

# Exercice d'analyse spatiale

## Bio 6075z

Pierre Legendre  
Avril 1992 / février 1994 / octobre 1996

L'exercice qui vous est proposé porte sur des données récoltées le 25 octobre 1988 dans l'étang de Thau. Ces données vous permettront de calculer des corrélogrammes (Moran et Geary), de réaliser des tests de Mantel simples et partiels et de calculer un corrélogramme de Mantel. Les données sont décrites succinctement au paragraphe suivant.

Dans le cadre d'un programme de recherche écologique portant sur un étang saumâtre du sud de la France (Amanieu *et al.*, 1989), un échantillonnage a été réalisé le 25 octobre 1988 à 20 stations d'échantillonnage (profondeur: 0,5 m). Ces stations furent sélectionnées, dans l'étang, à l'aide d'un raisonnement décrit par Legendre *et al.* (1989). Certaines des variables sont présentées au tableau suivant. Elles se trouvent également dans un fichier contenant 8 variables (séparées par des tabulations) disponible sur le site Web du cours Bio 6077.

Station no	Phytoplancton <sup>1</sup> ln(x + 1)	NH <sub>4</sub> <sup>2</sup> (ln)	Phaeopigments (ln)	ADN <sup>3</sup> (ln)	Coordonnées <sup>4</sup> X (km) Y (km)	
1	0.31	0.307	0.184	0.274	3	10
2	-0.31	0.207	0.212	0.213	5	9
3	-0.38	0.140	0.229	0.134	6	8
4	-0.89	1.371	0.287	0.177	6	10
5	-1.09	1.447	0.242	0.091	8	9
6	-0.43	0.668	0.531	0.272	9	10
7	0.49	0.300	0.948	0.460	10	7
8	0.89	0.329	1.389	0.253	11	6
9	0.41	0.207	0.765	0.235	12	5
10	0.23	0.223	0.737	0.362	12	7
11	-0.23	0.788	0.454	0.824	12	9
12	-0.06	1.112	0.395	0.419	13	8
13	-0.51	1.273	0.247	0.398	15	2
14	0.47	0.956	0.449	0.172	15	4
15	0.34	0.708	0.457	0.141	15	5
16	0.10	0.637	0.386	0.360	16	1
17	0.23	0.519	0.481	0.261	16	2
18	0.48	0.247	0.468	0.450	16	4
19	-0.11	1.664	0.321	0.287	17	6
20	0.06	0.182	0.380	0.510	18	4

<sup>1</sup> La quantité de phytoplancton est estimée par la densité de chlorophylle *a*, exprimée en mg/L. Cette variable a été normalisée à l'aide d'une transformation logarithmique [ $x' = \ln(x + 1)$ ].

<sup>2</sup> La variable NH<sub>4</sub>, exprimée en µmol/L, a été normalisée par une transformation log [ $x' = \ln(x)$ ].

<sup>3</sup> L'ADN bactérien fut mesuré par l'incorporation de thymidine tritiée dans les cellules en culture.

<sup>4</sup> La position géographique des stations d'échantillonnage est exprimée par leurs coordonnées sur un plan; les axes X et Y sont mesurés en km.

## Données

Vous pouvez réaliser l'exercice, soit à partir du tableau ci-dessus, soit à partir des fichiers de données correspondants qui sont à votre disposition dans le présent dossier. Un fichier unique contenant toutes les données est disponible sur le site Web du cours Bio 6077.

- 1) BNA,MA (20x2): Fichier ASCII contenant deux variables bactériologiques (BNA = bactéries hétérotrophes aérobies croissant sur milieu Nutrient Agar; MA = bactéries hétérotrophes aérobies croissant sur milieu Marine Agar).
- 2) Chla (20x1): Fichier ASCII contenant la variable Chlorophylle a, transformée par élimination de l'effet de la circulation (courants) (Legendre & Troussellier, 1993).
- 3) Env (20x3): Fichier ASCII contenant trois variables environnementales: NH<sub>4</sub>, Phéopigments, ADN(F).
- 4) Thau XY (20x2): Fichier ASCII contenant les coordonnées géographiques (X, Y) des 20 stations d'échantillonnage.

Préparez tout d'abord les matrices de distance suivantes à l'aide de SIMIL; si vous utilisez une mesure de similarité plutôt qu'une distance, transformez ensuite votre matrice de ressemblance en distances à l'aide de CONVERSION:

- 5) Esp/D1: Matrice de distances euclidiennes [ou toute autre distance de votre choix] entre les stations, calculée à partir des données d'espèces (BNA et MA) de la matrice (1).
- 6) Env/D1: Matrice de distances euclidiennes entre les stations, à partir des données environnementales de la matrice (3).
- 7) Spatial/D1: Matrice de distances géographiques (euclidiennes) entre les stations, à partir des coordonnées géographiques de la matrice (4).

## Programme Autocorrélation de Rv3; module Autocor de Rv4

La variable à analyser se trouve dans la matrice (2) Chla (20x1); les données sont quantitatives et non nominales. Les relations de distance entre les points sont fournies par la matrice SIMIL de distances géographiques (7) que vous aurez calculée; celle-ci vous permettra de calculer vos classes de distance. Vous aurez à choisir entre des classes équidistantes ou équi-fréquentes, et à décider du nombre de classes le plus approprié; rappelez-vous que pour un ensemble de  $n$  points, il y a  $n(n-1)/2$  distances dans la demi-matrice.

Dès le démarrage du programme, vous aurez à fournir un nom de fichier qui contiendra les résultats de vos calculs: division en classes et corrélogrammes.

Demandez au programme de conserver la matrice de CLASSES DE DISTANCE; cette matrice constituera votre fichier (8). Vous en aurez besoin pour calculer le corrélogramme de Mantel.

## Test de Mantel simple

Le programme MANTEL de Rv3 (ou le module Mantel de Rv4) vous permettra de comparer les distances dans les matrices (5) et (6), (5) et (7) ou (6) et (7). Les deux matrices jouent le même rôle dans cette analyse et peuvent donc, indifféremment, être déclarées comme matrice A ou B. Vous demanderez donc “le test de Mantel entre deux matrices”, déclarant que la matrice de distances est “une matrice de type SIMIL”. Puisqu’il s’agit de deux petites matrices, vous pouvez demander 499 ou 999 permutations pour ce problème, même sur un ordinateur peu puissant.

[ Ne pas utiliser ce programme sous le Système 7 de Macintosh; les messages demandant les noms de matrices ne s’afficheraient pas. Pour les tests de Mantel simples, le programme demande d’abord la matrice B, puis la matrice A. Pour les tests de Mantel partiels, le programme demande d’abord B, puis C, et enfin A.]

## Test de Mantel partiel (programme **Mantel** de Rv3 ou module **Mantel** de Rv4)

Dans ce calcul, il importe de bien déclarer quelle matrice joue le rôle de matrice A, B et C, car le calcul de la statistique de Mantel partiel (méthode de Smouse, Long & Sokal) réalise le calcul  $A \cdot B|C$ , c’est-à-dire la corrélation partielle entre A et B après avoir éliminé l’effet de C. Vous pourrez utiliser pour ce calcul les trois matrices de distance que vous avez calculées, (5), (6) et (7), dans l’ordre que vous jugerez opportun.

## Corrélogramme de Mantel (programme **Mantel** de Rv3 ou module **Mantel** de Rv4)

Ici encore, l’ordre des matrices est d’importance capitale. Après avoir demandé un “test de Mantel entre deux matrices”, vous déclarerez que la matrice A est une matrice “de classes de distance (pour corrélogramme)”; il s’agira de la matrice de classes de distance (8) que vous aurez obtenue lors du calcul de vos corrélogrammes de Mantel et de Geary. La matrice B sera l’une ou l’autre des autres matrices de distance que vous avez calculées: (5), (6) ou (7).

N’oubliez pas que le nombre de permutations que vous demanderez sera réalisé pour chacune des classes de distance; sur un ordinateur de faible puissance, il serait prudent de ne demander que 99 permutations, lors d’un premier essai. Les machines modernes permettent de réaliser le test de Mantel avec 999 ou 9999 permutations.

Le corrélogramme est inscrit dans un fichier dont vous aurez fourni le nom. Comme le programme vous le signale, le signe des coefficients d’autocorrélation est inversé si votre matrice B est une matrice de Distances; autrement dit, un signe négatif indique une autocorrélation positive, et vice-versa. Par contre, si votre matrice B est de type Similarités, alors un signe positif indique de l’autocorrélation positive. Pouvez-vous dire pourquoi?

## Références

Amanieu, M., P. Legendre, M. Troussellier & G.-F. Frisoni. 1989. Le programme Écothau: théorie écologique et base de la modélisation. *Oceanologica Acta* 12: 189-199.

Legendre, P. & M. Troussellier. 1993. Origin of spatial structures in aquatic bacterial communities: From hypotheses to numerical solutions. Pp. 353-358 *in*: Guerrero, R. & C. Pedrós-Alió (eds.) *Trends in Microbial Ecology*. Proc. 6th Intern. Symp. Microbial Ecology (ISME-6), Barcelona, 6-11 September 1992. Spanish Society for Microbiology.

Legendre, P., M. Troussellier, V. Jarry & M.-J. Fortin. 1989. Design for simultaneous sampling of ecological variables: from concepts to numerical solutions. *Oikos* 55: 30-42.