

Extraction of Spatial Features Using Factor Methods Illustrated on Stream Sediment Data¹

W. J. Krzanowski^{2,3} and T. C. Bailey²

Descriptive analysis of multivariate spatial data is considerably aided by preliminary extraction of underlying factors or components, followed by analysis of the derived factor or component scores. Much work has been done in this area, and a recent study proposed two contrasting types of such factors which differ in their spatial features. One type comprises orthogonal factors that exhibit lack of spatial cross-correlation across the study region, while the other focusses on extracting maximum spatial structure but results in cross-correlated factors. Many applications would benefit from a combination of these types, so the present paper shows how they can be synthesized in order to produce factors that are approximately orthogonal and also maximise spatial structure. An alternating least squares process is described for doing the computations and its efficacy is evaluated by Monte Carlo simulation. The proposed method is illustrated on a study of trace elements across Vancouver Island, where it is shown that the extracted factors are orthogonal, have approximately zero spatial cross-correlations, and when their scores are plotted on maps of the study region then there is a definite indication of spatial structure. The main features of the two separate previous types of factors have thus been combined successfully into a single set.

KEY WORDS: alternating least squares, common principal components, multivariate spatial analysis, spatial principal components.

INTRODUCTION

One common type of multivariate spatial data consists of measurements taken on p different variables at n locations in some spatial region or domain. For example, a subset of the National Geochemical Reconnaissance performed by the Geological Survey of Canada comprises measurements of the 5 elements zinc (Zn), copper (Cu), nickel (Ni), cobalt (Co) and manganese (Mn) taken at 916 stream locations in North Western Vancouver Island. Such data typically exhibit several different forms of covariance: the usual covariances between variables at each site, but also spatial autocovariances within variables and cross-covariances

¹Received 21 July 2004; accepted 28 March 2006; Published online: 21 February 2007.

²School of Engineering, Computer Science and Mathematics, University of Exeter, Exeter, UK.

³Department of Mathematical Sciences, University of Exeter, North Park Road, Exeter EX4 4QE, UK; e-mail: W.J.Krzanowski@exeter.ac.uk

between variables induced by proximity of neighbouring sites. The former may be termed *pointwise* cross-covariance, the latter *lagged* autocovariance and cross-covariance. Any inferences drawn about the data must take proper account of all these forms of covariance (or, equivalently, correlation).

Typical objectives of analysis are the identification and interpretation of any trends that exist across the region of interest, and perhaps the prediction of responses at sites not represented in the sample. Popular approaches involve simplification of the multivariate structure by preliminary extraction of underlying factors or components, followed by analysis of the derived factor or component scores. The simplest, and most readily available, approach is a principal component analysis of either raw or standardised data, but while the resultant principal component scores are pointwise uncorrelated nothing can be said in general about their lagged cross-correlations. Indeed, Goovaerts (1993) shows that even when principal components are computed with coregionalized variables, spatial orthogonality can only be expected in some special circumstances (such as those of intrinsic correlation or independent nugget components and uncorrelated subsets of variables). In general the cross-correlation between principal components may be nonnegligible, especially if the correlation structure of the raw variables changes according to the spatial scale considered, so more directly customized factors or components are needed. Bailey and Krzanowski (2000) have surveyed existing alternatives, and have proposed two different types of factor that can be used in such analysis. However, while each type of factor has undoubted benefits, both types have some shortcomings. The present contribution therefore aims to show how the best features of each type can be combined into a single analysis.

Background technical material is first summarised, the proposed scheme is then outlined, its efficacy is evaluated by Monte Carlo simulation, an illustrative example is provided, and a brief discussion concludes the paper.

TECHNICAL BACKGROUND

Suppose that $\{y(\mathbf{s}) = (y_1(\mathbf{s}), \dots, y_p(\mathbf{s}))', \mathbf{s} \in \mathcal{R}\}$, is a p -valued, second-order stationary, spatial process, defined over point locations, $\mathbf{s}$, within some spatial domain or region $\mathcal{R}$ and let $\Sigma = \text{cov}(y(\mathbf{s}))$ be the pointwise dispersion matrix of this process. Without loss of generality in what follows, we can assume that the expectations of the y_i are zero. Spatial factor methods aim to replace the y_i with a (smaller) number of components that exhibit 'interesting' spatial properties. In practice the focus has been more specifically on isolating components whose spatial covariance structure acts predominantly over a particular spatial range or scale.

Published methods date back to Webster (1978), and overviews have been provided by Goovaerts (1997) and Wackernagel (1998). See also Grunsky and

Agterberg (1988, 1991) and Nielsen and others (1997, 2000). A key approach is that by Switzer and Green (1984) and Switzer (1985), who seek p components $u_i(\mathbf{s}) = \mathbf{a}_i' \mathbf{y}(\mathbf{s})$, $i = 1, \dots, p$, such that $\text{cov}(u_i(\mathbf{s}), u_j(\mathbf{s})) = \text{cov}(u_i(\mathbf{s}), u_j(\mathbf{s} + \boldsymbol{\delta})) = 0$, $i \neq j$, and $\text{corr}(u_i(\mathbf{s}), u_i(\mathbf{s} + \boldsymbol{\delta})) \leq \text{corr}(u_j(\mathbf{s}), u_j(\mathbf{s} + \boldsymbol{\delta}))$, $i < j$, where $\boldsymbol{\delta}$ is some suitably chosen predetermined, fixed vector of length δ denoting spatial separation or lag. The underlying mathematics stems from the optimisation problem: find the coefficient vector $\mathbf{a}$ which minimises the autocorrelation $\text{corr}(u(\mathbf{s}), u(\mathbf{s} + \boldsymbol{\delta})) = \frac{\mathbf{a}' \mathbf{C}(\boldsymbol{\delta}) \mathbf{a}}{\mathbf{a}' \boldsymbol{\Sigma} \mathbf{a}}$, where $\mathbf{C}(\boldsymbol{\delta}) = \text{cov}(\mathbf{y}(\mathbf{s}), \mathbf{y}(\mathbf{s} + \boldsymbol{\delta}))$. This is mathematically analogous to the standard canonical variate set-up in classical multivariate analysis (Krzanowski, 2000, section 11.1), and leads to the equation $(\mathbf{C}(\boldsymbol{\delta}) - \lambda \boldsymbol{\Sigma}) \mathbf{a} = \mathbf{0}$. This equation has p distinct solutions $\mathbf{a}_i$, which are the normalised eigenvectors corresponding to eigenvalues $\lambda_i = \text{corr}(u_i(\mathbf{s}), u_i(\mathbf{s} + \boldsymbol{\delta}))$ of $\boldsymbol{\Sigma}^{-1} \mathbf{C}(\boldsymbol{\delta})$. The eigenvectors are typically arranged in ascending order of the λ_i , in which case the spatial factors have no pointwise correlations, no cross-correlations at the spatial scale defined by the chosen fixed lag $\boldsymbol{\delta}$, and successively minimise spatial auto-correlation at that lag.

Practical extraction of Switzer's components requires estimates of $\boldsymbol{\Sigma}$ and $\mathbf{C}(\boldsymbol{\delta})$ from sample data. $\boldsymbol{\Sigma}$ can be estimated by the conventional sample covariance matrix $\mathbf{S}$ for the p variables, but such an estimate may be biased (see, e.g., Percival, 1993) so a bias reduction technique such as the jackknife may be needed. A simple approach has been suggested by Quenouille (1957) in the time series context, and this can be equally applied to spatial data. Sample covariance matrices $\mathbf{S}_1, \mathbf{S}_2$ are computed from an arbitrary division of the data into two portions in addition to the matrix $\mathbf{S}$, and the bias-corrected estimate of $\boldsymbol{\Sigma}$ is given by $2\mathbf{S} - (\mathbf{S}_1 + \mathbf{S}_2)/2$.

However, direct estimation of $\mathbf{C}(\boldsymbol{\delta})$ will usually require corrections to achieve a symmetric and positive definite estimate, and this can cause problems. These problems can be avoided if we use the matrix of semi-variances, $\boldsymbol{\Gamma}(\boldsymbol{\delta}) = \frac{1}{2} \text{var}(\mathbf{y}(\mathbf{s} + \boldsymbol{\delta}) - \mathbf{y}(\mathbf{s}))$, instead of $\mathbf{C}(\boldsymbol{\delta})$. We note that $\boldsymbol{\Gamma}(\boldsymbol{\delta}) = \boldsymbol{\Sigma} - \mathbf{C}(\boldsymbol{\delta})$ for a stationary process satisfying the condition of reflection symmetry, i.e. $\mathbf{C}(\boldsymbol{\delta}) = \mathbf{C}(-\boldsymbol{\delta})$, so that

$$\begin{aligned} (\mathbf{C}(\boldsymbol{\delta}) - \lambda \boldsymbol{\Sigma}) \mathbf{a} &= \mathbf{0} \\ \Rightarrow (\boldsymbol{\Gamma}(\boldsymbol{\delta}) - (1 - \lambda) \boldsymbol{\Sigma}) \mathbf{a} &= \mathbf{0}. \end{aligned}$$

Resultant components are thus the same, but are now arranged in *descending* order of eigenvalues as these are $(1 - \text{corr}(u_i(\mathbf{s}), u_i(\mathbf{s} + \boldsymbol{\delta})))$. The standard geostatistical estimator can be used for $\boldsymbol{\Gamma}(\boldsymbol{\delta})$ without any of the foregoing problems in estimation of $\mathbf{C}(\boldsymbol{\delta})$. However, we return to this question in the last section below.

One further simple algebraic result is useful on computational grounds. If $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues of $\boldsymbol{\Sigma}$ and $\mathbf{L}$ is the matrix whose columns are the corresponding eigenvectors, then from the spectral decomposition of $\boldsymbol{\Sigma}$ we

can write $\Sigma = \mathbf{F}\mathbf{F}'$, where $\mathbf{F} = \mathbf{L}\sqrt{\Lambda}$. Thus

$$\begin{aligned} (\Gamma(\delta) - (1 - \lambda)\Sigma)\mathbf{a} &= \mathbf{0} \\ \Rightarrow (\mathbf{F}^{-1}\Gamma(\delta)(\mathbf{F}')^{-1} - (1 - \lambda)\mathbf{I})\mathbf{e} &= \mathbf{0}, \end{aligned}$$

where $\mathbf{e} = \mathbf{F}'\mathbf{a}$, so that $\lambda_i = (1 - \mu_i)$ and $\mathbf{a}_i = (\mathbf{F}')^{-1}\mathbf{e}_i$ where $(\mu_i, \mathbf{e}_i)$ are the eigenvalues and eigenvectors of $\mathbf{F}^{-1}\Gamma(\delta)(\mathbf{F}')^{-1}$. The advantage of this formulation is that we are working with a symmetric matrix, and extraction of its eigenvalues and eigenvectors is a much more stable computation than for the non-symmetric $\Sigma^{-1}\Gamma(\delta)$.

Switzer components thus have the two features: zero cross-correlations and ordered autocorrelations, but only at a single spatial separation δ . Bailey and Krzanowski (2000) therefore set out to generalise the technique and to derive a set of components that have the same features but now across a set of spatial separations $\delta_1, \delta_2, \dots, \delta_k$. For a reasonable size of k and a judicious set of separations, this can be used to approximate a whole area of interest. However, maximisation or minimisation of autocorrelations can now no longer be associated directly with lack of cross-correlation between components: maximising or minimising autocorrelations will lead to different components at each lag, while maximising or minimising a single function of autocorrelations will not necessarily produce zero cross-correlations at all lags.

Thus a choice must be made between two objectives, and two distinct sets of components can be produced. The first set arises when we look for components that have no cross-correlations at all lags, which requires the Switzer equations to be satisfied for a single set of components over a range $\delta_1, \delta_2, \dots, \delta_k$ of spatial separations, i.e.

$$\begin{aligned} (\mathbf{C}(\delta_i) - \lambda_i\Sigma)\mathbf{a} &= \mathbf{0} \quad (i = 1, 2, \dots, k) \\ \Rightarrow (\Gamma(\delta_i) - (1 - \lambda_i)\Sigma)\mathbf{a} &= \mathbf{0} \quad (i = 1, 2, \dots, k) \\ \Rightarrow (\mathbf{F}^{-1}\Gamma(\delta_i)(\mathbf{F}')^{-1} - (1 - \lambda_i)\mathbf{I})\mathbf{e} &= \mathbf{0} \quad (i = 1, 2, \dots, k), \end{aligned}$$

where $\mathbf{F}$ and $\mathbf{e}$ are defined as before.

These equations no longer have exact solution when $k > 1$, but approximate solutions can be found via a Common Principal Component Analysis (CPC; Flury, 1988) of $\mathbf{F}^{-1}\Gamma(\delta_i)(\mathbf{F}')^{-1}$ for $i = 1, 2, \dots, k$, using either maximum likelihood (Flury and Constantine, 1985) or least squares (Clarkson, 1988). The resulting components, which have low spatial cross-correlations, are thus given by $u_j(\mathbf{s}) = \mathbf{a}'_j\mathbf{y}(\mathbf{s})$, $j = 1, 2, \dots, p$, where $\mathbf{a}_i = (\mathbf{F}')^{-1}\mathbf{e}_i$ and $\mathbf{e}_i$ are the normalised common eigenvectors. Some examples are given in Bailey and Krzanowski (2000).

However, these components will have unpredictable autocorrelation properties. So, rather than requiring our components to have no spatial cross-correlations, we may prefer their autocorrelation functions to exhibit a variety of decay patterns over lag. Such a set of components can be produced if we consider a parameterised family $\rho^{(k)}(\delta; \theta^{(k)})$ of simple spatial autocorrelation functions, where δ is the lag and $\theta^{(k)}$ is the vector of parameters. For example the exponential family

$$\rho^{(k)}(\delta; \theta^{(k)}) = \begin{cases} 1 & \text{if } \delta = 0 \\ \alpha^{(k)} e^{-\beta^{(k)} \delta} & \text{if } \delta > 0 \end{cases}$$

has $\theta^{(k)} = (\alpha^{(k)}, \beta^{(k)})$, where $\beta^{(k)}$ reflects the range for spatial correlation and $\alpha^{(k)}$ provides for the possibility of a discontinuity, or ‘nugget effect,’ at the origin. We then look for orthogonal components whose spatial autocorrelation structures best approximate the $\rho^{(k)}(\delta; \theta^{(k)})$, $k = 1, \dots, K$. We can use least squares for each component; having found the optimum parameter values for a particular component, we project the data into the subspace orthogonal to all preceding components and repeat the process for the next component.

Thus in the exponential example above, for the first component we consider:

$$\begin{aligned} \min_{\mathbf{a}, \theta} & \left\{ \frac{\sum_i \left(\frac{\mathbf{a}' \Gamma(\delta_i) \mathbf{a}}{\mathbf{a}' \Sigma \mathbf{a}} - [1 - \rho(\delta_i; \theta)] \right)^2}{\sum_i \left(1 - \frac{\mathbf{a}' \Gamma(\delta_i) \mathbf{a}}{\mathbf{a}' \Sigma \mathbf{a}} \right)^2} \right\} \\ &= \min_{\mathbf{a}, \alpha, \beta} \left\{ \frac{\sum_i \left(\frac{\mathbf{a}' \Gamma(\delta_i) \mathbf{a}}{\mathbf{a}' \Sigma \mathbf{a}} - (1 - \alpha e^{-\beta \delta_i}) \right)^2}{\sum_i \left(1 - \frac{\mathbf{a}' \Gamma(\delta_i) \mathbf{a}}{\mathbf{a}' \Sigma \mathbf{a}} \right)^2} \right\} \end{aligned}$$

where the δ_i (>0) are chosen to cover the full distance over which one would expect the spatial autocorrelation of any component to act. (The denominator ensures that the recovered components are ordered in terms of their spatial autocorrelation across the distance range considered.) We can denote the resulting optimising parameters as $\mathbf{a}^{(1)}$, $\alpha^{(1)}$, $\beta^{(1)}$, project into the subspace orthogonal to this component, and repeat the process. Overall we obtain $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(K)}$ and $\alpha^{(1)}, \dots, \alpha^{(K)}$; $\beta^{(1)}, \dots, \beta^{(K)}$ for orthogonal components $(\mathbf{a}^{(k)})' \mathbf{y}(\mathbf{s})$, whose spatial autocorrelation successively best approximates $\alpha^{(k)} e^{-\beta^{(k)} \delta}$, with these functions being in decreasing order. Some examples are given in Bailey and Krzanowski (2000).

While each of these two sets of components are useful for spatial analysis, both exhibit shortcomings. The first set comprises components that have approximately zero cross-correlations at any spatial scale and so may be treated separately in any subsequent analysis, but which are arbitrarily ordered at each lag and do

not necessarily fit any useful autocorrelation model. On the other hand, the second set are systematically ordered in terms of spatial autocorrelation and are orthogonal, but nothing can be said about their lagged cross-correlations. Prediction, in particular, benefits if the components not only have predictive power through strong spatial structure but also have zero cross-correlations at all lags, so that each contributes independently to the overall prediction. An ideal solution would thus come from a synthesis of our previous two types of component, and in the next section we suggest a possible way to achieve this.

SYNTHESIS OF COMPONENTS

We would like to adjust the previous CPC analysis in order to maintain approximately zero cross-correlations everywhere, but to superimpose onto the resultant components an ordering in the range of their spatial autocorrelations through fit to a suitable parametric model. The development which follows is perfectly general, but to fix ideas we again use the exponential model as an exemplar. Let us therefore re-examine the CPC analysis above. Our derived components were given by $u_j(\mathbf{s}) = \mathbf{a}'_j \mathbf{y}(\mathbf{s})$, $j = 1, 2, \dots, p$, where $\mathbf{a}_i = (\mathbf{F}')^{-1} \mathbf{e}_i$ and $\mathbf{e}_i$ solved the equations

$$(\mathbf{F}^{-1} \boldsymbol{\Gamma}(\delta_i) (\mathbf{F}')^{-1} - (1 - \lambda_i) \mathbf{I}) \mathbf{e} = \mathbf{0} \quad (i = 1, 2, \dots, k).$$

For subsequent reference, write $\mathbf{E}$ as the matrix containing the orthonormal $\mathbf{e}_i$ as its columns.

Consider $\mathbf{z} = \mathbf{F}^{-1} \mathbf{y}$ (so that $u_j(\mathbf{s}) = \mathbf{e}'_j \mathbf{z}(\mathbf{s})$, $j = 1, 2, \dots, p$). Then

$$\begin{aligned} \text{cov}(\mathbf{z}(\mathbf{s})) &= \mathbf{F}^{-1} \text{cov}(\mathbf{y}(\mathbf{s})) (\mathbf{F}^{-1})' \\ &= \mathbf{F}^{-1} \boldsymbol{\Sigma} (\mathbf{F}')^{-1} \\ &= \boldsymbol{\Lambda}^{-1/2} \mathbf{L}^{-1} \mathbf{L} \boldsymbol{\Lambda} \mathbf{L}' (\mathbf{L}')^{-1} \boldsymbol{\Lambda}^{-1/2} \\ &= \mathbf{I}. \end{aligned}$$

So $\boldsymbol{\Omega}_i = (\mathbf{F}^{-1})' \mathbf{C}(\delta_i) (\mathbf{F}^{-1})'$ is therefore the cross-correlation matrix of $\mathbf{z}$ at lag δ_i . The CPC analysis can be viewed as producing a set of orthogonal components of $\mathbf{z}$, with coefficients in columns of $\mathbf{E}$, such that all $\mathbf{E}' \boldsymbol{\Omega}_i \mathbf{E}$ are approximately diagonal.

Thus if we now wish to find spatial components that not only have as nearly as possible zero cross-correlations at all lags but also fit as closely as possible an exponential model for their autocorrelations, then we want

- (i) off-diagonal elements of all $\mathbf{E}' \boldsymbol{\Omega}_i \mathbf{E}$ to be as nearly as possible zero, and

- (ii) diagonal elements $(\mathbf{E}'\mathbf{\Omega}_i\mathbf{E})_{jj}$ to be as nearly as possible equal to $\alpha_j e^{-\beta_j \delta_i}$ for all i and j .

Replacing $\mathbf{C}(\cdot)$ by $\mathbf{\Gamma}(\cdot)$ does not affect (i), but changes the function in (ii) to $1 - \alpha_j e^{-\beta_j \delta_i}$.

A least-squares criterion is clearly appropriate for judging the fit in (ii), so least-squares rather than maximum likelihood is the overall method of choice for finding the optimum components. Clarkson's (1988) algorithm finds the matrix $\mathbf{E}$ that reduces the off-diagonal elements of all $\mathbf{E}'\mathbf{\Omega}_i\mathbf{E}$ approximately to zero, i.e. minimises the sum of squares of all *off-diagonal* elements of $\mathbf{E}'\mathbf{\Omega}_i\mathbf{E}$. But from above we see that $\mathbf{E}'\mathbf{\Omega}_i\mathbf{E}$ is the cross-correlation matrix at lag δ_i of the u_j so has diagonal elements equal to 1. So if we write

$$\mathbf{\Delta}_i = \text{diag}(1 - \alpha_1 e^{-\beta_1 \delta_i}, 1 - \alpha_2 e^{-\beta_2 \delta_i}, \dots, 1 - \alpha_p e^{-\beta_p \delta_i}),$$

then our objective is to minimise the sum of squares ϕ of *all* elements of $\mathbf{E}'\mathbf{\Omega}_i\mathbf{E} - \mathbf{\Delta}_i$. But since $\mathbf{E}$ is orthonormal, we can replace $\mathbf{\Delta}_i$ by $(\mathbf{E}'\mathbf{E})\mathbf{\Delta}_i(\mathbf{E}'\mathbf{E})$, and hence to minimise ϕ we can equivalently minimise the sum of squares ϕ' of all elements of $\mathbf{E}'[\mathbf{\Omega}_i - \mathbf{E}\mathbf{\Delta}_i\mathbf{E}']\mathbf{E}$.

An alternating least-squares iterative scheme to achieve this can be set up as follows. Initial estimates $\mathbf{E}^{(0)}$, $\mathbf{\Delta}_i^{(0)}$ are obtained by minimising the sum of squares of off-diagonal elements of $\mathbf{E}'\mathbf{\Omega}_i\mathbf{E}$ using the Clarkson algorithm, and then fitting the exponential model $1 - \alpha_r e^{-\beta_r \delta_i}$ to the r th diagonal element of $\mathbf{E}^{(0)'}\mathbf{\Omega}_i\mathbf{E}^{(0)}$ for $r = 1, \dots, p$ and $i = 1, \dots, k$ by ordinary least squares.

Then in the j th iteration, $\mathbf{E}^{(j)}$ is found as the $\mathbf{E}$ which minimises the sum of squares of off-diagonal elements of $\mathbf{E}'[\mathbf{\Omega}_i - \mathbf{E}^{(j-1)}\mathbf{\Delta}_i^{(j-1)}\mathbf{E}^{(j-1)'}]\mathbf{E}$ while $\mathbf{\Delta}_i^{(j)}$ is found by fitting the exponential model to diagonal elements of $\mathbf{E}^{(j)'}\mathbf{\Omega}_i\mathbf{E}^{(j)}$.

Iterations are continued until two successive values of ϕ' are equal to within some small pre-set tolerance. Let $\mathbf{E}^{(f)}$ and $\mathbf{\Delta}_i^{(f)}$ be the values of $\mathbf{E}$ and $\mathbf{\Delta}_i$ at termination. Then the fitted exponential model parameters are given directly by $\mathbf{\Delta}^{(f)}$, while the coefficients of the j th required component u_j are given by the elements in the j th row of $\mathbf{a} = (\mathbf{F}')^{-1}\mathbf{E}^{(f)}$.

The end product of this iterative scheme is thus a set of coefficient vectors, which can be interpreted in the usual manner of multivariate components and which can be used for constructing scores that can be plotted for elucidation of underlying trends in the data.

EXPERIMENTAL VERIFICATION

To investigate the extent to which the proposed technique is able to recover structure built in to a set of data, we designed some Monte Carlo simulations. The

way we did this was essentially to start at the end and work backwards. Since the technique aims to isolate components whose spatial autocorrelations exhibit an ordering described by a parametric model, and whose cross-correlations are (approximately) zero at all lags, we started by defining components that had this structure. We generated observations for these components, and performed a series of transformations in order to convert them into a typical set of (correlated) spatial multivariate observations. We then applied the proposed technique to these data, and examined how accurately it could recover the known input structure.

Details of the data generation are as follows. Suppose that we wish to produce a data set comprising n observations on p variables, based on k underlying components that have the specified structure. Following the earlier development, we took the exponential family again as the parametric model of the autocorrelations but for simplicity used a one-parameter version by setting $\alpha^{(i)} = 1$ for all i . The first step was thus to decide on values for $\beta^{(i)}$ ($i = 1, \dots, k$), and then to generate n k -element data vectors whose population autocorrelations at lag δ were $(\mathbf{e}^{-\beta^{(1)}\delta}, \dots, \mathbf{e}^{-\beta^{(k)}\delta})$ and whose population cross-correlations were zero at all lags. The library *RandomFields* for the statistical software package *R* (<http://www.r-project.org/>) was used to generate n univariate observations from each of k Gaussian Markov Random Fields having means 0, variances 1 and specified autocorrelation functions $e^{-\beta^{(i)}\delta}$ for $i = 1, \dots, k$. This provided a series of k components, each with the required autocorrelation structure. However, they will not necessarily have approximately zero cross-correlations at all lags. To obtain the latter condition, we adopted a rejection sampling approach: components were cyclically rejected if any of their computed cross-correlations with existing components were too large, and a new component with the same autocorrelation structure was generated. Rejections were continued until the cross-correlations between components at all lags lay within a specified tolerance of zero. To ensure reasonable execution times we set the tolerance level at less than 0.1 in absolute value.

The above process therefore produced an $(n \times k)$ matrix $\mathbf{U}$ of observations on the spatially uncorrelated components for a given set of $\beta^{(i)}$ values. We then generated an arbitrary $k \times k$ orthonormal matrix $\mathbf{E}$ (by randomly generating k -element vectors and orthonormalising them using the Gram-Schmidt process) and transformed $\mathbf{U}$ to $\mathbf{Z} = \mathbf{UE}$. Finally, in order to generate a data matrix $\mathbf{Y}$ which had a non-trivial (i.e. non-identity) pointwise dispersion matrix $\mathbf{\Sigma}$, we chose values for this matrix, decomposed it into the product $\mathbf{FF'}$, and set $\mathbf{Y} = \mathbf{ZF'}$. The transformation that takes $\mathbf{U}$ into $\mathbf{Y}$ is $\mathbf{EF'}$, so the back-transformation that takes $\mathbf{Y}$ into $\mathbf{U}$ is $\mathbf{T} = (\mathbf{F'})^{-1}\mathbf{E'}$, and this matrix is easily computed in any run of the program.

Having generated the data matrix $\mathbf{Y}$, we then applied the proposed technique to it and judged the extent to which it had been successful by comparing the recovered values of $\beta^{(i)}$ with the input ones ($i = 1, \dots, k$), and the recovered coefficients $\mathbf{A}$ with the input ones $\mathbf{T}$. As an additional performance check, cross-correlations

of the recovered components were inspected to make sure they were close to zero, and the autocorrelations of the recovered components at each lag were compared with those of the input components.

Preliminary investigation of the data generating mechanism showed that the execution time rose extremely quickly as n , p and, particularly, k increased. We decided to generate spatial data across a unit square for normalisation, and to provide realistic sample sizes we settled on $n = 32 \times 32 = 1024$. In order to make the experiment tractable, therefore, we considered just three representative situations: $(k = 2, p = 3)$, $(k = 3, p = 3)$ and $(k = 3, p = 4)$. For $\beta(i)$ we chose $(10.0, 3.33)$ when $k = 2$ and $(10.0, 3.33, 2.0)$ when $k = 3$. These values produced autocorrelation functions exhibiting rapid, moderate and slow decay respectively. When $p = 3$ we set

$$\Sigma = \begin{pmatrix} 1.00 & 0.25 & 0.50 \\ 0.25 & 1.00 & -0.25 \\ 0.50 & -0.25 & 1.00 \end{pmatrix},$$

and when $p = 4$ we set

$$\Sigma = \begin{pmatrix} 1.00 & 0.25 & 0.50 & 0.25 \\ 0.25 & 1.00 & 0.75 & 0.50 \\ 0.50 & 0.75 & 1.00 & 0.75 \\ 0.25 & 0.50 & 0.75 & 1.00 \end{pmatrix}.$$

Ten lags, at values $0.1, 0.2, \dots, 0.9, 1.0$ were considered to be sufficient to cover the distance over which the spatial autocorrelations were acting, so these values were chosen for both the rejection sampling and the data analysis.

Results are given in Tables 1 (for $k = 2, p = 3$), 2 (for $k = 3, p = 3$) and 3 (for $k = 3, p = 4$). In each table we give the “target” (i.e. input to data generation) and “recovered” (i.e. output from data analysis) values of (i) $\beta^{(i)}$ for $(i = 1, \dots, k)$, (ii) coefficients, as columns of $\mathbf{T}$ and $\mathbf{A}$ respectively, and (iii) autocorrelations of the components in columns, up to lag 0.5.

Broadly speaking the results show that the proposed technique recovers the input structure very well, although there is some variability across the three situations. First, inspection of cross-correlations between recovered components showed them all to be below 0.1, indeed considerably so with only a very few exceeding 0.05. The technique is therefore highly successful at providing components that are spatially approximately uncorrelated at all lags, which is one of its two principal aims.

The second principal aim is to order the components in terms of the speed of decay of their autocorrelations as fitted to a parametric (here exponential) model.

Table 1. Comparison of Input and Output Feature Values for $k = 2, p = 3$

Features compared	Input		Output	
	cpt 1	cpt 2	cpt 1	cpt 2
$\beta^{(i)}$	10.00	3.33	10.24	5.45
Coefficients	0.633	0.150	0.627	0.163
	0.532	−0.619	0.554	−0.608
	0.243	0.603	0.222	0.608
Autocorrelations				
lag 0.1	0.359	0.632	0.362	0.630
lag 0.2	0.166	0.434	0.169	0.432
lag 0.3	0.040	0.307	0.042	0.306
lag 0.4	−0.065	0.172	−0.064	0.171
lag 0.5	−0.104	0.019	−0.105	0.019

There is some variability in respect of this aim across the three situations. For $k = 2, p = 3$, Table 1 shows that the aim is achieved excellently: the recovered $\beta(i)$ values are both close to the input values (to within the sort of sampling variation that might be expected) and, perhaps more crucially, the recovered autocorrelations at lags 0.1 to 0.5 match those of the input components almost exactly. This latter comparison also holds good for $k = 3, p = 4$ (Table 3), although here the recovered $\beta(i)$ values are somewhat more discrepant from their input values. The poorest recovery is shown for $k = 3, p = 3$ (Table 2), where only the recovered autocorrelations for component 3 show good accord with their input counterparts while the other components, and $\beta(i)$ values, exhibit some disagreements. These findings are mirrored by the comparison of recovered (**A**) and input (**T**) coefficients: excellent for all components in Tables 1 and 3, but only good in Table 2 for component 3. For a graphical illustration, Fig. 1 shows both the

Table 2. Comparison of Input and Output Values for $k = 3, p = 3$

Features compared	Input			Output		
	cpt 1	cpt 2	cpt 3	cpt 1	cpt 2	cpt 3
$\beta^{(i)}$	10.00	3.33	2.00	12.17	9.26	7.12
Coefficients	0.100	−0.232	1.266	0.011	−0.180	1.272
	0.744	0.762	−0.445	0.987	−0.103	−0.462
	−0.510	1.047	−0.558	0.420	1.076	−0.596
Autocorrelations						
lag 0.1	0.267	0.421	0.467	0.333	0.343	0.469
lag 0.2	0.117	0.148	0.256	0.068	0.189	0.257
lag 0.3	0.074	0.057	0.068	−0.039	0.108	0.138
lag 0.4	0.040	0.045	0.038	−0.037	0.070	0.056
lag 0.5	0.005	0.056	0.006	−0.044	0.064	0.061

Table 3. Comparison of Input and Output Values for $k = 3, p = 4$

Features compared	Input			Output		
	cpt 1	cpt 2	cpt 3	cpt 1	cpt 2	cpt 3
$\beta^{(i)}$	10.00	3.33	2.00	16.20	10.57	4.75
Coefficients	−0.349	0.505	0.822	−0.349	0.502	0.769
	−0.918	−0.375	−0.476	−0.930	−0.366	−0.504
	−0.162	0.285	−0.189	−0.120	0.287	−0.213
	0.793	0.658	−0.385	0.890	0.655	−0.383
Autocorrelations						
lag 0.1	0.210	0.486	0.795	0.206	0.486	0.791
lag 0.2	0.038	0.130	0.615	0.043	0.130	0.616
lag 0.3	−0.051	−0.178	0.412	−0.047	−0.178	0.419
lag 0.4	−0.088	−0.270	0.193	−0.091	−0.269	0.202
lag 0.5	−0.066	−0.229	−0.016	−0.071	−0.229	−0.011

input autocorrelation function (i.e. $e^{-\beta^{(i)}\delta}$ plotted against δ in a solid line) and the recovered autocorrelation function (i.e. $e^{-\hat{\beta}^{(i)}\delta}$ plotted against δ in a dotted line) for each component i (rows of figure) for each of the three situations (columns of the figure). The matches between input and recovered functions are good for all first components, but poorer for the later ones.

As a final point, it should be borne in mind that any recovered transformation may only be unique up to a rotation of axes, i.e. a further orthogonal transformation. It is therefore possible that application of Procrustes Analysis (e.g. Krzanowski, 2000, pp. 155–163) may yield a much closer match between input and recovered configurations, but this possibility was not pursued as the existing results seemed to be adequately convincing.

ILLUSTRATIVE EXAMPLE

For illustration we turn to the National Geochemical Reconnaissance mentioned in the introduction. The spatial coverage was quite dense; in general, the reconnaissance programme aims to achieve a sampling density of at least 1 sample per 13 km² and in this particular case the figure was more like 1 sample per 9 km². We consider the subset consisting of measurements on the 5 specified elements (Zn, Cu, Ni, Co, Mn) taken at 916 stream locations throughout North Western Vancouver Island. Preliminary inspection of the data revealed a feature typical of geochemical data measuring concentrations, namely highly skewed distributions of individual elements with a preponderance of generally low values, so we first transformed all variables to logarithms before conducting any analysis. Initial plots suggested reasonable stationarity, so we next inspected variograms of the individual elements and noted that all autocorrelations had stabilised by a lag of

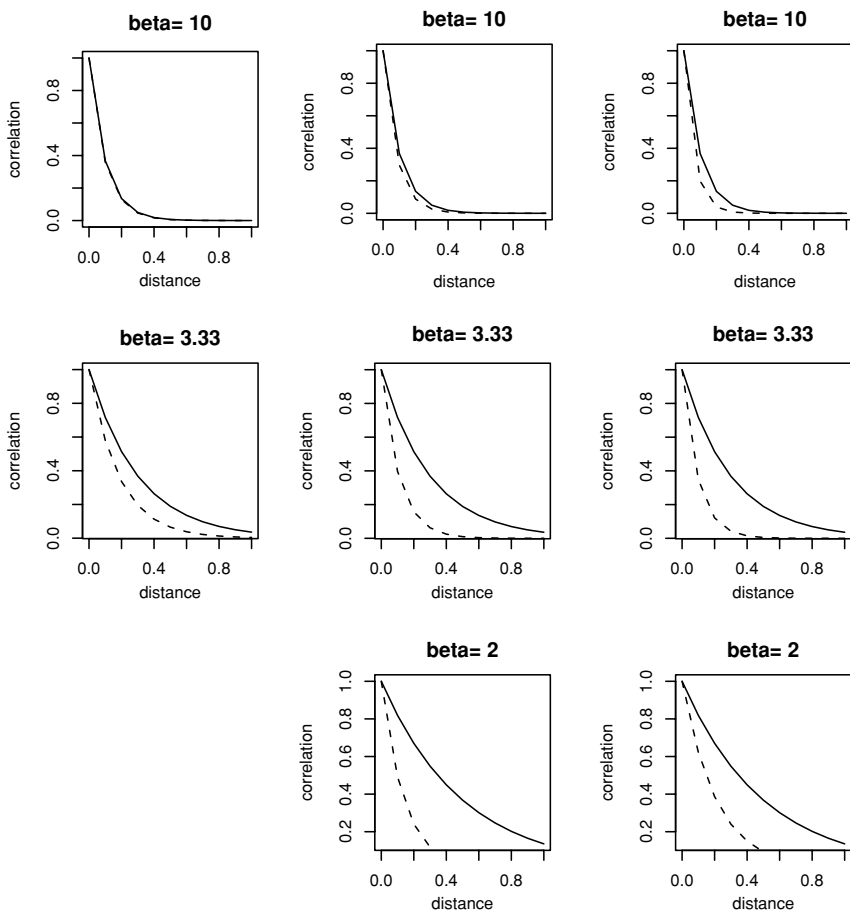


Figure 1. Input (solid line) and recovered (dotted line) autocorrelation functions for each component (rows) in each of the three experimental situations (columns).

around 20 kilometres, so for computation of the components described above we chose the 6 lag (δ_i) values 3, 6, 9, 12, 15 and 18 km.

With these choices we were then able to estimate Σ using the sample (point-wise) covariance matrix, and the matrices $\Gamma(\delta_i)$ at the 6 specified lags by calculating half the sample variance of the differences between all pairs of observations that lay within a small tolerance of the appropriate distance apart. Replacing population quantities in the methodology above by these sample estimates, we were able to find five spatially uncorrelated components. We label these components “dual” (as they are designed to optimise two aims), and now discuss their main

features for this set of data. For convenience we abbreviate the lagged distances 3, 6, 9 etc to lags 1, 2, 3 etc.

Inspection of matrices of cross-correlations between components showed that at lag zero they were all exactly zero, as expected, while the sum of squares of the 60 (10 correlations $\times$ 6 lags) values for lags 1–6 was just 0.310. So the first aim, of producing (approximately) spatially uncorrelated components, had certainly been met. But how effective was the technique in the second aim, i.e. imposing a spatial structure on these components? Figure 2 shows the correlograms of the five components, with fitted exponential functions superimposed on the graphs

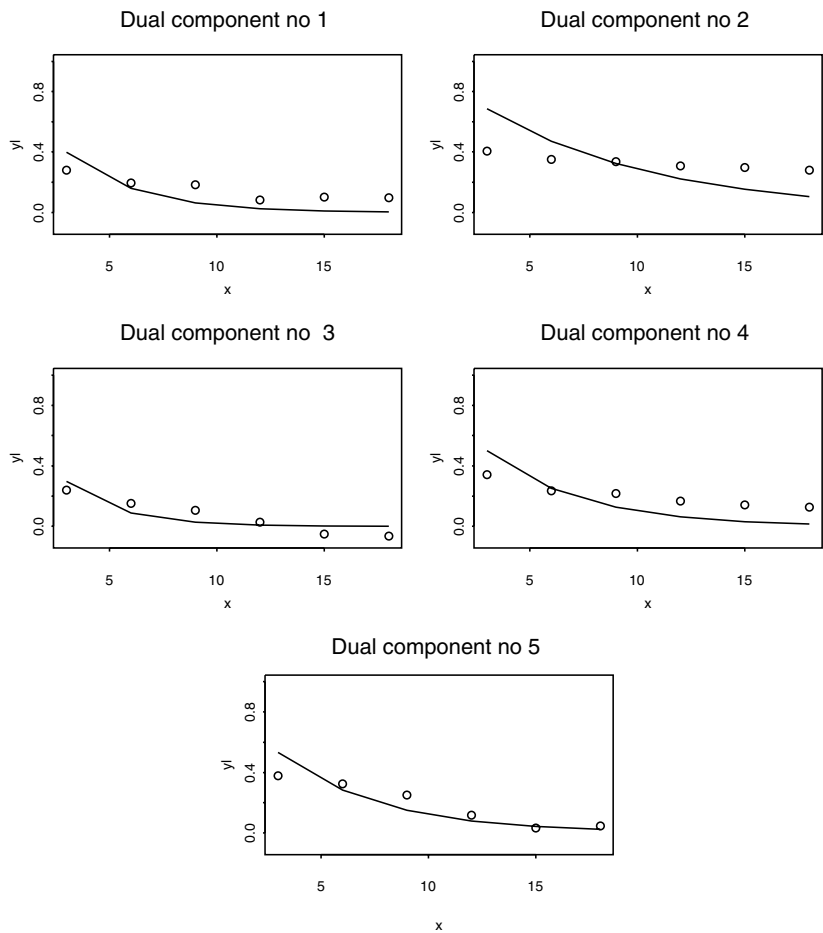


Figure 2. Autocorrelations at lag x for dual components, with fitted exponential model superimposed.

(but note that the order in which these components are produced is arbitrary, as discussed in the last section below). With the exception of component number 2 the dual components generally fit the postulated exponential model reasonably well, so they not only have (approximately) zero spatial cross-correlations but also have discernible spatial structure.

A further question to ask is to what extent the dual components provide a good summary of the data. To answer this question we formed scores from each set of components (by applying the matrix of component loadings to the original data matrix), and then plotted these scores on maps of the study region (Vancouver Island). To obtain interpolated maps we applied localized regression using the SPlus loess function of order 2 and with an appropriately chosen smoothing parameter (see, e.g. Cleveland, Grosse and Shyu, 1992). The five such plots are shown in Fig. 3. Increasingly darker shading represents increasing amount of the relevant combination of trace elements. We see that there is a definite indication of spatial structure, attesting to the utility of the proposed technique.

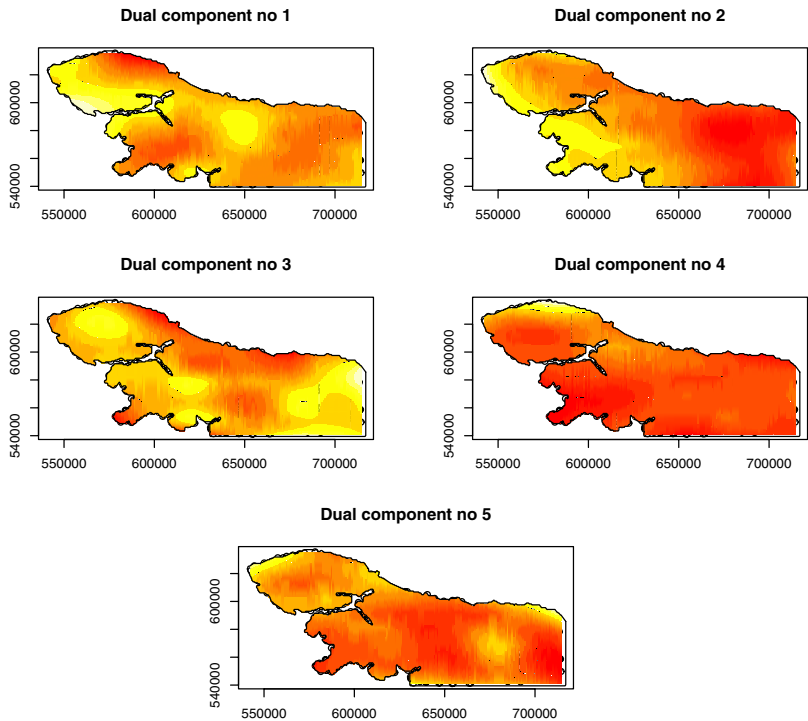


Figure 3. Maps of loess-smoothed scores computed from the dual components.

DISCUSSION

The simulation results and the illustrative example have shown that the proposed technique provides a useful addition to the toolkit for multivariate spatial analysis to set alongside existing factor methods, but there are of course various areas that require further discussion and investigation.

One such centres on the problems raised in estimation of central quantities such as covariance matrices and semi-variance matrices. Our mathematics above has been developed in terms of population quantities, and here we could have kept to $\mathbf{C}(\delta)$ throughout rather than transforming to $\mathbf{\Gamma}(\delta)$. The latter was introduced in order to avoid computational problems arising in estimation of the former but, as has already been mentioned, the resulting mathematics is only valid if the reflective symmetry assumption $\mathbf{C}(\delta) = \mathbf{C}(-\delta)$ is obeyed. This assumption has been discussed by various authors (e.g. Myers, 1982; Ver Hoef and Cressie, 1993), who point out that it is a difficult assumption to check and deal with in practice. In all our illustrations above this is not an issue, as the exponential autocorrelation model implies an isotropic structure. However, our view in general is that, whether or not reflective symmetry is valid, estimation of semi-variance matrices should always be more robust in practice than estimation of covariance matrices so should be preferred. Use of the semi-variance version formulated by Clark, Basinger and Harper (1989) has been shown to circumvent the reflective symmetry assumption in prediction situations (Ver Hoef and Cressie, 1993) so might also be used to advantage in the present case, but this is yet to be investigated.

A second area of interest is the interpretation of components and their role in dimensionality reduction. In principle, interpretation of the components derived above can proceed in the usual manner by identifying each component with those variables that have “large” coefficients and attaching an interpretation in terms of their sizes and signs. However, the fact that the mathematics is related to that of canonical variates indicates that there are potential scaling problems, and these need resolution before confidence can be placed on any interpretations. Similarly, a standard approach to dimensionality reduction is to discard those components that provide only a small contribution to the overall “goodness of fit” measure. The problem in the present case is that there are various potential aspects of goodness of fit of a component (e.g. lack of cross-correlation with other components, closeness of autocorrelation behaviour to a postulated model, variance explained), so the components cannot at present be uniquely ranked in order of “importance.” Further work is needed to derive a suitable overall measure of goodness of fit that can form the basis of such a ranking.

Two final points concern specific choices made in the above analyses that could be generalized with advantage. In our examples with the exponential model we have assumed that spatial correlation is isotropic, so generalization to anisotropic correlation structures would clearly be of interest. Secondly, we

have suggested that the user should decide at the outset of the analysis on the set of lags that are deemed to be satisfactory for spanning the space of investigation. The choice of number of lags and their spacing involves a similar set of considerations as those that arise in variogram estimation more generally (e.g. Wackernagel, 1998). Additional constraints imposed by the common principal component model underlying our techniques militates against use of a large number of lags. A simple rule of thumb would suggest using between about six and ten, but an automated system would benefit considerably from a built-in lag optimisation procedure. We aim to investigate these aspects in future work.

REFERENCES

- Bailey, T. C., and Krzanowski, W. J., 2000, Extensions to spatial factor methods with an illustration in geochemistry: *Math. Geol.*, v. 32, p. 657–682.
- Clark, I., Basinger, K. L., and Harper, W. V., 1989, MUCK—a novel approach to co-kriging, *in* Buxton, B. E., ed., *Proceedings of the Conference on Geostatistical Sensitivity and Uncertainty Methods for Ground-Water Flow and Radionuclide Transport Modelling*: Batelle Press, Columbus, p. 473–493.
- Clarkson, D. B., 1988, Algorithm ASR74, a least squares version of algorithm AS211, the F-G diagonalisation algorithm: *Appl. Stat.*, v. 37, p. 317–321.
- Cleveland, W. S., Grosse, E., and Shyu, W. M., 1992, Local regression models, *in* Chambers J. M. and Hastie T. J., eds., *Statistical Models in S*: Wadsworth and Brooks, Pacific Grove, California p. 309–376.
- Flury, B., 1988, *Common Principal Components and Related Multivariate Models*: John Wiley, New York, 285 p.
- Flury, B., and Constantine, G., 1985, Algorithm AS211, the F-G diagonalisation algorithm: *Appl. Stat.*, v. 34, p. 177–183.
- Goovaerts, P., 1993, Spatial orthogonality of the principal components computed from coregionalised variables: *Math. Geol.*, v. 25, p. 281–302.
- Goovaerts, P., 1997, *Geostatistics for Natural Resources Evaluation*: Oxford University Press, New York, 483 p.
- Grunsky, E. C., and Agterberg, F. P., 1988, Spatial and multivariate analysis of geochemical data from metavolcanic rocks in the Ben Nevis area, Ontario: *Math. Geol.*, v. 20, p. 825–861.
- Grunsky, E. C., and Agterberg, F. P., 1991, SPFAC: a FORTRAN program for spatial factor analysis of multivariate data: *Comput. & Geosci.*, v. 17, p. 133–160.
- Krzanowski, W. J., 2000, *Principles of Multivariate Analysis: a User's Perspective (Revised Edition)*: Oxford University Press, Oxford, 586 p.
- Myers, D. E., 1982, Matrix formulation of co-kriging: *Math. Geol.*, v. 14, p. 249–257.
- Nielsen, A. A., Conradsen, K., Pedersen, J. L., and Steenfelt, A., 1997, Spatial factor analysis of stream sediment geochemistry data from the South Greenland, *in* Pawlowsky-Glahn, V., ed., *Proceedings of the Third Annual Conference of the International Association for Mathematical Geology, IAMG97: Barcelona, 22–27 September 1997* p. 955–960.
- Nielsen, A. A., Conradsen, K., Pedersen, J. L., and Steenfelt, A., 2000, Maximum autocorrelation factorial kriging, *in* Kleingeld, W. J., and Krige, D. G., eds., *Proceedings of the 6th International Geostatistics Congress, Geostats 2000: Cape Town, 10–14 April 2000* p. 538–547.
- Percival, D. B., 1993, Three curious properties of the sample variance and autocovariance for stationary processes with unknown mean: *Am. Stat.*, v. 47, p. 274–276.

- Quenouille, M. H., 1957, *The Analysis of Multiple Time Series*: Griffin, London, 105 p.
- Switzer, P., 1985, Min/max autocorrelation factors for multivariate spatial imagery: *Computer Science and Statistics, Proceedings of the 16th symposium on the Interface*, p. 13–16.
- Switzer, P., and Green, A., 1984, Min/Max autocorrelation factors for multivariate spatial imagery: Technical Report No. 6, Department of Statistics, Stanford University, Stanford, CA.
- Ver Hoef, J. M., and Cressie, N., 1993, Multivariable spatial prediction: *Math. Geol.*, v. 25, p. 219–240.
- Wackernagel, H., 1998, *Multivariate Geostatistics*: Springer-Verlag, Berlin, 291 p.
- Webster, R., 1978, Optimally partitioning soil transects: *J. Soil Sci.*, v. 45, p. 205–218.