

1 - Structure spatio-temporelle des variables en
océanographie. Problèmes d'analyse numérique et méthodes
d'analyse spatiale.

Pierre Legendre

Département des Sciences Biologiques, Université de Montréal,
C.P. 6128, Succursale "A", Montréal, Québec H3C 3J7, Canada.

Louis Legendre

Département de Biologie, Université Laval, Québec, Québec
G1K 7P4, Canada.

Le texte ci-après est un résumé de la Conférence Plénière présentée
par les auteurs lors de la Journée "Biométrie et Océanographie" (Journées
de Statistique, Université de Lille, mai 1986).

AUTOCORRÉLATION SPATIO-TEMPORELLE DES VARIABLES OCÉANOGRAPHIQUES

La presque totalité des variables échantillonnées en océanographie
présentent une forte hétérogénéité dans leur répartition spatio-temporelle,
laquelle est une conséquence de l'hétérogénéité spatio-temporelle des
apports énergétiques dans le milieu marin. Celle-ci entraîne, au sein de la
masse océanique, des mouvements de convection (sur la verticale) ainsi que
d'advection horizontale des eaux, qui conduisent à la formation de discon-
tinuités spatio-temporelles, souvent qualifiées d'interfaces.

La structuration spatio-temporelle du milieu physique conduit à
une organisation, tant spatiale que temporelle, des êtres vivants ainsi que
des processus biologiques. Les zones d'interfaces sont le siège d'une forte
activité biologique (LEGENDRE *et al.*, 1986). Même en dehors de ces zones, on
observe souvent une répartition spatiale contagieuse des organismes marins,
sous forme de taches (e.g. phytoplancton), de nuages (zooplancton), de
bancs (e.g. poissons) ou de troupeaux (e.g. cétacés). Dans le domaine tem-
porel, de telles répartitions contagieuses existent aussi, par exemple les
poussées (blooms) de phytoplancton ou encore les explosions d'organismes
animaux (e.g. méduses). Il s'ensuit que la plupart, sinon la totalité des
variables océanographiques font montre d'une forte autocorrélation spatio-
temporelle. Il y a autocorrélation lorsqu'une valeur mesurée en un certain
point ou/et à un certain moment permet de prévoir approximativement les
valeurs de la même variable en d'autres points de l'espace et/ou du temps,
la forme de la fonction d'autocorrélation (par exemple, un gradient) étant
connue. Cette définition s'applique tout aussi bien à l'autocorrélation
positive que négative.

Les phénomènes naturels étudiés en océanographie ne sont donc pas *a priori* répartis de façon aléatoire dans l'espace-temps. LEGENDRE *et al.* (1986 : Tableau 2) ont d'ailleurs dégagé les échelles spatio-temporelles caractéristiques d'un certain nombre de niveaux trophiques. De telles échelles sont bien sûr différentes suivant l'axe considéré (vertical, horizontal ou temporel). Quelle que soit la stratégie d'échantillonnage retenue (voir, à ce propos, FRONTIER 1982), les observations océanographiques ne constituent donc généralement pas des réalisations indépendantes des variables aléatoires étudiées. Il s'ensuit une violation de l'hypothèse d'indépendance des observations, dont les conséquences sur les tests statistiques classiques seront discutées dans la prochaine section. Il existe par ailleurs un ensemble de méthodes numériques, encore relativement peu utilisées par les océanographes, qui permettent d'analyser les données dans un contexte d'autocorrélation spatio-temporelle. Ces méthodes, qui ont été développées pour l'analyse d'observations réparties sur des surfaces, peuvent être dans bien des cas étendues aux volumes ainsi qu'aux structures spatio-temporelles.

METHODES D'ANALYSE SPATIALE

En écologie, l'étude des relations spatiales entre phénomènes biologiques se limite souvent à ce qui est désigné en géographie quantitative sous le nom d'étude des patrons de points. Le chercheur souhaite alors établir si les organismes (considérés comme des points) sont répartis dans l'espace de façon agrégée, aléatoire ou régulière. Les méthodes numériques correspondantes peuvent être notamment trouvées dans CICERI *et al.* (1977) ou dans RILEY (1981).

Si, au contraire, une valeur est associée aux points d'observation, ou encore si ceux-ci représentent l'intégration d'une surface, d'autres méthodes sont à la disposition du chercheur pour déterminer l'influence de l'espace sur la répartition des variables mesurées. Ces méthodes, qui ont été développées dans des domaines aussi divers que l'épidémiologie, la géographie, le génie minier, la génétique des populations, l'écologie et la psychométrie, peuvent être classifiées en fonction des objectifs visés.

Objectif 1. Test de la présence d'autocorrélation spatiale. Cet objectif peut correspondre à deux points de vue diamétralement opposés.

(1) Dans un premier cas, le chercheur espère montrer qu'il n'y a pas d'autocorrélation spatiale, car il désire faire appel à la statistique paramétrique. En effet, tel que mentionné plus haut, l'autocorrélation spatiale influence les tests statistiques classiques. La violation de l'hypothèse d'indépendance des observations a pour effet de modifier la distribution de référence des statistiques calculées et, par conséquent, de modifier les résultats des tests effectués au moyen des distributions théoriques habituelles (BIVAND, 1980 ; CLIFF & ORD, 1981). A titre d'exemple, le coefficient de corrélation linéaire r calculé entre deux variables à forte autocorrélation spatiale conduit à une sous-estimation de la variance de r ; par conséquent, on rejette trop souvent l'hypothèse principale, reconnaissant ainsi comme significatives des corrélations qui ne le sont pas (test trop libéral). L'autocorrélation devrait donc être compensée par une perte de degrés de liberté.

(2) Le chercheur peut, au contraire, souhaiter démontrer l'existence d'une structure spatiale et chercher à en dégager la forme. Trois méthodes distinctes peuvent servir à tester la présence d'autocorrélation spatiale : le test de Mantel, le corrélogramme spatial (univariable) et le corrélogramme de Mantel (multivariable).

Le test de MANTEL (1967) permet d'établir s'il existe un gradient spatial, pour une variable seule ou pour une multivariable. Ce test mesure la relation entre les changements des valeurs de la variable (ou de la multivariable) et la proximité spatiale des observations. On teste la statistique de Mantel par permutations, ou encore par approximation à une distribution de t . Le test peut aussi être employé pour vérifier s'il existe une relation entre une multivariable et un modèle (par exemple, un modèle constitué par une classification obtenue à partir d'un autre ensemble de données).

Le corrélogramme spatial univariable utilise la statistique de Moran ou celle de Geary. Le corrélogramme permet de vérifier l'existence d'une structure spatiale, de préciser la forme de cette structure, enfin de tester des hypothèses particulières quant au(x) phénomène(s) générateur(s) de cette structure. Pour la mise en oeuvre de cette méthode, consulter SOKAL & ODEN (1978), LEGENDRE & LEGENDRE (1984b), ODEN (1984) ou SOKAL & THOMSON (1987).

Le corrélogramme de Mantel, récemment décrit par SOKAL *et al.* (sous presse), poursuit des objectifs similaires à ceux du corrélogramme spatial univariable. Il permet toutefois de traiter une multivariable échantillonnée dans l'espace.

Objectif 2. Test de modèles de causalité incluant l'espace. DOW & CHEVERUD (1985), HUBERT (1985) ainsi que SMOUSE *et al.* (1986) ont récemment proposé des méthodes permettant d'effectuer un test de Mantel partiel. Le test de Mantel partiel est réalisé à partir de trois matrices, dont la matrice des distances spatiales entre les observations. Les deux autres matrices contiennent des distances écologiques, univariables ou multivariables. Un tel test est analogue à la corrélation partielle, laquelle permet, dans le contexte de l'analyse des coefficients de direction, le test de modèles simples de causalité (LEGENDRE & LEGENDRE, 1984a). Les modèles de causalité dérivés du test de Mantel partiel sont limités, pour le moment, aux relations entre deux variables (ou multivariables) et l'espace.

Objectif 3. Reconstruction d'une surface ou estimation d'une ressource. Le chercheur souhaite généralement étendre à l'ensemble de la surface échantillonnée l'information provenant d'un nombre fini de points d'observation. Il peut, dans ce but, avoir recours aux cartes interpolées, au krigeage ou encore au groupement avec contrainte de contiguïté.

Les cartes interpolées ne permettent de représenter qu'une seule variable à la fois. Il existe plusieurs méthodes d'interpolation, incluant l'analyse des surfaces théoriques ("trend surface analysis"). Les programmes de calcul fournissent une estimation de la variable en tout point de la surface considérée et, souvent aussi, une carte des résidus ou de la variation non décrite par la carte interpolée.

Le krigeage permet de réaliser une cartographie souvent plus fine que les simples méthodes d'interpolation. Les programmes de calcul produisent de plus une carte de l'erreur type des estimations, laquelle peut servir entre autres à identifier les régions où il convient d'intensifier l'échantillonnage.

Le groupement avec contrainte de contiguïté permet, quant à lui, de cartographier des données multivariables. Ce groupement peut être réalisé pour des données réparties selon une, deux ou plusieurs dimensions. Les groupes contigus produits par l'algorithme de groupement sont ensuite représentés sur une carte. Une revue du sujet a été préparée par LEGENDRE (1987) ; voir aussi LEGENDRE & LEGENDRE, 1984c et LEGENDRE *et al.* (1985).

Les traitements numériques ci-dessus requièrent l'utilisation de l'ordinateur. Dans certains cas, les programmes nécessaires sont disponibles commercialement. Parmi ces derniers on trouve, pour les cartes interpolées, le progiciel SYMAP¹ et, pour les variogrammes et pour le krigeage, BLUEPACK², GEOSTAT³ ainsi que UNIMAP⁴. Par ailleurs, le progiciel "R" d'analyse des données multivariées, disponible auprès du premier auteur, contient des programmes pour les tests de Mantel (simples et partiels), le corrélogramme de Mantel, les corrélogrammes de Moran et de Geary ainsi que le groupement avec contrainte de contiguïté temporelle (une seule dimension) ou spatiale (deux dimensions ou plus).

CONCLUSION

Il découle de la discussion ci-dessus que les échantillons océanographiques ne constituent généralement pas des réalisations indépendantes de la variable aléatoire étudiée. Ceci ne dépend nullement du plan d'échantillonnage, mais provient plutôt de la nature même des processus qui structurent le milieu marin. Il est donc illusoire d'espérer lever cette contrainte en modifiant la stratégie d'échantillonnage. Dans un tel contexte, l'océanographe n'est pas justifié de supposer *a priori* que les variables échantillonnées en mer présentent des valeurs réparties de façon aléatoire. Il s'ensuit que les caractéristiques spatio-temporelle de l'échantillonnage constituent une variable, qui doit intervenir dans l'analyse des données océanographiques.

Les méthodes d'analyse spatiale fournissent un cadre général qui permet de traiter explicitement la structure spatio-temporelle lors de l'analyse des données. Ces méthodes sont de toute évidence appelées à un avenir brillant en océanographie. Toutefois, comme toujours en écologie numérique, le passage du numérique à l'écologique constitue une étape cruciale, où il importe qu'océanographes et spécialistes du traitement des données explorent ensemble les conditions d'applications des méthodes en tenant compte des caractéristiques propres aux données océanographiques. Comme toujours pour les écologistes, les traitements numériques ne constituent pas une fin en soi mais plutôt un moyen pour mieux appréhender la réalité complexe du milieu naturel.

-
- 1- Laboratory for Computer Graphics and Spatial Analysis, Harvard University, 520 Gund Hall, 48 Quincy Street, Cambridge, Massachusetts 02138, USA.
 - 2- Centre de Géostatistique et de Morphologie Mathématique, 35 rue Saint-Honoré, F-77305 Fontainebleau Cedex, France.
 - 3- Geostat Systems International Inc., 5180 Chemin de la Reine-Marie, Montréal, Québec H3W 3E7, Canada.
 - 4- European Software Contractors A/S, Norregade, DK-2800 Lyngby, Danemark.

RÉFÉRENCES BIBLIOGRAPHIQUES COMMENTÉES

- BIVAND (R.), 1980 - A Monte Carlo study of correlation coefficient estimation with spatially autocorrelated observations. *Questiones geographicae*, 6 : 5-10.
(Dangers de l'inférence à partir de coefficients de corrélation en présence d'autocorrélation spatiale)
- CICERI (M.-F.), MARCHAND (B.) & RIMBERT (S.), 1977 - *Introduction à l'analyse de l'espace*. Masson, Paris, 173 pp.
(Méthodes d'étude des patrons de points, etc.)
- CLIFF (A.D.) & ORD (J.K.), 1981 - *Spatial processes : models and applications*. Pion Ltd, London, 266 pp.
(Autocorrélation spatiale)
- DAVID (M.) - *Geostatistical ore reserve estimation*. Elsevier, Amsterdam, 364 pp.
(Manuel sur le krigeage)
- DOW (M.M.) & CHEVERUD (J.M.), 1985 - Comparison of distance matrices in studies of population structure and genetic microdifferentiation : quadratic assignment. *Am. J. phys. Anthropol.*, 68 : 367-373.
(Test de Mantel partiel : méthode de Dow et Cheverud).
- FRONTIER (S.) édit., 1982 - *Stratégies d'échantillonnage en écologie*. Masson, Paris et Presses de l'Université Laval, Québec, 494 pp.
(Plans d'échantillonnage en écologie)
- HUBERT (L.J.), 1985 - Combinatorial data analysis : association and partial association. *Psychometrika*, 50 : 449-467.
(Test de Mantel partiel : méthode de Hubert)
- LEGENDRE (L.), DEMERS (S.) & LEFAIVRE (D.), 1986 - Biological production at marine ergoclines. 1-29 in J.C. Nihoul édit., *Marine interfaces ecohydrodynamics*. Elsevier, Amsterdam, 670 pp.
(Production biologique aux interfaces énergétiques marines)
- LEGENDRE (L.) & LEGENDRE (P.), 1948a - *Ecologie numérique*, 2^e édit.. Tome 1 : *Le traitement multiple des données écologiques*. Masson, Paris et Presses de l'Université du Québec, 335 pp.
(Autocorrélation spatiale)
- LEGENDRE (P.), 1987 - Constrained clustering. 289-307. in P. Legendre & L. Legendre (édit.), *Developments in numerical ecology*. NATO Advanced Study Institute Series G (Ecological Sciences), Springer Verlag, Berlin, 585 pp.
(Revue des méthodes de groupement avec contrainte de contiguïté)
- LEGENDRE (P.), DALLOT (S.) & LEGENDRE (L.), 1985 - Succession of species within a community : chronological clustering, with applications to marine and freshwater zooplankton. *Amer. Nat.*, 125 : 257-288.
(Groupement avec contrainte de contiguïté temporelle)
- LEGENDRE (P.) & LEGENDRE (V.), 1984c - Postglacial dispersal of freshwater fishes in the Québec peninsula. *Can. J. Fish. aquat. Sci.*, 41 : 1781-1802.
(Groupement avec contrainte de contiguïté spatiale)

- MANTEL (N.), 1967 - The detection of disease clustering and a generalized regression approach. *Cancer Res.*, 27 : 209-220.
(Test de Mantel)
- ODEN (N.L.), 1984 - Assessing the significance of spatial correlograms. *Geogr. Anal.*, 16 : 1-16.
(Autocorrélation spatiale)
- RIPLEY (B.D.), 1981 - *Spatial statistics*. John Wiley & Sons, New York, 252 pp.
(Méthodes d'étude des patrons de points)
- SMOUSE (P.E.), LONG (J.C.) & SOKAL (R.R.), 1986 - Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.*, 35 : 627-637.
(Test de Mantel partiel : méthode de Smouse, Long et Sokal)
- SOKAL (R.R.) & ODEN (N.L.), 1978 - Spatial autocorrelation in biology. 1. Methodology. *Biol. J. Linnean Soc.*, 10 : 199-228.
(Autocorrélation spatiale)
- SOKAL (R.R.), ODEN (N.L.) & BARKER (J.S.F.), 1987 - Spatial structure in *Drosophila buzzatii* populations : simple and directional spatial autocorrelation. *Amer. Nat.*, 129 : 122-142.
(Corrélogramme multidimensionnel de Mantel et autres méthodes dérivées du test de Mantel)
- SOKAL (R.R.) & THOMSON (J.D.), 1987 - Applications of spatial autocorrelation in ecology. 431-466 in P. Legendre & L. Legendre (édit.), *Developments in numerical ecology*. NATO Advanced Study Institute Series G (Ecological Sciences), Springer-Verlag, Berlin. 585 pp.
(Revue des méthodes d'autocorrélation spatiale)