Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 1402–1415
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Modified F tests for assessing the multiple correlation between one spatial process and several others

Pierre Dutilleul[a,*], Bernard Pelletier[a,b], Gülhan Alpargu[c]

[a]Department of Plant Science, McGill University, Macdonald Campus, Ste-Anne-de-Bellevue, Que., Canada H9X 3V9
[b]Department of Natural Resource Sciences, McGill University, Macdonald Campus, Ste-Anne-de-Bellevue, Que., Canada H9X 3V9
[c]Mathematics Department, California State University, Fullerton, CA 92834, USA

Received 5 August 2005; accepted 11 June 2007
Available online 3 July 2007

Abstract

The modified t test presented by Dutilleul [1993. Modifying the t test for assessing the correlation between two spatial processes. Biometrics 49, 305–314] to assess the simple correlation between two spatial processes is built on the sample correlation coefficient calculated from partial realizations, and requires the estimation of an effective sample size appropriately defined. The autocovariances of processes are taken into account via the effective sample size, both in the evaluation of the test statistic and in the calculation of the probability of significance.

In this article, we present modified F tests to assess the multiple correlation between one spatial process and several (i.e., q) others, from partial realizations collected at the same sampling locations. An effective sample size is used in all these modified F tests. Its theoretical expression is obtained from the expected value of the coefficient of determination in the absence of multiple correlation, under the bivariate Gaussian and first-order stationarity assumptions for the 'dependent process' and the 'predicted process' (i.e., the ordinary least-squares (OLS) regression predictor defined from the q other processes). The modified F tests presented differ in the procedure (i.e., model-based versus model-free) followed to estimate the autocovariance matrices appearing in the expression of the effective sample size. The validity and power of the new testing procedures are studied through simulations, for various sampling schemes and combinations of the autocovariance structure and number of sampling locations. An example with real data is given. In closing, we discuss the theoretical and practical aspects of our modified F tests in multiple correlation analysis, in comparison with those used in the repeated measures ANOVA.

Keywords: Effective sample size in F tests; Model-based versus model-free estimation; Multiple correlation; Sampling designs; Spatial processes and autocovariances; Validity and power of significance tests

1. Introduction

When the conditions of application of inferential methods are not met by sample data, essentially two options are offered to the data analyst: transform the data so that the transformed data satisfy the conditions, or develop a modified inferential method in which the conditions that are not met by the raw data are relaxed at the estimation stage or deviations from them are taken into account in the test of significance. In simple correlation analysis, the

* Corresponding author. Tel.: +1 514 3987870; fax: +1 514 3987897.
E-mail addresses: pierre.dutilleul@mcgill.ca (P. Dutilleul), bernard.pelletier@mcgill.ca (B. Pelletier), galpargu@fullerton.edu (G. Alpargu).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2007.06.022

autocovariances of stochastic processes are susceptible to bias the variance of sample correlation coefficients computed between partial realizations (Jenkins and Watts, 1968). In the spatial framework, Clifford et al. (1989) and Dutilleul (1993) therefore proposed modified t tests where the number of sampling locations N is replaced by an effective sample size M in the denominator of the test statistic and the associated number of degrees of freedom. This replacement aims at taking the autocovariances of the two spatial processes into account when both processes are autocorrelated and Pearson's product–moment correlation coefficient r is calculated from raw data. Hereafter, we concentrate on Dutilleul's (1993) modified t test, in which the complete expression for M is used.

In the context of multiple correlation analysis, consider the situation in which one spatial process is correlated with several (i.e., q) others and partial realizations have been collected for the $1+q$ processes at the same sampling locations. In such a situation, how to build a valid and powerful test on R, the sample multiple correlation, or equivalently, on R^2, the coefficient of determination? To answer this question, we followed the approach described above, by extending the modified t test into a modified F test. To do that, we basically used a multiple linear regression model in which the q explanatory variables, together with the dependent variable and the error, are random and indexed by space.

In linear models in general (e.g., Searle, 1971) and in spatially adjusted models of ANOVA and multiple regression in particular (Griffith, 1978; Cook and Pocock, 1983; Upton and Fingleton, 1985), the autocorrelation of errors plays an important role in the inefficiency of slope estimators and the invalidity of significance tests. This is clear in fixed ANOVA models and regression models with fixed explanatory variables, where errors provide the only random term on the right-hand side of the model equation. When regressors are random, their own autocorrelation may have an equally important role (Alpargu, 2001; Alpargu and Dutilleul, 2001, 2003a, b, 2006). To get some insight into this, consider a simple linear regression $Y(\mathbf{u}) = a + bX(\mathbf{u}) + \varepsilon(\mathbf{u})$, where the random terms are indexed by $\mathbf{u}$, which denotes a location in 2-D space. When $X(\mathbf{u})$ and $X(\mathbf{u}')$ with $\mathbf{u} \neq \mathbf{u}'$ are random and i.i.d., whether or not the errors are autocorrelated, the unmodified t test of significance of slope b is valid. In fact, the equivalent t test of significance of the simple correlation between the dependent variable and the regressor is then valid, with $M = N$. However, when the two spatial processes underlying $X(\mathbf{u})$ and $\varepsilon(\mathbf{u})$ are autocorrelated and their autocovariances follow some monotonic decreasing functions, both unmodified t tests (i.e., for the slope and for the simple correlation) are invalid.

In Section 2, we present the general but theoretical form of the modified F test that we developed to assess the multiple correlation between one spatial process and q others, from partial realizations collected at the same sampling locations. Since the effective sample size that appears in the denominator of the test statistic via the associated number of degrees of freedom needs to be estimated in practice, different estimation procedures are presented in Section 3, providing as many variants of the modified F test. The results of two simulation studies designed to analyze the validity and power of tests for various sampling schemes and combinations of the autocovariance structure and number of sampling locations are presented in Section 4. For comparison purposes, the classical, unmodified F test and the REstricted Maximum Likelihood (REML) likelihood-ratio χ^2 test are also studied through simulations. An example with real data is given in Section 5. Theoretical and practical aspects of our modified F tests are discussed in Section 6.

2. Modification of the F test

Below, we derive the theoretical expression of the effective sample size to be used in the modified F test that we propose for multiple correlation analysis with spatial data. The main steps are: (i) the effective sample size is defined under the null hypothesis of no multiple correlation between the 'dependent process' and the q 'explanatory processes'; (ii) we use a reparameterized form of Wishart's (1931, p. 356) formula for the expected value of the squared sample multiple correlation under the null hypothesis in the i.i.d. case, as a starting point; (iii) in the presence of spatial autocorrelation, the number of sampling locations is replaced by the effective sample size in the formula; (iv) the equality between the squared sample multiple correlation and the squared sample simple correlation between the dependent process and the 'predicted process' (i.e., the OLS predictor of multiple regression) allows the use of results for simple correlation analysis with spatial data, which simplifies the mathematical developments.

Let $\{\mathbf{Z}(\mathbf{u}) \mid \mathbf{u} \in D \subset \mathbb{R}^2\}$ be a $(1+q)$-variate Gaussian, first-order stationary spatial process, with $\mathbf{Z}(\mathbf{u}) = \{Y(\mathbf{u}), X_1(\mathbf{u}), \ldots, X_q(\mathbf{u})\}^T$ (cf. Dutilleul, 1993 for the modified t test when $q = 1$). Note that $D \subset \mathbb{R}$ would correspond to 1-D space or time, and first-order stationarity is sufficient for the theoretical derivation in this section but further assumptions may be required for the estimation of autocovariance matrices (Section 3). Hereafter, $\{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$ denotes a set of N sampling locations.

Consider a situation in which a partial realization of the $(1+q)$-variate spatial process provided sample data $\{y(\mathbf{u}_i), x_1(\mathbf{u}_i), \ldots, x_q(\mathbf{u}_i); i = 1, \ldots, N\}$, to correlate the first component and the other q. Let $R_{YX_1\ldots X_q} = (\mathbf{S}_{YX_1\ldots X_q} \mathbf{S}^{-1}_{X_1\ldots X_q X_1\ldots X_q} \mathbf{S}^{\mathrm{T}}_{YX_1\ldots X_q}/S^2_Y)$ denote the sample multiple correlation, with $\mathbf{S}_{YX_1\ldots X_q}$, the $1\times q$ sample covariance vector, S^2_Y, the sample variance for the first component, and $\mathbf{S}_{X_1\ldots X_q X_1\ldots X_q}$, the $q\times q$ sample variance–covariance matrix for the others. The coefficient of determination is given by $R^2_{YX_1\ldots X_q}$.

If the population multiple correlation $\rho_{YX_1\ldots X_q}(\mathbf{u})$ between $Y(\mathbf{u})$ and $X_1(\mathbf{u}), \ldots, X_q(\mathbf{u})$ is zero (i.e., the null hypothesis is true) and the random variables $Y(\mathbf{u}), Y(\mathbf{u}')$ with $\mathbf{u} \neq \mathbf{u}'$ are uncorrelated and homoscedastic (or the random variables $X_1(\mathbf{u}), X_1(\mathbf{u}'), \ldots, X_q(\mathbf{u}), X_q(\mathbf{u}')$ are so), then the expected value of the coefficient of determination is given by

$$E(R^2_{YX_1\ldots X_q}) = q/(N-1), \tag{1}$$

which can be rewritten as

$$N = 1 + q/E(R^2_{YX_1\ldots X_q}). \tag{2}$$

Eq. (2) above extends Eq. (3.3) of Dutilleul (1993) for simple correlation, since $E(R^2_{YX_1\ldots X_q}) = \mathrm{Var}(R_{YX_1\ldots X_q})$ if $\rho_{YX_1\ldots X_q}(\mathbf{u}) = 0$. The equalities in Eq. (2) here and in Eq. (3.3) in Dutilleul (1993) are exact because they are written under the null hypothesis of no correlation. It is when the expected value of the squared sample correlation coefficient is calculated under the alternative hypothesis that equalities are approximate (Wishart, 1931, p. 356; see also Morrison, 1990, p. 108).

When the autocovariances of processes are such that Eq. (1) does not hold, the left-hand side of Eq. (2) must be changed for an effective sample size given by

$$M = 1 + q/E(R^2_{YX_1\ldots X_q}). \tag{3}$$

In Eq. (3), the expected value of the coefficient of determination incorporates the autocovariances of the $1+q$ processes. However, the calculation of $E(R^2_{YX_1\ldots X_q})$ from $R^2_{YX_1\ldots X_q} = \mathbf{S}_{YX_1\ldots X_q}\mathbf{S}^{-1}_{X_1\ldots X_q X_1\ldots X_q}\mathbf{S}^{T}_{YX_1\ldots X_q}/S^2_Y$, by developing the inverse of $\mathbf{S}^{-1}_{X_1\ldots X_q X_1\ldots X_q}$ and by using the properties of quadratic and bilinear forms in normal vectors (Searle, 1971), rapidly becomes a challenge with increasing q. A more practical way to proceed is to take advantage of the equality between the coefficient of determination and the square of some sample simple correlation, as shown below.

If $r_{Y\hat{Y}}$ denotes the sample simple correlation (i.e., Pearson's product–moment coefficient) between the dependent process Y and the predicted process $\hat{Y}$, then $R^2_{YX_1\ldots X_q} = r^2_{Y\hat{Y}}$ and the results of simple correlation analysis with spatial data (Dutilleul, 1993) readily apply by replacing X_1 (i.e., $q = 1$) by $\hat{Y}$. Under the bivariate Gaussian and first-order stationarity assumptions for the dependent and predicted processes, it follows that

$$M = 1 + 1/E(r^2_{Y\hat{Y}}). \tag{4}$$

If $\mathbf{\Sigma}_Y$ and $\mathbf{\Sigma}_{\hat{Y}}$ denote the $N\times N$ theoretical autocovariance matrices of the dependent and predicted processes considered at the same N sampling locations and $\mathbf{B} = N^{-1}(\mathbf{I}_N - N^{-1}\mathbf{J}_N)$ with $\mathbf{I}_N$ and $\mathbf{J}_N$ the $N\times N$ identity matrix and the $N\times N$ matrix of ones, then $E(r^2_{Y\hat{Y}})$ (i.e., the expected value of a ratio) can be approximated by $\mathrm{tr}(\mathbf{B\Sigma}_Y\mathbf{B\Sigma}_{\hat{Y}})/\{\mathrm{tr}(\mathbf{B\Sigma}_Y)\mathrm{tr}(\mathbf{B\Sigma}_{\hat{Y}})\}$ (i.e., the ratio of expected values), so that

$$M \approx 1 + \left\{\mathrm{tr}(\mathbf{B\Sigma}_Y)\mathrm{tr}(\mathbf{B\Sigma}_{\hat{Y}})\right\}/\mathrm{tr}(\mathbf{B\Sigma}_Y\mathbf{B\Sigma}_{\hat{Y}}). \tag{5}$$

Note that the first-order stationarity and Gaussian distribution of the $(1+q)$-variate process imply the first-order stationarity and Gaussian joint distribution of the dependent and predicted processes because these are, respectively, one of the components and one linear combination of the components of the $(1+q)$-variate process.

Finally, we propose the statistic

$$(R^2_{YX_1\ldots X_q}/q)/\{(1-R^2_{YX_1\ldots X_q})/(M-q-1)\} \tag{6}$$

to test H_0: $\rho^2_{YX_1\ldots X_q}(\mathbf{u}) = 0$ against H_1: $\rho^2_{YX_1\ldots X_q}(\mathbf{u}) > 0$, and the Fisher–Snedecor distribution $F(q, M-q-1)$ to approximate the distribution of (6) under H_0, where M is given by the right-hand side of (5). The rationale behind the proposed approximation is the following. The expression of test statistic (6) is identical to that of the classical, unmodified F test, except M replaces N in the expression of the denominator. For the rest, the modified F test statistic is built as a ratio of two mean squares derived from quadratic forms in normal vectors, leading to a natural Fisher–Snedecor distribution approximation. Using (3), it is easy to show that the expected value of the numerator (i.e., $1/(M-1)$) is equal to that of the denominator under the null hypothesis, that is, the ratio of expected values is 1.0 under H_0. The decision-making rule reads as follows: if the observed value of test statistic (6) is greater than $F_{1-\alpha}(q, M-q-1)$, then one rejects H_0 at level α; otherwise, one accepts H_0 at level α. In the following sections, we elaborate on the options available for the estimation of M in practice, and study the validity and power of the corresponding variants of the modified F test empirically, using simulations.

3. Estimation of the effective sample size

In applications, the effective sample size M must be estimated from the data (i.e., the directly observable, partial realization of the dependent process and the partial realization of the predicted process obtained by OLS multiple regression of $y(\mathbf{u}_i)$ on $x_1(\mathbf{u}_i), \ldots, x_q(\mathbf{u}_i)$; $i = 1, \ldots, N$). In general terms, this estimation will be performed by using autocovariance matrix estimates $\hat{\mathbf{\Sigma}}_Y$ and $\hat{\mathbf{\Sigma}}_{\hat{Y}}$ to evaluate the right-hand side of (5), which provides an estimated effective sample size

$$\hat{M} = 1 + \{\mathrm{tr}(\mathbf{B}\hat{\mathbf{\Sigma}}_Y)\mathrm{tr}(\mathbf{B}\hat{\mathbf{\Sigma}}_{\hat{Y}})\}/\mathrm{tr}(\mathbf{B}\hat{\mathbf{\Sigma}}_Y\mathbf{B}\hat{\mathbf{\Sigma}}_{\hat{Y}}). \tag{7}$$

If the homoscedasticity condition applies to the autocovariance structures of the processes, that is, $\mathbf{\Sigma}_Y = \sigma_Y^2\mathbf{V}_Y$ and $\mathbf{\Sigma}_{\hat{Y}} = \sigma_{\hat{Y}}^2\mathbf{V}_{\hat{Y}}$, where $\mathbf{V}_Y$ and $\mathbf{V}_{\hat{Y}}$ denote the theoretical autocorrelation matrices of Y and $\hat{Y}$ at the N sampling locations, then it is sufficient to estimate $\mathbf{V}_Y$ and $\mathbf{V}_{\hat{Y}}$ and replace $\hat{\mathbf{\Sigma}}_Y$ and $\hat{\mathbf{\Sigma}}_{\hat{Y}}$ by the autocorrelation matrix estimates $\hat{\mathbf{V}}_Y$ and $\hat{\mathbf{V}}_{\hat{Y}}$ in (7); the variances which appear in both the numerator and the denominator of the ratio are canceled out. Under second-order stationarity and isotropy, Moran's (1950) I statistic evaluated at a certain number of distances or distance classes can be used to fill $\hat{\mathbf{V}}_Y$ and $\hat{\mathbf{V}}_{\hat{Y}}$,without modeling the autocorrelation functions. Under slightly different assumptions, the autocovariance estimator derived from Matheron's classical semivariance estimator and one minus Geary's (1954) c statistic (i.e., a standardized semivariance estimator) can be used similarly. From the three resulting 'model-free variants' of the modified F test, the second is studied in the next sections. In the geostatistical approach, variograms are usually modeled, and we focus on this approach below.

Different 'model-based variants' of the modified F test can be thought of, depending on the variogram model (e.g., the number of basic functions) and the procedure followed for its fitting. In all cases, the estimated effective sample size must be evaluated under the null hypothesis, that is, in the absence of cross correlation between the dependent and predicted processes. Accordingly, the cross variogram is excluded from the fitting. A linear model of regionalization (LMR) (Goovaerts, 1997, pp. 95–97) is fitted separately to the direct experimental variograms of the dependent and predicted processes in the 'LMR-based variant' of the modified F test studied here. In each variogram fitting, the ranges (i.e., nonlinear parameters) are estimated first and then the sills (i.e., linear parameters) are estimated by using the range estimates. Such fitting in two stages follows the least squares estimation procedure designed by Golub and Pereyra (1973) for nonlinear models whose parameters separate. Here, ranges are estimated by weighted least squares (Zhang et al., 1995), and sills by estimated generalized least squares (EGLS) (Pelletier et al., 2004). Finally, the ranges of spatial autocorrelation and the vectors of sills estimated for the respective basic variogram functions are used to define $\hat{\mathbf{\Sigma}}_Y$ and $\hat{\mathbf{\Sigma}}_{\hat{Y}}$ in (7). Matheron's classical semivariance estimator is used to compute experimental variograms.

In Section 6, we discuss other ways to estimate autocovariance matrices for the application of the modified F test in and outside the spatial framework. The theoretical definition of M requires no particular assumption on autocovariances, but the estimation of spatial autocovariances will generally require some type of stationarity or isotropy.

4. Validity and power analyses

4.1. Design of simulation studies

We have conducted two simulation studies (denoted **I** and **II** hereafter), to assess the effects of a number of factors on the performance of different variants of the modified F test. Among these factors are the number of sampling locations

(N) and the sampling scheme. Values of N are 225, 400 in study **I**, where the scheme is a regular or irregular 2-D grid. Regular grids are square (i.e., 15×15 for $N = 225$ and 20×20 for $N = 400$). Three types of irregular grid with same area as the corresponding regular grid with same N are considered: completely random, stratified random (9 strata for $N = 225$ versus 16 strata for $N = 400$; the area of one stratum $=$ the area covered by a 5×5 subgrid of a regular grid), and 'hybrid' (a 15×15 regular grid plus a completely random grid with $N = 175$). Values of N are 75, 150 in study **II**, where sampling is supposed to be performed at equally spaced locations on a transect. The value of q (i.e., the number of explanatory processes) is 2 in both simulation studies. Other simulation parameters are: the theoretical significance level (α), the squared population multiple correlation at same location ($\rho^2_{YX_1\ldots X_q}(\mathbf{u})$, denoted $\rho^2_{YX_1X_2}$ hereafter), the number of random structural components, the range or practical range of spatial autocorrelation, the basic variogram functions, and the corresponding sill values; it must be noted that spatial autocorrelations of the dependent and explanatory processes are the same in study **I**, but are different in study **II**.

The main objectives of study **I**, which is broader than study **II**, are to analyze the validity and, when justified, the power of different variants of the modified F test in a number of situations to which users can relate. Estimated p-values are provided by the rejection rates of the null hypothesis. In addition to the LMR-based and model-free variants (Section 3), one theoretical variant with sills and basic variogram functions specified correctly and a model-based variant with sills to be estimated while the range of spatial autocorrelation is assumed known are studied for comparison purposes. The classical, unmodified F test and the REML likelihood-ratio χ^2 test are also studied; for the latter, equations and formulas from Diggle et al. (1996, pp. 92–95) were adjusted for 2-D spatial data, by using the Euclidean distances between sampling locations to build variance–covariance matrices. As for study **II**, the classical, unmodified F test is compared under the null hypothesis to three variants of the modified F test: theoretical, LMR-based, and model-free.

In study **I**, there are two random structural components: a nugget effect and a spherical variogram model. Values of the range of spatial autocorrelation are 2, 3, 4, 5, and 6. The absence of spatial autocorrelation (i.e., pure nugget effect or range $= 1$ for the regular grid) is also considered to study the performance of the modified F test in the conditions of application of the unmodified F test; in this case as in the cases with spatial autocorrelation, two structures were used for the LMR-based variant of the modified F test only if the sill estimates for the spherical component reached the 0.05 threshold. The value of $\rho^2_{YX_1X_2}$ varies from 0.0 to 0.9 by steps of 0.1, and the values of population simple correlations at same location involving Y are the same at each structure. In study **II**, the dependent and explanatory processes are characterized by one random structural component defined by an exponential variogram model (i.e., a 1-D spatial AR(1) model). The autoregressive coefficient for X_1, X_2 is equal to minus that for Y, which is positive (cf. Dutilleul, 1993, for $q = 1$). Accordingly, the theoretical variant of the modified F test was adjusted by calculating the spatial autocovariances of the predicted process $\hat{Y}$, conditionally to the partial realizations of X_1 and X_2. The specific objective of **II** is to assess the validity of the modified F test in a situation where the effective sample size is expected to be greater than N because of the opposition of sign between autoregressive coefficients. The practical range values varying from 2 to 6, a total of 15 combinations of autoregressive coefficient values are considered. In both simulation studies, the values of α are 0.01, 0.05, and 0.10; without loss of generality, the two explanatory processes are uncorrelated; and the number of simulation runs for a given combination of parameter values is 5000.

The main part of our procedure to simulate spatial data is identical to that of Pelletier et al. (2004). That is, the same matrix algebra tools and function of normally distributed random values (i.e., randn) were used (The MathWorks, 2006).

4.2. Simulation results

Validity and power analysis results for study **I** are presented in Figs. 1 and 2, respectively; power is depicted only for valid testing procedures. Validity analysis results for study **II** are presented in Fig. 3. To interpret Figs. 1 and 3, we use the following criterion: for a given combination of simulation parameter values, a test is said to be valid at the theoretical significance level α if the corresponding empirical significance level $\hat{p}$ is smaller than α or the approximate 95% confidence interval $\hat{p} \pm 2\sqrt{\hat{p}(1-\hat{p})/5000}$ contains the value of α. Descriptive statistics for the estimated effective sample size are reported in Tables 1–3. The number and content of tables can be explained as follows. True distances were used to compute experimental variograms in the case of the regular sampling grid (Table 1). For all types of sampling grid (i.e., regular, completely random, stratified random, and hybrid) considered in study **I**, equally spaced

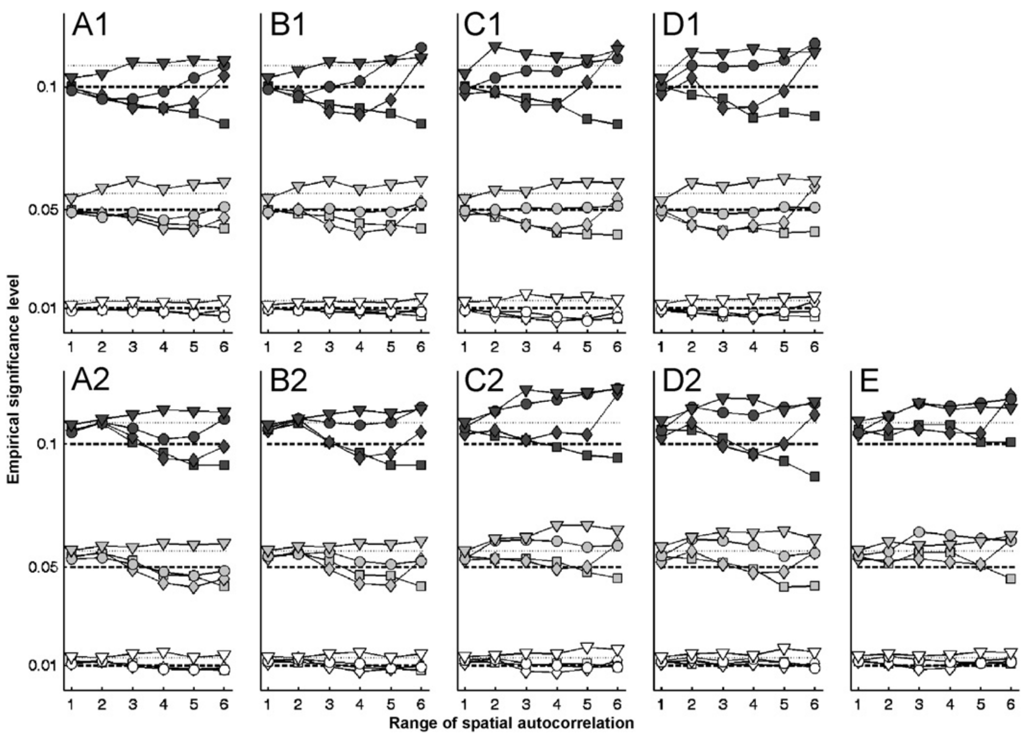

Fig. 1. Validity analysis results for simulation study **I**, in relation to the type of sampling grid, the number of sampling locations, and the use of true distances versus distance classes in the computation of direct experimental variograms when possible. The sampling grids are: square regular, (A1)–(A2) and (B1)–(B2); completely random, (C1)–(C2); stratified random, (D1)–(D2); and hybrid, (E). The number of sampling locations is 225 in panels (A1), (B1), (C1), and (D1), and 400 in the other panels. True distances were used in the computation of direct experimental variograms in (A1)–(A2), whereas distance classes were used in (B1)–(B2). With the exception of the REML likelihood-ratio χ^2 test, all tests are variants of the modified F test, which are represented by squares (theoretical), diamonds (model-based, two estimated ranges), and circles (model-free); the REML likelihood-ratio χ^2 test is represented by downward pointing triangles. Dark gray symbols, $\alpha = 0.10$; light gray symbols, $\alpha = 0.05$; and empty symbols, $\alpha = 0.01$. The highest value of empirical significance level for which the approximate 95% confidence interval centered on it contains the corresponding value of α is indicated by a fine dotted horizontal line.

distance classes defined from 1 to 7 by steps of 1 were used to compute experimental variograms prior to estimating the effective sample size in the model-based and model-free variants of the modified F test (Table 2). Because of space constraints and some redundancy, descriptive statistics for the estimated effective sample size in the case of the hybrid sampling grid are not reported and results for the model-based variant with known range are reported only in Table 1. Distances from 1 to 7 by steps of 1 were used in the computation of experimental variograms in study **II**, where equally spaced distance classes matched true distances (Table 3). Since the homoscedasticity assumption holds in both simulation studies, only the autocorrelation of spatial processes, as modeled by a given autocovariance structure and measured by the range or practical range, matters for the validity of tests. Below, the results obtained for study **I** and study **II** are commented separately.

In study **I**, the modified F tests are valid in the great majority of cases, but the validity of some of them appears to be dependent on the sampling scheme and the theoretical significance level. At $\alpha = 0.01$, all modified F tests are found to be valid, with one exception in $9 \times 6 \times 4 = 216$ cases for the model-free variant with range $= 2$ for the hybrid grid. At $\alpha = 0.05$, 9 of the 15 invalidity cases for modified F tests are for the model-free variant for irregular sampling grids with $N = 400$; the six others are for the model-based variant with known range (3 cases) and the LMR-based variant (3 cases). At $\alpha = 0.10$, there are 36 invalidity cases for modified F tests; 25 of them are for the model-free variant and the other eleven are distributed almost evenly (5 against 6) between the same two model-based variants as for $\alpha = 0.05$ (essentially for range $= 2$ for the former and for range $= 6$ for the latter). It must be noted that the model-free variant of the modified F test was found to be invalid only once in 36 cases when true distances were used for the regular sampling grid, compared to 5 when distance classes were used for the same type of sampling grid and values of α, N

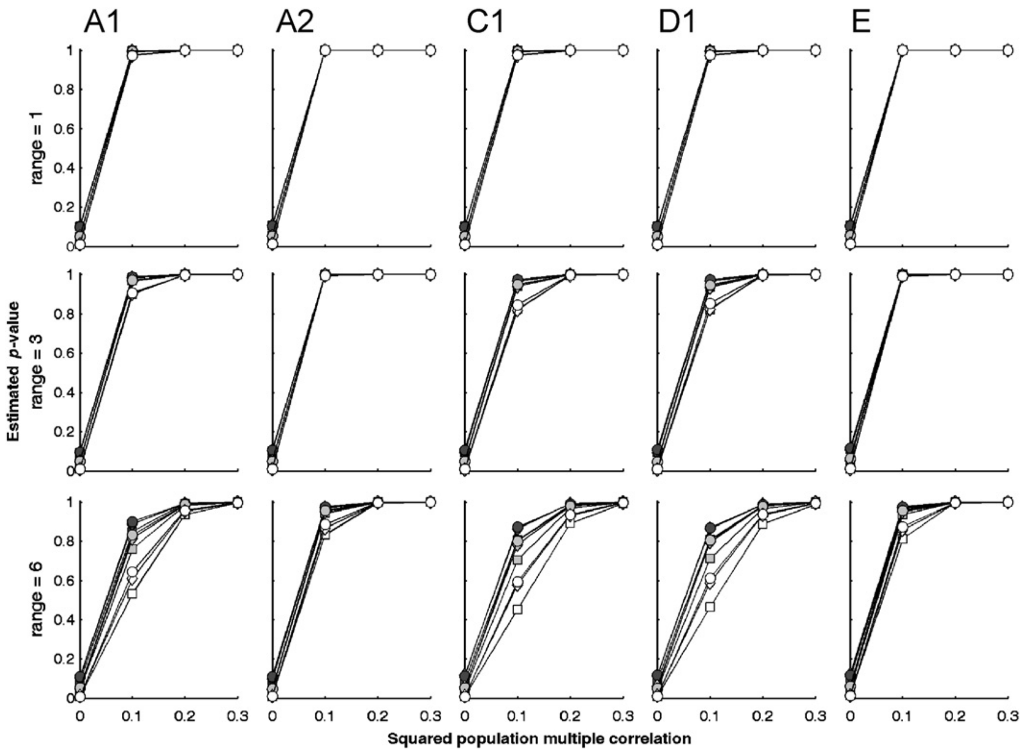

Fig. 2. Power analysis results for simulation study **I**. See legend of Fig. 1 for the definition of panels and the meaning of symbols and gray tones.

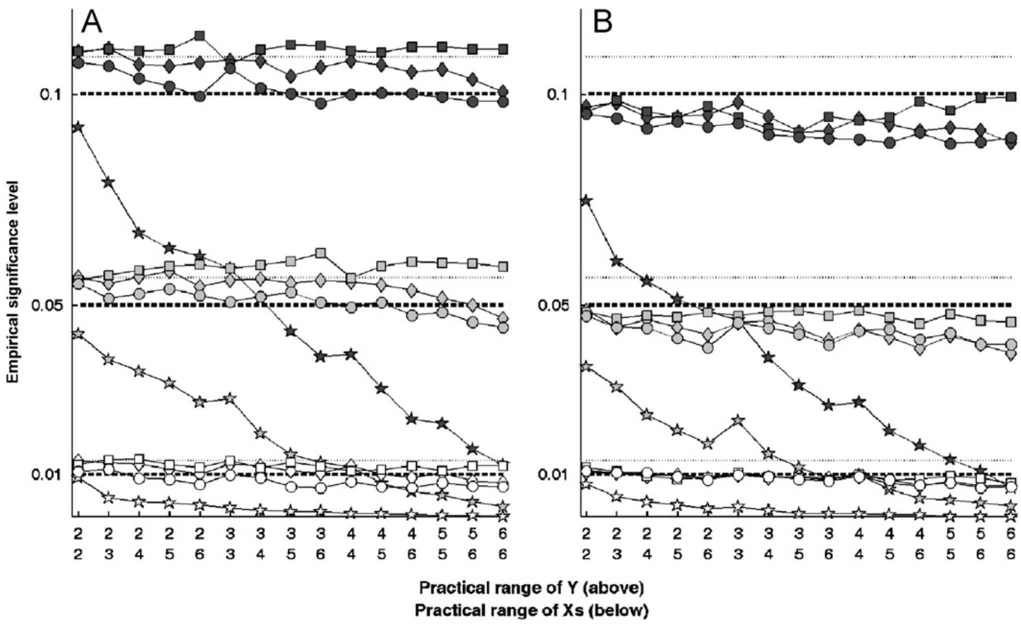

Fig. 3. Validity analysis results for simulation study **II**. Three of the four tests studied are variants of the modified F test, which are represented by squares (theoretical), diamonds (model-based, two estimated ranges), and circles (model-free); the fourth test is the classical, unmodified F test, which is represented by stars. The two panels correspond to different numbers of sampling locations regularly spaced on a transect: (A) $N = 75$ and (B) $N = 150$. The definition of gray tones and dotted lines is the same as in Fig. 1.

Table 1
Descriptive statistics for the estimated effective sample size in four variants of the modified F test (by using true distances for all variants) in the case of the square regular sampling grid in simulation study **I**

	$N = 225$ (i.e., 15×15)						$N = 400$ (i.e., 20×20)					
Range of spherical component	1[a]	2	3	4	5	6	1	2	3	4	5	6
Theoretical												
Mean	225.0	204.4	164.0	129.3	103.5	85.2	400.0	362.4	288.3	223.9	175.9	141.4
Std	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Model-based, known range												
Mean	225.0	205.3	166.8	133.9	109.7	92.0	400.0	363.0	290.5	228.0	181.4	147.6
Std	0.0	9.3	15.6	19.3	21.0	21.6	0.0	12.0	20.5	24.7	26.0	25.7
Min	225.0	171.4	119.4	83.2	60.9	46.5	400.0	316.0	222.3	159.8	110.3	83.3
Max	225.0	225.0	225.0	219.2	224.1	225.0	400.0	400.0	373.0	335.1	317.5	305.4
Model-based, two estimated ranges												
Mean	224.2	206.7	166.2	134.3	113.1	100.2	399.1	364.8	291.0	229.5	185.9	157.2
Std	2.1	12.8	19.6	25.2	27.7	28.4	2.3	17.5	26.4	34.4	38.1	38.1
Min	196.0	140.1	60.4	47.6	39.0	40.6	355.2	295.9	187.9	122.6	93.5	77.2
Max	225.0	225.0	223.0	223.9	222.5	223.2	400.0	400.0	394.0	357.8	331.7	322.8
Model-free												
Mean	225.1	206.3	169.7	139.6	118.4	104.4	400.0	364.0	294.6	236.5	194.5	165.7
Std	9.1	13.0	19.5	24.0	26.4	27.6	9.3	15.5	25.6	32.1	35.4	37.1
Min	190.6	149.1	89.5	57.3	40.8	32.5	350.7	299.0	196.2	132.2	91.3	65.9
Max	268.2	264.6	259.8	234.4	224.5	221.6	441.4	413.7	381.7	353.8	321.5	319.0

[a]Here and in Table 2, there is no spherical component or there is a pure nugget effect for this value of range.

and range. In the end, the theoretical modified F test was invalid for none of the sampling schemes and none of the values of α, N and range considered in study **I**; all variants of the modified F test were valid, without exception, in the conditions of application of the classical, unmodified F test (i.e., the absence of spatial autocorrelation); and when some of the variants of the modified F test were found to be invalid in the presence of spatial autocorrelation, the lower bound of the confidence interval built on $\hat{p}$ was, in general, slightly greater than α.

The above validity analysis results can be related to the behavior of the estimated effective sample size, depending on the variant of the modified F test, the range of autocorrelation of the spatial processes, the sampling scheme, the number of sampling locations, and the use of true distances versus distance classes (Tables 1 and 2).

(i) Overall, the Mean value of estimated effective sample sizes for the model-free variant of the modified F test diverges the most from the corresponding Theoretical value. This is reflected by Max values that can exceed N, due to raw autocorrelation estimates of different sign for the dependent and predicted processes at several lags in the model-free variant (Table 1). The model-free variant is also the only one for which the autocovariance matrix estimates are not necessarily positive definite, although this is not a prerequisite for trace properties (Graybill, 1983, Theorem 9.1.23, p. 304; see also Section 6 here). One might expect the difference between theoretical and empirical Mean values to increase from top to bottom in the tables, that is, from model-based with known range to model-free, as knowledge about the autocovariance structures decreases and there are less modeling assumptions. This is true only for very short and very long ranges, possibly because of ergodic fluctuations for ranges of intermediate length; there is no guarantee that working with known range is then better than working with estimated ranges.

(ii) As expected, the estimated effective sample sizes decrease with the range of autocorrelation of the spatial processes when the theoretical autocorrelation values of Y and X_1, X_2 are of same sign, as in study **I**. All other factors being equal, the stronger the autocorrelation is, or the longer the range is, the lower the Mean value of estimated effective sample sizes, but the higher the Std.

(iii) The pattern described in (ii) applies to all sampling schemes, with some differences. For the same values of N and range, the Mean values of estimated effective sample sizes in the presence of autocorrelation tend to be lower

Table 2
Descriptive statistics for the estimated effective sample size in three variants of the modified F test (by using true distances for the Theoretical variant and distance classes for the two others) in simulation study **I**

	$N = 225$						$N = 400$					
Range of spherical component	1	2	3	4	5	6	1	2	3	4	5	6
Square regular sampling grid												
Theoretical												
Mean	225.0	204.4	164.0	129.3	103.5	85.2	400.0	362.4	288.3	223.9	175.9	141.4
Std	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Model-based, two estimated ranges												
Mean	224.1	208.8	165.3	133.8	113.9	102.8	399.0	368.6	289.3	228.4	186.3	161.5
Std	2.3	14.3	20.0	24.9	27.1	27.1	2.4	21.4	26.3	34.6	37.7	36.3
Model-free												
Mean	224.8	207.0	172.2	142.3	120.5	106.2	399.8	366.0	299.7	241.4	198.3	168.6
Std	4.8	10.6	18.3	23.4	26.2	27.6	5.0	13.6	24.7	32.0	35.6	37.4
Completely random sampling grid												
Theoretical												
Mean	225.0	174.6	139.1	110.6	89.7	74.8	400.0	310.7	246.1	193.2	153.8	125.1
Std	0.0	2.1	2.0	1.7	1.4	1.2	0.0	2.8	2.7	2.3	1.9	1.6
Model-based, two estimated ranges												
Mean	221.8	178.0	141.2	116.9	101.7	94.1	396.0	316.7	248.7	200.1	166.6	148.1
Std	6.7	18.5	23.0	26.1	28.0	28.9	9.4	27.2	31.0	36.5	38.4	38.2
Model-free												
Mean	224.9	184.0	151.2	125.7	108.1	96.9	399.8	325.8	263.9	214.2	178.3	153.5
Std	4.8	17.5	23.5	27.1	29.0	30.2	4.9	23.3	31.7	36.1	38.3	39.1
Stratified random sampling grid												
Theoretical												
Mean	225.0	175.7	140.4	111.8	90.7	75.6	400.0	313.0	248.9	195.8	155.9	126.8
Std	0.0	2.0	1.9	1.6	1.3	1.1	0.0	2.6	2.4	2.0	1.7	1.4
Model-based, two estimated ranges												
Mean	222.2	180.0	143.3	118.6	102.9	95.4	396.1	317.8	250.7	201.7	167.9	148.9
Std	6.3	17.9	22.3	25.7	27.7	28.7	9.1	26.4	30.3	36.1	38.5	38.0
Model-free												
Mean	224.8	185.3	152.7	127.3	109.3	97.6	399.9	328.0	265.8	215.4	179.3	154.4
Std	4.7	17.1	22.7	26.2	28.1	29.4	4.9	22.1	30.7	35.5	38.0	39.3

for irregular sampling grids. More specifically, the decrease of $\hat{M}$ with increasing range length is faster for the completely random grid than for the stratified random grid and the regular grid in this order, showing that the complete or partial randomness of sampling is taken into account in the evaluation of the estimated effective sample size. The behavior of $\hat{M}$ in the case of the hybrid grid with $N = 400$ (numerical results not reported) falls between the stratified random grid and the regular grid. Standard errors follow a reverse pattern, and are generally higher for the completely random grid.

(iv) The number of sampling locations N seems to affect the model-free variant of the modified F test, as more than two thirds of its invalidity cases are for $N = 400$. In view of Table 2, the differences between empirical Mean values of $\hat{M}$ for the model-free variant and the other variants of the modified F test appear larger for $N = 400$ than for $N = 225$, particularly for the completely random and stratified random grids. With one exception (i.e., for $N = 400$), the model-free variant is valid at $\alpha = 0.01$.

(v) Besides some computational advantages, the use of distance classes instead of true distances in the case of the regular sampling grid provides estimated effective sample sizes with lower Std values at short ranges for the model-free variant, at the expense of higher Mean values at these and other ranges, which may explain the few invalidity cases of that variant at $\alpha = 0.10$. No other noticeable difference is to be reported (Tables 1 and 2).

Table 3
Descriptive statistics for the estimated effective sample size in three variants of the modified F test (by using true distances for the three variants) in the case of the regular transect sampling in simulation study **II**

Practical range of dependent process	2	3	4	5	6	3	4	5	6	4	5	6	5	6	6
Practical range of explanatory processes[a]	2	2	2	2	2	3	3	3	3	4	4	4	5	5	6
$N = 75$															
Theoretical															
Mean	82.1	87.1	91.0	94.0	96.3	96.0	103.2	108.9	113.5	113.1	121.4	128.1	131.6	140.3	150.5
Std	2.9	3.0	3.0	3.0	2.9	5.8	6.0	6.1	6.2	9.1	9.5	9.7	12.8	13.2	16.7
Model-based															
Mean	82.6	87.7	91.8	94.8	96.9	96.2	103.4	108.8	112.7	113.1	120.7	126.2	129.4	136.1	142.9
Std	6.5	9.2	11.4	13.0	14.0	12.6	15.4	17.3	18.6	18.2	20.1	21.3	21.5	22.3	22.2
Model-free															
Mean	81.9	86.9	90.8	94.0	96.6	95.7	102.7	108.3	112.9	112.6	120.4	126.8	130.4	138.4	148.4
Std	7.8	10.3	12.9	15.3	17.6	12.9	15.8	18.6	21.3	19.0	22.2	25.2	25.6	28.9	32.7
$N = 150$															
Theoretical															
Mean	164.9	175.5	183.6	189.8	194.7	194.4	209.7	221.8	231.4	231.0	248.7	263.1	270.9	289.8	312.3
Std	4.2	4.3	4.3	4.2	4.2	8.2	8.5	8.7	8.7	12.9	13.4	13.7	18.2	18.9	24.0
Model-based															
Mean	165.1	175.9	184.2	190.5	194.9	194.8	210.2	222.1	230.7	231.9	249.1	261.6	270.1	286.2	303.3
Std	9.5	13.6	17.1	19.8	21.5	18.5	22.8	26.1	28.1	27.9	31.7	33.7	35.1	36.7	36.4
Model-free															
Mean	164.9	175.3	183.5	189.9	195.1	194.3	209.2	221.2	230.9	230.7	247.8	261.8	270.0	288.0	310.3
Std	9.9	13.7	17.6	21.3	24.7	17.5	21.8	26.2	30.4	26.5	31.2	35.9	36.4	41.6	47.2

[a]For the dependent process, spatial autocorrelation is modeled to be non-negative and decrease exponentially. For the explanatory processes, it is modeled to alternate in sign (i.e., to be negative at odd lags and positive at even lags) and to decrease exponentially in absolute value.

In study **I**, the unmodified F test is valid in the absence of spatial autocorrelation, but rapidly becomes excessively invalid when the range of the spherical structure increases. Accordingly, this test is absent from Fig. 1. The test found to be the second most invalid in study **I** is the REML likelihood-ratio χ^2 test. In fact, with a few exceptions in the absence of spatial autocorrelation and when the range length is greater than or equal to 2 for the 15×15 regular grid and the $N = 225$ completely random grid, this test is invalid (Fig. 1). Accordingly, it was not included in the power analysis. This result is in agreement with results obtained by Alpargu and Dutilleul (2006) for the REML likelihood-ratio χ^2 test in stepwise regression with random explanatory variables and errors that were temporally autocorrelated following a spherical autocovariance structure.

In view of Fig. 2, the power of all modified F tests reaches 1.0 relatively rapidly, that is, before or at $\rho^2_{YX_1X_2} = 0.3$, depending on the sampling scheme and the values of N and range. In the absence of spatial autocorrelation (top panels), the power of modified F tests is very similar to that of the classical, unmodified F test; this could be expected from the corresponding estimated effective sample sizes (Tables 1 and 2). For obvious reasons, the power is higher for $N = 400$ (Fig. 2(A2)) than for $N = 225$ (Fig. 2(A1)), all other factors (i.e., sampling scheme, values of α and range) being the same. For given sampling scheme and N value, the power of modified F tests slowly decreases with increasing range length, that is, with decreasing estimated effective sample sizes (Tables 1 and 2). Without being really low, the power is the lowest when a completely random grid is used (Fig. 2C), because of smaller estimated effective sample sizes for this sampling scheme (Table 2). Differences in the estimated effective sample size among modified F tests increase with increasing range length, and so do the differences in power among the tests (bottom panels of Fig. 2). When valid, the model-free variant is the most powerful, but this modified F test is not always valid, especially at $\alpha = 0.10$ (see above). The least powerful modified F test, without being conservative, is the theoretical variant, which was included here for comparative purposes, and was never found to be invalid in study **I**. Compared to the theoretical variant, the variant with known range and the LMR-based variant are slightly more powerful, but can be invalid occasionally. For given sampling scheme and range value, differences in power among modified F tests are smaller for larger N.

When the theoretical autocorrelations of Y are all positive while those of X_1, X_2 alternate in sign, as in study **II**, patterns are completely reversed. The unmodified F test in **II** shows a drop-off in rejection rate under the null hypothesis, with increasingly long ranges for the dependent and explanatory processes (Fig. 3). More specifically, with $N=75$, the empirical significance level of the unmodified F test starts at about 0.09 (double range of 2) to decrease at 0.01 (double range of 6), while $\alpha=0.10$; at $\alpha=0.01$, the empirical significance level of the test is near zero from the double range of 4 on; and with $N=150$, the situation is even worse. The three modified F tests succeed in remedying the situation, via estimated effective sample sizes greater than N or $2N$ (Table 3). For $N=75$, there are even several excessive values of $\hat{p}$ at $\alpha=0.05$ and 0.10 for the theoretical modified F test, while the two other modified F tests are valid, with just a few exceptions. For $N=150$, there is no invalidity case to report and the empirical significance levels of all three modified F tests match the theoretical significance level, especially when $\alpha=0.01$. In study **II**, unlike study **I**, both Mean and Std values of estimated effective sample sizes increase with increasing range lengths (Table 3). In accordance with expectations, the dispersion of $\hat{M}$ increases from the theoretical to the model-free variant of the modified F test, with the LMR-based variant generally in-between.

5. An example with real data

The spatial data analyzed here were collected in a micro-watershed of central Malawi in the fall of 1996, as part of a study investigating factors influencing soil quality and maize yield in agro-ecosystems (Pelletier, 2000). The study area was about 320 m × 1700 m. Among the variables sampled, maize yield (Y), pH (X_1), extractable nitrogen (X_2), microbial biomass (X_3), and carbon in floating particulate organic matter (X_4) are used for an agro-environmental example. There are $N=152$ sampling sites located on an irregular grid. To improve the normality of their distribution, the data were submitted to a Box–Cox transformation, which was followed by a standardization to zero mean and unit variance.

Distance classes 25-m wide were used to compute direct experimental variograms. We followed the guidelines of Goulard and Voltz (1992) for the identification of structures to be included in linear models of (co)regionalization. One structure (i.e., one basic variogram function) was not sufficient to describe the spatial distribution of variables in our example, and three was excessive (i.e., the dependent and predicted processes did not appear to possess three structures). Therefore, two structures were considered: a nugget effect, which incorporated the measurement error and spatial sources of variation at distances smaller than the shortest sampling distance, and a spherical component, which accounted for spatial autocorrelation up to a range to be estimated. A spherical component was chosen because it provided a better goodness of fit than exponential and Gaussian basic variogram functions. Estimated range values obtained by fitting a LMR to each of the two direct experimental variograms separately are 89 m for maize yield and 67 m for predicted maize yield.

The smaller value of N in simulation study **I** was 225, and the number of sampling locations was not greater than 150 in simulation study **II**, where sampling was supposed to be conducted in 1-D space (i.e., on a transect) instead of 2-D space. Moreover, the number of explanatory variables was two in both simulation studies, and is four here. Therefore, we have run supplementary simulations, using the same sampling scheme as in our example and similar spatial autocovariance structures and multiple correlations values. The results obtained with 5000 simulation runs showed that in a framework like that of our example, the theoretical, LMR-based, and model-free variants of the modified F test are valid at $\alpha=0.01$, 0.05, and 0.10, and their power reach 1.0 by the squared population multiple correlation value of 0.3.

Eventually, the results of significance tests (on real data) can be summarized as follows: $R^2=0.07$; $P=0.027$ with the classical, unmodified F test; with modified F tests, $P=0.095$ and 0.197 for the LMR-based variant ($\hat{M}=111.0$) and the model-free variant ($\hat{M}=85.5$), respectively. In conclusion, the unmodified F test exaggerates the significance of the multiple correlation, one modified F test merely keeps its significance at $\alpha=0.10$, and the association of maize yield with the four soil variables is declared non-significant by the other modified F test. MATLAB (The MathWorks, 2006) codes implementing the modified F tests with real data are available from http://environmetricslab.mcgill.ca.

6. Discussion

The definition of effective sample size is an exercise that must be repeated whenever the inferential situation changes. When population means are the parameters of interest in fixed or generalized linear models, the definition of effective sample size is based on the variance of mean estimators, in particular the denominator of that variance written as a ratio. This is the case in the modified t test for the comparison of two means (Cliff and Ord, 1975), the confidence interval estimation for one mean (Cressie, 1993, Chapter 1), and the calculation of the sample size required to achieve the nominal power when covariates are fixed in linear regression with repeated measures (Liu and Liang, 1997). In the repeated measures ANOVA, the definition of the multiplicative factor applied to the numbers of df in the modified F test, called 'epsilon' and due to Box (1954a, b), follows from properties of the χ^2 distribution of ANOVA mean squares under the appropriate assumptions. In correlation analyses, the variance of the correlation estimator, or equivalently the expected value of its square under the null hypothesis of no correlation, is the basis for the definition of the effective sample size. In all those situations, the effective sample size is a function of the autocorrelation and heteroscedasticity of the processes via their autocovariances, but is not and cannot be dependent on the population parameters that are the object of inference (e.g., differences between means, correlations between variables).

In this article, we extended to multiple correlation the definition of effective sample size used in Dutilleul's (1993) modified t test for simple correlation. This extension resulted in a modified F test. We showed through simulations that unlike the REML likelihood-ratio χ^2 test, the model-based and model-free variants of the modified F test are generally valid and powerful. We have shown this for various sampling schemes and different numbers of sampling locations and autocovariance structures. We also showed that for the same range of spatial autocorrelation, large samples do not prevent inflated rejection rates for the classical, unmodified F test when the dependent and explanatory processes possess autocovariances of same sign; when of different sign, large samples do not necessarily ensure sufficient rejection rates to it.

As briefly mentioned by Dutilleul (1993, p. 310), Box's epsilon used in the repeated measures ANOVA is related to the effective sample size used in correlation analysis with autocorrelated and heteroscedastic sample data, as the product of Box's epsilon by $(a-1)$, where a denotes the number of repeated measures, can be seen as an effective sample size. However, whereas the product of Box's epsilon by $(a-1)$ cannot be greater than $a-1$ in theory, M in correlation analysis can be greater than N. The former constraint results from a property relating the square of the trace of a symmetric matrix $\mathbf{A}$ and the trace of its square: $\{\mathrm{trace}(\mathbf{A})\}^2/\mathrm{trace}(\mathbf{A}^2) \leqslant \mathrm{rank}(\mathbf{A})$ (Graybill, 1983, Theorems 9.1.22 and 9.1.23, pp. 303–305). Moreover, the approach we followed in our article, with the definition of a theoretical modified F test first and then model-free and model-based variants for applications, resembles that of Box (1954a, b), who defined the epsilon factor theoretically and was followed by Greenhouse and Geisser (1959) and Huynh and Feldt (1976) for its estimation in practice. It can be shown (Dutilleul, 2008) that circularity (Huynh and Feldt, 1970; Rouanet and Lépine, 1970) is generally involved in a weak sufficient condition for valid unmodified t testing in correlation analysis, while it is the most general necessary and sufficient condition for unmodified F testing in the repeated measures ANOVA.

In a pioneering study, Bivand (1980) suggested that the problem of autocorrelation in the correlation analysis with spatial data was worse for irregular grids than for regular grids. Our simulation results confirm this suggestion with estimated effective sample sizes smaller for the completely random sampling grid than for the other sampling grids (especially the regular one), all other factors being equal in study **I**. It means that to be valid, the modified F tests need a smaller number of df in the denominator, to compensate for the complete randomness of the sampling scheme. The use of distance classes instead of true distances in the computation of experimental variograms does not seem to be a prevailing factor.

In the context of coregionalization analysis, the direct and cross experimental variograms of the dependent and explanatory processes would be jointly modeled with the same basic functions under a linear model of coregionalization (Journel and Huijbregts, 1978, p. 172). Within the limits of this model, structural coefficients of correlation and determination can be evaluated, and modified t and F tests aimed at assessing their significance are presently the object of another study. Such extensions, however, are not straightforward, and special attention then needs to be paid to the interference between random structural components within a bounded sampling domain and the resulting uncertainty of range and sill estimates (Stein, 1999, p. 13; Larocque et al., 2007).

A broad domain of application can be foreseen for the modified F test in the analysis of multiple correlation between one process and several others. Its spatial version was presented here. The development of temporal and spatio-temporal versions as well as a version for repeated measures would require adjustments of moderate importance, as it was the case for the modified t test in simple correlation analysis (Dutilleul and Pinel-Alloul, 1996; Dutilleul et al., 1998; Alpargu and Dutilleul, 2003b). When the first-order stationarity assumption does not hold, means and autocovariances should be analyzed jointly. This is possible with repeated measures data, without having to model the autocovariances when the profile vectors can be used as replicates. Beyond hypothesis testing, the effective sample size defined in our modified F test can be used to build confidence intervals on $R^2_{YX_1\ldots X_q}$ (Moschopoulos and Mudholkar, 1983) and on the adjusted coefficient, $R^2_{adj} = R^2_{YX_1\ldots X_q} - (1 - R^2_{YX_1\ldots X_q})q/(N - q - 1)$.

Acknowledgments

This study was supported by the Natural Sciences and Engineering Research Council of Canada and Le Fonds québécois de la recherche sur la nature et les technologies. For the simulation studies, we have benefited from computational facilities made available to us, thanks to a Canada Foundation for Innovation grant. We are grateful to two reviewers and the Coordinating Editor for constructive comments on the manuscript.

References

Alpargu, G., 2001. Estimation and testing in quantitative linear models with autocorrelated errors. Ph.D. Thesis, Department of Mathematics and Statistics, McGill University.

Alpargu, G., Dutilleul, P., 2001. Efficiency analysis of ten estimation procedures for quantitative linear models with autocorrelated errors. J. Statist. Comput. Simulation 69, 257–275.

Alpargu, G., Dutilleul, P., 2003a. Efficiency and validity analyses of two-stage estimation procedures and derived testing procedures in quantitative linear models with AR(1) errors. Comm. Statist. Simulation Comput. 32, 799–833.

Alpargu, G., Dutilleul, P., 2003b. To be or not to be valid in testing the significance of the slope in simple quantitative linear models with autocorrelated errors. J. Statist. Comput. Simulation 73, 165–180.

Alpargu, G., Dutilleul, P., 2006. Stepwise regression in mixed quantitative linear models with autocorrelated errors. Comm. Statist. Simulation Comput. 35, 79–104.

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

Box, G.E.P., 1954a. Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification. Ann. Math. Statist. 25, 290–302.

Box, G.E.P., 1954b. Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification. Ann. Math. Statist. 25, 484–498.

Cliff, A.D., Ord, J.K., 1975. The comparison of means when samples consist of spatially autocorrelated observations. Environ. Planning A 7, 725–734.

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

Cook, D.G., Pocock, S.J., 1983. Multiple regression in geographical mortality studies with allowance for spatially correlated errors. Biometrics 39, 361–371.

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York.

Diggle, P.J., Liang, K.-Y., Zeger, S.L., 1996. Analysis of Longitudinal Data. Clarendon Press, Oxford.

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

Dutilleul, P., 2008. A note on sufficient conditions for valid unmodified t testing in correlation analysis with autocorrelated and heteroscedastic sample data. Commun. Statist. Theory Methods 37, in press.

Dutilleul, P., Pinel-Alloul, B., 1996. A doubly multivariate model for statistical analysis of spatio–temporal environmental data. Environmetrics 7, 551–566.

Dutilleul, P., Herman, M., Avella-Shaw, T., 1998. Growth rate effects on correlations among ring width, wood density and mean tracheid length in Norway spruce (*Picea abies* (L.) Karst). Canad. J. Forest Res. 28, 56–68.

Geary, R.C., 1954. The contiguity ratio and statistical mapping. Inc. Statist. 5, 115–145.

Golub, G.H., Pereyra, V., 1973. The differentiation of pseudo-inverses and nonlinear least-squares problems whose variables separate. SIAM J. Numer. Anal. 10, 413–432.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Goulard, M., Voltz, M., 1992. Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. Math. Geol. 24, 269–286.

Graybill, F.A., 1983. Matrices with Applications in Statistics. Wadsworth, Belmont, CA.

Greenhouse, S.W., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrika 24, 95–112.

Griffith, D., 1978. A spatially adjusted ANOVA model. Geogr. Anal. 10, 296–301.

Huynh, H., Feldt, S., 1970. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. J. Amer. Statist. Assoc. 65, 1582–1589.

Huynh, H., Feldt, S., 1976. Estimation of the Box correction for degrees of freedom for sample data in randomised block and split-plot designs. J. Ed. Statist. 1, 69–82.

Jenkins, G.M., Watts, D.G., 1968. Spectral Analysis and its Applications. Holden-Day, San Francisco, CA.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London.

Larocque, G., Dutilleul, P., Pelletier, B., Fyles, J.W., 2007. Characterization and quantification of uncertainty in coregionalization analysis. Math. Geol. 39, 263–288.

Liu, G., Liang, K.-Y., 1997. Sample size calculations for studies with correlated observations. Biometrics 53, 937–947.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.

Morrison, D.F., 1990. Multivariate Statistical Methods. McGraw-Hill, New York.

Moschopoulos, P.G., Mudholkar, G.S., 1983. A likelihood ratio based normal approximation for the non-null distribution of the multiple correlation. Comm. Statist. Simulation Comput. 12, 355–371.

Pelletier, B., 2000. Management practices, soil quality and maize yield in smallholder farming systems of central Malawi. Ph.D. Thesis, Department of Natural Resource Sciences, McGill University.

Pelletier, B., Dutilleul, P., Larocque, G., Fyles, J.W., 2004. Fitting the linear model of coregionalization by generalized least squares. Math. Geol. 36, 323–343.

Rouanet, H., Lépine, D., 1970. Comparison between treatments in a repeated-measures design: ANOVA and multivariate methods. British J. Math. Statist. Psych. 23, 147–163.

Searle, S.R., 1971. Linear Models. Wiley, New York.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

The MathWorks, 2006. MATLAB Version R2006. The MathWorks, Natick.

Upton, G.J.G., Fingleton, B., 1985. Spatial Data Analysis by Example: Point Pattern and Quantitative Data. Wiley, Chichester, UK.

Wishart, J., 1931. The mean and second moment coefficient of the multiple correlation coefficient, in samples from a normal population. Biometrika 22, 353–361.

Zhang, X.F., Van Eijkeren, J.C.H., Heemink, A.W., 1995. On the weighted least-squares method for fitting a semivariogram model. Comput. Geosci. 21, 605–608.