

## Pratique de l'analyse de l'autocorrélation spatiale en géomorphologie : définitions opératoires et tests

## Spatial autocorrelation analysis in geomorphology: Definitions and tests

## Aplicación del método de análisis de autocorrelación espacial en Geomorfología: funcionamiento y pruebas estadísticas

Philippe Aubry et Hervé Piégay

Volume 55, numéro 2, 2001

URI : <https://id.erudit.org/iderudit/008297ar>

DOI : <https://doi.org/10.7202/008297ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0705-7199 (imprimé)

1492-143X (numérique)

[Découvrir la revue](#)

Citer cet article

Aubry, P. & Piégay, H. (2001). Pratique de l'analyse de l'autocorrélation spatiale en géomorphologie : définitions opératoires et tests. *Géographie physique et Quaternaire*, 55(2), 111–129. <https://doi.org/10.7202/008297ar>

Résumé de l'article

L'autocorrélation spatiale peut être définie comme la ressemblance des valeurs prises par une variable, exprimée en fonction de leur localisation géographique. L'analyse de l'autocorrélation permet de quantifier la régularité spatiale d'un phénomène (une forme de complexité spatiale) et de déterminer la portée de la dépendance spatiale afin, notamment, de définir un dispositif d'échantillonnage garantissant l'indépendance des données, autorisant ainsi l'utilisation des tests statistiques usuels. L'article aborde trois points : i) définition opératoire de l'autocorrélation pour des variables quantitatives ou qualitatives ; ii) utilisation des tests de randomisation pour tester l'hypothèse nulle d'absence d'autocorrélation ; iii) illustration des deux points précédents par des exemples. Trois jeux de données simulés sont présentés pour illustrer les différences entre les statistiques : le premier ne présente aucune structure spatiale, le deuxième est caractérisé par une structure spatiale périodique, le troisième est un gradient linéaire. Deux jeux de données géomorphologiques sont également analysés : i) une série de segments fluviaux élémentaires, distribués longitudinalement, et sur lesquels ont été mesurées la largeur et l'incision d'un lit fluvial ; ii) la cartographie des formes d'érosion d'un bassin versant, traitée comme une image matricielle. Dans le premier cas, la structure emboîtée des tronçons géomorphologiquement homogènes est mise en lumière, et ce à différentes échelles spatiales. Le second exemple montre qu'une analyse omnidirectionnelle peut conduire à sous-estimer la portée de l'autocorrélation lorsque le phénomène étudié présente une orientation géographique privilégiée. Dans ce cas, il peut s'avérer impossible de définir un échantillon de données spatialement indépendantes, répondant aux exigences des tests statistiques classiques.

# PRATIQUE DE L'ANALYSE DE L'AUTOCORRÉLATION SPATIALE EN GÉOMORPHOLOGIE : DÉFINITIONS OPÉRATOIRES ET TESTS

Philippe AUBRY et Hervé PIÉGAY\* : UMR CNRS 5558 Biométrie et biologie évolutive, 43, boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France (actuellement : Addinsoft, 40, rue Damrémont, 75018 Paris, France); UMR CNRS 5600 Environnement, Ville, Société, 18, rue Chevreul, 69362 Lyon Cedex 07, France.

**RÉSUMÉ** L'autocorrélation spatiale peut être définie comme la ressemblance des valeurs prises par une variable, exprimée en fonction de leur localisation géographique. L'analyse de l'autocorrélation permet de quantifier la régularité spatiale d'un phénomène (une forme de complexité spatiale) et de déterminer la portée de la dépendance spatiale afin, notamment, de définir un dispositif d'échantillonnage garantissant l'indépendance des données, autorisant ainsi l'utilisation des tests statistiques usuels. L'article aborde trois points : i) définition opératoire de l'autocorrélation pour des variables quantitatives ou qualitatives ; ii) utilisation des tests de randomisation pour tester l'hypothèse nulle d'absence d'autocorrélation ; iii) illustration des deux points précédents par des exemples. Trois jeux de données simulés sont présentés pour illustrer les différences entre les statistiques : le premier ne présente aucune structure spatiale, le deuxième est caractérisé par une structure spatiale périodique, le troisième est un gradient linéaire. Deux jeux de données géomorphologiques sont également analysés : i) une série de segments fluviaux élémentaires, distribués longitudinalement, et sur lesquels ont été mesurées la largeur et l'incision d'un lit fluvial ; ii) la cartographie des formes d'érosion d'un bassin versant, traitée comme une image matricielle. Dans le premier cas, la structure emboîtée des tronçons géomorphologiquement homogènes est mise en lumière, et ce à différentes échelles spatiales. Le second exemple montre qu'une analyse omnidirectionnelle peut conduire à sous-estimer la portée de l'autocorrélation lorsque le phénomène étudié présente une orientation géographique privilégiée. Dans ce cas, il peut s'avérer impossible de définir un échantillon de données spatialement indépendantes, répondant aux exigences des tests statistiques classiques.

**ABSTRACT** *Spatial autocorrelation analysis in geomorphology: Definitions and tests.* Spatial autocorrelation can be defined as the similarity of values of a given variable, in relation with their spatial location. Autocorrelation functions are used to quantify the spatial regularity of a phenomenon (a form of spatial complexity) and to assess the lag of spatial dependence in order to design a sampling procedure for which the data are independent, which permits the use of traditional statistical tests. Three points have been developed: i) Definition of autocorrelation functions used for categorical and quantitative variables: Geary's *c*, Moran's *I*, semivariogram, non ergodic covariance and correlation, *J* statistics; ii) Definition of the randomization tests used to test the null hypothesis of no autocorrelation; iii) examples illustrating the two previous objectives. Three sets of simulated data were used to compare different autocorrelation functions (Geary's *c*, non ergodic covariance and correlation): the first one has no spatial structure, the second one has a periodic spatial structure whereas the third one is characterized by a linear gradient. Spatial autocorrelation has also been assessed on measured geomorphological data. Two sets were studied: i) a set of elementary channel segments of 500 m in length on which mean active channel width and degradation have been measured, ii) a set of pixels of an image of two basins representing the different types of hillslope erosion forms. In the first case, the nested structure of homogeneous geomorphological reaches is highlighted at different spatial scales. The second example, which illustrates autocorrelation assessment on a categorical variable, shows that omnidirectional analysis can underestimate the autocorrelation lag when the studied phenomenon is characterized by a preferential geographical orientation. In this particular case, it may not be possible to define a sample of data which are spatially independent and on which it is possible to use classical statistical tests.

**RESUMEN** *Aplicación del método de análisis de autocorrelación espacial en Geomorfología: funcionamiento y pruebas estadísticas.* El método de correlación espacial puede definirse como la semejanza de valores que adopta una variable expresada en función de su localización geográfica. Dicho método permite cuantificar la regularidad espacial de un fenómeno (un tipo de complejidad espacial) y determinar el alcance de la dependencia espacial a fin de definir una estrategia de muestreo que garantice la independencia de los datos y que permita la utilización de pruebas estadísticas usuales. El presente trabajo aborda los siguientes puntos: i) Explicación del funcionamiento de la autocorrelación aplicada a variables discretas o continuas; ii) utilización de pruebas aleatorias para aceptar o rechazar la hipótesis nula de ausencia de autocorrelación; iii) ilustración de los dos puntos anteriores por medio de ejemplos. Tres grupos de datos ficticios son presentados para ejemplificar las diferencias entre las estadísticas: el primero carece de estructura espacial, el segundo se caracteriza por una estructura espacial periódica y el tercero representa un gradiente lineal. Así mismo se analizan dos grupos de datos geomorfológicos: i) una serie de segmentos fluviales elementales distribuidos longitudinalmente y de los cuales se ha medido la amplitud e incisión del lecho fluvial; ii) la cartografía de los tipos de erosión de la pendiente de la cuenca tratado como imagen matricial. En el primero de los casos, la estructura constituida de elementos geomorfológicos homogéneos y puesta en evidencia a distintas escalas espaciales. El segundo ejemplo muestra que un análisis omnidireccional puede conducir a subestimar el alcance de la autocorrelación cuando el fenómeno estudiado muestra una orientación geográfica privilegiada. En este caso, puede ser imposible definir una muestra de datos espacialmente independientes, respondiendo así a las exigencias estadísticas clásicas.

## INTRODUCTION

La géomorphologie peut être considérée comme un domaine de recherche situé à l'interface entre plusieurs disciplines scientifiques. En effet, le géomorphologue fait appel à une variété de démarches et d'outils élaborés dans des disciplines connexes, principalement la géologie et l'écologie. La géologie, et surtout l'écologie, figurent parmi les disciplines scientifiques qui font le plus appel aux méthodes statistiques, qu'il s'agisse de l'analyse des données ou de la modélisation des phénomènes. Cependant, l'utilisation des méthodes statistiques est encore rare dans le domaine de la géomorphologie, en particulier la géomorphologie fluviale (Clément et Piégay, sous presse). Cette situation peut s'expliquer en partie par une raison historique. En effet, cette discipline présente de fortes racines déterministes et mécanistes, qui font référence à l'observation et à la mesure des processus sur le terrain, afin d'identifier les lois physiques qui les régissent. Néanmoins, ces lois peuvent faire intervenir des paramètres qui sont déterminés de façon empirique et non pas de façon théorique, et qui sont donc susceptibles de présenter une variabilité qui peut être appréhendée statistiquement. Le faible recours aux méthodes statistiques peut s'expliquer également par le fait que la géomorphologie, et en particulier la géomorphologie fluviale, produit souvent des jeux de données de petite taille. Chaque mesure correspondant à un événement, une crue, un glissement de terrain, les données comptent rarement plus de 10 à 20 valeurs, ce qui est généralement considéré comme trop faible pour pouvoir appliquer les outils statistiques classiques. En conséquence, les études de géomorphologie font peu appel aux outils statistiques, exception faite des méthodes de régression.

Toutefois, les méthodes statistiques, descriptives ou inférentielles, se révèlent aujourd'hui pertinentes en géomorphologie dans la mesure où :

1. les progrès métrologiques permettent d'acquérir un grand nombre de données. Par exemple, en géomorphologie fluviale, il devient aisé et relativement bon marché de mesurer le débit, voire les concentrations de matière en suspension, à l'échelle de plusieurs stations et d'esquisser ainsi une analyse des phénomènes à petite échelle. Il est donc possible d'étendre géographiquement des recherches auparavant conduites essentiellement au niveau stationnel, ce qui se traduit par une augmentation de la quantité des données à traiter (Rickenmann, 1997 ; Liébault *et al.*, 1999). Les progrès métrologiques permettent également de mieux comprendre les processus à grande échelle. Des outils statistiques sont ainsi utilisés pour étudier comment se distribuent les vitesses dans la masse d'eau et le lien entre vitesses, débit et début d'entraînement des particules (Clifford *et al.*, 1992 ; Kirkbride et Ferguson, 1995 ; Biron *et al.*, 1998) ;

2. l'analyse systémique (ou holistique) conduit à élargir le champ géographique des problématiques. Dans le cas de la géomorphologie fluviale, ce type d'analyse conduit à concevoir des recherches intégrées au sein de bassins versants (Schumm, 1977 ; Newson, 1994). Dans ce cadre, des jeux de données de grande taille peuvent être produits afin de mettre en lumière les liens bassin versant – chenal. Ces données se

présentent généralement sous une forme simplifiée pour pouvoir être collectées sur un très grand nombre de stations, dans le but de répondre à des problématiques appliquées ou interdisciplinaires (Downs, 1994). Par exemple, au lieu de mesurer précisément la granulométrie au laboratoire, il s'avère souvent préférable de considérer une variable nominale à  $k$  modalités identifiables sur le terrain, visuellement ou au touché (Piégay *et al.*, sous presse) ;

3. l'approche systémique en géomorphologie ainsi que les progrès technologiques en matière d'acquisition et de traitement de l'information expliquent le recours croissant à la télé-détection et aux SIG (systèmes d'information géographique) (Mertès *et al.*, 1995 ; Andriamahefa, 1999 ; Gilvear *et al.*, 1999) ; ces outils permettent d'analyser simultanément un ensemble de couches d'information géographique, par exemple en géomorphologie fluviale, les formes d'érosion, la position du réseau hydrographique, la pente, l'altitude ou la couverture végétale, afin de mettre en lumière les contrôles géomorphologiques à l'échelle du bassin versant, établir des scénarios et les simuler.

La plupart des variables considérées en géomorphologie ont pour caractéristique commune d'être « régionalisées », c'est-à-dire localisées dans l'espace géographique. En géomorphologie fluviale, cette localisation est unidimensionnelle, le long d'un continuum fluvial ou le long d'un transect (Ichim et Radoane, 1990 ; Gurnell, 1997), ou bidimensionnelle, au sein d'un bassin versant ou d'un lit fluvial (Montgomery *et al.*, 1998a). En conséquence, l'analyse des données géomorphologiques s'effectue naturellement au sein des SIG (Carrara *et al.*, 1991 ; Latulipe et Peiry, 1996 ; Montgomery *et al.*, 1998b). Toutefois, il convient de reconnaître que les SIG se sont développés indépendamment de l'analyse statistique des données spatiales, ce qui se traduit actuellement par une demande croissante d'intégration des deux (Arbia, 1993 ; Haining *et al.*, 1996). Dans le cadre d'une analyse statistique, les SIG sont le plus souvent utilisés pour stocker les données, les exporter vers un programme statistique externe, puis les résultats sont importés et visualisés. L'interfaçage entre les principaux SIG et logiciels statistiques du marché semble plus fort qu'il y a quelques années (Croft et Kessler, 1996), mais encore faut-il que les outils statistiques utilisés se révèlent adaptés à la nature régionalisée des variables traitées. En effet, le problème fondamental posé par l'analyse statistique des variables régionalisées réside dans le fait que les procédures statistiques usuelles (corrélation linéaire, régression, analyse de variance, etc.) ne tiennent pas compte de l'*autocorrélation spatiale* des valeurs (Annexe I).

L'autocorrélation spatiale peut être définie de façon intuitive comme la ressemblance des valeurs en fonction de leur localisation géographique. L'autocorrélation spatiale est positive lorsque les valeurs mesurées dans des stations voisines se ressemblent davantage qu'elles ne ressemblent aux autres valeurs. On parle d'autocorrélation négative dans le cas d'une dissemblance. Le terme *autocorrélation spatiale* est classiquement utilisé comme synonyme de *dépendance spatiale* (Gatrell, 1979 ; Cliff et Ord, 1981). L'autocorrélation traduit le degré de structuration spatiale d'une variable régionalisée, autrement dit une forme de complexité spatiale. La première

étape du traitement statistique d'une variable régionalisée consiste à décrire sa structure d'autocorrélation spatiale.

L'intérêt de l'analyse de l'autocorrélation en géomorphologie est multiple puisqu'elle permet :

- de déterminer la portée de l'autocorrélation spatiale, c'est-à-dire la distance à partir de laquelle l'hétérogénéité des valeurs peut être considérée comme indépendante de la localisation géographique. Connaissant la portée, on peut envisager de définir un dispositif d'échantillonnage garantissant l'indépendance spatiale des données, afin de pouvoir utiliser les tests statistiques usuels (test du coefficient de corrélation de Pearson, test d'indépendance du  $\chi^2$ , tests de l'ANOVA, etc.),

- de quantifier la régularité spatiale du phénomène, autrement dit, mesurer une forme de complexité spatiale. Cette démarche est essentiellement élaborée en écologie du paysage (par exemple, Turner *et al.*, 1991) mais peut être appliquée à des problématiques géomorphologiques. Par exemple, à partir des données d'un profil en long (distance longitudinale, altitude) ou de l'analyse d'un tracé en plan (longueur d'onde des méandres), il est possible d'identifier une tendance ainsi que des ruptures morphologiques, et de définir la longueur et l'amplitude des principales unités (Ferguson, 1975; Clifford *et al.*, 1992; Madej, 1999),

- de modéliser l'autocorrélation spatiale afin de pouvoir en tenir compte dans les procédures statistiques d'estimation spatiale, de test, ou de simulation.

L'objet de cet article est de traiter de l'analyse de l'autocorrélation spatiale des variables régionalisées en géomorphologie, et plus particulièrement en géomorphologie fluviale. La thématique de la modélisation de l'autocorrélation spatiale n'est pas abordée. L'article est organisé en trois sections : i) exposé de quelques définitions opératoires de l'autocorrélation spatiale pour des variables quantitatives ou nominales; ii) application des tests de randomisation afin d'évaluer l'hypothèse nulle d'absence d'autocorrélation spatiale; iii) illustration des deux points précédents au moyen d'exemples simulés et de données réelles utilisées en géomorphologie fluviale.

## PRÉSENTATION DES DÉFINITIONS OPÉRATOIRES

Une variable régionalisée se note sous la forme d'une fonction  $z(x)$  établissant la correspondance entre la valeur mesurée  $z_i$  et le support spatial de la mesure  $x_i$ , pour  $i=1, \dots, n$ ; toute statistique mesurant l'autocorrélation spatiale repose sur la mise en relation des deux informations constitutives de la variable régionalisée  $z(x)$  : a) le voisinage des supports  $x_i$ ; b) la ressemblance entre les valeurs  $z_i$ .

Dans l'espace des supports (espace géographique) et dans l'espace des valeurs, il est possible de définir différents types de proximités. En outre, il existe plusieurs approches pour mettre en relation les proximités entre supports et les proximités entre valeurs, ce qui conduit à de nombreuses définitions opératoires de l'autocorrélation spatiale.

## DÉFINITIONS PROPOSÉES EN GÉOGRAPHIE STATISTIQUE

Les statistiques mesurant la dépendance spatiale sont essentiellement issues de la géographie (revue dans Cliff et Ord, 1973, 1981), domaine largement dominé par les tests de l'autocorrélation spatiale (Ripley, 1988). Actuellement, ce sont encore les géographes qui produisent le plus de mesures et de tests d'autocorrélation spatiale (par exemple, Anselin, 1988, 1995; Getis, 1989; Hubert et Arabie, 1991; Deichman et Anselin, 1994; Ord et Getis, 1995; Tiefelsdorf et Boots, 1997; Simon, 1997). Les statistiques dont l'usage est le plus répandu en géographie sont le  $c$  de Geary (Geary, 1954) et le  $I$  de Moran (Moran, 1950).

$c$  de Geary

La forme originelle du  $c$  de Geary (Geary, 1954) peut être généralisée en considérant une matrice  $\mathbf{W}$  de poids  $\omega_{ij}$  affectés à tous les couples de supports  $(x_i, x_j)$ , notés plus simplement  $(i, j)$  (Cliff et Ord, 1981) :

$$c = \frac{(n-1) \sum_{i,j} \omega_{ij} (z_i - z_j)^2}{2W \sum_i (z_i - \bar{z})^2} \quad (1)$$

avec  $W = \sum_{i,j} \omega_{ij}$ . La somme du numérateur est étendue à tous les couples de supports en faisant varier les indices  $i$  et  $j$ , la somme du dénominateur étant étendue à tous les supports.

Le  $c$  de Geary n'admet pas de borne supérieure mais admet une borne inférieure évidente  $\min(c)=0$  qui correspond à l'autocorrélation spatiale maximale, c'est-à-dire lorsque les valeurs comparées sont identiques. Par construction, le  $c$  de Geary est un rapport d'estimations de la variance respectivement spatiale et a-spatiale. Par conséquent, la valeur attendue du  $c$  de Geary en absence d'autocorrélation spatiale est  $E[c]=1$  (Geary, 1954; Cliff et Ord, 1981). Les valeurs comprises dans l'intervalle  $[0,1[$  témoignent donc d'une autocorrélation spatiale positive tandis que les valeurs dans  $]1,+\infty[$  traduisent la présence d'autocorrélation négative.

Les poids de la matrice  $\mathbf{W}$  traduisent la manière dont on connecte les supports entre eux, c'est-à-dire la relation de voisinage des supports. Il existe deux situations selon que le voisinage est défini en faisant intervenir : a) un graphe de voisinage des supports (éventuellement pondéré); b) les vecteurs formés par les couples de supports (le plus souvent dans un espace muni de la métrique euclidienne).

Si les distances euclidiennes n'ont pas beaucoup de sens, par exemple lorsque la densité relative des supports varie fortement et/ou que les erreurs de localisation ne sont pas négligeables, il convient d'utiliser à la place un graphe de voisinage représentant la topologie des connexions entre supports (Legendre et Fortin, 1989). Le problème est alors de savoir si la question posée par le géomorphologue permet de définir ce graphe. Dans le cas de supports ponctuels repérés le

long d'une rivière ou d'une ligne de côté, le graphe est naturellement réduit à une simple chaîne (Royaltey *et al.*, 1975; Chessel et Sabatier, 1993; Thioulouse *et al.*, 1995). Dans le cas de supports ponctuels répartis dans le plan, la question posée peut parfois indiquer le graphe le plus approprié, mais en pratique il est généralement difficile de présenter des arguments en faveur d'un graphe particulier (Gordon et Finden, 1985). Le choix de la matrice  $\mathbf{W}$  conditionne la nature de l'autocorrélation spatiale testée (Gabriel et Sokal, 1969; Sokal et Oden, 1978). La question est alors de savoir ce que l'on doit inférer si l'autocorrélation s'avère statistiquement significative en utilisant une certaine définition opératoire (par exemple, une certaine matrice de poids  $\mathbf{W}$ ) et non significative avec une autre définition (Gatrell, 1979). La réponse est simplement qu'un résultat plus ou moins significatif met en évidence l'adéquation plus ou moins bonne du modèle de variation spatiale imposé a priori par  $\mathbf{W}$  aux données (Jumars *et al.*, 1977; Cliff et Ord, 1981, p. 168, 174). Dans cette optique, plutôt que de tester différents modèles de variation spatiale, il est possible de chercher directement la matrice  $\mathbf{W}$  qui maximise la statistique utilisée (Kooijman, 1976; Boots et Dufournaud, 1994). Cependant, il reste ensuite à interpréter la signification du modèle de variation spatiale obtenu, ce qui peut s'avérer assez difficile.

#### Liens entre le $c$ de Geary et d'autres statistiques

Le  $c$  de Geary (Geary, 1954) est lié à d'autres statistiques définies notamment dans le domaine des séries temporelles et en écologie statistique (Annexe II), ainsi qu'au variogramme de la géostatistique. En effet, en considérant uniquement les couples de supports  $(i, j)$  séparés par un vecteur  $\mathbf{h}_{ij} = \mathbf{h}$ , la matrice  $\mathbf{W}$  définit une classe de vecteurs, et le  $c$  de Geary s'écrit comme la fonction :

$$c(\mathbf{h}) = \frac{\hat{\gamma}(\mathbf{h})}{s_{n-1}^2} \quad (2)$$

avec  $s_{n-1}^2 = (n-1)^{-1} \sum_i (z_i - \bar{z})^2$  l'estimateur sans biais de la variance théorique, et le (demi) variogramme « expérimental » défini en géostatistique par Matheron (1965). Le variogramme peut s'écrire (Isaaks et Srivastava, 1989) :

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{(i,j)|\mathbf{h}_{ij}=\mathbf{h}} (z_i - z_j)^2 \quad (3)$$

avec  $N(\mathbf{h})$  le nombre de couples de supports séparés par le vecteur  $\mathbf{h}$ . En considérant uniquement la norme du vecteur  $\mathbf{h}$ , notée  $h$ , la valeur  $2\hat{\gamma}(h)$  n'est pas autre chose que la moyenne des carrés des écarts des valeurs dont les supports sont séparés par la distance  $h$ . Sous cette forme, le  $c$  de Geary apparaît comme le variogramme expérimental standardisé par une estimation de la variance théorique. En conséquence, la représentation graphique du  $c$  de Geary est strictement identique à celle du variogramme, seule l'échelle des valeurs prises par la statistique étant modifiée par la standardisation par la variance.

Dans ce qui suit, nous considérons exclusivement le  $c$  de Geary sous la forme d'une fonction de vecteurs inter-supports (équ. 2).

#### $I$ de Moran

De même que pour le  $c$  de Geary, la forme originelle du  $I$  de Moran (Moran, 1950) peut être généralisée en considérant une matrice  $\mathbf{W}$  de poids  $\omega_{ij}$  affectés à tous les couples de supports  $(x_i, x_j)$  (Cliff et Ord, 1981) :

$$I = \frac{n}{2W} \frac{\sum_{i,j} \omega_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2} \quad (4)$$

avec  $W = \sum_{i,j} \omega_{ij}$ . La somme du numérateur est étendue à tous les couples de supports en faisant varier les indices  $i$  et  $j$ , la somme du dénominateur étant étendue à tous les supports.

Le  $I$  de Moran ne s'interprète pas exactement comme un coefficient de corrélation classique en ce qu'il n'est pas restreint à l'intervalle  $[-1, +1]$  et ne vaut pas exactement 0 en absence d'autocorrélation spatiale. En général, la borne supérieure  $|I|$  est inférieure à 1, bien qu'elle puisse dépasser 1 dans certains cas particuliers (Cliff et Ord, 1981, p. 21). La valeur attendue du  $I$  de Moran en absence d'autocorrélation spatiale est  $E[I] = -(n-1)^{-1}$  (Moran, 1950; Cliff et Ord, 1981). Une valeur inférieure à  $E[I]$  traduit de l'autocorrélation négative tandis qu'une valeur supérieure témoigne d'une autocorrélation positive.

#### DÉFINITIONS PROPOSÉES EN GÉOSTATISTIQUE

D'autres mesures de la dépendance spatiale que le  $c$  de Geary et le  $I$  de Moran ont été introduites en géostatistique à des fins de modélisation de la structure d'autocorrélation spatiale (revue dans Isaaks et Srivastava, 1989; Deutsch et Journel, 1992; Myers, 1997). La statistique la plus utilisée est le variogramme expérimental (Matheron, 1965), défini dans la section consacrée au  $c$  de Geary (équ. 3). Plus récemment, Isaaks et Srivastava (1988) ont introduit deux fonctions de covariance et de corrélation dites *non ergodiques* qui s'avèrent être d'un grand intérêt pour l'analyse de l'autocorrélation spatiale.

La différence entre la covariance spatiale classique (équ. 13, annexe III) et la covariance dite *non ergodique* tient dans la prise en compte des moyennes calculées sur deux ensembles formés par les valeurs situées de part et d'autre du vecteur qui les sépare, au lieu d'une moyenne globale qui ne tient pas compte de la localisation des valeurs. La covariance non ergodique s'écrit (Isaaks et Srivastava, 1989) :

$$\hat{C}_D(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{(i,j)|\mathbf{h}_{ij}=\mathbf{h}} (z_i - m_{-\mathbf{h}})(z_j - m_{+\mathbf{h}}) \quad (5)$$

avec  $m_{-h}$  la moyenne de toutes les valeurs dont les supports sont situés à  $-h$  des autres, c'est-à-dire à une extrémité du vecteur  $h$  :

$$m_{-h} = \frac{1}{N(h)} \sum_{i|h_j=h} z_i \quad (6)$$

et  $m_{+h}$  la moyenne de toutes les valeurs dont les supports sont situés à  $+h$  des autres, c'est-à-dire à l'autre extrémité du vecteur  $h$  :

$$m_{+h} = \frac{1}{N(h)} \sum_{j|h_j=h} z_j \quad (7)$$

Les valeurs de  $m_{-h}$  et de  $m_{+h}$  sont généralement différentes (Isaaks et Srivastava, 1989, p. 59).

De même que la corrélation classique est définie comme la covariance entre deux variables, divisée par le produit des écarts types de chacune des variables, une fonction d'auto-corrélation spatiale standardise la covariance spatiale d'une variable avec elle-même (autocovariance spatiale), en la divisant par sa variance (c'est-à-dire le produit de son écart type par lui-même). La corrélation dite *non ergodique* est formée de la même façon, en divisant la covariance non ergodique, non plus cette fois par la variance globale — qui ne tient pas compte de la localisation des valeurs — mais par le produit d'écarts types calculés sur deux ensembles formés par les valeurs situées de part et d'autre du vecteur qui les sépare. La corrélation non ergodique s'écrit donc :

$$\hat{\rho}_D(h) = \frac{\hat{C}_D(h)}{\sigma_{-h}\sigma_{+h}} \quad (8)$$

avec  $\sigma_{-h}^2$  la variance de toutes les valeurs dont les supports sont situés à  $-h$  des autres, c'est-à-dire à une extrémité du vecteur  $h$  :

$$\sigma_{-h}^2 = \frac{1}{N(h)} \sum_{i|h_j=h} (z_i - m_{-h})^2 \quad (9)$$

et  $\sigma_{+h}^2$  la variance de toutes les valeurs dont les supports sont situés à  $+h$  des autres, c'est-à-dire à l'autre extrémité du vecteur  $h$  :

$$\sigma_{+h}^2 = \frac{1}{N(h)} \sum_{j|h_j=h} (z_j - m_{+h})^2 \quad (10)$$

De même que pour les moyennes, les valeurs de  $\sigma_{-h}^2$  et de  $\sigma_{+h}^2$  sont généralement différentes (Isaaks et Srivastava, 1989, p. 60). Ainsi, la nouveauté introduite par les fonctions non ergodiques est la prise en compte d'éventuelles différences : a) entre les moyennes locales, grâce à la distinction entre  $m_{-h}$  et  $m_{+h}$ ; b) entre les variances locales, grâce à la distinction entre  $\sigma_{-h}^2$  et  $\sigma_{+h}^2$ .

Pour comparer ces deux fonctions entre elles et avec le variogramme standardisé  $c(h)$  il faut utiliser les transformations suivantes (Rossi *et al.*, 1992; Liebhold *et al.*, 1993) :

a)  $c^{(1)}(h) = \{s^2 - \hat{C}_D(h)\} / s^2$  pour la covariance non ergodique; b)  $c^{(2)}(h) = 1 - \hat{\rho}_D(h)$  pour la corrélation non ergodique.

Après transformation, les fonctions varient dans le même sens que le variogramme standardisé ( $c$  de Geary) et ont même espérance sous l'hypothèse nulle d'absence d'auto-corrélation ( $E[c(h)]=1$ ). Selon le couple de fonctions considéré, une superposition imparfaite des représentations graphiques montre donc des différences de moyennes locales ( $c(h)$  et  $c^{(1)}(h)$ ), de variances locales ( $c^{(1)}(h)$  et  $c^{(2)}(h)$ ) ou éventuellement des deux à la fois ( $c(h)$  et  $c^{(2)}(h)$ ).

La géostatistique définit également d'autres fonctions telles que le rodogramme, ou le madogramme (Journel, 1988; Deutsch et Journel, 1992, p. 42; Walker *et al.*, 1997), ainsi que différents types de variogrammes relatifs (Deutsch et Journel, 1992, p. 41-42; Myers, 1997, p. 297-300). Dans ce qui suit, nous considérons exclusivement le variogramme standardisé ou  $c$  de Geary  $c(h)$ , la covariance non ergodique et la corrélation non ergodique sous leurs formes  $c^{(1)}(h)$  et  $c^{(2)}(h)$ .

#### DÉFINITION POUR DONNÉES NOMINALES

Le  $c$  de Geary, la covariance non ergodique et la corrélation non ergodique peuvent être appliqués aussi bien à des données quantitatives qu'à des données binaires, bien que des statistiques spécifiques aient été développées dans le cas des données binaires (par exemple, Moran, 1948). En revanche, dans le cas des données nominales à plusieurs modalités (par exemple, des types de sols, différentes formes d'érosion, ou différents couverts végétaux), il convient d'utiliser une autre statistique pour mesurer la dépendance spatiale. En effet, la statistique doit être adaptée à la structure algébrique des données nominales qui permet uniquement de tester l'égalité ou l'inégalité stricte de deux modalités. Afin d'obtenir une mesure qui s'interprète de façon similaire au variogramme, nous proposons la statistique suivante :

$$J(h) = \frac{1}{N(h)} \sum_{(i,j)|h_{ij}=h} \delta'(i,j) \quad (11)$$

avec

$$\delta'(i,j) = \begin{cases} 0 & \text{si } z_i = z_j \\ 1 & \text{sinon} \end{cases}$$

La statistique  $J(h)$  n'est pas autre chose que la statistique  $a(h)$  proposée antérieurement par Deichman et Anselin (1994), mais exprimée sous la forme d'une dissimilarité plutôt que sous la forme d'une similarité, la ressemblance maximale moyenne étant atteinte pour 0 et la dissemblance maximale moyenne pour 1. Il convient de noter que les modalités prises en compte dans le calcul de  $J(h)$  doivent correspondre à un même niveau hiérarchique dans la description des phénomènes géomorphologiques car une variable nominale ne permet pas de distinguer différents niveaux de précision dans la description.

## TEST DES FONCTIONS D'AUTOCORRÉLATION

### TESTS PARAMÉTRIQUES

Le  $c$  de Geary et le  $I$  de Moran sont tous les deux asymptotiquement normalement distribués sous l'hypothèse nulle  $H_0$  d'absence d'autocorrélation. Afin de tester les valeurs observées en faisant référence à la loi normale, les moments des deux statistiques sont donnés par Cliff et Ord (1981). En ce qui concerne le  $c$  de Geary, la validité de l'approximation normale est discutée dans Cliff et Ord (1981). La normalité de la distribution du  $I$  de Moran sous  $H_0$  a donné lieu à une littérature plus abondante que dans le cas du  $c$  de Geary, citons notamment Cliff et Ord (1971), Sen (1976), Cliff et Ord (1977), Sen (1977), Cliff et Ord (1981) et plus récemment, Terui et Kikuchi (1994), et Waldhör (1996).

Les tests paramétriques du variogramme, de la covariance et de la corrélation non ergodiques n'ont pas été développés, l'objet des géostatisticiens n'étant pas de tester l'autocorrélation spatiale, mais de la modéliser. À notre connaissance, il n'existe pas non plus de test paramétrique concernant la statistique  $J$  pour données nominales.

### TESTS DE RANDOMISATION

Dans le cas où le nombre de supports est faible ou lorsque nous ne disposons pas d'un test paramétrique, il est particulièrement intéressant d'utiliser un test permutatif. Le principe d'un test permutatif consiste à générer toutes les valeurs possibles d'une statistique — c'est-à-dire ici pour toutes les permutations des valeurs parmi les supports — afin de calculer la  $p$ -value exacte associée à la valeur observée de la statistique (détails dans l'annexe IV). Le test permutatif peut être envisagé pour  $n \leq 10$ . Dans la plupart des cas ( $n > 10$ ), l'énumération de toutes les permutations est impraticable et le test permutatif est estimé par un test de randomisation (Edgington, 1986, 1987).

Un test de randomisation consiste à réaliser aléatoirement un nombre limité de permutations au lieu de réaliser de façon exhaustive toutes les permutations possibles. Il est conseillé de générer un nombre élevé de permutations aléatoires afin d'obtenir une estimation précise de la  $p$ -value. Pour un jeu de données comportant une centaine de valeurs, un nombre de permutations  $m = 10\ 000$  semble constituer un minimum raisonnable en deçà duquel il convient de ne pas descendre. Lorsque les données sont plus nombreuses, il conviendrait également d'augmenter le nombre des permutations aléatoires, en restant toutefois dans des limites raisonnables en termes de temps de calcul. Outre le calcul d'une  $p$ -value, la distribution empirique des valeurs de la statistique calculées pour l'ensemble des permutations aléatoires permet de construire un intervalle de confiance pour la valeur observée.

### Statistiques équivalentes

Dans le cadre d'un test de randomisation, le variogramme est une statistique équivalente au  $c$  de Geary et la fonction de covariance est une statistique équivalente au  $I$  de Moran. En

effet, le  $c$  de Geary et le  $I$  de Moran sont équivalents, respectivement : a) au variogramme standardisé par  $s_{n-1}^2$  (équ. 2) ; b) à la fonction de covariance, standardisée par  $s_n^2$  (équ. 12, Annexe III).

Or, le calcul de la variance des données est invariant par permutation de l'ordre des valeurs, de sorte que la standardisation par  $s_{n-1}^2$  (ou par  $s_n^2$ ) ne change pas les résultats des tests de randomisation.

### Intérêt des tests de randomisation

L'intérêt des tests de randomisation est de ne pas faire appel à une hypothèse distributionnelle concernant la statistique testée, ce qui évite notamment d'avoir à développer analytiquement les moments de la statistique sous l'hypothèse nulle  $H_0$  d'absence d'autocorrélation, et permet ainsi de tester toute statistique, y compris une statistique qui serait imaginée par le géomorphologue afin de répondre aux besoins d'une étude particulière.

Les  $p$ -values étant calculées pour chaque classe de vecteurs  $h$ , il est possible de fixer un seuil arbitraire (par exemple,  $\alpha = 0,05$ ) pour identifier les valeurs d'autocorrélation qui sont significatives à ce seuil. La  $p$ -value étant vue comme la force de l'évidence contre l'hypothèse nulle, il est préférable de visualiser toutes les  $p$ -values afin d'interpréter avec finesse la fonction d'autocorrélation spatiale associée. À cet effet, Walker *et al.* (1997) proposent de représenter les  $p$ -values comme des fonctions ou  $p$ -grammes, conjointement aux fonctions d'autocorrélation spatiale, qu'il s'agisse du variogramme, d'un covariogramme ou d'un corrélogramme. La distribution approximative de la statistique sous  $H_0$  étant connue, il est également possible de tracer des intervalles de confiance approximatifs des fonctions d'autocorrélation spatiale, pour différents seuils arbitraires tels que  $\alpha = 0,10$ ,  $\alpha = 0,05$  ou  $\alpha = 0,01$ .

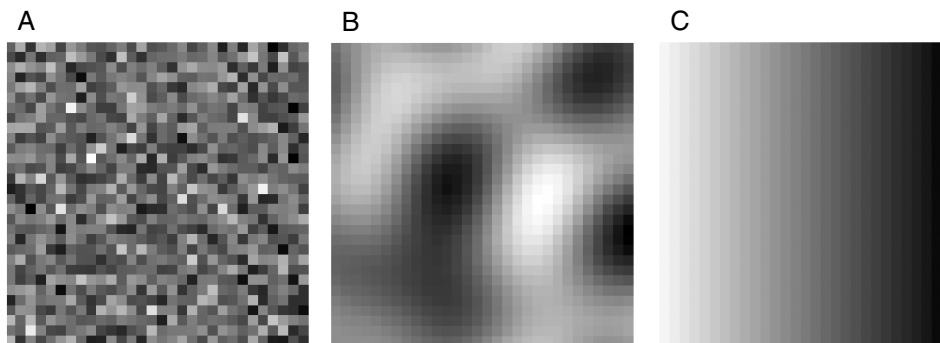
## IDENTIFICATION DE LA PORTÉE DE L'AUTOCORRÉLATION : EXEMPLES DE DONNÉES SIMULÉES ET RÉELLES

Au-delà de la question de savoir si l'autocorrélation est significative ou pas, ce qui intéresse le géomorphologue c'est de mesurer la portée de l'autocorrélation afin de déterminer l'échelle de la dépendance spatiale. Lorsque l'autocorrélation positive décroît avec la distance puis s'annule complètement, la portée est définie de façon stricte comme la distance à partir de laquelle l'autocorrélation est non significative. Lorsque l'autocorrélation décroît, s'annule, puis change de signe, la notion de portée perd de sa signification.

Classiquement, les fonctions issues de la géostatistique sont considérées comme utiles pour mesurer la portée de l'autocorrélation spatiale (Rossi *et al.*, 1992). En fait, le corrélogramme du  $I$  de Moran peut tout aussi bien être utilisé qu'un variogramme ou que la covariance non ergodique, mais dans tous les cas, la définition opératoire de la portée constitue un problème délicat. Il convient d'abord de remarquer que la définition opératoire de la portée dépend du type de fonction d'autocorrélation et qu'un résultat différent peut être obtenu selon

FIGURE 1. Images en niveaux de gris 30 x 30 pixels des variables quantitatives simulées. A) Absence de structure spatiale. B) Structure spatiale périodique. C) Gradient linéaire orienté E-O.

*Panchromatic images 30 x 30 pixels of simulated quantitative variables. A) No spatial structure. B) Periodic spatial structure. C) Linear gradient E-W oriented.*



que l'on utilise le  $c$  de Geary (ou ce qui revient au même, le variogramme) ou le  $I$  de Moran (ou ce qui revient au même, la covariance, Annexe III) : la notion de portée n'échappe évidemment pas au problème de la définition opératoire de l'autocorrélation spatiale elle-même. Par exemple, dans leur étude de la teneur en cuivre et en zinc dans les sédiments du bassin du Yangtze (Chine), Zhang et Selinus (1997) mettent en évidence une portée d'environ 1 000 km avec le variogramme et de 500 km avec le corrélogramme du  $I$  de Moran.

Nous proposons, à l'instar de Walker *et al.* (1997), d'utiliser le  $p$ -gramme afin de déterminer le plus objectivement possible la portée de l'autocorrélation spatiale. Le choix d'un seuil arbitraire  $\alpha$  au-delà duquel l'autocorrélation spatiale serait jugée non significative a peu de sens. En effet, le choix de  $\alpha$  influence évidemment la définition de la portée (Walker *et al.*, 1997). Il est plus judicieux d'examiner la variation des  $p$ -values au sein du  $p$ -gramme : lorsque plusieurs  $p$ -values successives plaident en faveur de l'hypothèse  $H_0$ , alors c'est que la portée de l'autocorrélation a été atteinte ou a été légèrement dépassée.

## DONNÉES SIMULÉES

### Protocole

Trois images ont été simulées sur une grille de 30 x 30 pixels de pas  $\Delta = 1$  (fig. 1). La première image a été obtenue en permutant aléatoirement les pixels d'une image spatialement structurée et ne présente par conséquent plus aucune structure spatiale comme peut en témoigner l'absence de continuité spatiale des valeurs des pixels (fig. 1A). Au contraire, l'image de la figure 1B montre une structure spatialement régulière qui présente un caractère périodique avec alternance de zones foncées et de zones claires, séparées par des zones de transition graduelle. Enfin, un gradient orienté E-O est représenté sur la figure 1C.

Pour chaque image, un échantillon de 100 pixels a été prélevé selon une grille 10 x 10 centrée dans la grille originelle de 30 x 30 pixels. Pour chaque échantillon, le  $c$  de Geary ainsi que la covariance et la corrélation non ergodiques ont été calculés pour un ensemble de 7 distances séparées par des

intervalles de largeur  $d = 3$ . Le nombre de distances est limité à 7 afin de ne pas dépasser la distance maximale de représentativité de la fonction, c'est-à-dire la distance au-delà de laquelle seuls les pixels situés en périphérie sont comparés entre eux. Soit  $L_{\max}$  la distance maximale dans l'image (ici  $L_{\max} = 30\sqrt{2} \approx 42,4$ ) ; la distance maximale de représentativité de la fonction vaut  $L_{\max}/2$ , soit ici  $L_{\max}/2 \approx 21,2$  (Journel, 1985 ; Rossi *et al.*, 1992). Le cas du gradient étant un exemple de structure parfaitement anisotrope, les fonctions ont été calculées uniquement dans la direction E-O, pour des multiples exacts de  $d$ . Pour les deux autres échantillons, les fonctions ont été calculées dans toutes les directions, les structures sous-jacentes ne montrant pas de direction privilégiée (isotropie). En outre, les valeurs ont été calculées pour des classes de distances définies pour des multiples de  $d$ , avec une tolérance  $\varepsilon = 1,5$ . L'usage d'une tolérance permet de lisser la fonction tout en augmentant le nombre de couples de pixels intervenant dans le calcul de chaque valeur de la fonction d'autocorrélation (David, 1977 ; Journel et Huijbregts, 1978). La tolérance  $\varepsilon = 0,5 \times d$  permet d'utiliser toute l'information disponible et s'avère d'usage courant (par exemple, Isaaks et Srivastava, 1989, p. 148). Enfin, des intervalles de confiance au seuil de 1 % ont été établis dans chaque cas sur la base de 10 000 permutations aléatoires.

### Résultat

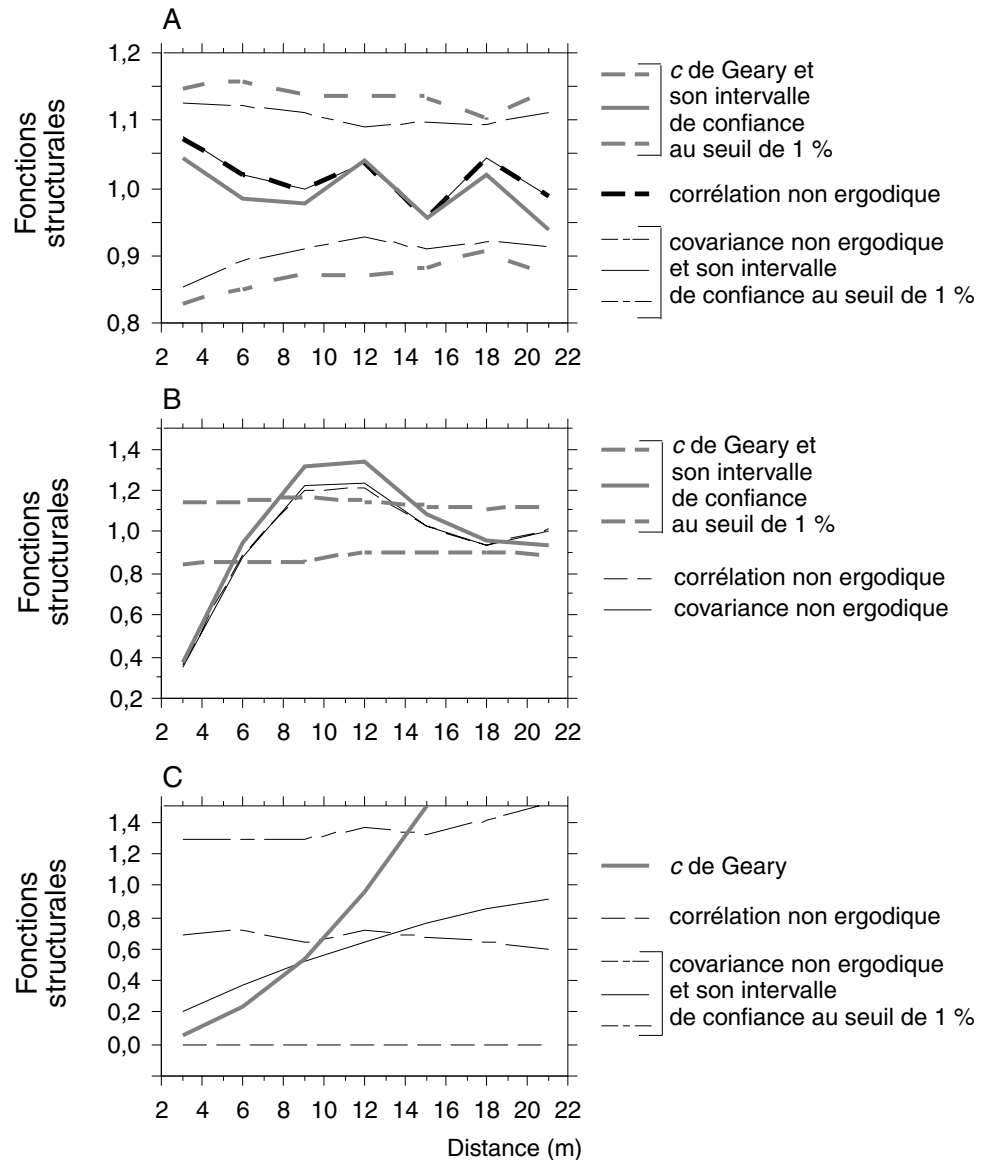
Dans le cas de l'échantillon de l'image aléatoire, la superposition de la covariance et de la corrélation non ergodiques est parfaite, ce qui montre l'absence de variation de la variance locale. Il existe toutefois de légères différences de moyennes locales, sauf pour les distances 12 et 15 (fig. 2A). L'examen des intervalles de confiance à 1 % du  $c$  de Geary et de la covariance non ergodique montre que l'autocorrélation n'est pas significative, quelle que soit la distance (fig. 2A).

Dans le cas de l'échantillon de l'image montrant une structure périodique, les trois fonctions calculées sont pratiquement identiques (fig. 2B). Ceci montre l'absence de différences marquées à la fois des moyennes et des variances locales. La nature périodique de la structure spatiale sous-jacente se traduit par une oscillation amortie des fonctions d'autocorrélation autour de leur espérance sous  $H_0$  ( $E[c(\mathbf{h})] = 1$ ). L'intervalle



FIGURE 2. Fonctions d'autocorrélation des variables quantitatives simulées, superposition du  $c$  de Geary, de la covariance et de la corrélation non ergodiques. A) Absence de structure spatiale. B) Structure spatiale périodique. C) Gradient linéaire orienté E-O.

Examples of three autocorrelation functions of simulated quantitative variables, the Geary's  $c$ , the non ergodic covariance and correlation. A) No spatial structure. B) Periodic spatial structure. C) Linear gradient E-W oriented.



de confiance du  $c$  de Geary montre que l'autocorrélation spatiale décroît, puisqu'elle est d'abord positive (de 3 à 6), non significative (de 6 à 8), puis négative (entre 8 et 12), avant de croître à nouveau (entre 12 et 15) et de s'amortir autour de  $E[c(\mathbf{h})] = 1$ .

Les trois fonctions calculées pour l'échantillon du gradient sont nettement distinctes (fig. 2C). En effet, le  $c$  de Geary croît comme une branche de parabole tandis que la covariance non ergodique semble bornée par un plateau correspondant à l'espérance sous  $H_0$  ( $E[c(\mathbf{h})] = 1$ ) et que la corrélation non ergodique est parfaitement horizontale et située à 0. L'allure du  $c$  de Geary indique typiquement la présence d'une tendance, ce qui est confirmé par l'allure très différente de la covariance non ergodique qui tient compte des différences de moyennes locales. L'intervalle de confiance de la covariance non ergodique permet de définir une portée située entre 15 et 18, soit approximativement la moitié de la largeur de la grille 30 x 30. Ceci indique qu'une fois tenu compte des différences

de moyennes entre les extrémités des vecteurs inter-supports, des valeurs séparées par une bande de 15 m peuvent être considérées comme indépendantes. Enfin, la corrélation non ergodique est nulle quelle que soit la distance considérée, ce qui témoigne d'une ressemblance parfaite à toutes les distances et l'absence d'une source de variabilité autre que les différences de moyennes et de variances calculées aux extrémités des vecteurs inter-supports.

#### Discussion

Les images simulées (fig. 1) correspondent à des situations bien tranchées qui permettent de faire le lien entre la structure spatiale telle qu'elle apparaît visuellement et l'allure du  $c$  de Geary, de la covariance et de la corrélation non ergodiques.

L'aspect plat des fonctions d'autocorrélation dans le cas de l'image aléatoire (fig. 1A) traduit le fait que la variabilité

des valeurs est indépendante de la distance ; autrement dit, la variabilité n'est pas spatialement structurée. L'absence de structure spatiale est connue dans la littérature géostatistique sous le nom d' « effet de pépite pur » (*pure nugget effect*) (Journel et Huijbregts, 1978). L'effet de pépite pur constitue donc l'hypothèse nulle du test de l'autocorrélation spatiale et peut être démontré sans ambiguïté à l'aide d'un test de randomisation (fig. 2A). En pratique, l'absence d'autocorrélation spatiale significative aux faibles distances constitue l'exception plutôt que la règle dans le cas des phénomènes géomorphologiques.

L'homogénéité spatiale des moyennes et des variances locales observée dans le cas de la structure périodique (fig. 2B) provient du fait que l'image a été obtenue en calculant une réalisation d'un modèle stochastique (fonction aléatoire) stationnaire à l'ordre 2 (voir Journel et Huijbregts, 1978 ; Myers 1989). L'homogénéité spatiale des moyennes et des variances locales ainsi que l'isotropie constituent les hypothèses structurelles les plus simples, auxquelles il est fait référence par défaut lors de l'analyse de l'autocorrélation spatiale. L'écart à ces hypothèses par défaut peut être mis en évidence de façon assez fine en comparant le  $c$  de Geary, la covariance et la corrélation non ergodique, calculées selon différentes directions.

L'exemple du gradient parfait (fig. 2C) illustre l'allure non bornée du  $c$  de Geary en présence d'une tendance, ainsi que les différentes étapes du « filtrage » de la variabilité effectué par la covariance et la corrélation non ergodiques. En pratique, même dans le cas d'une forte ressemblance, la corrélation non ergodique ne peut pas se maintenir à zéro pour toutes les distances, parce qu'il existe toujours une variabilité résiduelle.

Les variables régionalisées rencontrées dans l'étude des phénomènes géomorphologiques combinent selon différents degrés les situations archétypales illustrées par les figures 1A, 1B et 1C. En effet, les structures spatiales ne sont jamais parfaitement aléatoires (fig. 1A), ni parfaitement homogènes et isotropes (fig. 1B) et encore moins constituées d'une tendance simple, sans variabilité résiduelle (fig. 1C). Un des objectifs de l'analyse de l'autocorrélation spatiale est précisément de contribuer à décrire la structure spatiale, en termes d'écart à l'absence de structure spatiale, d'écart à l'homogénéité (présence d'une tendance) et d'écart à l'isotropie (anisotropie).

#### EXEMPLES DE VARIABLES RÉGIONALISÉES UNIDIMENSIONNELLES : INCISION ET LARGEUR DU LIT DE LA DRÔME (FRANCE)

##### Zone d'étude

La Drôme est une rivière d'une longueur de près de cent kilomètres, affluent de rive gauche du Rhône moyen (fig. 3). Elle draine un bassin de 1 640 km<sup>2</sup>, principalement dessiné dans les pré-alpes calcaires et marneuses du Vercors et du Diois. Avec un module de 18,6 m<sup>3</sup>s<sup>-1</sup> à Saillans, cette rivière à fond graveleux présente un style fluvial en tresses, tout à fait caractéristique dans sa basse vallée. Depuis les années 1945,

la Drôme enregistre une métamorphose : son lit se rétracte et s'incise, en raison d'une modification des pratiques agricoles en lit majeur et du développement de l'activité extractive en lit mineur (Piégay et Landon, 1997 ; Landon *et al.*, 1998).

À la suite de l'étude intégrée conduite en 1995 sur ce bassin, une base de données géomorphologiques a été constituée. Deux cent douze segments de rivière d'une longueur de 500 m ont été définis. Pour chacun d'eux, près de 30 variables géomorphologiques et biogéographiques ont été mesurées (Landon *et al.*, 1995). Ces données ont permis de mettre en lumière des tendances géomorphologiques à la fois synchroniques, amont - aval, et diachroniques, certaines variables ayant été collectées à différentes dates.

Deux variables clés — l'incision du lit entre 1928 et 1996 et sa largeur en 1991 — ont été retenues pour mesurer leur autocorrélation spatiale et décrire leur variabilité amont - aval. L'incision et la largeur du lit tendent à augmenter nettement vers l'aval. Ces tendances ne sont cependant pas parfaitement régulières, plusieurs secteurs s'individualisant (fig. 4).

##### Incision

Le  $c$  de Geary et la covariance non ergodique ne se superposent pas parfaitement, ce qui indique la présence de différences de moyennes locales qui s'organisent selon une tendance (fig. 5A). Cette tendance est du reste bien visible dans la représentation graphique des données (fig. 4A). Les covariogramme et corrélogramme non ergodiques présentent un tracé approximativement parallèle : ceci témoigne de la présence de différences de variances locales, ces différences étant toutefois approximativement constantes quelle que soit la distance considérée (tracés parallèles). Les différences de variance locale constituent une source de variabilité qui est « filtrée » par la corrélation non ergodique, ce qui explique que le corrélogramme soit situé en dessous du covariogramme. Comme cette source de variabilité est quasiment indépendante de la distance, elle n'est pas intéressante à considérer du point de vue de l'analyse de la structure géomorphologique.

Le tracé du  $c$  de Geary témoigne d'une tendance globale et d'un emboîtement de quatre niveaux de structuration : les ruptures se situent approximativement aux kilomètres 13, 25 et 40 (fig. 5A). Cette structure emboîtée est relativement difficile à mettre en évidence par un simple examen des données. Toutefois, on distingue assez clairement deux segments d'une longueur de 40 km qui correspondent à deux tronçons particuliers de la Drôme, de part et d'autre du km 40 (Saillans). Le secteur aval a été le plus intensément exploité par l'industrie des granulats et présente ainsi une forte incision. Il se distingue du secteur amont, plus éloigné des centres de consommation de granulats (Valence, vallée du Rhône), moins exploité et donc moins incisé. La structure d'environ 20-25 km de longueur correspond grosso modo à des unités géomorphologiques élémentaires : km 0 – km 20 (secteur à tressage actif, fortement incisé), km 20 – km 40 (secteur divagant dont le fond de lit est ponctuellement sous la commande des affleurements calcaires), km 40 – km 60 (gorge centrale où le lit est moins exploité et moins sensible à l'enfoncement par érosion

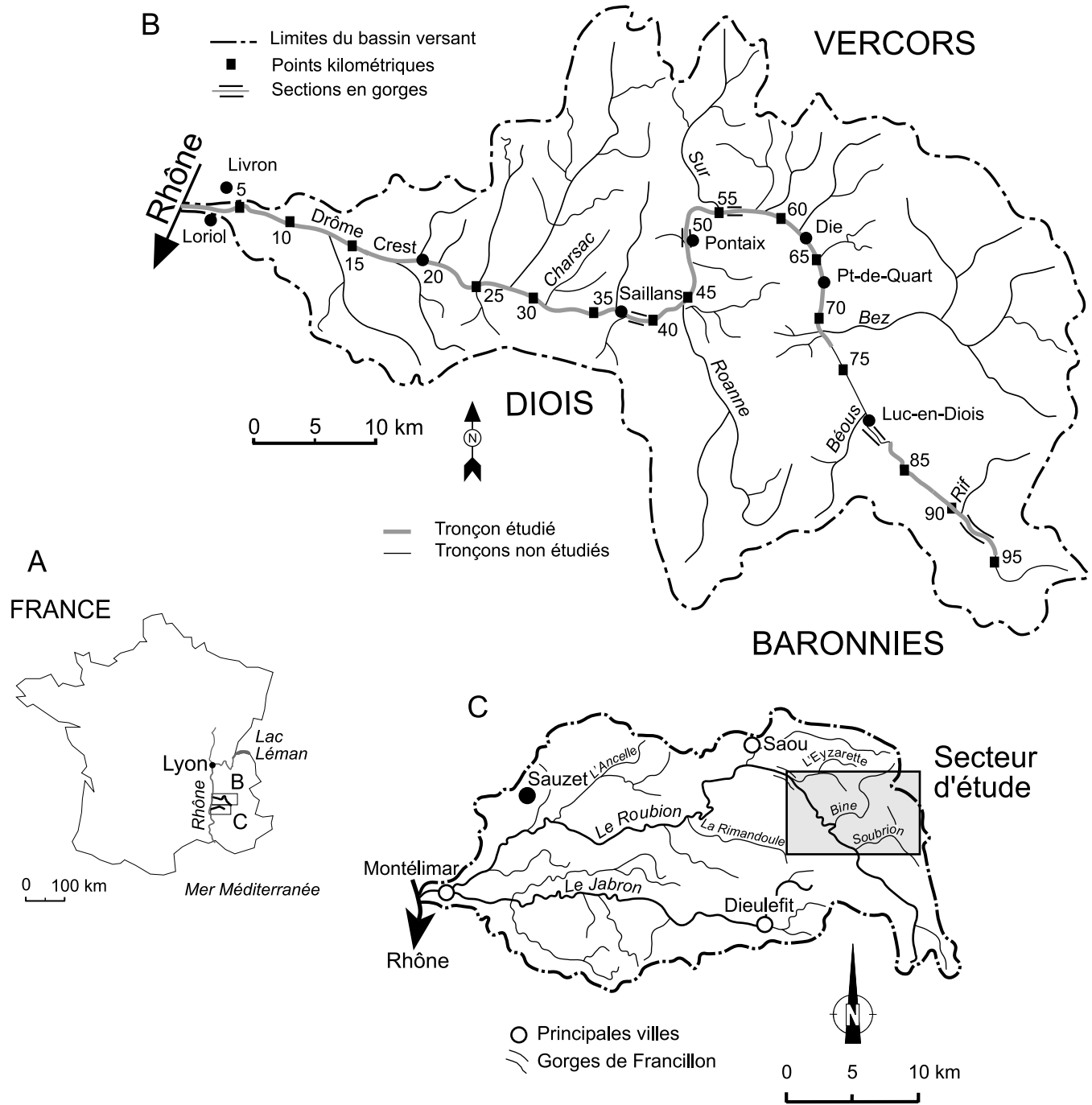


FIGURE 3. Localisation des jeux de données réelles. A) Localisation des sites de la Drôme (B) et du Roubion (C) dans le couloir rhodanien (France). B) Tronçon de la Drôme, depuis sa confluence avec le Rhône jusqu'au km 95. C) Sous-bassins de la Bine et du Soubriion dans le bassin du Roubion.

A) Location of the two studied sites in the southeastern France. B) The reach of the Drôme river from its confluence with the Rhône to km 95. C) The sub-basins of the Bine and the Soubriion within the Roubion basin.

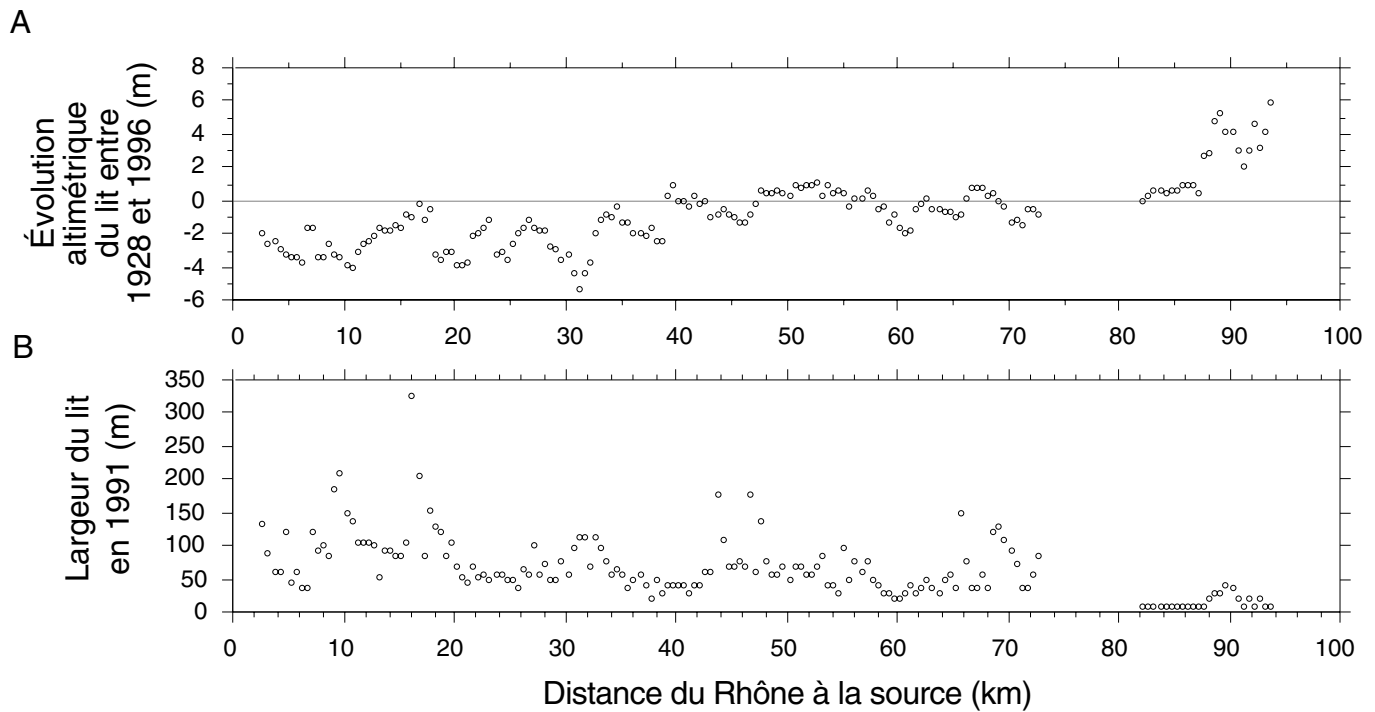


FIGURE 4. Valeurs moyennes de l'incision et de la largeur du lit de la Drôme, depuis sa confluence avec le Rhône jusqu'au km 95, par segments de 500 m. A) Incision du lit de la Drôme entre 1928 et 1996. B) Largeur du lit de la Drôme en 1991.

Mean degradation and width of the active channel of the Drôme river, from its confluence with the Rhône to km 95, per segment of 500 m of river length. A) Active channel degradation between 1928 and 1996. B) Active channel width in 1991.

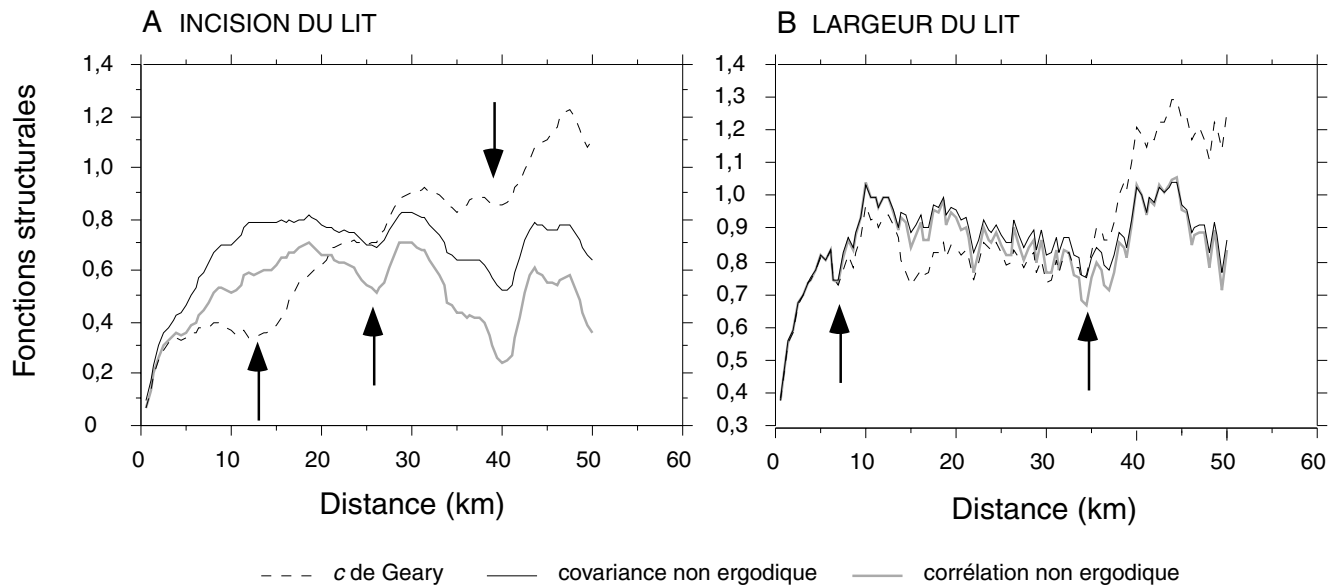


FIGURE 5. Superposition des représentations graphiques du  $c$  de Geary, de la covariance et de la corrélation non ergodiques en fonction de la distance entre segments (km). A) Incision du lit de la Drôme entre 1928 et 1996. B) Largeur du lit de la Drôme en 1991. Les flèches noires indiquent des ruptures de tendance décrites dans le texte.

Evolution of the Geary's  $c$ , non ergodic covariance and correlation according to the distance between the channel segments (km). A) Active channel degradation between 1928 and 1996. B) Active channel width in 1991. The black arrows indicate trend breaking described in the text.

régressive ou progressive), amont km 60 (plaine alluviale intramontagnarde touchée ponctuellement par l'incision) (fig. 3B). La structure la plus élémentaire (10-15 km) distingue au sein de chaque grand segment géomorphologique des sous-ensembles, différemment touchés par l'incision. Le tronçon km 0 - km 10 correspond à une zone de tressage fortement incisée alors que le tronçon km 10 - km 20 est marqué par un gradient : la zone aval est endiguée et soumise à une érosion régressive qui s'atténue à l'amont dans une zone plus naturelle. Entre les km 40 et 50, le contexte est similaire : un gradient oppose la gorge proprement dite, à l'aval, peu touchée par l'incision car le lit est protégé par des blocs provenant des versants, et le tronçon amont, alluvial et plus touché par l'incision. Le km 90 sépare également deux segments :

une plaine alluviale très stable à l'aval et une zone en gorge à l'amont et au sein de laquelle le lit enregistre un exhaussement. La structure spatiale la plus élémentaire distingue ainsi clairement des unités morphologiques d'une dizaine de kilomètres de longueur en fonction de leur sensibilité à l'incision, de la position des zones d'extraction ou des points fixes (affleurements rocheux, seuils artificiels, pavage) déterminant l'enfoncement du lit.

La covariance non ergodique montre également les ruptures identifiées en examinant le  $c$  de Geary. En effet, après la prise en compte des différences de moyennes locales, une portée située entre 10 et 15 km peut être identifiée en examinant l'intervalle de confiance au seuil de 1 % (fig. 6A). À une distance de 25 km, puis de 40 km, la covariance non

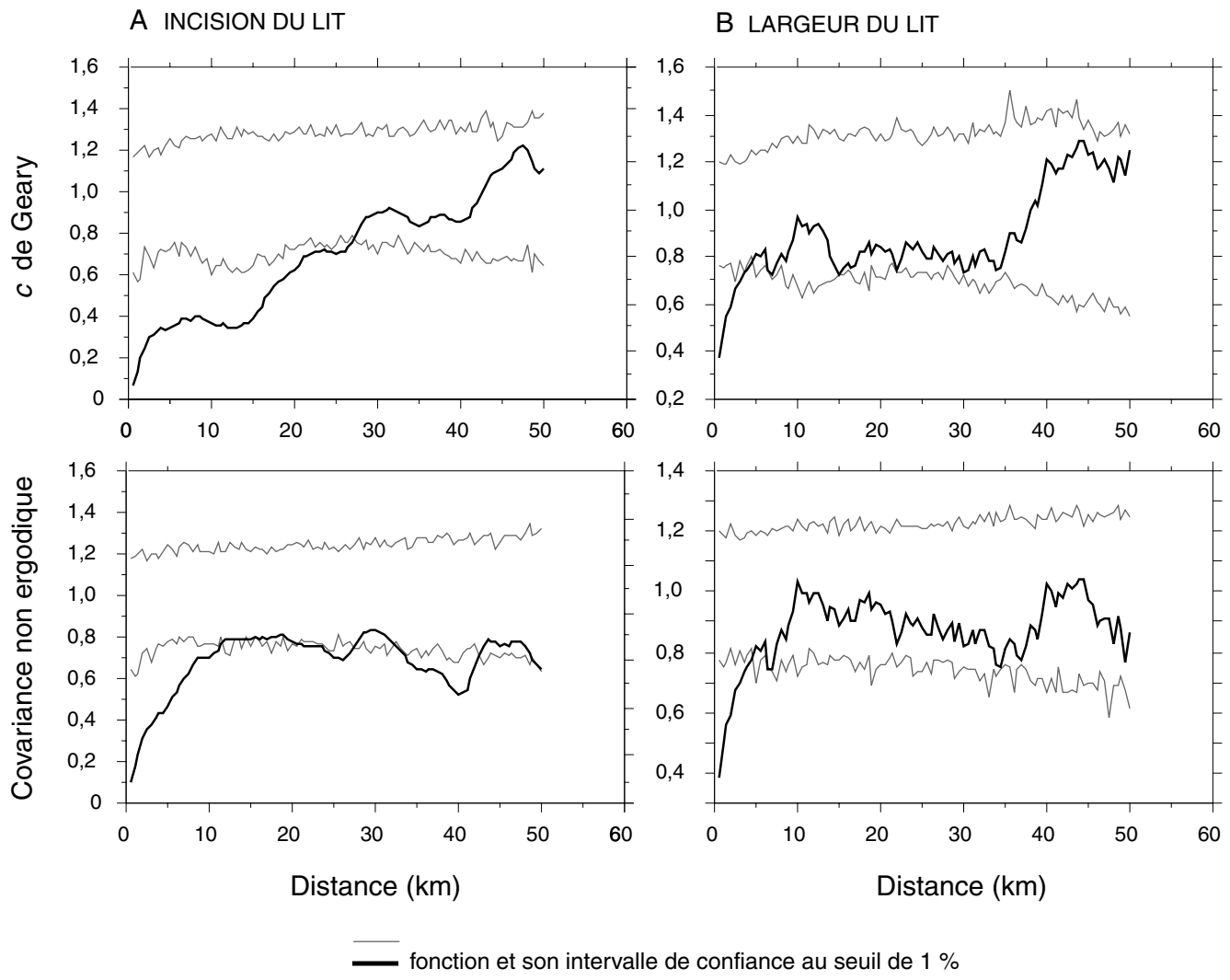


FIGURE 6. Représentation graphique du  $c$  de Geary, de la covariance non ergodique et de la corrélation non ergodique en fonction de la distance entre segments (km) et intervalles de confiance sous  $H_0$  au seuil de 1 % (10 000 permutations aléatoires). A) Incision du lit de la Drôme entre 1928 et 1996. B) Largeur du lit de la Drôme en 1991.

*Evolution of the Geary's  $c$ , the non ergodic covariance and correlation according to the distance between the channel segments (km) and confidence interval under  $H_0$  for a threshold of 1 % (10 000 random permutations). A) Active channel degradation between 1928 and 1996, B) Active channel width in 1991.*

ergodique présente deux pics d'autocorrélation spatiale positive, significative au seuil de 1 % (fig. 6A).

#### Largeur du lit

Dans le cas de la largeur, le  $c$  de Geary met en lumière une tendance (fig. 5B), mais celle-ci est moins nette que celle observée précédemment (fig. 5A). La différence entre le  $c$  de Geary et le couple covariance / corrélation non ergodiques se manifeste essentiellement à partir de 35 km (fig. 5B). En effet, l'examen des données montre que les tronçons séparés par une distance supérieure à 35 km ont des moyennes différentes : par exemple, les segments situés aux alentours du km 85 ont une largeur très différente de ceux localisés vers le km 50, qui ont eux-mêmes une largeur différente de ceux localisés vers le km 15 (fig. 4B). De fait, le  $c$  de Geary distingue clairement le secteur aval, très large, qui correspond à un tronçon de plaine alluviale à l'aval des gorges de Saillans marqué par une tendance à l'élargissement, et le secteur amont qui se caractérise par une alternance de secteurs très étroits (gorges, zones endiguées) et de secteurs plus larges (plaine alluviale plus naturelle). La covariance et la corrélation non ergodiques se superposent, ce qui témoigne de l'absence d'une différence de variance locale. La portée, déterminée à l'aide des intervalles de confiance au seuil de 1 %, est identique, que l'on considère le  $c$  de Geary ou la covariance non ergodique : elle peut être située entre 5 et 10 km (fig. 6B). Ce segment élémentaire est ainsi beaucoup plus court que celui calculé avec les valeurs d'incision. De fait, la largeur semble distinguer des sous-secteurs au sein des segments élémentaires identifiés par l'incision : km 10 - km 20, distinction nette entre la zone endiguée aval (km 10 - km 15) et la zone des ramières (km 15 - km 20), km 40 - km 50, distinction nette entre la gorge aval (km 40 - km 44) et la zone alluviale amont plus large (km 44 - km 50).

#### Discussion

L'analyse conjointe des trois statistiques décrivant la dépendance spatiale permet de mettre en évidence la présence éventuelle d'une tendance, de structures emboîtées (fig. 5A) ou de différences de moyennes locales sur une portion particulière de l'axe des distances (fig. 5B). Avec l'exemple de l'incision du lit, il apparaît que la définition de la portée (au sens large) dépend de la statistique utilisée, puisqu'elle peut être évaluée à environ 28 km dans le cas du  $c$  de Geary, et située entre 10 et 15 km dans le cas de la covariance non ergodique. Avec le  $c$  de Geary, la définition de la portée inclut la tendance, tandis qu'avec la covariance non ergodique, la portée est définie en filtrant les effets de la tendance. Dans le cas de la largeur du lit, la portée définie au sens du  $c$  de Geary ou de la covariance non ergodique est semblable dans la mesure où il n'existe pas de différence de moyenne locale jusqu'à cette distance (fig. 5B).

L'analyse des deux variables (incision et largeur) montre des différences de régularité spatiale : en effet, le tracé des fonctions concernant l'incision est lisse (fig. 5A et 6A), alors que celui concernant la largeur est nettement plus irrégulier (fig. 5B et 6B). Ceci correspond à l'existence de segments

présentant une largeur très différente de celle des segments voisins (par exemple, km 10, km 17, km 44, fig. 4B), tandis que l'incision est un phénomène effectivement beaucoup plus régulier (fig. 4A).

L'analyse de l'autocorrélation des variables élémentaires décrivant la variabilité longitudinale de la géométrie des lits fluviaux permet de mettre en lumière les tendances géomorphologiques d'ensemble, tout en distinguant à différentes échelles spatiales des segments homogènes. De fait, dans le cadre d'une analyse longitudinale, le  $c$  de Geary constitue un outil pertinent permettant de mettre en évidence l'allométrie (au sens de Church et Mark, 1980) au sein du système, notamment l'augmentation de la largeur du lit avec l'augmentation de la taille du bassin et donc du débit, tout en soulignant l'importance des facteurs locaux (affleurements rocheux, morphologie du fond de vallée, position des sites d'extraction et des tronçons endigués) qui contrôlent la géométrie du lit, notamment sa largeur et son niveau d'incision. Ceci montre que dans les bassins versants dont la géologie est contrastée et où les actions humaines s'expriment variablement le long du continuum fluvial, les facteurs locaux — au-delà d'un paramètre de taille tel que la superficie du bassin — jouent un rôle clé, expliquant le fait que les tendances longitudinales s'expriment modestement.

Cette démarche permet ainsi de montrer la complexité longitudinale de l'agencement des formes géomorphologiques — notamment en raison de la distribution complexe des facteurs de contrôle le long du continuum — tout en indiquant au-delà de quelle distance les supports peuvent donner lieu à des données spatialement indépendantes, utilisables dans des analyses statistiques classiques. Dans le cadre de l'étude longitudinale de la géométrie hydraulique, l'analyse de l'autocorrélation spatiale permet ainsi de pouvoir mieux échantillonner les segments sur lesquels une analyse de régression peut ensuite être réalisée afin de prévoir l'évolution de la géométrie du lit en fonction de la taille du bassin versant ou de la distance à la source.

#### EXEMPLE DE VARIABLES RÉGIONALISÉES BIDIMENSIONNELLES : FORMES D'ÉROSION DES SOUS-BASSINS DU ROUBION (FRANCE)

##### Zone d'étude et problématique

Le Roubion est également un affluent du Rhône moyen, dont le bassin est limitrophe de celui de la Drôme (fig. 3C). Il draine le même compartiment géologique que ce dernier : les zones en relief, dont l'altitude ne dépasse pas 1 600 m, sont dessinées dans des calcaires massifs alors que les zones en creux présentent une lithologie dominée par les marnes et les marno-calcaires. Il s'agit d'une région sédimentaire plissée, soumise à d'importantes déformations et présentant un relief inverse typique. Deux de ses sous-bassins, la Bine et le Soubrion, ont fait l'objet d'une étude détaillée pour évaluer quelles étaient les potentialités de recharge du Roubion en sédiments grossiers (Liébault *et al.*, 1998). L'un des problèmes clé sur ce bassin — comme sur celui de la Drôme — est de gérer le déficit sédimentaire du bras principal causé par les

extractions de graviers en lit mineur et d'évaluer la capacité d'autorestoration de la rivière compte tenu des potentialités de son bassin.

L'un des enjeux est de pouvoir identifier quels sont les bassins versants qui présentent des stocks sédimentaires abondants et qui peuvent alimenter durablement le Roubion en sédiments. Dans ce cadre, prévoir la distribution géographique des formes d'érosion à partir de différents descripteurs du bassin, comme la lithologie, les formations superficielles, le couvert végétal, la pente, l'orientation, ou encore l'altitude, pourrait être utile pour évaluer à une échelle régionale plus petite quelles sont les potentialités du bassin en matière d'alimentation du Roubion en charge de fond. Mesurer l'autocorrélation spatiale est ainsi une première étape permettant de savoir s'il est possible de définir un échantillon de valeurs spatialement indépendantes et d'établir un modèle de régression, de façon classique (modèle d'erreurs indépendantes, utilisation des tests).

#### Formes d'érosion

Une cartographie des phénomènes d'érosion se manifestant sur les versants a été réalisée à partir de l'analyse de photographies aériennes et d'une reconnaissance de terrain (Liébault *et al.*, 1998). Ont ainsi été cartographiés les glissements, les loupes de solifluxion, les éboulements et les éboulis, les seconds se distinguant des premiers par le processus de formation, la forme et la taille des particules, ou encore, le ruissellement diffus et le ravinement, ce dernier caractérisant un ruissellement concentré et mieux connecté que le

précédant au réseau hydrographique. La carte des zones d'érosion, réalisée sur Arc Info/Arc View, a été convertie en une image matricielle de 12 000 pixels de 50 m de côté dont un échantillon aléatoire simple de 1 000 pixels a été extrait.

Comme cette région se dessine dans une structure sédimentaire plissée, les formes de relief ont une orientation dominante. Dans le bassin de la Bine et du Soubrion, la lithologie, et par voie de conséquence la topographie, présentent une orientation préférentielle NO-SE (fig. 7A). De fait, la répartition géographique des zones d'érosion, tous processus confondus, répond au contexte topographique et lithologique et les formes de ces zones présentent elles-mêmes une orientation préférentielle (fig. 7B).

La variable décrivant les processus d'érosion est une variable nominale à 8 modalités (fig. 7B). Pour étudier l'autocorrélation spatiale de cette variable, nous utilisons la statistique  $J$ . La modalité 8 correspondant aux entités non soumises à l'érosion, n'est pas prise en compte dans les calculs. La statistique  $J$  est calculée pour un ensemble de 70 classes de distances définies pour des multiples de  $d = 50$  m, avec une tolérance de  $\pm 25$  m. Les calculs sont d'abord effectués en considérant toutes les directions à la fois (fonction omnidirectionnelle). Cependant, la structure spatiale sous-jacente n'étant manifestement pas isotrope, des fonctions directionnelles sont également calculées, pour les directions S-N, E-O, NE-SO et NO-SE, en utilisant une tolérance angulaire de  $\pm 10^\circ$ . Des intervalles de confiance ont été établis pour les fonctions directionnelles, sur la base de 10 000 permutations aléatoires.

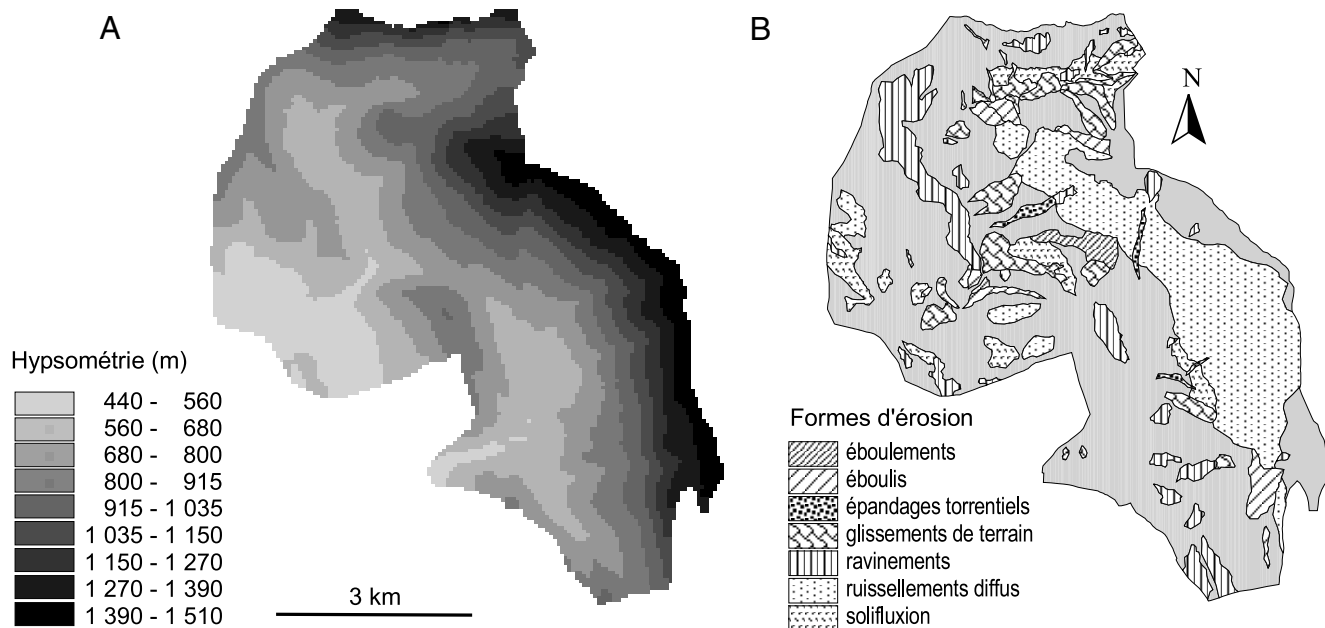


FIGURE 7. Cartographies des sous-bassins de la Bine et du Soubrion. A) Hypsométrie. B) Formes d'érosion.

The sub-basins of the Bine and the Soubrion. A) Relief architecture. B) Distribution of erosion forms.

Compte tenu de la taille du jeu de données traité (1000 pixels), un échantillon de 10 000 permutations aléatoires ne permet pas d'estimer avec précision les queues de distribution sous  $H_0$  de la statistique  $J$ . En conséquence, les intervalles de confiance ont été calculés en utilisant un seuil de 10 %.

L'examen de la fonction  $J$  omnidirectionnelle montre que l'autocorrélation spatiale des processus d'érosion décroît régulièrement, jusqu'à atteindre un palier situé à 1,5 km, avant de croître à nouveau au-delà de 2,5 km (fig. 8A). Effectivement, l'intervalle de confiance au seuil de 10 % confirme qu'en moyenne, les processus d'érosion sont autocorrélés positivement jusqu'à une distance d'environ 1,5 km et autocorrélés négativement au-delà de 2,5 km (fig. 8A).

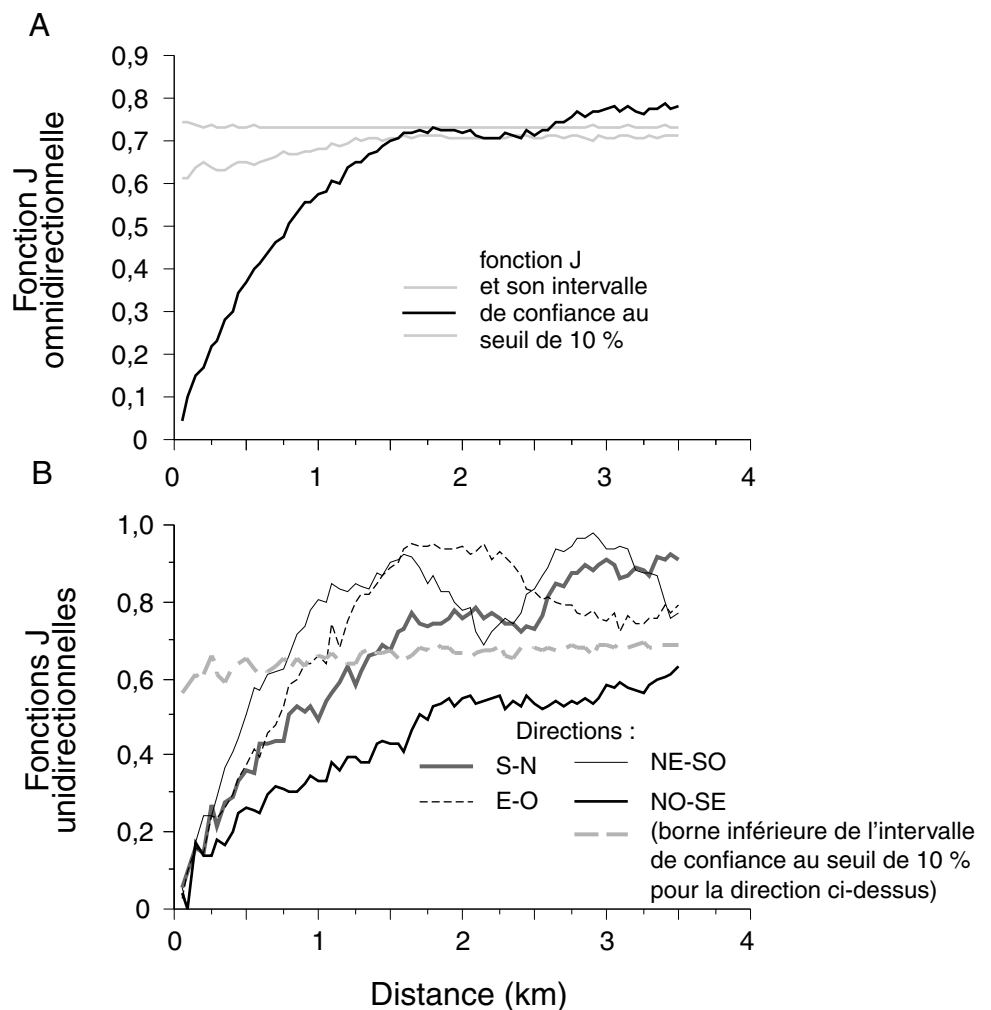
La carte des formes d'érosion montre que la taille des zones homogènes varie fortement, mais excède rarement 2,5 km dans leur plus grande longueur (éboulements et éboulis). Plus les points sont proches, plus ils ont de chance de correspondre au même processus d'érosion (autocorrélation positive). Deux points distants l'un de l'autre d'une distance

comprise entre 1,5 km et 2,5 km ont, en moyenne, autant de chance de correspondre au même processus d'érosion qu'à deux processus différents (autocorrélation non significative). Au-delà de 2,5 km de séparation, les deux points sont très souvent situés dans des zones délimitant des formes d'érosion différentes (autocorrélation négative).

L'analyse directionnelle de l'autocorrélation montre une forte anisotropie de la distribution spatiale des processus d'érosion. En effet, la représentation graphique de la statistique  $J$  diffère selon l'orientation (fig. 8B). Parmi les quatre orientations envisagées (S-N, E-O, NE-SO et NO-SE), la direction NO-SE présente une fonction  $J$  nettement distincte des autres qui correspond à l'orientation préférentielle des formes de relief, et par conséquent, à celle des processus d'érosion. La fonction  $J$  de direction NO-SE n'atteint jamais de portée, l'autocorrélation positive étant toujours significative. En revanche, la fonction  $J$  de direction NE-SO — direction perpendiculaire à l'orientation préférentielle des formes de relief — a la portée la plus courte, les deux autres fonctions  $J$  présentant une allure intermédiaire.

FIGURE 8. Représentation graphique de la statistique  $J$  pour la variable régionalisée « formes d'érosion » en fonction de la distance entre supports. A) Fonction  $J$  omnidirectionnelle et intervalle de confiance sous  $H_0$  au seuil de 10 % (10 000 permutations aléatoires). B) Fonction  $J$  directionnelle pour les directions suivantes : S-N, E-O, NE-SO, NO-SE (tolérance angulaire de  $\pm 10^\circ$ ) et intervalle de confiance sous  $H_0$  au seuil de 10 % (10 000 permutations aléatoires) pour la direction NO-SE.

*Evolution of the J statistics for the variable "erosion forms" in relation with the distance between pixels. A) Omnidirectional J function and confidence interval under  $H_0$  for a threshold of 10 % (10 000 random permutations). B) Directional J function for the following orientations: S-N, E-W, NE-SW, NW-SE (angular tolerance  $\pm 10^\circ$ ) and confidence interval under  $H_0$  for a threshold of 10 % (10 000 random permutations) for the NW-SE orientation.*





## Discussion

La conversion en une image matricielle de la carte des formes d'érosion des bassins versants de la Bine et du Soubrion a généré une variable régionalisée nominale dont l'autocorrélation spatiale a pu être mesurée à l'aide de la statistique  $J$ ; ceci illustre le fait que ce type d'analyse statistique n'est pas limité aux seules données quantitatives.

Dans le cas de données bidimensionnelles, une analyse directionnelle est indispensable afin de vérifier la validité de l'hypothèse d'isotropie. Une allure nettement distincte des fonctions d'autocorrélation témoigne de l'existence d'une anisotropie. En l'occurrence, la direction préférentielle NO-SE apparaît très clairement sur la figure 8B, conformément à l'appréciation visuelle de la carte des formes d'érosion (fig. 7B).

Dans le cadre de la définition d'un échantillon visant à garantir l'indépendance spatiale des données, le recours à une fonction omnidirectionnelle afin de déterminer la portée de l'autocorrélation spatiale n'est valide qu'en situation d'isotropie. En l'occurrence, la direction préférentielle NO-SE des processus d'érosion ne permet pas de définir une portée maximale compatible avec la taille du domaine d'étude, qui puisse garantir l'indépendance spatiale quelle que soit la direction. Cet exemple montre qu'il n'est pas toujours possible de procéder de façon simple à un échantillonnage garantissant l'indépendance spatiale des données. En conséquence, une approche alternative au problème posé par l'autocorrélation spatiale des données consiste à la modéliser afin d'en tenir compte dans les procédures statistiques elles-mêmes.

## CONCLUSION

La nature régionalisée des variables étudiées par le géomorphologue l'oblige à tenir compte de l'autocorrélation spatiale de ses données avant de construire un modèle statistique prédictif. Les exemples étudiés dans cet article ont montré l'intérêt d'une analyse de l'autocorrélation spatiale mettant en œuvre simultanément plusieurs statistiques, afin de mettre en évidence différentes caractéristiques des structures spatiales. Cette pratique, renforcée par le recours aux intervalles de confiance établis par une procédure de permutation aléatoire des données, permet de mieux appréhender la complexité spatiale des formes et leur structuration à différentes échelles. Toutefois, la définition objective de la portée de l'autocorrélation spatiale constitue un problème délicat. En effet, la portée est définie conditionnellement au type de fonction d'autocorrélation, de sorte qu'un résultat différent peut être obtenu selon la statistique qui est utilisée. Le second exemple réel exposé a notamment illustré le fait que l'isotropie des phénomènes est rare en géomorphologie. En conséquence, toute étude menée dans un espace bidimensionnel doit s'appuyer sur une analyse directionnelle de l'autocorrélation spatiale. Si le principal objectif de cette analyse consiste à déterminer un pas d'échantillonnage garantissant l'indépendance spatiale des données, il convient alors de s'appuyer sur la direction pour laquelle la portée est la plus élevée. Cependant, le pas d'échantillonnage qui est déterminé de cette façon n'est pas nécessairement compatible avec la taille du domaine d'étude et la taille de l'échantillon à constituer. Dans ce cas, il

est difficile de générer un jeu de données spatialement indépendantes et il devient nécessaire de remplacer les procédures classiques qui supposent l'indépendance spatiale, par des procédures statistiques tenant compte de l'autocorrélation. L'analyse de l'autocorrélation spatiale se révèle donc indispensable, comme préalable à la construction d'un modèle statistiquement valide.

## REMERCIEMENTS

Les auteurs remercient vivement Norbert Landon, Frédéric Liébault et Fabrice Taillefumier de l'UMR 5600 du CNRS pour avoir participé aux études conduites sur la Drôme (1995) ou sur le Roubion (1998) et à la construction des jeux de données réelles utilisés dans cet article. Ils remercient également Pierre Richard, Michael Church, Benoît Saint-Onge et un lecteur anonyme, dont les commentaires constructifs ont permis d'améliorer le manuscrit.

## RÉFÉRENCES

- Andriamahefa, H., 1999. Les hydro-écorégions du bassin de la Loire : morphologie, hydrologie, pressions anthropiques sur les cours d'eau et les bassins versants. Thèse de Géographie et Aménagement, Université Jean Monnet de Saint-Étienne, 272 p.
- Anselin, L., 1988. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, 20 : 1-17.
- \_\_\_\_\_, 1995. Local indicators of spatial association. *Geographical Analysis*, 27 : 93-115.
- Arbia, G., 1993. The use of GIS in spatial statistical surveys. *International Statistical Review*, 61 : 339-359.
- Biron, P.M., Lane, S.N., Roy, A.G., Bradbrook, K.F. et Richards, K.S., 1998. Sensitivity of bed shear stress estimated from vertical velocity profiles : The problem of sampling resolution. *Earth Surface Processes and Landforms*, 16 : 427-445.
- Boots, B.N. et Dufournaud, C., 1994. A programming approach to minimizing and maximizing spatial autocorrelation statistics. *Geographical Analysis*, 26 : 54-66.
- Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V. et Reichenback, P., 1991. GIS techniques and statistical models in evaluating landslide hazard. *Earth Surface Processes and Landforms*, 16 : 427-445.
- Cerrioli, A., 1997. Modified tests of independence in 2 x 2 tables with spatial data. *Biometrics*, 53 : 619-628.
- Chessel, D., 1981. The spatial autocorrelation matrix. *Vegetatio*, 46 : 177-180.
- Chessel, D. et Sabatier, R., 1993. Couplage de triplets statistiques et graphes de voisinage, p. 28-37. In B. Asselain, M. Boniface, C. Duby, C. Lopez, J.P. Masson et J. Tranchefort, édit., *Biométrie et analyse de données spatio-temporelles*. Société française de Biométrie, ENSA, Rennes, 123 p.
- Church, M. et Mark, D.M., 1980. On size and scale in geomorphology. *Progress in Physical Geography*, 4(3) : 342-390.
- Clément, P. et Piégay, H., sous presse. Statistics and fluvial geomorphology. In M.G. Kondolf et H. Piégay, édit., *Tools in Fluvial Geomorphology*, Wiley, Chichester.
- Cliff, A.D. et Ord, J.K., 1969. The problem of spatial autocorrelation, p. 25-55. In A.J. Scott, édit., *London Papers in Regional Science*. 1. *Studies in Regional Science*. Pion, London, 216 p.
- \_\_\_\_\_, 1971. Evaluating the percentage points of a spatial autocorrelation coefficient. *Geographical Analysis*, 3 : 51-61.
- \_\_\_\_\_, 1973. *Spatial Autocorrelation*. Pion, London, 178 p.
- \_\_\_\_\_, 1977. Large sample-size distribution of statistics used in testing for spatial correlation: A comment. *Geographical Analysis*, 9 : 297-299.
- \_\_\_\_\_, 1981. *Spatial Processes. Models and Applications*. Pion, London, 266 p.

- Clifford, N.J., Robert, A. et Richards, K.S., 1992. Estimation of flow resistance in gravel-bedded rivers: a physical explanation of the multiplier of roughness length. *Earth Surface Processes and Landforms*, 17 : 111-126.
- Croft, F. et Kessler, B., 1996. Remote sensing, image processing, and GIS. Trends and forecasts. *Journal of Forestry*, 94 : 31-35.
- David, M., 1977. *Geostatistical Ore Reserve Estimation*. Elsevier, New York, 364 p.
- Deichman, U. et Anselin, L., 1994. Exploratory spatial data analysis of categorical variables: An application to African farming systems data, p. 2107-2116. *In* J.J. Harts, H.F.L. Ottens, H.J. Scholten et J. Van Arragon, édité., *EGIS/MARI '94. Fifth European Conference and Exhibition on Geographical Information Systems*. EGIS Foundation, Utrecht.
- Deutsch, C.V. et Journel, A.G., 1992. *GSLIB. Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 340 p.
- Downs, P.W., 1994. Characterization of river channel adjustments in the Thames basin, south-east England. *Regulated rivers. Research and Management*, 9 : 151-175.
- Durbin, J. et Watson, G.S., 1950. Testing for serial correlation in least squares regression. I. *Biometrika*, 37 : 409-428.
- Dutilleul, P., 1993. Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, 49 : 305-314.
- Edgington, E.S., 1986. Randomization tests, p. 530-538. *In* S. Kotz et N.L. Johnson, édité., *Encyclopedia of Statistical Sciences*, vol. 7. Wiley, New York, 714 p.
- \_\_\_\_\_. 1987. *Randomization Tests*. Second edition. Marcel Dekker, New York, 341 p.
- Ferguson, R.I., 1975. Meander irregularity and wavelength estimation. *Journal of Hydrology*, 26 : 315-333.
- Gabriel, K.R. et Sokal, R.R., 1969. A new statistical approach to geographic variation analysis. *Systematic Zoology*, 18 : 259-278.
- Gatrell, A.C., 1979. Autocorrelation in spaces. *Environment and Planning A*, 11 : 507-516.
- Geary, R.C., 1954. The contiguity ratio and statistical mapping (with discussion). *Incorporated Statistician (The)*, 5 : 115-145.
- Getis, A., 1989. A spatial association model approach to the identification of spatial dependence. *Geographical Analysis*, 21 : 251-259.
- Gilvear, D.J., Bryant, R. et Hardy, T., 1999. Remote sensing of channel morphology and in-stream fluvial processes. *Progress in Environmental Science*, 1 : 257-284.
- Gordon, A.D. et Finden, C.R., 1985. Classification of spatially-located data. *Computational Statistics Quarterly*, 2 : 315-328.
- 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.
- Gurnell, A.M., 1997. Adjustments in river channel geometry associated with hydraulic discontinuities across the fluvial-tidal transition of a regulated river. *Earth Surface Processes and Landforms*, 22 : 967-985.
- Haining, R., Ma, J. et Wise, S., 1996. Design of a software system for interactive spatial statistical analysis linked to a GIS. *Computational Statistics Quarterly*, 11 : 449-466.
- Hill, M.O., 1973. The intensity of spatial pattern in plant communities. *Journal of Ecology*, 61 : 225-235.
- Hubert, L.J. et Arabie, P., 1991. The assessment of spatial autocorrelation through constrained multiple regression. *Geographical Analysis*, 23 : 95-111.
- Ichim, I. et Radoane, M., 1990. Channel sediment variability along a river: A case study of the Siret River (Romania). *Earth Surface Processes and Landforms*, 15 : 211-225.
- Isaaks, E.H. et Srivastava, R.M., 1988. Spatial continuity for probabilistic and deterministic geostatistics. *Mathematical Geology*, 20 : 313-341.
- \_\_\_\_\_. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York, 561 p.
- Journel, A.G., 1985. The deterministic side of geostatistics. *Mathematical Geology*, 17 : 1-15.
- \_\_\_\_\_. 1988. New distance measures: the route toward truly non-Gaussian geostatistics. *Mathematical Geology*, 20 : 459-475.
- Journel, A.G. et Huijbregts, C.J., 1978. *Mining Geostatistics*. Academic Press, London, 600 p.
- Jumars, P.A., Thistle, D. et Jones, M.L., 1977. Detecting two-dimensional spatial structure in biological data. *Oecologia*, 28 : 109-123.
- Kirkbride, A.D. et Ferguson, R., 1995. Turbulent flow structure in a gravel-bed river: Markov chain analysis of the fluctuating velocity profile. *Earth Surface Processes and Landforms*, 20 : 721-733.
- Kooijman, S.A.L.M., 1976. Some remarks on the statistical analysis of grids especially with respect to ecology. *Annals of Systems Research*, 5 : 113-132.
- Landon, N., Piégay, H. et Bravard, J.P., 1995. Mission d'expertise géomorphologique : propositions pour une gestion physique équilibrée du lit de la Drôme. *Commission locale de l'Eau du Schéma d'Aménagement et de Gestion des Eaux de la Drôme, Rapport*, 105 p.
- \_\_\_\_\_. 1998. The Drôme River incision (France): From assessment to management. *Landscape and Urban Planning*, 43 : 119-131.
- Latulippe, C. et Peiry, J.-L., 1996. Essai de hiérarchisation des zones de production de matières en suspension dans le bassin-versant d'un grand cours d'eau de l'Isère en amont de Grenoble. *Revue de Géographie alpine*, 2 : 29-44.
- Legendre, P. et Fortin, M.-J., 1989. Spatial pattern and ecological analysis. *Vegetatio*, 80 : 107-138.
- Liébault, F., Clément, P., Piégay, H. et Landon, N., 1999. Assessment of bedload supply potentiality from the tributary watersheds of a degraded river: The Drôme (France). *Arctic, Antarctic and Alpine Research*, 31 : 108-117.
- Liébault, F., Piégay, H. et Taillefumier, F., 1998. Évaluation des potentialités de recharge du Haut-Roubion à partir d'une analyse fine des caractéristiques géographiques du bassin et de deux affluents : la Bine et le Soubriou. *Syndicat mixte d'Aménagement rural de la Drôme et Syndicat intercommunal du Bassin du Roubion, Rapport*, 90 p.
- Liebold, A.M., Rossi, R.E. et Kemp, W.P., 1993. Geostatistics and Geographic Information Systems in applied insect ecology. *Annual Review of Entomology*, 38 : 303-327.
- Ludwig, J.A. et Goodall, D.W., 1978. A comparison of paired-with-blocked-quadrat variance methods for the analysis of spatial pattern. *Vegetatio*, 38 : 49-59.
- Madej, M.A., 1999. Temporal and spatial variability in thalweg profiles of a gravel-bed river. *Earth Surface Processes and Landforms*, 24 : 1153-1169.
- Matheron, G., 1965. *Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux sciences de la nature*. Masson, Paris, 305 p.
- Mertès, L.A.K., Daniel, D.L., Melack, J.M., Nelson, B., Martinelli, A. et Forsberg, B.R., 1995. Spatial patterns of hydrology, geomorphology, and vegetation on the floodplain of the Amazon river in Brazil from a remote sensing perspective. *Geomorphology*, 13 : 215-232.
- Montgomery, D.R., Dietrich, W.E. et Sullivan, K., 1998a. The role of GIS in watershed analysis, p. 241-261. *In* S.N. Lane, K.S. Richards et J.H. Chandler, édité., *Landform, Monitoring, Modelling and Analysis*, Wiley, Chichester, 454 p.
- Montgomery, D.R., Sullivan, K. et Greenberg, H.M., 1998b. Regional test of a model for shallow landsliding. *Hydrological Processes*, 12 : 943-955.
- Moran, P.A.P., 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society Series B*, 10 : 243-251.
- \_\_\_\_\_. 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37 : 17-23.
- Myers, D.E., 1989. To be or not to be stationary? That is the question. *Mathematical Geology*, 21 : 347-362.
- Myers, J.C., 1997. *Geostatistical Error Management. Quantifying Uncertainty for Environmental Sampling and Mapping*. Van Nostrand Reinhold, New York, 571 p.

- Newson, M., 1994. Sustainable integrated development and the basin sediment system: guidance from fluvial geomorphology, p. 1-10. *In* C. Kirby et W.R. White, édit., *Integrated River Basin Development*, Wiley, Chichester, 537 p.
- Okabe, A., 1976. A note on Geary's spatial contiguity ratio. *Geographical Analysis*, 8 : 315-318.
- Ord, J.K. et Getis, A., 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, 27 : 286-306.
- Pettitt, A.N., 1982. Durbin-Watson test, p. 426-428. *In* S. Kotz et N.L. Johnson, édit., *Encyclopedia of Statistical Sciences*, Vol. 2. Wiley, New York, 613 p.
- Piégay, H., Bornette, G. et Grante, P., sous presse. Assessment of silting-up dynamics of 11 cut-off channel plugs on a free-meandering river (the Ain River), France. *In* R.J. Allison, édit., *Applied Geomorphology, Theory and Practice*, Wiley, Chichester.
- Piégay, H. et Landon, N., 1997. Promoting an ecological management of riparian forests on the Drôme River, France. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 7 : 287-304.
- Rickenmann, D., 1997. Sediment transport in Swiss torrents. *Earth Surface Processes and Landforms*, 22 : 937-951.
- Ripley, B.D., 1988. Spatial data analysis, p. 570-573. *In* S. Kotz et N.L. Johnson, édit., *Encyclopedia of Statistical Sciences*, vol. 8. Wiley, New York, 870 p.
- Rossi, R.E., Mulla, D.J., Journel, A.G. et Franz, E.H., 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, 62 : 277-314.
- Royaltey, H.H., Astrachan, E. et Sokal, R.R., 1975. Tests for patterns in geographic variation. *Geographical Analysis*, 7 : 369-395.
- Schumm, S.A., 1977. *The Fluvial System*. Wiley, New York, 338 p.
- Sen, A., 1976. Large sample-size distribution of statistics used in testing for spatial correlation. *Geographical Analysis*, 8 : 175-184.
- \_\_\_\_\_, 1977. "Large sample-size distribution of statistics used in testing for spatial correlation": A reply. *Geographical Analysis*, 9 : 300.
- Simon, G., 1997. An angular version of spatial correlations, with exact significance tests. *Geographical Analysis*, 29 : 267-278.
- Sokal, R.R., 1979. Ecological parameters inferred from spatial correlograms, p. 167-196. *In* G.P. Patil et M. Rosenzweig, édit., *Contemporary quantitative ecology and related econometrics*. International Co-operative Publishing House, Fairland, Maryland, 695 p.
- Sokal, R.R. et Oden, N.L., 1978. Spatial autocorrelation in biology. 1 Methodology. *Biological Journal of the Linnean Society*, 10 : 199-228.
- Terui, N. et Kikuchi, M., 1994. The size-adjusted critical region of Moran's I test statistics for spatial autocorrelation and its application to geographical areas. *Geographical Analysis*, 26 : 213-227.
- Thioulouse, J., Chessel, D. et Champely, S., 1995. Multivariate analysis of spatial patterns: A unified approach to local and global structures. *Environmental and Ecological Statistics*, 2 : 1-14.
- Tiefelsdorf, M. et Boots, B., 1997. A note on the extremities of local Moran's  $I_s$  and their impact on the global Moran's  $I^*$ . *Geographical Analysis*, 29 : 248-257.
- Turner, S.J., O'Neill, R.V., Conley, W., Conley, M.R. et Humphries, H.C., 1991. Pattern and scale: statistics for landscape ecology, p. 17-49. *In* M.G. Turner et R.H. Gardner, édit., *Quantitative Methods in Landscape Ecology*. Ecological Studies, Vol. 82, Springer-Verlag, New York, 536 p.
- Ver Hoef, J.M., Cressie, N.A.C. et Glenn-Lewin, D.C., 1993. Spatial models for spatial statistics: Some unification. *Journal of Vegetation Science*, 4 : 441-452.
- von Neumann, J., 1941. Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics*, 12 : 367-395.
- Waldhör, T., 1996. The spatial autocorrelation coefficient Moran's I under heteroscedasticity. *Statistics in Medicine*, 15 : 887-892.
- Walker, D.D., Loftis, J.C. et Mielke, P.W., 1997. Permutation methods for determining the significance of spatial dependence. *Mathematical Geology*, 29 : 1011-1024.
- Walter, S.D., 1974. On the detection of household aggregation of disease. *Biometrics*, 30 : 525-538.

- Zhang, C.S. et Selinus, O., 1997. Spatial analyses for copper, lead and zinc contents in sediments of the Yangtze River basin. *Science of the Total Environment*, 204 : 251-262.

## ANNEXE I

### LOGICIELS STATISTIQUES ET DONNÉES SPATIALES

Certains logiciels statistiques commerciaux offrent des possibilités en matière d'analyse et de modélisation spatiale, c'est le cas de Systat, qui comporte, au même titre que des outils d'analyse des séries chronologiques, un module de *Spatial statistics*, en fait strictement limité à la géostatistique. C'est également le cas de S+SpatialStats qui constitue un module supplémentaire de S+, permettant notamment de calculer certaines statistiques d'autocorrélation spatiale telles que le  $c$  de Geary et le  $I$  de Moran. Toutefois, ces logiciels n'offrent ni toutes les mesures exposées dans cet article ni la possibilité de réaliser des tests de randomisation et des intervalles de confiance non paramétriques comme nous le proposons.

Par ailleurs, à notre connaissance, il n'existe aucun logiciel statistique commercialisé qui adapte les procédures statistiques classiques au traitement des variables régionalisées autocorrélées, par exemple le test de la corrélation entre deux variables quantitatives (Dutilleul, 1993), ou bien un test d'association pour variables binaires (Cerioli, 1997).

## ANNEXE II

### LIENS ENTRE LE $c$ DE GEARY ET D'AUTRES STATISTIQUES

Le  $c$  de Geary (Geary, 1954) est une généralisation à l'espace bidimensionnel du rapport  $\eta$  de von Neumann défini pour les séries temporelles (von Neumann, 1941). Par ailleurs, le  $\eta$  de von Neumann est algébriquement identique à la statistique  $d$  introduite par Durbin et Watson (1950) pour tester l'autocorrélation sérielle du terme d'erreur dans un modèle linéaire (Pettitt, 1982). Ainsi, le  $c$  de Geary est également lié au  $d$  de Durbin-Watson (Sokal et Oden, 1978). Okabe (1976) propose une interprétation du  $c$  de Geary basée sur l'analyse de variance et Chessel (1981) mentionne les relations entre le  $c$  de Geary, l'ANOVA hiérarchique proposée par Greig-Smith (1952) dans le domaine de l'écologie statistique, et la statistique de Walter (1974).

En considérant uniquement les couples de supports  $(i, j)$  séparés par un vecteur  $\mathbf{h}_{ij} = \mathbf{h}$ , la matrice  $\mathbf{W}$  définit une classe de vecteurs, et sous cette forme, le  $c$  de Geary n'est pas autre chose que le variogramme expérimental standardisé par une estimation de la variance théorique. En conséquence, à l'instar des liens étroits existant entre le  $c$  de Geary et l'ANOVA hiérarchique proposée par Greig-Smith (Chessel, 1981), le variogramme — équivalent au  $c$  de Geary, à un facteur de standardisation près — présente de profondes analogies avec des statistiques proposées en écologie telles que la TTLV (*Two-Term Local Variance*) de Hill (1973), ou la PQV (*Paired-Quadrat Variance*) de Ludwig et Goodall (1978) (Ver Hoef *et al.*, 1993).

### ANNEXE III

#### LIENS ENTRE LE I DE MORAN ET D'AUTRES STATISTIQUES

En considérant uniquement les couples de supports  $(i, j)$  séparés par un vecteur  $\mathbf{h}_{ij} = \mathbf{h}$ , la matrice  $\mathbf{W}$  définit une classe de vecteurs, et le  $I$  de Moran s'écrit comme la fonction :

$$I(\mathbf{h}) = \frac{\hat{C}(\mathbf{h})}{s_n^2} \quad (12)$$

avec  $s_n^2 = n^{-1} \sum_i (z_i - \bar{z})^2$  la variance des valeurs, et  $\hat{C}(\mathbf{h})$  la fonction de covariance spatiale empirique :

$$\hat{C}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{(i,j) \mid \mathbf{h}_{ij} = \mathbf{h}} (z_i - \bar{z})(z_j - \bar{z}) \quad (13)$$

Sous cette forme, le  $I$  de Moran apparaît comme la covariance spatiale standardisée par la variance des valeurs.

Le lien entre le  $I$  de Moran et le  $c$  de Geary a été étudié par Sokal (1979). Il s'avère que si le  $c$  de Geary montre une autocorrélation spatiale positive moins forte que le  $I$  de Moran ou une autocorrélation spatiale négative plus forte que le  $I$  de Moran, cela peut indiquer que les valeurs extrêmes sont spatialement structurées, au sens de la matrice de pondération  $\mathbf{W}$  utilisée (Sokal, 1979).

Par ailleurs, des résultats donnés dans Cliff et Ord (1969, 1981) indiquent que la variance du  $I$  de Moran est moins affectée par la distribution statistique des valeurs que le  $c$  de Geary. L'expérience montre que le  $c$  de Geary et le  $I$  de Moran conduisent à des résultats similaires dans le cas d'une distribution symétrique, mais qu'une forte asymétrie peut entraîner un désaccord entre les deux statistiques. Ce désaccord est d'autant plus marqué que l'autocorrélation spatiale est tenue et le  $c$  de Geary s'avère davantage affecté par l'asymétrie de la distribution que le  $I$  de Moran.

L'influence du choix de la statistique sur la conclusion du test de l'autocorrélation spatiale est un phénomène inévitable mais indésirable. En conséquence, il est judicieux d'utiliser les deux statistiques conjointement : a) une conclusion simi-

laire constitue l'assurance que le résultat n'est pas artefactuel et est bien indépendant de la statistique utilisée ; b) une conclusion contradictoire doit nécessairement conduire le géomorphologue à approfondir son analyse.

### ANNEXE IV

#### PRINCIPE DES TESTS DE PERMUTATION ET DE RANDOMISATION

Soit une statistique  $Z$  et soit  $\Pi$  l'ensemble des  $n!$  permutations des indices  $i = 1, 2, \dots, n$  ; la distribution de  $Z$  sous  $\mathbf{H}_0$  est obtenue en assignant une probabilité égale de  $1/n!$  aux valeurs de  $Z$  calculées pour les  $n!$  permutations des valeurs  $\{z_i \mid i = 1, \dots, n\}$ . La distribution de  $Z$  sous  $\mathbf{H}_0$  étant établie conditionnellement aux données, le test ne concerne que les données elles-mêmes et pas une population statistique dont seraient issues les données. À partir de la distribution exacte de  $Z$  sous  $\mathbf{H}_0$ , la  $p$ -value associée à la valeur observée  $Z_{obs}$  est calculée comme la proportion de valeurs supérieures ou égales à  $Z_{obs}$  (ou inférieures ou égales, selon l'hypothèse alternative considérée) :

$$p = \frac{\text{Card}(\{Z \mid Z \geq Z_{obs}\})}{n!} \quad (14)$$

Le test permutatif peut être envisagé pour  $n \leq 10$ . Dans la plupart des cas ( $n > 10$ ), l'énumération de toutes les permutations est impraticable et le test permutatif est estimé par un test de randomisation (Edgington, 1986, 1987), en construisant un ensemble de valeurs  $\Omega$  obtenu par échantillonnage aléatoire et avec remise de  $\Pi$ . La  $p$ -value obtenue dans le test de randomisation est une estimation sans biais de la  $p$ -value exacte :

$$\hat{p} = \frac{\text{Card}(\{Z \mid Z \in \Omega, Z \geq Z_{obs}\})}{\text{Card}(\Omega)} \quad (15)$$

avec la valeur  $Z_{obs}$  figurant dans l'ensemble  $\Omega$  pour les permutations considérées. La distribution d'échantillonnage de cette  $p$ -value est asymptotiquement normale, de moyenne  $p$  et de variance  $p(1-p)/m$  où  $m = \text{Card}(\Omega)$ .