ECOGRAPHY 25: 601–615, 2002

The consequences of spatial structure for the design and analysis of ecological field surveys

Pierre Legendre, Mark R. T. Dale, Marie-Josée Fortin, Jessica Gurevitch, Michael Hohn and Donald Myers

Legendre, P., Dale, M. R. T., Fortin, M.-J., Gurevitch, J., Hohn, M. and Myers, D. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. – Ecography 25: 601–615.

In ecological field surveys, observations are gathered at different spatial locations. The purpose may be to relate biological response variables (e.g., species abundances) to explanatory environmental variables (e.g., soil characteristics). In the absence of prior knowledge, ecologists have been taught to rely on systematic or random sampling designs. If there is prior knowledge about the spatial patterning of the explanatory variables, obtained from either previous surveys or a pilot study, can we use this information to optimize the sampling design in order to maximize our ability to detect the relationships between the response and explanatory variables?

The specific questions addressed in this paper are: a) What is the effect (type I error) of spatial autocorrelation on the statistical tests commonly used by ecologists to analyse field survey data? b) Can we eliminate, or at least minimize, the effect of spatial autocorrelation by the design of the survey? Are there designs that provide greater power for surveys, at least under certain circumstances? c) Can we eliminate or control for the effect of spatial autocorrelation during the analysis? To answer the last question, we compared regular regression analysis to a modified t-test developed by Dutilleul for correlation coefficients in the presence of spatial autocorrelation.

Replicated surfaces (typically, 1000 of them) were simulated using different spatial parameters, and these surfaces were subjected to different sampling designs and methods of statistical analysis. The simulated surfaces may represent, for example, vegetation response to underlying environmental variation. This allowed us 1) to measure the frequency of type I error (the failure to reject the null hypothesis when in fact there is no effect of the environment on the response variable) and 2) to estimate the power of the different combinations of sampling designs and methods of statistical analysis (power is measured by the rate of rejection of the null hypothesis when an effect of the environment on the response variable has been created).

Our results indicate that: 1) Spatial autocorrelation in both the response and environmental variables affects the classical tests of significance of correlation or regression coefficients. Spatial autocorrelation in only one of the two variables does not affect the test of significance. 2) A broad-scale spatial structure present in data has the same effect on the tests as spatial autocorrelation. When such a structure is present in one of the variables and autocorrelation is found in the other, or in both, the tests of significance have inflated rates of type I error. 3) Dutilleul's modified t-test for the correlation coefficient, corrected for spatial autocorrelation, effectively corrects for spatial autocorrelation in the data. It also effectively corrects for the presence of deterministic structures, with or without spatial autocorrelation. The presence of a broad-scale deterministic structure may, in some cases, reduce the power of the modified t-test.

P. Legendre (pierre.legendre@umontreal.ca), Dépt de Sciences Biologiques, Univ. de Montréal, C.P. 6128, succ. Centre-ville, Montréal, Québec, Canada H3C 3J7. – M.R.T. Dale, Dept of Biological Sciences, Univ. of Alberta, Edmonton, AB, Canada T6G 2E9. – M.-J. Fortin, Dept of Zoology, Univ. of Toronto, Toronto, ON, Canada M5S 3G5. – J. Gurevitch, Dept of Ecology and Evolution, State Univ. of New York, Stony Brook, NY 11794-5245, USA. – M. Hohn, West Virginia Geology and Economy Survey, P.O. Box 879, Morgantown, WV 26507-0879, USA. – D. Myers, Dept of Mathematics, Univ. of Arizona, P.O. Box 210089, Tucson, AZ 85721, USA.

Accepted 11 March 2002

ISSN 0906-7590

In field surveys, ecologists make observations at different spatial locations, hereafter referred to as "sites". The aim is often to relate biological response variables (e.g., the growth of individuals, the abundance of a species, or the structure of ecological communities) to explanatory environmental variables (e.g., soil characteristics, herbivore abundance).

Ecologists know from experience that physical as well as biological variables observed in nature display spatial patterns. Patterns may result either from deterministic processes or from processes causing spatial autocorrelation, or both; the distinction is explained below. As we will see in this paper, some types of spatial pattern make it difficult to accurately detect and quantify the biological responses, which are of primary interest. The effects of spatial structures in general, and of spatial autocorrelation in particular, on statistical tests of significance, have been described elsewhere (see, e.g., Bivand 1980, Cliff and Ord 1981, Haining 1990, or Legendre and Legendre 1998 for a review). It is our purpose here to determine the consequences of particular sampling designs on the detection of biological responses in the face of some common types of spatial patterning in the data.

Rather than analysing actual data sets, we chose the simulation approach because the use of simulations allows us to compare the outcome of the analysis to "the truth", which we know because we have generated it. The analysis of real data sets is usually limited in terms of the number of data sets available with all the necessary variables; it is also limited by the fact that one does not know whether the null hypothesis (H_0: there is no effect of the environmental variable on the response variable) or the alternative hypothesis (H_1: there is an effect) is true in any particular case. Monte Carlo simulations allow researchers to know the exact relationship between variables in the data. No doubt exists as to whether it is H_0 or H_1 which is true in each particular simulated data set (Milligan 1996).

We used stochastic simulations designed to produce surfaces of responses. These surfaces may incorporate 1) deterministic (e.g., physical) spatial patterns in the environmental variable to which the biological entities are responding, 2) spatial autocorrelation in the underlying environmental variable, and 3) spatial autocorrelation in the biotic responses. Each pair of surfaces was generated using a particular set of parameter values, and was replicated 1000 times. By analysing these replicates, we can explore the consequences of using various sampling designs for our ability to detect true biotic responses, and to conclude that there is no biotic response when none is present while different types of spatial pattern are present.

The simulations were used in this paper to address three questions formulated below. In these simulations, we generated the processes described above that may give rise to spatial structures: deterministic spatial structures and spatial autocorrelation in the explanatory variables, plus spatial autocorrelation in the ecological response variable.

The questions addressed in this paper are: a) What is the effect (in terms of type I error) of spatial autocorrelation on the statistical tests of correlation and regression analysis, which are commonly used by ecologists to analyse field survey data? b) Can we eliminate the effect of spatial autocorrelation by the design of the survey? Which designs provide the most power? c) Can we eliminate or control for the effect of spatial autocorrelation during the analysis? To answer this question, we used ordinary correlation analysis and compared it to a modified t-test developed by Dutilleul for correlation coefficients in the presence of spatial autocorrelation.

Sampling designs

In the absence of prior knowledge about the systems that they intend to sample, ecologists have been taught to rely on systematic or random sampling designs. When designing a sampling plan, scientists should make use of information, obtained from either previous surveys or a pilot study, about the nature of the underlying spatial structure of the variables. This is unfortunately not always the case. When prior knowledge about the spatial structure does exist, it is often not clear how to use it to optimize the design. The objective of this optimization should be to increase the power to detect real patterns in the response (minimizing the probability of type II error) while reducing the likelihood of false detection of responses that do not exist (which would be a type I error). Analysis of the simulations, reported in this paper, will give indications to that effect.

Spatial dependence versus spatial autocorrelation

The spatial patterns that are the most commonly encountered in nature are gradients and patches. The processes that may have produced the observed spatial structures are the subject of many debates in the literature. For response variables such as the size of plants, or species abundances, it is useful to recognize that spatial patterns may originate in two different ways, as specified by the following models (Legendre and Legendre 1998).

1. Spatial dependence. – This model implies that the response variable is spatially structured because it depends upon explanatory (e.g., physical) variables that are themselves spatially structured by their own generating processes. This is an extension of the environmental control model developed during the 1950s by Whittaker (1956), Hutchinson (1957), and Bray and Curtis (1957). The equation is the following:

$$y_j = \mu_y + f(\text{explanatory variables}_j) + \varepsilon_j \quad (1)$$

The model implies that the value taken by a dependent variable y at site j is the overall regional mean μ_y of the variable, modulated by adding the local effect of the explanatory variables at site j, plus a random error component ε_j.

2. Spatial autocorrelation. – In this model, the value of response variable y at site j is assumed to result from some dynamic process within variable *y* itself. Spatial autocorrelation actually refers to the lack of independence among the error components of field data, as a function of geographic distance among the sites. The equation describing this model is the following:

$$y_j = \mu_y + \sum_i f(y_i - \mu_y) + \varepsilon_j \quad (2)$$

This equation implies that the value of variable y at site j is the overall regional mean μ_y of the variable, plus a weighted sum of the centred values $(y_i - \mu_y)$ of the same variable at sites i that surround j, plus an independent error term ε_j. Sites i are those that are within the zone of spatial influence of the process generating the autocorrelation. In the simulation model described below, the extent of this zone of influence is determined by the range of the spherical variogram model used to generate spatial autocorrelation in the data. Note that the total error term of this model is $[\Sigma_i f(y_i - \mu_y) + \varepsilon_j]$; it contains a spatially-structured and a spatially-independent portion.

Real-case field observations often result from a combination of models 1 and 2, model 1 providing for the large-scale spatial structuring and model 2 for the smaller-scale structure:

$$y_i = \mu_y + f_1(\text{explanatory variables}_j) + \sum_i f_2(y_i - \mu_y) + \varepsilon_j \quad (3)$$

Data generated under this model may not be stationary if the explanatory variable manifests large-scale spatial structuring. Model 2 assumes second-order stationarity, which means that the mean (μ_y) and variance are the same in any portion of the study area and that the spatial autocovariance, which is the same all over the area, is a function of the separation vector rather than a function of the locations of the two points (Cressie 1993). Lack of stationarity occurs when the scale of dependence in model 1 approaches the size of the area under study. For example, the environmental variable might be spatially structured at a small spatial scale and remain stationary at the scale of the study area. Therefore, it is possible to have a relationship that can be modelled by eq. (3) and is also stationary.

An added complexity is that the physical explanatory variables of ecological models may themselves be the result of deterministic spatial structures, plus autocorrelation generated by the processes that have given rise to the environmental variables.

Methods

The model

In our simulations, a response variable (R) measured during a field survey is considered to represent the sum of separate effects (Fig. 1): the influence of an explanatory environmental variable (E), spatial autocorrelation in the response variable (SA_R), and a spatially unstructured random error component (ε_R) taking independent values for each observation i:

$$R_i = f(E_i) + SA_{Ri} + \varepsilon_{Ri} \quad (4)$$

The environmental variable, in turn, may possess a deterministic structure (D) plus a spatially autocorrelated error component (SA_E) and a spatially unstructured random error (ε_E):

$$E_i = D_i + SA_{Ei} + \varepsilon_{Ei} \quad (5)$$

Assuming a linear response function of the ecological to the environmental variables, the model for the response variable R may be written as follows:

$$R_i = \beta E_i + SA_{Ri} + \varepsilon_{Ri} = \beta(D_i + SA_{Ei} + \varepsilon_{Ei}) + SA_{Ri} + \varepsilon_{Ri} \quad (6)$$

The assumptions of this model are the following: a) all environmental effects can be summarized by a single variable whose effect on R is linear; the effect is thus modelled by multiplying E by an effect-size (regression-type) parameter β (referred to below as the "transfer parameter"). b) The error component ε, which takes independent values (i.e., not spatially autocorrelated) for each observation i, is modelled as a normal error term whose variance (Var_ε) is fixed by a parameter provided for each simulation. A normal error can legitimately be assumed for a natural phenomenon that results from a large number of factors acting independently, whose random effects are cumulative, if the variance of the phenomenon produced by each factor is small (Galton 1898, Scherrer 1984).

Numerical simulations

Simulations have been performed to check the type I error and estimate the power of the tests of significance in the presence of different types of sampling designs and spatial structures. Type I error occurs when the null hypothesis is rejected while the data conform to

H_0. A test of statistical significance is valid if the rejection rate is not larger than the significance level α, for any value of α, when the null hypothesis is true (Edgington 1995). A test of significance should also be able to reject the null hypothesis in most instances when H_0 is false. The ability to reject H_0 in these circumstances is referred to as the power of the test. In the present study, power is the empirical rate of rejection of the null hypothesis when H_0 is false by construct. High power is a desirable property. When two or more procedures are available (sampling designs or tests of statistical significance), one should use the procedure that has the highest power.

A simulation run, using the computer program described below, consists of the following steps: 1) Specify the number of simulations and the size of the experimental field, which is given in number of points from west to east and from north to south. 2) Specify a sampling design and the number of sampling units. 3) Specify the characteristics of the environmental variable E: the type of deterministic structure D; the parameters of the spherical variogram model describing the autocorrelation function SA_E for the environmental variable; and the slope parameter (beta) through which the environmental component will carry on to the response variable. The variance of the normal error component ε_E is set to be 1. 4) Specify the parameters of the spherical variogram model describing the autocorrelation function SA_R for the response variable, as well as the variance of the normal error component ε_R for the response variable.

Simulation setup

A computer program has been written to carry out the numerical simulations. A simulation consists of 1) the generation of an explanatory environmental surface and a response surface according to the parameters specified for the simulation run, 2) extraction of the explanatory variable (E) and response variable (R) following a sampling design, and 3) analysis of the relationship between E and R. A simulation run is the process through which a set of (typically 1000) replicated pairs of surfaces (E, R) are generated using a given combination of parameters, and analysed. For each simulated pair of variables, the program conducts a correlation analysis and produces a probability associated with the t-statistic. Results are accumulated over all simulations of a run.

The final statistic for a simulation run is the proportion of rejections of the null hypothesis (H_0: no effect of the environmental variable on the response variable) across the simulations. A 95% confidence interval for the rejection rate is also computed. The program allows users to obtain output files containing individual simulated surfaces, which can be drawn as maps.

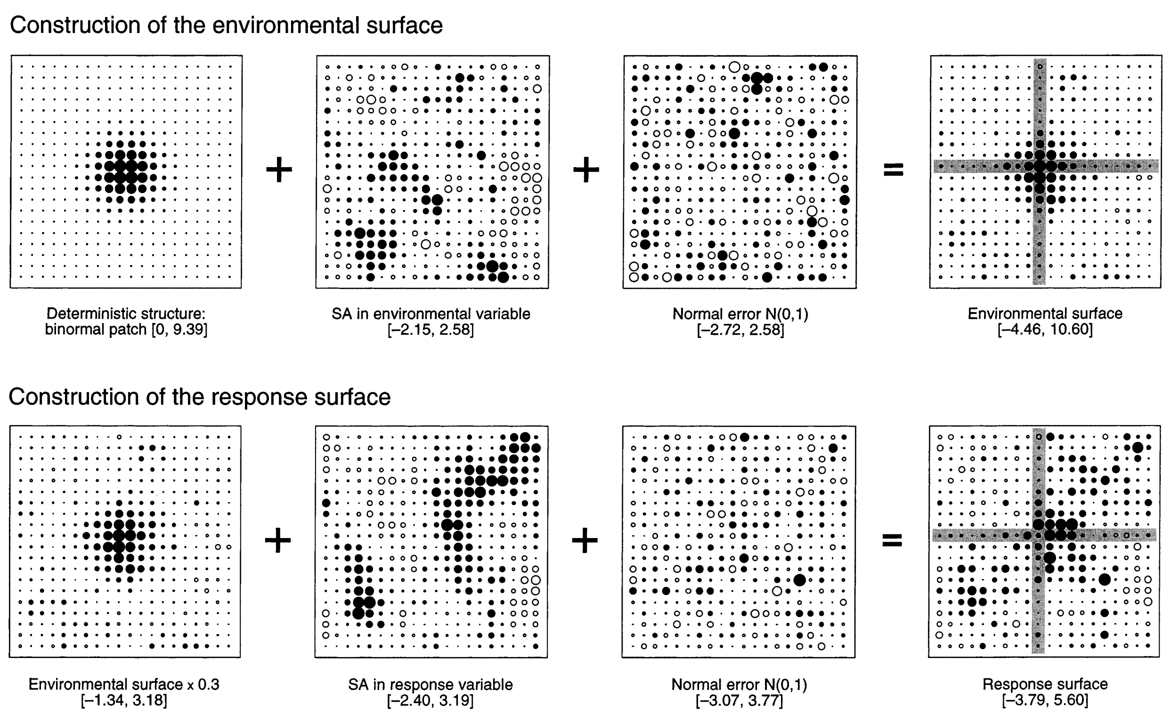


Fig. 1. Construction of the environmental and response surfaces during the simulations. See text. Bigger empty circles represent larger negative values; bigger black circles, larger positive values. The range of values in each graph is shown in brackets underneath. The sampling design is a cross in this example; the 39 sampled points are shown in grey.

Table 1. Simulations carried out to study the rate of type I error.

Deterministic structure in the environmental variable	Sampling designs	Variogram ranges for variables E and R	Total number of runs, each runs involving 1000 simulations
No deterministic structure ("deterministic" = 0)	1, 2, 3, 8, 10, 11, 12, 13	0, 20 and 50 (total of 9 combinations)	72
	6, 7, 9	0, 4 and 16 (total of 9 combinations)	27
Gradients in two directions	1, 2, 3	0, 20 and 50 (total of 9 combinations)	27
("deterministic" = 2)	6, 7, 9	0, 4 and 16 (total of 9 combinations)	27
One big patch	1, 2, 3	0, 20 and 50 (total of 9 combinations)	27
("deterministic" = 3)		0, 4 and 16 (total of 9 combinations)	27
Four waves	1, 2, 3	0, 20 and 50 (total of 9 combinations)	27
("deterministic" = 4)	6, 7, 9	0, 4 and 16 (total of 9 combinations)	27
Two regions/zones	1, 2, 3, 8	0, 20 and 50 (total of 9 combinations)	36
("deterministic" = 5)	6, 7, 9	0, 4 and 16 (total of 9 combinations)	27
Grand total			324

Spatially autocorrelated surfaces with given variogram parameters were generated using the Conditional Simulation method, as implemented in subroutine Sgsim of the geostatistical software library GSLib (Deutsch and Journel 1992). We restricted our simulations to spherical variograms with range specified in the run parameters, without or with anisotropy.

The following parameters were used in the simulations reported in this paper.

a) Surfaces were generated in a field containing 100×100 points (or nodes).

b) Six types of underlying spatial structures were available in the program (simulation parameter "deterministic"): 0) no deterministic structure (random normal error only); 1) linear gradient from north (low values) to south (high values), not used in this study; 2) linear gradients from north to south and from west to east, so that the lowest values are in the upper left-hand corner and the highest values are in the lower right-hand corner of the map; 3) one big bi-normal patch in the centre of the field; 4) four waves across the field with their crests parallel to the upper and lower frames of the map; 5) two zones (north and south portions of the field) separated by a sharp discontinuity.

c) There are thirteen different choices for the sampling designs (simulation parameter "design") available in the simulation program. Different numbers of sampling units, depending on the design, were used in generating the simulations: 1) Simple random sampling (100 units). 2) Systematic (100 units). 3) Aggregates of 5 sampling units in a systematic pattern (80 or 125 units). 4) Horizontal transect with a single sampling interval (design not used in this paper). 5) Horizontal transect with 2 sampling intervals (design not used in this paper). 6) Vertical transect with a single sampling interval (50 units, interval of 2 points). 7) Vertical transect with two sampling intervals (50 units; alternating intervals of 1 and 2 points; the transect started on row 1 and reached down to row 74 of the 100×100 field). 8) Stratified vertically into 2 strata; simple random sampling within each stratum (100 units). 9) Cross consisting of two crossing transects (99 units; interval of 2 points between units). 10) Two parallel vertical transects distributed evenly (100 units, interval of 2 points). 11) Three parallel vertical transects distributed evenly (99 units, interval of 3 points). 12) Two parallel vertical transects distributed at random (100 units, interval of 2 points). 13) Three parallel vertical transects distributed at random (99 units, interval of 3 points).

d) Spatial autocorrelation in the environmental variable E and the response variable R was specified by spherical variogram models with nugget values of 0. The ranges were {20, 50} points for sampling designs 1, 2, 3, 8, 10, 11, 12 or 13, and {4, 16} points for sampling designs 6, 7 and 9. These values were chosen in such a way that there might be an effect on the rate of type I error for the ordinary t-tests, when testing the correlation coefficient.

A simulation study can never explore all parameter combinations. Choices have to be made in order to obtain publishable results in finite time. The simulation effort reported in this paper was the following: 1) Type I error study – the simulations are described in Table 1. Total: 324 runs, each one involving 1000 simulations. 2) Power study – the same simulation runs were repeated using the value 0.3 for the transfer parameter β. The value 0.3 was chosen because it obtained differences in power among simulation runs; it was high enough to produce a measurable response in many simulations, yet not so high as to produced detection of an effect in all simulations. Total: 324 runs, each one involving 1000 simulations.

Illustration

Figure 1 illustrates how the environmental and response surfaces were created and the sampling was conducted. In this example, sampling is conducted in a field containing 20×20 points. The surface for variable E (environmental) was constructed as the sum of a deterministic structure, plus a spatially autocorrelated error component (SA) generated using a spherical variogram model with range of 5 in both directions, plus a spatially unstructured random error component N(0, 1). Note that the first two components may or may not be present in the results reported below; a choice of 5 deterministic components are available in the program.

The response variable is the sum of three separate effects: the influence of the environmental variable multiplied by an effect-size parameter, spatial autocorrelation (SA) in the response variable generated using a spherical variogram model with range of 5 in both directions, and a spatially unstructured random error component N(0, 1). The first two components may or may not be present in the results reported below. One of several sampling designs, available in the program, was applied to the data field. The environmental and response variables were measured at these sites, producing the data file which was then analysed by the program using an ordinary t-test of the significance of the correlation coefficients as well as a t-test modified to take spatial autocorrelation into account.

Statistics

Since we are dealing with a single environmental and response variable, simple linear regression can be replaced by correlation analysis; the F-test of the regression coefficient is equivalent to a two-tailed t-test of the Pearson correlation coefficient between these two variables. The first test used in the present study is thus a regular t-test of the Pearson correlation coefficient. The second one is Dutilleul's (1993) modified t-test which corrects the variance of the test statistic as well as the degrees of freedom (df) in the presence of spatial autocorrelation. This test is a generalized and exact form of the approximate procedure proposed by Clifford et al. (1989). We will examine how well Dutilleul's modified t-test is able to compensate for SA in the environmental and response variables.

Predicted results

The rate of type I error is computed as the proportion of rejection of the null hypothesis when the data conform to it. In our simulations, H_0 is true if the environmental (E) and response (R) variables are not linked by the transfer parameter β. A test can be said to have correct rate of type I error if, across the simulations, the rejection rate is approximately equal to the significance level α used to make the statistical decision.

Based upon the simulation results reported by Bivand (1980) for autocorrelated variables in the case of the Pearson correlation coefficient, we made the following prediction: in field surveys, the rate of type I error should be inflated (i.e., the rejection rate should be higher than α) when spatial autocorrelation is present in both the controlling environmental variable (E) and the response variable (R), if the spatial autocorrelation is not explicitly taken into account in the course of the analysis. The statistical reasons for this predicted behaviour are summarized in Legendre and Legendre (1998).

Simulation results

In the simulation runs for studying the rate of type I error, surfaces were generated in such a way that the null hypothesis was true. This means that the transfer parameter β of eq. (5) was set to zero, so that no relationship was created between the environmental and response variables. The alternative hypothesis was made true in the power study by setting the transfer parameter β to 0.3. The statistic subjected to a test of significance was the correlation between the environmental (E) and response (R) variables.

The rate of type I error is the most important aspect in the comparison of the 11 sampling designs used in this study and the two ways of analysing the data (classical Pearson *r* and test corrected for spatial autocorrelation). Here are the results for the various cases that were studied.

1. No deterministic structure in the environmental variable ("deterministic" = 0)

Type I error

Under simple random sampling, the ordinary t-test has correct α significance level when there is no autocorrelation, or when autocorrelation is present in only one of the variables (Fig. 2A). The rate of type I error is inflated when autocorrelation is present in both variables. This is in agreement with the results reported by Bivand (1980). The same remains true, by and large, for the other random sampling designs investigated in this study.

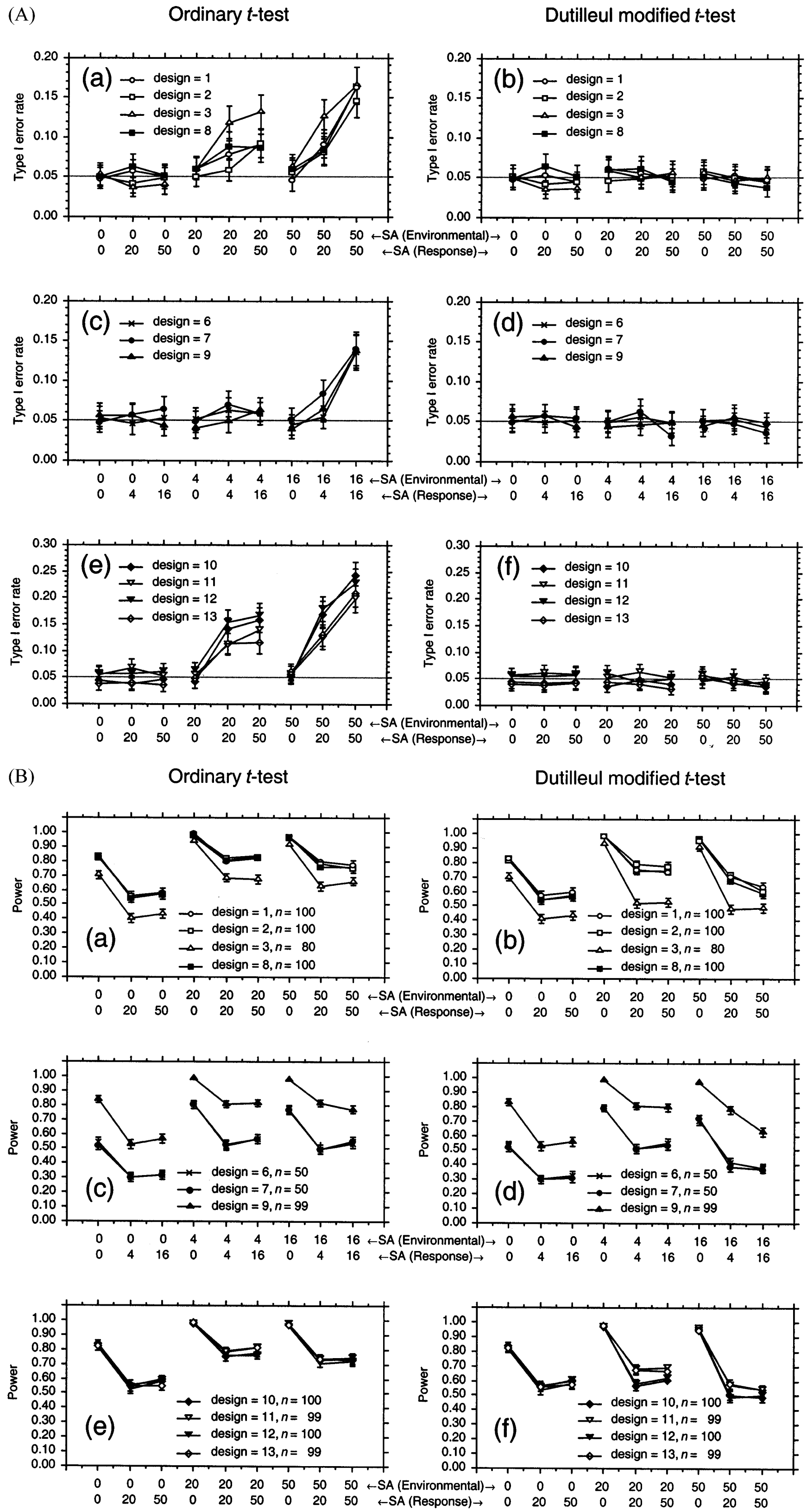
Dutilleul's modified t-test, which compensates for spatial autocorrelation, results in correct rate of type I error with all sampling designs investigated here. With some designs, the modified t-test has a slightly reduced rate of type I error; the test thus remains valid.

Power

The power study (Fig. 2B) shows that power is the same for all sampling designs. The only differences in power are most likely due to differences in sample size (n = 80 for sampling design 3, n = 50 for designs 6 and 7; n = 99 or 100 for all other designs). The presence of SA in the environmental variable seems to increase power slightly.

When the unmodified t-test has correct rate of type I error, its power is the same as Dutilleul's modified

Fig. 2. A) Type I error rates and 95% confidence intervals (error bars) at $\alpha = 0.05$ of the ordinary t-test (left) and Dutilleul's modified t-test (right) for increasing values (along the abscissa) of the ranges of the variograms determining spatial autocorrelation (SA) in the environmental and response variables. Each error rate estimate results from analysing 1000 simulated data sets for which the null hypothesis was true. There was no deterministic structure in the simulated data sets. The sampling designs are described in the Methods. B) Same as A), power study.



t-test. The apparent greater power of the unmodified t-test, in some cases, is due to the increase in rate of type I error shown in Fig. 2A. The unmodified t-test is invalid when it has inflated rate of type I error; it should not be used in these cases. The presence of any amount of SA in the response variable reduces the power of both tests.

Lesson learned

In the absence of SA, there is no change in the degrees of freedom in Dutilleul's modified t-test. Results are identical to those of the unmodified t-test. Dutilleul's modified t-test can thus be used in all situations, i.e., in the presence or absence of SA. The results depend, however, on the quality of the estimation of SA; this problem is saved for the Discussion.

2. Gradient in the environmental variable ("deterministic" = 2)

Gradients are the most commonly encountered spatial structures in nature.

Type I error

A broad-scale gradient in the environmental variable has the same effect on the ordinary t-tests as if SA was present in that variable (Fig. 3A). Thus, it is only when SA is not present in the response variable that the regular t-test is valid, having correct or conservative (i.e., deflated) rate of type I error.

Dutilleul's modified t-test reacts to the presence of a deterministic spatial structure by over-correcting the F-statistic and degrees of freedom; this produces reduced rates of type I error. The test thus remains valid.

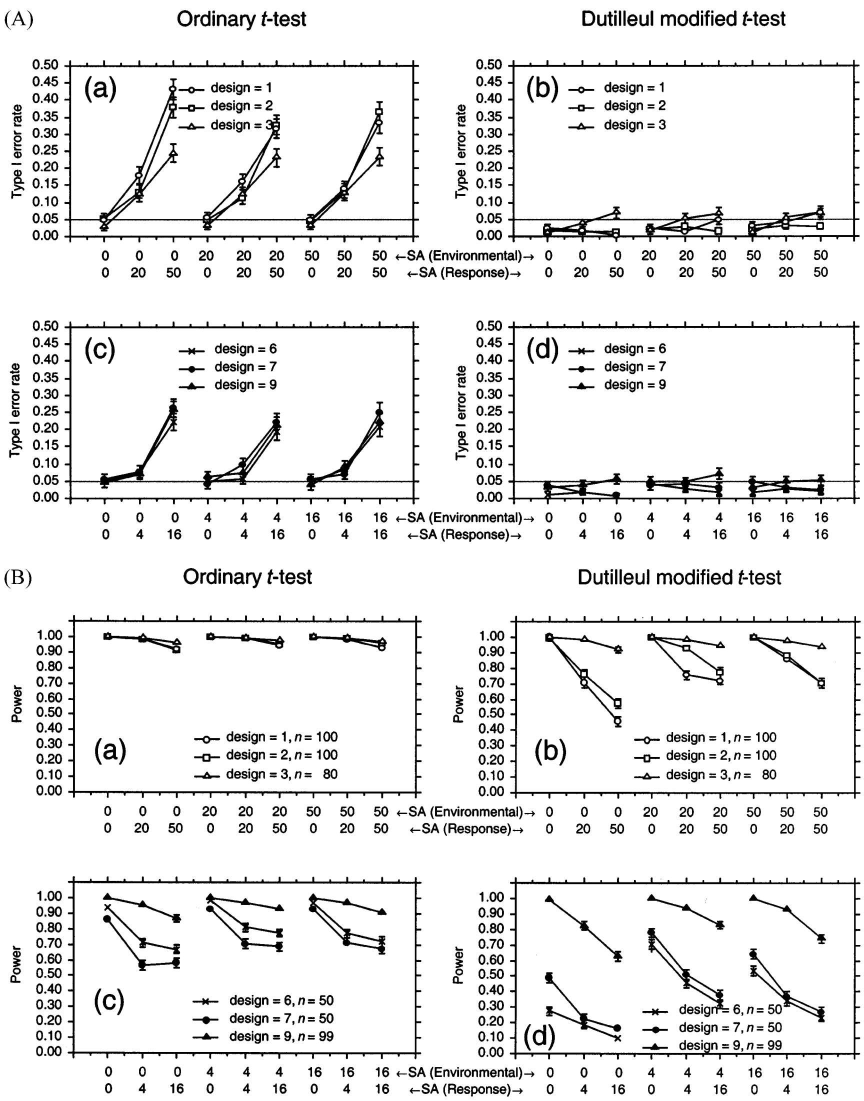


Fig. 3. A) Type I error rates and 95% confidence intervals at $\alpha = 0.05$ of the ordinary t-test (left) and Dutilleul's modified t-test (right) for increasing values of the ranges of the variograms determining spatial autocorrelation in the environmental and response variables. See Fig. 2A for details. A gradient ("deterministic" = 2) was included in the environmental variable. B) Same as A), power study.

The modified t-test interprets the spatial gradient as an indicator of broad-scale spatial autocorrelation, as if the sampled area was a portion of a broader-scale autocorrelated structure. Aggregated sampling (design = 3) may produce slightly inflated rate of type I error when SA is strong in the response variable, in addition to the spatial gradient.

Power
The power study (Fig. 3B) shows that for ordinary t-tests, power is no problem when the test is valid, i.e., when there is no autocorrelation in the response variable. The seemingly smaller power for the two transect designs (design = 6 and 7) is due to smaller sample size (n = 50).

For Dutilleul's modified t-test, some designs have poor power with some or all combinations of SA in the environmental and response variables. Aggregated sampling (design = 3) retains high power because it allows a better estimation of autocorrelation in the first distance classes of the correlograms computed in Dutilleul's method. The other designs have approximately equal power; the transect designs (design = 6 and 7) seem to have smaller power, but this is due to the smaller sample size used in the simulations.

Lesson learned
Instead of analysing data containing a broad-scale gradient, a better way is to look for, and identify, the gradient in the environmental variable. It can be explicitly included in the regression model in the form of a linear trend-surface equation of the site coordinates, together with the environmental and response variables, as will be shown in the Discussion.

To analyse data containing an un-filtered broad-scale gradient, aggregated sampling (design = 3) is the best overall design if the analysis is done using Dutilleul's modified t-test.

3. Large patch in the environmental variable ("deterministic" = 3)

Patches are perhaps the second most commonly encountered spatial structures in nature.

Type I error
The observations are essentially the same as in the case of gradients. A large patch in the environmental variable has the same effect on regular t-tests as if SA was present (Fig. 4A). The effect of this type of spatial structure on Dutilleul's modified t-test is negligible.

Power
The power study (Fig. 4B) shows that for ordinary t-tests, power is no problem with sampling designs 1, 2, 3 and 9 when the test is valid, which is the case when there is no autocorrelation in the response variable. For Dutilleul's modified t-test, the transect and cross designs have poor power with some or all combinations of SA in the environmental and response variables. This may be linked to violation of the stationarity assumption by the large patch structure. The best designs in terms of power are aggregated (design = 3), systematic (design = 2), and simple random sampling (design = 1).

Lesson learned
Rather than analysing data containing a large patch, it seems preferable to positively identify the patch structure during a pilot study (or during the actual study) and remove its effect by including the terms of a polynomial trend-surface equation in the regression analysis; see example in the Discussion. Short of that, to analyse data containing a large patch, simple random, aggregated, and systematic sampling (design = 1, 2 and 3) are the best overall designs if the analysis is done using Dutilleul's modified t-test.

4. Waves in the environmental variable ("deterministic" = 4)

Regular waves are found, in nature, when the deposition of a material (e.g., sand) was controlled by the flow of a fluid, commonly air or water, but also ice, magma, etc.

Type I error
With this type of spatial structure, the sampling designs covering the whole sampling field (i.e., random and systematic) are doing better in the presence of spatial autocorrelation in the response variable than the other designs (Fig. 5A). The worst results were obtained using the aggregated design, even though we made sure that the aggregates were not in phase with the waves of the deterministic structure. The best design without Dutilleul's modified t-test is the systematic. By and large, Dutilleul's modified test has correct rate of type I error, except for a slight inflation in the case of the systematic design, in the presence of strong autocorrelation in the response variable.

Power
The power study (Fig. 5B) shows that all methods have equally good power when they are valid. With Dutilleul's modified t-test, transects (designs 6 and 7) seem to have lower power in the presence of spatial autocorrelation in the response variable, but this is due to the smaller sample size used in the transect simulations.

Lesson learned
With this type of spatial structure, the best thing to do is to use a systematic sampling design and carry out a test without modification; or to use any type of sampling and carry out a test with Dutilleul's modification.

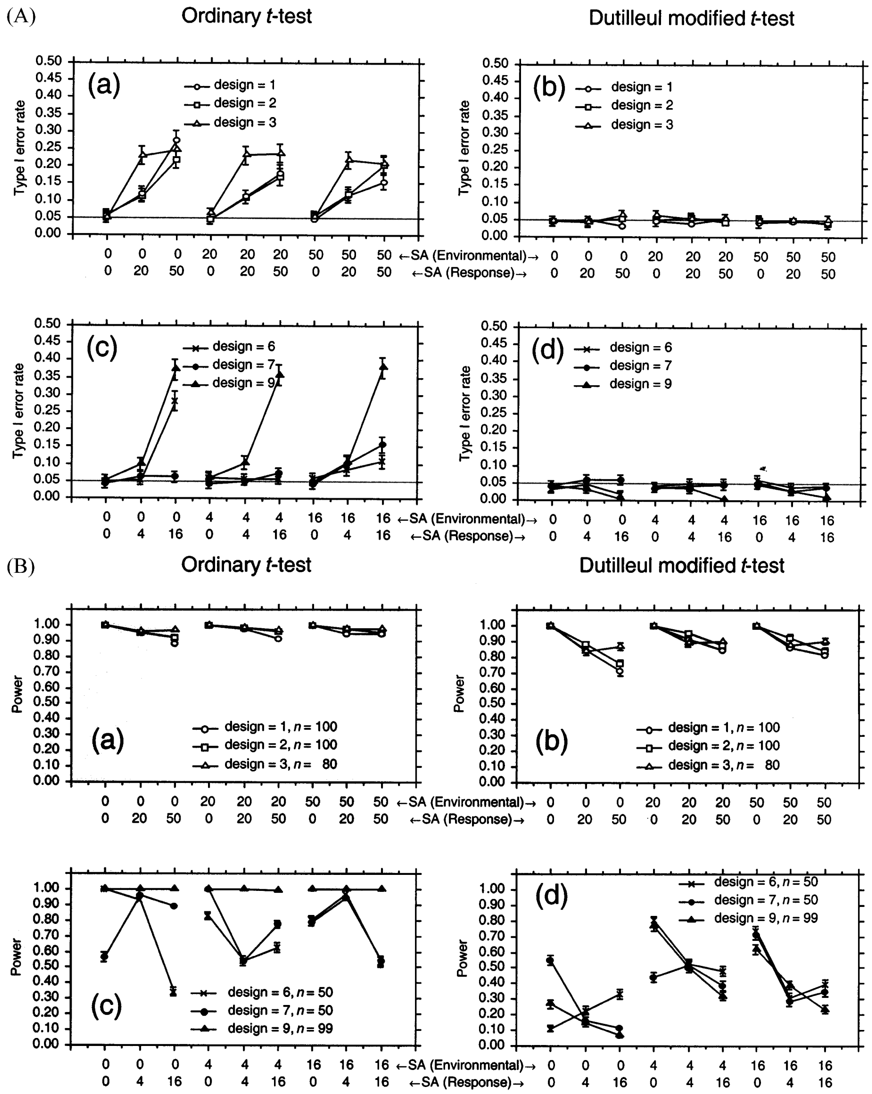


Fig. 4. A) Type I error rates and 95% confidence intervals at $\alpha = 0.05$ of the ordinary t-test (left) and Dutilleul's modified t-test (right) for increasing values of the ranges of the variograms determining spatial autocorrelation in the environmental and response variables. See Fig. 2A for details. A big bi-normal patch ("deterministic" = 3) was included in the environmental variable. B) Same as A), power study.

5. Two clearly separate zones in the environmental variable ("deterministic" = 5)

This type of structure is encountered, for instance, in vegetation surveys that cover two geologically different zones. Another example is found in limnology, when the littoral and pelagic zones of a lake are analysed together.

Type I error

The observations are the same as in the case of a gradient (Fig. 6A).

Power

The power study (Fig. 6B) shows that all designs have equally good power, except for vertical transects with a single sampling interval (design 6: nearly no power at all) and the transect with two sampling intervals (design 7: low power) in the case of Dutilleul's modified t-test. This is clearly due to violation of the stationarity assumption by the two-zone structure.

Lesson learned

If the strata have been chosen in such a way as to fit the natural divisions of the environmental variable in the field, one should use a covariable, representing the strata, in the analysis of the results (partial correlation or partial regression), to control for the effect of the strata means on the analysis.

Discussion

Dutilleul's modified t-test which takes the effect of spatial autocorrelation into account is a major breakthrough for the analysis of survey data. However, it requires good estimates of the spatial autocorrelation present in the variables under study. As sample size increases, the estimates become more accurate. Conversely, with small sample size, the accuracy of the modification to the F-statistic and the degrees of freedom is reduced. This is an inherent limitation of any method of modification based upon estimates of spatial autocorrelation obtained from the data themselves. The take-home message for ecologists is that, if the variables under study are spatially autocorrelated, the sample size (n) should be as large as possible – for instance: $n = 100$ in the case of strongly autocorrelated data.

For Dutilleul's modified t-test, the reduction of the rate of type I error due to the presence of a deterministic structure actually depends on the scale of that structure. The effect on the rate of type I error is stronger for broader-scale deterministic spatial structures such as gradients, and smaller for smaller structures such as big bumps in the centre of the field. With some combinations of SA in the environmental and response variables, designs utilizing transects have poorer power than the simple random, systematic or aggregated designs, probably because the assumption of stationarity, required by the correlograms computed in the Dutilleul procedure, is violated by the presence of these structures.

Taking broad-scale spatial structures and finer-scale SA into account

The simulations have shown that, when a broad-scale spatial structure is present in the environmental variable E, this structure makes even Dutilleul's modified t-test have reduced rate of type I error. Although the test remains valid, its power is reduced, so this is an

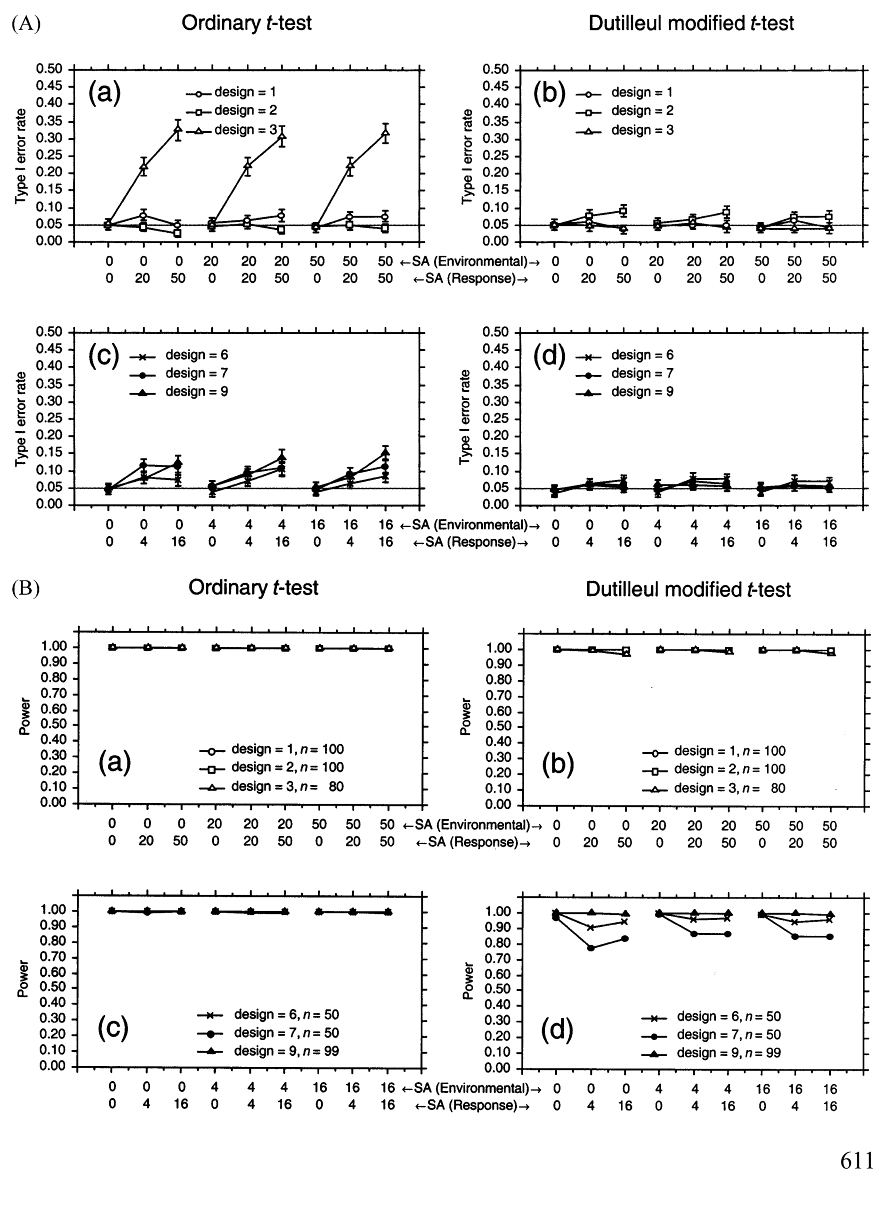


Fig. 5. A) Type I error rates and 95% confidence intervals at $\alpha = 0.05$ of the ordinary t-test (left) and Dutilleul's modified t-test (right) for increasing values of the ranges of the variograms determining spatial autocorrelation in the environmental and response variables. See Fig. 2A for details. Waves ("deterministic" = 4) were included in the environmental variable. B) Same as A), power study.

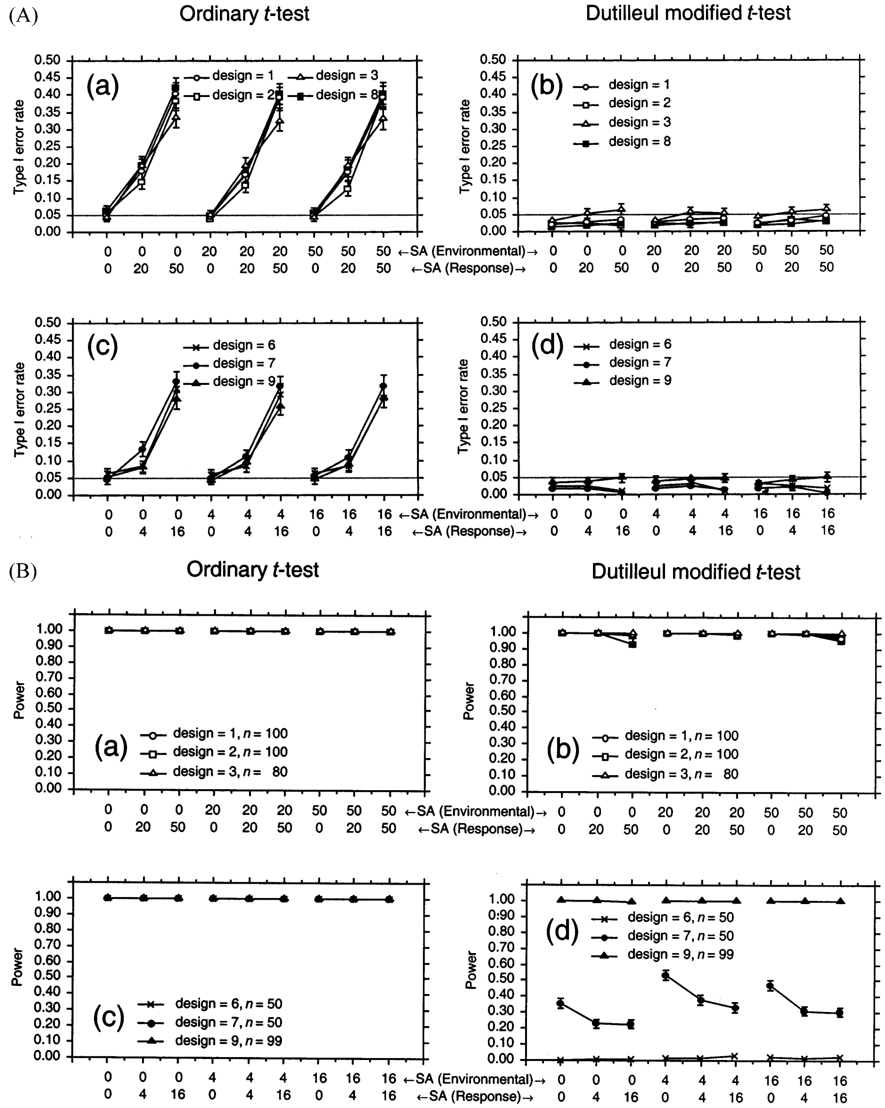


Fig. 6. A) Type I error rates and 95% confidence intervals at $\alpha = 0.05$ of the ordinary t-test (left) and Dutilleul's modified t-test (right) for increasing values of the ranges of the variograms determining spatial autocorrelation in the environmental and response variables. See Fig. 2A for details. A two-zone structure ("deterministic" = 5) was included in the environmental variable. B) Same as A), power study.

undesirable property. Instead of analysing data containing a broad-scale spatial structure, a better way is to look for, and identify, the gradient in the environmental variable using some form of spatial modelling. The method of analysis is the following.

1) Is there a broad-scale spatial component in the environmental variable E? One can use the results of a pilot study or field observations to answer this question. If so, this structure must be identified and "peeled off" the data before studying the relationship between the environmental and response variables. In some cases, the broad-scale spatial component can be modelled using a linear or polynomial trend-surface equation. Trend-surface analysis is a classical form of spatial modelling; it is explained in several textbooks, including Legendre and Legendre (1998, Section 13.2.1). In other instances, the broad-scale component can be hypothesized to have other functional forms. For instance, a patch can be modelled by a Gaussian response function (an example is given below) which can be modelled by a normal density function through nonlinear regression; a discontinuity can be modelled by a dummy variable in linear regression. k is the number of parameters required to fit the broad-scale spatial model to variable E.

2) Calculate the partial correlation between R and E. According to our hypothesis, if a broad-scale spatial structure is present in the data, it is caused by the broad-scale spatial structuring of the environmental variable. 2.1) Compute the vector of residuals $Res(E_i)$ of the environmental variable E after fitting the broad-scale spatial model. 2.2) Compute the vector residuals $Res(R_i)$ of the regression of the response variable R on the fitted broad-scale spatial model. 2.3) Compute the

correlation r between Res(E) and Res(R). This correlation is actually the partial correlation between E and R after controlling for the broad-scale spatial model.

3) Compute the associated probability: 3.1) To take into account the spatial autocorrelation potentially present in E and R, compute the Dutilleul-corrected number of degrees of freedom, ν_{Dut} A program to compute the modified t-test for the Pearson correlation coefficient corrected for spatial autocorrelation, following Dutilleul (1993), is available at URL ⟨http://www.fas.umontreal.ca/biol/legendre/⟩. This program can be used to estimate the partial correlation r (the same value is obtained as in step 2.3 above) as well as the corrected number of degrees of freedom ν_{Dut}.

Note that in some cases, after removing the broad-scale spatial structure, spatial autocorrelation analysis may not detect any significant autocorrelation remaining in one, the other, or both residuals Res(E) and Res(R). In that case, one does not have to compute a modified number of degrees of freedom using Dutilleul's method: when autocorrelation is present in only one of the variables under study, or in none of them, the rate of type I error is not modified, as shown by Bivand (1980) and illustrated in Fig. 2A.

3.2) Compute the modified partial t-statistic from the partial correlation coefficient r:

$$t_c = \frac{r\sqrt{\nu_c}}{\sqrt{1 - r^2}} \qquad (7)$$

In this formula, use a corrected number of degrees of freedom $\nu_c = \nu_{Dut} - k$ where k is the number of parameters required in step 1 (above) to fit the broad-scale spatial model to variable E. If no spatial autocorrelation is present in the data (or, at least, in one of the residual variables), $\nu_c = (n - 2) - k$; t is then the classical statistic for testing the significance of a partial correlation coefficient. This value is identical to the t-statistic used for testing the significance of a partial regression coefficient in multiple regression.

3.3) Find the probability associated with the t-statistic in a one-tailed or two-tailed test, for ν_c degrees of freedom.

One should check that there is no other broad-scale spatial component in the response variable R, besides the one modelled for E. If this happened, it would mean that some other environmental variable E′ containing a broad-scale spatial structure is also an important determinant of R; the model should be redesigned to include this variable.

Example 1

Let us illustrate this procedure using two numerical examples. We used a sampling field of size (100 × 100 points). Using our simulation program, we generated a first pair of variables similar to those of Fig. 1, but without any effect of the environmental variable E (Fig. 7a) on the response variable R (Fig. 7b); to do so, we simply set the simulation transfer parameter (β) to 0. The environmental variable was made to contain a large patch in the centre of the field, as in Fig. 1, plus spatially-autocorrelated error and non-spatially-structured normal error. The response variable only contained spatially-autocorrelated error and non-spatially-structured error. For both variables, the spatially-autocorrelated error component was generated using a spherical variogram model with range of 25 in both directions. We ran a horizontal transect through

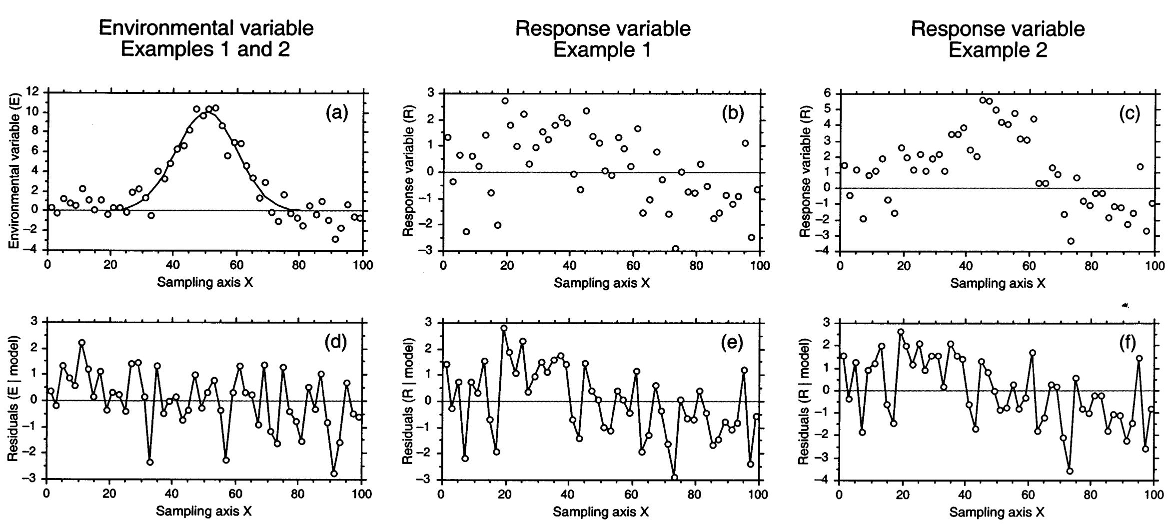


Fig. 7. Illustration of the data used in Examples 1 and 2. In Example 1, R is independent of E. In Example 2, a dependence between E and R was created by setting the response parameter beta to 0.4. E is the same in the two examples. (a, b, c) Raw data. The Gaussian density function fitted to E, called Fit(E) in the text, is also shown (curve). (d, e, f) Plot of the residuals. Adjacent values are linked by lines to make it easier to appreciate the autocorrelation remaining in the residuals.

the centre of the patch (like the horizontal arm of the cross design in Fig. 1) and measured the two variables at 50 equispaced points along the transect.

We will now assume that we don't know how the variables were generated and analyse them as we would do for field data. Our task is to deconstruct the data, peeling off the broad-scale spatial component, and find whether or not the residuals are significantly related to each other. So, we will proceed to spatial modelling of variable E. Figure 7a indicates the presence of a bump in the E values observed along the transect. A bump would be difficult to model using a polynomial trend-surface equation, and would require many terms to approximate its shape. We chose to model it using a Gaussian function (normal density):

$$E_i = c\left[\frac{1}{\sqrt{2\pi b}}\, e^{\frac{-(x_i - a)^2}{2b}}\right] \tag{8}$$

where X_i represents the position of site i along the transect, a represents the estimate of the mean, b is that of the variance, and *c* is a vertical scale parameter. The model was fitted to the E data using nonlinear regression (curve in Fig. 7a, $R^2 = 0.91$); $k = 3$ parameters were estimated to fit the model. The parameter estimates were $a = 50.44$, $b = 92.54$, and $c = 242.81$. The fitted values of this model, Fit(E), were computed and used in the sequel as our estimates of the broad-scale deterministic spatial structure identified in E. The residuals Res(E) of this model were also calculated (Fig. 7d).

The next step is to test the hypothesis that the broad-scale spatial structure found in E may have been passed on to the response variable R. The regression of R on vector Fit(E) was computed and the residuals Res(R) were calculated (Fig. 7e); as expected, this regression explained very little of the response data ($R^2 = 0.08$) since the data had been generated with a beta coefficient of 0.

The correlation between the two vectors of residuals was $r = 0.1654$. Dutilleul's modified t-test for the correlation coefficient, corrected for spatial autocorrelation, was computed; among other information, the program provided the number of degrees of freedom corrected for spatial autocorrelation ($\nu_{Dut} = 28.65$). For $\nu_c = \nu_{Dut} - k = 25.65$, the corrected t-statistic was $t_c = (r\sqrt{\nu_c})/\sqrt{(1 - r^2)} = 0.8496$ and the associated probability was $p = 0.4032$. We can now compare this answer to the results one would have obtained from the calculation of a correlation coefficient between the two original variables E and R: $r(E,R) = 0.3178$, $p = 0.0245$. At significance level 0.05, one would have drawn the erroneous conclusion that there was a significant relationship between R and E. This would have been due to the inflated type I error rate of the test in the presence of autocorrelation (Fig. 2A) and of a broad-scale deterministic structure (Fig. 4A) in the data (Table 2, central column). As we observed in Fig. 2A, the test would have been too liberal ($p = 0.2509$ in the central portion of the Table) if we had not applied Dutilleul's modified t-test, which corrects for the spatial autocorrelation present in the data.

Table 2. Three estimates of the significance of the correlation between a response (R) and an environmental (E) variable. In Example 1, R is independent of E. In Example 2, a dependence between E and R was created by setting the response parameter beta to 0.4. *k* is the number of parameters required to fit the broad-scale spatial model to variable E; $k = 3$ in these examples.

	Example 1 beta = 0	Example 2 beta = 0.4
Correlation between R and E		
r(R,E)	0.3178	0.8172
$\nu = n - 2$	48	48
p	0.0245*	<0.0001***
Correlation between residuals after controlling for effect of broad-scale spatial structure		
r[Resid(R), Resid(E)]	0.1654	0.4466
$\nu = n - 2$	48	48
p	0.2509 N.S.	0.0012**
Correlation between residuals using Dutilleul's modified t-test		
r[Resid(R), Resid(E)]	0.1654	0.4466
$\nu_c = \nu D_{Dut} - k$	25.65	22.15
p	0.4032 N.S.	0.0202**

*** : $p \le 0.001$; ** : $p \le 0.01$; * : $p \le 0.05$; N.S.: not significant.

Example 2

A second pair of variables was simulated, but this time there was an effect of E (Fig. 7a) on R (Fig. 7c) that was generated by setting the simulation transfer parameter (β) to 0.4. Except for that, the deterministic structure in E, and the SA and normal error components in R and E, were the same as in the first example, so that the estimated broad-scale deterministic structure in E, Fit(E), as well as the residuals Res(E), were the same as in Example 1.

In the second step, R was regressed on Fit(E) and the residuals Res(R) were calculated (Fig. 7f); this time, the regression explained an important portion of the variance of the response data ($R^2 = 0.5868$) since the data had been generated with a beta coefficient of 0.4.

The correlation between the two vectors of residuals was $r = 0.4466$. Dutilleul's modified t-test for the correlation coefficient, corrected for spatial autocorrelation, was computed; the program provided the number of degrees of freedom corrected for spatial autocorrelation ($\nu_{Dut} = 25.15$). For $\nu_c = \nu_{Dut} - k. = 22.15$, the modified t-statistic was $t_c = (r\sqrt{\nu_c})/\sqrt{(1 - r^2)} = 2.5034$ and the associated probability was $p = 0.0202$. We can now compare this answer to the results one would have obtained from the calculation of a correlation coefficient between the two original variables E and R:

$r(E,R) = 0.8172$, $p < 0.0001$. The statistical conclusion drawn from this result would have been correct, but the probability is far too small. The incorrect and correct results are summarized in the right-hand column of Table 2. Again, and as in Fig. 4A, the test would have been too liberal ($p = 0.0012$ in the central portion of the Table) if we had not applied Dutilleul's modified t-test, which corrects for the spatial autocorrelation present in the data.

Conclusion

1) Spatial autocorrelation in both variables disturbs the classical tests of significance of correlation or regression coefficients. Spatial autocorrelation in a single variable does not affect the test of significance.

2) A broad-scale spatial structure present in data has the same effect on the tests as spatial autocorrelation. When such a structure is present in one of the variables and autocorrelation is found in the other, or in both, the tests of significance have inflated rate of type I error.

3) Dutilleul's modified t-test for the correlation coefficient, corrected for spatial autocorrelation, effectively corrects for spatial autocorrelation in the data. It also effectively corrects for the presence of some structures, with or without spatial autocorrelation; the test is always valid. The presence of a broad-scale deterministic structure may, in some cases, reduce the power of the modified t-test: Dutilleul's modified t-test uses correlograms computed to estimate the corrected number of degrees of freedom, and correlograms assume second-order stationarity. This assumption is clearly violated, for instance, by the presence of a big patch in the centre of the field.

The simulation program used in this paper constitutes one of the end products of this work. The source code, written in FORTRAN, is available at URL ⟨http://www.fas.umontreal.ca/biol/legendre/⟩ to users who may want to develop subroutines allowing the comparison of different methods of analysis of the data in terms of type I error and power, or simply generate spatially-structured random data sets. A user's manual is distributed with the program.

Acknowledgements – The work reported in this paper was conducted as part of the Working Group "Integrating the Statistical Modeling of Spatial Data in Ecology" supported by the National Center for Ecological Analysis and Synthesis (NCEAS), a Center funded by NSF (Grant # DEB-94-21535), the Univ. of California at Santa Barbara, and the State of California. For the simulations reported in this paper, program SimSSD was launched on a large number of computers using the Sepia system for distributed computing developed by Philippe Casgrain (work supported by NSERC grant number OGP0007738 to P. Legendre). We are thankful to Marie-Hélène Duplain and Univ. de Montréal for making an undergraduate computer room available during several nights for the Sepia-based simulations.

References

Bivand, R. 1980. A Monte Carlo study of correlation coefficient estimation with spatially autocorrelated observations. – Quaest. Geogr. 6: 5–10.

Bray, R. J. and Curtis, J. T. 1957. An ordination of the upland forest communities of southern Wisconsin. – Ecol. Monogr. 27: 325–349.

Cliff, A. D. and Ord, J. K. 1981. Spatial processes – models and applications. – Pion, London.

Clifford, P., Richardson, S. and Hémon, D. 1989. Assessing the significance of the correlation between two spatial processes. – Biometrics 45: 123–134.

Cressie, N. A. C. 1993. Statistics for spatial data, revised ed. – Wiley.

Deutsch, C. V. and Journel, A. G. 1992. GSLIB – Geostatistical software library and user's guide. – Oxford Univ. Press.

Dutilleul, P. 1993. Modifying the t test for assessing the correlation between two spatial processes. – Biometrics 49: 305–314.

Edgington, E. S. 1995. Randomization tests, 3rd ed. – Marcel Dekker, NY.

Galton, F. 1898. Natural inheritance. – Macmillan.

Haining, R. 1990. Spatial data analysis in the social and environmental sciences. – Cambridge Univ. Press.

Hutchinson, G. E. 1957. Concluding remarks. – Cold Spring Harbor Symp. Quant. Biol. 22: 415–427.

Legendre, P. and Legendre, L. 1998. Numerical ecology, 2nd English ed. – Elsevier.

Milligan, G. W. 1996. Clustering validation – results and implications for applied analyses. – In: Arabie, P., Hubert, L. J. and De Soete, G. (eds), Clustering and classification. World Scientific Publ. Co, River Edge, NJ, pp. 341–375.

Scherrer, B. 1984. Biostatistique. – Gaëtan Morin Ed., Boucherville.

Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. – Ecol. Monogr. 26: 1–80.