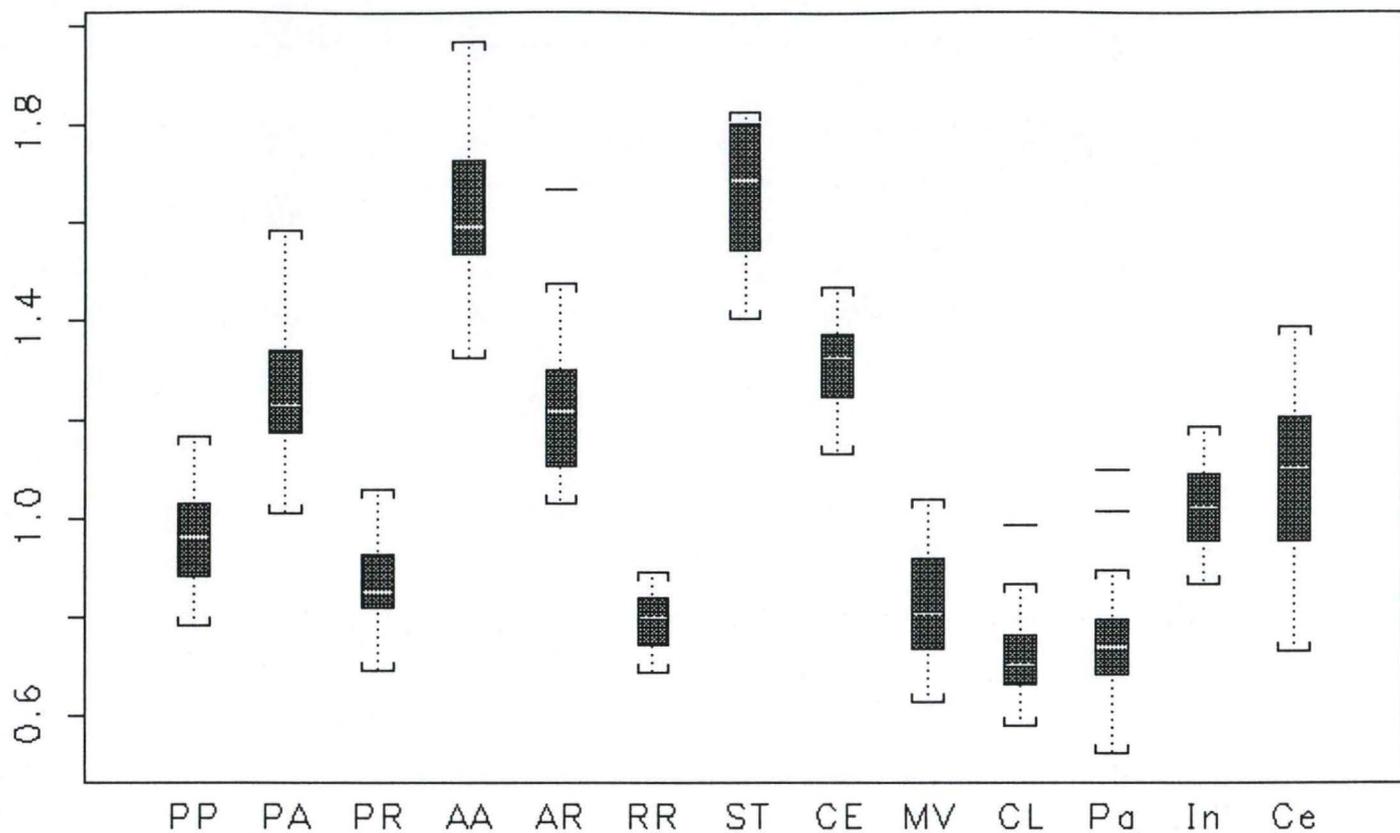


Figure 3.21



— moyenne sur 10 repetitions
..... ecart type

Figure 3.22



Methode du Sup de l'Inf

Figure 3.23

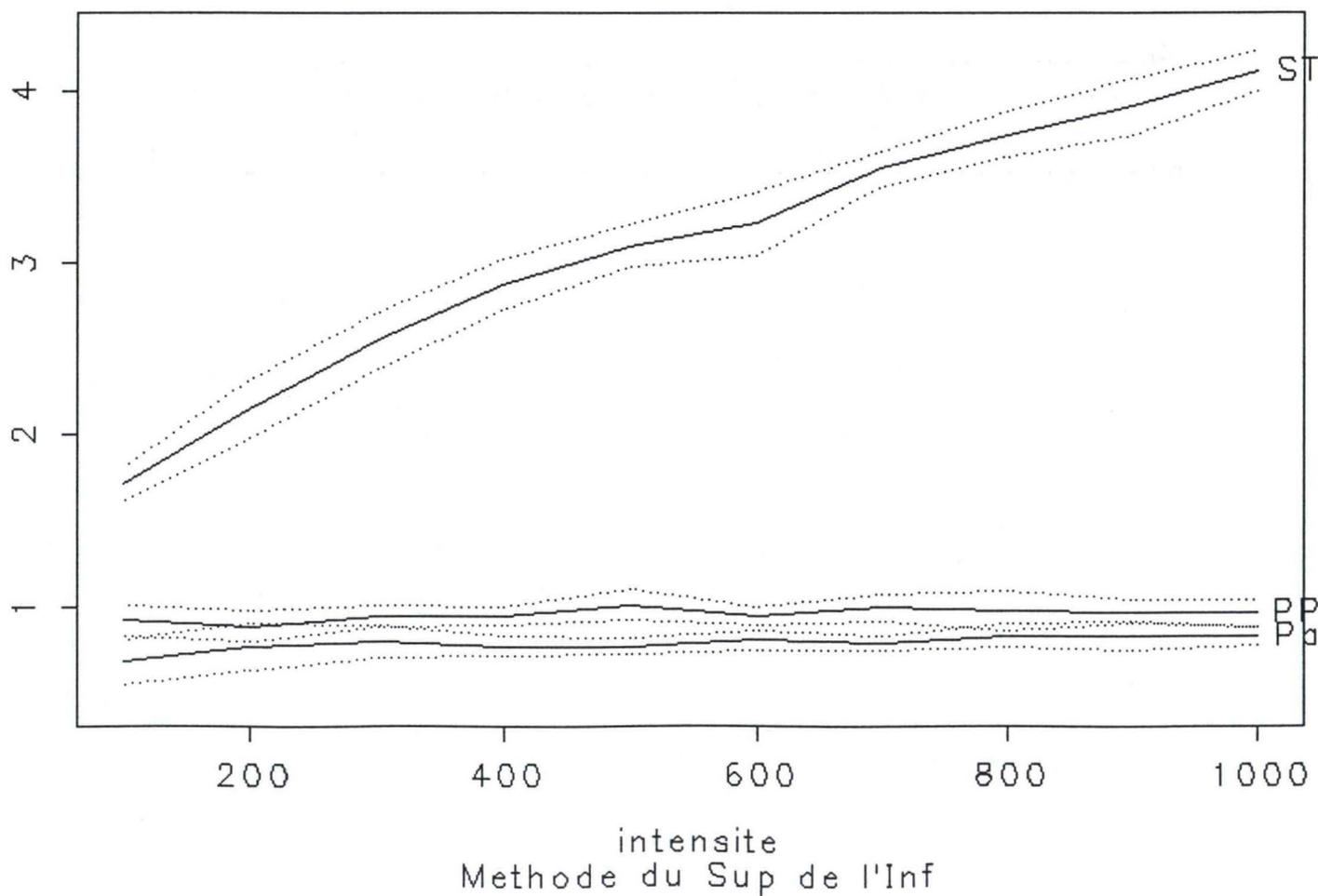
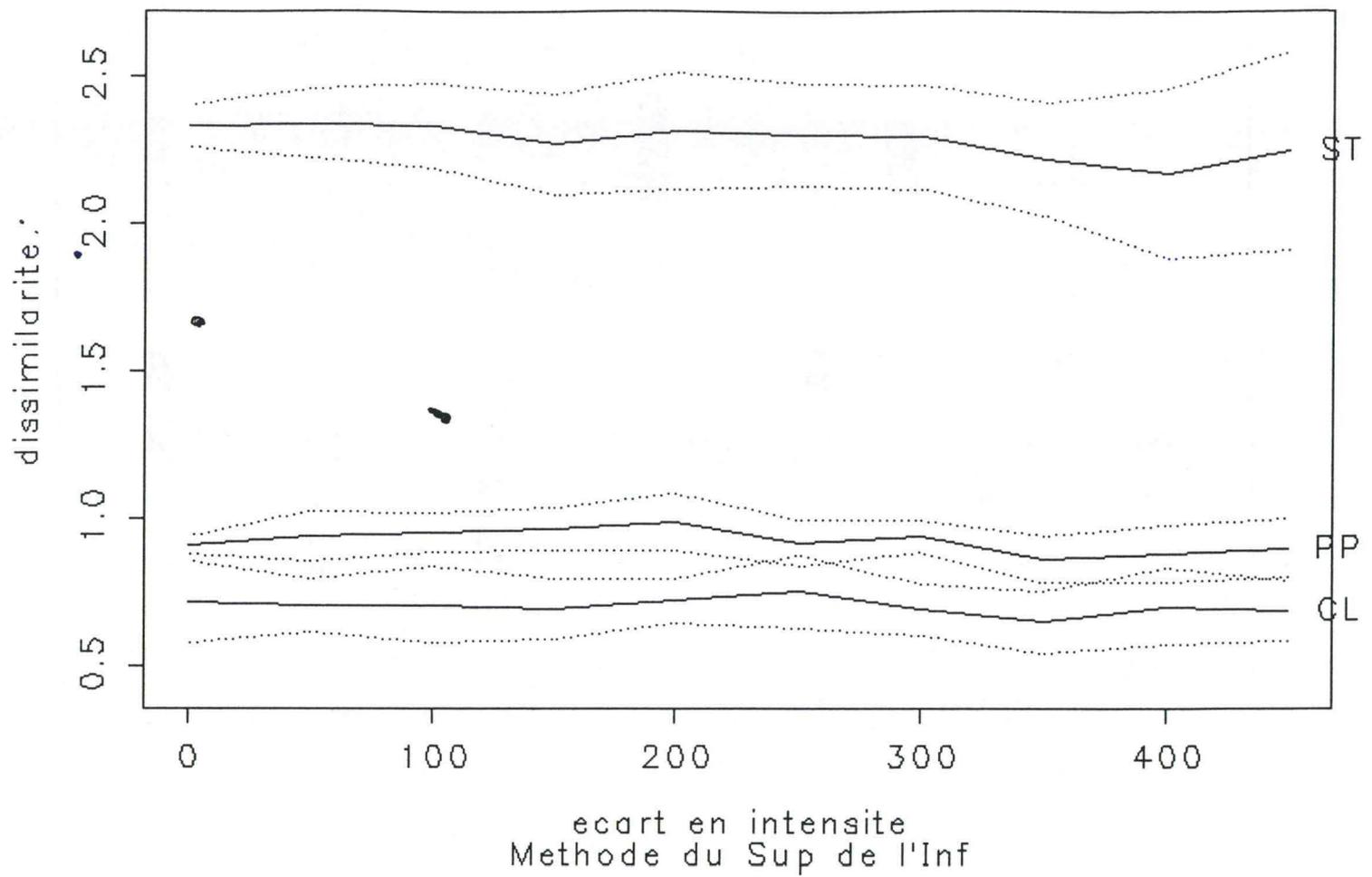
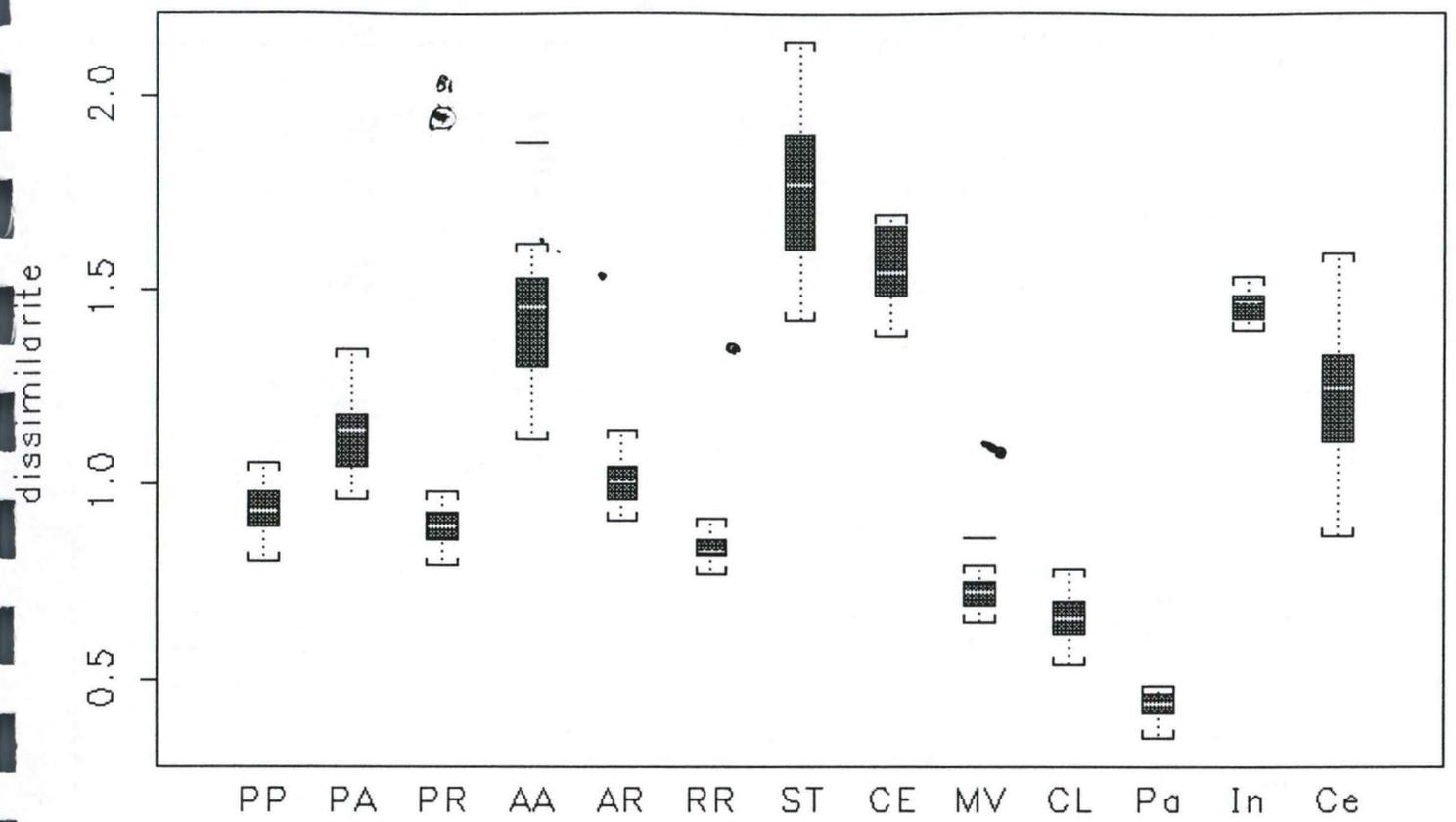


Figure 3.24



— moyenne sur 10 repetitions
..... ecart type

Figure 3.25



Methode des plus proches voisins

Figure 3.26

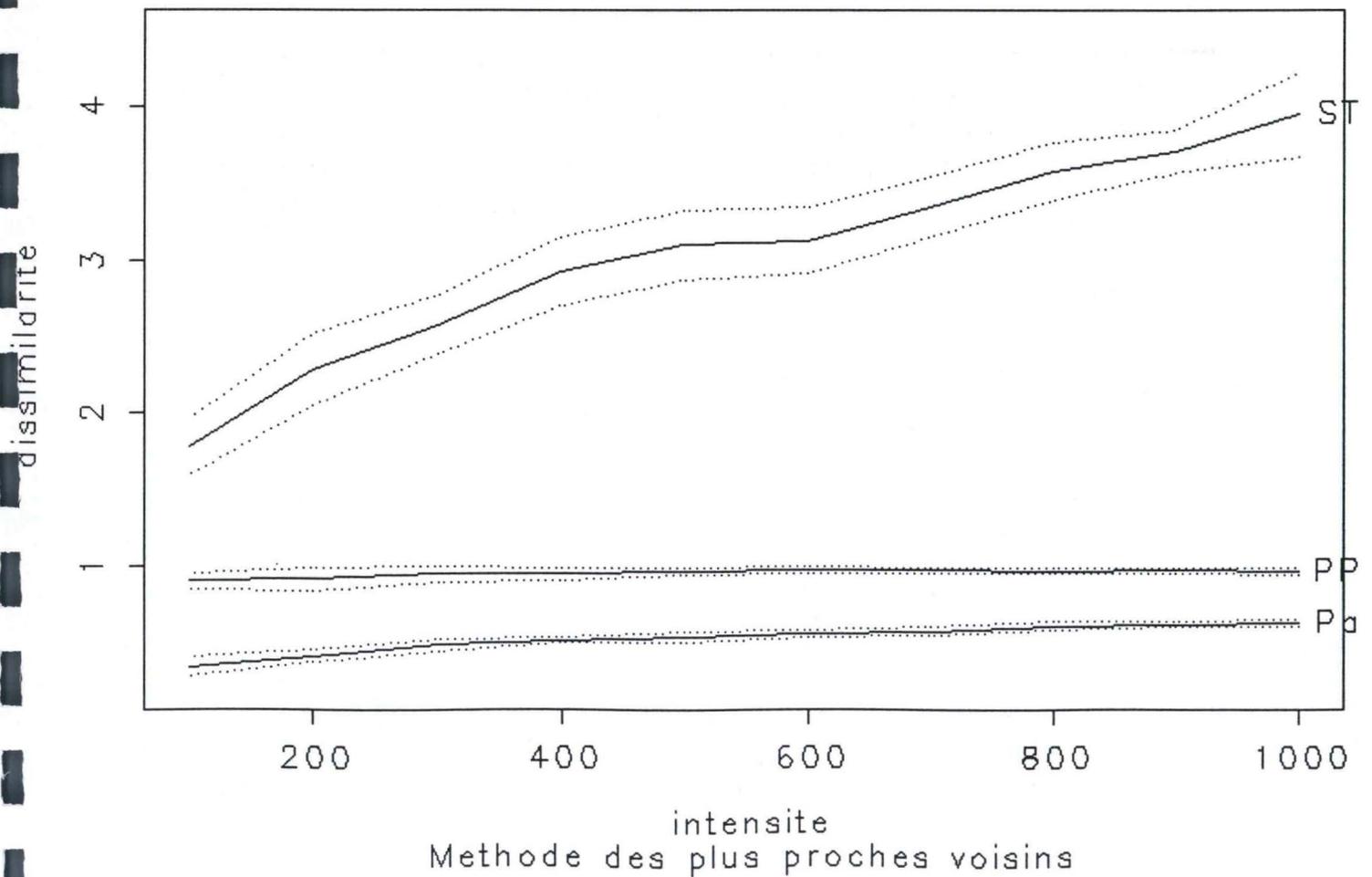
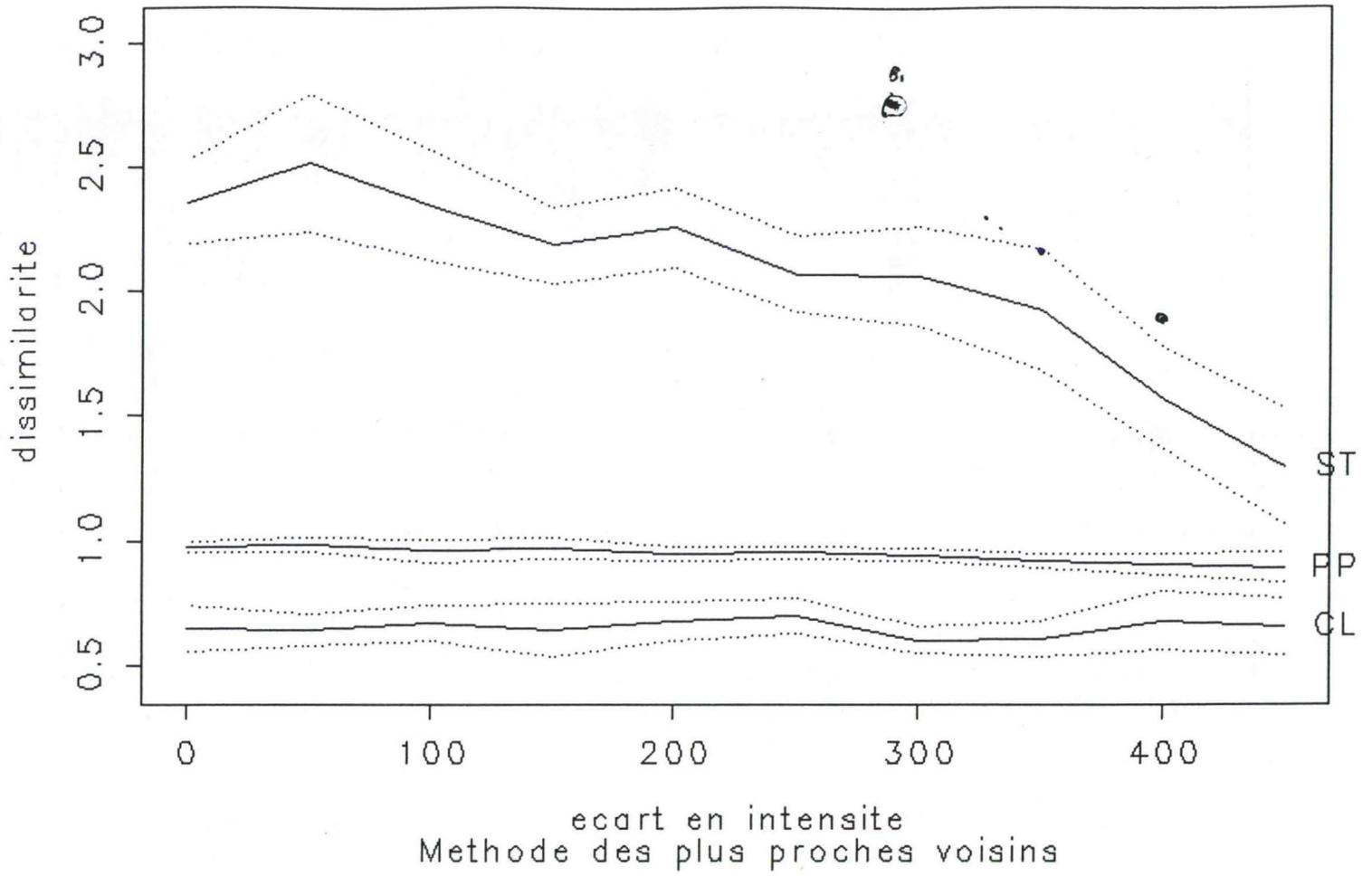
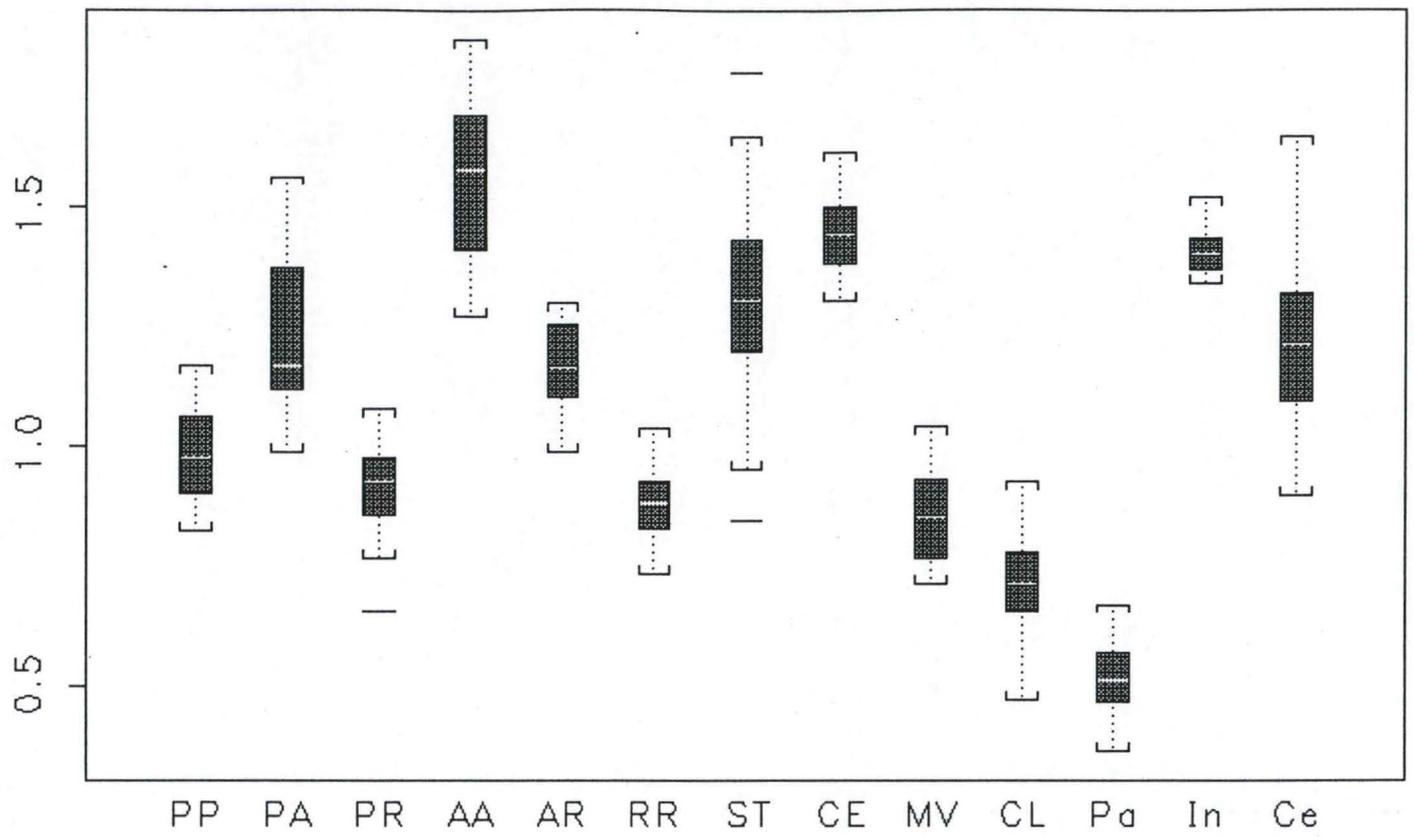


Figure 3.27



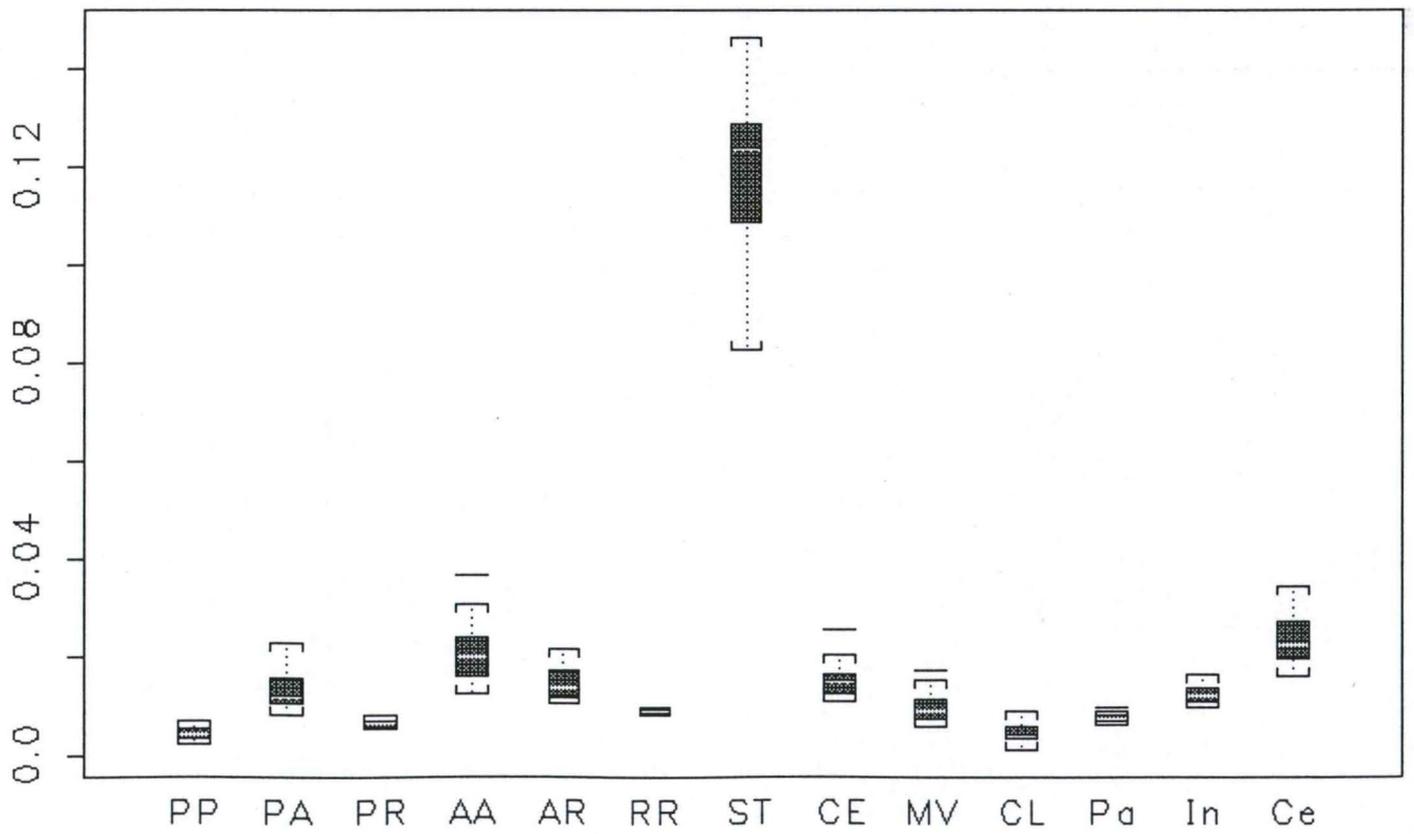
— moyenne sur 10 repetitions
..... ecart type

Figure 3.28



ppv avec echantillonnage partiel

Figure 3.29



Methode 2 de Ripley

4. APPLICATIONS : ÉTUDE DE RÉPARTITIONS BIVARIÉS DANS DES PEUPEMENTS FORESTIERS DE LA CÔTE NORD OUEST DU PACIFIQUE ET DE GUYANE

Les pseudo-indices de dissimilarité qui viennent d'être étudiés ont été calculés pour caractériser des répartitions spatiales bivariées réelles. Deux formations forestières très différentes ont ainsi servi d'application : d'une part une forêt de la côte Nord Ouest du Pacifique près de Vancouver, Canada, d'autre part la forêt tropicale humide de Paracou en Guyane française.

La forêt canadienne présente un double intérêt : d'une part c'est une *forêt naturelle*, et l'on peut donc espérer que la répartition spatiale des arbres reflète des phénomènes biologiques et non des travaux de sylviculture ; d'autre part c'est un système comportant naturellement très peu d'espèces d'arbres, généralement *deux espèces dominantes*. On est donc bien dans le domaine d'application des méthodes statistiques d'étude de répartitions bivariées.

A l'opposé la forêt guyanaise se distingue par sa richesse en espèces d'arbres. Les interactions spatiales entre espèces ne se ramènent alors sûrement pas à des interactions deux à deux, et il est donc aussi intéressant de voir ce que donnent les indices dans une telle situation.

4.1. Etude de la forêt de Vancouver

4.1.1. Description des peuplements

Huit peuplements ont été étudiés. Leur localisation et leur composition spécifique sont résumées dans le tableau suivant :

| Parcelle n° | Localisation | Espèces principales | Espèces minoritaires |
|-------------|---------------------|---------------------|----------------------|
| 1 | Capilano watershed | Ab, Hw | Cw |
| 2 | " | Ab, Hw | Cw |
| 3 | " | Ab, Cw, Hw | Df |
| 4 | " | Cw, Hw | Df |
| 5 | Coquitlam watershed | Ab, Hw | Cw |
| 6 | " | Cw, Hw | Df |
| 7 | " | Cw, Hw | Ab |
| 8 | Furry Creek | Ab, Hw | - |

Le bassin versant de Capilano se situe à une quinzaine de kilomètres au nord de Vancouver. Le bassin versant de Coquitlam se situe à une trentaine de kilomètres au nord est de Vancouver. Furry Creek se trouve à 40 km environ au nord de Vancouver, dans le district de Squamish. Les bassins versants de Capilano et de Coquitlam, qui alimentent Vancouver en eau potable, sont des zones protégées.

Ab représente *Abies amabilis* (Dougl.) Forbes (nom vernaculaire: silver fir), Hw représente *Tsuga heterophylla* (Raf.) Sarg. (nom vernaculaire: western hemlock), Cw représente *Thuja plicata* Donn ex D. Don in Lamb (nom vernaculaire: western red cedar), et Df représente *Pseudotsuga menziesii* (Mirb.) Franco (nom vernaculaire: Douglas fir).

Sur les huit peuplements, quatre comprennent les deux espèces western hemlock et silver fir, trois comprennent western hemlock et western red cedar, et un comprend les trois espèces simultanément. Outre ces espèces majoritaires, quelques arbres épars d'espèces marginales sont

généralement présents aussi sur les parcelles. N'ont été pris en compte que les espèces d'arbres atteignant la canopée.

Les huit peuplements sont issus de la régénération naturelle de zones exploitées dans les années 1920. Ils appartiennent à la même zone biogéoclimatique, dénommée Coastal Western Hemlock par Krajina [39]. Les huit parcelles, qui mesurent chacune environ 10 ha, ont été choisies de manière à ce que chaque parcelle soit homogène, et qu'il y ait des différences d'une parcelle à l'autre. Leur relief peut être décrit qualitativement de la façon suivante :

Parcelle 1 : au bord de la rivière Capilano. Le relief est plat, en terrasses successives vers la rivière. Le matériau est constitué de bancs d'alluvions.

Parcelle 2 : sur pente régulière moyenne. Des petits torrents creusent des talwegs peu profonds parallèlement à la pente.

Parcelle 3 : cuvette large, au fond de laquelle se trouve un ruisseau.

Parcelle 4 : pente régulière douce. Site apparemment très homogène.

Parcelle 5 : sommet arrondi.

Parcelle 6 : juxtaposition de collines, aux pentes relativement fortes.

Parcelle 7 : juxtaposition de collines aux pentes douces. Les vallons entre les collines sont des zones humides.

Parcelle 8 : pente régulière moyenne. Des petits torrents creusent des talwegs peu profonds.

4.1.2. Echantillonnage

Ne disposant que d'un temps limité pour faire les mesures, nous avons opté pour un échantillonnage partiel, ce qui élimine d'emblai les méthodes de Kershaw et de Ripley. Par ailleurs, il nous paraissait intéressant de disposer d'un indice qui mesurât un écart à l'indépendance et d'un autre qui mesurât un écart à l'association. Les méthodes retenues furent donc finalement celles des quadrats, de Diggle & Cox (avec prise en compte des quatre plus proches voisins), et des plus proches voisins.

Sur chaque parcelle, 100 points d'échantillonnage ont été placés de façon régulière, avec un espacement tel qu'un même arbre ne soit pas pris en compte plus d'une fois. En chaque point d'échantillonnage ont été mesurés

- les nombres d'arbres de chaque espèce présents dans un rayon de 5 ou 8 m (selon la densité du peuplement) autour du point d'échantillonnage.
- les distances $X_E^{(1)}, X_E^{(2)}, X_E^{(3)}, X_E^{(4)}$ du point d'échantillonnage au premier, deuxième, troisième, quatrième plus proche arbre de l'espèce E ($E=Ab, Cw, Hw$). Ne sont pris en compte que les arbres d'un diamètre à 1,30 m supérieur ou égal à 7,5 cm.
- les distances $Y_{EE'}$ de l'arbre de l'espèce E ($E=Ab, Cw, Hw$) le plus proche du point d'échantillonnage à son plus proche voisin de l'espèce E' ($E'=Ab, Cw, Hw$).

4.1.3. Etude des répartitions univariées

Bien que l'objet de cette étude ne soit pas la caractérisation des répartitions univariées, une brève description des répartitions univariées aide à mieux comprendre les relations entre espèces.

Densités

La densité est estimée comme le nombre moyen d'arbres par quadrat divisé par l'aire du quadrat. Elle est exprimée en arbre/ha :

| Parcelle n° | Ab | Cw | Hw | Total |
|-------------|-----|-----|-----|-------|
| 1 | 224 | - | 522 | 746 |
| 2 | 76 | - | 378 | 454 |
| 3 | 522 | 401 | 641 | 1564 |
| 4 | - | 280 | 393 | 673 |
| 5 | 261 | - | 349 | 609 |
| 6 | - | 554 | 374 | 928 |
| 7 | - | 547 | 620 | 1167 |
| 8 | 463 | - | 950 | 1413 |

Le hemlock est toujours l'espèce majoritaire par le nombre, excepté dans la parcelle 6. La densité moyenne varie du simple au triple entre la parcelle la moins dense et la parcelle la plus dense.

Caractérisation de la répartition spatiale

Quatre tests ont été calculés pour tester si la répartition est aléatoire ou non : le test de Thompson, le test de Holgate, le test de Hines & Hines et le test du χ^2 . Ces tests et leurs résultats sont présentés en annexe G. Les quatre tests donnent des résultats différents. Les tests de Thompson, de Hines & Hines et du χ^2 ont tendance à rejeter l'hypothèse d'une répartition aléatoire (pour une répartition agrégée la plupart du temps), tandis que le test de Holgate a tendance à l'accepter. Si on suit les conclusions données par une majorité de tests, les résultats suivants sont obtenus :

| | |
|------------|--|
| Parcelle 1 | Ab: répartition agrégée Hw: répartition aléatoire |
| Parcelle 2 | Ab: répartition agrégée Hw: répartition aléatoire |
| Parcelle 3 | Ab: répartition agrégée Cw: ? (deux tests acceptent \mathcal{H}_0 , deux tests le rejettent) Hw: répartition agrégée |
| Parcelle 4 | Cw: répartition aléatoire Hw: ? (deux tests acceptent \mathcal{H}_0 , deux tests le rejettent) |
| Parcelle 5 | Ab: répartition agrégée Hw: ? |
| Parcelle 6 | Cw: répartition agrégée Hw: ? |
| Parcelle 7 | Cw: ? Hw: ? |
| Parcelle 8 | Ab: répartition agrégée Hw: répartition aléatoire |

Ainsi le silver fir est toujours agrégé, tandis que le hemlock est presque toujours réparti de façon aléatoire. Il est difficile de conclure quant à la répartition du western red cedar.

4.1.4. Etude des répartitions bivariées

Test de l'indépendance

La répartition bivariée observée est une réalisation d'un processus bivarié. L'hypothèse nulle est que les processus marginaux du processus bivarié sont indépendants. Parmi les méthodes retenues, la méthode des quadrats et celle de Diggle & Cox permettent de tester l'indépendance des répartitions marginales :

1. Test basé sur les quadrats : soit X^2 la distance du χ^2 définie sur le tableau de contingence des présences / absences dans les quadrats (voir annexe F). Sous \mathcal{H}_0 , X^2 est distribué comme un χ^2 à un degré de liberté.
2. Test de Diggle & Cox : soit $X_{1(k)}$ et $X_{2(k)}$ les distances d'un point d'échantillonnage au k^{eme} plus proche arbre des espèces 1 et 2. Sous \mathcal{H}_0 , $X_{1(k)}$ et $X_{2(k)}$ sont des variables indépendantes, dont l'indépendance est testée avec le coefficient de corrélation des rangs τ de Kendall.

Les résultats des tests sont donnés en annexe G. Le tableau suivant les résume :

| Parcelle | Test des quadrats | Test de Diggle & Cox |
|-----------|-------------------|--------------------------------|
| 1 | indépendance | indépendance |
| 2 | " | " |
| 3 (Ab/Cw) | " | " (sauf pour $k = 3$) |
| 3 (Ab/Hw) | " | " |
| 3 (Cw/Hw) | " | " |
| 4 | " | " (sauf pour $k = 4$) |
| 5 | répulsion | répulsion ($k = 1, 2, 3, 4$) |
| 6 | " | " ($k = 2, 3$) |
| 7 | " | " ($k = 1, 2, 3, 4$) |
| 8 | " | " ($k = 1, 2, 3, 4$) |

Les deux tests concordent : il y a indépendance entre les répartitions marginales pour les peuplements 1,2,3,4, et il y a une association négative entre les deux répartitions marginales pour les peuplements 5,6,7,8.

Indices de dissimilarité

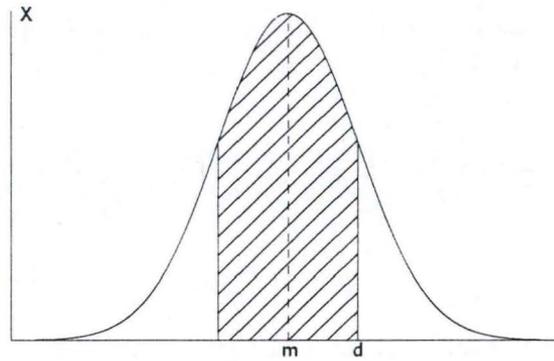
Trois des pseudo indices de dissimilarité présentés dans la partie II ont été calculés : celui des quadrats, celui de Diggle & Cox avec prise en compte des quatre plus proches voisins, et l'indice des plus proches voisins. Les valeurs des indices sont données en annexe. Elles ont été confrontées à des valeurs obtenues par simulation. Chacun des treize processus ponctuels définis en partie I a été simulé 20 fois avec une valeur de l'intensité égale à celle observée au Canada, et la méthode d'échantillonnage réalisée au Canada a été reproduite dans la simulation. Les résultats sont présentés sur les figures 4.4, 4.5 et 4.6. Les boxplots correspondent aux simulations tandis que les traits horizontaux représentent les valeurs des indices observés au Canada sur chacune des parcelles.

On peut chercher à mesurer de combien les valeurs des indices calculés pour une répartition s'écartent des valeurs obtenues par simulation en procédant de la façon suivante : supposons que l'on dispose de I indices de dissimilarité. Une répartition bivariée sera donc décrite par I valeurs d_1, d_2, \dots, d_I . Soit N un processus ponctuel simulé ; on peut estimer la valeur moyenne m_i et l'écart type σ_i ($i = 1 \dots I$) de chaque indice pour ce processus. En approximant la distribution des indices pour N par une loi normale, l'écart de la répartition observée à N selon l'indice i peut alors être mesuré par

$$\delta_i = 1 - 2 \Pr [X_i < m_i - |m_i - d_i|]$$

où X_i est une variable aléatoire qui suit une loi normale $\mathcal{N}(m_i, \sigma_i)$; δ_i est égal à l'aire de la surface

hachurée :



Finalement, l'écart de la répartition observée à N peut être mesuré par une moyenne des δ_i , par exemple $\Delta = \left(\sum_{i=1}^I \delta_i^2 \right)^{1/2}$.

Ces calculs ont été faits pour chacun des treize processus ponctuels simulés, en tenant compte des trois indices calculés. Pour chacun des peuplements, celui des processus parmi les treize simulés dont l'écart Δ à la répartition observée est minimum est indiqué dans le tableau suivant :

| | | | | | | | | | | |
|-----------|----|----|-----------|-----------|-----------|----|----|----|----|----|
| Parcelle | 1 | 2 | 3 (Ab/Cw) | 3 (Ab/Hw) | 3 (Cw/Hw) | 4 | 5 | 6 | 7 | 8 |
| Processus | PA | AR | AA | AA | AA | PR | AR | MV | MV | In |

Conclusion sur les simulations Le nombre de processus simulés étant restreint, et les processus simulés étant de plus des processus simples, il convient de nuancer les résultats des simulations : la proximité de la valeur d'un indice calculé pour une répartition observée aux valeurs de cet indice calculé pour des simulations d'un processus n'indiquera qu'une tendance de la répartition observée à être décrite par le processus.

Les figures 4.4 et 4.5 ne font que représenter les résultats des tests d'indépendance des répartitions marginales : les peuplements 1, 2, 3, 4 tombent dans le domaine des processus avec indépendance des processus marginaux, tandis que les peuplements 5, 6, 7, 8 s'en écartent vers les processus avec répulsion à grande ou moyenne échelle des processus marginaux (processus CE ou In).

La figure 4.6 apporte de l'information supplémentaire. Comme les valeurs de l'indice des plus proches voisins calculé pour les parcelles 5 et 8 sont toujours proches des valeurs obtenues par simulation des processus CE et In, la répartition dans ces deux parcelles reste compatible avec une association négative des deux espèces à grande ou moyenne échelle. Mais ce n'est plus le cas pour les parcelles 6 et 7 : la répartition dans ces parcelles doit être plus compliquée.

Si l'on s'intéresse aux répartitions bivariées avec indépendance des répartitions marginales (parcelles 1-4), la figure 4.6 montre les résultats suivants :

- la répartition dans la parcelle 1 est compatible avec les processus PA et AR
- la répartition dans la parcelle 2 est compatible avec les processus PA et AR
- la répartition dans la parcelle 3 (Ab / Cw) est compatible avec les processus AA
- la répartition dans la parcelle 3 (Ab / Hw) est compatible avec les processus AA
- la répartition dans la parcelle 3 (Cw / Hw) est compatible avec les processus PA et AR
- la répartition dans la parcelle 4 est compatible avec les processus PP.

Ces résultats sont cohérents avec ceux de l'étude des répartitions univariées.

Les écarts des répartitions aux processus amènent à des conclusions identiques dans le cas des parcelles 1, 2, 3, 4, 8, mais amènent à des conclusions très différentes dans le cas des parcelles 6 et 7 : pour ces deux parcelles en effet, le processus qui apparaît comme le plus proche est MV, c'est-à-dire un processus avec attraction à petite échelle des processus marginaux, alors que les méthodes de Diggle & Cox et des quadrats indiquent au contraire qu'il y a association négative entre les deux répartitions marginales. Il donc certainement plus prudent de raisonner indice par indice.

La densité peut-elle être reliée aux caractéristiques des répartitions ?

On peut chercher à voir enfin s'il y a un lien entre la densité et la dépendance / indépendance des répartitions marginales. Par exemple :

| | Dépendance des répartitions marginales | Indépendance des répartitions marginales |
|-------------------------------------|---|---|
| Densité du hemlock (arbre/ha) | 349 (parcelle 5) | 522 (parcelle 1) |
| | 374 (parcelle 6) | 378 (parcelle 2) |
| | 620 (parcelle 7) | 641 (parcelle 3) |
| | 950 (parcelle 8) | 393 (parcelle 4) |

Un test de Wilcoxon-Mann-Whitney appliqué à ces deux échantillons montre qu'il n'y a pas de lien entre la densité du hemlock et la dépendance / indépendance des répartitions marginales ($U = 9$; $\Pr[U \geq 9] = 0,443$).

De la même manière, aucun lien n'apparaît entre la densité totale et la dépendance / indépendance des répartitions marginales ($U = 5$; $\Pr[U \geq 5] = 0,829$), pas plus qu'entre la densité de l'espèce autre que le hemlock et la dépendance / indépendance des répartitions marginales ($U = 6$; $\Pr[U \geq 6] = 0,757$).

4.1.5. Etude de la répartition trivariée (parcelle 3)

Seule la parcelle 3 rassemble les trois espèces silver fir, western red cedar et western hemlock simultanément. Nous avons tenté de voir sur cette parcelle si l'interaction entre les trois espèces se ramène ou non aux interactions deux à deux. Nous venons de voir déjà qu'il y a indépendance entre les espèces prises deux à deux.

La répartition trivariée a été étudiée à l'aide d'un tableau de contingence à trois entrées sur la présence / absence des espèces :

| | Hw présent | | |
|------------|----------------------|----------------|----------------|
| | Cd présent | Cd absent | |
| Ab présent | $n_{111} = 11$ | $n_{121} = 18$ | $n_{1.1} = 29$ |
| Ab absent | $n_{211} = 22$ | $n_{221} = 16$ | $n_{2.1} = 38$ |
| | $n_{.11} = 33$ | $n_{.21} = 34$ | $n_{..1} = 67$ |
| | Hw absent | | |
| | Cd présent | Cd absent | |
| Ab présent | $n_{112} = 7$ | $n_{122} = 8$ | $n_{1.2} = 15$ |
| Ab absent | $n_{212} = 13$ | $n_{222} = 5$ | $n_{2.2} = 18$ |
| | $n_{.12} = 20$ | $n_{.22} = 13$ | $n_{..2} = 33$ |
| | Hw présent ou absent | | |
| | Cd présent | Cd absent | |
| Ab présent | $n_{11.} = 18$ | $n_{12.} = 26$ | $n_{1..} = 44$ |
| Ab absent | $n_{21.} = 35$ | $n_{22.} = 21$ | $n_{2..} = 56$ |
| | $n_{.1.} = 53$ | $n_{.2.} = 47$ | $n = 100$ |

Cent quadrats circulaires de rayon $\sqrt{3/\lambda\pi}$, où λ est la densité totale d'arbres, ont été placés sur la zone d'étude ; n_{ijk} est le nombre de quadrats qui contiennent aucun ($i = 2$) ou au moins un ($i = 1$) silver fir, aucun ($j = 2$) ou au moins un ($j = 1$) western red cedar, et aucun ($k = 2$) ou au moins un ($k = 1$) western hemlock.

L'annexe D donne un aperçu de la façon d'analyser un tableau de contingence $2 \times 2 \times 2$. L'hypothèse nulle la plus forte est celle d'indépendance totale, qui signifie que la répartition de chacune des espèces est indépendante des deux autres.

Le test d'indépendance totale appliquée à la parcelle 3 conduit à accepter cette hypothèse ($G^2 = 5,863$; $\text{Pr}[\chi^2 < G^2] = 0,790$: voir annexe D). D'autres tests d'hypothèses moins fortes confirment ce résultat.

En fin de compte, les occurrences des trois espèces sur la parcelle 3 peuvent être considérées comme complètement indépendantes les unes des autres.

4.1.6. Lien avec les caractéristiques écologiques des espèces

Aucune explication biologique des répartitions observées n'a été testée dans le cadre de cette étude. Néanmoins, compte tenu des connaissances que l'on a sur l'écologie des trois espèces, des hypothèses peuvent être avancées pour expliquer les répartitions observées.

Le western hemlock, le western red cedar et le silver fir sont toutes trois des espèces tolérantes à l'ombre. Mais les rangs de tolérance varient d'un auteur à l'autre [6, 63, 66]. Ces divergences s'expliquent par le fait que les auteurs n'estiment pas de la même façon la tolérance à l'ombre, que la tolérance à l'ombre n'est pas indépendante d'autres facteurs comme l'humidité du sol, la richesse en minéraux du sol [66], et qu'il y a une grande variabilité intraspécifique pour la tolérance à l'ombre [6].

Si l'on cherche à comparer le western hemlock au western red cedar, il apparaît que le western red cedar supporte mieux les sites secs que le hemlock [28, 63]. De plus le western red cedar supporte également mieux les zones humides mal drainées que le hemlock [6, 28]. Sa capacité de reproduction végétative par marcottage lui donne un avantage supplémentaire sur le hemlock dans les zones humides [28]. Le western red cedar possède également une amplitude écologique plus large que le hemlock en ce qui concerne les besoins en sels minéraux : les pousses de western red cedar supportent mieux un faible niveau de sels minéraux que celles de hemlock [46]; Gregory [24] observe également que l'abondance du red cedar par rapport au hemlock augmente quand les sols deviennent plus pauvres. Le hemlock a donc une amplitude écologique plus restreinte que le red cedar, mais à l'intérieur de son rang écologique il est plus compétitif : Messier [46] montre en effet que lorsque l'humidité du sol est convenable et que le niveau en sels minéraux est suffisant, les pousses de hemlock croissent plus vite que celles de red cedar. Cela rejoint les observations de plusieurs auteurs [24, 28] comme quoi dans les phénomènes de succession, le red cedar apparaît en premier, puis est progressivement éliminé par le hemlock. Daubenmire (*in* [28]) estime en fait que le red cedar ne peut se maintenir que sur des habitats qui se situent en dehors du rang écologique du hemlock. Habeck [28] observe des peuplements mélangés red cedar et hemlock qui restent stables en composition, et en regardant de plus près il remarque en effet que le red cedar se maintient dans des micro dépressions mal drainées, tandis que le hemlock domine sur les pentes bien drainées. Klinka (communication personnelle) fait également remarquer que dans les peuplements mélangés red cedar / hemlock, le red cedar se trouve plutôt dans les microhabitats constitués des dépressions mal drainées et des sommets de buttes.

Ces données peuvent expliquer ce que l'on observe sur les parcelles 6 et 7 : le relief de ces parcelles est en effet constitué de petites collines ; les vallons entre les collines sont des zones humides et les sommets de collines sont des zones sèches peu favorables au hemlock, et où donc le red cedar se maintient ; les pentes sont au contraire dominées par le hemlock. Les deux espèces apparaissent alors dissociées. Sur les parcelles 3 et 4, où le relief est beaucoup plus homogène, les deux espèces apparaissent au contraire réparties indépendamment l'une de l'autre.

Il est en revanche plus difficile de proposer une explication à la relation que l'on observe entre le hemlock et le silver fir. Plusieurs auteurs [48, 66] observent que le silver fir, malgré une croissance plus lente, finit par remplacer le hemlock dans les peuplements où les deux espèces coexistent. Zobel & Antos [66] montrent que les pousses de silver fir sont plus tolérantes que celles de hemlock à la fois à l'ombre, à la sécheresse, et à l'enterrement par la litière. La répartition systématiquement agrégée des silver firs que l'on observe peut s'expliquer par la faible capacité de dispersion des graines. Klinka (communication personnelle) estime que les graines de silver fir ne se dispersent pas à plus de 15 m, tandis que les graines de hemlock, plus légères, peuvent aller jusqu'à 3 km. Cela peut aussi être relié à la répartition souvent aléatoire des hemlocks.

4.2. Etude de la forêt de Paracou

4.2.1. Description du site

Le dispositif sylvicole de Paracou est situé à proximité de la commune de Sinnamary en Guyane française. Il se compose de douze parcelles carrées de 6,25 ha chacune, sur lesquelles ont été relevés chaque année depuis 1994 la position, l'essence et le diamètre des tiges de plus de 10 cm de diamètre. La description de ce dispositif a déjà fait l'objet de plusieurs publications (par exemple [2, 3]).

4.2.2. Choix des espèces

Pour l'étude des répartitions bivariées, trois espèces ont été retenues, qui sont représentées par un nombre suffisamment élevé d'individus sur les parcelles, et qui présentent un intérêt écologique. Il s'agit de *Dicorynia guianensis* (l'angélique), de *Pradosia cohlearia* (le kimbotto) et d'*Eperua falcata* (le wapa).

L'angélique a une répartition agrégée [Hélène Dessart, thèse en cours], comme il apparaît sur la figure 4.1. Le dispositif de Paracou contient 576 individus de cette espèce, avec un minimum de 25 individus sur la parcelle 5 et un maximum de 101 individus sur la parcelle 9.

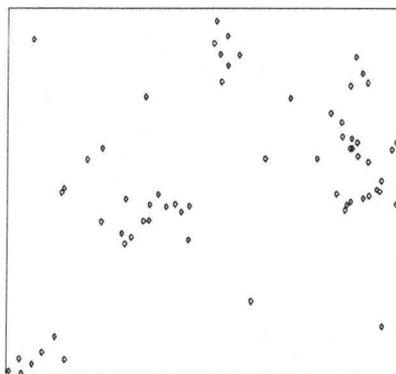


Figure 4.1: Répartition de l'angélique sur la parcelle 8

L'angélique, tout comme *Eperua falcata*, est une espèce plus héliophile que le kimbotto (Vincent Favrichon, communication personnelle). Les données sur ses exigences édaphiques sont contradictoires : Lescure & Boulet [41] estiment que l'espèce est favorisée sur sol à drainage vertical bloqué et sur sol hydromorphe ; Frédérique Collinet (thèse en cours) note au contraire un impact négatif de l'hydromorphie sur sa densité. Il se pourrait que l'espèce n'ait pas d'exigences édaphiques particulières et que sa dynamique de distribution en agrégats soit le principal facteur de sa répartition spatiale.

Le kimbotto a une répartition plus proche d'une répartition aléatoire (voir fig. 4.2), bien qu'il soit beaucoup plus présent sur certaines parcelles que sur d'autres. Sa densité varie entre 17 individus sur la parcelle 2 et 139 individus sur la parcelle 4, pour un total de 949 individus sur l'ensemble du dispositif.

C'est une espèce sciaphile. Frédérique Collinet note une abondance légèrement moindre de l'espèce sur les sols hydromorphes, mais ce comportement n'est pas très tranché.

51

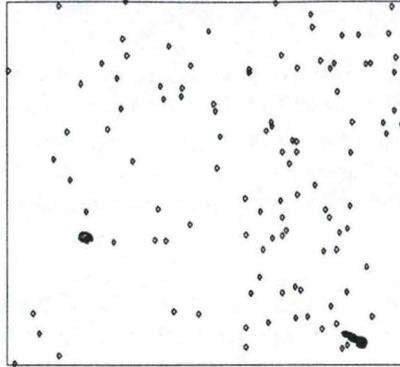


Figure 4.2: Répartition du kimboto sur la parcelle 8

Le wapa (*Eperua falcata*) a une répartition par plaques de grandes tailles qui chevauchent les parcelles du dispositif de Paracou [2](voir aussi fig. 4.3). C'est une espèce abondante : 1893 individus sont présents sur le site, avec un minimum de 1 individu sur la parcelle 1 et un maximum de 283 individus sur la parcelle 6.

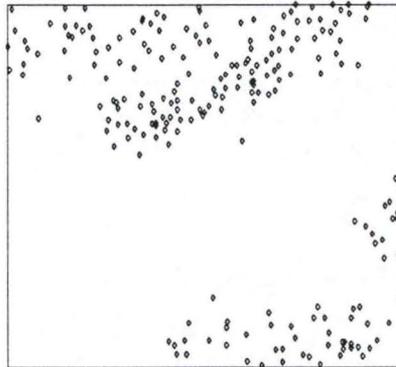


Figure 4.3: Répartition du wapa sur la parcelle 4

La répartition spatiale du wapa apparaît fortement liée au sol, comme de nombreuses études l'ont fait ressortir [2, 3, 58, 59]. Le wapa se retrouve ainsi sur des sols à fortes contraintes physico-hydriques et chimiques : mauvaise aération (bas-fonds engorgés), présence d'horizons compacts et faibles réserves hydriques (sols à drainage superficiel), toxicité aluminique éventuellement marquée.

4.2.3. Tests de l'indépendance

Quatre tests de l'indépendance des répartitions marginales ont été calculés :

1. Test des quadrats : soit X^2 la distance du χ^2 calculée sur le tableau de contingence des présences / absences des espèces dans les quadrats (voir annexe F). S'il y a indépendance des

répartitions marginales, X^2 suit approximativement un χ^2 à un degré de liberté. La méthode d'échantillonnage est la même que pour les simulations.

2. Test de Diggle & Cox : soit $X_{i(k)}$ et $X_{2(k)}$ les distances d'un point d'échantillonnage au $k^{\text{ème}}$ plus proche arbre des espèces 1 et 2. Si les processus sont homogènes et qu'il y a indépendance des répartitions marginales, alors $X_{i(k)}$ et $X_{2(k)}$ sont des variables indépendantes dont l'indépendance est testée avec le coefficient de corrélation τ de Kendall. La méthode d'échantillonnage est la même que pour les simulations, en tenant compte des quatre plus proches voisins ($k = 1, 2, 3, 4$).
3. Test de Goodall : soit X_i la distance d'un point arbitraire à l'arbre de l'espèce i le plus proche, et Y_{ji} la distance d'un arbre arbitraire de l'espèce j à son plus proche voisin de l'espèce i . Si les processus sont homogènes et qu'il y a indépendance des répartitions marginales, alors X_i et Y_{ji} sont identiquement distribués, ce qui peut être testé par la statistique U de Wilcoxon. La méthode d'échantillonnage est la même que pour les simulations, c'est-à-dire qu'on réalise en fait quatre tests de Wilcoxon.
4. Test à l'aide de la fonction K_{12} de Ripley [30, 42, 50] : la fonction K_{12} de Ripley est calculée pour la répartition observée, et comparée à la fonction K_{12} obtenue par translation arbitraire d'une des répartitions marginales par rapport à l'autre. Ce test ne constitue pas en fait un test d'indépendance des processus marginaux : il s'agit d'un test d'indépendance des processus marginaux conditionnellement aux répartitions marginales observées. En pratique, un nombre n (nous avons pris $n = 99$) de translations aléatoires sont réalisées, ce qui fournit une enveloppe aux fonctions K_{12} , dans laquelle s'inscrit ou ne s'inscrit pas la fonction K_{12} de la répartition observée selon que l'on accepte ou que l'on rejette l'indépendance des répartitions marginales.

Les tests ont été calculés sur chaque parcelle où les effectifs des espèces dépassaient 10 individus, puis globalement en réunissant les données tirées de chaque parcelle (excepté pour le test de Ripley). Les résultats des tests sont résumés dans les tableaux suivants :

Angélique et kimbotó

| Parcelle | Tests | | | |
|----------|---------------------|--------------|-----------------|--------------|
| | Diggle & Cox | quadrats | Goodall | Ripley |
| 1 | attraction pour k=1 | indépendance | indépendance | indépendance |
| 2 | indépendance | " | " | " |
| 3 | " | " | " | " |
| 4 | " | " | " | " |
| 5 | " | " | répulsion (1/4) | " |
| 6 | " | " | indépendance | " |
| 7 | " | " | " | " |
| 8 | " | " | " | " |
| 9 | " | " | " | " |
| 10 | " | " | " | " |
| 11 | " | " | " | " |
| 12 | " | " | " | " |
| Global | indépendance | indépendance | indépendance | |

Les tests s'accordent, excepté sur la parcelle 1 où le test de Diggle & Cox indique une attraction à grande échelle et sur la parcelle 5 où l'un des quatre tests de Goodall indique une répulsion. On peut conclure à l'indépendance des répartitions de l'angélique et du kimbotó sur l'ensemble des parcelles.

Angélique et wapa

| Parcelle | Tests | | | |
|----------|---------------------------|--------------|--------------|--------------------------|
| | Diggle & Cox | quadrats | Goodall | Ripley |
| 3 | attraction pour k=1 | indépendance | indépendance | indépendance |
| 4 | attraction pour k=3 | " | " | " |
| 5 | indépendance | " | " | " |
| 6 | attraction pour k=3,4 | " | " | " |
| 7 | indépendance | " | " | " |
| 8 | " | " | " | répulsion pour $r > 50m$ |
| 9 | " | " | " | indépendance |
| 10 | attraction pour k=2 | " | " | " |
| 11 | indépendance | " | " | " |
| 12 | attraction pour k=1,2,3,4 | " | " | " |
| Global | attraction pour k=4 | indépendance | indépendance | |

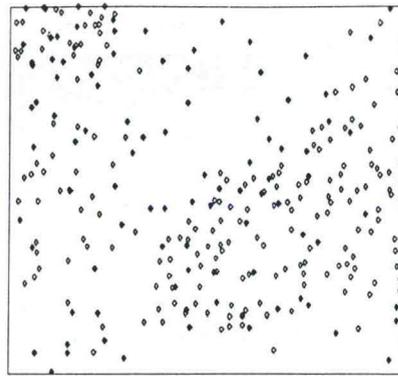
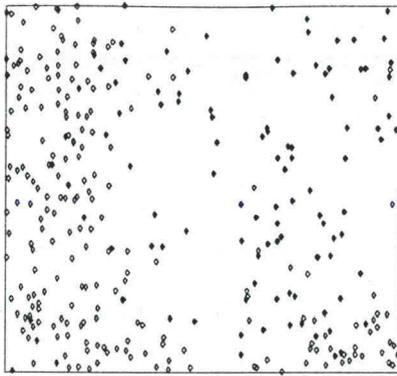
Les tests sont moins concluants que pour l'angélique et le kimbotto. Les quatre tests indiquent l'indépendance des répartitions marginales sur les parcelles 5,7,9,11. Les désaccords entre tests sont faibles sur les parcelles 3,4,6,10, et on peut supposer qu'il y a aussi indépendance sur ces parcelles. En revanche il est plus étonnant que le test de Ripley indique une répulsion sur la parcelle 8 et que le test de Diggle & Cox indique une attraction sur la parcelle 12, alors qu'aucun des autres tests ne l'indique.

Kimbotto et wapa

| Parcelle | Tests | | | |
|----------|--------------------|--------------|------------------|-------------------------------------|
| | Diggle & Cox | quadrats | Goodall | Ripley ¹ |
| 3 | indépendance | indépendance | indépendance | indépendance |
| 4 | " | " | " | " |
| 5 | " | " | " | " |
| 6 | " | " | répulsion (1/4) | répulsion ($600 \leq r \leq 800$) |
| 7 | " | " | indépendance | indépendance |
| 8 | répulsion pour k=2 | répulsion | " | répulsion ($r \geq 300$) |
| 9 | indépendance | " | " | indépendance |
| 10 | " | " | " | " |
| 11 | " | " | " | " |
| 12 | " | " | " | " |
| Global | indépendance | indépendance | attraction (1/4) | |

Les tests s'accordent là encore plus ou moins. Il y a indépendance sur les parcelles 3,4,5,7,9,10,11,12, une tendance à la répulsion sur la parcelle 6, et répulsion sur la parcelle 8. La figure ci-dessous, qui représente les répartitions spatiales des deux espèces, montre en effet que celles-ci se répartissent sur la parcelle 8 comme deux nuages de points qui ne se recouvrent pas. Par comparaison, sur la parcelle 5 où il y a indépendance, le wapa est toujours réparti par plaques, mais le kimbotto est réparti de façon beaucoup plus uniforme.

¹Les fonctions $L_{12} = \sqrt{K_{12}/\pi}$ de Ripley pour les parcelles 5 (indépendance) et 8 (répulsion) sont présentées à titre d'exemple sur la figure 4.7.



Parcelle 8
Répartition du wapa (point blancs) et du kimboto (points noirs)

Parcelle 5

4.2.4. Indices de dissimilarité

Les indices de dissimilarité ont été calculés sur chaque parcelle et comparés à des simulations de processus ponctuels. Les figures résultantes sont rassemblées à la fin de cette partie (fig. 4.8 à 4.31).

Angélique et kimboto. Les figures relatives aux méthodes des quadrats, de Diggle & Cox, de Goodall, de Kershaw et de Ripley (fig. 4.8 à 4.12) ne font qu'illustrer les résultats des tests d'indépendance : les valeurs des indices apparaissent en effet compatibles avec celles obtenues par simulation de processus ponctuels dont les processus marginaux sont indépendants.

Les indices mesurant un écart à l'association, à savoir les indices de Pielou, du Sup de l'Inf et des plus proches voisins (fig. 4.13 à 4.15), montrent que les processus avec indépendance des processus marginaux qui s'approchent le plus (parmi les six processus simulés) des répartitions observées sont PA et AR. Cela est cohérent avec la répartition agrégée de l'angélique et la répartition à peu près aléatoire du kimboto.

De plus les indices affichent une certaine cohérence entre eux. Si l'on prend l'exemple de la parcelle 12 : les indices mesurant un écart à l'indépendance placent toujours cette parcelle parmi celles dont la répulsion est la plus marquée (fig. 4.8 à 4.12); les indices mesurant un écart à l'association placent toujours cette parcelle parmi celles qui se rapprochent le plus du processus AA (fig. 4.13 à 4.15).

Angélique et wapa. Intéressons nous aux parcelles qui posent problème, c'est-à-dire la 8 et la 12. La parcelle 8 apparaît comme l'une des parcelles où l'indépendance est nette pour les indices des quadrats, de Goodall et de Diggle & Cox (fig. 4.16 à 4.18). En revanche les indices de Ripley et de Kershaw font bien ressortir une répulsion sur cette parcelle (fig. 4.19 et 4.20). Quant à la parcelle 12, les indices des quadrats, de Diggle & Cox et de Kershaw (fig. 4.16, 4.17 et 4.19) font ressortir une attraction tandis que les indices de Goodall et de Ripley (fig. 4.18 et 4.20) laissent plutôt supposer qu'il y a indépendance.

Par comparaison, le comportement des indices sur la parcelle 10 est moins ambigu : tous les indices mesurant l'écart à l'indépendance font ressortir cette parcelle comme l'une de celles où l'attraction est la plus marquée.

Kimboto et wapa. Si l'on s'intéresse aux parcelles 8, 6 et 5 : la répulsion sur la parcelle 8 ressort bien (à l'exception de l'indice de Goodall, tous les indices mesurant un écart à l'indépendance placent cette parcelle comme étant celle où la répulsion est la plus marquée). De même la parcelle

5 ressort bien comme une parcelle où il y a indépendance des répartitions marginales (fig. 4.24 à 4.28).

En revanche les indices tempèrent les résultats des tests pour la parcelle 6, les indices semblent indiquer (fig. 4.24 à 4.28) qu'il y a plutôt indépendance des répartitions marginales sur la parcelle 6, contrairement à la tendance à la répulsion que les tests ont fait ressortir.

4.2.5. Conclusion

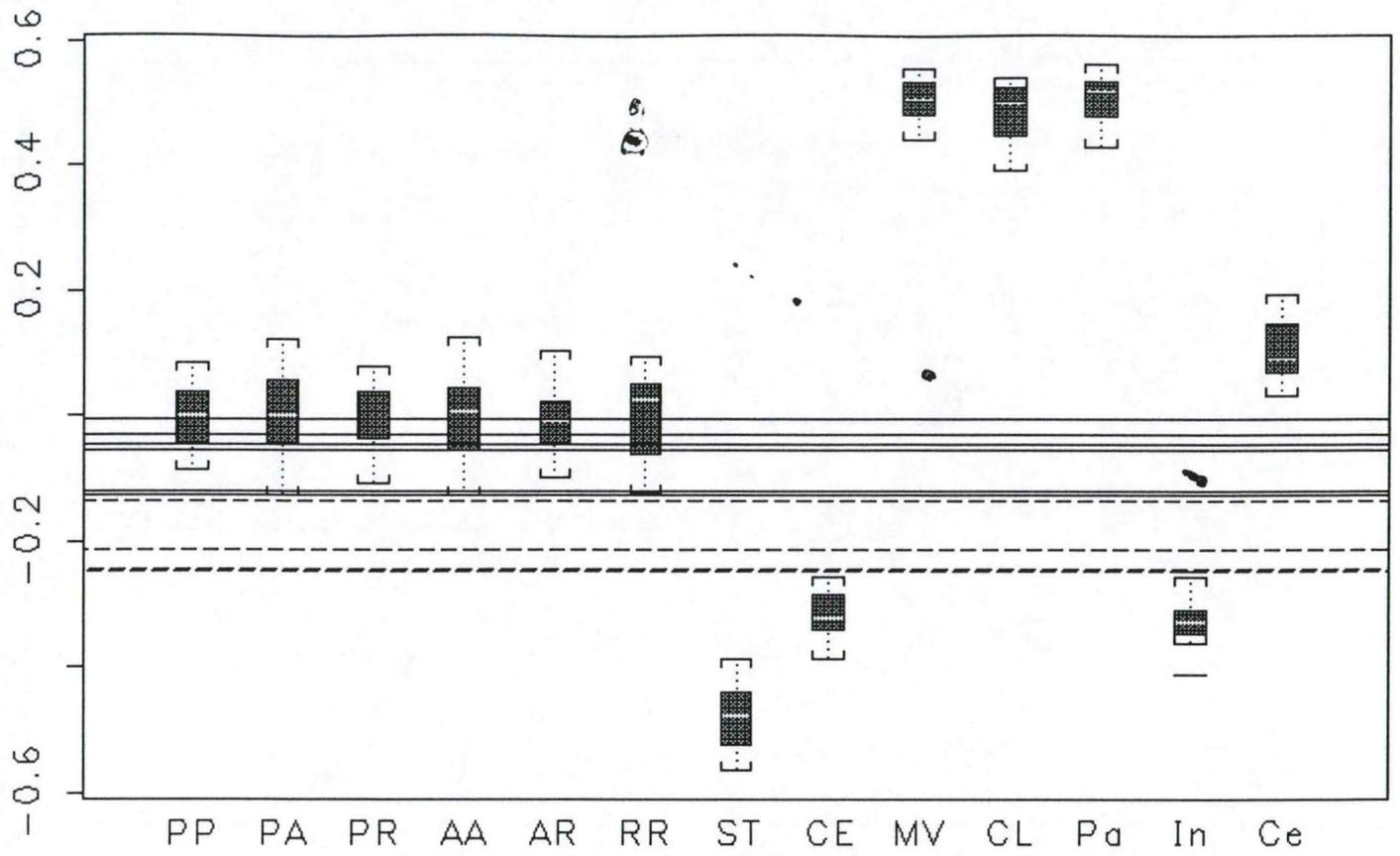
Les résultats sont beaucoup moins nets pour la forêt guyanaise que pour la forêt canadienne : les tests ne concordent pas complètement, et amènent plus à dégager des tendances que des conclusions ; les indices reflètent cette incertitude. Cela n'a rien d'étonnant,

- pour une raison statistique : les effectifs des espèces sur les parcelles de Paracou sont faibles (du moins en comparaison avec le Canada), même pour les espèces les plus abondantes, et les méthodes statistiques sont donc moins puissantes. Quand on regroupe toutes les parcelles cependant, on atteint des effectifs comparables à ceux du Canada. Les tests indiquent alors qu'il y a indépendance des répartitions marginales.

- pour une raison biologique : la forêt guyanaise est bien sûr beaucoup plus complexe que la forêt canadienne. La richesse spécifique laisse supposer des interactions complexes entre espèces, et regarder simplement les interactions entre espèces deux à deux est certainement trop partiel.

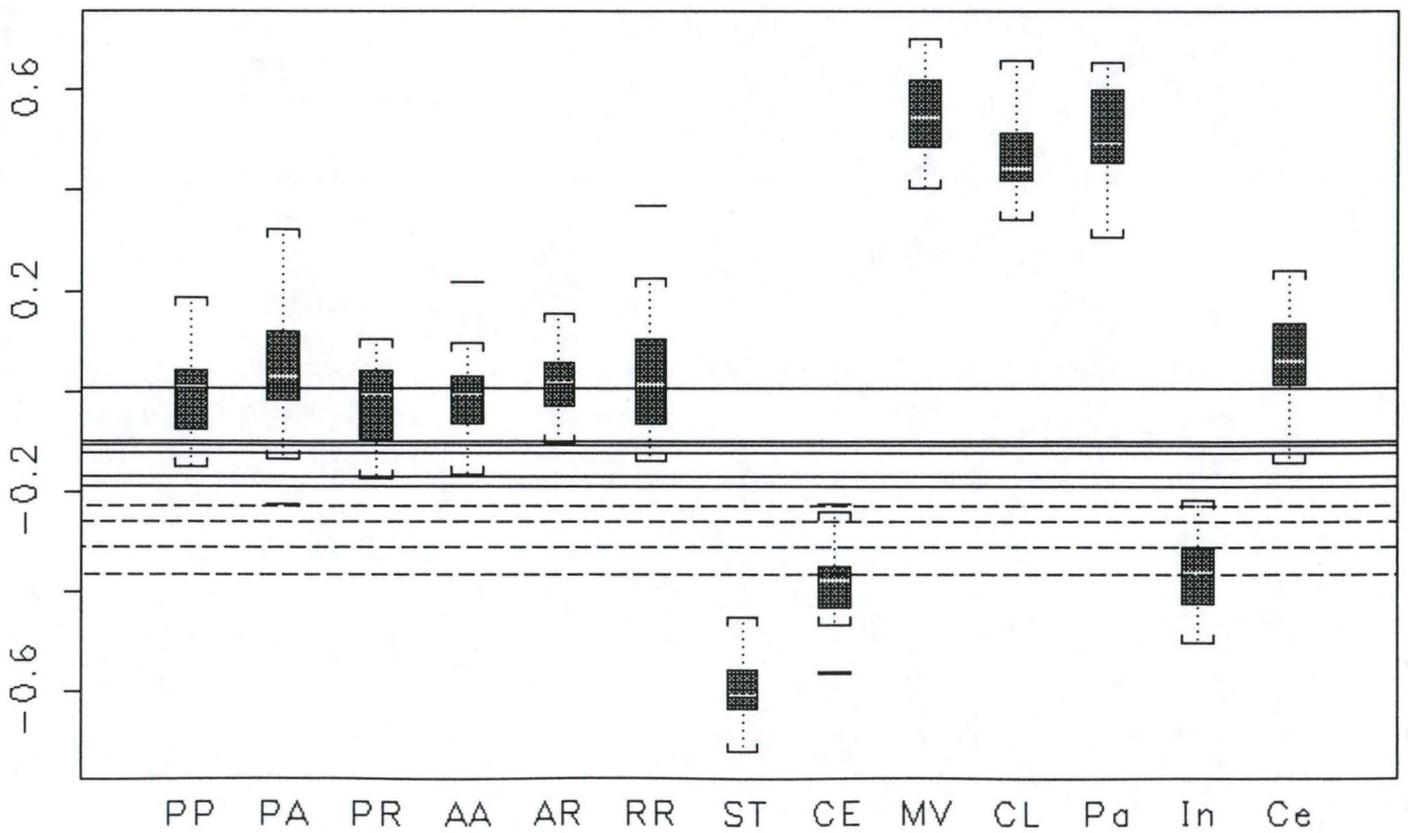
Néanmoins si l'on s'en tient aux résultats obtenus, l'indépendance entre deux quelconques des trois espèces étudiées ressort presque toujours. Seule la parcelle 8 montre une répulsion entre le kimboto et le wapa. Que cette répulsion soit le fruit d'une coïncidence, ou qu'elle traduise une divergence écologique entre les deux espèces, l'examen d'une répartition spatiale ne permet pas de trancher.

Figure 4.4



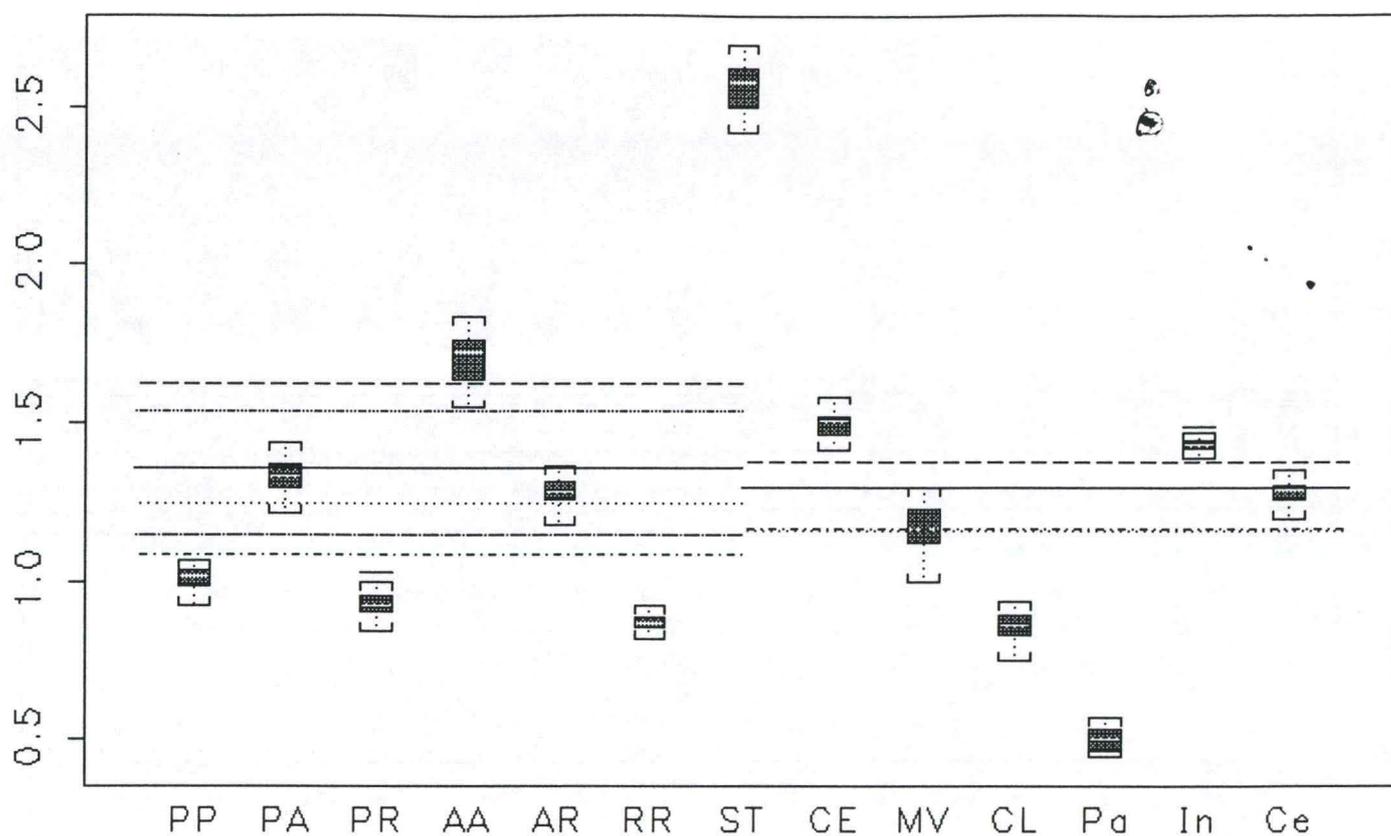
Indice de Diggle & Cox

Figure 4.5



Indice des quadrats

Figure 4.6



Indice des plus proches voisins

Legende des figures 4.4, 4.5, 4.6

- | | | | |
|-------|-------------------|-------|--------------------|
| ———— | parcelles 1,2,3,4 | ———— | parcelle 1 |
| ----- | parcelles 5,6,7,8 | | parcelle 2 |
| | | ----- | parcelle 4 |
| | | ———— | parcelle 5 |
| | | | parcelle 6 |
| | | ----- | parcelle 7 |
| | | ----- | parcelle 8 |
| | | ----- | parcelle 3 (Ab/Cw) |
| | | | parcelle 3 (Ab/Hw) |
| | | ----- | parcelle 3 (Cw/Hw) |

Figure 4.7

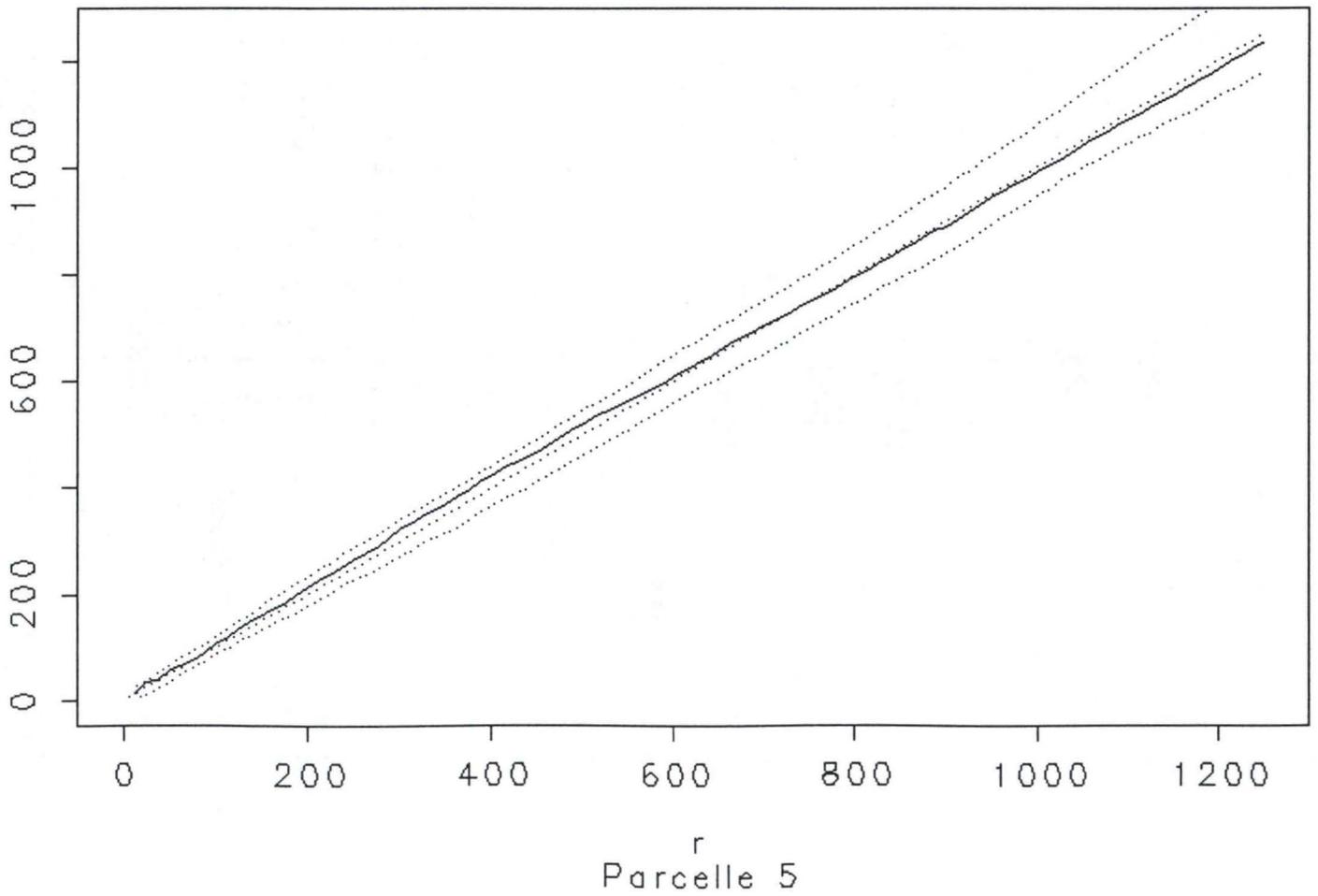
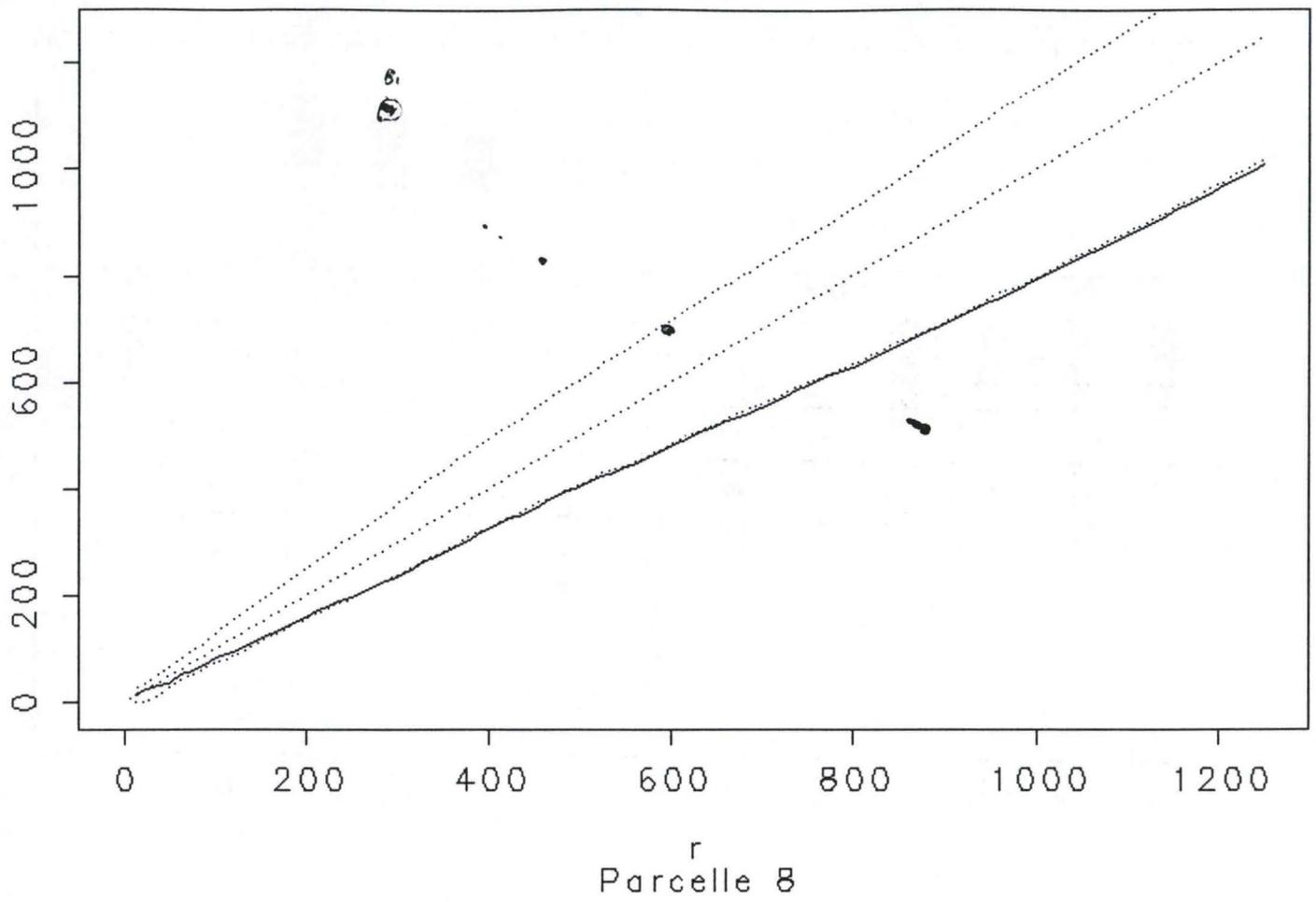
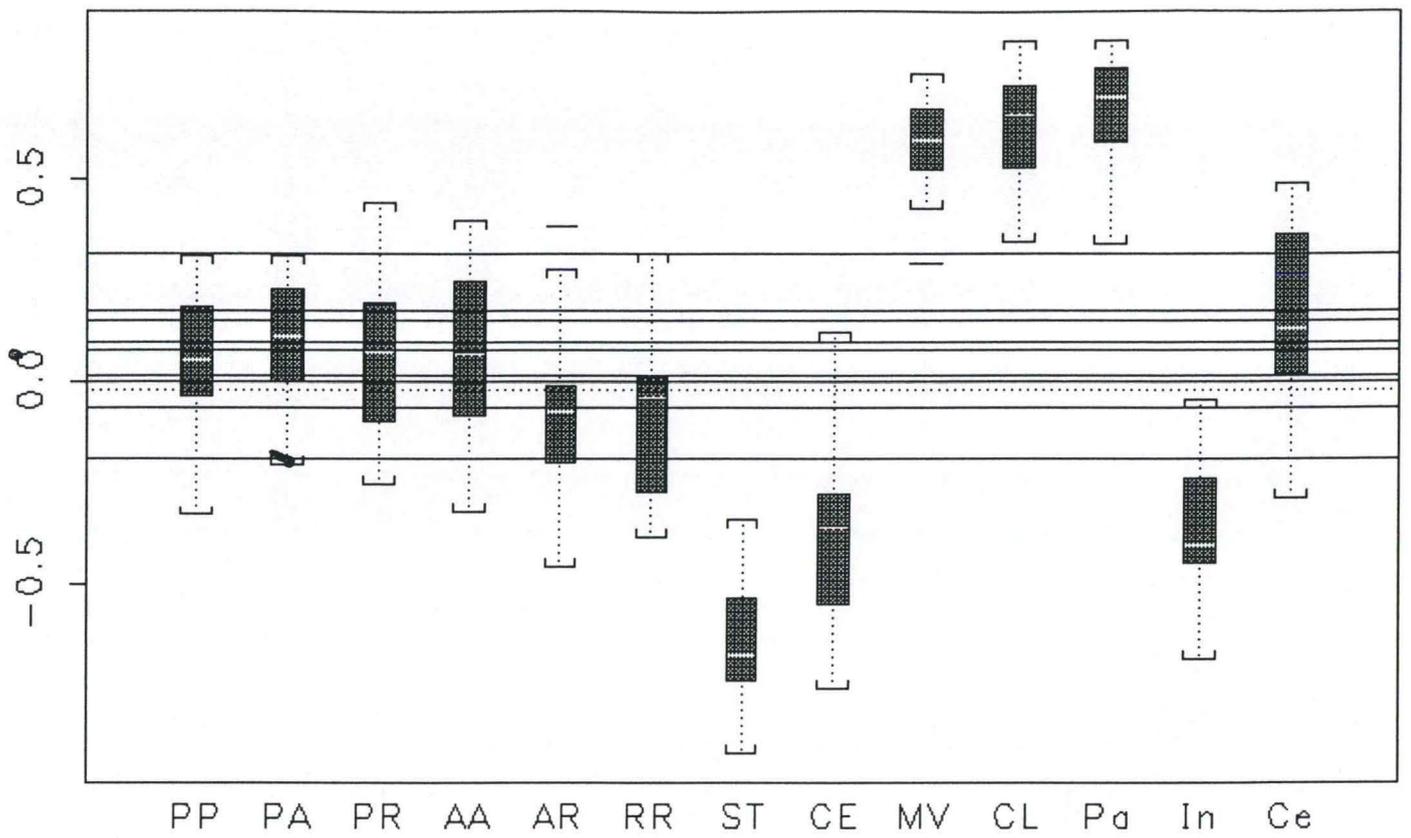
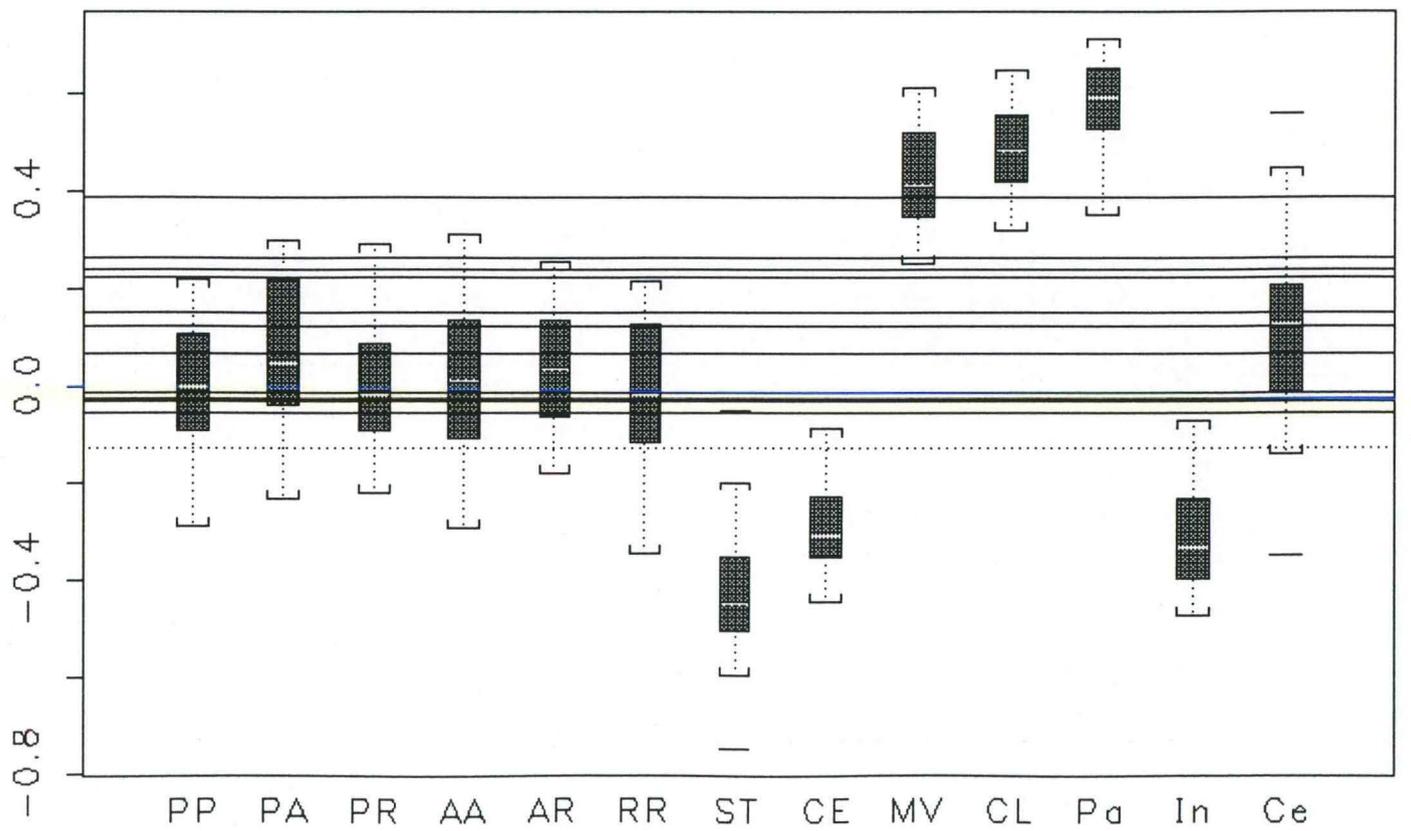


Figure 4.8



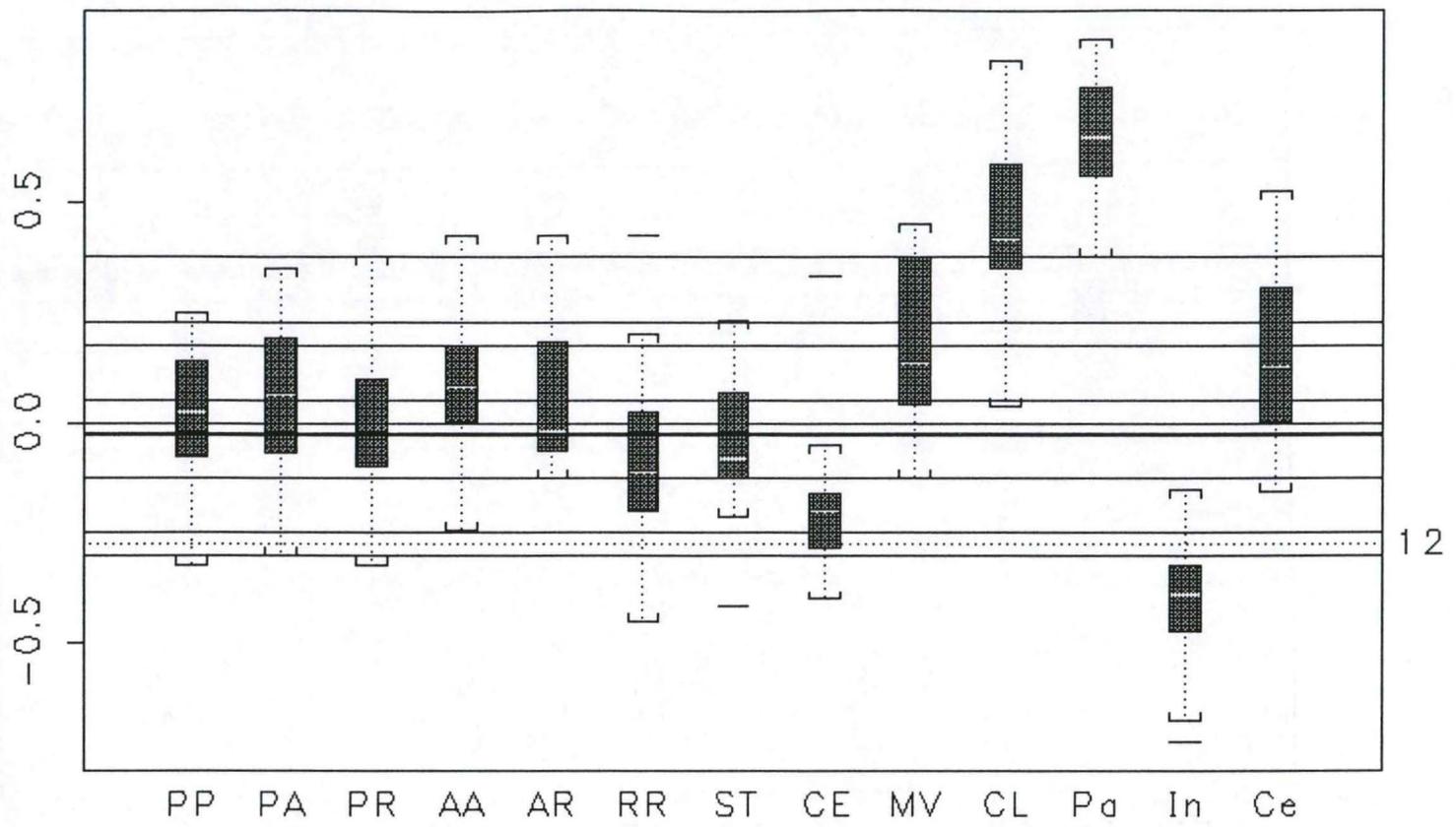
Angelique & kimboto. Methode des quadrats

Figure 4.9



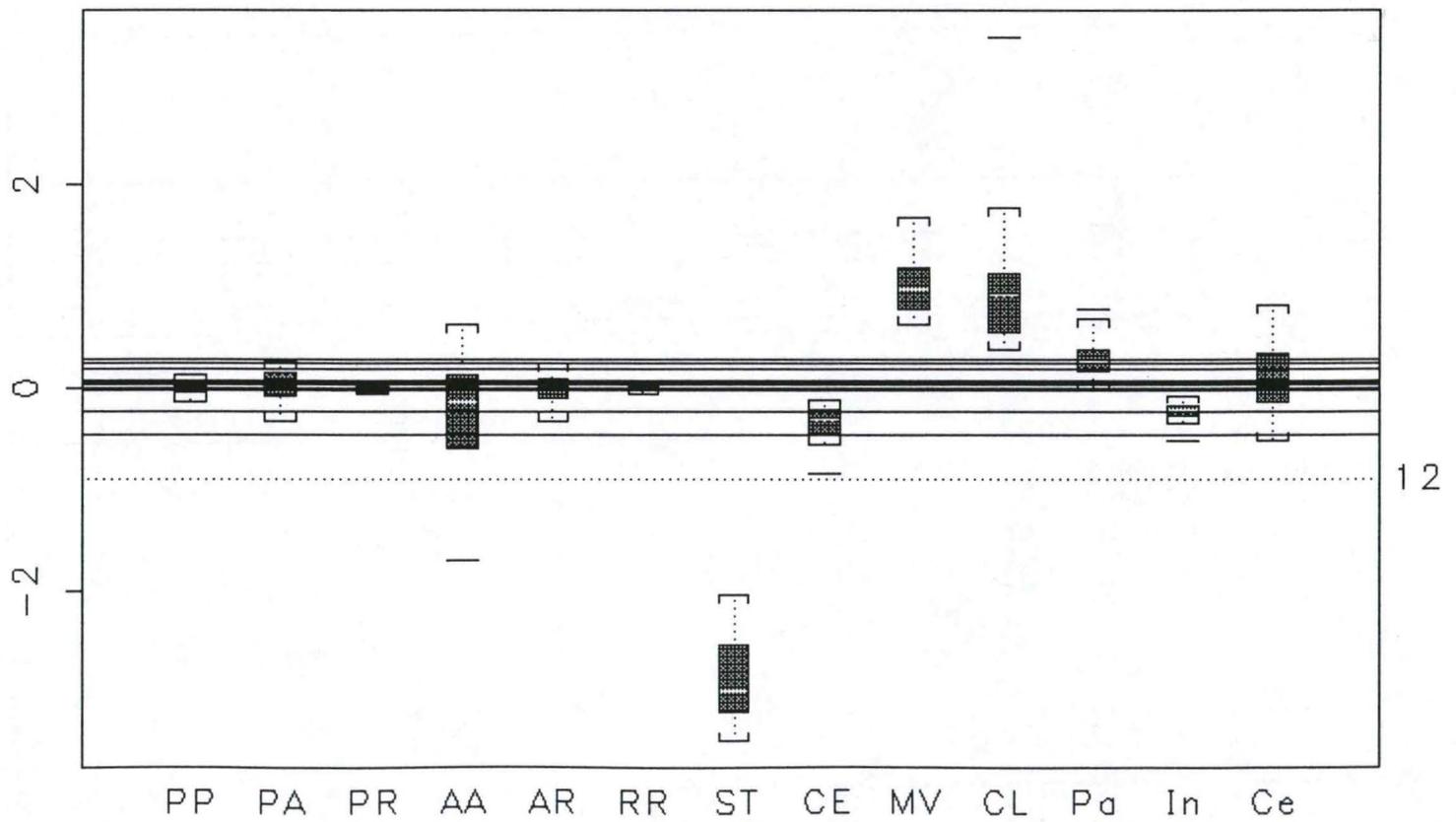
Angelique & kimboto. Methode de Diggle & Cox

Figure 4.10



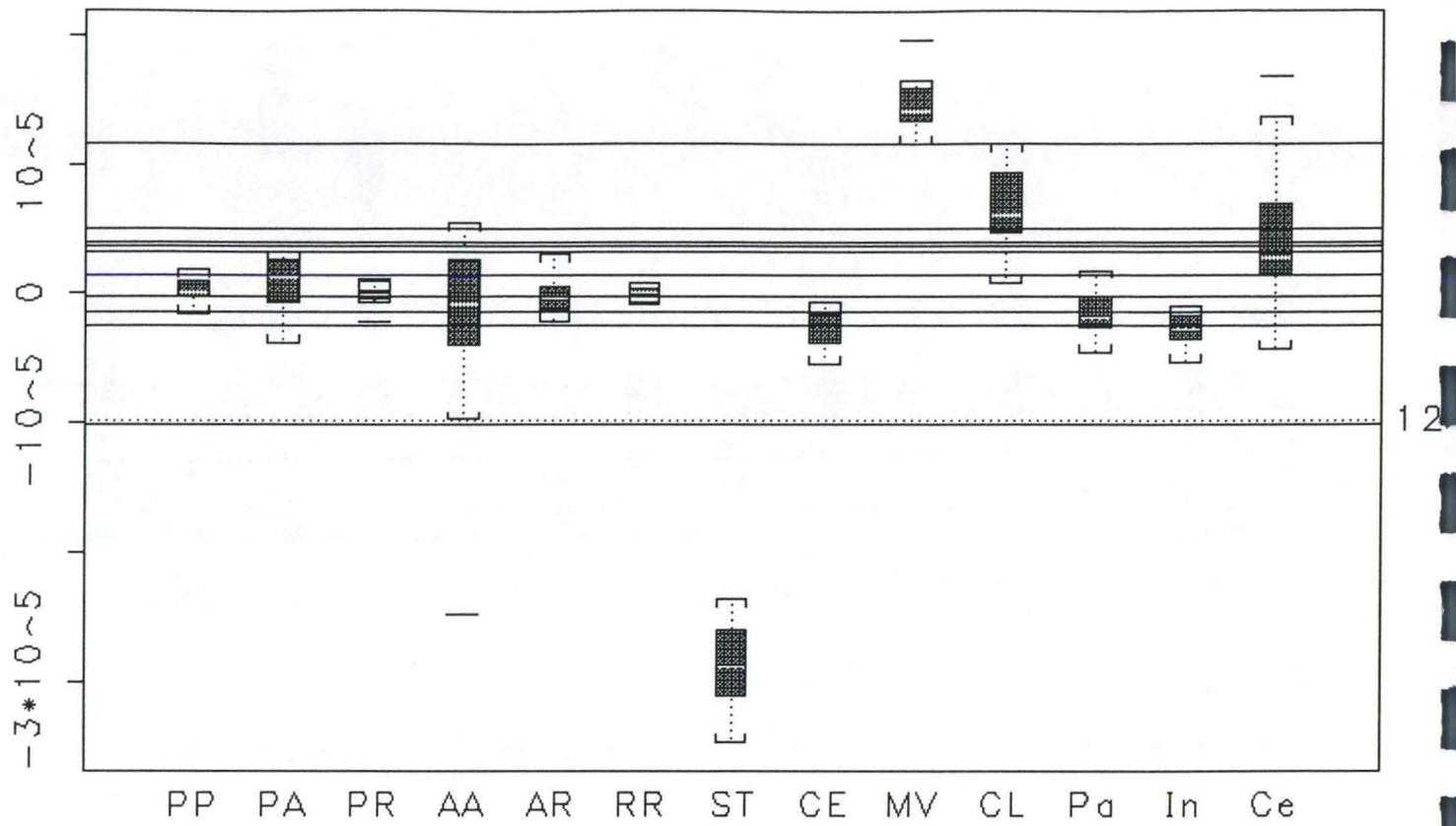
Angelique & kimboto. Methode de Goodall

Figure 4.11



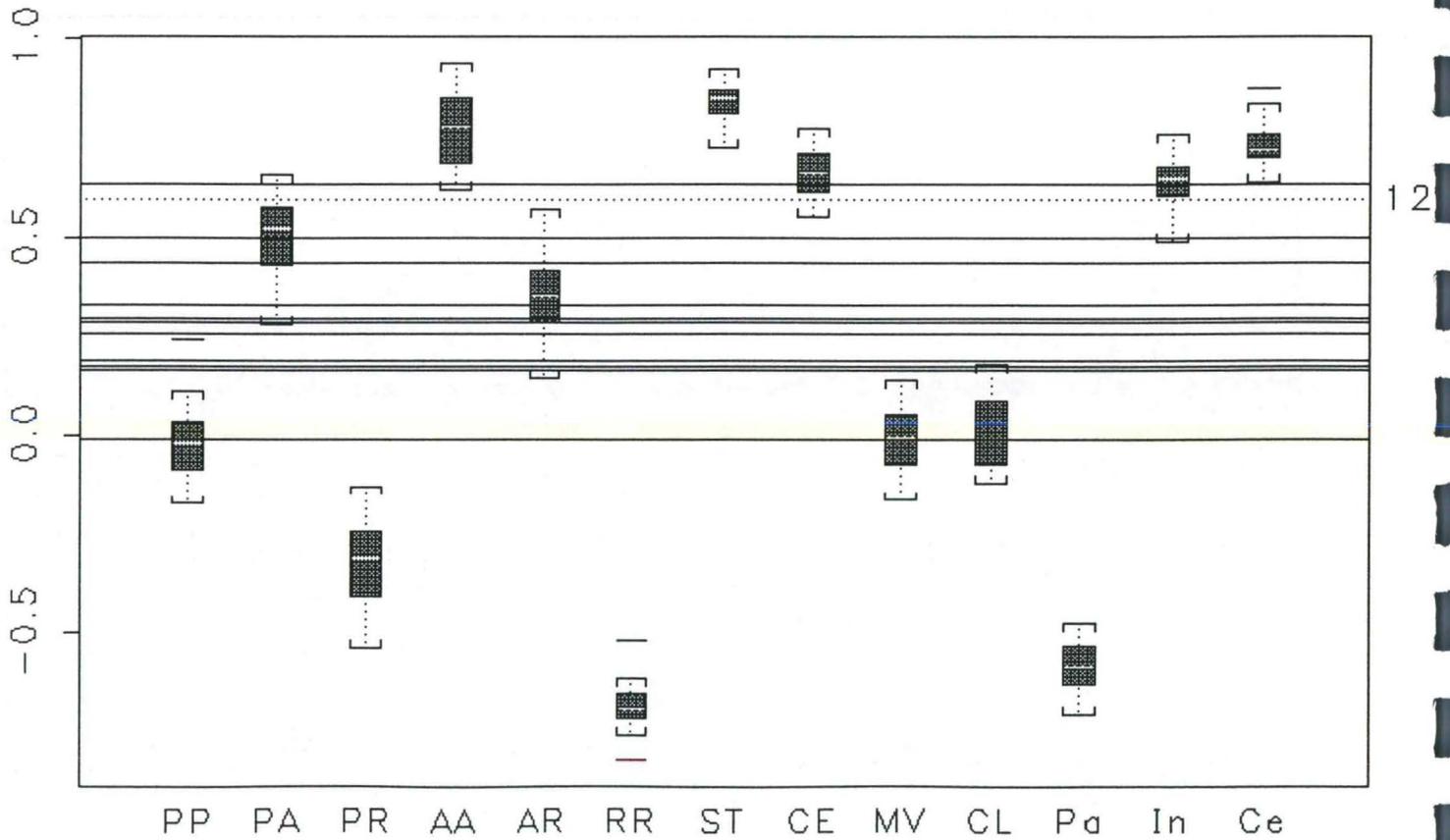
Angelique & kimboto. Methode de Kershaw

Figure 4.12



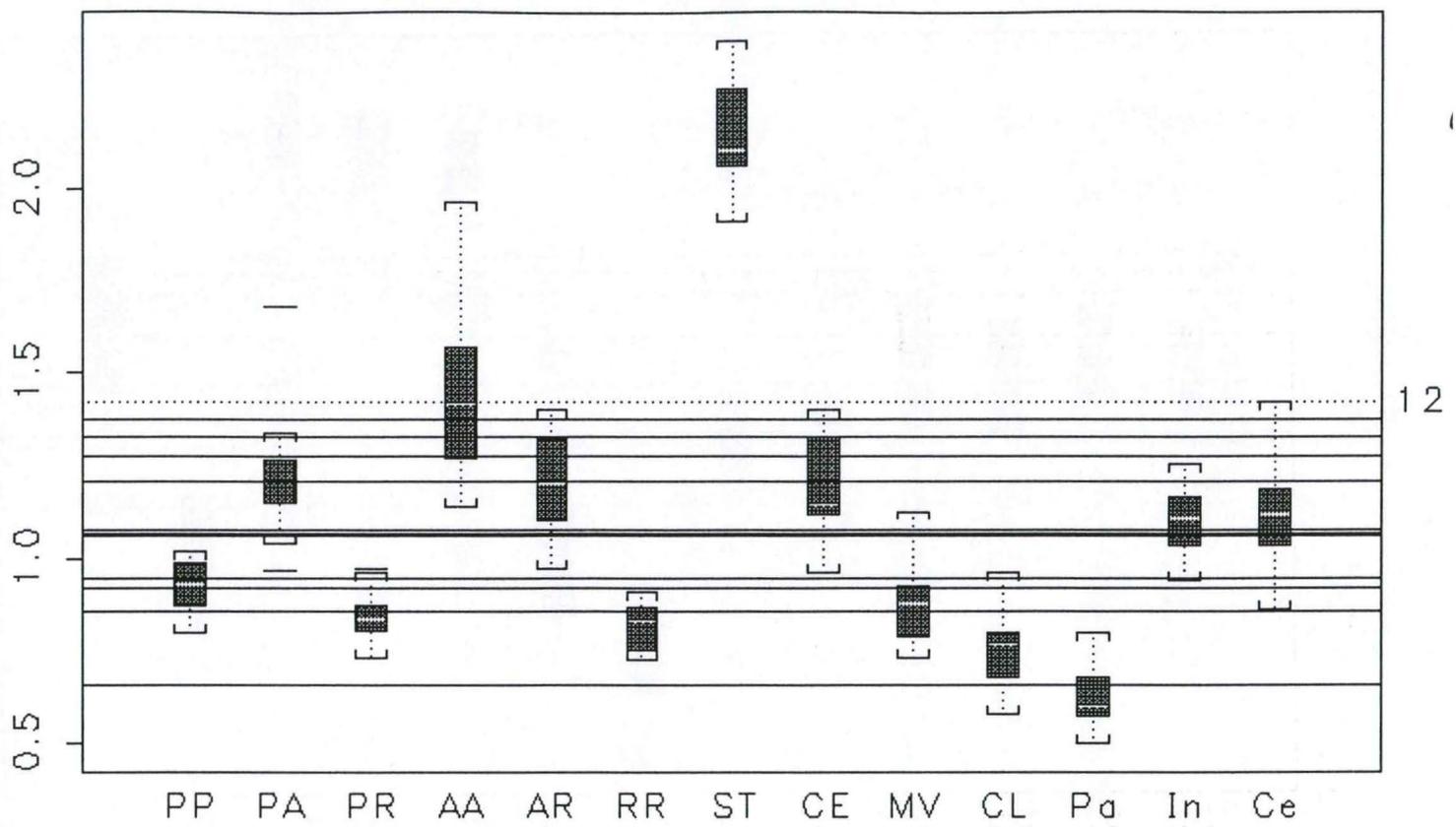
Angelique & kimboto. Methode de Ripley

Figure 4.13



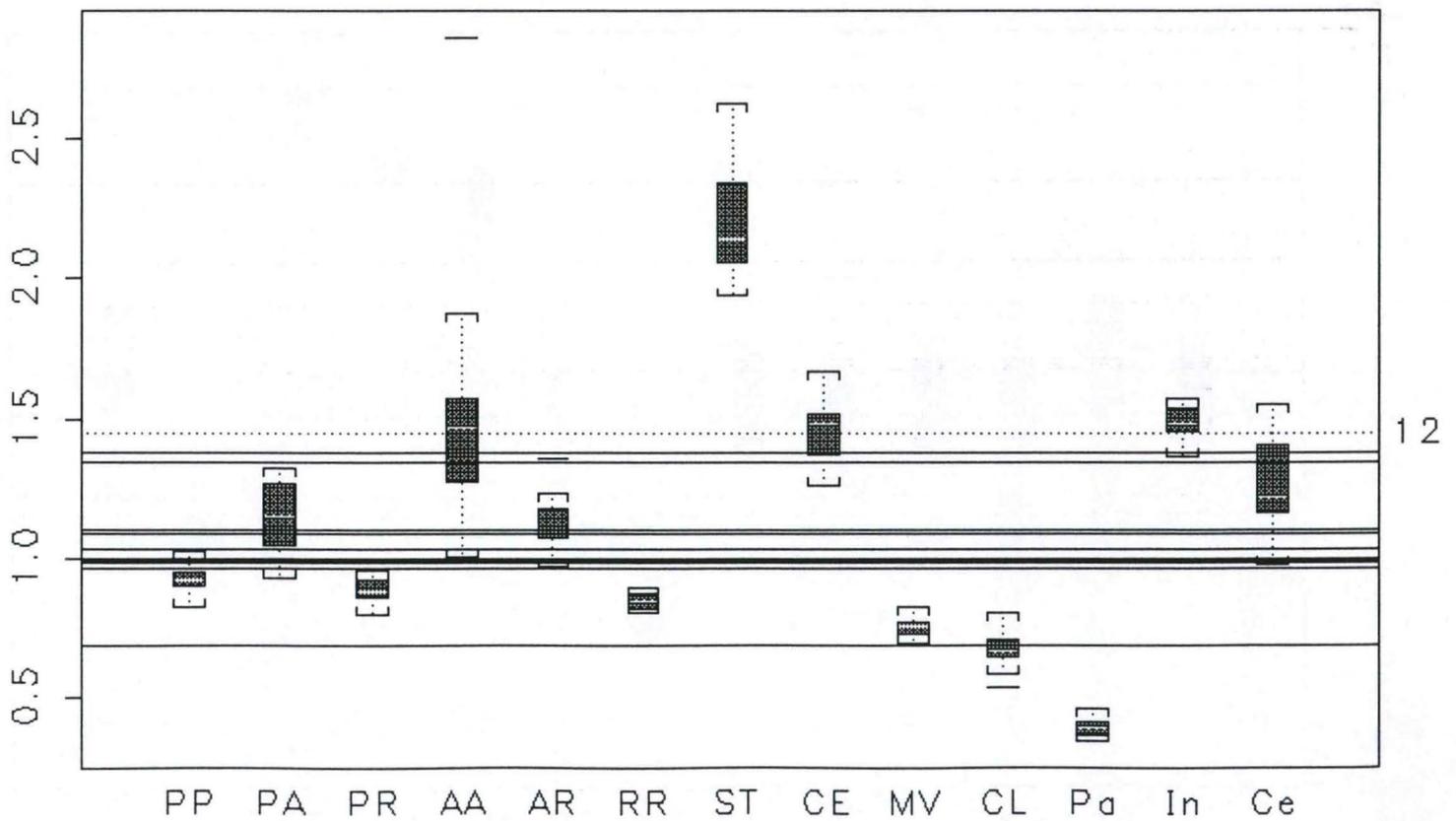
Angelique & kimboto. Methode de Pielou

Figure 4.14



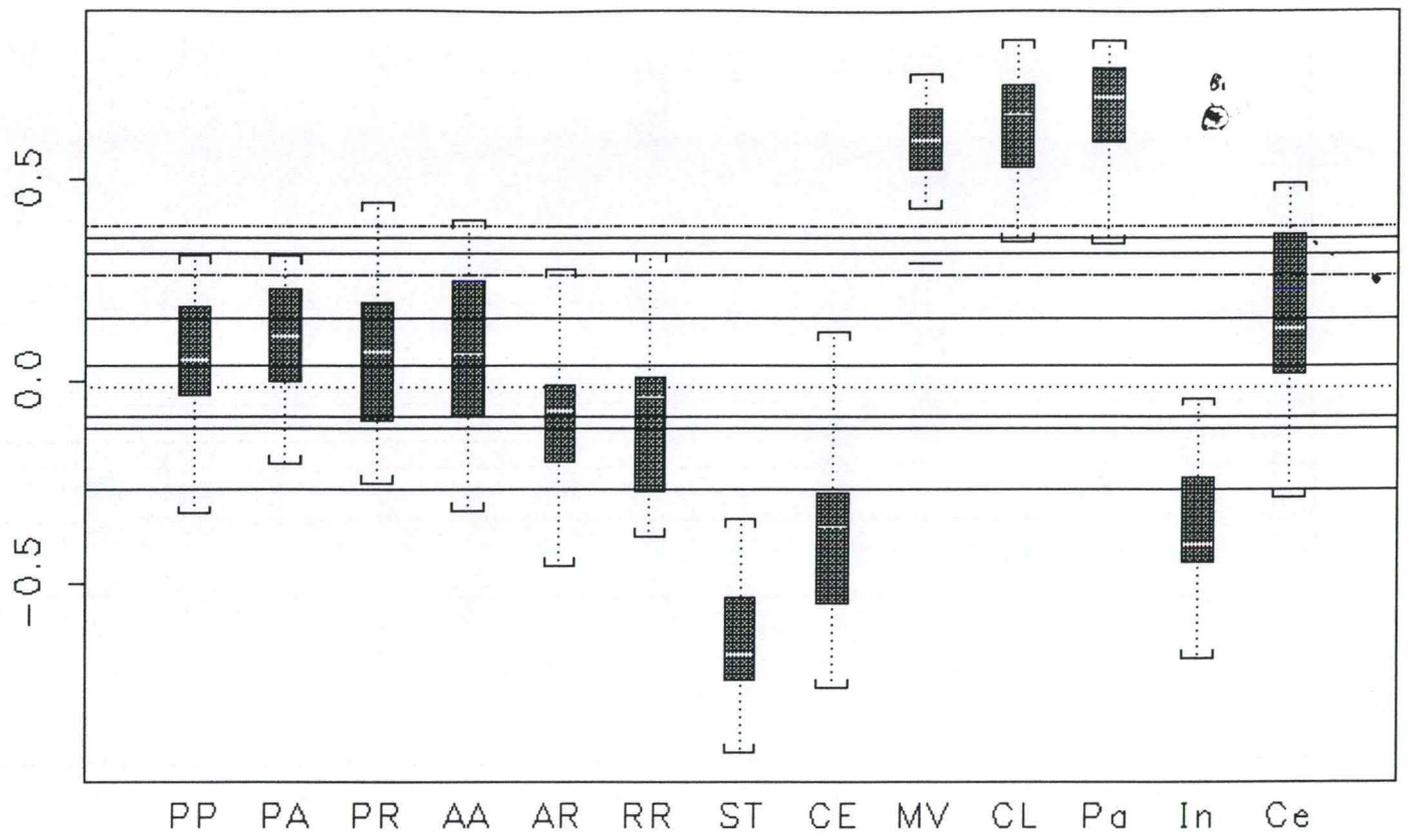
Angelique & kimboto. Methode du Sup de l'Inf

Figure 4.15



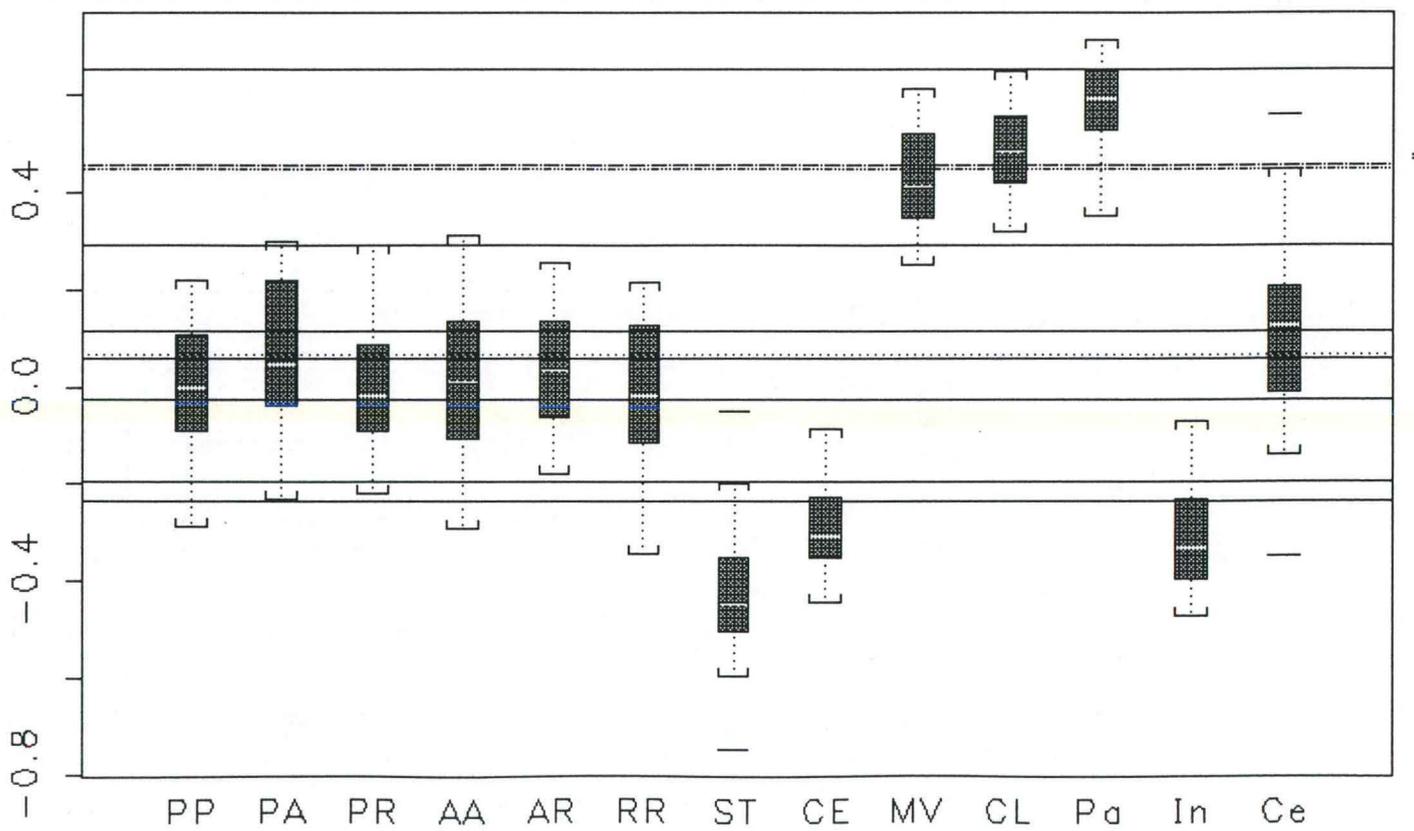
Angelique & kimboto. Methode des plus proches voisins

Figure 4.16



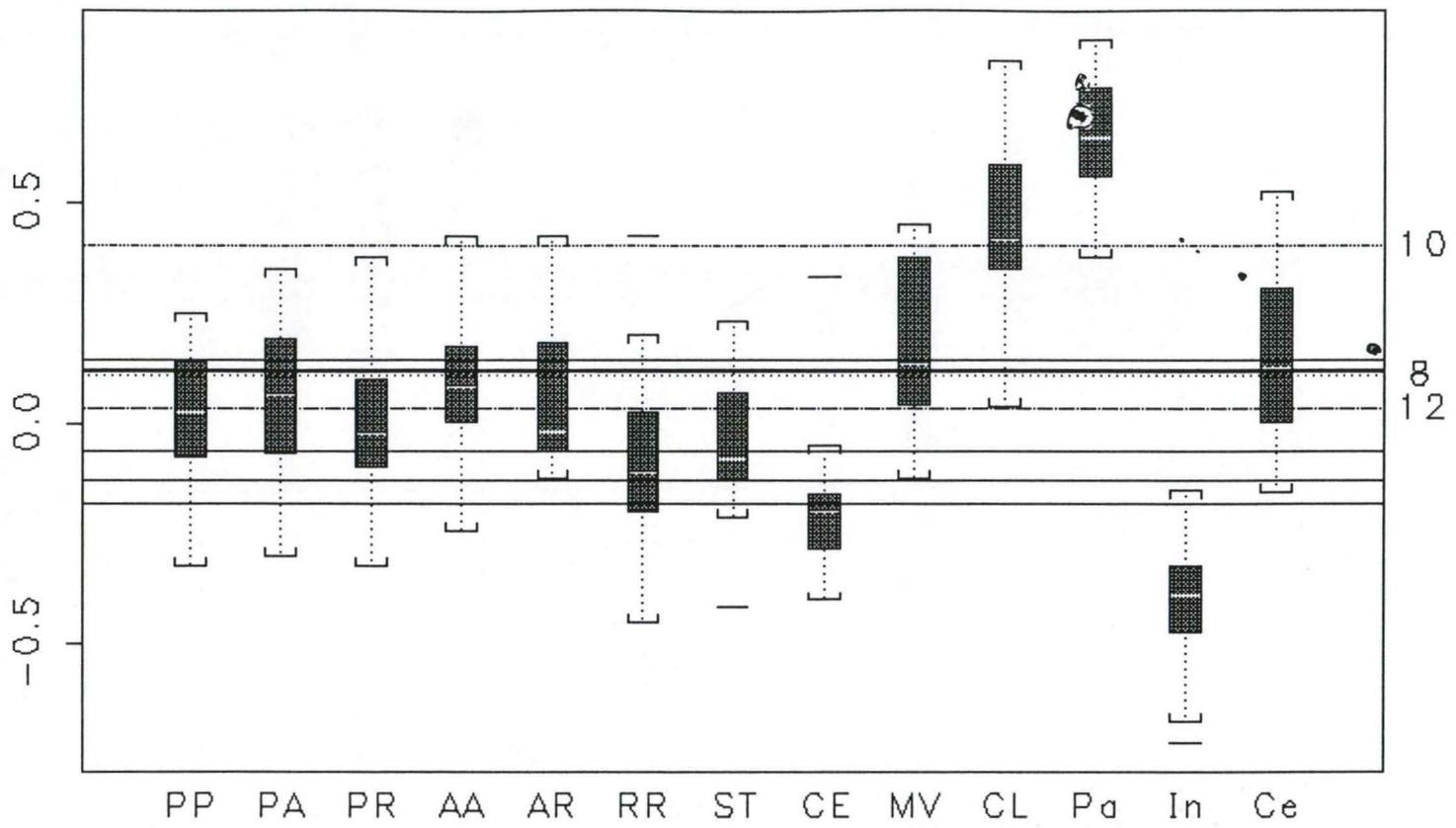
Angelique et wapa. Methode des quadrats

Figure 4.17



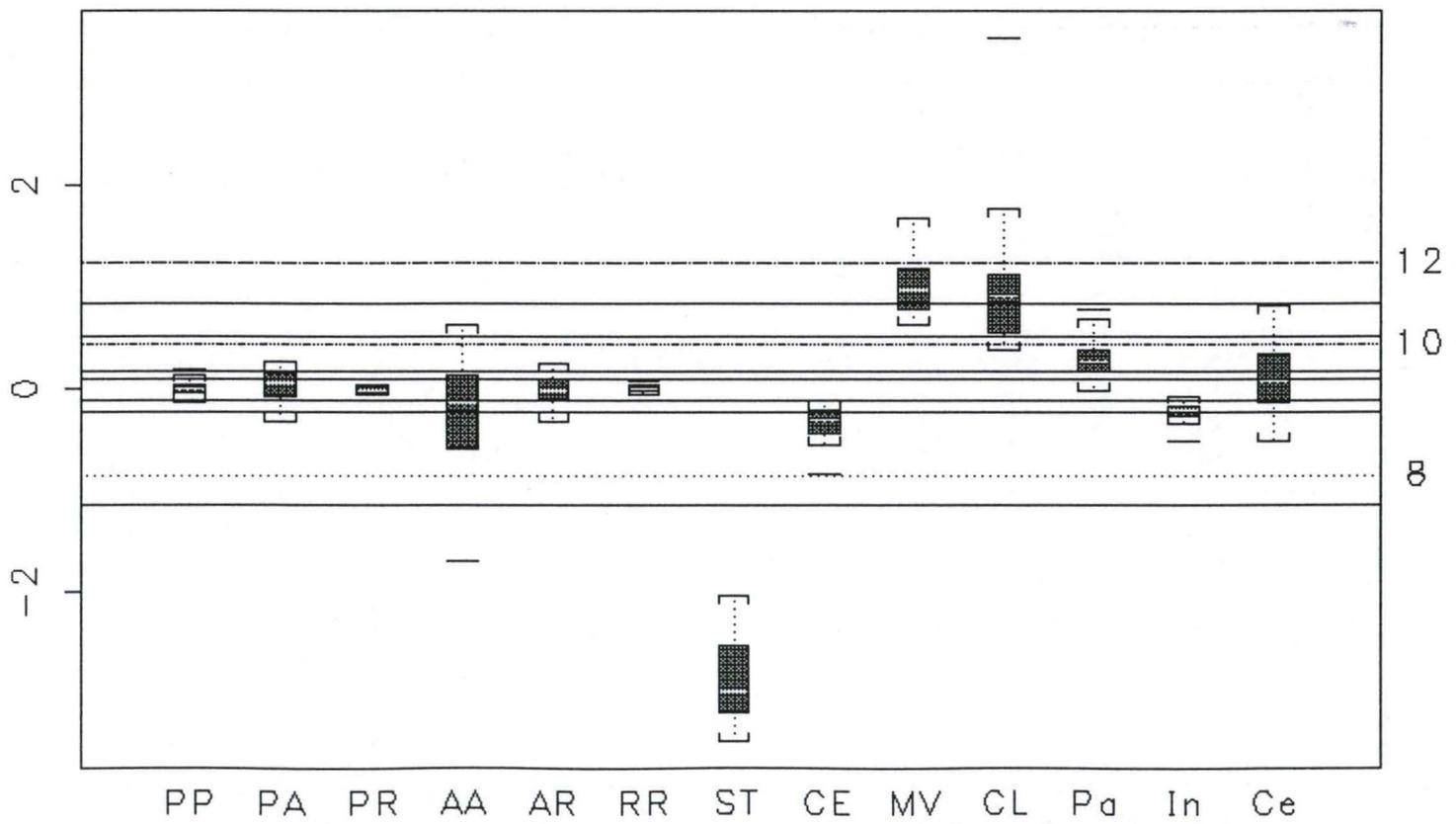
Angelique et wapa. Methode de Diqgle & Cox

Figure 4.18



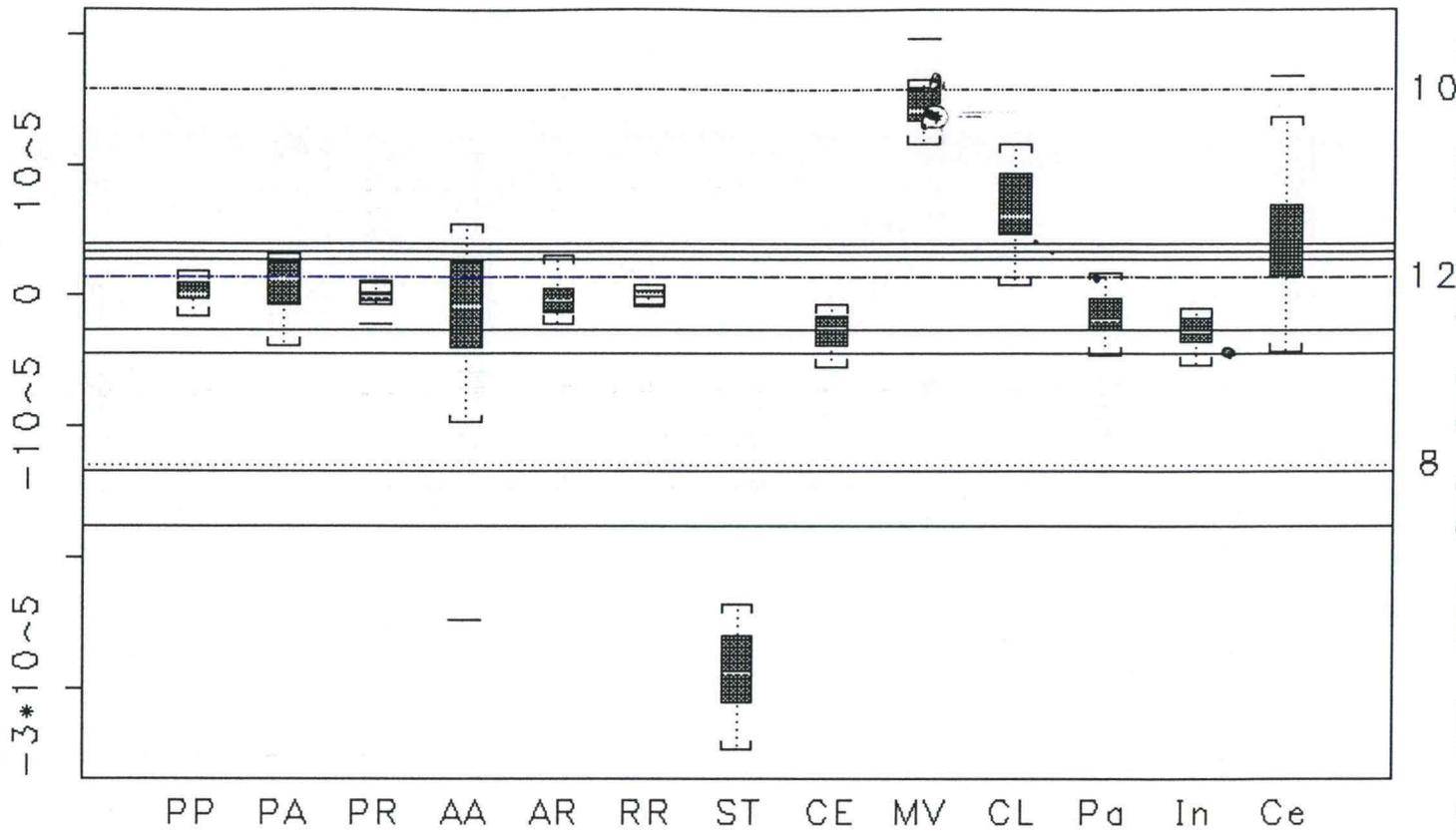
Angelique et wapa. Methode de Goodall

Figure 4.19



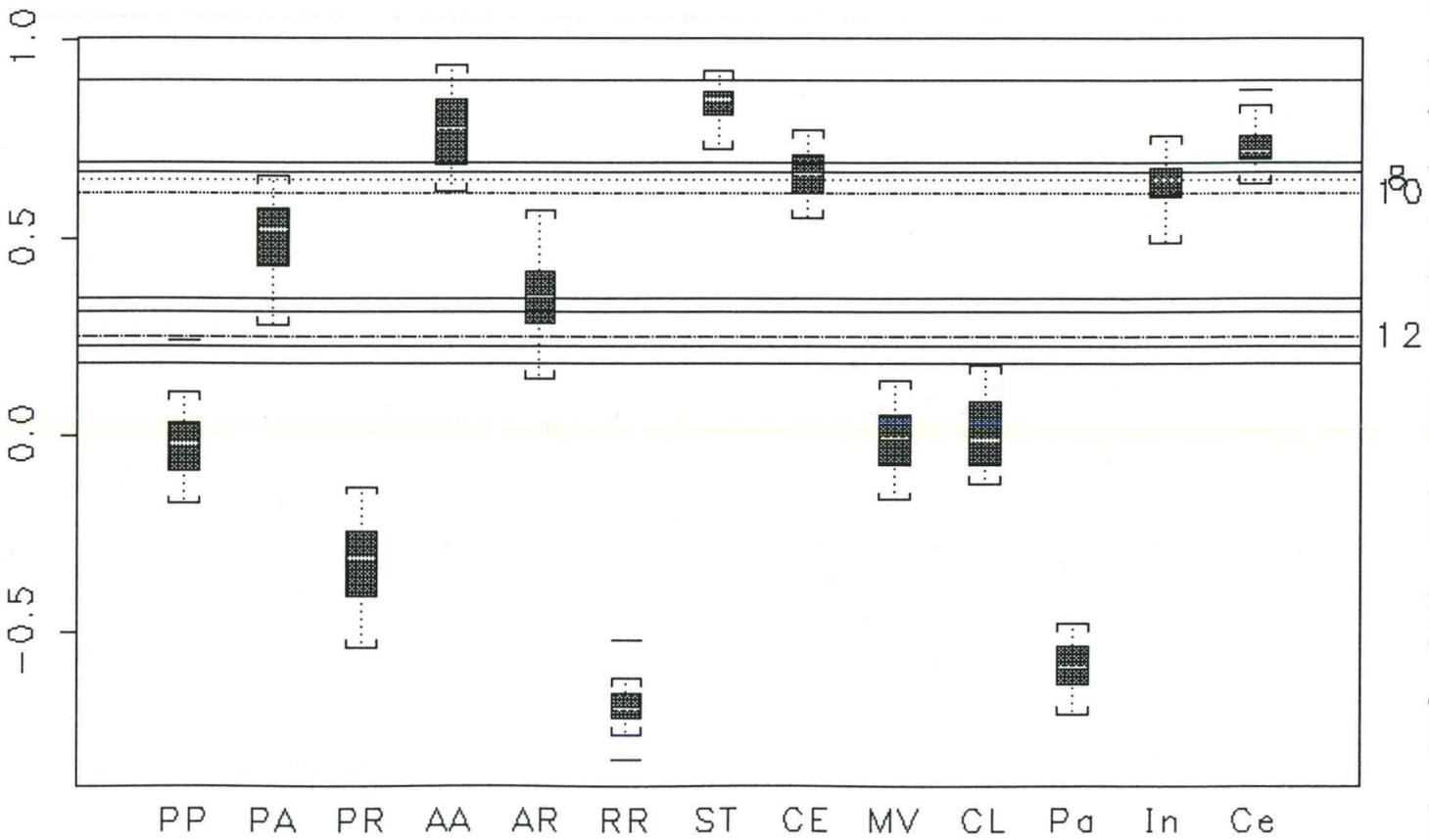
Angelique et wapa. Methode de Kershaw

Figure 4.20



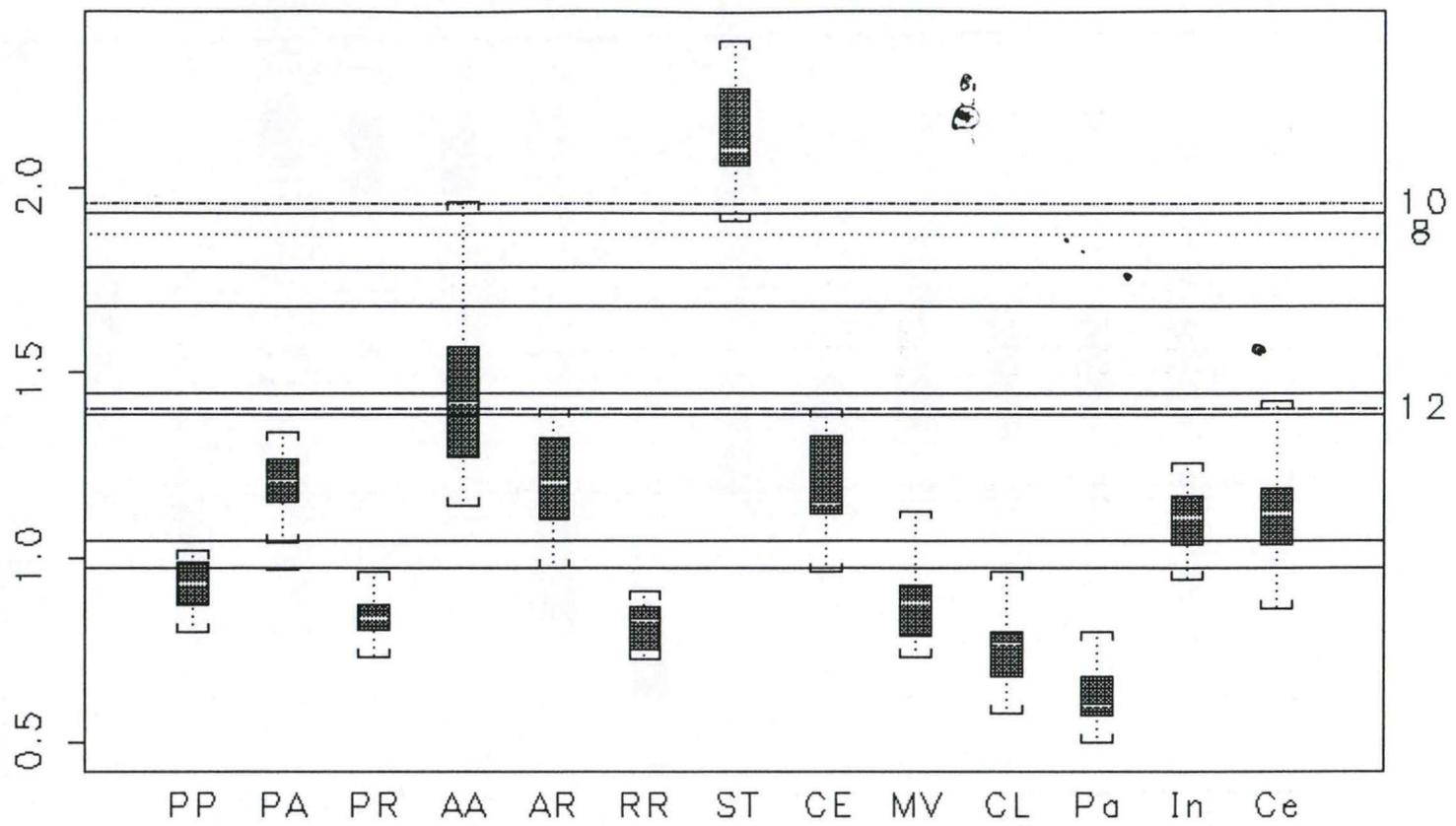
Angelique et wapa. Methode de Ripley

Figure 4.21



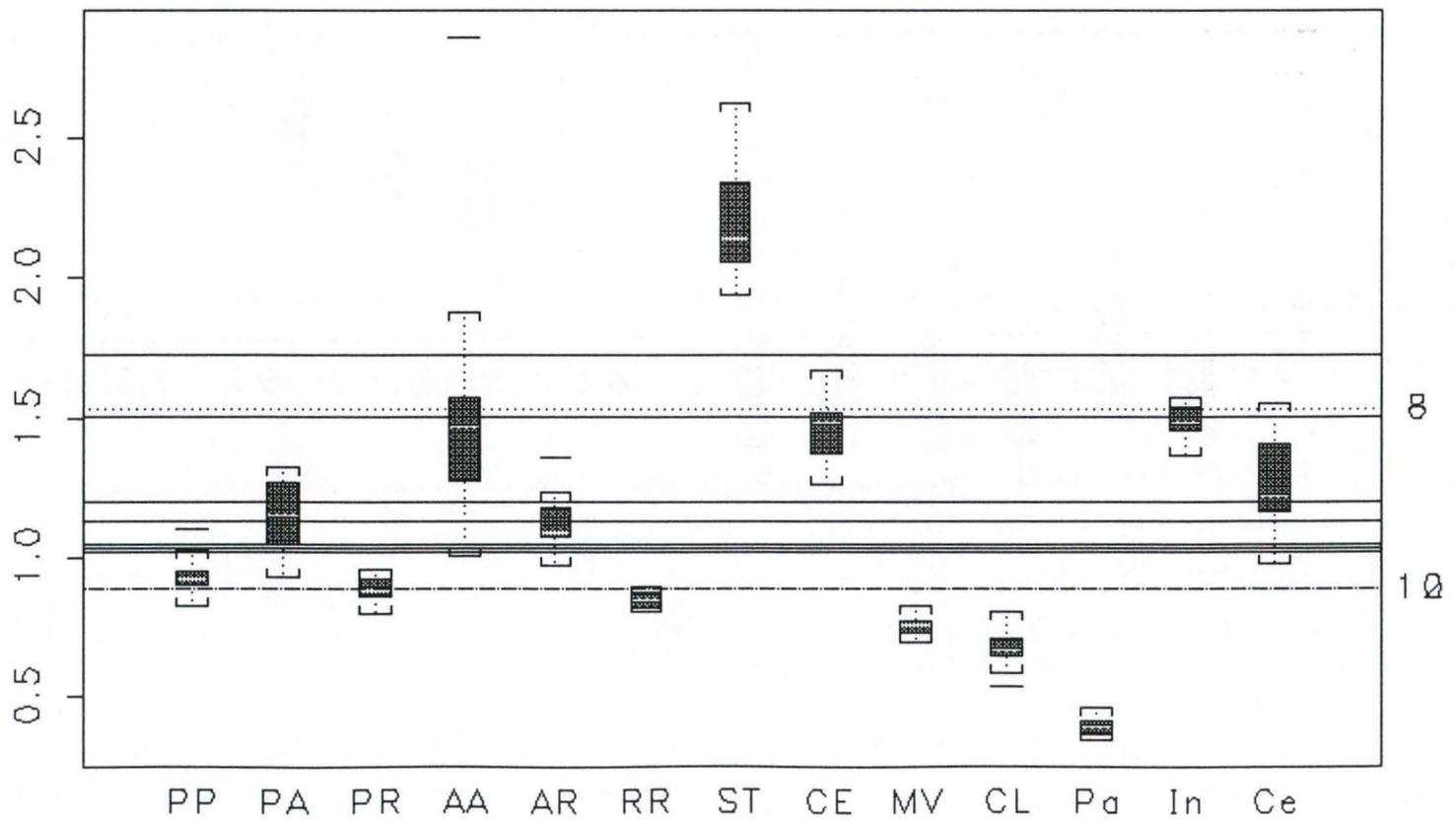
Angelique et wapa. Methode de Pielou

Figure 4.22



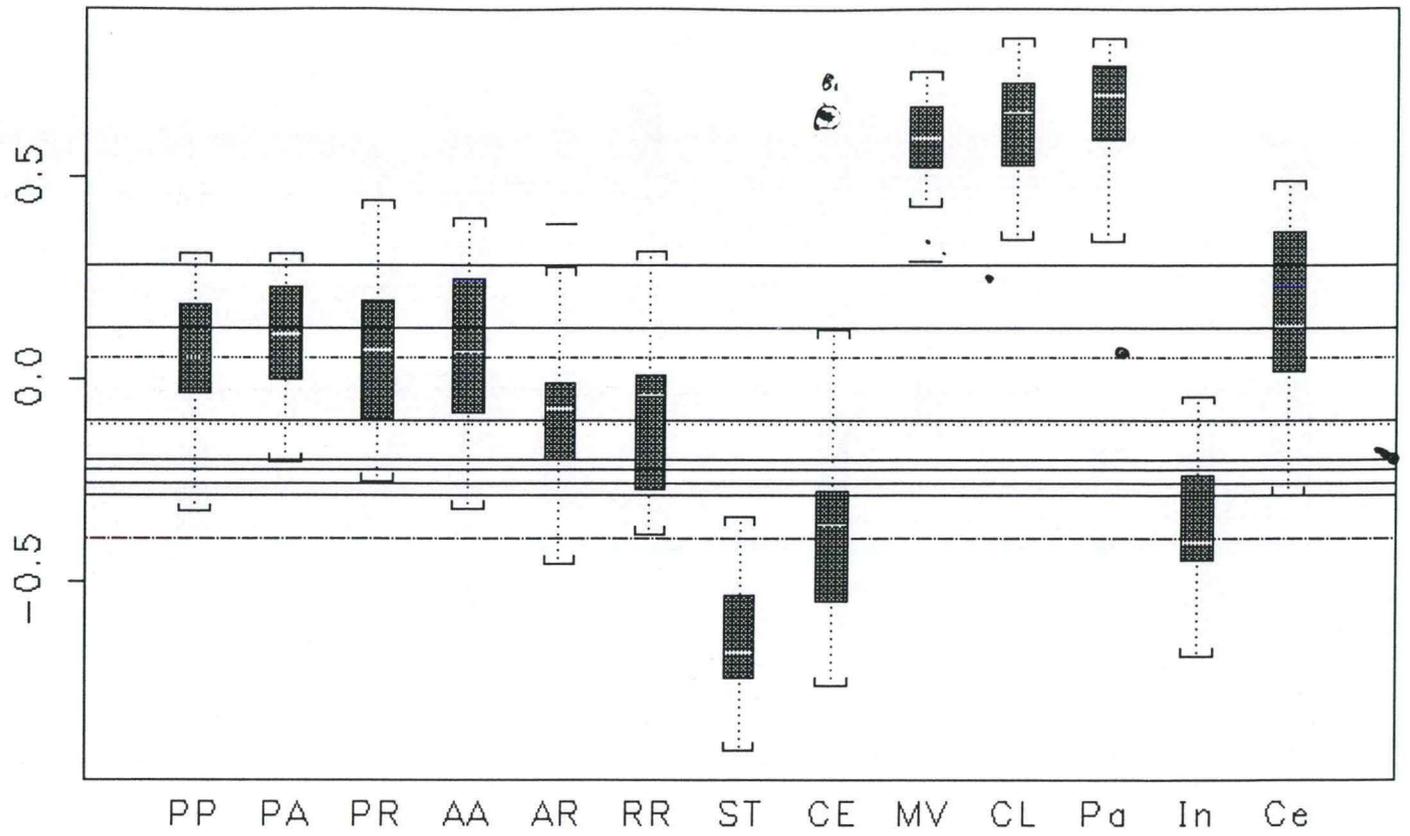
Angelique et wapa. Methode du Sup de l'Inf

Figure 4.23



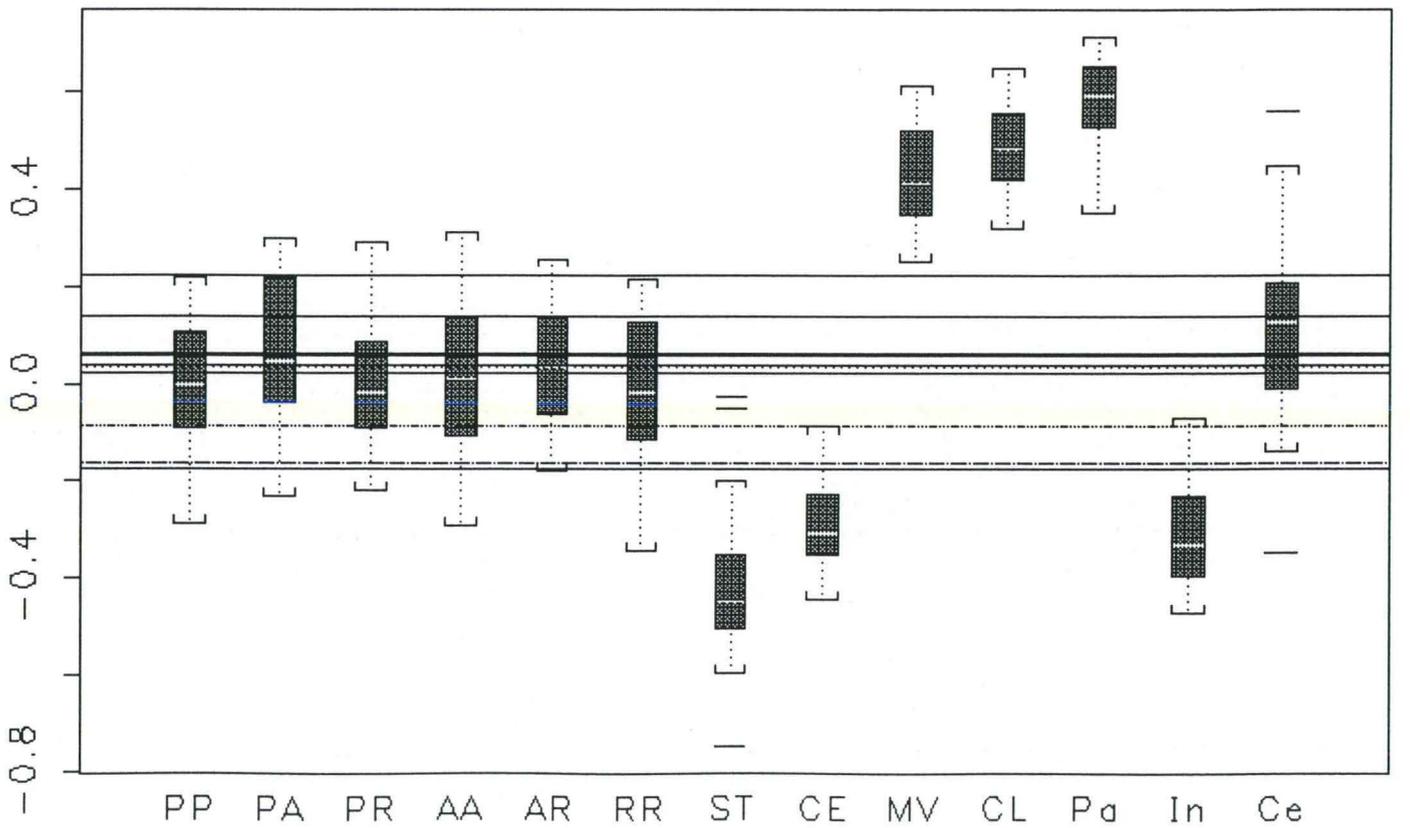
Angelique et wapa. Methode des plus proches voisins

Figure 4.24



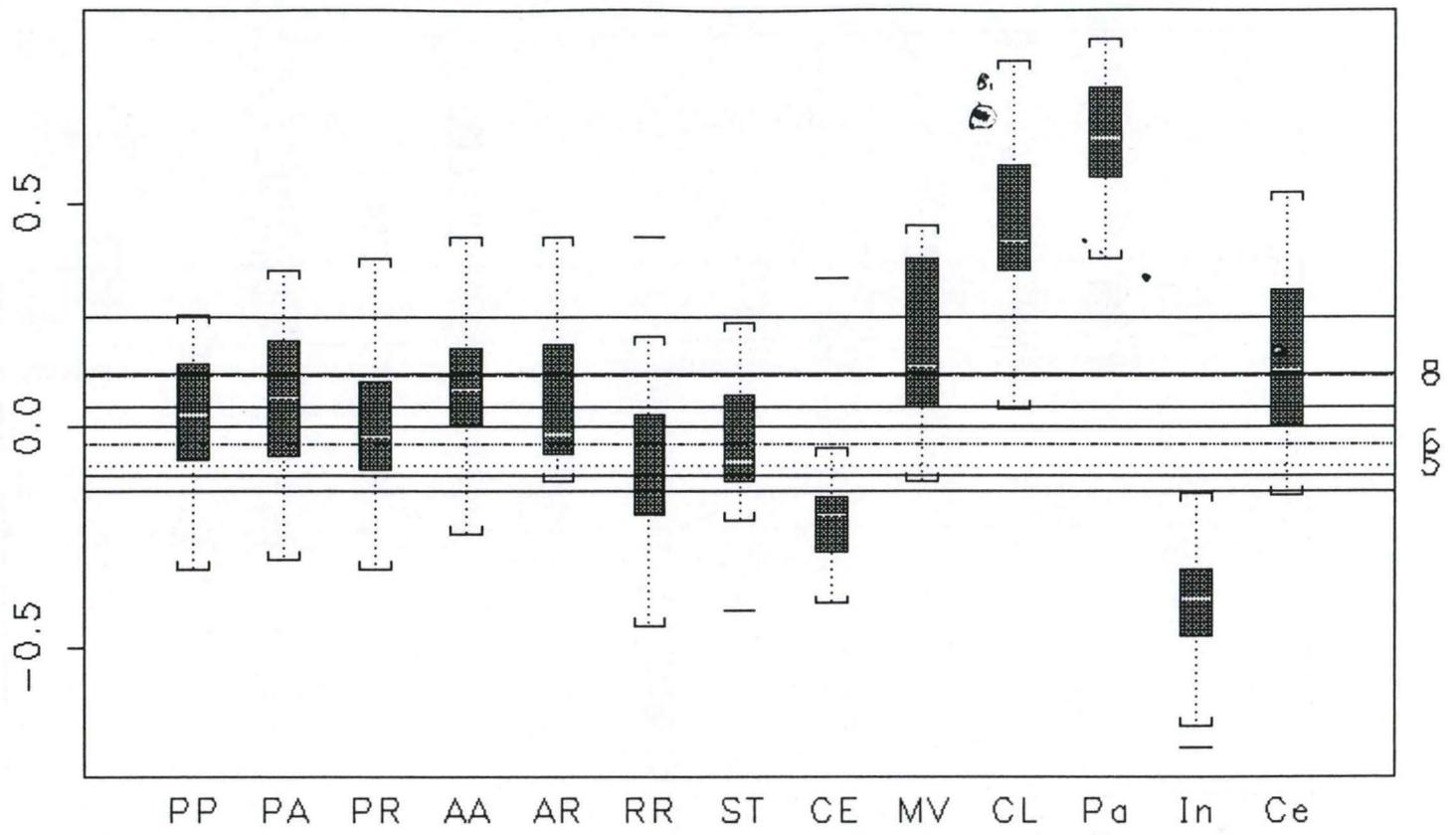
Kimboto & wapa. Methode des quadrats

Figure 4.25



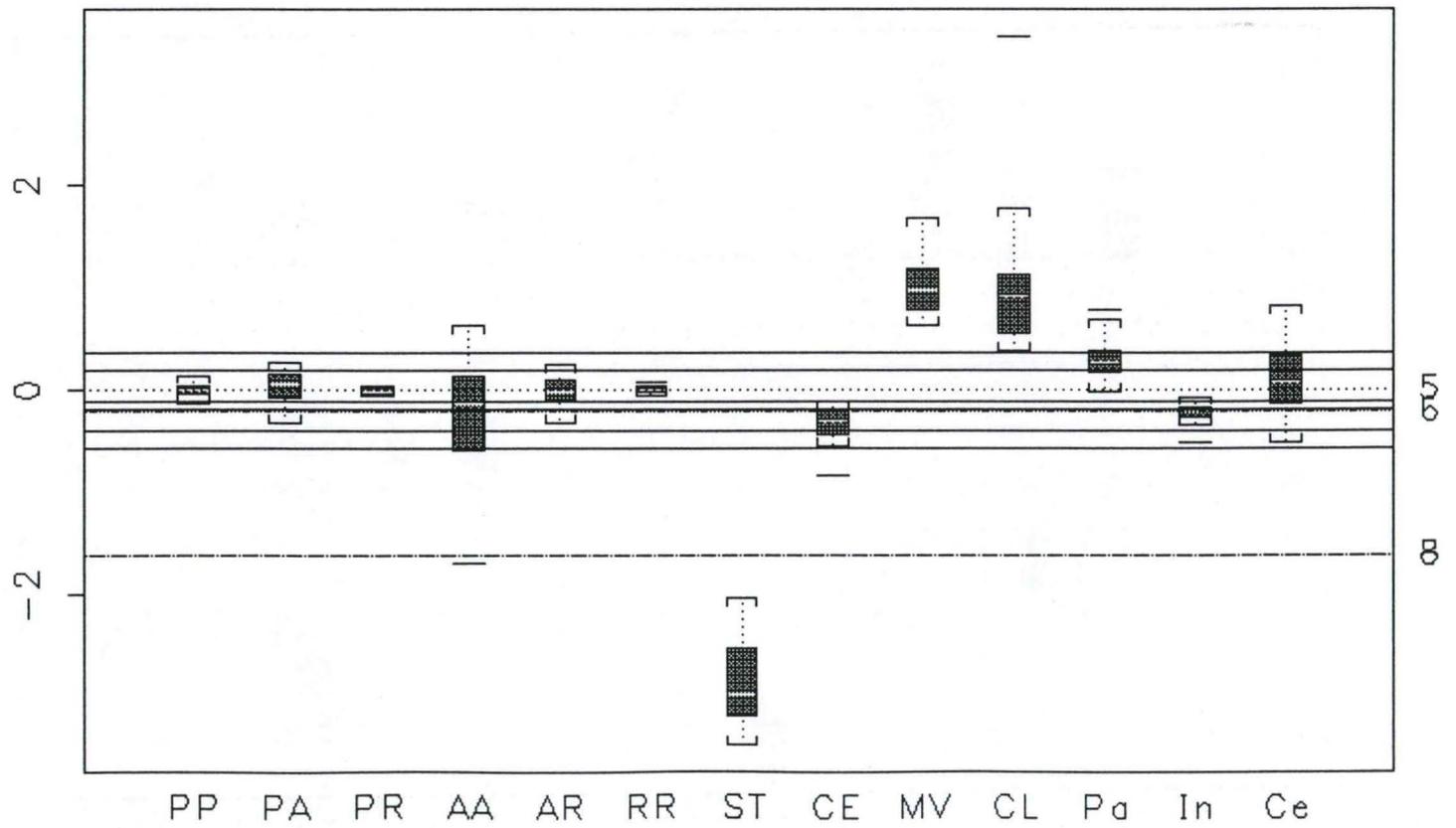
Kimboto & wapa. Methode de Diqale & Cox

Figure 4.26



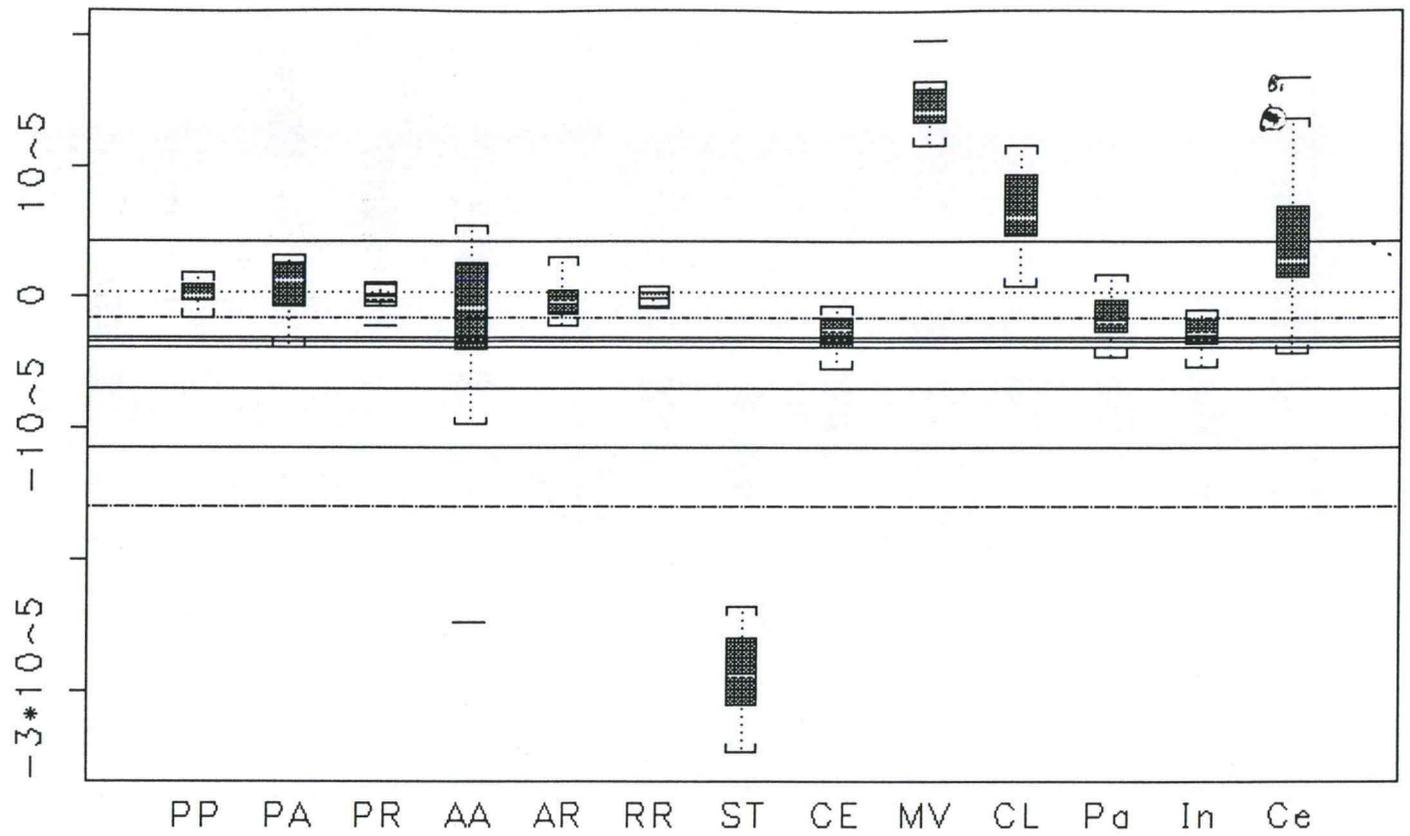
Kimboto & wapa. Methode de Goodall

Figure 4.27



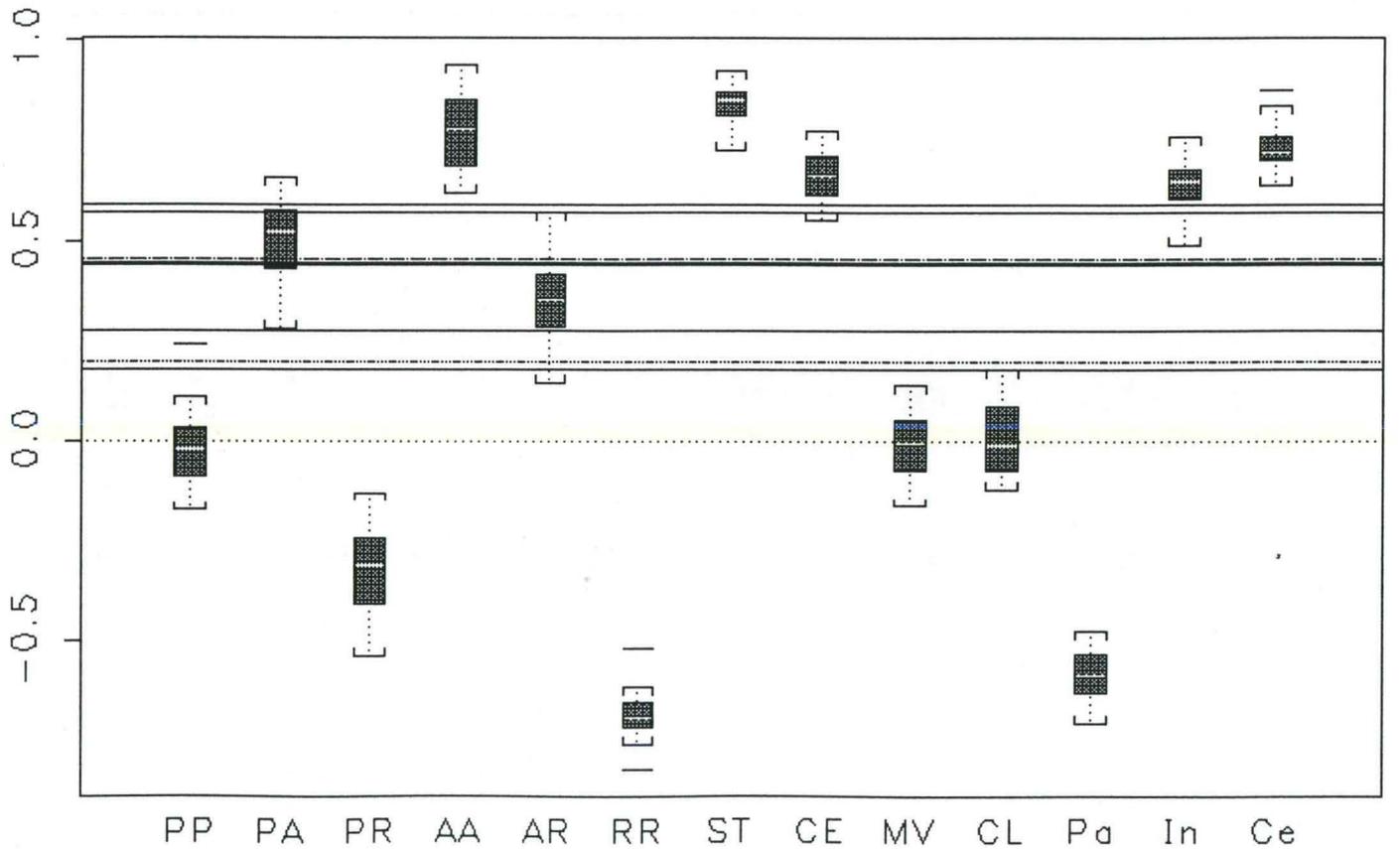
Kimboto & wapa. Methode de Kershaw

Figure 4.28



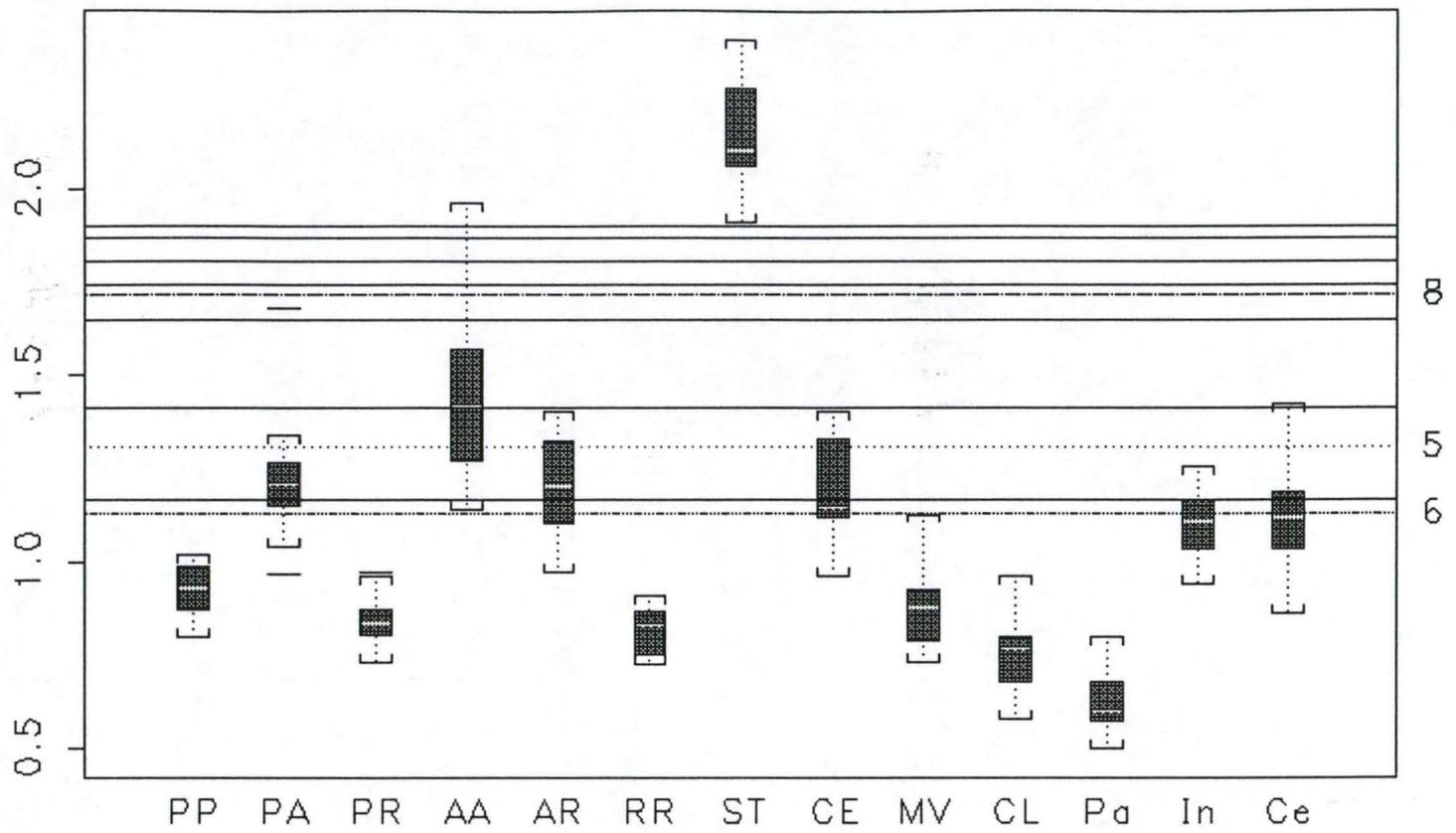
Kimboto & wapa. Methode de Ripley

Figure 4.29



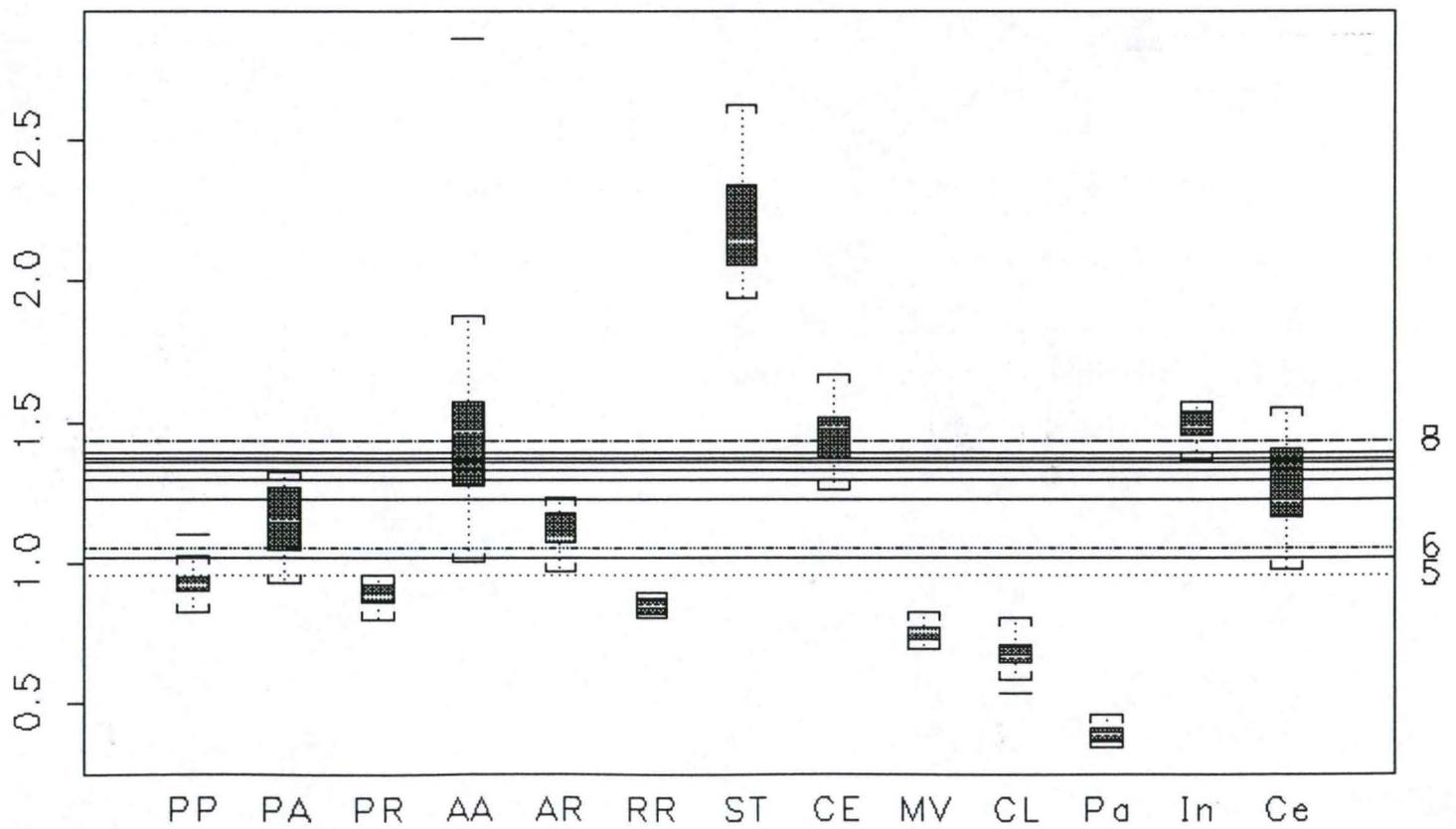
Kimboto & wapa. Methode de Pielou

Figure 4.30



Kimboto & wapa. Methode du Sup de l'Inf

Figure 4.31



Kimboto & wapa. Methode des plus proches voisins

5. CONCLUSION

Les indices étudiés ne répondent pas aux mêmes questions et demandent des informations de différentes natures, comme cela est résumé dans le tableau suivant :

| | Indices mesurant un écart à l'association | Indices mesurant un écart à l'indépendance |
|------------------------|--|---|
| Méthodes par quadrats | | quadrats Kershaw |
| Méthodes par distances | Pielou plus proches voisins Sup de l'Inf | Diggle & Cox Goodall Ripley (méthode 1) |

La démarche pour caractériser la répartition spatiale de deux espèces consiste d'abord à tester s'il y a indépendance des répartitions marginales, à l'aide d'un des indices mesurant un écart à l'indépendance. S'il y a indépendance, chaque répartition marginale peut être étudiée séparément. Il paraît intéressant dans tous les cas de combiner l'information apportée par plusieurs indices. Dans la mesure où les indices mesurant un écart à l'indépendance et les indices mesurant un écart à l'association apportent des informations de natures différentes, un indice au moins de chaque type devrait être calculé.

Les indices ne sont par ailleurs pas tous aussi performants. La méthode des quadrats donne des résultats plus satisfaisants que la méthode de Kershaw, et est de plus plus facile à mettre en oeuvre. Parmi les indices par distances mesurant un écart à l'indépendance, la méthode de Ripley semble la plus performante, mais est lourde à mettre en oeuvre. La méthode de Diggle & Cox offre un bon compromis entre performances et facilité de mise en oeuvre. C'est une méthode supérieure à la méthode de Goodall, et finalement assez comparable par les résultats qu'elle donne à la méthode des quadrats. Parmi les indices mesurant un écart à l'association, l'indice de Pielou semble le plus intéressant, suivi de l'indice des plus proches voisins.

Ces indices donnent des résultats satisfaisants pour des systèmes simples comme celui de la forêt candienne. Pour des systèmes plus complexes comme la forêt guyanaise, la confrontation des différents indices est nécessaire et ne permet de dégager que des tendances.

A. INDICES DE DISSIMILARITÉ

Soit E un ensemble d'objets. Un *indice de dissimilarité* d est une application de $E \times E$ dans \mathbb{R}^+ vérifiant les propriétés suivantes :

- (1) non-négativité : $\forall (a, b) \in E^2, d(a, b) \geq 0$
- (2) symétrie : $\forall (a, b) \in E^2, d(a, b) = d(b, a)$
- (3) normalisation : $\forall (a, b) \in E^2, a = b \implies d(a, b) = 0$

On peut demander à l'indice de satisfaire une propriété de normalisation (3') plus forte que (3)

$$(3') \quad \forall (a, b) \in E^2, a = b \Leftrightarrow d(a, b) = 0$$

On demande aussi parfois à l'indice de satisfaire des propriétés supplémentaires :

- (4) inégalité triangulaire : $\forall (a, b, c) \in E^3, d(a, b) \leq d(a, c) + d(c, b)$
- (5) inégalité ultramétrique : $\forall (a, b, c) \in E^3, d(a, b) \leq \max [d(a, c), d(c, b)]$

La propriété (5) implique (4).

Selon que l'indice vérifie ou non ces différentes propriétés, il porte différents noms :

| | | Propriétés | | | | | |
|--------------------------------|------------------------|------------|-----|-----|------|-----|-----|
| | | (1) | (2) | (3) | (3') | (4) | (5) |
| Indices de dissimilarité | écart semi-métrique | X | X | X | | | |
| | écart métrique | X | X | X | | X | |
| | écart ultramétrique | X | X | X | | | X |
| | indice de distance | X | X | | X | | |
| | distance métrique | X | X | | X | X | |
| | distance ultramétrique | X | X | | X | | X |

Une transformation permet de passer d'un indice de distance à une distance métrique [8, p.53]. Le passage d'un indice de dissimilarité à une distance ultramétrique est un problème que cherche à résoudre la classification hiérarchique.

B. QUELQUES RAPPELS SUR LA THÉORIE DES PROCESSUS PONCTUELS

B.1. Définitions

De façon non formelle, un *processus ponctuel* est un modèle aléatoire qui contrôle les localisations $\{\vec{x}_i\}$ d'événements dans un ensemble $X \subset \mathbb{R}^d$. On se limitera au cas $d = 2$. Une répartition est une réalisation d'un processus ponctuel. La façon la plus naturelle de définir une répartition est grâce aux coordonnées $\vec{x}_1, \vec{x}_2, \dots$ des événements dans X . Une façon équivalente de définir une répartition est de se donner pour tout sous-ensemble $B \subset X$ le nombre $\phi(B)$ d'événements qui y sont situés. De façon formelle : soit \mathcal{X} la tribu borélienne de X . Une mesure de comptage ϕ localement finie sur X est une application de \mathcal{X} dans \mathbb{N} telle que $\phi(B) < \infty$ pour tout $B \in \mathcal{X}$ borné. Soit Φ une collection de mesures de comptage localement finies sur X , et \mathcal{N} la plus petite tribu engendrée par les ensembles de la forme $\{\phi \in \Phi / \phi(B) = n\}, \forall B \in \mathcal{X}, \forall n \in \mathbb{N}$. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

Définition. Un processus ponctuel N sur X est une application mesurable de $(\Omega, \mathcal{A}, \mathbb{P})$ dans (Φ, \mathcal{N}) .

Le processus ponctuel N définit donc une probabilité Π_N sur (Φ, \mathcal{N}) :

$$\forall Y \in \mathcal{N}, \Pi_N(Y) = \Pr(\{\omega \in \Omega / N(\omega) \in Y\})$$

Un *processus ponctuel marqué* est, de façon non formelle, un processus ponctuel tel qu'à chaque événement localisé en \vec{x} est associé une marque aléatoire $M(\vec{x})$. La marque peut être par exemple la hauteur de l'arbre ou son diamètre. Lorsque la marque est une variable qualitative, avec un nombre fini de modalités, on parle plutôt de *processus ponctuel multivarié*. Par la suite, on ne considérera que des processus ponctuels bivariés, les deux modalités de la marque correspondant aux deux espèces d'arbres. De façon formelle : \mathcal{X} étant la tribu borélienne de X et $\phi = (\phi_1, \phi_2)$ étant une mesure de comptage bivariée, c'est-à-dire une application de \mathcal{X} dans $\mathbb{N} \times \mathbb{N}$, soit Φ une collection de mesures de comptage bivariées localement finies, et \mathcal{N} la plus petite tribu engendrée par les ensembles de la forme $\{\phi \in \Phi / (\phi_1(B_1), \phi_2(B_2)) = (n_1, n_2)\}, \forall (B_1, B_2) \in \mathcal{X}^2, \forall (n_1, n_2) \in \mathbb{N}^2$. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

Définition. Un processus ponctuel bivarié $\mathbf{N} = (N_1, N_2)$ sur X est une application mesurable de $(\Omega, \mathcal{A}, \mathbb{P})$ dans (Φ, \mathcal{N}) .

\mathbf{N} définit donc une probabilité $\Pi_{\mathbf{N}}$ sur (Φ, \mathcal{N}) .

B.2. Propriétés

B.2.1. Processus simple

Un processus ponctuel N est *simple* si $\Pi_N(\{\phi \in \Phi / \exists \vec{x} \in X, \phi(\{\vec{x}\}) > 1\}) = 0$. Par la suite on ne considérera que des processus ponctuel simple, conformément à l'idée intuitive qu'en un même point il ne peut pas pousser plus d'un arbre.

B.2.2. Stationnarité

Pour tout $\vec{s} \in \mathbb{R}^2$, soit $\tau_{\vec{s}} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'opérateur de translation $\tau_{\vec{s}}(\vec{x}) = \vec{x} - \vec{s}$. Soit $N : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\Phi, \mathcal{N})$ un processus ponctuel, et soit $N_{\vec{s}}$ le processus ponctuel défini de $(\Omega, \mathcal{A}, \mathbb{P})$ dans (Φ, \mathcal{N}) par

$N_{\vec{s}}(\omega) = N(\omega) \circ \tau_{\vec{s}}^{-1}, \forall \omega \in \Omega$. Le processus ponctuel N est *stationnaire* (on dit aussi *homogène*) si $\Pi_N(Y) = \Pi_{N_{\vec{s}}}(Y), \forall Y \in \mathcal{N}, \forall \vec{s} \in \mathbb{R}^2$. Intuitivement, une répartition issue d'un processus ponctuel stationnaire présentera la même configuration de points quelque soit la région du plan d'où on l'observe.

B.2.3. Isotropie

De même, $\rho_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ étant la rotation d'angle θ autour de l'origine, N est *isotrope* si $\Pi_N(Y) = \Pi_{N_\theta}(Y), \forall Y \in \mathcal{N}, \forall \theta \in [0, 2\pi[$, avec $N_\theta(\omega) = N(\omega) \circ \rho_\theta^{-1}, \forall \omega \in \Omega$.

B.2.4. Indépendance

Deux processus ponctuels N_1 et N_2 sont *indépendants* si $\Pi_{\mathbf{N}}(Y_1, Y_2) = \Pi_{N_1}(Y_1) \times \Pi_{N_2}(Y_2)$, avec $\mathbf{N} = (N_1, N_2)$.

B.3. Caractéristiques du premier ordre

B.3.1. Mesure de premier moment

La mesure de premier moment μ_N pour un processus ponctuel N est l'espérance de la variable aléatoire $N(B)$ représentant le nombre d'événements survenus dans la région $B \subset \mathbb{R}^2$:

$$\forall B \in \mathcal{X}, \mu_N(B) = E(N(B)) = \int_{\Phi} \phi(B) \Pi_N(d\phi)$$

Pour un processus bivarié \mathbf{N} on définit de même :

$$\forall B \in \mathcal{X}, \mu_{N_i}(B) = E(N_i(B)) = \int_{\Phi} \phi_i(B) \Pi_{\mathbf{N}}(d\phi) \quad , i = 1, 2$$

B.3.2. Intensité

L'intensité λ du processus ponctuel N est définie par :

$$\lambda(\vec{x}) = \lim_{\nu(d\vec{x}) \rightarrow 0} \frac{\mu_N(d\vec{x})}{\nu(d\vec{x})}$$

(si cette limite existe), où ν est la mesure de Lebesgue. Si le processus est simple, on a de plus :

$$\lambda(\vec{x}) = \lim_{\mathcal{B} \rightarrow \{\vec{x}\}} \frac{\Pr[N(\mathcal{B}) = 1]}{\nu(\mathcal{B})} = \lim_{\mathcal{B} \rightarrow \{\vec{x}\}} \frac{\Pr[N(\mathcal{B}) > 0]}{\nu(\mathcal{B})}$$

où \mathcal{B} est une boule centrée en \vec{x} . De même pour un processus ponctuel bivarié :

$$\lambda_i(\vec{x}) = \lim_{\nu(d\vec{x}) \rightarrow 0} \frac{\mu_{N_i}(d\vec{x})}{\nu(d\vec{x})}$$

Pour un processus stationnaire, l'intensité est constante : $\lambda(\vec{x}) = \lambda$ et $\mu_N(B) = \lambda\nu(B)$.

B.4. Caractéristiques du second ordre

B.4.1. Mesure de second moment

La mesure de second moment $\mu_N^{(2)}$ pour un processus ponctuel N est :

$$\forall (B_1, B_2) \in \mathcal{X}^2, \mu_N^{(2)}(B_1 \times B_2) = E(N(B_1)N(B_2)) = \int_{\Phi} \phi(B_1)\phi(B_2) \Pi_N(d\phi)$$

De même pour un processus ponctuel bivarié N :

$$\forall (B_1, B_2) \in \mathcal{X}^2, \mu_{N_i N_j}^{(2)}(B_1 \times B_2) = E(N_i(B_1)N_j(B_2)) = \int_{\mathfrak{F}} \phi_i(B_1)\phi_j(B_2)\Pi_N(d\phi) \quad , i, j = 1, 2$$

On définit également la *mesure de second moment factorisé* $\alpha_N^{(2)}$:

$$\forall (B_1, B_2) \in \mathcal{X}^2, \alpha_N^{(2)}(B_1 \times B_2) = \mu_N^{(2)}(B_1 \times B_2) - \mu_N^{(2)}(B_1 \cap B_2)$$

En particulier : $\forall B \in \mathcal{X}, \alpha_N^{(2)}(B \times B) = E[N(B)(N(B) - 1)]$.

B.4.2. Densité produit de second ordre

La densité produit de second ordre $\rho^{(2)}$ d'un processus ponctuel N est :

$$\rho^{(2)}(\vec{x}, \vec{y}) = \lim_{\nu(d\vec{x}) \rightarrow 0, \nu(d\vec{y}) \rightarrow 0} \frac{\mu_N^{(2)}(d\vec{x} \times d\vec{y})}{\nu(d\vec{x})\nu(d\vec{y})}$$

dans la mesure où cette limite existe. Intuitivement, $\rho^{(2)}(\vec{x}, \vec{y})d\vec{x}d\vec{y}$ est la probabilité que la région infinitésimale $d\vec{x}$ centrée en \vec{x} et la région infinitésimale $d\vec{y}$ centrée en \vec{y} contiennent chacune un événement. Pour un processus stationnaire isotrope, $\rho^{(2)}$ n'est fonction que de la distance $r = \|\vec{x} - \vec{y}\|$. On définit alors la *fonction de corrélation de paires* :

$$g(r) = \frac{\rho^{(2)}(r)}{\lambda^2}$$

où λ est l'intensité du processus. De même pour un processus ponctuel bivarié, on définit la *densité produit croisée de second ordre* $\rho_{ij}^{(2)}$:

$$\rho_{ij}^{(2)}(\vec{x}, \vec{y}) = \lim_{\nu(d\vec{x}) \rightarrow 0, \nu(d\vec{y}) \rightarrow 0} \frac{\mu_{N_i N_j}^{(2)}(d\vec{x} \times d\vec{y})}{\nu(d\vec{x})\nu(d\vec{y})} \quad , i, j = 1, 2$$

Intuitivement, $\rho_{ij}^{(2)}(\vec{x}, \vec{y})d\vec{x}d\vec{y}$ est la probabilité que la région infinitésimale $d\vec{x}$ centrée en \vec{x} contienne un événement de type i et que la région infinitésimale $d\vec{y}$ centrée en \vec{y} contienne un événement de type j . Lorsque le processus est stationnaire isotrope, on définit la *fonction de corrélation de paires croisée* :

$$g_{12}(r) = \frac{\rho_{12}^{(2)}(r)}{\lambda_1 \lambda_2} \text{ avec } r = \|\vec{x} - \vec{y}\|$$

B.4.3. Mesure réduite de second moment

On montre [9, p.632] que pour un processus ponctuel stationnaire isotrope N d'intensité λ , il existe une mesure \mathcal{K}_N sur (X, \mathcal{X}) , appelée mesure réduite de second moment, telle que :

$$\alpha_N(A \times B) = \lambda^2 \int_X \nu(A \cap B_{-\vec{h}}) \mathcal{K}_N(d\vec{h}) \quad , \forall (A, B) \in \mathcal{X}^2$$

où $B_{-\vec{h}} = \{\vec{x} - \vec{h}/\vec{x} \in B\}$. La fonction K de Ripley est alors définie par :

$$K(r) = \mathcal{K}(B(\vec{0}, r))$$

où $B(\vec{0}, r)$ est la boule fermée de rayon r centrée à l'origine. Intuitivement, $\lambda K(r)$ est le nombre moyen d'événements situés à une distance inférieure ou égale à r d'un événement arbitraire. On montre aussi [9] que pour un processus stationnaire, isotrope et simple :

$$g(r) = \frac{1}{2\pi r} \frac{d}{dr} K(r)$$

De même pour un processus ponctuel bivarié stationnaire isotrope d'intensité (λ_1, λ_2) , on peut définir des fonctions $K_{ij}(r)$, $i, j = 1, 2$, telles que $\lambda_j K_{ij}(r)$ est le nombre moyen d'événements de type j situés à une distance inférieure ou égale à r d'un événement arbitraire de type i , et :

$$g_{ij}(r) = \frac{1}{2\pi r} \frac{d}{dr} K_{ij}(r)$$

Par ailleurs, $K_{12}(r) = K_{21}(r)$.

B.5. Processus indépendants et étiquetage aléatoire

B.5.1. Indépendance

Si N_1 et N_2 sont deux processus ponctuels indépendants stationnaires isotropes, alors

$$K_{12}(r) = \pi r^2 \quad \text{et} \quad g_{12}(r) = 1 \quad ,$$

la réciproque étant fautive [9, p.700]. Soit G_{ij} la fonction de répartition de la distance d'un événement arbitraire de type i à l'événement de type j le plus proche, et F_i la fonction de répartition de la distance d'un point arbitraire à l'événement de type i le plus proche ($i, j = 1, 2$). Soit F la fonction de répartition de la distance d'un point arbitraire à l'événement le plus proche, quel que soit son type ; $1 - F(r)$ s'interprète aussi comme la probabilité qu'un cercle de rayon r centré en un point arbitraire ne contienne aucun événement. Sous les mêmes hypothèses, on a également [14, p.92] :

$$\begin{aligned} G_{ij}(x) &= F_j(x) \quad , \forall x \in \mathbb{R}^+ \quad , \quad i \in \{1, 2\} \quad , \quad j = 3 - i \\ 1 - F(x) &= (1 - F_1(x))(1 - F_2(x)) \quad , \forall x \in \mathbb{R}^+ \end{aligned}$$

B.5.2. Dépendance

Si N_1 et N_2 sont deux processus ponctuels stationnaires isotropes non indépendants, on parle d'*attraction* entre les deux processus marginaux lorsque $K_{12}(r) > \pi r^2$, et de *répulsion* lorsque $K_{12}(r) < \pi r^2$.

B.5.3. Etiquetage aléatoire

Soit N un processus ponctuel stationnaire isotrope, et N_1 le processus obtenu en retenant chaque événement de N indépendamment des autres avec une probabilité $p \in [0, 1]$ ('random thinning'). Si l'on considère à la fois N_1 et le processus N_2 formé des événements de N non retenus, cela correspond à un étiquetage aléatoire ('random labelling'). On a alors [16] :

$$K_{11}(r) = K_{22}(r) = K_{12}(r) = K(r) \quad , \forall r$$

où K est la fonction de Ripley du processus ponctuel N . L'étiquetage aléatoire n'est équivalent à l'indépendance des processus ponctuels marginaux N_1 et N_2 que si N_1 et N_2 sont des processus de Poisson homogènes [14, p.93].

B.6. Processus de Poisson univarié

Les processus ponctuels de Poisson ayant un rôle central dans la théorie des processus ponctuels, un paragraphe leur est consacré. Les autres processus ponctuels univariés ne seront pas présentés ; les principaux sont décrits par exemple dans [9, pp.650-695].

Définition [60]. Un processus de Poisson homogène est un processus ponctuel simple, homogène, et qui possède la propriété suivante : quels que soient k ($k \in \mathbb{N}$) ensembles de \mathbb{R}^2 disjoints deux à deux, les k variables aléatoires donnant le nombre d'événements présents dans les k ensembles sont statistiquement indépendantes.

Il découle de cette définition une propriété très souvent utilisée des processus Poissoniens homogènes : le nombre d'événements présents dans un ensemble B suit une loi de Poisson de paramètre $\lambda\nu(B)$, où λ est l'intensité du processus :

$$\forall B \in \mathcal{X}, \Pr[N(B) = n] = \frac{1}{n!} (\lambda\nu(B))^n e^{-\lambda\nu(B)}$$

De plus, sachant que B contient n événements, ces n événements sont indépendamment et identiquement distribués uniformément sur B . Le processus de Poisson homogène modélise donc la répartition aléatoire.

Les caractéristiques de second ordre d'un processus de Poisson homogène se calculent par ailleurs facilement :

$$\begin{aligned} \rho^{(2)}(\vec{x}, \vec{y}) &= \lambda^2 \\ K(r) &= \pi r^2 \end{aligned}$$

B.7. Processus ponctuels bivariés

B.7.1. Processus de Poisson bivariés

Un processus de Poisson bivarié est un processus ponctuel bivarié dont les deux processus marginaux N_1 et N_2 sont des processus de Poisson homogènes, tout en n'étant pas forcément indépendants. De tels processus modélisent des situations biologiques dans lesquelles les deux espèces interagissent directement dans un habitat homogène.

Processus des paires liées [17]

C'est un processus de Poisson bivarié montrant de l'attraction entre les processus marginaux. Le processus est défini de la façon suivante :

1. Les événements de type 1 sont réalisés d'après un processus de Poisson homogène d'intensité λ .
2. A chaque événement de type 1 est associé un événement de type 2 dont la position relative par rapport à l'événement de type 1 suit indépendamment une loi symétrique radialement autour de l'origine (typiquement une loi normale symétrique radialement, de variance σ). Une variante de ce processus, qui rend moins nette l'attraction entre les deux répartitions marginales consiste à rajouter une troisième étape :
3. On casse les "liens" événement de type 1 / événement de type 2 avec une probabilité $p \in]0, 1[$, et pour chaque lien cassé on remplace l'un des deux événements au hasard.

Processus de Brown *et al.*

C'est un processus de Poisson bivarié montrant au contraire de la répulsion entre les deux répartitions marginales. Le processus est défini de la façon suivante :

1. Les événements d'un processus primaire sont réalisés d'après un processus de Poisson homogène d'intensité ρ .
2. Les cellules de Dirichlet sont construites autour de chaque événement du processus primaire.
3. Pour chaque cellule de Dirichlet C , les nombres n_1 et n_2 d'événements de types 1 et 2 sont déterminés de la façon suivante : n_1 est généré par une loi de Poisson de paramètre $\lambda_1\nu(C)$, où ν est la mesure de Lebesgue. Le nombre n_2 est alors déterminé de sorte que

$$\Pr[N_2(C) \leq n_2] = 1 - \Pr[N_1(C) \leq n_1]$$

où $N_1(C)$ et $N_2(C)$ sont des variables aléatoires qui suivent des lois de Poisson de paramètres respectifs $\lambda_1\nu(C)$ et $\lambda_2\nu(C)$.

4. Conditionnellement à n_1 et n_2 , les événements des types 1 et 2 sont identiquement et indépendamment distribués d'après une loi uniforme sur \mathcal{C} .
5. Le processus final est constitué des événements des types 1 et 2 seulement.

B.7.2. Processus parents-enfants de Dixon [20]

Ce processus ponctuel bivarié montre de l'attraction entre les événements "parents" et les événements "enfants". C'est un processus homogène anisotrope, susceptible de modéliser la dispersion des graines. Comme le processus des paires liées, il est généré en deux étapes :

1. Les événements "parents" sont réalisés d'après un processus de Poisson homogène.
2. Chaque parent a deux enfants, dont les coordonnées relatives $(\Delta X_i, \Delta Y_i)_{i=1,2}$ sont des variables aléatoires indépendantes identiquement distribuées suivant une loi normale.

B.7.3. Processus d'inhibition mutuelle [17]

C'est un processus montrant de la répulsion entre les répartitions marginales, et qui peut modéliser la compétition entre deux espèces dans un habitat homogène. Contrairement au processus de Brown *et al.* où les répartitions marginales sont aléatoires, les répartitions marginales sont ici agrégées. Les événements de types 1 et 2 sont générés un par un en alternance. Chaque événement est provisoirement placé suivant une distribution uniforme sur la zone d'étude ; s'il se trouve à moins d'une distance δ d'un événement du type opposé déjà en place, une nouvelle localisation provisoire est générée et la procédure est répétée, sinon la localisation est retenue. La procédure s'achève quand un nombre prédéfini d'événements ont été placés, ou quand il n'est plus possible de poursuivre. Ce processus peut être généralisé de plusieurs façons. On peut d'abord distinguer deux distances d'inhibition δ_1 et δ_2 pour chacune des espèces. On peut aussi rendre la répulsion moins marquée en définissant une fonction de survie :

$$p(r) = \begin{cases} \left(\frac{r}{\delta}\right)^\beta & \text{si } r < \delta \\ 1 & \text{si } r \geq \delta \end{cases}$$

Le $(m+1)^{\text{ème}}$ individu de l'espèce 1 reste à la place provisoire qui lui a été tirée avec la probabilité $\prod_{j=1}^m p(r_j)$, où r_j est la distance de cet individu au $j^{\text{ème}}$ des m individus de l'espèce 2 déjà en place.

B.7.4. Processus de Poisson isolé [29]

C'est un processus ponctuel bivarié impliquant la répulsion entre les deux processus marginaux et pouvant donc modéliser la compétition entre deux espèces dans un habitat homogène. Contrairement au processus d'inhibition mutuelle et au processus de Brown *et al.*, l'une des répartitions marginale est régulière et l'autre agrégée. Le processus est défini de la façon suivante :

1. Un processus ponctuel de Poisson homogène d'intensité λ est généré.
2. Chaque événement de ce processus est marqué de la façon suivante : si la distance de l'événement à son plus proche voisin est plus grande que δ , il est du type 1, sinon il est du type 2.

B.7.5. Processus de Cox bivarié [19]

Les processus de Cox bivariés constituent une famille de processus appropriée pour décrire des répartitions qui résultent de la variabilité du milieu, plutôt que des interactions directes entre individus. Un processus de Cox bivarié est dirigé par un champ aléatoire bivarié $\Lambda(\vec{x}) = \{\Lambda_1(\vec{x}), \Lambda_2(\vec{x})\}$; le processus $N = (N_1, N_2)$ est alors tel que, étant donné une réalisation $\{\lambda_1, \lambda_2\}$ du champ Λ , N_1

et N_2 sont des processus de Poisson *indépendants inhomogènes*, d'intensités respectives $\lambda_1(\vec{x})$ et $\lambda_2(\vec{x})$. Si Λ est stationnaire isotrope, alors N est stationnaire isotrope ; on a alors :

$$\begin{aligned}\rho_{ij}(u) &= \gamma_{ij}(u) + \alpha_i \alpha_j \text{ avec } \alpha_i = E[\Lambda_i(\vec{x})] \\ \gamma_{ij}(\|\vec{x} - \vec{y}\|) &= \text{Cov}[\Lambda_i(\vec{x}), \Lambda_j(\vec{y})] \\ K_{ij}(t) &= \pi t^2 + \frac{2\pi}{\alpha_i \alpha_j} \int_0^t \gamma_{ij}(u) u \, du\end{aligned}$$

Processus de Cox lié

Le champ aléatoire Λ d'un processus de Cox lié vérifie $\Lambda_1 = \nu \Lambda_2$, avec $\nu < 0$ fixé. Il y a alors attraction entre les deux processus marginaux, et :

$$K_{11}(t) = K_{22}(t) = K_{12}(t) = \pi t^2 + \frac{2\pi}{\alpha_2^2} \int_0^t \gamma_{22}(u) u \, du$$

Les processus de Cox liés modélisent des situations biologiques dans lesquelles les deux espèces sont convergentes dans un habitat inhomogène.

Processus de Cox équilibré

Le champ aléatoire Λ d'un processus de Cox équilibré vérifie $\Lambda_1 + \Lambda_2 = \nu$, avec $\nu < 0$ fixé. Il y a alors répulsion entre les deux processus marginaux, et :

$$K_{12}(t) = \pi t^2 - \frac{2\pi}{\alpha_1 \alpha_2} \int_0^t \gamma_{22}(u) u \, du$$

Les processus de Cox équilibrés modélisent des situations biologiques dans lesquelles les deux espèces sont divergentes dans un habitat inhomogène.

B.7.6. Processus ponctuel bivarié éclairci [19]

Le processus d'éclaircie permet de définir un processus ponctuel N à partir d'un processus de base N_0 . Il permet de rendre compte de la mort et de la survie des individus. Un processus ponctuel bivarié éclairci, basé sur le processus ponctuel bivarié $N_0 = \{N_{01}, N_{02}\}$ et dirigé par le champ aléatoire bivarié $Z = (Z_1, Z_2)$ est le processus $N = \{N_1, N_2\}$ obtenu, pour une réalisation $z = (z_1, z_2) \in [0, 1]^2$ de Z , en retenant un événement de N_0 en \vec{x} avec la probabilité $z(\vec{x})$, indépendamment de la rétention des autres événements. Le mécanisme d'éclaircie peut être appliqué à n'importe quel type de processus ponctuel. En particulier les processus de Cox peuvent être générés en appliquant des processus d'éclaircie aux processus de Poisson homogènes. De plus la composition d'un processus de Cox et d'un processus d'éclaircie donne à nouveau un processus de Cox. Un exemple intéressant de processus ponctuel bivarié éclairci est celui donné par Diggle & Milne [19]. Le processus ponctuel de base N_0 est un processus de Cox lié, dirigé par le champ aléatoire $\Lambda = \{\Lambda_1, \Lambda_2\}$ tel que :

$$\Lambda_1(\vec{x}) = \Lambda_2(\vec{x}) = \nu \sum_{i=1}^{\infty} h(\vec{x} - \vec{X}_i)$$

où les \vec{X}_i sont les événements d'un processus de Poisson homogène d'intensité μ/ν , et h est la densité de probabilité de la distribution normale symétrique radialement, de variance ξ . Le processus N est obtenu à partir de N_0 par éclaircie suivant le champ $Z = (Z_1, Z_2)$ défini de la façon suivante :

$$\begin{aligned}Z_1(\vec{x}) &= \begin{cases} 1 - \beta & \text{pour } \vec{x} \in \bigcup_{i=1}^{\infty} \mathcal{B}(\vec{Y}_i, \delta) \\ \beta & \text{pour } \vec{x} \notin \bigcup_{i=1}^{\infty} \mathcal{B}(\vec{Y}_i, \delta) \end{cases} \\ Z_2(\vec{x}) &= 1 - Z_1(\vec{x})\end{aligned}$$

où les \vec{Y}_i sont les événements d'un processus de Poisson homogène d'intensité ω et δ un réel positif. Le processus final N est un processus de Cox bivarié. Il permet de traduire (pour des valeurs appropriées des paramètres) une association positive à petite échelle et une association négative à grande échelle.

B.7.7. Processus de Markov bivarié [14, p.102]

Définition. Un processus ponctuel défini sur un borné $A \subset \mathbb{R}^2$ est un processus de Markov de rang ρ si son intensité en \vec{x} conditionnellement à la réalisation du processus sur $A - \{\vec{x}\}$ ne dépend que des événements dans $\mathcal{B}(\vec{x}, \rho) - \{\vec{x}\}$, où $\mathcal{B}(\vec{x}, \rho)$ est la boule fermée de rayon ρ centrée en \vec{x} .

Les processus de Markov sont bien adaptés pour la modélisation des interactions entre individus. Soit $(\mathcal{X}, \mathcal{Y}) = (\{\vec{x}_1 \dots \vec{x}_{n_1}\}, \{\vec{y}_1 \dots \vec{y}_{n_2}\})$ la réalisation d'un processus de Markov bivarié sur une zone bornée A . La vraisemblance de $(\mathcal{X}, \mathcal{Y})$ par rapport à une paire de processus de Poisson homogènes indépendants d'intensités unité peut se mettre sous la forme [9, p.705] :

$$f(\mathcal{X}, \mathcal{Y}) = \frac{n_1!n_2!}{C} \exp \left\{ - \sum_{i=1}^{n_1} \xi_1(\vec{x}_i) \sum_{j=1}^{n_2} \xi_2(\vec{x}_j) - \sum_{1 \leq i < j \leq n_1} \psi_{11}(\|\vec{x}_i - \vec{x}_j\|) \right. \\ \left. - \sum_{1 \leq i < j \leq n_2} \psi_{22}(\|\vec{y}_i - \vec{y}_j\|) - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \psi_{12}(\|\vec{x}_i - \vec{y}_j\|) \right\}$$

où C est une constante de normalisation. Les potentiels ξ_1 et ξ_2 décrivent les effets à petite échelle. Les potentiels ψ_{11} et ψ_{22} décrivent les interactions entre événements de même type, tandis que ψ_{12} décrit les interactions entre événements de type 1 et événements de type 2.

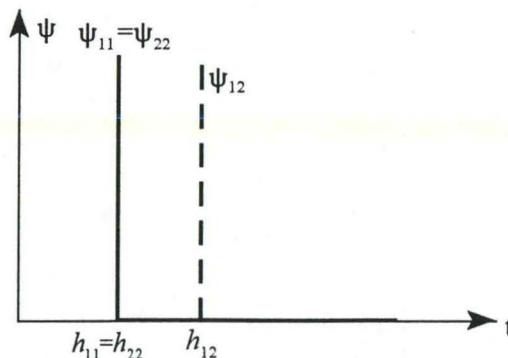
Processus bivariés d'interactions de paires

Les processus bivariés d'interactions de paires ('pairwise interaction process') sont tels que :

$$f(\mathcal{X}, \mathcal{Y}) = \alpha \beta_1^{n_1} \beta_2^{n_2} \exp \left\{ - \sum_{1 \leq i < j \leq n_1} \psi_{11}(\|\vec{x}_i - \vec{x}_j\|) - \sum_{1 \leq i < j \leq n_2} \psi_{22}(\|\vec{y}_i - \vec{y}_j\|) \right. \\ \left. - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \psi_{12}(\|\vec{x}_i - \vec{y}_j\|) \right\}$$

La constante de normalisation α n'est pas assurée d'être finie :

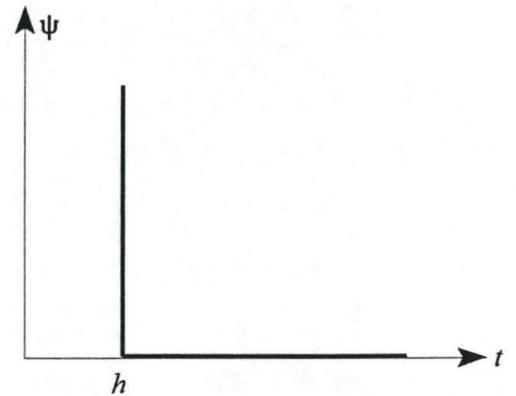
- Une condition suffisante, mais pas nécessaire, pour que α soit finie est que les énergies potentielles $\exp\{-\psi_{ij}(t)\}$ soient bornées et identiquement nulles à partir d'une distance donnée h_{ij} . Les distances h_{ij} sont appelées distances de noyau dur ('hard core distances') et représentent les distances minimales entre deux événements. Un exemple de tels potentiels est :



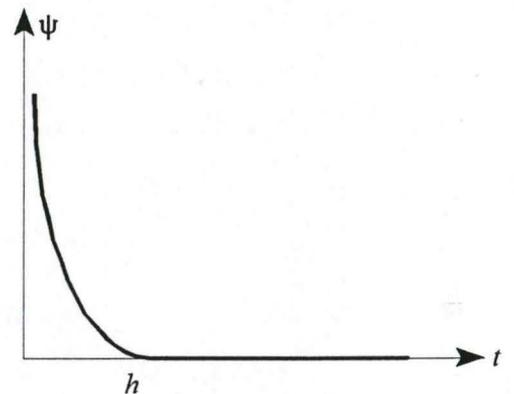
Ce processus à noyaux durs crée une répulsion entre les deux processus analogue à celle d'un processus de Cox équilibré, alors que les hypothèses biologiques sous-jacentes (milieu hétérogène / compétition entre individus) n'ont rien à voir.

- Une autre condition suffisante [18, p.100] est que les fonctions d'interaction ψ_{ij} soient positives ou nulles, et décroissent sur \mathbb{R}^+ . Les processus résultants montrent de la répulsion. Des exemples de potentiels sont donnés par Diggle *et al.* :

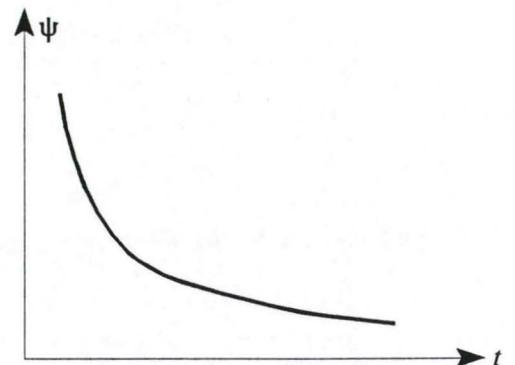
$$\psi(t) = \begin{cases} \theta & \text{si } t \leq h \\ 0 & \text{si } t > h \end{cases} \text{ (processus de Strauss)}$$



$$\psi(t) = \begin{cases} -\ln \left[1 - \frac{(1 - (t/h)^2)}{0} \right] & \text{si } t \leq h \\ 0 & \text{si } t > h \end{cases}$$



$$\psi(t) = -\ln(1 - \exp(-t^2/h^2)) \text{ ('Very Soft Core model')}$$



En fin de compte, les processus d'interactions de paires sont bien adaptés pour décrire la régularité ($\psi > 0$) mais sont mal adaptés pour l'agrégation ($\psi < 0$).

Processus de Strauss bivariés [9, p.705]

Ce sont des processus de Markov bivariés tels que :

$$f(\mathcal{X}, \mathcal{Y}) = \frac{e^{-2v(A)}}{\alpha} \beta_1^{n_1} \beta_2^{n_2} \gamma_{11}^{Y(11)} \gamma_{22}^{Y(22)} \gamma_{12}^{Y(12)}$$

où $Y(ij)$ est le nombre de paires distinctes d'événements situés dans la région A , telles que l'un des événements de la paire soit du type i et l'autre du type j , et que la distance entre les deux événements soit inférieure à ρ (processus de Markov de rang ρ). Les potentiels correspondants sont :

$$\begin{aligned} \xi_i &= 0 \\ \psi_{ij}(t) &= \begin{cases} \ln \gamma_{ij} & \text{si } t \leq \rho \\ 0 & \text{si } t > \rho \end{cases} \end{aligned}$$

C. CALCULS POUR LES INDICES DE DISSIMILARITÉ

C.1. Méthode par présence / absence dans des quadrats

Soit N le processus bivarié de séparation totale défini de la façon suivante : on génère un processus de Poisson homogène d'intensité λ ; tous les événements dont l'abscisse est inférieure à $\frac{1}{2}$ sont de l'espèce 1, les autres sont de l'espèce 2.

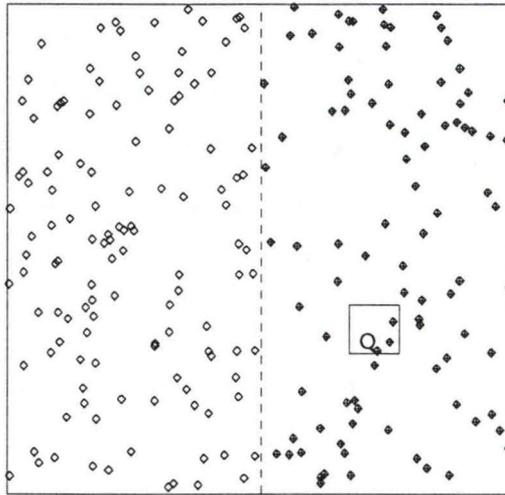


Figure C.1: Réalisation du processus de séparation totale

La zone d'étude est le carré $[0, 1] \times [0, 1]$ de \mathbb{R}^2 . On place au hasard sur la zone d'étude un quadrat Q carré de côté a . Soit

$$A = \Pr [N_1(Q) > 0 \text{ et } N_2(Q) > 0]$$

$$B = \Pr [N_1(Q) > 0 \text{ et } N_2(Q) = 0]$$

$$C = \Pr [N_1(Q) = 0 \text{ et } N_2(Q) > 0]$$

$$D = \Pr [N_1(Q) = 0 \text{ et } N_2(Q) = 0]$$

où $N_i(Q)$ est la variable aléatoire donnant le nombre d'événements du type i présents dans Q .

$$A = \int_0^{1-a} \Pr [N_1(Q) > 0 \text{ et } N_2(Q) > 0 / X_Q = x] \times \Pr [x \leq X_Q < x + dx]$$

où X_Q est l'abscisse du coin inférieur gauche de Q . Le quadrat étant placé au hasard sur la zone

d'étude, X_Q suit une loi uniforme sur $[0, 1 - a]$. Donc

$$\begin{aligned} A &= \frac{1}{1-a} \int_0^{1-a} \Pr [N_1(Q) > 0 \text{ et } N_2(Q) > 0 / X_Q = x] dx \\ (1-a)A &= \int_0^{\frac{1}{2}-a} \Pr [N_1(Q) > 0 \text{ et } N_2(Q) > 0 / X_Q = x] dx \\ &\quad + \int_{\frac{1}{2}-a}^{\frac{1}{2}} \Pr [N_1(Q) > 0 \text{ et } N_2(Q) > 0 / X_Q = x] dx \\ &\quad + \int_{\frac{1}{2}}^{1-a} \Pr [N_1(Q) > 0 \text{ et } N_2(Q) > 0 / X_Q = x] dx \end{aligned}$$

Comme $0 \leq X_Q < \frac{1}{2} - a \implies N_2(Q) = 0$ et $\frac{1}{2} \leq X_Q < 1 - a \implies N_1(Q) = 0$

$$\begin{aligned} (1-a)A &= \int_{\frac{1}{2}-a}^{\frac{1}{2}} \Pr [N_1(Q) > 0 \text{ et } N_2(Q) > 0 / X_Q = x] dx \\ (1-a)A &= \int_{\frac{1}{2}-a}^{\frac{1}{2}} \left(1 - e^{-\lambda a(\frac{1}{2}-x)}\right) \left(1 - e^{-\lambda a(x+a-\frac{1}{2})}\right) dx \end{aligned}$$

L'intégration conduit à

$$A = \frac{1}{1-a} \left[a \left(1 + e^{-\lambda a^2}\right) + \frac{2}{\lambda a} \left(e^{-\lambda a^2} - 1\right) \right]$$

De même

$$\begin{aligned} (1-a)B &= \int_0^{\frac{1}{2}-a} \Pr [N_1(Q) > 0 \text{ et } N_2(Q) = 0 / X_Q = x] dx \\ &\quad + \int_{\frac{1}{2}-a}^{\frac{1}{2}} \Pr [N_1(Q) > 0 \text{ et } N_2(Q) = 0 / X_Q = x] dx \\ &\quad + \int_{\frac{1}{2}}^{1-a} \Pr [N_1(Q) > 0 \text{ et } N_2(Q) = 0 / X_Q = x] dx \\ (1-a)B &= \int_0^{\frac{1}{2}-a} \left(1 - e^{-\lambda a^2}\right) \times 1 \times dx \\ &\quad + \int_{\frac{1}{2}-a}^{\frac{1}{2}} \left(1 - e^{-\lambda a(\frac{1}{2}-x)}\right) e^{-\lambda a(x+a-\frac{1}{2})} dx \\ &\quad + \int_{\frac{1}{2}}^{1-a} 0 \times e^{-\lambda a^2} dx \\ B &= \frac{1}{1-a} \left[\frac{1}{2} - a - \frac{e^{-\lambda a^2}}{2} - \frac{1}{\lambda a} \left(e^{-\lambda a^2} - 1\right) \right] \end{aligned}$$

Par symétrie $C = B$. Enfin

$$\begin{aligned} (1-a)D &= \int_0^{\frac{1}{2}-a} \Pr [N_1(Q) = 0 \text{ et } N_2(Q) = 0 / X_Q = x] dx \\ &\quad + \int_{\frac{1}{2}-a}^{\frac{1}{2}} \Pr [N_1(Q) = 0 \text{ et } N_2(Q) = 0 / X_Q = x] dx \\ &\quad + \int_{\frac{1}{2}}^{1-a} \Pr [N_1(Q) = 0 \text{ et } N_2(Q) = 0 / X_Q = x] dx \\ (1-a)D &= \int_0^{\frac{1}{2}-a} e^{-\lambda a^2} \times 1 \times dx \\ &\quad + \int_{\frac{1}{2}-a}^{\frac{1}{2}} e^{-\lambda a(\frac{1}{2}-x)} e^{-\lambda a(x+a-\frac{1}{2})} dx \\ &\quad + \int_{\frac{1}{2}}^{1-a} 1 \times e^{-\lambda a^2} dx \\ D &= e^{-\lambda a^2} \end{aligned}$$

C.2. Méthode de Kershaw

On considère le même processus bivarié de séparation totale que précédemment, et on conserve les mêmes notations. Calculons $\text{Cov}[N_1(Q), N_2(Q)]$.

$$E(N_1(Q)N_2(Q)) = E[E(N_1(Q)N_2(Q) / X_Q = x)]$$

2

$$\begin{aligned}
 &= \int_0^{1-a} E(N_1(Q)N_2(Q) / X_Q = x) \frac{dx}{1-a} \\
 &= \int_{\frac{1}{2}-a}^{\frac{1}{2}} E(N_1(Q)N_2(Q) / X_Q = x) \frac{dx}{1-a}
 \end{aligned}$$

car $0 \leq X_Q < \frac{1}{2} - a \implies N_2(Q) = 0$ et $\frac{1}{2} \leq X_Q < 1 - a \implies N_1(Q) = 0$. Lorsque $X_Q = x \in [\frac{1}{2} - a, \frac{1}{2}]$, $N_1(Q)$ suit une loi de Poisson de paramètre $\lambda a (\frac{1}{2} - x)$, et $N_2(Q)$ suit indépendamment de $N_1(Q)$ une loi de Poisson de paramètre $\lambda a (x - \frac{1}{2} + a)$. Donc

$$\begin{aligned}
 E(N_1(Q)N_2(Q)) &= \int_{\frac{1}{2}-a}^{\frac{1}{2}} \lambda a \left(\frac{1}{2} - x\right) \times \lambda a \left(x - \frac{1}{2} + a\right) \frac{dx}{1-a} \\
 &= \frac{\lambda^2 a^5}{6(1-a)}
 \end{aligned}$$

Par ailleurs

$$\begin{aligned}
 E(N_1(Q)) &= E[E(N_1(Q) / X_Q = x)] \\
 &= \int_0^{1-a} E(N_1(Q) / X_Q = x) \frac{dx}{1-a}
 \end{aligned}$$

Lorsque $X_Q \in [0, \frac{1}{2} - a]$, $N_1(Q)$ suit une loi de Poisson de paramètre λa^2 ; lorsque $X_Q \in [\frac{1}{2} - a, \frac{1}{2}]$, $N_1(Q)$ suit une loi de Poisson de paramètre $\lambda a (\frac{1}{2} - X_Q)$; lorsque $X_Q \in [\frac{1}{2}, 1 - a]$, $N_1(Q) = 0$, donc

$$\begin{aligned}
 E(N_1(Q)) &= \int_0^{\frac{1}{2}-a} \lambda a^2 \frac{dx}{1-a} + \int_{\frac{1}{2}-a}^{\frac{1}{2}} \lambda a \left(\frac{1}{2} - x\right) \frac{dx}{1-a} \\
 &= \frac{\lambda a^2}{2}
 \end{aligned}$$

Par symétrie, $E(N_2(Q)) = E(N_1(Q))$. Finalement

$$\begin{aligned}
 \text{Cov}[N_1(Q), N_2(Q)] &= E(N_1(Q)N_2(Q)) - E(N_2(Q))E(N_1(Q)) \\
 &= \frac{\lambda^2 a^4}{1-a} \left(\frac{5a-3}{12}\right)
 \end{aligned}$$

C.3. Méthode de Pielou

Soit \mathbf{N} un processus bivarié homogène, E un événement pris au hasard, et $\text{ppv}(E)$ le plus proche voisin de E . Soit

$$\begin{aligned}
 A &= \Pr[E \text{ est de type 1 et } \text{ppv}(E) \text{ est de type 1}] \\
 B &= \Pr[E \text{ est de type 1 et } \text{ppv}(E) \text{ est de type 2}] \\
 C &= \Pr[E \text{ est de type 2 et } \text{ppv}(E) \text{ est de type 1}] \\
 D &= \Pr[E \text{ est de type 2 et } \text{ppv}(E) \text{ est de type 2}]
 \end{aligned}$$

Soit λ_1 et λ_2 les intensités des processus marginaux de \mathbf{N} .

$$\begin{aligned}
 A &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \Pr[\text{ppv}(E) \text{ est de type 1} / E \text{ est de type 1}] \\
 B &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \Pr[\text{ppv}(E) \text{ est de type 2} / E \text{ est de type 1}] \\
 C &= \frac{\lambda_2}{\lambda_1 + \lambda_2} \Pr[\text{ppv}(E) \text{ est de type 1} / E \text{ est de type 2}] \\
 D &= \frac{\lambda_2}{\lambda_1 + \lambda_2} \Pr[\text{ppv}(E) \text{ est de type 2} / E \text{ est de type 2}]
 \end{aligned}$$

Soit X_i la distance d'un point arbitraire à l'arbre de l'espèce i le plus proche, et Y_{ij} la distance d'un arbre arbitraire de l'espèce i à son plus proche voisin de l'espèce j .

$$\Pr [\text{ppv}(E) \text{ est de type } i / E \text{ est de type } j] = \Pr [Y_{ji} < Y_{j(3-i)}], (i, j) \in \{1, 2\}^2$$

C.3.1. Superposition de deux processus de Poisson homogènes indépendants

Si \mathbf{N} est composé de deux processus de Poisson homogènes indépendants, alors

$$\begin{aligned} \Pr [Y_{ji} < Y_{j(3-i)}] &= \Pr [X_i < X_{(3-i)}] \\ &= \int_0^\infty \Pr [x < X_{(3-i)} \text{ et } x \leq X_i < x + dx] \\ &= \int_0^\infty \Pr [x < X_{(3-i)}] \Pr [x \leq X_i < x + dx] \end{aligned}$$

car X_i et $X_{(3-i)}$ sont alors des variables aléatoires indépendantes. Pour un processus de Poisson homogène d'intensité λ

$$\begin{aligned} \Pr [x < X] &= e^{-\lambda\pi x^2} \\ \Pr [x \leq X < x + dx] &= 2\lambda\pi e^{-\lambda\pi x^2} dx \end{aligned}$$

donc

$$\Pr [\text{ppv}(E) \text{ est de type } i / E \text{ est de type } j] = \int_0^\infty 2\lambda_i\pi x e^{-\lambda_i\pi x^2} e^{-\lambda_{(3-i)}\pi x^2} dx$$

L'intégration conduit à

$$\Pr [\text{ppv}(E) \text{ est de type } i / E \text{ est de type } j] = \frac{\lambda_i}{\lambda_1 + \lambda_2}$$

Finalement

$$\begin{aligned} A &= \frac{\lambda_1^2}{(\lambda_1 + \lambda_2)^2} \\ B &= \frac{\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)^2} = C \\ D &= \frac{\lambda_2^2}{(\lambda_1 + \lambda_2)^2} \end{aligned}$$

C.3.2. Superposition de deux réseaux réguliers indépendants

Si \mathbf{N} est composé de deux réseaux carrés de côté a , réguliers, indépendants, alors

$$\Pr [Y_{ii} < Y_{i(3-i)}] = \Pr [Y_{ii} < X_{(3-i)}]$$

et Y_{ii} a une distribution un peu particulière

$$\begin{cases} Y_{ii} = a & \text{avec la probabilité } 1 \\ Y_{ii} \neq a & \text{avec la probabilité } 0 \end{cases}$$

Pour un réseau carré régulier de côté a , $X \leq a\frac{\sqrt{2}}{2}$, donc

$$\Pr [Y_{11} < X_2] = \Pr [Y_{22} < X_1] = 0$$

Finalement

$$\begin{aligned} A &= D = 0 \\ B &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \quad C = \frac{\lambda_2}{\lambda_1 + \lambda_2} \end{aligned}$$

Remarque : $\frac{AD-BC}{\sqrt{(A+B)(A+C)(B+D)(C+D)}} = -1$ alors qu'il y a indépendance des processus marginaux : la méthode de Pielou ne mesure pas un écart à l'indépendance.

D. TABLEAUX DE CONTINGENCE $2 \times 2 \times 2$

Soit trois caractères à deux modalités chacun $\mathcal{A} = (A_1, A_2)$, $\mathcal{B} = (B_1, B_2)$, $\mathcal{C} = (C_1, C_2)$. Dans une population de taille N , les individus se ventilent entre les 8 caractéristiques possibles obtenues par croisement des trois caractères. Cela peut être résumé dans un tableau de contingence à trois entrées :

| | | C_1 | | | | C_2 | | | | C_1 ou C_2 | | | | |
|-------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------------|-----------|-----------|-----------|-----|
| | | B_1 | B_2 | | | B_1 | B_2 | | | B_1 | B_2 | | | |
| A_1 | n_{111} | n_{121} | $n_{1.1}$ | A_1 | n_{112} | n_{122} | $n_{1.2}$ | A_1 | $n_{11.}$ | $n_{12.}$ | $n_{1..}$ | | | |
| A_2 | n_{211} | n_{221} | $n_{2.1}$ | A_2 | n_{212} | n_{222} | $n_{2.2}$ | A_2 | $n_{21.}$ | $n_{22.}$ | $n_{2..}$ | | | |
| | | $n_{.11}$ | $n_{.21}$ | $n_{..1}$ | | | $n_{.12}$ | $n_{.22}$ | $n_{..2}$ | | | $n_{.1.}$ | $n_{.2.}$ | N |

n_{ijk} est le nombre d'individus ayant les caractères A_i , B_j et C_k . Le problème est d'analyser les interactions entre les trois caractères \mathcal{A} , \mathcal{B} et \mathcal{C} .

On note p_{ijk} la fréquence théorique de la caractéristique (A_i, B_j, C_k) , et un point à la place d'un indice signifie la sommation sur cet indice ($p_{...} = 1$).

D.1. Mesure de l'interaction par le rapport θ

Tableau de contingence 2×2 Pour un tableau de contingence croisant deux caractères à deux modalités chacun, l'interaction entre les deux caractères peut être mesurée par le rapport θ défini par :

| | | B_1 | B_2 | | | |
|-------|----------|----------|----------|----------|--|--|
| A_1 | n_{11} | n_{12} | $n_{1.}$ | | | |
| A_2 | n_{21} | n_{22} | $n_{2.}$ | | | |
| | | $n_{.1}$ | $n_{.2}$ | $n_{..}$ | | |

$$\theta = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

On définit de même un rapport $\pi = p_{11}p_{22}/p_{12}p_{21}$ pour les fréquences théoriques.

L'indépendance entre les deux caractéristiques est définie par

$$p_{ij} = p_{i.}p_{.j}$$

On montre que l'indépendance entre les deux caractères est équivalente à $\pi = 1$. Lorsque $\pi > 1$, les deux caractères sont associés ; lorsque $\pi < 1$, les deux caractères sont dissociés.

Tableau de contingence $2 \times 2 \times 2$ Pour un tableau de contingence à trois entrées, plusieurs rapports θ peuvent être considérés : le lien entre \mathcal{A} et \mathcal{B} conditionnellement à $\mathcal{C} = C_1$ est mesuré par

$$\theta_{C_1} = \frac{n_{111}n_{221}}{n_{121}n_{211}}$$

et le lien entre \mathcal{A} et \mathcal{B} conditionnellement à $\mathcal{C} = C_2$ est mesuré par

$$\theta_{C_2} = \frac{n_{112}n_{222}}{n_{122}n_{212}}$$

On définit de même $\theta_{A_1} = n_{111}n_{122}/n_{112}n_{121}$ et $\theta_{A_2} = n_{211}n_{222}/n_{212}n_{221}$ qui mesurent le lien entre \mathcal{B} et \mathcal{C} conditionnellement à \mathcal{A} , et $\theta_{B_1} = n_{111}n_{212}/n_{112}n_{211}$ et $\theta_{B_2} = n_{121}n_{222}/n_{122}n_{221}$ qui mesurent le lien entre \mathcal{A} et \mathcal{C} conditionnellement à \mathcal{B} .

D.2. Absence d'interaction à trois facteurs

Si le lien (mesuré par le rapport θ) entre deux quelconques des caractères est le même que la valeur du troisième caractère, alors on dit qu'il n'y a pas d'interaction à trois facteurs. L'absence d'interaction à trois facteurs s'écrit donc :

$$\begin{cases} \theta_{A_1} = \theta_{A_2} \\ \theta_{B_1} = \theta_{B_2} \\ \theta_{C_1} = \theta_{C_2} \end{cases}$$

Comme $\frac{n_{111}n_{221}}{n_{121}n_{211}} = \frac{n_{112}n_{222}}{n_{122}n_{212}} \iff \frac{n_{111}n_{122}}{n_{112}n_{121}} = \frac{n_{211}n_{222}}{n_{212}n_{221}} \iff \frac{n_{111}n_{212}}{n_{112}n_{211}} = \frac{n_{121}n_{222}}{n_{122}n_{221}}$, il suffit en fait d'avoir une seule de ces égalités.

Test de l'absence d'interaction à trois facteurs Le test de Bartlett [35, 40] permet de tester s'il y a interaction à trois facteurs. L'hypothèse nulle est

$$\mathcal{H}_0 : \frac{p_{111}p_{221}}{p_{121}p_{211}} = \frac{p_{112}p_{222}}{p_{122}p_{212}}$$

Le test consiste à résoudre l'équation du troisième degré en x

$$\frac{(n_{111} + x)(n_{221} + x)}{(n_{121} - x)(n_{211} - x)} = \frac{(n_{112} - x)(n_{222} - x)}{(n_{122} + x)(n_{212} + x)}$$

Sous \mathcal{H}_0 ,

$$X^2 = \sum_{ijk} \frac{x^2}{n_{ijk} + \delta_{ijk}x}$$

suit approximativement un χ^2 à un degré de liberté, avec $\delta_{ijk} = +1$ si $k = 1$ et $i = j$ ou si $k = 2$ et $i \neq j$, et $\delta_{ijk} = -1$ si $k = 2$ et $i = j$ ou si $k = 1$ et $i \neq j$.

D.3. Indépendance conditionnelle simple

Une interaction entre les trois caractères plus contraignante que l'absence d'interaction à trois facteurs est l'indépendance conditionnelle simple. Il y a indépendance conditionnelle simple lorsque deux des caractères sont indépendants conditionnellement au troisième.

Supposons par exemple que \mathcal{A} et \mathcal{B} sont indépendants conditionnellement à \mathcal{C} . L'indépendance conditionnelle simple s'écrit alors :

$$\begin{cases} \theta_{C_1} = \theta_{C_2} = 1 \\ \theta_{B_1} = \theta_{B_2} \neq 1 \\ \theta_{A_1} = \theta_{A_2} \neq 1 \end{cases}$$

ou de façon équivalente

$$n_{ijk} = \frac{n_{.jk}n_{i.k}}{n_{..k}}$$

Lorsque \mathcal{A} et \mathcal{B} sont indépendants conditionnellement à \mathcal{C} , $\theta_{A_1} = \theta_{A_2} = \theta_{\mathcal{A}}$, où $\theta_{\mathcal{A}} = n_{.11}n_{.22}/n_{.12}n_{.21}$ mesure le lien entre \mathcal{B} et \mathcal{C} sans tenir compte de \mathcal{A} . Le lien entre \mathcal{B} et \mathcal{C} est donc le même que l'on prenne \mathcal{A} en compte ou pas, c'est-à-dire que le tableau de contingence $2 \times 2 \times 2$ peut être écrasé en un tableau de contingence 2×2 en éliminant \mathcal{A} lorsque l'on s'intéresse au lien entre \mathcal{B} et \mathcal{C} .

De même lorsque \mathcal{A} et \mathcal{B} sont indépendants conditionnellement à \mathcal{C} , $\theta_{B_1} = \theta_{B_2} = \theta_{\mathcal{B}}$, et le tableau de contingence $2 \times 2 \times 2$ peut être écrasé en un tableau de contingence 2×2 en éliminant \mathcal{B} .

L'indépendance conditionnelle simple implique l'absence d'interaction à trois facteurs, et la réciproque est fautive.

Test de l'indépendance conditionnelle [34, p.51] L'hypothèse nulle (en supposant par exemple que l'on veut tester l'indépendance de \mathcal{A} et \mathcal{B} conditionnellement à \mathcal{C}) est

$$\mathcal{H}_0 : p_{ijk} = \frac{p_{.jk}p_{i.k}}{p_{..k}}$$

Sous \mathcal{H}_0 ,

$$G^2 = \sum_{ijk} \frac{\left(n_{ijk} - \frac{n_{.jk}n_{i.k}}{n_{..k}} \right)^2}{\frac{n_{.jk}n_{i.k}}{n_{..k}}}$$

suit approximativement un χ^2 à 2 degrés de liberté.

D.4. Indépendance partielle simple

Une interaction entre les trois caractères plus contraignante que l'indépendance conditionnelle simple est l'indépendance partielle simple. Il y a indépendance partielle simple lorsque deux des caractères sont indépendants du troisième.

Supposons par exemple que \mathcal{B} et \mathcal{C} sont indépendants de \mathcal{A} . L'indépendance partielle simple s'écrit alors :

$$\begin{cases} \theta_{C_1} = \theta_{C_2} = 1 \\ \theta_{B_1} = \theta_{B_2} = 1 \\ \theta_{A_1} = \theta_{A_2} \neq 1 \end{cases}$$

ou de façon équivalente

$$n_{ijk} = n_{.jk} \frac{n_{i..}}{N}$$

L'indépendance partielle simple implique l'indépendance conditionnelle simple, et la réciproque est fautive. L'indépendance partielle simple équivaut en fait à une double indépendance conditionnelle.

Test de l'indépendance partielle simple [34, p.51] L'hypothèse nulle (en supposant par exemple que l'on veut tester l'indépendance de \mathcal{B} et \mathcal{A} , et de \mathcal{C} et \mathcal{A}) est

$$\mathcal{H}_0 : p_{ijk} = p_{.jk}p_{i..}$$

Sous \mathcal{H}_0 ,

$$G^2 = \sum_{ijk} \frac{\left(n_{ijk} - n_{.jk} \frac{n_{i..}}{N} \right)^2}{n_{.jk} \frac{n_{i..}}{N}}$$

suit approximativement un χ^2 à 3 degrés de liberté.

D.5. Indépendance totale

L'hypothèse la plus forte que l'on peut faire sur l'interaction entre les trois caractères est l'indépendance totale. Il y a indépendance totale lorsque chacun des trois caractères est indépendant des deux autres, ce qui s'écrit :

$$\begin{cases} \theta_{C_1} = \theta_{C_2} = 1 \\ \theta_{B_1} = \theta_{B_2} = 1 \\ \theta_{A_1} = \theta_{A_2} = 1 \end{cases}$$

ou de façon équivalente

$$n_{ijk} = \frac{n_{i..}n_{.j.}n_{..k}}{N^2}$$

L'indépendance totale implique l'indépendance partielle simple, et la réciproque est fautive. L'indépendance totale équivaut en fait à une double indépendance partielle ou à une triple indépendance partielle.

Test de l'indépendance totale [34, p.48] L'hypothèse nulle est

$$\mathcal{H}_0 : p_{ijk} = p_{i..}p_{.j.}p_{..k}$$

Sous \mathcal{H}_0 ,

$$G^2 = \sum_{ijk} \frac{\left(n_{ijk} - \frac{n_{i..}n_{.j.}n_{..k}}{N^2} \right)^2}{\frac{n_{i..}n_{.j.}n_{..k}}{N^2}}$$

suit approximativement un χ^2 à 4 degrés de liberté.

D.6. Indépendance marginale simple

Il y a indépendance marginale simple lorsque deux des caractères sont indépendants lorsque l'on ne tient pas compte du troisième caractère (c'est-à-dire en raisonnant sur le tableau de contingence 2×2 qui croise ces deux caractères).

Supposons par exemple que \mathcal{A} et \mathcal{B} sont indépendants lorsque l'on ne prend pas en compte \mathcal{C} . L'indépendance marginale simple s'écrit :

$$\theta_C = 1$$

ou de façon équivalente

$$n_{ij.} = \frac{n_{i..}n_{.j.}}{N}$$

L'indépendance totale implique la triple indépendance marginale ($\theta_A = 1$, $\theta_B = 1$, $\theta_C = 1$). Réciproquement, la triple indépendance marginale plus l'absence d'interaction à trois facteurs impliquent l'indépendance totale.

Test de l'indépendance marginale simple C'est le test habituel du χ^2 sur un tableau de contingence 2×2 . L'hypothèse nulle (en supposant par exemple que l'on veut tester l'indépendance de \mathcal{A} et \mathcal{B} sans prendre en compte \mathcal{C}) est

$$\mathcal{H}_0 : p_{ij.} = p_{i..}p_{.j.}$$

Sous \mathcal{H}_0 ,

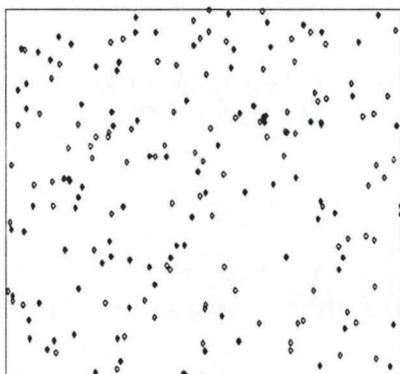
$$X^2 = \sum_{ij} \frac{\left(n_{ij.} - \frac{n_{i..}n_{.j.}}{N} \right)^2}{\frac{n_{i..}n_{.j.}}{N}}$$

suit approximativement un χ^2 à un degré de liberté.

E. PRÉSENTATION DES TREIZE PROCESSUS PONCTUELS BIVARIÉS DE RÉFÉRENCES

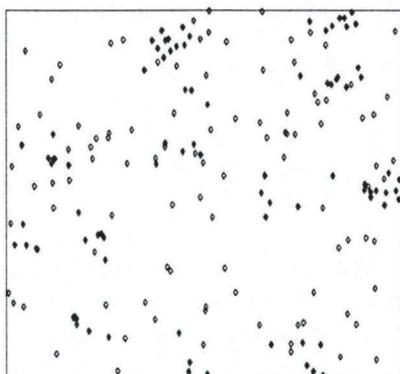
Tous les processus ont été générés sur le carré $[0, 1] \times [0, 1]$ de \mathbb{R}^2 . Le logiciel utilisé pour les programmer est S-PLUS de AT&T.

E.1. Superposition de deux processus de Poisson homogènes indépendants (processus PP)



Le processus de Poisson homogène est défini en annexe B. Les intensités sont $(\lambda_1, \lambda_2) = (100, 100)$.

E.2. Superposition d'un processus de Poisson homogène et d'un processus de Neyman-Scott indépendants (processus PA)

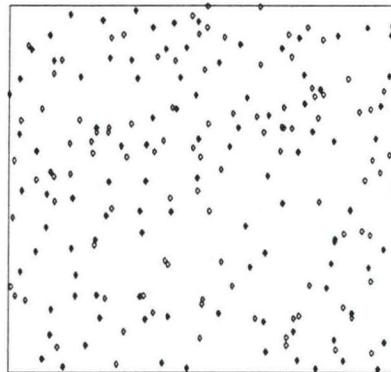


Le processus (univarié) de Neyman-Scott [9, p.661] est défini de la façon suivante :

1. Les points parents sont générés par un processus de Poisson homogène d'intensité λ .
2. Chaque parent a un nombre d'enfants qui est donné par une loi de Poisson de paramètre f .
3. Les coordonnées polaires (R, θ) des points enfants par rapport à leur parent sont des variables aléatoires indépendantes identiquement distribuées de la façon suivante : θ suit une loi uniforme sur $[0, 2\pi[$; R suit une loi normale d'espérance nulle et d'écart type σ .
4. Le processus ne retient que les points enfants.

L'intensité du processus de Poisson est $\lambda = 100$. Les valeurs des paramètres pour le processus de Neyman-Scott sont : $\lambda = 25$, $f = 7$, $\sigma = 0,05$.

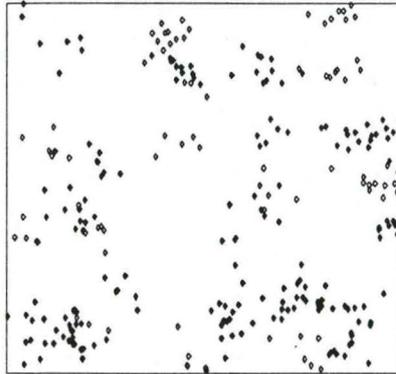
E.3. Superposition d'un processus de Poisson homogène et d'un processus d'inhibition séquentiel (processus PR)



Le processus d'inhibition séquentielle [15] est défini de la façon suivante : lorsque i événements ont été placés, la position du $(i + 1)^{\text{eme}}$ est tirée au hasard. Si le point ainsi tiré est au moins à une distance h des i événements en place, alors il est retenu, sinon on tire à nouveau une position au hasard, etc. L'algorithme s'achève quand n points ont été placés.

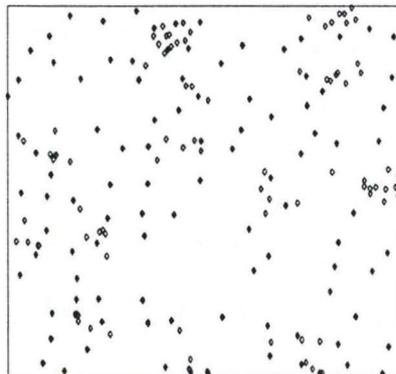
L'intensité du processus de Poisson est $\lambda = 100$. Les paramètres pour le processus d'inhibition séquentielle valent : $n = 100$, $h = 0,05$.

**E.4. Superposition de deux processus de Neyman-Scott indépendants
(processus AA)**



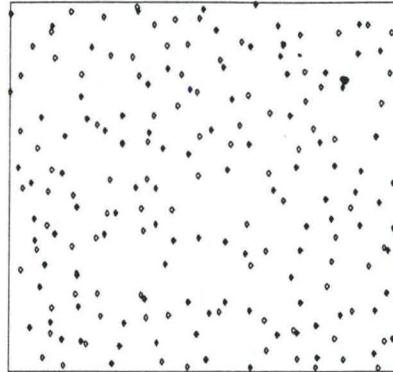
Les processus sont définis comme précédemment. Les valeurs des paramètres sont les mêmes.

**E.5. Superposition d'un processus de Neyman-Scott et d'un processus
d'inhibition séquentiel indépendants (processus AR)**



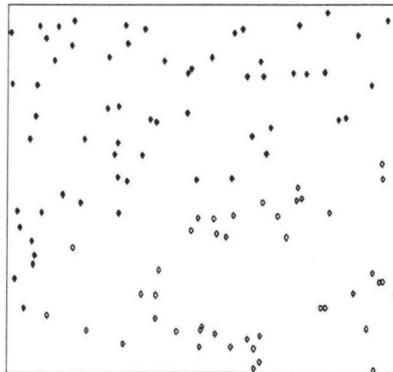
Les processus sont définis comme précédemment. Les valeurs des paramètres sont les mêmes.

E.6. Superposition de deux processus d'inhibition séquentiels (processus RR)



Les processus sont définis comme précédemment. Les valeurs des paramètres sont les mêmes.

E.7. Processus de Cox équilibré dégénéré (processus ST)



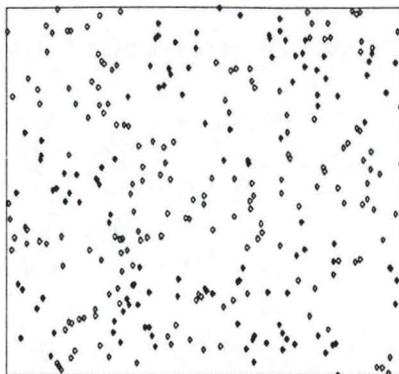
Le processus est défini de la façon suivante : les événements de types 1 et 2 sont générés suivant des processus de Poisson homogènes d'intensités respectives λ_1 et λ_2 . Un événement de type 1 situé en $\vec{x} = (x_1, x_2)$ survit avec la probabilité $p_1(\vec{x})$ définie par :

$$f(x) = \sqrt{x} \quad p_1 = \begin{cases} 0 & \text{si } x_2 \leq f(x_1) - \Delta/2 \\ \frac{x_2 - f(x_1) + \Delta/2}{\Delta} & \text{si } f(x_1) - \Delta/2 < x_2 \leq f(x_1) + \Delta/2 \\ 1 & \text{si } x_2 > f(x_1) + \Delta/2 \end{cases}$$

où Δ est une distance fixée. Un événement de type 2 situé en \vec{x} survit avec la probabilité $p_2(\vec{x}) = 1 - p_1(\vec{x})$.

Les valeurs des paramètres sont les suivants : $(\lambda_1, \lambda_2) = (100, 100)$, $\Delta = 0, 2$.

E.8. Processus de Cox équilibré (processus CE)



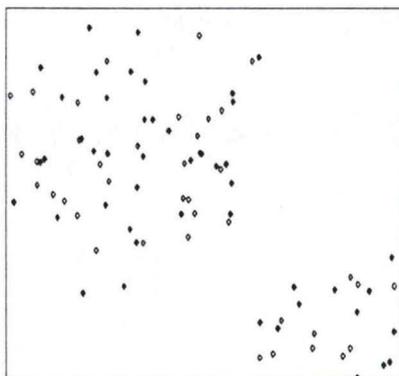
Le processus est défini de la façon suivante [19] : les événements de types 1 et 2 sont générés suivant des processus de Poisson homogènes d'intensités respectives λ_1 et λ_2 . Un événement de type 1 situé en \vec{x} survit avec la probabilité $p_{eq1}(\vec{x})$ définie par :

$$r = \min_i \left(\|\vec{x} - \vec{X}_i\| \right) \quad p_{eq1} = \begin{cases} 1 - \beta & \text{si } r \leq \delta \\ \beta & \text{si } r > \delta \end{cases}$$

où δ est une distance fixée, et les \vec{X}_i sont les événements d'un processus de Poisson homogène d'intensité ω . Un événement de type 2 situé en \vec{x} survit avec la probabilité $p_{eq2}(\vec{x}) = 1 - p_{eq1}(\vec{x})$.

Les valeurs des paramètres sont les suivants : $(\lambda_1, \lambda_2) = (300, 300)$, $\omega = 40$, $\delta = 0,075$, $\beta = 0$.

E.9. Processus de Cox lié dégénéré (processus MV)



Le processus est défini de la façon suivante : les événements de types 1 et 2 sont générés suivant des processus de Poisson homogènes d'intensités respectives λ_1 et λ_2 . Un événement situé en \vec{x} , quelque soit son type, survit avec la probabilité $p(\vec{x})$ définie par :

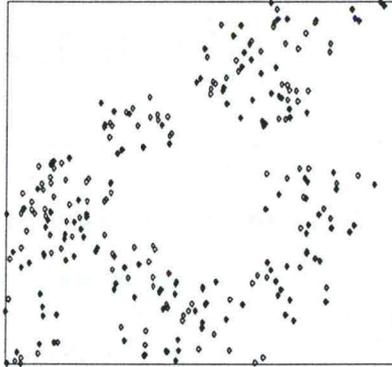
$$\vec{a} = \begin{pmatrix} 0,3 \\ 0,65 \end{pmatrix} \quad \vec{b} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad r = \min \left(\|\vec{x} - \vec{a}\|, \|\vec{x} - \vec{b}\| \right)$$

$$p = \begin{cases} 1 & \text{si } 0 \leq r < R \\ 1 - \frac{r-R}{\Delta R} & \text{si } R \leq r < R + \Delta R \\ 0 & \text{si } r \geq R + \Delta R \end{cases}$$

où R et ΔR sont des rayons fixés.

Les valeurs des paramètres sont les suivants : $(\lambda_1, \lambda_2) = (100, 100)$, $R = 0,25$, $\Delta R = 0,2$.

E.10. Processus de Cox lié (processus CL)



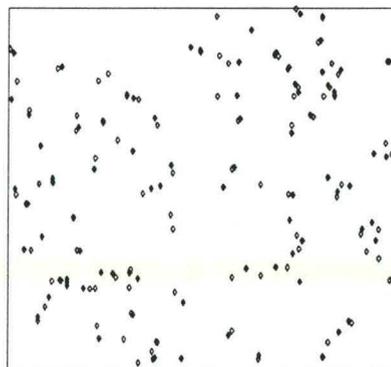
Le processus est défini de la façon suivante [19] : les événements de types 1 et 2 sont générés suivant des processus de Poisson homogènes d'intensités respectives λ_1 et λ_2 . Un événement situé en \vec{x} , quelque soit son type, survit avec la probabilité $p_{lié}(\vec{x})$ définie par :

$$r = \min_i (\|\vec{x} - \vec{X}_i\|) \quad p_{lié} = \begin{cases} 1 & \text{si } 0 \leq r < R \\ 1 - \frac{r-R}{\Delta R} & \text{si } R \leq r < R + \Delta R \\ 0 & \text{si } r \geq R + \Delta R \end{cases}$$

, et les \vec{X}_i sont les événements d'un processus de Poisson homogène d'intensité μ .

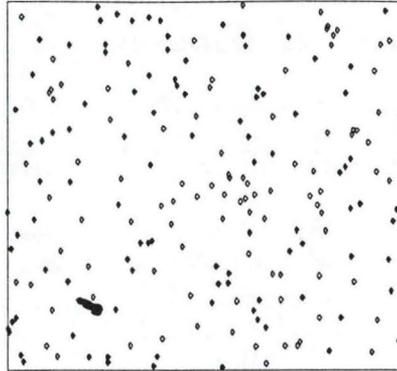
Les valeurs des paramètres sont les suivants : $(\lambda_1, \lambda_2) = (300, 300)$, $\mu = 15$, $R = 0,1$, $\Delta R = 0,02$.

E.11. Processus des paires liées (processus Pa)



Ce processus est défini en annexe B. Les valeurs des paramètres sont les suivants : $\lambda = 100$, $\sigma = 0,0224$, $p = 0,2$.

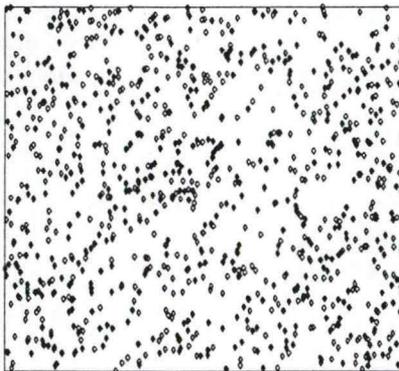
E.12. Processus d'inhibition mutuelle (processus In)



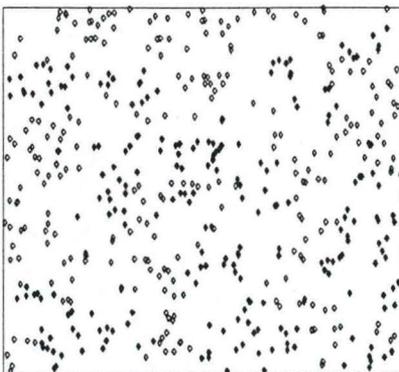
Ce processus est défini en annexe B. Les valeurs des paramètres sont les suivants : $n = 100$, $\delta = 0,05$.

E.13. Processus de Cox composé (processus Ce)

Il résulte de la composition des processus de Cox équilibré et de Cox lié présentés précédemment :

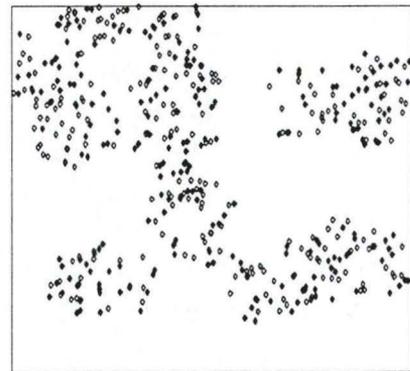


↓ survie p_{eq}

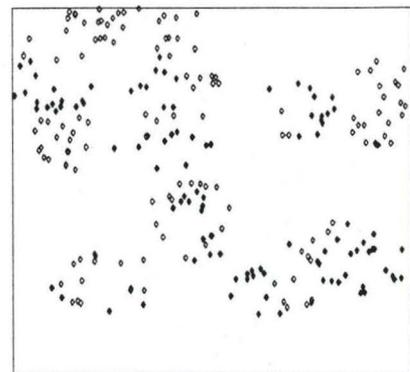


survie $p_{lié}$

↘ $p_{eq} \circ p_{lié}$



↓ survie p_{eq}



survie $p_{lié}$

Les valeurs des paramètres sont les suivants : $\lambda_1 = \lambda_2 = 500$, $\mu = 15$, $R = 0,1$, $\Delta R = 0,02$, $\omega = 40$, $\delta = 0,075$, $\beta = 0$.

E.14. Caractéristiques des processus

| Processus | Stationnarité | Interaction entre processus marginaux | Nature des répartitions marginales | Situation biologique modélisée |
|-----------|---------------|---|---|---|
| PP | homogène | indépendance | aléatoires | Pas d'interaction dans un milieu homogène |
| PA | homogène | indépendance | aléatoire et agrégée | idem |
| PR | homogène | indépendance | aléatoire et régulière | idem |
| AA | homogène | indépendance | agrégées | idem |
| AR | homogène | indépendance | agrégée et régulière | idem |
| RR | homogène | indépendance | régulières | idem |
| ST | inhomogène | répulsion à petite échelle | par plaques, aléatoires à l'intérieur des plaques | Habitat inhomogène, espèces divergentes |
| CE | homogène | répulsion à moyenne échelle | agrégées | idem |
| MV | inhomogène | attraction à petite échelle | par plaque, aléatoires à l'intérieur des plaques | Habitat inhomogène, espèces convergentes |
| CL | homogène | attraction à moyenne échelle | agrégées | idem |
| Pa | homogène | attraction à grande échelle | aléatoires | Commensalisme dans habitat homogène |
| In | homogène | répulsion à grande échelle | agrégées | Compétition dans habitat homogène |
| Ce | homogène | attraction à petite échelle, répulsion à grande échelle | agrégées | Habitat inhomogène |

F. PRÉSENTATION DES INDICES

F.1. Méthodes par quadrats

Un quadrat est une surface échantillon de taille et forme fixées. Les méthodes par quadrats supposent que l'information collectée est le nombre d'individus présents dans les quadrats, sans préoccupation de leur position spatiale. Deux méthodes par quadrats ont été étudiées.

F.1.1. Méthode par présence / absence dans des quadrats [26, 57, 64]

Un nombre N de quadrats sont répartis sur la zone d'étude. Dans chaque quadrat, on regarde si les deux espèces sont présentes, ou si l'une seulement est présente, ou si aucune des deux n'est présente. L'information peut être rassemblée dans un tableau de contingence :

| | | espèce 1 | | |
|----------|----------|-------------|-------------|-------------|
| | | présente | absente | |
| espèce 2 | présente | a | b | $m = a + b$ |
| | absente | c | d | $n = c + d$ |
| | | $r = a + c$ | $s = b + d$ | N |

où a est le nombre de quadrats où les deux espèces sont simultanément présentes, b est le nombre de quadrats où l'espèce 1 est absente et l'espèce 2 présente, etc. L'indice de dissimilarité peut être définie par une distance du χ^2 ,

$$X^2 = \frac{(N-1)(|ad-bc| - N/2)^2}{mnr s}$$

ou par une distance de corrélation de points

$$\rho = \frac{ad-bc}{\sqrt{mnr s}}$$

Si les répartitions des deux espèces sont indépendantes, alors a, b, c, d sont peu différents de $\frac{rm}{N}$, $\frac{ms}{N}$, $\frac{rn}{N}$, $\frac{sn}{N}$, et X^2 et ρ s'annulent ; s'il y a association positive, a et d sont grands devant b et c , donc X^2 et ρ sont significativement plus grands que zéro ; s'il y a association négative, b et c sont grands devant a et d , donc X^2 est significativement plus grand que zéro et ρ est significativement inférieur à zéro. Il y a intérêt à choisir plutôt la distance de corrélation de points ρ comme indice de dissimilarité, puisque ρ contrairement à X^2 change de signe quand l'association change de signe, et à utiliser X^2 pour tester l'indépendance.

Choix du quadrat

Plusieurs questions se posent à ce niveau : combien de quadrats placer ? Comment les répartir ? Quelle taille doivent avoir les quadrats ? Quelle forme doivent-ils avoir ?

On supposera que la forme des quadrats a peu d'incidence sur les résultats, ce qui n'est vrai en toute rigueur que pour des processus homogènes. Pour les simulations, les quadrats sont carrés, tous orientés dans le même sens.

Le bon fonctionnement de la méthode nécessite que les comptages dans les quadrats soient indépendants. Suivant les recommandations d'Upton & Fingleton [64, p.229], les quadrats ont été

positionnés aux noeuds d'un quadrillage carré de côté R , la surface des quadrats n'excédant pas $\frac{R^2}{4}$. Cela suppose donc que l'autocorrélation spatiale est négligeable au delà de la distance R . Une autre façon de procéder serait de placer les quadrats au hasard, mais le taux de couverture par les quadrats devra alors être inférieur à 25%.

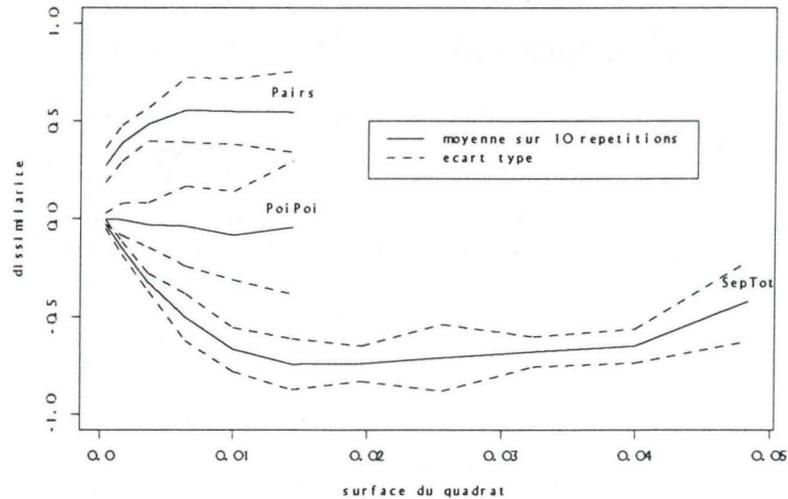


Figure F.1: Influence de la taille du quadrat sur la valeur de l'indice

Le nombre N de quadrats placés sur la zone d'étude est donc directement fonction de la taille des quadrats. Le choix de la taille du quadrat n'est pas anodin. La figure F.1 montre pour trois processus ponctuels d'intensité totale 200 que l'indice de dissimilarité peut considérablement varier selon la taille choisie (quand le quadrat devient trop grand, il vient un moment où ρ n'est plus défini car b , c et d s'annulent). La figure F.2 montre la dissimilarité calculée en fonction de la taille du quadrat, pour différentes intensités du processus, dans le cas du processus ponctuel de séparation totale défini en annexe C. La taille optimale du quadrat, qui correspond au minimum de la courbe dissimilarité / taille du quadrat, est fonction de l'intensité du processus, comme le montre la figure F.3.

En prenant un quadrat de surface $\frac{2}{\lambda}$, où λ est l'intensité du processus, on ne s'écarte pas trop de cet optimum. Dans le cas d'un processus homogène, cette expression s'interprète comme le fait qu'il y a en moyenne $\mu(Q) = 2$ événements dans un quadrat.

Lien avec les processus ponctuels

Dans la mesure où les quadrats sont indépendants, a, b, c, d sont les réalisations d'une loi multinomiale de paramètres N et A, B, C, D , avec :

$$A = \Pr [N_1(Q) > 0 \text{ et } N_2(Q) > 0]$$

$$B = \Pr [N_1(Q) > 0 \text{ et } N_2(Q) = 0]$$

$$C = \Pr [N_1(Q) = 0 \text{ et } N_2(Q) > 0]$$

$$D = \Pr [N_1(Q) = 0 \text{ et } N_2(Q) = 0]$$

où $N_i(Q)$ est la variable aléatoire donnant le nombre d'individus de l'espèce i présents dans le quadrat Q . S'il y a indépendance des processus marginaux :

$$A = \Pr [N_1(Q) > 0] \times \Pr [N_2(Q) > 0]$$

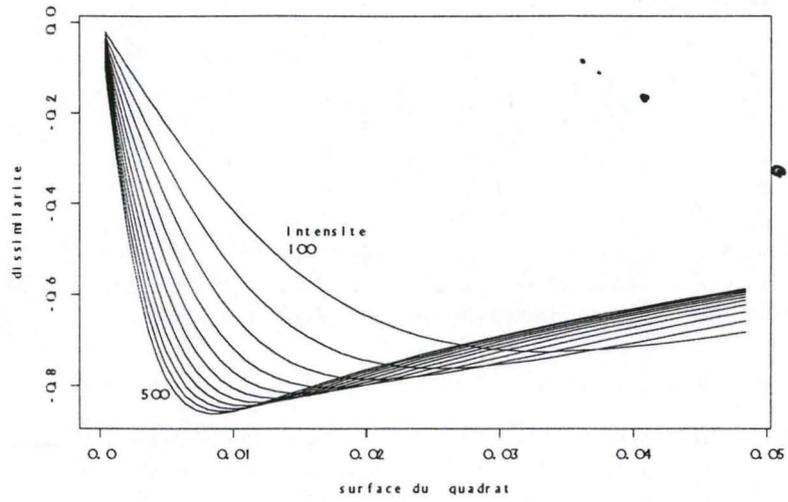


Figure F.2: Influence de la taille du quadrat sur la valeur de l'indice (cas théorique)

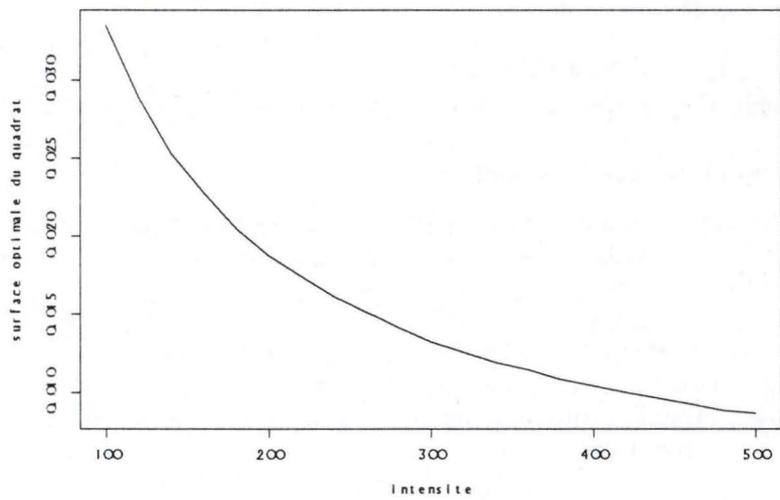


Figure F.3: Taille optimale du quadrat pour le processus de séparation totale

etc.

En particulier pour la superposition de deux processus de Poisson homogènes indépendants d'intensités respectives λ_1 et λ_2 :

$$\begin{aligned} A &= (1 - e^{-\lambda_1 S}) (1 - e^{-\lambda_2 S}) \\ B &= (1 - e^{-\lambda_1 S}) e^{-\lambda_2 S} \\ C &= e^{-\lambda_1 S} (1 - e^{-\lambda_2 S}) \\ D &= e^{-(\lambda_1 + \lambda_2) S} \end{aligned}$$

où S est la surface du quadrat. Un autre exemple de calcul de A, B, C, D est donné en annexe C. De façon générale, connaissant le processus ponctuel, on peut calculer A, B, C, D et remonter à la loi de a, b, c, d . Les calculs sont rapidement compliqués.

Variantes de cette méthode

Si les nombres d'individus $N_1(Q)$ et $N_2(Q)$ de chaque espèce présents dans les quadrats sont élevés et que l'on observe rarement des quadrats vides (ce qui correspond donc à des quadrats de plus grande taille), on peut définir une distance en estimant la covariance de $N_1(Q)$ et $N_2(Q)$. Greig-Smith [26] suggère de calculer le coefficient de corrélation linéaire sur les échantillons $n_1(Q_j)$ et $n_2(Q_j)$, $j = 1..N$, des nombres d'individus des espèces 1 et 2 présents dans les N quadrats $Q_1...Q_N$, mais il n'y a aucune raison de supposer la relation linéaire.

Si l'une des espèces est abondante dans les quadrats tandis que l'autre espèce peut être décrite par sa présence / absence :

| | | | |
|---|-------------|-------------|-------------|
| | espèce 1 | | |
| | présente | absente | |
| espèce 2 : nombre d'individus comptés dans les quadrat | $x_1...x_r$ | $y_1...y_s$ | $r + s = N$ |

Greig-Smith suggère de comparer la moyenne des échantillons $(x_1...x_r)$ et $(y_1...y_s)$ par un test des espérances de Student, ce qui suppose que les comptages sont indépendants et que l'on peut approximer les x_i par une loi normale.

Un exemple d'application de ces méthodes est donné par Kershaw [37]

F.1.2. Méthode de Kershaw [38]

Le principe de cette méthode est de calculer une estimation de la covariance de $N_1(Q)$ et $N_2(Q)$, où $N_i(Q)$ est la variable aléatoire représentant le nombre d'événements de type i présents dans le quadrat Q .

Si les processus marginaux sont indépendants : $\text{Cov}(N_1(Q), N_2(Q)) = 0$

S'il y a association positive : $\text{Cov}(N_1(Q), N_2(Q)) > 0$

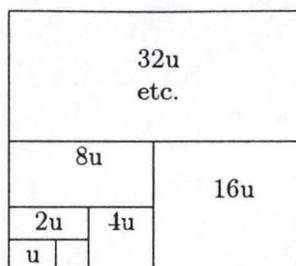
S'il y a association négative : $\text{Cov}(N_1(Q), N_2(Q)) < 0$

Soit $N(Q)$ la variable aléatoire représentant le nombre d'événements (sans distinction d'espèce) présents dans le quadrat Q .

$$N(Q) = N_1(Q) + N_2(Q) \implies \text{Cov}(N_1(Q), N_2(Q)) = \frac{1}{2} [\text{Var}(N(Q)) - \text{Var}(N_1(Q)) - \text{Var}(N_2(Q))]$$

Pour estimer $\text{Cov}(N_1(Q), N_2(Q))$, il suffit donc de savoir estimer la variance du nombre d'événements présents dans un quadrat pour un processus ponctuel univarié. La méthode de Kershaw consiste à estimer $\text{Cov}(N_1(Q), N_2(Q))$ en utilisant comme estimation de $\text{Var}(N(Q))$ la variance locale V_k de Greig-Smith, définie de la façon suivante : la zone d'étude est divisée selon une grille régulière de 2^n (pour les simulations, $n = 8$) quadrats de surface u ; les quadrats sont regroupés successivement

deux par deux en 2^{n-1} quadrats de taille $2u$, 2^{n-2} quadrats de taille $4u$, etc., jusqu'à un quadrat de taille $2^n u$, comme indiqué sur la figure :



Soit $x_{i,k}$ le nombre d'événements présents dans le i^e quadrat ($i = 1 \dots 2^{n-k}$) de taille $2^k u$. Par définition :

$$V_k = \frac{1}{2^n} \left(\sum_{i=1}^{2^{n-k}} 2x_{i,k}^2 - \sum_{i=1}^{2^{n-k-1}} x_{i,k+1}^2 \right), \quad k = 0, 1, \dots, (n-1)$$

(V_k est en fait une estimation de $\text{Var}(N(Q)) \times \frac{u}{\nu(Q)}$, où $\nu(Q)$ est la surface du quadrat Q).

On obtient donc une courbe représentant $\text{Cov}(N_1(Q), N_2(Q)) \times \frac{u}{\nu(Q)}$ en fonction de $\nu(Q)$, pour des valeurs de $\frac{\nu(Q)}{u}$ puissances de deux.

Définition de la dissimilarité

La question se pose alors de définir un indice de dissimilarité à partir de cette courbe. La figure F.4 montre la courbe $\text{Cov}(N_1(Q), N_2(Q)) / \nu(Q)$ versus $\nu(Q)$ pour le processus de séparation totale défini en annexe C : pour de petites tailles de quadrats, la covariance est quasi-nulle ; la courbe passe par un minimum négatif (association négative) pour une taille de quadrat correspondant à la taille des plaques monospécifiques, puis croît brutalement. Il n'est donc pas possible de définir la distance comme la valeur extrême de la courbe : dans le cas présent, la valeur extrême de la courbe est positive alors qu'il y a association négative ! Il n'est pas non plus possible de définir la distance comme la valeur de la courbe en une taille de quadrat donnée, puisque la taille de quadrat pour laquelle la covariance a une valeur significative dépend de l'échelle à laquelle est perceptible la dépendance des processus marginaux.

La distance a donc été définie comme la moyenne de la covariance estimée pour les différentes tailles de quadrats :

$$d(A, B) = \frac{1}{2^n} \sum_{k=0}^{n-1} (V_{(A \cup B), k} - V_{A, k} - V_{B, k})$$

où $V_{X, k}$ est la variance locale de Greig-Smith calculée en prenant en compte les événements de type X .

Lien avec les processus ponctuels

Il est en général assez facile de calculer $\text{Cov}(N_1(Q), N_2(Q))$ pour un processus ponctuel bivarié donné. Un exemple de calcul est donné en annexe C pour le processus de séparation totale évoqué ci-dessus.

Il est en revanche très difficile de calculer la distribution de $d(A, B)$, même pour les processus les plus simples. Pour un processus de Poisson homogène d'intensité λ , on remarquera simplement que $2^{n-k-1} \frac{V_k}{\lambda u}$ suit approximativement une loi du χ^2 à 2^{n-k-1} degrés de liberté [62].

Variantes de cette méthode

Il existe autant de variantes de cette méthode qu'il y a de façons d'estimer $\text{Var}(N(Q))$ par une variance locale. Dale & Blundon [11] et Dale & Powell [12] proposent plusieurs expressions, assez compliquées, de la variance locale lorsque les quadrats sont répartis sur une grille bidimensionnelle.

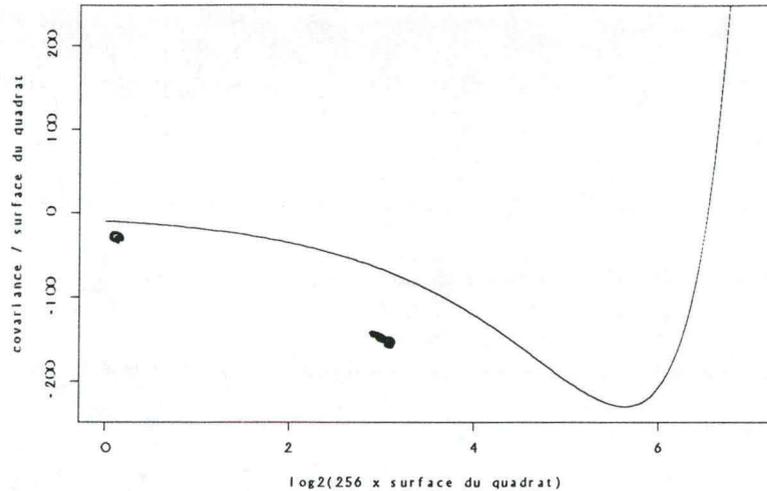


Figure F.4: Courbe covariance versus taille du quadrat pour le processus de séparation totale

Lorsque les quadrats sont répartis non plus sur une grille bidimensionnelle mais en ligne, on pourra utiliser comme expression de la variance locale celle de Kershaw [36], de Usher [65], ou celles de Hill [31]. Toutes ces méthodes reposent sur le regroupement de quadrats en quadrats de plus en plus grands. Une autre façon de procéder est d'utiliser des quadrats de taille constante, mais de plus en plus éloignés les uns des autres. On obtient alors l'expression de la variance locale de Goodall [23] et de Ludwig & Goodall [43], cette dernière expression étant en fait celle du variogramme.

Par ailleurs, Besag [4] décrit une autre méthode utilisant des quadrats répartis sur une grille bidimensionnelle régulière. L'idée, inspirée du test univarié de Mead [44], consiste à comparer la situation observée à celle que l'on observerait en permutant les quadrats au hasard à l'intérieur de sous-blocs de taille variable.

F.2. Méthodes par distances

Dans les méthodes par distances, l'information collectée est des distances, soit entre arbres, soit entre des points et des arbres. Certaines méthodes ne nécessitent qu'un échantillonnage partiel de distances. D'autres au contraire nécessitent un échantillonnage complet ; il est alors avantageux de réaliser une cartographie de la zone d'étude, en notant la position spatiale de tous les individus.

F.2.1. Méthode de Diggle & Cox [17, 64, p.250]

Soit X_i la distance d'un point arbitraire à l'arbre de l'espèce i le plus proche ($i = 1, 2$). Le principe de la méthode de Diggle & Cox est de comparer X_1 à X_2 . En effet pour des processus ponctuels bivariés homogènes

- si les processus marginaux sont indépendants : X_1 et X_2 sont des variables aléatoires indépendantes
- s'il y a attraction entre les processus marginaux : X_1 et X_2 sont positivement corrélés
- s'il y a répulsion entre les processus marginaux : X_1 et X_2 sont négativement corrélés.

La dissimilarité est définie comme la valeur du coefficient de corrélation des rangs τ de Kendall calculé sur un échantillon $(x_{1k}, x_{2k})_{k=1..n}$ de taille n de (X_1, X_2)

- $\tau = 0$ correspond à l'indépendance des répartitions marginales
- $\tau = 1$ correspond à une association positive entre les répartitions marginales
- $\tau = -1$ correspond à une association négative entre les répartitions marginales.

Comment réaliser l'échantillonnage ?

La méthode d'échantillonnage suivie est celle proposée par Diggle & Cox : n points $O_1 \dots O_n$ sont placés sur la zone d'étude selon un réseau régulier. De chaque point O_k sont mesurées les distances x_{1k} et x_{2k} aux arbres les plus proches des espèces 1 et 2. Les distances x_{1k} et x_{2k} ne sont en fait retenues que si elles sont inférieures à la distance de O_k au bord de la zone d'étude le plus proche.

La figure F.5 montre les variations de la dissimilarité avec le nombre n de points d'échantillonnage pour quatre processus ponctuels d'intensité $\lambda = 2000$. Pour les simulations, n a été pris égal à 5% du nombre total d'arbres présents dans la zone d'étude, comme le suggèrent Diggle & Cox.

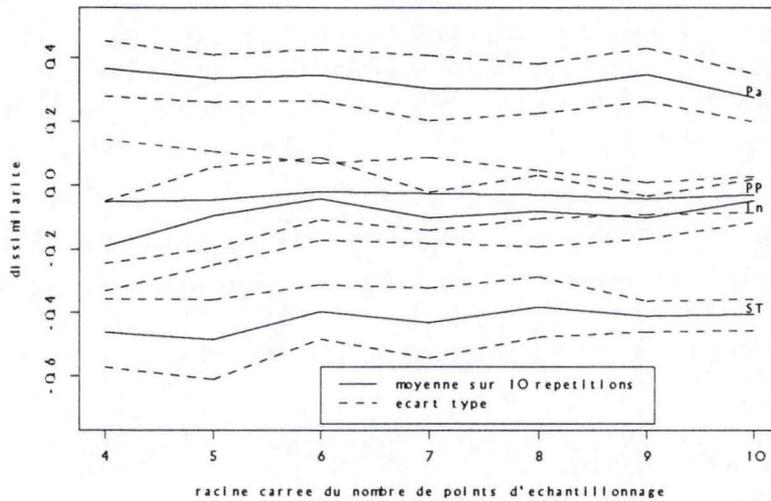


Figure F.5: Variations de l'indice de Diggle & Cox avec le nombre de points d'échantillonnage

Variantes de cette méthode

On peut chercher d'abord à améliorer la méthode de Diggle & Cox en collectant plus d'information en chaque point d'échantillonnage. Deux façons de procéder peuvent être envisagées :

- en chaque point d'échantillonnage O_k sont mesurées les distances $x_{ik}^{(l)}$ au l^{eme} plus proche arbre de l'espèce i ($i = 1, 2; l = 1 \dots L$). Le coefficient de corrélation des rangs τ de Kendall est alors calculé L fois sur les échantillons $(x_{1k}^{(l)})_{k=1..n}$ et $(x_{2k}^{(l)})_{k=1..n}$, et la dissimilarité est définie comme la moyenne de ces L valeurs.
- le plan autour de chaque point d'échantillonnage est découpé en L secteurs égaux. Soit $x_{ik}^{(l)}$ la distance de O_k à l'arbre de l'espèce i le plus proche dans le l^{eme} secteur du plan. La

dissimilarité est alors définie comme la valeur du coefficient de corrélation des rangs τ de Kendall calculé sur les échantillons $(x_{1k}^{(l)})_{k=1..n, l=1..L}$ et $(x_{2k}^{(l)})_{k=1..n, l=1..L}$.

Ces deux variantes ont également été étudiées pour $L = 4$. Là encore, le taux d'échantillonnage a été choisi égal pour les simulations à 5% du nombre total d'arbres présents dans la zone d'étude.

F.2.2. Méthode de Goodall [9, 17, 22, 64]

Soit X_i la distance d'un point arbitraire à l'arbre de l'espèce i le plus proche, et Y_{ij} la distance d'un arbre de l'espèce i pris au hasard à son plus proche voisin de l'espèce j . Soit F_i la fonction de répartition de X_i , et G_{ij} celle de Y_{ij} ($i = 1, 2; j = 3 - i$).

Le principe de la méthode de Goodall est de comparer F_i à G_{ij} . En effet, pour des processus ponctuels bivariés *homogènes* :

- si les processus marginaux sont indépendants : $\forall x, F_1(x) = G_{21}(x)$ et $F_2(x) = G_{12}(x)$
- s'il y a attraction entre les processus marginaux : $\exists x / F_1(x) > G_{21}(x)$ ou $F_2(x) > G_{12}(x)$
- s'il y a répulsion entre les processus marginaux : $\exists x / F_1(x) < G_{21}(x)$ ou $F_2(x) < G_{12}(x)$

L'indice de dissimilarité est donc défini de la façon suivante : soit $(x_{ik})_{k=1..n}$ un échantillon de taille n de X_i et $(y_{ijk})_{k=1..m}$ un échantillon de taille m de Y_{ij} , tels que $(x_{ik})_{k=1..n}$ et $(y_{ijk})_{k=1..m}$ soient des échantillons indépendants. Soit u_i la valeur de la statistique U de Wilcoxon calculée sur $(x_{ik}, y_{jik'})_{k=1..n, k'=1..m}$ ($i = 1, 2; j = 3 - i$)

- $u_i = 0$ correspond une association négative des répartitions marginales
- $u_i = mn$ correspond à une association positive des répartitions marginales
- $u_i = \frac{mn}{2}$ correspond à l'indépendance des répartitions marginales

La dissimilarité entre les deux répartitions marginales est donc posée égale à

$$d(1, 2) = \frac{1}{2} \left[\left(\frac{2u_1}{mn} - 1 \right) + \left(\frac{2u_2}{mn} - 1 \right) \right] = \frac{u_1 + u_2}{mn} - 1$$

Comment réaliser l'échantillonnage ?

Plusieurs questions se posent à ce niveau : quelle doit être la taille des échantillons ? Comment placer les points d'échantillonnage ? Comment choisir un arbre au hasard ? Comment traiter les effets de bord ?

Nous avons suivi pour les simulations la méthode d'échantillonnage proposée par Diggle & Cox [17, p.17] : $2n_1$ points $O_1, O_2 \dots O_{n_1}, O_1^*, O_2^* \dots O_{n_1}^*$ sont répartis sur la zone d'étude selon un réseau régulier. De chaque point O_k est mesurée la distance x_{1k} à l'arbre de l'espèce 1 le plus proche, ainsi que la distance y_{21k} de l'arbre de l'espèce 2 le plus proche de O_k à son plus proche voisin de l'espèce 1. On procède de même pour les points O_k^* . On obtient alors deux jeux d'échantillons (x_{1k}) et (y_{21k}^*) d'une part, (x_{1k}^*) et (y_{21k}) d'autre part, qui peuvent être considérés comme indépendants dans la mesure où les distances entre points d'échantillonnage sont grandes devant la distance à laquelle est perceptible l'autocorrélation spatiale.

De même $2n_2$ points d'échantillonnage $Q_1, Q_2 \dots Q_{n_2}, Q_1^*, Q_2^* \dots Q_{n_2}^*$ sont répartis sur la zone d'étude selon un réseau régulier. De chaque point Q_k est mesurée la distance x_{2k} à l'arbre de l'espèce 2 le plus proche, ainsi que la distance y_{12k} de l'arbre de l'espèce 1 le plus proche de Q_k à son plus proche voisin de l'espèce 2. On procède de même pour les points Q_k^* . On obtient alors deux jeux d'échantillons (x_{2k}) et (y_{12k}^*) d'une part, (x_{2k}^*) et (y_{12k}) d'autre part.

Il y a alors quatre calculs possibles de la statistique U de Wilcoxon. La dissimilarité est prise égale à la moyenne des résultats de ces quatre calculs.

Pour respecter l'indépendance des échantillons tout en gardant une puissance suffisante, Diggle & Cox suggèrent de prendre un nombre de points d'échantillonnage n_i égal à 10% du nombre d'arbres de l'espèce i présents dans la zone d'étude. Comme rien n'empêche en fait de confondre les points O et Q , le nombre total de points d'échantillonnage est égal à $\max(n_1, n_2)$, et varie entre 5 et 10% du nombre total d'arbres présents dans la zone d'étude. La figure F.6 montre les variations de la dissimilarité avec le nombre total $n = \max(n_1, n_2)$ de points d'échantillonnage, pour quatre processus ponctuels bivariés d'intensités marginales $(\lambda_1, \lambda_2) = (1000, 1000)$: l'écart type de la distance diminue lorsque n augmente.

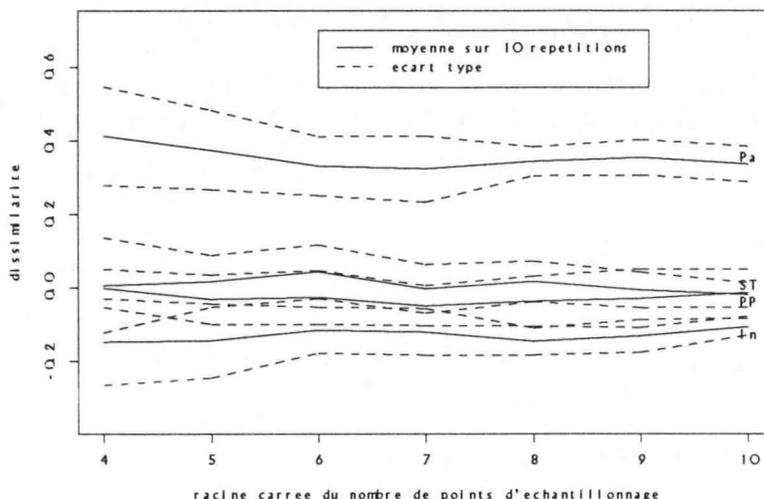


Figure F.6: Variation de l'indice de Goodall avec le nombre de points d'échantillonnage

Les effets de bord ont été traités de la façon suivante : la distance x_{ik} (respectivement y_{ijk}) n'est prise en compte que si elle est inférieure à la distance de O_k (respectivement de l'arbre de l'espèce i le plus proche de O_k) au bord le plus proche.

Lien avec les processus ponctuels

Les fonctions F_i et G_{ij} sont connues pour les processus ponctuels les plus courants (voir par exemple [51]).

Variantes de cette méthode

Il existe plusieurs statistiques permettant de comparer les fonctions F_i et G_{ij} . La méthode initialement proposée par Goodall [22] consistait en fait à comparer les échantillons $(\sqrt{x_{1k}})_{k=1..n}$ et $(\sqrt{y_{21k}})_{k=1..n}$ d'une part, $(\sqrt{x_{2k}})_{k=1..n}$ et $(\sqrt{y_{12k}})_{k=1..n}$ d'autre part par un test des espérances de Student. La méthode proposée ici est une variante non paramétrique proposée par Diggle & Cox.

Diggle & Cox [17] proposent en outre d'utiliser la statistique de Kolmogorov-Smirnov $\max_x \left| \tilde{F}_i(x) - \tilde{G}_{ji}(x) \right|$ $1, 2; j = 3 - i$, où \tilde{F}_i et \tilde{G}_{ij} sont des estimateurs de F_i et G_{ij} obtenus à partir des échantillons de X_i et Y_j (Upton & Fingleton [64] proposent des expressions pour ces estimateurs).

F.2.3. Méthode des plus proches voisins

Comme précédemment, on note Y_{ij} la distance d'un arbre arbitraire de l'espèce i à son plus proche voisin de l'espèce j . L'idée intuitive sur laquelle repose cette méthode est que si deux nuages de points sont bien séparés, comme sur la figure 2.2a, alors les distances Y_{12} et Y_{21} vont être grandes devant les distances moyennes entre individus des espèces 2 et 1, alors que si les deux nuages de points sont bien mélangés, comme sur la figure 2.2c, les distances Y_{12} et Y_{21} vont être du même ordre de grandeur que les distances moyennes entre individus des espèces 2 et 1. Et s'il y a association positive entre les deux nuages de points, comme sur la figure 2.2b, alors les distances Y_{12} et Y_{21} vont être petites devant les distances moyennes entre individus des espèces 2 et 1.

La dissimilarité est donc définie de la façon suivante : soit $\{E_{i1}, E_{i2} \dots E_{in_i}\}$ l'ensemble des individus de l'espèce i présents dans la zone d'étude, et y_{ijk} la distance de E_{ik} à son plus proche voisin de l'espèce j ($i = 1, 2; j = 3 - i; k = 1 \dots n_i$). Pour éliminer les effets de bord, on ne prend en compte que les distances y_{ijk} inférieures à la distance de E_{ik} au bord de la zone d'étude le plus proche. Soit $\{E_{i1}^*, E_{i2}^* \dots E_{in_i}^*\}$ l'ensemble des individus de l'espèce i vérifiant cette condition. La dissimilarité $d(1, 2)$ entre les deux nuages de points est alors définie par :

$$d(1, 2) = \frac{2}{n_1^* + n_2^*} \left(\sqrt{\frac{n_2}{A}} \sum_{k=1}^{n_1^*} y_{12k}^* + \sqrt{\frac{n_1}{A}} \sum_{k=1}^{n_2^*} y_{21k}^* \right)$$

où A est la surface de la zone d'étude. La raison de la multiplication par $\sqrt{n_1/A}$ et $\sqrt{n_2/A}$ va être exposée maintenant.

Lien avec les processus ponctuels

Les moyennes $\left(\frac{1}{n_1^*} \sum_{k=1}^{n_1^*} y_{12k}^*\right)$ et $\left(\frac{1}{n_2^*} \sum_{k=1}^{n_2^*} y_{21k}^*\right)$ peuvent être considérées comme des estimations de $E(Y_{12})$ et de $E(Y_{21})$. Les nombres n_1 et n_2 peuvent être considérés comme des estimations des intensités λ_1 et λ_2 des processus ponctuels marginaux. La dissimilarité $d(1, 2)$ peut donc être considérée comme une estimation de la moyenne pondérée par n_1^* et n_2^* de $\frac{E(Y_{12})}{(2\sqrt{\lambda_2})^{-1}}$ et de $\frac{E(Y_{21})}{(2\sqrt{\lambda_1})^{-1}}$.

Connaissant le processus ponctuel bivarié, $E(Y_{12})$ et $E(Y_{21})$ sont connus. En particulier, si les processus marginaux sont indépendants : $G_{12} = F_2$ et $G_{21} = F_1$, donc $E(Y_{12}) = E(X_2)$ et $E(Y_{21}) = E(X_1)$, où X_i est la distance d'un point arbitraire à l'arbre de l'espèce i le plus proche. Pour un processus de Poisson homogène d'intensité λ , $E(X) = \frac{1}{2\sqrt{\lambda}}$.

Donc pour un processus ponctuel bivarié formé de la superposition de deux processus de Poisson homogènes indépendants

$$\frac{n_1^*}{n_1^* + n_2^*} \frac{E(Y_{12})}{(2\sqrt{\lambda_2})^{-1}} + \frac{n_2^*}{n_1^* + n_2^*} \frac{E(Y_{21})}{(2\sqrt{\lambda_1})^{-1}} = 1$$

De même pour un processus univarié agrégé : $E(X) > \frac{1}{2\sqrt{\lambda}}$
 pour un processus univarié régulier : $E(X) < \frac{1}{2\sqrt{\lambda}}$

Le comportement de la distance dans les six cas de processus ponctuels bivariés ayant des processus marginaux indépendants peut donc être résumé dans le tableau suivant :

| dissimilarité | processus 1 | | |
|-----------------------|-------------|-----------|----------|
| | agrégé | aléatoire | régulier |
| processus 2 agrégé | $\gg 1$ | > 1 | ? |
| processus 2 aléatoire | | 1 | < 1 |
| processus 2 régulier | | | $\ll 1$ |

Contrairement à toutes les distances vues jusqu'à maintenant, cette méthode donne des résultats différents quand il y a indépendance des processus marginaux selon la nature des processus marginaux.

Si le calcul de $E(Y_{ij})$ est généralement faisable, en revanche le calcul de la distribution de la distance $d(1,2)$ est très compliqué. En particulier, comme on prend en compte tous les arbres, les distances y_{12k} et y_{21k} ne peuvent pas être considérées comme les réalisations de variables aléatoires indépendantes.

Variante de cette méthode

Plutôt que de prendre en compte tous les arbres présents dans la zone d'étude, on peut faire un échantillonnage partiel de n arbres de chaque espèce, et mesurer la distance à leur plus proche voisin de l'espèce opposée. Pour rendre cette méthode plus précise, on peut comme précédemment collecter plus d'information de deux façons :

- on échantillonne n arbres de chaque espèce ; soit $y_{ijk}^{(l)}$ la distance du $k^{\text{ème}}$ arbre ($k = 1 \dots n$) de l'espèce i ($i = 1, 2$) à son $l^{\text{ème}}$ ($l = 1 \dots L$) plus proche voisin de l'espèce j ($j = 3 - i$). Pour définir la distance on prendra comme précédemment le processus ponctuel bivarié constitué de deux processus de Poisson homogènes indépendants comme processus de référence. Etant donné un processus ponctuel bivarié homogène, soit $Y_{ij}^{(l)}$ la distance d'un arbre arbitraire de l'espèce i à son $l^{\text{ème}}$ plus proche voisin de l'espèce j , et $X_i^{(l)}$ la distance d'un point arbitraire au $l^{\text{ème}}$ plus proche arbre de l'espèce i . Si les processus marginaux sont indépendants, $Y_{ij}^{(l)}$ et $X_j^{(l)}$ sont des variables aléatoires identiquement distribuées.

Pour un processus de Poisson homogène d'intensité λ :

$$E(X^{(l)}) = \frac{1}{\sqrt{\lambda\pi}} \frac{\Gamma(l + \frac{1}{2})}{\Gamma(l)}$$

donc dans le cas de la superposition de deux processus de Poisson homogènes indépendants

$$E(Y_{12}^{(l)}) = \frac{1}{\sqrt{\lambda_2\pi}} \frac{\Gamma(l + \frac{1}{2})}{\Gamma(l)} \quad \text{et} \quad E(Y_{21}^{(l)}) = \frac{1}{\sqrt{\lambda_1\pi}} \frac{\Gamma(l + \frac{1}{2})}{\Gamma(l)}$$

La dissimilarité est donc définie dans ce cas par :

$$d(1,2) = \frac{\sqrt{\pi}}{2L} \sum_{l=1}^L \frac{\Gamma(l)}{\Gamma(l + \frac{1}{2})} \left(\sqrt{\hat{\lambda}_2} \frac{1}{n} \sum_{k=1}^n y_{12k}^{(l)} + \sqrt{\hat{\lambda}_1} \frac{1}{n} \sum_{k=1}^n y_{21k}^{(l)} \right)$$

où $\hat{\lambda}_1$ et $\hat{\lambda}_2$ sont des estimateurs de l'intensité des processus marginaux.

- on échantillonne n arbres de chaque espèce ; autour de chaque arbre le plan est découpé en L secteurs égaux. Soit $y_{ijk}^{(l)}$ la distance du $k^{\text{ème}}$ arbre ($k = 1 \dots n$) de l'espèce i ($i = 1, 2$) à son plus proche voisin de l'espèce j ($j = 3 - i$) situé dans le $l^{\text{ème}}$ secteur du plan. Pour définir la distance, on prendra encore le processus ponctuel constitué de la superposition de deux processus de Poisson homogène indépendant comme processus de référence. Etant donné un processus ponctuel bivarié homogène, soit $Y_{ij}^{(l)}$ la distance d'un arbre arbitraire de l'espèce i à son plus proche voisin de l'espèce j situé dans le $l^{\text{ème}}$ secteur du plan, et $X_i^{(l)}$ la distance d'un point arbitraire au plus proche arbre de l'espèce i situé dans le $l^{\text{ème}}$ secteur du plan. Si les processus marginaux sont indépendants, $Y_{ij}^{(l)}$ et $X_j^{(l)}$ sont des variables aléatoires identiquement distribuées.

Pour un processus de Poisson homogène d'intensité λ :

$$E(X^{(l)}) = \frac{\sqrt{L}}{2\sqrt{\lambda}}$$

donc dans le cas de la superposition de deux processus de Poisson homogènes indépendants

$$E\left(Y_{12}^{(l)}\right) = \frac{\sqrt{L}}{2\sqrt{\lambda_2}} \text{ et } E\left(Y_{21}^{(l)}\right) = \frac{\sqrt{L}}{2\sqrt{\lambda_1}}$$

La dissimilarité est donc définie dans ce cas par :

$$d(1,2) = \frac{1}{L^{3/2}} \sum_{l=1}^L \left(\sqrt{\hat{\lambda}_2} \frac{1}{n} \sum_{k=1}^n y_{12k}^{(l)} + \sqrt{\hat{\lambda}_1} \frac{1}{n} \sum_{k=1}^n y_{21k}^{(l)} \right)$$

Ces deux variantes ont également été étudiées. La méthode d'échantillonnage des arbres est la suivante : n points sont répartis sur la zone d'étude selon un réseau régulier ; les n arbres d'échantillonnage de l'espèce i sont les arbres de l'espèce i les plus proches de ces n points d'échantillonnage. Les effets de bord sont traités comme précédemment. On prend comme estimation des intensités des processus marginaux le nombre d'arbres de chaque espèce présents sur la zone d'étude. Le nombre n de points d'échantillonnage a été pris égal pour les simulations à 5% du nombre total d'arbres présents sur la zone d'étude.

F.2.4. Méthode du Sup de l'Inf [5]

Cette repose sur la même idée que la méthode des plus proches voisins. Si l'on note comme précédemment $\{E_{i1}^* \dots E_{in_i}^*\}$ l'ensemble des individus de l'espèce i plus proches de leur plus proche voisin de l'espèce j ($j = 3 - i$) que du bord de la zone d'étude, et y_{ijk}^* la distance de E_{ik}^* à son plus proche voisin de l'espèce j , la dissimilarité est définie non plus comme une moyenne pondérée des y_{ijk}^* , mais à partir du maximum des y_{ijk}^* :

$$d(1,2) = \frac{1}{2} \left(\frac{1}{C(n_1^*, n_2/A)} \max_{k=1..n_1^*} \{y_{12k}^*\} + \frac{1}{C(n_2^*, n_1/A)} \max_{k=1..n_2^*} \{y_{21k}^*\} \right)$$

avec $C(a,b) = \frac{2a}{\sqrt{\pi b}} \int_0^\infty u^2 e^{-u^2} (1 - e^{-u^2})^{a-1} du \quad , \forall (a,b) \in \mathbb{N}^* \times \mathbb{R}^*$

où n_i est le nombre d'arbres de l'espèce i présents dans la zone d'étude, et A la surface de la zone d'étude. La raison de la normalisation par le facteur C est de rendre la distance moins sensible aux variations d'intensité. Le facteur est en fait calculé de sorte que que $d(1,2) \approx 1$ pour un processus formé de la superposition de deux processus de Poisson homogènes indépendants d'intensités respectives λ_1 et λ_2 , quels que soient λ_1 et λ_2 .

Lien avec les processus ponctuels

Conditionnellement à la présence de n_i^* individus de l'espèce i dans la zone d'étude, soit Y_{ijk}^* ($k = 1 \dots n_i^*$) la distance du k^{eme} de ces individus à son plus proche voisin de l'espèce j ($j = 3 - i$)

$$\Pr \left[\max_{k=1..n_i^*} Y_{ijk}^* \leq y \right] = \Pr \left[Y_{ij1}^* \leq y \text{ et } Y_{ij2}^* \leq y \dots \text{ et } Y_{ijn_i^*}^* \leq y \right]$$

Pour un processus formé de la superposition de deux processus de Poisson homogènes indépendants d'intensités respectives λ_1 et λ_2 , les Y_{ijk}^* (i et j fixés, k variable) sont des variables indépendantes identiquement distribuées, donc :

$$\Pr \left[\max_{k=1..n_i^*} Y_{ijk}^* \leq y \right] = (\Pr [Y_{ij} \leq y])^{n_i^*}$$

où Y_{ij} est la distance d'un individu arbitraire de l'espèce i à son plus proche voisin de l'espèce j . Les processus marginaux étant indépendants, Y_{ij} est distribué comme la distance X_j d'un point arbitraire à l'individu de l'espèce j le plus proche.

Pour un processus de Poisson homogène d'intensité λ , $\Pr[X \leq x] = 1 - e^{-\lambda\pi x^2}$

$$\text{d'où } \Pr \left[\max_{k=1..n_i^*} Y_{ijk}^* \leq y \right] = \left(1 - e^{-\lambda_j \pi y^2} \right)^{n_i^*}$$

Conditionnellement à n_i^* , la densité de probabilité de $\Delta_{ij} = \max_{k=1..n_i^*} Y_{ijk}^*$ est donc :

$$f_{ij}(y) = n_i^* \left(1 - e^{-\lambda_j \pi y^2} \right)^{n_i^*-1} 2\lambda_j \pi y e^{-\lambda_j \pi y^2}$$

D'où conditionnellement à n_i^* :

$$E(\Delta_{ij}) = \int_0^\infty y f_{ij}(y) dy = \frac{2n_i^*}{\sqrt{\lambda_j \pi}} \int_0^\infty u^2 e^{-u^2} \left(1 - e^{-u^2} \right)^{n_i^*-1} du = C(n_i^*, \lambda_j)$$

et donc pour un processus bivarié formé de la superposition de deux processus de Poisson homogènes indépendants d'intensités respectives λ_1 et λ_2 , conditionnellement à la prise en compte de n_1^* individus de l'espèce 1 et de n_2^* individus de l'espèce 2 :

$$\frac{1}{2} \left(\frac{E(\Delta_{12})}{C(n_1^*, \lambda_2)} + \frac{E(\Delta_{21})}{C(n_2^*, \lambda_1)} \right) = 1$$

Les calculs deviennent très compliqués pour d'autres types de processus, car on ne peut plus considérer les $(Y_{ijk}^*)_{k=1..n_i^*}$ comme des variables indépendantes.

F.2.5. Méthode de Pielou [20, 45, 54, 64, p.242]

Cette méthode permet de comparer le voisinage (c'est-à-dire la nature du plus proche voisin) des arbres de l'espèce 1 et le voisinage des arbres de l'espèce 2. Pour chaque arbre présent dans la zone d'étude, on relève son espèce et l'espèce de son plus proche voisin. Cette information peut être résumée dans un tableau de contingence :

| | | plus proche voisin | |
|---------------------|----------|--------------------|----------|
| | | espèce 1 | espèce 2 |
| individu de base | espèce 1 | a | b |
| | espèce 2 | c | d |

Les effets de bord sont pris en compte en ne retenant que les individus qui sont plus près de leur plus proche voisin que du bord de la zone d'étude. Le nombre a représente le nombre d'individus de l'espèce 1 ayant comme plus proche voisin un individu de l'espèce 1, etc.

La dissimilarité est définie par une distance de corrélation de points sur le tableau de contingence :

$$\rho = \frac{ad - bc}{\sqrt{(a+c)(a+b)(c+d)(b+d)}}$$

Lorsque les proportions d'individus ayant comme plus proche voisin un individu de l'espèce 1 ou 2 sont les mêmes pour les deux espèces, ρ est voisin de zéro ; ρ devient significativement supérieur à zéro lorsque les proportions d'individus ayant comme plus proche voisin un individu de la même espèce dépasse les proportions d'individus ayant comme plus proche voisin un individu de l'espèce opposée ; ρ est significativement inférieur à zéro dans la situation inverse de la situation précédente.

Il faut être bien conscient que ρ ne mesure pas un écart à l'indépendance des processus marginaux. On peut avoir ρ non significativement différent de zéro pour un processus bivarié dont les processus marginaux sont fortement dépendants. Dixon [20] en donne un exemple frappant. On peut avoir à l'inverse ρ significativement différent de zéro pour un processus bivarié dont les processus marginaux sont indépendants. Upton & Fingleton [64, p.244] en donnent un exemple. En fait ρ ne mesure un écart à l'indépendance des processus marginaux que dans le cas où les processus marginaux sont des processus de Poisson homogènes.

Lien avec les processus ponctuels

Si au lieu de relever tous les individus présents dans la zone d'étude, on en tire N au hasard de sorte qu'ils puissent être considérés comme des tirages indépendants : a, b, c, d sont alors les réalisations d'une loi multinomiale de paramètres N et A, B, C, D , avec :

$$\begin{aligned} A &= \Pr [E \in e_1 \text{ et } \text{ppv}(E) \in e_1] \\ B &= \Pr [E \in e_1 \text{ et } \text{ppv}(E) \in e_2] \\ C &= \Pr [E \in e_2 \text{ et } \text{ppv}(E) \in e_1] \\ D &= \Pr [E \in e_2 \text{ et } \text{ppv}(E) \in e_2] \end{aligned}$$

où E désigne un arbre tiré au hasard, $\text{ppv}(E)$ son plus proche voisin, et e_i l'ensemble des individus de l'espèce i .

$$\begin{aligned} A &= \Pr [\text{ppv}(E) \in e_1 / E \in e_1] \times \Pr [E \in e_1] \\ \Pr [E \in e_1] &= \frac{n_1}{n_1 + n_2} \\ \Pr [\text{ppv}(E) \in e_1 / E \in e_1] &= \Pr [Y_{11} < Y_{12}] \end{aligned}$$

où n_i désigne le nombre d'arbres de l'espèce i présents sur la zone d'étude, et Y_{ij} est la distance d'un arbre arbitraire de l'espèce i à son voisin de l'espèce j le plus proche.

Dans le cas général, Y_{11} et Y_{12} ne sont pas des variables aléatoires indépendantes, et les calculs sont alors compliqués. Néanmoins :

- Si les processus marginaux sont indépendants, alors il y a indépendance de Y_{11} et Y_{12} , et :

$$\Pr [\text{ppv}(E) \in e_1 / E \in e_1] = \Pr [Y_{11} < X_2]$$

où X_i est la distance d'un point arbitraire à l'arbre de l'espèce i le plus proche.

- Si le processus marginal de l'espèce 1 est un processus de Poisson homogène, alors il y a indépendance de Y_{11} et Y_{12} , et :

$$\Pr [\text{ppv}(E) \in e_1 / E \in e_1] = \Pr [X_1 < Y_{12}]$$

On obtient des expressions similaires pour B, C et D . Les calculs sont menés jusqu'au bout pour deux exemples en annexe B.

Variante de cette méthode

Dixon [20] propose une méthode dérivée de celle de Pielou pour tester si la répartition observée résulte d'un étiquetage aléatoire. Comme Pielou, il s'appuie sur un tableau de contingence sur les plus proches voisins, et suggère de calculer

$$\begin{aligned} C &= \frac{z_1^2 + z_2^2 - 2rz_1z_2}{1 - r^2} \\ \text{avec } r &= \frac{\text{Cov}(A, D)}{\sqrt{\text{Var } A \cdot \text{Var } D}} \quad z_1 = \frac{A - E(A)}{\sqrt{\text{Var } A}} \quad z_2 = \frac{D - E(D)}{\sqrt{\text{Var } D}} \end{aligned}$$

où A (respectivement D) est le nombre d'individus de l'espèce 1 (respectivement 2) ayant un individu de la même espèce comme plus proche voisin. Lorsque l'étiquetage "espèce 1", "espèce 2" est aléatoire, C suit approximativement un χ^2 à deux degrés de liberté.

Cuzick & Edwards [10] cherchent à tester également si la répartition observée peut résulter d'un étiquetage aléatoire. Comme pour la méthode de Pielou, l'information récoltée est la nature des individus dans le voisinage d'un individu donné. Plus précisément, leur test est basé sur le nombre d'individus de l'espèce 1 parmi les k plus proches voisins d'un individu de base de l'espèce 1, et sur le nombre d'individus de l'espèce 1 qui sont plus proches de cet individu de base que le $k^{\text{ème}}$ plus proche individu de l'espèce 2.

F.2.6. Méthodes de Ripley

Les méthodes de Ripley sont basées sur la fonction K de Ripley. Intuitivement pour un processus ponctuel bivarié homogène, λ_j étant l'intensité du processus N_j , $\lambda_j K_{ij}(r)$ est le nombre moyen d'arbres de l'espèce j situés à une distance inférieure ou égale à r d'un arbre arbitraire de l'espèce i . On a toujours $K_{12} = K_{21}$.

Choix d'un estimateur de la fonction K

Plusieurs estimateurs de K_{ij} ont été proposés [9, 42, 64], la principale difficulté venant des effets de bord. Celui qui a été utilisé dans les simulations est l'estimateur proposé par Ripley :

$$\hat{K}_{ij}(r) = \frac{A}{n_1 n_2} \sum_{i=1..n_1} \sum_{j=1..n_2} \frac{\delta_{ij}(r)}{w(\vec{x}_i, \vec{y}_j)}$$

avec A surface de la zone d'étude
 n_i nombre d'arbres de l'espèce i présents dans la zone d'étude
 \vec{x}_i coordonnées du i^e arbre de l'espèce 1
 \vec{y}_j coordonnées du j^e arbre de l'espèce 2
 $\delta_{ij}(r) = \begin{cases} 1 & \text{si } \|\vec{x}_i - \vec{y}_j\| \leq r \\ 0 & \text{sinon} \end{cases}$
 $w(\vec{a}, \vec{b})$ proportion de la circonférence du cercle centré en \vec{a} et passant par \vec{b} qui se trouve à l'intérieur de la zone d'étude.

Le rayon r varie entre 0 et r_{\max} , où r_{\max} vaut la moitié du plus petit diamètre de la zone d'étude.

Définition de la dissimilarité

Les fonctions K permettent de déceler aussi bien un écart à l'indépendance des processus marginaux qu'un écart à l'étiquetage aléatoire.

Écart à l'indépendance des processus marginaux (méthode 1) [19, 47, 50]

L'indépendance des processus marginaux implique : $K_{12}(r) = K_{21}(r) = \pi r^2$. Toute distance entre les deux fonctions K_{12} (ou K_{21}) et $r \mapsto \pi r^2$ donne donc une mesure de l'écart à l'indépendance. On a toutefois préféré définir la dissimilarité entre les deux répartitions marginales par

$$d(1, 2) = \int_0^{r_{\max}} \left(\sqrt{\frac{\hat{K}_{12}(r)}{\pi}} - r \right) dr$$

qui ne constitue pas une distance entre les fonctions $\hat{L}_{12}(r) = \sqrt{\frac{\hat{K}_{12}(r)}{\pi}}$ et $r \mapsto r$, mais qui a l'avantage de changer de signe selon qu'il y a attraction ou répulsion. En effet, s'il y a attraction entre les processus marginaux à l'échelle $1/r$, $K_{12}(r) = K_{21}(r) > \pi r^2$; s'il y a répulsion entre les processus marginaux à l'échelle $1/r$, $K_{12}(r) = K_{21}(r) < \pi r^2$.

Écart à l'étiquetage aléatoire (méthode 2) [16]

S'il y a étiquetage aléatoire, alors $K_{12}(r) = K_{21}(r) = K_{11}(r) = K_{22}(r)$. Comme l'on s'intéresse à la relation entre les deux processus marginaux (écart de K_{12} ou K_{21} à K_{11} et K_{22}), et non pas à la comparaison des deux processus marginaux (écart de K_{11} à K_{22}), la dissimilarité entre les deux répartitions marginales a alors été définie par :

$$d(1, 2) = \frac{1}{2} \left(\sqrt{\int_0^{r_{\max}} (\hat{L}_{12}(r) - \hat{L}_{11}(r))^2 dr} + \sqrt{\int_0^{r_{\max}} (\hat{L}_{12}(r) - \hat{L}_{22}(r))^2 dr} \right)$$

$$\text{avec } \hat{L}_{ij}(r) = \sqrt{\frac{\hat{K}_{ij}(r)}{\pi}}, \quad i, j = 1, 2$$

F.3. Autres méthodes

D'autres méthodes encore ont été proposées pour étudier les répartitions spatiales de deux espèces.

La méthode des transects notamment [56] se distingue des méthodes évoquées jusqu'à présent par la nature de l'information qui est collectée. Elle consiste à suivre un transect en forêt en notant l'espèce de chaque individu rencontré successivement sur le transect. On obtient donc une suite du type 111 2 1 22 111 etc., où des séquences de 1 alternent avec des séquences de 2. S'il y a étiquetage aléatoire des individus par leur espèce, alors la longueur des séquences de 1 (respectivement 2) suit une loi géométrique de paramètre p (respectivement q), et p (respectivement q) est égal à la proportion d'individus de l'espèce 1 (respectivement 2) dans la population totale.

Des estimations \tilde{p} et \tilde{q} des paramètres de la loi géométrique peuvent être calculées à partir de la suite, et l'écart à la situation d'étiquetage est alors mesuré par $\tilde{p} + \tilde{q} - 1$.

D'autres pistes, qui n'ont pas encore été explorées, pourraient être aussi parcourues :

- une approche de type analyse d'image pourrait être tentée. On pourrait par exemple construire les polygones de Voronoï en affectant une couleur à chaque espèce, et chercher à analyser la mosaïque bicolore ainsi obtenue.
- la méthode du "petit chaperon rouge" est une méthode par distances qui consiste à parcourir la forêt d'arbre en arbre, en se fixant des règles, puis à étudier la suite des distances ainsi obtenue.

G. RÉSULTATS POUR LA FORÊT DE VANCOUVER

G.1. Etude des répartitions univariées : tests de la répartition aléatoire

La répartition est considérée comme une réalisation d'un processus ponctuel N . L'hypothèse nulle est \mathcal{H}_0 : " N est un processus de Poisson homogène d'intensité λ ". Les tests sont :

1. Test de Thompson [61]: Soit $X_{(k)}$ la distance d'un point d'échantillonnage au $k^{\text{ème}}$ plus proche arbre, et $(X_{(k)i})_{i=1..n}$ un échantillon de n distances indépendantes ; sous \mathcal{H}_0 , $2\lambda\pi \sum_{i=1}^n X_{(k)i}^2$ est distribué comme un χ^2 à $2nk$ degrés de liberté.
2. Test de Holgate [33]: Soit $X_{(s)}$ et $X_{(t)}$ ($s < t$) les distances d'un point d'échantillonnage aux $s^{\text{ème}}$ et $t^{\text{ème}}$ plus proches arbres, et $(X_{(s)i}, X_{(t)i})_{i=1..n}$ un échantillon de n couples de distances indépendants ; sous \mathcal{H}_0 , $(\sum_{i=1}^n X_{(s)i}^2) / (\sum_{i=1}^n X_{(t)i}^2)$ est distribué comme une variable β à $2ns, 2n(t-s)$ degrés de liberté.
3. Test de Hines & Hines [32]: Soit $X_{(1)}$ la distance d'un point d'échantillonnage au plus proche arbre, et $(X_{(1)i})_{i=1..n}$ un échantillon de n distances indépendantes ; sous \mathcal{H}_0 , $A = n (\sum_{i=1}^n X_{(1)i}^2) / (\sum_{i=1}^n X_{(1)i})^2$ a une distribution indépendante de λ , dont les valeurs critiques sont tabulées.
4. Test du χ^2 : Soit n quadrats indépendants de surface S placés sur la zone d'étude ; soit $\{[i_0 = 0, i_1 - 1], [i_1, i_2 - 1] \dots [i_{I-1}, i_I - 1], [i_I, i_{I+1} = +\infty]\}$ une partition de N en $(I + 1)$ classes, et soit N_k ($k = 0 \dots I$) le nombre de quadrats qui contiennent entre i_k et $i_{k+1} - 1$ arbres ($\sum_{k=0}^I N_k = n$). Soit $p_k = \sum_{m=i_k}^{i_{k+1}-1} e^{-\lambda S} \frac{(\lambda S)^m}{m!}$, $\forall k \in \{0, 1, \dots, I\}$. Sous \mathcal{H}_0 , $\sum_{i=0}^I (N_i - np_i)^2 / np_i$ est distribué asymptotiquement (lorsque $n \mapsto \infty$) comme un χ^2 à $I - 1$ degrés de liberté.

Pratiquement, la partition de N est choisie de sorte que $N_k \geq 5$, $\forall k$ et que I soit aussi grand que possible.

Les résultats des test sont donnés dans les tableaux suivants :

Test de Thompson Les nombres représentent $\Pr [X^2 \leq A]$, où X^2 est une variable aléatoire distribuée comme un χ^2 à $200k$ degrés de liberté, et A est la valeur de la statistique calculée à partir des distances observées.

| | | | | |
|------------|-----------|--------|--------|--------|
| Parcelle 1 | k=1 | k=2 | k=3 | k=4 |
| | Ab 1.000* | 1.000* | 1.000* | 1.000* |
| | Hw 0.948 | 0.686 | 0.884 | 0.965* |
| Parcelle 2 | k=1 | k=2 | k=3 | k=4 |
| | Ab 1.000* | 1.000* | 1.000* | 1.000* |
| | Hw 0.777 | 0.538 | 0.791 | 0.872 |
| Parcelle 3 | k=1 | k=2 | k=3 | k=4 |
| | Ab 1.000* | 1.000* | 1.000* | 1.000* |
| | Cw 0.797 | 0.955* | 0.988* | 1.000* |
| | Hw 0.948 | 0.995* | 0.996* | 1.000* |

| | | | | | |
|------------|-----|--------|--------|--------|--------|
| Parcelle 4 | k=1 | k=2 | k=3 | k=4 | |
| | Cw | 0.997* | 0.864 | 0.908 | 0.945 |
| | Hw | 0.900 | 0.802 | 0.691 | 0.770 |
| Parcelle 5 | k=1 | k=2 | k=3 | k=4 | |
| | Ab | 1.000* | 1.000* | 1.000* | 1.000* |
| | Hw | 0.980* | 1.000* | 0.999* | 1.000* |
| Parcelle 6 | k=1 | k=2 | k=3 | k=4 | |
| | Cw | 1.000* | 1.000* | 1.000* | 1.000* |
| | Hw | 0.999* | 1.000* | 1.000* | 1.000* |
| Parcelle 7 | k=1 | k=2 | k=3 | k=4 | |
| | Cw | 0.994* | 0.998* | 1.000* | 1.000* |
| | Hw | 0.992* | 1.000* | 0.999* | 1.000* |
| Parcelle 8 | k=1 | k=2 | k=3 | k=4 | |
| | Ab | 1.000* | 1.000* | 1.000* | 1.000* |
| | Hw | 0.677 | 0.759 | 0.634 | 0.926* |

* : \mathcal{H}_0 est rejeté à 5% pour une répartition agrégée.

Test de Holgate Les nombres représentent $\Pr [B \leq A]$, où B est une variable aléatoire qui suit une loi β à $200s, 200(t - s)$ degrés de liberté, et A est la valeur de la statistique calculée à partir des distances observées.

| | | | | | | | | |
|------------|-------|---------|---------|---------|-----|-------|-------|---------|
| Parcelle 1 | Ab | s=1 | s=2 | s=3 | Hw | s=1 | s=2 | s=3 |
| | t=2 | 0.084 | | | t=2 | 0.969 | | |
| | t=3 | 0.159 | 0.660 | | t=3 | 0.871 | 0.201 | |
| | t=4 | 0.073 | 0.279 | 0.059 | t=4 | 0.792 | 0.139 | 0.232 |
| Parcelle 2 | Ab | s=1 | s=2 | s=3 | Hw | s=1 | s=2 | s=3 |
| | t=2 | 0.118 | | | t=2 | 0.837 | | |
| | t=3 | 0.344 | 0.912 | | t=3 | 0.636 | 0.167 | |
| | t=4 | 0.171 | 0.512 | 0.044** | t=4 | 0.585 | 0.163 | 0.357 |
| Parcelle 3 | Ab | s=1 | s=2 | s=3 | Hw | s=1 | s=2 | s=3 |
| | t=2 | 1.000* | | | t=2 | 0.383 | | |
| | t=3 | 1.000* | 1.000* | | t=3 | 0.546 | 0.767 | |
| | t=4 | 1.000* | 1.000* | 1.000* | t=4 | 0.283 | 0.272 | 0.024** |
| | Cd | s=1 | s=2 | s=3 | | | | |
| | t=2 | 0.302 | | | | | | |
| t=3 | 0.284 | 0.390 | | | | | | |
| t=4 | 0.084 | 0.046** | 0.007** | | | | | |
| Parcelle 4 | Cd | s=1 | s=2 | s=3 | Hw | s=1 | s=2 | s=3 |
| | t=2 | 0.998* | | | t=2 | 0.827 | | |
| | t=3 | 0.992* | 0.501 | | t=3 | 0.887 | 0.773 | |
| | t=4 | 0.986* | 0.475 | 0.441 | t=4 | 0.851 | 0.672 | 0.381 |
| Parcelle 5 | Ab | s=1 | s=2 | s=3 | Hw | s=1 | s=2 | s=3 |
| | t=2 | 0.699 | | | t=2 | 0.283 | | |
| | t=3 | 0.980* | 1.000* | | t=3 | 0.611 | 0.948 | |
| | t=4 | 0.914 | 0.944 | 0.065 | t=4 | 0.311 | 0.472 | 0.015** |
| Parcelle 6 | Cd | s=1 | s=2 | s=3 | Hw | s=1 | s=2 | s=3 |
| | t=2 | 0.420 | | | t=2 | 0.828 | | |
| | t=3 | 0.435 | 0.493 | | t=3 | 0.841 | 0.625 | |
| | t=4 | 0.492 | 0.588 | 0.645 | t=4 | 0.804 | 0.550 | 0.395 |
| Parcelle 7 | Cd | s=1 | s=2 | s=3 | Hw | s=1 | s=2 | s=3 |
| | t=2 | 0.759 | | | t=2 | 0.471 | | |
| | t=3 | 0.770 | 0.588 | | t=3 | 0.715 | 0.895 | |
| | t=4 | 0.559 | 0.232 | 0.059 | t=4 | 0.416 | 0.433 | 0.028** |

| Parcelle 8 | Ab | s=1 | s=2 | s=3 | Hw | s=1 | s=2 | s=3 |
|------------|-----|--------|--------|--------|-----|-------|-------|---------|
| | t=2 | 0.929 | | | t=2 | 0.470 | | |
| | t=3 | 0.990* | 0.978* | | t=3 | 0.620 | 0.764 | |
| | t=4 | 0.998* | 0.998* | 0.981* | t=4 | 0.380 | 0.325 | 0.042** |

* : \mathcal{H}_0 est rejeté à 5% pour une répartition agrégée.

** : \mathcal{H}_0 est rejeté à 5% pour une répartition régulière.

Test de Hines & Hines Les nombres sont les valeurs de la statistique A calculées à partir des distances observées.

| Parcelle | Ab | Cd | Hw |
|----------|--------|---------|---------|
| 1 | 1.481* | - | 1.154** |
| 2 | 1.418* | - | 1.160** |
| 3 | 1.652* | 1.485* | 1.375* |
| 4 | - | 1.253 | 1.144** |
| 5 | 1.562* | - | 1.322 |
| 6 | - | 1.467* | 1.382* |
| 7 | - | 1.201** | 1.264 |
| 8 | 1.428* | - | 1.271 |

* : \mathcal{H}_0 est rejeté à 5% pour une répartition agrégée.

** : \mathcal{H}_0 est rejeté à 5% pour une répartition régulière.

Test du χ^2 Les nombres représentent $\Pr[X^2 \leq A]$, où X^2 est une variable distribuée comme un χ^2 , et A est la valeur de la statistique calculée à partir des comptages observés.

| Parcelle | Ab | Cd | Hw |
|----------|--------|--------|--------|
| 1 | 0.000* | - | 0.222 |
| 2 | 0.035* | - | 0.931 |
| 3 | 0.000* | 0.514 | 0.000* |
| 4 | - | 0.125 | 0.023* |
| 5 | 0.013* | - | 0.013* |
| 6 | - | 0.000* | 0.235 |
| 7 | - | 0.000* | 0.001* |
| 8 | 0.000* | - | 0.000* |

* : \mathcal{H}_0 est rejeté à 5%.

G.2. Etude des répartition bivariables

G.2.1. Résultats des tests de l'indépendance

1. Test basé sur les quadrats : les nombres sont $\Pr[X^2 \leq A]$, où X^2 est une variable aléatoire distribuée comme un χ^2 à un degré de liberté, et A est la valeur de la statistique calculée à partir des comptages observés. Le tableau contient aussi le coefficient de corrélation de

points ρ .

| Parcelle | Pr $X^2 \leq A$ | ρ |
|-------------|-----------------|--------|
| 1 | 0.669 | -0.121 |
| 2 | 0.883 | -0.189 |
| 3 (Ab / Cw) | 0.860 | -0.169 |
| 3 (Ab / Hw) | 0.103 | 0.009 |
| 3 (Cw / Hw) | 0.606 | -0.107 |
| 4 | 0.562 | -0.100 |
| 5 | 0.961* | -0.228 |
| 6 | 0.982* | -0.260 |
| 7 | 0.996* | -0.311 |
| 8 | 0.999* | -0.363 |

* : \mathcal{H}_0 est rejeté à 5%. Comme $\rho < 0$, il y a association négative entre les deux répartitions marginales.

2. Test de Diggle & Cox : les nombres sont $\Pr [T > |t|]$, où T est une variable aléatoire distribuée comme le coefficient de corrélation τ de Kendall pour deux échantillons indépendants, et t est la valeur de τ calculée à partir des distances observées. Le tableau contient aussi t .

| Parcelle 1 | Pr $[T > t]$ | t | Parcelle 4 | Pr $[T > t]$ | t |
|------------|----------------|--------|------------|----------------|--------|
| k=1 | 0.335 | -0.065 | k=1 | 0.501 | -0.046 |
| k=2 | 0.854 | -0.013 | k=2 | 0.106 | -0.109 |
| k=3 | 0.986 | -0.001 | k=3 | 0.162 | -0.095 |
| k=4 | 0.520 | -0.043 | k=4 | 0.001* | -0.229 |

| Parcelle 2 | Pr $[T > t]$ | t | Parcelle 5 | Pr $[T > t]$ | t |
|------------|----------------|--------|------------|----------------|--------|
| k=1 | 0.394 | -0.058 | k=1 | 0.003* | -0.205 |
| k=2 | 0.821 | -0.015 | k=2 | 0.001* | -0.232 |
| k=3 | 0.388 | -0.059 | k=3 | 0.000* | -0.274 |
| k=4 | 0.224 | -0.082 | k=4 | 0.000* | -0.267 |

| Parcelle 3 (Ab / Cw) | Pr $[T > t]$ | t | Parcelle 6 | Pr $[T > t]$ | t |
|----------------------|----------------|--------|------------|----------------|--------|
| k=1 | 0.055 | -0.130 | k=1 | 0.053 | -0.131 |
| k=2 | 0.075 | -0.121 | k=2 | 0.049* | -0.134 |
| k=3 | 0.026* | -0.151 | k=3 | 0.027* | -0.149 |
| k=4 | 0.131 | -0.102 | k=4 | 0.064 | -0.126 |

| Parcelle 3 (Ab / Hw) | Pr $[T > t]$ | t | Parcelle 7 | Pr $[T > t]$ | t |
|----------------------|----------------|--------|------------|----------------|--------|
| k=1 | 0.849 | 0.013 | k=1 | 0.006* | -0.185 |
| k=2 | 0.981 | 0.002 | k=2 | 0.020* | -0.158 |
| k=3 | 0.770 | -0.020 | k=3 | 0.001* | -0.219 |
| k=4 | 0.844 | -0.013 | k=4 | 0.000* | -0.297 |

| Parcelle 3 (Cw / Hw) | Pr $[T > t]$ | t | Parcelle 8 | Pr $[T > t]$ | t |
|----------------------|----------------|--------|------------|----------------|--------|
| k=1 | 0.138 | -0.101 | k=1 | 0.001* | -0.218 |
| k=2 | 0.155 | -0.097 | k=2 | 0.011* | -0.172 |
| k=3 | 0.872 | -0.011 | k=3 | 0.000* | -0.274 |
| k=4 | 0.761 | -0.021 | k=4 | 0.000* | -0.334 |

* : \mathcal{H}_0 est rejeté à 5%. Comme $t < 0$, il y a association négative entre les deux répartitions marginales.

G.2.2. Valeurs des indices de dissimilarité

| Parcelle 1 | ρ | τ | d |
|-------------|--------|--------|-------|
| 1 | -0.121 | -0.031 | 1.356 |
| 2 | -0.189 | -0.046 | 1.245 |
| 3 (Ab / Cw) | -0.169 | -0.126 | 1.626 |
| 3 (Ab / Hw) | +0.009 | -0.005 | 1.534 |
| 3 (Cw / Hw) | -0.107 | -0.057 | 1.147 |
| 4 | -0.100 | -0.120 | 1.087 |
| 5 | -0.228 | -0.245 | 1.293 |
| 6 | -0.260 | -0.135 | 1.160 |
| 7 | -0.311 | -0.215 | 1.167 |
| 8 | -0.363 | -0.249 | 1.375 |

ρ est l'indice des quadrats, τ est l'indice de Diggle & Cox, d est l'indice des plus proches voisins.

BIBLIOGRAPHIE

- [1] AGRESTI, A. (1984) *Analysis of Ordinal Categorical Data*. John Wiley & Sons, New York, 287 p.
- [2] BARITEAU, M. (1992) Régénération naturelle de la forêt tropicale humide de Guyane : étude de la répartition spatiale de *Qualea rosea* Aublet, *Eperua falcata* Aublet et *Symphonia globulifera* Linnaeus f. *Annales des Sciences Forestières* **49**, 359-382
- [3] BARTHES, B. (1988) Où pousse cet arbre ? Premiers résultats à l'étude des relations sol-végétation en forêt de Paracou (Guyane), cas du wapa et du gonfolo. Rapport interne Orstom-CTFT, 33p.
- [4] BESAG, J. (1977) Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute* **47**, 77-92
- [5] CAILLIEZ, F. & PAGES, J. P. (1976) *Introduction à l'analyse de données*. SMASH, 615 p.
- [6] CARTER, R. E. & KLINKA, K. (1992) Variation in shade tolerance of Douglas fir, western hemlock, and western red cedar in coastal British Columbia. *Forest Ecology and Management* **55**, 87-105
- [7] CASSIE, R. M. (1962) Frequency distribution models in the ecology of plankton and other organisms. *Journal of Animal Ecology* **31**, 65-91
- [8] CHANDON, J. L. & PINSON, S. (1981) *Analyse Typologique, Théories et Applications*. Masson, Paris, 254 p.
- [9] CRESSIE, N. (1991) *Statistics for Spatial Data*. John Wiley & Sons, New York, 900 p.
- [10] CUZICK, J. & EDWARDS, R. (1990) Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society, Series B* **52**, 73-104
- [11] DALE, M. R. T. & BLUNDON, D. J. (1991) Quadrat covariance analysis and the scales of interspecific association during primary succession. *Journal of Vegetation Science* **2**, 103-112
- [12] DALE, M. R. T. & POWELL, R. D. (1994) Scales of segregation and aggregation of plants of different kinds. *Canadian Journal of Botany* **72**, n°4, 448-53
- [13] DIGGLE, P. J. (1977) The detection of random heterogeneity in plant populations. *Biometrics* **33**, 390-394
- [14] DIGGLE, P. J. (1983) *Statistical Analysis of Spatial Point Patterns*. Academic Press, London, 148 p.
- [15] DIGGLE, P. J., BESAG, J. & GLEAVES, J. T. (1976) Statistical analysis of spatial point patterns by means of distance methods. *Biometrics* **32**, 659-667
- [16] DIGGLE, P. J. & CHETWYND, A. G. (1991) Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics* **47**, 1156-63

- [17] DIGGLE, P. J. & COX, T. F. (1983) Some distance-based tests of independence for sparsely-sampled multivariate spatial point pattern. *International Statistical Review* **51**, 11-23
- [18] DIGGLE, P. J., FIKSEL, T., GRABARNIK, P., OGATA, Y., STOYAN, D. & TANEMURA, M. (1994) On parameter-estimation for pairwise interaction point-processes. *International Statistical Review* **62**, n°1, 99-117
- [19] DIGGLE, P. J. & MILNE, R. K. (1983) Bivariate Cox processes : some models for bivariate spatial point patterns. *Journal of the Royal Statistical Society, Series B* **45**, 11-21
- [20] DIXON, P. (1994) Testing spatial segregation using a nearest-neighbor contingency table. *Ecology* **75**, n°7, 1940-8
- [21] GALIANO, E. F. (1982) Pattern detection in plant populations through the analysis of plant-to-all plants distances. *Vegetatio* **49**, 39-43
- [22] GOODALL, D. W. (1965) Plot-less tests of interspecific association. *Journal of Ecology* **53**, 197-210
- [23] GOODALL, D. W. (1974) A new method for the analysis of spatial pattern by random pairing of quadrats. *Vegetatio* **29**, 135-46
- [24] GREGORY, R. A. (1957) Some silvicultural characteristics of western redcedar in Southeast Alaska. *Ecology* **38**, n°4, 646-649
- [25] GREIG-SMITH, P. (1952) The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany* **16**, 293-316
- [26] GREIG-SMITH, P. (1964) *Quantitative Plant Ecology (2nd edition)*. Butterworths, London, 255 p.
- [27] GREIG-SMITH, P. (1979) Pattern in vegetation. *Journal of Ecology* **67**, 755-779
- [28] HABECK, J. R. (1968) Forest succession in the Glacier Park cedar-hemlock forests. *Ecology* **49**, n°5, 872-880
- [29] HANISCH, K. H. & STOYAN, D. (1979) Formulas for the second-order analysis of marked point processes. *Mathematische Operationsforschung und Statistik, Series Statistics* **10**, 555-560
- [30] HARKNESS, R. D. & ISHAM, V. (1983) A bivariate spatial point pattern of ants' nests. *Applied Statistics* **32**, n°3, 293-303
- [31] HILL, M. O. (1973) The intensity of spatial pattern in plant communities. *Journal of Ecology* **61**, 225-35
- [32] HINES, W. G. S. & HINES, R. J. O'HARA (1979) The Eberhardt statistic, and the detection of nonrandomness of spatial point distribution. *Biometrika* **66**, 73-79
- [33] HOLGATE, P. (1965) Some new tests of randomness. *Journal of Ecology* **53**, 261-6
- [34] JOBSON, J. D. (1992) *Applied Multivariate Data Analysis. Vol. II : Categorical and Multivariate Methods*. Springer-Verlag, New York, 730 p.
- [35] KENDALL, M. (1975) *Multivariate Analysis*. Charles Griffin & Co., 210 p.
- [36] KERSHAW, K. A. (1957) The use of cover and frequency in the detection of pattern in plant communities. *Ecology* **38**, 291-99
- [37] KERSHAW, K. A. (1959) An investigation of the structure of a grassland community. II. III. *Journal of Ecology* **47**, 31-53

- [38] KERSHAW, K. A. (1960) The detection of pattern and association. *Journal of Ecology* **48**, 233-242
- [39] KRAJINA, V. J. (1965) Biogeoclimatic zones in British Columbia. *Ecology of Western North America* **1**, 1-17
- [40] KU, H. H., VARNER, R. N., KULLBACK, S. (1971) On the analysis of multidimensional contingency tables. *Journal of the American Statistical Association* **66**, n°333, 55-64
- [41] LESCURE, J. P. & BOULET, R. (1985) Relationships between soil and vegetation in a tropical rain forest in French Guiana. *Biotropica* **17**, n°2, 155-164
- [42] LOTWICK, H. W. & SILVERMAN, B. W. (1982) Methods of analysing spatial processes of several types of points. *Journal of the Royal Statistical Society, Series B* **44**, 406-13
- [43] LUDWIG, J. A. & GOODALL, D. W. (1978) A comparison of paired with blocked-quadrat variance methods for the analysis of spatial pattern. *Vegetatio* **38**, 49-59
- [44] MEAD, R. (1974) A test for spatial pattern at several scales using data from a grid of contiguous quadrats. *Biometrics* **30**, 295-307
- [45] MEAGHER, T. R. & BURDICK, D. S. (1980) The use of nearest neighbor frequency analyses in studies of association. *Ecology* **61**, 1253-55
- [46] MESSIER, C. (1993) Factors limiting early growth of western redcedar, western hemlock and Sitka spruce seedlings on ericaceous-dominated clearcut sites in coastal British Columbia. *Forest Ecology and Management* **60**, 181-206
- [47] MOEUR, M. Characterizing spatial patterns of trees using stem-mapped data. *Forest Science* **39**, n°4, 756-775
- [48] MUNGER, T. T. (1940) The cycle from Douglas fir to hemlock. *Ecology* **21**, n°4, 451-459
- [49] PATIL, G. P. & STITELER, W. (1974) Concepts of aggregation and their quantification : a critical review with some new results and applications. *Researches on Population Ecology* **4**, 1-7
- [50] PENTTINEN, A., STOYAN, D. & HENTTONEN, H. M. (1992) Marked point-processes in forest statistics. *Forest Science* **38**, n°4, 806-824
- [51] PERSSON, O. (1964) Distance methods. *Studia Forestalia Suecica* **15**, 1-68
- [52] PETERSON, C. H. (1976) Measurements of community pattern by indices of local segregation and species diversity. *Journal of Ecology* **64**, 157-169
- [53] PIELOU, E. C. (1960) A single mechanism to account for regular, random & aggregated populations. *Journal of Ecology* **48**, 575-84
- [54] PIELOU, E. C. (1961) Segregation and symmetry in two-species populations as studied by nearest-neighbour relationships. *Journal of Ecology* **49**, 255-69
- [55] PIELOU, E. C. (1962) The use of plant-to-neighbor distances for the detection of competition. *Journal of Ecology* **50**, 357-367
- [56] PIELOU, E. C. (1962) Runs of one species with respect to another in transects through plant populations. *Biometrics* **18**, 579-93
- [57] PIELOU, E. C. (1969) *An Introduction to Mathematical Ecology*. John Wiley & Sons, New York, 290 p.

102

- [58] SABATIER, D. & PREVOST (1990) Variations du peuplement forestier à l'échelle stationnelle : le cas de la station des Nouragues en Guyane française. Actes de l'atelier MAB, IUFRO, Cayenne
- [59] STEEGE, H. ter, JETTEN, V., POLAK, M. & WERGER, M. (1993) Tropical rain forest types and soils of a watershed in Guyana, South America. *Tropenbos Serie 3*, 47-72
- [60] STOYAN, D. & STOYAN, H. (1994) *Fractals, Random Shapes and Point Fields*. John Wiley & Sons, Chichester, 390 p.
- [61] THOMPSON, H. R. (1956) Distribution of distance to n^{th} nearest neighbour in a population of randomly distributed individuals. *Ecology* **37**, 391-4
- [62] THOMPSON, H. R. (1958) The statistical study of plant distribution patterns using a grid of quadrats. *Australian Journal of Botany* **6**, 322-43
- [63] TURNER, D. P. (1985) Successional relationships and a comparison of biological characteristics among six northwestern conifers. *Bulletin of the Torrey Botanical Club* **112**, n°4, 421-428
- [64] UPTON, G. & FINGLETON, B. (1985) *Spatial data analysis by example. Vol. I. Point pattern and quantitative data*. John Wiley & Sons, Chichester, 410 p.
- [65] USHER, M. B. (1969) The relation between mean square and block size in the analysis of similar patterns. *Journal of Ecology* **57**, 505-14
- [66] ZOBEL, D. B. & ANTOS, J. A. (1991) Growth and development of natural seedlings of *Abies* and *Tsuga* in old-growth forest. *Journal of Ecology* **79**, 985-998