Origine des structures spatiales en écologie
Conséquences des structures spatiales pour les tests statistiques

Pierre Legendre

Cours *Biostatistique et écologie numérique*

Ifremer, Brest, 6-10 juin 2005
Types of spatial structures

A spatial structure may appear in a variable y because the process that has produced the values of y is spatial and has generated autocorrelation in the data; or it may be caused by the dependence of y upon one or several causal variables x which are spatially structured; or both processes. In both cases, spatial correlation will be found when analyzing the data (Chapter 11). The spatially-structured causal variables x may be explicitly identified in the model, or not; see Table 13.3.

Model 1: autocorrelation — The value \( y_j \) observed at site j on the geographic surface is assumed to be the overall mean of the process \( \mu_y \) plus a weighted sum of the centred values \( (y_i - \mu_y) \) at surrounding sites i, plus an independent error term \( \epsilon_j \):

\[
y_j = \mu_y + \sum f (y_i - \mu_y) + \epsilon_j
\] (1.1)

The \( y_i \)'s are the values of y at other sites i located within the zone of spatial influence of the process generating the autocorrelation (Fig. 1.4). The influence of neighbouring sites may be given, for instance, by weights \( w_i \) which are function of the distance between sites i and j (eq. 13.19); other functions may be used. The total error term is \( \{ \sum f (y_i - \mu_y) + \epsilon_j \} \); it contains the autocorrelated component of variation. As written here, the model assumes stationarity (Subsection 13.1.1). Its equivalent in time series analysis is the autoregressive (AR) response model (eq. 12.30).

Model 2: spatial dependence — If one can assume that there is no autocorrelation in the variable of interest, the spatial structure may result from the influence of some explanatory variable(s) exhibiting a spatial structure. The model is the following:

\[
y_j = \mu_y + f (\text{explanatory variables}) + \epsilon_j
\] (1.2)

where \( y_j \) is the value of the dependent variable at site j and \( \epsilon_j \) is an error term whose value is independent from site to site. In such a case, the spatial structure, called
Origin of spatial structures for a variable of interest $y$

Consider an explanatory variable $x$ and a response variable $y$

**Model 1:** No autocorrelation in $x$
- No autocorrelation in $y$
- $y$ independent of $x$ (uncorrelated)

**Model 2:** No autocorrelation in $x$
- Autocorrelated spatial structure in $y$
- $y$ independent of $x$ (uncorrelated)

More generally for a 1- or 2-dimensional surface with bi-directional effects: autocorrelation

$$y_j = \mu_y + \sum f(y_i - \mu_y) + \epsilon_j$$

where the $y_i$'s are the values of $y$ at other sites $i$ located within the zone of spatial influence of the process generating autocorrelation in $y$ (Fig. 1.4).
Model 3: No autocorrelation in $x$ (centred values)
No autocorrelation in $y$
y depends on $x$

\[
\begin{align*}
\rho_x &= 0 \\
\rho_y &= 0 \\
\rho_y &= 0 \\
\rho_y &= 0
\end{align*}
\]

Functional dependence of $y$ on $x$
\[y_j = \mu_y + \beta x_j + \epsilon_j\]

Model 4: Autocorrelated spatial structure in $x$ (centred values)
No autocorrelation in $y$
y depends on $x$

\[
\begin{align*}
\rho_x &= \rho_x \\
\rho_y &= \rho_y \\
\rho_y &= \rho_y \\
\rho_y &= \rho_y
\end{align*}
\]

Induced spatial dependence: the spatial structure of $y$ results from the dependence of $y$ on $x$ which is spatially autocorrelated

Directional autoregressive model: \[x_j = \rho_x x_{j-1} + \xi_j, \quad y_j = \beta x_j + \epsilon_j\]

More generally for a surface: \[y_j = \mu_y + \beta x_j + \epsilon_j\]

Model 5: Autocorrelated spatial structure in $x$ (centred values)
Autocorrelated spatial structure in $y$
y depends on $x$

\[
\begin{align*}
\rho_x &= \rho_x \\
\rho_y &= \rho_y \\
\rho_y &= \rho_y \\
\rho_y &= \rho_y
\end{align*}
\]

Autocorrelation plus induced spatial dependence: the spatial structure of $y$ results from autocorrelation in $y$ and from the dependence of $y$ on $x$ which is spatially autocorrelated

Directional autoregressive model: \[x_j = \rho_x x_{j-1} + \xi_j, \quad y_j = \rho_y y_{j-1} + \beta x_j + \epsilon_j\]

More generally for a surface: \[y_j = \mu_y + \sum f(y_i - \mu_y) + \beta x_j + \epsilon_j\]
Construction of the environmental surface

Deterministic structure: binormal patch [0, 9.39] + SA in environmental variable [-2.15, 2.58] + Normal error N(0,1) [-2.72, 2.58] = Environmental surface [-4.46, 10.60]

Legendre et al. 2002, Fig. 1

Confidence intervals of a correlation coefficient

Confidence interval corrected for spatial autocorrelation: \( r \) is not significantly different from zero

Confidence interval computed from the usual tables: \( r \neq 0 \) *

**Figure 1.5.** Effect of positive spatial autocorrelation on tests of correlation coefficients; * means that the coefficient is declared significantly different from zero in this example.
Ordinary $t$-test

Dutilleul modified $t$-test

Legendre et al. 2002, Fig. 2A

Figure 13.1  Components of a sampling design are the grain size, sampling interval, and extent. In the Figure, the sampling units are represented by squares.
Figures 10 and 11 from: