Biometrika (2008), **95**, 4, *pp*. 813–829 doi: 10.1093/biomet/asn053

Printed in Great Britain

Testing the covariance structure of multivariate random fields

BY BO LI

Department of Statistics, Purdue University, West Lafayette, Indiana 47907, U.S.A.
boli@purdue.edu

MARC G. GENTON

Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.
genton@stat.tamu.edu

AND MICHAEL SHERMAN

Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.
sherman@stat.tamu.edu

SUMMARY

There is an increasing wealth of multivariate spatial and multivariate spatio-temporal data appearing. For such data, an important part of model building is an assessment of the properties of the underlying covariance function describing variable, spatial and temporal correlations. In this paper, we propose a methodology to evaluate the appropriateness of several types of common assumptions on multivariate covariance functions in the spatio-temporal context. The methodology is based on the asymptotic joint normality of the sample space-time cross-covariance estimators. Specifically, we address the assumptions of symmetry, separability and linear models of coregionalization. We conduct simulation experiments to evaluate the sizes and powers of our tests and illustrate our methodology on a trivariate spatio-temporal dataset of pollutants over California.

Some key words: Covariance; Linear model of coregionalization; Separability; Space and time; Symmetry.

1. INTRODUCTION

Multivariate data obtained at space-time locations have been increasingly employed in various scientific areas. For instance, trivariate atmospheric pollution (CO, NO, NO_2) in California has been analysed by Schmidt & Gelfand (2003) and Majumdar & Gelfand (2007) with the goal of predicting the pollution vector at unmonitored sites. In another context, the selling prices and rents of commercial real estate in Chicago, Dallas and San Diego were investigated by Gelfand et al. (2004) with the purpose of providing implications for real estate finance and investment analysis. As another example, the racial distribution of residents in southern Louisiana was examined by Sain & Cressie (2007) to find whether locations of toxic facilities impact the distribution of white people and minorities. All these analyses take advantage of covariances among variables as well as covariances between locations. However, the multivariate data structure often introduces a cumbersome covariance matrix due to the multiple levels of covariances, which potentially restricts the implementation of traditional spatial data analysis methods such as kriging and cokriging. This situation becomes more serious as the number of variables and/or number of observation locations increases.

Assumptions such as symmetry and separability defined in the multivariate context can greatly reduce the dimension of the covariance matrix, yet their validity should be assessed before being adopted. For example, ignoring the asymmetric structure can lead to missing important characteristics of the underlying process and result in inferior prediction (Ver Hoef & Cressie, 1992; Wackernagel, 2003). For univariate data, much recent research on choice of parsimonious covariance models has appeared; see, for example, Scaccia & Martin (2002, 2005), Guan et al. (2004), Mitchell et al. (2005, 2006), or Fuentes (2006). For univariate space-time observations, Li et al. (2007) described the concepts of space-time symmetry and space-time separability, and also proposed a method to assess those properties. A separable model is usually desired for its simplicity and it has been used widely; for instance, the univariate separable model is a special case of the first-order multivariate autoregressive model. In contrast to the univariate context, much less work has addressed the simplification issue for multivariate space-time covariance modelling. Wackernagel (2003) recommended several approaches based on graphical techniques to examine the symmetry and separable properties for multivariate random fields. While these techniques are certainly useful, they are somewhat subjective. The complex interaction between space-time locations and variables and the lack of solid testing methods point to the need for a more rigorous procedure to assess the appropriateness of common assumptions made on multivariate random fields.

Consider a p-dimensional multivariate random field $Y(x) = \{Y_1(x), \ldots, Y_p(x)\}^{\mathrm{T}}$, $x \in D$, for a region $D \subset \mathbb{R}^q$, $q \geqslant 1$. A common framework for modelling multivariate data is

$$Y(x) = \mu(x) + Z(x) + \epsilon(x), \tag{1}$$

where $\mu(x) = A_x\beta$, with A_x a matrix of covariates depending on coordinates or other location-specific variables, β is an unknown vector of parameters, and $\epsilon(x)$ denotes a white noise vector, that is, $\epsilon(x) \sim F(0, \Omega)$, where F is a p-variate distribution function with diagonal covariance matrix Ω. Assuming that the mean function $\mu(x)$ in equation (1) can be adequately modelled, we focus on the covariance structure of the process $Z(x)$ in this work. We assume a second-order stationary multivariate random field for $Z(x)$, that is, $\mathrm{E}\{Z(x)\} = 0$ and

$$C(k) = \mathrm{cov}\{Z(x), Z(x+k)\} = \mathrm{E}\{Z(x)Z(x+k)^{\mathrm{T}}\}, \tag{2}$$

where $C(k) = [C_{ij}(k)]$ is a $p \times p$ cross-covariance matrix for each lag vector $k \in \mathbb{R}^q$. This formulation allows for several settings of great interest. For example,

(1) $p = 1$, $x = s = (s_1, s_2)$ in $\mathbb{R}^2$ for scalars s_i. This represents a univariate spatial random field with equally or unequally spaced locations;
(2) $p = 1$, $x = (s, t)$ in $\mathbb{R}^{d+1}$ for a scalar t representing a time index. In this case, we have a univariate spatio-temporal random field;
(3) for general p, $x = (s, t)$ in $\mathbb{R}^{d+1}$, we observe a multivariate space-time random field.

Although the matrix $C(k)$ for any specific lag k is in general neither positive nor negative definite, in order that equation (2) be a valid cross-covariance function, the $np \times np$ covariance matrix of the random vector $\{Z(x_1)^{\mathrm{T}}, \ldots, Z(x_n)^{\mathrm{T}}\}^{\mathrm{T}} \in \mathbb{R}^{np}$ must be nonnegative definite for any positive integer n and points $x_1, \ldots, x_n$ in $\mathbb{R}^q$. In recent years, various constructions of valid cross-covariance functions have been proposed. The most straightforward model, in the form of a product of variable covariances and spatial covariance, was proposed by Mardia & Goodall

(1993). This separable model is relatively easy to implement in practice but at the price of imposing often strong restrictions on the covariance structure. A more flexible type of model known as the linear model of coregionalization (Wackernagel, 2003; Gelfand et al., 2004) eases this restriction while still remaining in a concise form, and can be considered as a generalization of the separable model. Ver Hoef & Barry (1998) presented a moving-average approach, also known as kernel convolution, to establish valid stationary cross-covariograms, and later Majumdar & Gelfand (2007) discussed building multivariate spatial cross-covariances through a convolution of p stationary one-dimensional covariance functions.

Our goal is to assess the appropriateness of the above assumptions made on the covariance functions of multivariate random fields. Such assumptions greatly simplify the modelling of multivariate data. For example, consider bivariate observations ($p = 2$) at $n = 5$ irregularly spaced spatial locations, $x_1, \ldots, x_5$. This leads to ten distinct spatial lags. The number of parameters needed to describe the covariances among $z_1(x_1), \ldots, z_1(x_5), z_2(x_1), \ldots, z_2(x_5)$ is 43. Under a symmetry assumption, this number reduces to 33, and further reduces to 13 under an assumption of space-variable separability. Beyond that, under the separability assumption, we can often find an adequate spatial covariance model to fit our data due to the availability of rich classes of spatial covariance models; see Gneiting et al. (2007) for a recent review. For example, if we find an exponential covariance model to be appropriate, then the required number of parameters drops to 4. Such gains in model simplicity become more crucial with n or p large. In general, for p-dimensional observations at n irregularly spaced locations, the symmetry condition reduces the number of parameters from $p\{1 + p + pn(n-1)\}/2$ to $\{p(p+1)(2 - n + n^2)\}/4$, and separability further reduces it to $\{p(p+1) + n(n-1)\}/2$. Adding a temporal index to observations reinforces the need for simplification of multivariate covariance functions. Another benefit from assessing the covariance structure occurs when cokriging is considered. If the covariance functions are separable, cokriging is equivalent to separate kriging for each variable if all variables have been measured at all sample locations.

2. Hypotheses of interest

A specific hypothesis of great interest is symmetry. Symmetry notions arise from the observation that $C_{ij}(k) \neq C_{ij}(-k)$ in general, or equivalently, $C_{ij}(k) \neq C_{ji}(k)$. Such asymmetric behaviour is often observed when the response of one variable affects another variable delayed in time. Wackernagel (2003, p. 147) illustrates this phenomenon concerning the gas input and the CO_2 output of a furnace. The fluctuation of CO_2 delays with respect to the fluctuation of gas input due to the chemical reaction time for gas to produce CO_2. Thus the cross-covariance between these two variables exhibits an asymmetric pattern. This example also demonstrates that in the multivariate context, the maximal cross-covariance is not necessarily located at $k = 0$. This is in stark contrast to the univariate case, where this particular notion of asymmetry does not exist. Another situation where such asymmetry occurs is when the cross-covariance is defined between a random function and its derivative, for instance, the cross-covariance between water head and transmissivity in hydrology, where the latter is the derivative of the former (Dagan, 1985). In such situations, the cross-covariance is an odd function of k. Despite the wide range of asymmetries, there are many cases where the cross-covariance is symmetric or approximately symmetric. Once being recognized, the symmetry property has the attractive features of simplicity and parsimony, and enables us to further assess the separable property of the cross-covariance function. Therefore, we seek to test the hypotheses $H_0 : C_{ij}(k) = C_{ij}(-k)$, or equivalently $H_0 : C_{ij}(k) = C_{ji}(k)$, for all i, j and all lags $k \in \mathbb{R}^q$.

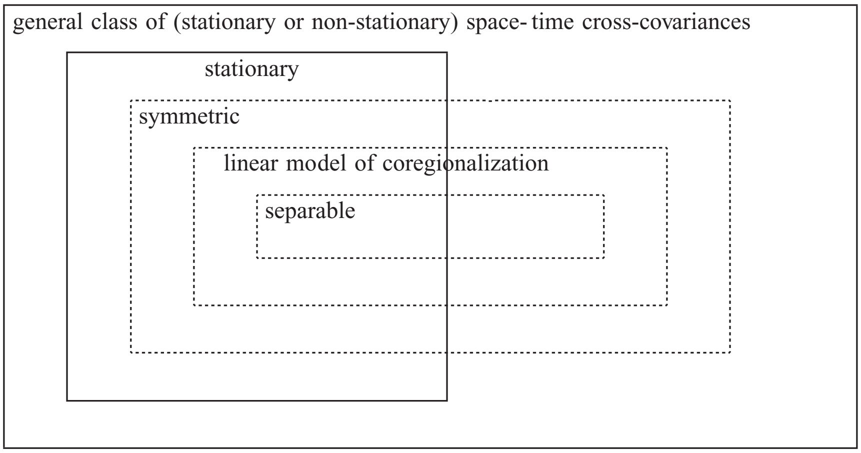

Fig. 1. Schematic representation of various cross-covariance structures of interest.

The cross-covariance is said to be separable if $C(k) = \rho(k)T$ for a scalar correlation function ρ and a positive definite $p \times p$ matrix T, for all lags $k \in \mathbb{R}^q$. When this assumption holds, the covariance functions for different variables are proportional, in that

$$C_{ij}(k_1)/C_{ij}(k_2) = \rho(k_1)/\rho(k_2), \tag{3}$$

or equivalently $C_{ij}(k_1)\rho(k_2) = C_{ij}(k_2)\rho(k_1)$, for all lags $k_1 \in \mathbb{R}^q$, $k_2 \in \mathbb{R}^q$ and all variable pairs (i, j). The right-hand side in equation (3) is independent of i, j. It is seen that symmetry holds if separability holds. Hence, the fundamental hypothesis of separability can be tested once we draw the conclusion of symmetry.

Wackernagel (2003, p. 156–7) presented two approaches to explore whether a separable model is adequate for a given multivariate dataset. One of them calls for examining the constancy of a ‘codispersion coefficient,’ defined as a ratio of variograms. The other is based on the value of cross-variograms between the principal components of variables. A zero value of such cross-variograms indicates the adequacy of a separable model. Both approaches yield useful information, yet there are no clear guidelines on how to make the judgment of constancy or zero proximity.

A more general hypothesis than that of separability is that the multivariate random field follows a linear model of coregionalization, namely that

$$C(k) = \sum_{g=1}^{r} \rho_g(k) T_g \tag{4}$$

for some positive integer $r \leqslant p$ and for all lags $k \in \mathbb{R}^q$, where ρ_g are scalar correlation functions and T_g are $p \times p$ matrices. The above separability is the special case of a linear model of coregionalization when $r = 1$. The linear model of coregionalization has been widely employed in many applications; see, for example, Gelfand et al. (2004). If the data strongly suggest that separability does not hold, it is then of interest to determine if the random field allows for a linear model of coregionalization with small r. Figure 1 summarizes the various aforementioned structures.

3. Asymptotics

In this section, we give asymptotic theory necessary to develop tests of the hypotheses discussed in § 2. In many situations, the observations are taken at a fixed number of locations in a region $S \subset \mathbb{R}^2$ at regular times $T_n = \{1, \ldots, n\}$. In this particular case, to quantify temporal dependence we define the mixing coefficient (Ibragimov & Linnik, 1971, p. 306) as

$$\alpha(u) = \sup_{A,B} \{|P(A \cap B) - P(A)P(B)|, A \in \mathfrak{F}_{-\infty}^{0}, B \in \mathfrak{F}_{u}^{\infty}\},$$

where $\mathfrak{F}_{-\infty}^{0}$ denotes the σ-algebra generated by the past time process until $t = 0$, and $\mathfrak{F}_{u}^{\infty}$ is the σ-algebra generated by the future time process from $t = u$. We assume that the mixing coefficient $\alpha(u)$ satisfies the mixing condition

$$\alpha(u) = O(u^{-\epsilon}), \tag{5}$$

for some $\epsilon > 0$. An autoregressive model of order 1 satisfies equation (5) with Gaussian, double exponential or Cauchy errors. For Gaussian errors, $\alpha(u) \leqslant C|\rho|^u$, where ρ denotes the autoregressive parameter. For other Gaussian processes, from Dhoukan (1994) we have that if $\mathrm{cov}\{Z(t), Z(t+u)\} = O(u^{-\beta})$, then $\alpha(u) = O(u^{1-\beta})$. Thus, equation (5) holds when $\beta > 1$, which in turn implies that the temporal covariances are summable.

Let Λ be a set of user-chosen space-time lags and let c denote the cardinality of Λ. Define $G = \{C_{ij}(h,u) : (h,u) \in \Lambda, i, j = 1, \ldots, p\}$ to be the length cp^2 vector of cross-covariances at spatio-temporal lags $k = (h,u)$ in Λ. Let $\widehat{C}_{ij,n}(h,u)$ denote the estimator of $C_{ij}(h,u)$ based on observations in $S \times T_n$, and let $\widehat{G}_n = \{\widehat{C}_{ij,n}(h,u), (h,u) \in \Lambda, i, j = 1, \ldots, p\}$. Let $S(h) = \{s : s \in S, s + h \in S\}$ and let $|S(h)|$ denote the number of elements in $S(h)$.

We further assume the following moment condition:

$$\sup_n \mathrm{E}\big[||T_n|^{1/2}\{\widehat{C}_{ij,n}(h,u) - C_{ij}(h,u)\}|^{2+\delta}\big] \leqslant C_\delta, \tag{6}$$

for some $\delta > 0$, $C_\delta < \infty$ and any $i, j = 1, \ldots, p$, and define the estimator of $C_{ij}(h,u)$ as

$$\widehat{C}_{ij,n}(h,u) = \frac{1}{|S(h)||T_n|} \sum_{S(h)} \sum_{t=1}^{n-u} Z_i(s,t) Z_j(s+h, t+u).$$

This is the natural empirical estimator of $C_{ij}(h,u)$.

Proposition 1. *Let $\{Z(s,t) \in \mathbb{R}^p, s \in \mathbb{R}^2, t \in \mathbb{Z}\}$ be a zero-mean strictly stationary spatio-temporal multivariate random field observed in $D_n = S \times T_n$, where $S \subset \mathbb{R}^2$ and $T_n = \{1, \ldots, n\}$. Assume*

$$\sum\nolimits_{t \in \mathbb{Z}} |\mathrm{cov}\{Z_i(0,0)Z_j(h_1,u_1), Z_l(s,t)Z_m(s+h_2, t+u_2)\}| < \infty, \tag{7}$$

for all $h_1 \in S(h_1)$, $h_2 \in S(h_2)$, $s \in S$, any finite $u_1, u_2 \in \mathbb{Z}$ and $i, j, l, m = 1, \ldots, p$. Then $\Sigma = \lim_{n\to\infty} |T_n| \mathrm{cov}(\widehat{G}_n, \widehat{G}_n)$ exists. The element of Σ corresponding to the covariance $\mathrm{cov}\{\widehat{C}_{ij,n}(h_p,u_p), \widehat{C}_{lm,n}(h_q,u_q)\}$ is

$$\frac{1}{|S(h_p)||S(h_q)|} \times \sum\nolimits_{S(h_p)} \sum\nolimits_{S(h_q)} \sum\nolimits_{t\in\mathbb{Z}} \mathrm{cov}\{Z_i(s_1,0)Z_j(s_1+h_p,u_p), Z_l(s_2,t)Z_m(s_2+h_q,t+u_q)\}.$$

If we further assume that Σ is positive definite and that conditions (5) and (6) hold, then $|T_n|^{1/2}(\widehat{G}_n - G) \longrightarrow N_{cp^2}(0, \Sigma)$ in distribution as $n \to \infty$.

The proof is given in the Appendix. Other asymptotic regimes and data structures for univariate spatio-temporal random fields were studied by Li et al. (2008). The strict stationarity in space S can be relaxed if only location-specific spatial covariances are concerned. For example, in § 6 we consider the covariances for each specific pair of stations instead of estimating the covariance at arbitrary lags for the irregularly spaced monitoring stations. Condition (7) holds, for example, if $E\{|Z_i(s,t)|^{4+\delta}\} < C_\delta$ for each variable $i = 1, \ldots, p$, for some $\delta > 0$, $C_\delta < \infty$, and $\alpha(u) = O(u^{-\epsilon(4+\delta)/\delta})$ for some $\epsilon > 1$ (Ibragimov & Linnik, 1971, Theorem 18·5·3). No assumption of Gaussianity is required. However, if Z is Gaussian, then equation (7) reduces to $\sum_{t\in\mathbb{Z}} |C_{il}(h_1,t)C_{jm}(h_2,t+u)| < \infty$ for all $h_1 \in S(h_1)$, $h_2 \in S(h_2)$, any finite $u \in \mathbb{Z}$ and $i, j, l, m = 1, \ldots, p$. Similar bounds could be derived for equation (6), although explicit results for general $\delta > 0$ seem difficult.

Our tests involve ratios or products of elements in $\widehat{G}_n$ and for this reason we need the following result for smooth functions of our estimators. Given a real-valued function $f = (f_1, \ldots, f_b)^{\mathrm{T}}$ defined on G such that f is differentiable at G, we have via the multivariate delta theorem (Mardia et al., 1979, p. 52), that

$$|T_n|^{1/2}\{f(\widehat{G}_n) - f(G)\} \longrightarrow N_b(0, B^{\mathrm{T}}\Sigma B)$$

in distribution as $n \to \infty$, where $B_{ij} = \partial f_j/\partial G_i$ $(i = 1, \ldots, cp^2,\ j = 1, \ldots, b)$. Then for a matrix A such that $Af(G) = 0$ under the null hypothesis, it holds that

$$\text{TS} = |T_n|\{Af(\widehat{G}_n)\}^{\mathrm{T}}(AB^{\mathrm{T}}\Sigma BA^{\mathrm{T}})^{-1}\{Af(\widehat{G}_n)\} \longrightarrow \chi_a^2 \tag{8}$$

in distribution as $n \to \infty$, where a is the row rank of the matrix A and TS denotes our test statistic.

4. TESTING PROCEDURES

4·1. *Testing separability and symmetry*

We define the proper function, f, and contrast matrix, A, according to the specific hypotheses of interest. To test for separability, we define f to give pairwise products. Clearly, f can also be defined as pairwise ratios, but nevertheless we adopt the product form for convenience. Under the null hypothesis of separability, we can determine a matrix A such that $Af(G) = 0$. For example, if $p = 2$ and $\Lambda = \{k_1, k_2, k_3, k_4\}$, where $k_i = (h_i, u_i)$ for $i = 1, 2, 3, 4$, we have

$$G = \{C_{12}(k_1), C_{12}(k_2), C_{12}(k_3), C_{12}(k_4), \rho(k_1), \rho(k_2), \rho(k_3), \rho(k_4)\}^{\mathrm{T}}.$$

The presence of $\rho(k)$ rather than $\rho_{11}(k)$ or $\rho_{22}(k)$ is justified by the null hypothesis. We define

$$f(G) = \{C_{12}(k_1)\rho(k_2), C_{12}(k_2)\rho(k_1), C_{12}(k_3)\rho(k_4), C_{12}(k_4)\rho(k_3)\}^{\mathrm{T}},$$

and

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

Under the null hypothesis of separability, $Af(G) = 0$. We note that there is no unique set of elements in G or $f(G)$. For example, we can extend $f(G)$ and A accordingly by including $C_{12}(k_1)\rho(k_3)$ and $C_{12}(k_3)\rho(k_1)$ in $f(G)$. However, we find that the repeated use of the same lags, k_1 and k_3, tends to increase the size of the test. A possible reason for this is the extra dependency between elements in $f(G)$ introduced by using certain testing lags repeatedly. The tests for symmetry properties are analogous to the test for separability, but the symmetry test turns out to be simpler as f is the identity function in this case. The correlation $\rho(k)$ appearing in the test statistic is estimated by the average over $\rho_{ij}(k)$, where $\rho_{ij}(k)$ denotes the correlation between the ith and the jth variables at lag $k = (h, u)$.

4·2. *Testing the order of the linear model of coregionalization*

Consider a multivariate spatio-temporal random field $Z(s, t)$ under a linear model of coregionalization of order r defined by equation (4). Our goal is to determine the smallest r necessary to accurately model $C(k)$, where $k = (h, u)$. If the hypothesis of separability is not rejected, then the order is $r = 1$.

If our separability test indicates nonseparability in $C(k)$, that is, $r > 1$, then we perform a test of the hypothesis $H_0 : r = 2$. If this hypothesis is rejected, then we move to $H_0 : r = 3$ and continue sequentially until there is no significant evidence against the null hypothesis or until r reaches $p - 1$. We develop our testing approach based on the fact that the linear model of coregionalization can be hierarchically constructed through univariate conditional random fields (Royle & Berliner, 1999; Berliner, 2000). This hierarchical approach models dependence of variables through conditional means, thus it allows us to avoid the difficult estimation of ρ_g in equation (4) while still taking the variable dependence into account.

We illustrate our testing approach with a linear model of coregionalization with $p = 3$ variables and

$$Z(s, t) = AW(s, t) = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} W_1(s, t) \\ W_2(s, t) \\ W_3(s, t) \end{pmatrix}. \tag{9}$$

The lower triangular form of A is assumed without loss of generality (Gelfand et al., 2004, p. 272). We denote the component random field and its correlation as $W_g(s, t)$ and $\rho_g(k)$, $g = 1, 2, 3$, and denote a_g as the gth column of A. For simplicity, we assume that the $W_g(s, t)$ have zero mean and unit variance. This assumption has no effect on the covariance study, since the spatial variance term can be absorbed into A. Assuming $\rho_g(k) = \rho(k)$ results in a separable covariance model, $C(k) = \rho(k)T$, where $T = AA^{\mathrm{T}}$. This is the separable case of $r = 1$. We obtain $r = 2$ by assuming that there are only two distinct functions among the $\rho_g(k)$. Without loss of generality, we assume $\rho_2(k) = \rho_3(k)$, hence $C(k) = \sum_{g=1}^{2} \rho_g(k)T_g$, where $T_1 = a_1 a_1^{\mathrm{T}}$ and $T_2 = a_2 a_2^{\mathrm{T}} + a_3 a_3^{\mathrm{T}}$. If all of the $\rho_g(k)$ are distinct functions, we have $r = 3$ with $C(k) = \sum_{g=1}^{3} \rho_g(k)T_g$, where $T_g = a_g a_g^{\mathrm{T}}$.

The joint distribution of $Z(s, t) = \{Z_1(s, t), Z_2(s, t), Z_3(s, t)\}^{\mathrm{T}}$ can be written as

$$\mathrm{pr}\{Z(s, t)\} = \mathrm{pr}\{Z_1(s, t)\}\mathrm{pr}\{Z_2(s, t) \mid Z_1(s, t)\}\mathrm{pr}\{Z_3(s, t) \mid Z_1(s, t), Z_2(s, t)\},$$

where $Z_i(s, t)$ $(i = 1, 2, 3)$ represents the component random field corresponding to each variable. With this probability formula, we can develop $Z(s, t)$ through univariate conditional random fields (Schmidt & Gelfand, 2003; Gelfand et al., 2004). Specifically,

$$\begin{aligned} Z_1(s, t) &= a_{11} W_1(s, t), \\ Z_2(s, t) \mid Z_1(s, t) &= \frac{a_{21}}{a_{11}} Z_1(s, t) + a_{22} W_2(s, t), \\ Z_3(s, t) \mid Z_1(s, t), Z_2(s, t) &= \left(\frac{a_{31}}{a_{11}} - \frac{a_{21}a_{32}}{a_{11}a_{22}} \right) Z_1(s, t) + \frac{a_{32}}{a_{22}} Z_2(s, t) + a_{33} W_3(s, t). \end{aligned}$$

The associated unconditional form is exactly equation (9). If $\rho_2(k) = \rho_3(k)$, that is, $r = 2$, we can express $Z_3(s, t)$ given $Z_1(s, t)$ as

$$Z_3(s, t) \mid Z_1(s, t) = \frac{a_{31}}{a_{11}} Z_1(s, t) + a_{32} W_2(s, t) + a_{33} W_3(s, t).$$

Letting $Z_2'(s,t) = Z_2(s,t) - a_{21}/a_{11} Z_1(s,t)$ and $Z_3'(s,t) = Z_3(s,t) - a_{31}/a_{11} Z_1(s,t)$, we obtain a separable bivariate random field $Z'(s,t) = \{Z_2'(s,t), Z_3'(s,t)\}^{\mathrm{T}}$. Hence the test of $H_0 : r = 2$ is equivalent to the separability test $H_0 : C'(k) = \rho'(k)T'$, where $C'(k)$, $\rho'(k)$ and T' are the functions defined over $Z'(s,t)$. To estimate a_{21}/a_{11} and a_{31}/a_{11}, we first estimate T by $\widehat{T} = \widehat{C}(0)$ and then decompose $\widehat{T}$ into $\widehat{A}\widehat{A}^{\mathrm{T}}$, from which we obtain plug-in estimators for the two ratios.

4·3. *Covariance estimation*

To use the asymptotic result (8) in § 3, estimators of the matrices B, when f is not the identity function in the test, and Σ are required. The matrix B can be estimated empirically by means of $\widehat{G}_n$. Noting the large number of elements in Σ, we apply a subsampling technique to estimate it. Subsampling has proven to be a successful approach in a wide variety of contexts. For the specific task of estimating the covariance matrix of covariance estimators, see Guan et al. (2004) and Li et al. (2007). In particular, consider the implementation of subsampling with overlapping sub-blocks in a fixed space domain and an increasing time domain. In the univariate setting, Carlstein (1986) calculated the optimal sub-block length $l(n)$ for the sample mean computed from an order 1 autoregressive time series based on non-overlapping subseries. Adjusting this slightly for the use of overlapping subseries gives $l(n) = \{2\widehat{\gamma}/(1-\widehat{\gamma}^2)\}^{2/3}(3n/2)^{1/3}$, where $\widehat{\gamma}$ denotes the estimated order 1 autoregressive parameter. However, in the multivariate setting, a unique temporal correlation estimator may not exist, since the correlation function of each variable may be different. To estimate γ, we simply average the first-order temporal correlation estimators across all the variables. Although this procedure is not necessarily optimal, it works reasonably well in all cases we have considered.

5. Monte Carlo simulations

5·1. *Symmetry*

We examine the size and power of our test of symmetry for a variety of sample sizes in the multivariate space-time context. We use 1000 simulation replicates in each setting considered. To assess the performance of our test, we first give a situation which clearly distinguishes between symmetric and asymmetric covariance structures. In order to do this, we generate a bivariate, $p = 2$, random field from a relatively simple model. Assume $Z_1(s,t) = Z_2(s, t+\Delta_t) + \epsilon(s,t)$, where $Z_2(s,t)$ is a strictly stationary random field with $C_{22}(h,u) = \mathrm{cov}\{Z_2(s,t), Z_2(s+h, t+u)\}$, $\epsilon(s,t)$ denotes random noise with mean zero and variance σ^2, and Δ_t denotes a temporal differential. It follows that

$$C_{11}(h,u) = \mathrm{cov}\{Z_1(s,t), Z_1(s+h,t+u)\} = C_{22}(h,u) + \sigma^2,$$
$$C_{12}(h,u) = \mathrm{cov}\{Z_1(s,t), Z_2(s+h,t+u)\} = C_{22}(h,u-\Delta_t),$$
$$C_{21}(h,u) = \mathrm{cov}\{Z_2(s,t), Z_1(s+h,t+u)\} = C_{22}(h,u+\Delta_t).$$

It can be seen that $C_{12}(h,u) \neq C_{21}(h,u)$ when $\Delta_t \neq 0$. We also have $C_{12}(h,u) \neq C_{12}(h,-u)$, $C_{21}(h,u) \neq C_{21}(h,-u)$ for $\Delta_t \neq 0$. Principally, we can perform three tests with $H_0^1 : C_{12}(h,u) = C_{21}(h,u)$, $H_0^2 : C_{12}(h,u) = C_{12}(h,-u)$ and $H_0^3 : C_{21}(h,u) = C_{21}(h,-u)$ as their respective null hypotheses.

For the purpose of illustrating how our test of symmetry performs, we assume $C_{22}(h,u)$ to be space-time separable and $C_{22}(h,u) = C_{22}(\|h\|, |u|)$. We also performed all tests under nonseparable space-time covariance structures with qualitatively similar results as those to follow. In the specific setting above, we observe $C_{12}(h,-u) = C_{21}(h,u)$ and $C_{21}(h,-u) = C_{12}(h,u)$. Hence, in this case, our hypothesis essentially tests $H_0 : C_{12}(\|h\|, u) = C_{21}(\|h\|, u)$.

Table 1. *Empirical sizes (%) of the test for symmetry/separability*

Grid size	γ	Sizes for symmetry $\lvert T_n \rvert$ 200	500	1000	Sizes for separability $\lvert T_n \rvert$ 200	500	1000
	0·4	4	4	3	6	6	6
3×3	0·6	5	5	4	8	8	6
	0·8	11	6	5	9	9	8
	0·4	4	4	4	6	6	4
5×5	0·6	6	6	3	7	6	5
	0·8	10	8	5	11	8	7
	0·4	5	4	3	5	6	5
7×7	0·6	8	4	4	7	7	6
	0·8	10	6	6	11	8	6

Nominal level is 5%. Sizes for symmetry are obtained by setting $\phi = 3$ and using contrasts $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$ defined in § 5·1. Sizes for separability are obtained by setting $\phi_1 = \phi_2 = 3$ and using contrasts $\mathcal{C}_1$ and $\mathcal{C}_2$ defined in § 5·2. The largest standard error for sizes is 1·0%.

We generate Z_2 via a vector autoregressive model of order 1 for $\{Z_2(s_1, t), \ldots, Z_2(s_m, t)\}^{\mathrm{T}}$ with Gaussian random noise and a spatial exponential covariance function on a spatial grid of m observations; see, for example, de Luna & Genton (2005).The variance of the random noise of the vector autoregressive model and σ^2 are both set equal to 1. Let ϕ denote the spatial exponential range parameter, let γ be the temporal correlation parameter in the vector autoregressive model and denote the number of temporal replicates by $|T_n|$. We choose four lags k: $k_1 = (\|h\| = 1, u = 1)$, $k_2 = (\|h\| = 2^{1/2}, u = 1)$, $k_3 = (\|h\| = 2, u = 1)$, $k_4 = (\|h\| = 5^{1/2}, u = 1)$, which form four contrasts $\mathcal{C}_i = C_{12}(k_i) - C_{21}(k_i)$ $(i = 1, 2, 3, 4)$.

Fixing $\phi = 3$ and $\Delta_t = 0$, we estimate the probability of a type I error. Table 1 gives the empirical sizes. Overall the sizes approach the nominal level of 0·05 as grid size and/or temporal length increase. Fixing $\phi = 3$ and $\gamma = 0{\cdot}4$, we have examined the powers under grids of sizes 3×3, 5×5 and 7×7, with $\Delta_t = 1, 2$ and 3, and $|T_n| = 200$, 500 and 1000 using contrasts $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$ defined above. The powers increase as grid size and/or temporal length increase. Powers diminish, however, as Δ_t increases, which is to be expected, as k contains only lags with $u = 1$, which cannot capture the strongest cross-correlation when $\Delta_t > 1$. The larger Δ_t, the further $C(k)$ departs from strong cross-correlation. This provides a guidance on choosing testing lags. Specifically, we should ensure that k covers the possible delays when considering a true physical process.

5·2. *Separability*

We address the ability of our test to detect the difference between separable and nonseparable covariance structures between variables. Towards this goal, we generate data from separable and nonseparable models as follows, and we use 1000 simulation replicates.

We generate a bivariate, $p = 2$, random field using a linear model of coregionalization $Z(s, t) = AW(s, t)$, where A is a 2×2 lower triangular matrix such that $AA^{\mathrm{T}} = T$, $W(s, t) = \{W_1(s, t), W_2(s, t)\}^{\mathrm{T}}$, with $W_1(s, t)$ and $W_2(s, t)$ independent vector autoregressive models of order 1 with Gaussian random noise and a spatial exponential covariance function. Each vector autoregressive random field is separable in terms of its space-time covariance, nevertheless, as noted in § 5·1 this is not essential to our results. The temporal correlation parameter is γ. We

Table 2. *Empirical powers (%) of the test for separability*

Grid size	γ	$\|T_n\| = 200$ (a)	(b)	(c)	$\|T_n\| = 500$ (a)	(b)	(c)	$\|T_n\| = 1000$ (a)	(b)	(c)	ϕ_2
	0·4	27	19	18	58	52	39	89	86	68	2
3×3	0·6	20	20	39	43	42	78	71	72	98	3
	0·8	15	19	58	25	29	94	40	45	100	4
	0·4	65	56	55	96	96	92	100	100	100	2
5×5	0·6	48	49	93	86	88	100	99	100	100	3
	0·8	33	39	99	56	64	100	85	90	100	4
	0·4	86	82	87	100	100	100	100	100	100	2
7×7	0·6	70	72	100	97	98	100	100	100	100	3
	0·8	43	55	100	80	86	100	98	99	100	4

Columns (a) and (b) are powers varying with γ by setting $\phi_1 = 2$, $\phi_2 = 4$. Column (a) uses contrasts $\mathcal{C}_1$ and $\mathcal{C}_2$, while column (b) uses contrasts $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$. Column (c) are powers varying with ϕ_2 by setting $\gamma = 0{\cdot}4$ and $\phi_1 = 1$ and using contrasts $\mathcal{C}_1$ and $\mathcal{C}_2$. Nominal level is 5%, and $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$ are defined in § 5·2. The largest standard error for powers is 1·6.

denote the spatial range parameters of $W_1(s, t)$ and $W_2(s, t)$ by ϕ_1 and ϕ_2. We set

$$T = \begin{pmatrix} 1 & 0{\cdot}5 \\ 0{\cdot}5 & 1 \end{pmatrix}.$$

We choose four lags k: $k_1 = (\|h\| = 1, u = 0)$, $k_2 = (\|h\| = 2^{1/2}, u = 0)$, $k_3 = (\|h\| = 2, u = 0)$, $k_4 = (\|h\| = 5^{1/2}, u = 0)$, which form two contrasts,

$$\mathcal{C}_1 = C_{12}(k_1)\rho(k_3) - C_{12}(k_3)\rho(k_1), \quad \mathcal{C}_2 = C_{12}(k_2)\rho(k_4) - C_{12}(k_4)\rho(k_2),$$

where ρ represents space-time correlation. To study the size of our test, we set the spatial range parameters equal, $\phi_1 = \phi_2 = 3$. These sizes are given in Table 1. We see that the sizes become closer to nominal as the amount of temporal and/or spatial information increases. However, as the temporal correlation increases, the empirical sizes depart more from the nominal level of 0·05.

To study the power, we consider three settings using different contrasts: $\phi_1 = 2$, $\phi_2 = 4$, using contrasts $\mathcal{C}_1$ and $\mathcal{C}_2$; $\phi_1 = 2$, $\phi_2 = 4$, using contrasts $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$, where

$$\mathcal{C}_3 = C_{12}(k'_1)\rho(k'_3) - C_{12}(k'_3)\rho(k'_1), \quad \mathcal{C}_4 = C_{12}(k'_2)\rho(k'_4) - C_{12}(k'_4)\rho(k'_2),$$

with $k'_1 = (\|h\| = 1, u = 1)$, $k'_2 = (\|h\| = 2^{1/2}, u = 1)$, $k'_3 = (\|h\| = 2, u = 1)$, $k'_4 = (\|h\| = 5^{1/2}, u = 1)$; and temporal correlation fixed at $\gamma = 0{\cdot}4$, $\phi_1 = 1$, but ϕ_2 varies, using contrasts $\mathcal{C}_1$ and $\mathcal{C}_2$.

The powers from these three settings are given in Table 2. The patterns of the results are quite clear. As the amount of temporal data increases, the power increases in all cases. However, as the temporal correlation increases, the empirical powers drop, as is to be expected. In cases where spatial correlation is much stronger than temporal correlation, adding space-time lags does not help to increase power compared to using purely spatial lags. However, if temporal correlation becomes relatively stronger, then including the space-time lags becomes more beneficial. The essential idea is to use the most correlated lags in order to obtain the highest power.

5·3. *Linear model of coregionalization*

We generate the random field using a linear model of coregionalization $Z(s, t) = AW(s, t)$ as in equation (9), where A is a lower triangular matrix with $AA^{\mathrm{T}} = T$. The vector $W(s, t) = \{W_1(s, t), W_2(s, t), W_3(s, t)\}^{\mathrm{T}}$, where $W_1(s, t)$, $W_2(s, t)$ and $W_3(s, t)$ are independent vector autoregressive models of order 1 with Gaussian random noise and spatial exponential

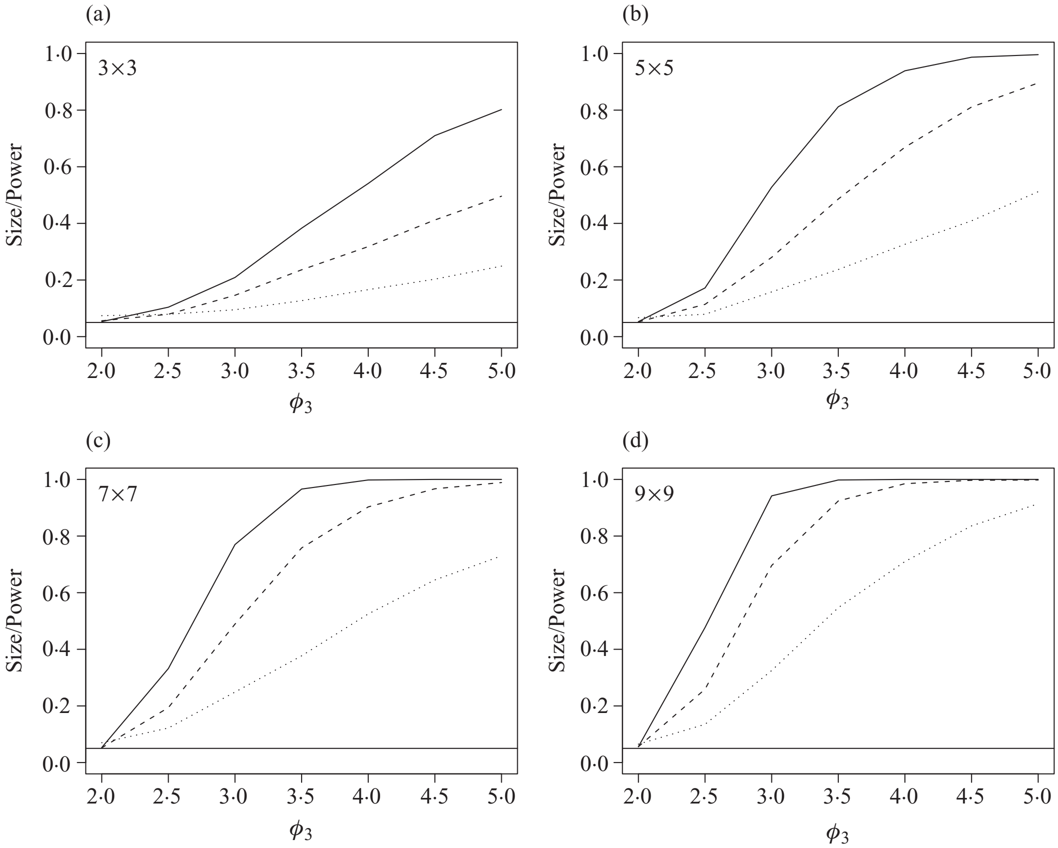

Fig. 2. Empirical sizes and powers of the linear model of coregionalization test for different spatial grid sizes and temporal sizes. For each plot, the spatial grid size is given in the top-left corner; the dotted line, dashed line and the solid line represent $|T_n| = 200, 500, 1000$, respectively. The horizontal line indicates the nominal level of 0·05. The parameters are $\phi_1 = 1$ and $\phi_2 = 2$. The value $\phi_3 = 2{\cdot}0$ corresponds to the size of the test.

covariance functions. Each vector autoregressive random field is separable in terms of its space-time covariance function. Again, the nonseparable structure does not alter our basic conclusions about the testing performance. The temporal correlation parameter is γ. The spatial range parameters of $W_1(s,t)$, $W_2(s,t)$ and $W_3(s,t)$ are denoted by ϕ_1, ϕ_2 and ϕ_3. We set

$$T = \begin{pmatrix} 1 & 0{\cdot}5 & 0{\cdot}5 \\ 0{\cdot}5 & 1 & 0{\cdot}5 \\ 0{\cdot}5 & 0{\cdot}5 & 1 \end{pmatrix},$$

and use the same previous testing lags k_1, k_2, k_3 and k_4, as well as $\mathcal{C}_1$ and $\mathcal{C}_2$, from § 5·2. By fixing $\phi_1 = 1$ and $\phi_2 = 2$ while varying ϕ_3, we obtain empirical sizes and powers for $H_0 : r = 2$ under various experimental settings. We use 1000 simulation replicates. Figure 2 shows that the empirical sizes are reasonably close to the nominal level of 0·05, and that the powers increase as grid size increases and/or temporal length increases. Moreover, power increases as ϕ_3 increases as we expect with the associated increasing departure from the null hypothesis.

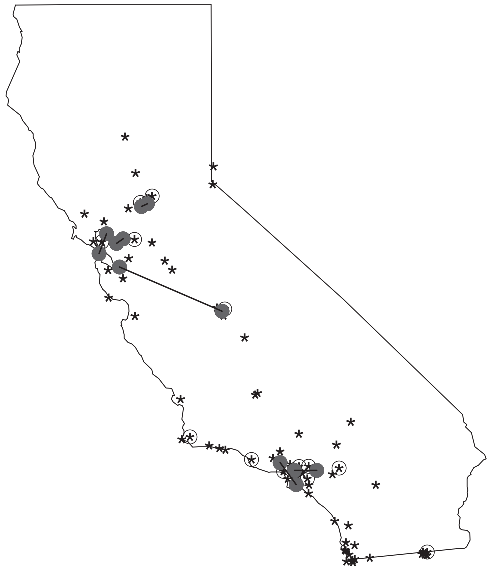

Fig. 3. Eighty-three monitoring stations (stars and solid circles) in California that have measurements of CO, NO and NO_2 in 1999. Twenty-three monitoring stations (circles) have consecutive daily pollutants among the 83 stations. The six pairs composed from 12 stations used in testing are denoted as connected solid circles.

6. TRIVARIATE POLLUTANT DATA

We apply our testing procedures to a trivariate atmospheric pollution dataset, of CO, NO and NO_2 levels, from the California Air Resources Board in 1999. During that year, there were 83 monitoring stations with hourly measurements of CO, NO and NO_2, but not all of them were recorded continuously in time. There are 23 sites with consecutive daily average measurements of the three pollutants. Figure 3 shows the locations of these 23 monitoring stations. The daily average of the three pollutants over 68 locations on 16 July 1999 was analysed originally by Schmidt & Gelfand (2003) and later by Majumdar & Gelfand (2007). In their analyses, two different methodologies were presented to build valid multivariate covariance models for the three correlated pollutants, but both suggested different spatial ranges for all three components. We perform the symmetry, separability and linear model of coregionalization tests on this dataset.

Six pairs composed from 12 stations shown in Fig. 3 were selected for our test. This choice of testing pairs attempts to achieve a wide spatial coverage of stations in addition to seeking pairs with large correlation. This choice of lags may not be optimal but does give reasonable coverage over the state. To stabilize the variance, we take the logarithm of the daily averages as in Schmidt & Gelfand (2003) and Majumdar & Gelfand (2007). Following the detrending technique in Haslett & Raftery (1989), we estimate the seasonal effect by calculating the mean of the logarithm of the daily averages over all 12 stations for each day of the year, and then regress the resulting time series on a set of annual harmonics. Figure 4 displays the seasonal effect of the three pollutants, respectively. Subtraction of the estimated seasonal effect from the logarithm of the daily averages for each station yields approximately stationary residuals. The underlying random fields of these residuals arise as the components of $Z(x)$ in the framework of equation (1) and serve as the processes of our interest.

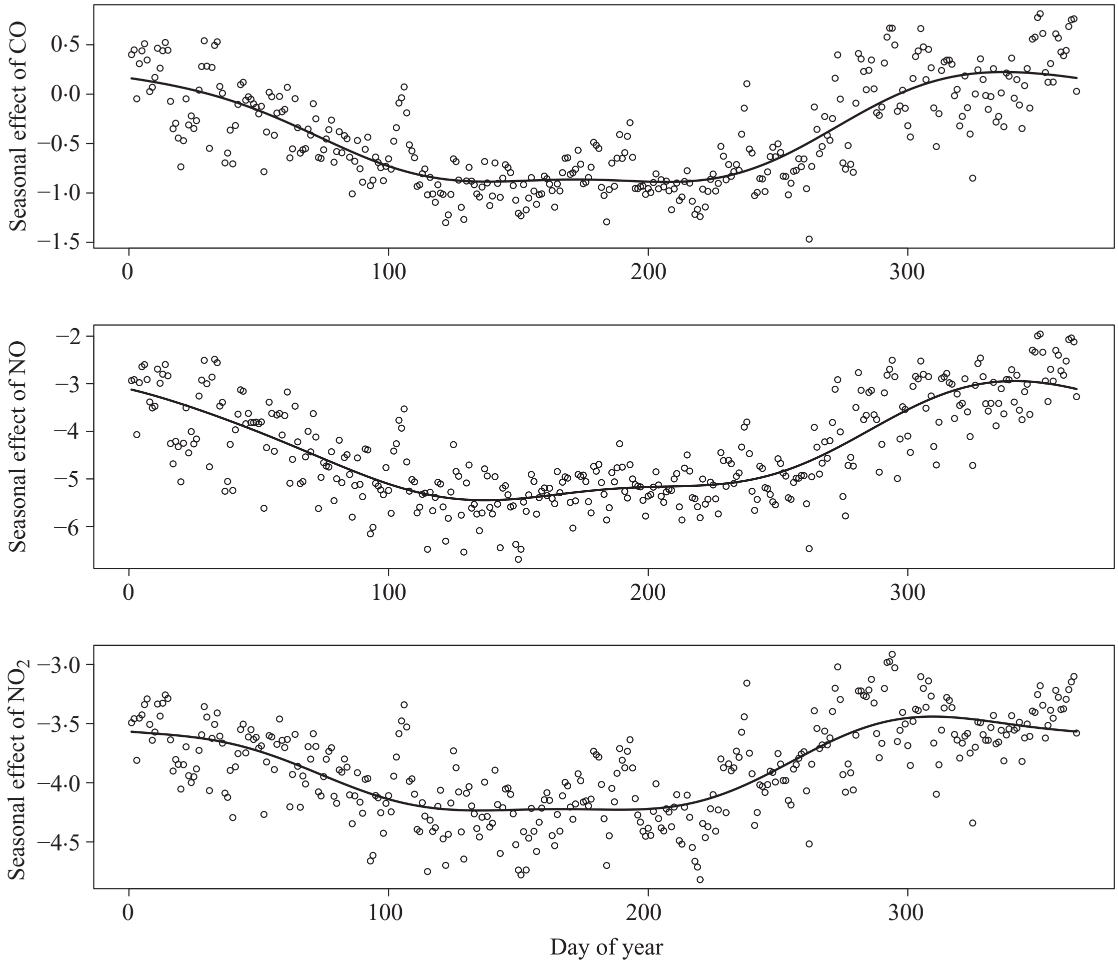

Fig. 4. Seasonal effects for CO, NO and NO_2, respectively. The dots show the mean of the logarithm of daily averages over 12 stations used in the test for each day of the year. The curves are the estimated seasonal effect.

Noting that the largest empirical cross-correlations occur at temporal lag $u = 0$ for all variable combinations, we choose $u = 0$ for all testing lags. Three types of pollutants and six pairs of stations generate 18 testing pairs in the symmetry test, and consequently the degree of freedom for the symmetry test is $a = 18$. We obtain an observed test statistic of $\text{TS} = 103{\cdot}5$ with a corresponding p-value $= 4{\cdot}97 \times 10^{-14}$. To further understand the structure of cross-covariance of the three pollutants, it is interesting to assess the symmetry for each grouping of two pollutants separately. The tests give p-value $= 0{\cdot}00047$ between CO and NO, p-value $= 2{\cdot}19 \times 10^{-8}$ between CO and NO_2 and p-value $= 0{\cdot}00041$ between NO and NO_2. These tests indicate that overall the cross-covariance of the three pollutants displays asymmetric properties, while CO and NO_2 depart most from symmetry compared to the other two pairs. This suggests that it is more appropriate to employ an asymmetric covariance model to respect the asymmetric feature of the cross-covariances displayed in the pollution data. In particular, a linear model of coregionalization does not appear to be suitable for this dataset.

Once the test detects an asymmetric covariance structure, testing the separability and the order of the linear model of coregionalization turns out to be unnecessary. Here we discuss these tests for the purpose of illustration. We form three contrasts using the six pairs of stations in our separability test, and thus obtain a test statistic from nine contrasts with degree of freedom $a = 9$. We obtain TS = 129·4 and the approximate p-value $= 9{\cdot}48 \times 10^{-21}$. Splitting the global separability test into three components, we have p-value $= 0{\cdot}00033$ between CO and NO, p-value $= 5{\cdot}27 \times 10^{-12}$ between CO and NO_2 and p-value $= 2{\cdot}20 \times 10^{-12}$ between NO and NO_2. Then we perform the linear model of coregionalization test of $H_0 : r = 2$ still based on the six station pairs. In principle, we test the separability of the two residual random fields resulting from the three component random fields. For illustration purposes, one residual random field is produced by subtracting CO from NO, and the other one by subtracting CO from NO_2. Both symmetry and separability are tested for the two residual random fields, giving p-value $= 0{\cdot}027$ for the symmetry test and p-value $= 0{\cdot}0014$ for the separability test. Thus as expected, we reject all the separability hypotheses and the linear model of coregionalization. Comparing the symmetry and separability tests in the same scenario, the more significant p-value for separability is reasonable as it suggests more evidence against the more focused hypothesis of separability.

7. Discussion

Modelling multivariate data observed at space-time locations often leads to a complex covariance model involving a large number of parameters. The symmetric or separable covariance structure eases this complexity by regulating the interactions between variables and locations, and thus are commonly assumed in data modelling. However, making assumptions without justification often makes resulting inferences less reliable. In this article, we developed an approach to testing symmetry and separability based on the asymptotic joint normality of cross-covariance estimators and properties of covariances under the hypothesis. Our simulation studies demonstrate this testing approach to be accurate and powerful.

Enlightened by the conditional version of building multivariate random fields, we also proposed an approach to find the most parsimonious yet sufficient model among the linear model of coregionalization family. Again we support this particular test by simulations. The empirical sizes and powers behave reasonably. Although we illustrate this approach based on $p = 3$, similar results can be obtained for a general $p > 3$ with some algebra. Principally, linear models of coregionalization for nonseparable multivariate space-time covariance functions can be constructed by superimposing $r \leqslant p$ separable covariances, and are introduced as a tool for dimension reduction in, for example, Grzebyk & Wackernagel (1994). Identifying the smallest number r of such separable covariances needed is always desirable. Our linear model of coregionalization test can be considered as a reliable approach for dimension reduction and model choice.

Although the tests are developed in terms of covariance estimators, they retain the same properties if we use correlation estimators defined, for instance, in Gelfand et al. (2004) in place of covariance estimators due to the unit-free large-sample χ^2 family of distributions for the test statistics.

In the current context, it does not seem reasonable to always adhere to rejecting hypotheses when the associated p-value is smaller than $0{\cdot}05$, say. Large datasets often lead to formal rejection of null hypotheses and moderate departures from the null may be acceptable to allow for more parsimonious models and the associated easier covariance estimation. In our example, however, we thoroughly rejected symmetry, and thus separability and the linear model of coregionalization, so no simplification was possible.

Once a strong departure from an assumption is detected, it is more appropriate to turn to a model that respects the data features. For example, Sain & Cressie (2007) recently proposed a general and flexible covariance model for multivariate lattice data that allows for asymmetric covariances between different variables at different locations. Current research on the construction of flexible parametric cross-covariance models is underway.

Our testing methodology was developed for stationary random fields, yet it is worth noting that the notions of symmetry and separability exist in a more general context which does not depend on stationarity or any structure for the indexing sets. For example, these notions are also well defined for nonstationary random fields as illustrated in Fig. 1 and discussed by Genton (2007) and Gneiting et al. (2007). Further, any random field with indexing sets that can be partitioned into two or more parts is legitimate for the concepts of symmetry and separability.

Acknowledgement

The authors are grateful to the editor, an associate editor and two anonymous referees for their valuable comments. This research was sponsored by the U.S. National Science Foundation.

Appendix

Proof of Proposition 1

Let $T_n(u) = \{t : t \in T_n, t + u \in T_n\}$. First, we expand $\text{cov}\{\widehat{C}_{ij,n}(h_p, u_p), \widehat{C}_{lm,n}(h_q, u_q)\}$ as a summation of $\text{cov}\{Z_i(s_1, t_1)Z_j(s_1 + h_p, t_1 + u_p), Z_l(s_2, t_2)Z_m(s_2 + h_q, t_2 + u_q)\}$ over all $s_1 \in S(h_p), t_1 \in T_n(u_p), s_2 \in S(h_q), t_2 \in T_n(u_q)$ divided by $Q = |S(h_p)||S(h_q)||T_n|^2$. Then we simplify this average to the summation of $\text{cov}\{Z_i(s_1, 0)Z_j(s_1 + h_p, u_p), Z_l(s_2, t)Z_m(s_2 + h_q, t + u_q)\}$ with a factor of $|T_n(u_p) \cap \{T_n(u_q) - t\}|/Q$ over all $s_1 \in S(h_p), s_2 \in S(h_q)$ and $t \in T_n(u_q) - T_n(u_p)$.

By applying condition (7) and Kronecker's lemma, $|T_n|\text{cov}\{\widehat{C}_{ij,n}(h_p, u_p), \widehat{C}_{lm,n}(h_q, u_q)\}$ converges to

$$\frac{1}{|S(h_p)||S(h_q)|} \sum\nolimits_{S(h_p)} \sum\nolimits_{S(h_q)} \sum\nolimits_{t \in \mathbb{Z}} \text{cov}\{Z_i(s_1, 0)Z_j(s_1 + h_p, u_p), Z_l(s_2, t)Z_m(s_2 + h_q, t + u_q)\}$$

as $n \to \infty$.

Letting $A_n = |T_n|^{1/2}\{\widehat{C}_{ij,n}(h, u) - C_{ij}(h, u)\}$, we prove $A_n \to N(0, \sigma^2)$ in distribution as $n \to \infty$ by applying a blocking technique and telescope arguments. Specifically, we divide the whole field into blocks along time, that is, there is no division in space S.

Let $l(n) = n^\alpha$ and let $m(n) = n^\alpha - n^\eta$, for some $1/(1 + \epsilon) < \eta < \alpha < 1$. Divide the original field D_n into non-overlapping cylinders, $D^b_{l(n)} = S \times T^b_{l(n)}, b = 1, \ldots, k_n$, where $|T^b_{l(n)}| = l(n)$; within each cylinder, further obtain $D^b_{m(n)} = S \times T^b_{m(n)}$, which shares the same center as $D^b_{l(n)}$. Thus distance $(D^b_{m(n)}, D^{b'}_{m(n)}) \geqslant n^\eta$ for $b \neq b'$.Let $a_n = k_n^{-1/2} \sum_{b=1}^{k_n} a^b_n$, $a'_n = k_n^{-1/2} \sum_{b=1}^{k_n} (a^b_n)'$, where $a^b_n = \{m(n)\}^{1/2}(\widehat{C}^b_{ij,n} - C_{ij})$, $\widehat{C}^b_{ij,n}$ is $\widehat{C}_{ij,n}$ estimated over $D^b_{m(n)}$ and $(a^b_n)'$ have the same marginal distributions as a^b_n but are independent. Let $\phi'_n(x)$ and $\phi_n(x)$ be the characteristic functions of a'_n and a_n, respectively. The proof consists of the following three steps. As $n \to \infty$: S1, $A_n - a_n \to 0$ in probability; S2, $\phi'_n(x) - \phi_n(x) \to 0$; S3, $a'_n \to N(0, \sigma^2)$ in distribution.

Proof of S1. Let $D^{m(n)}$ denote the union of all $D^b_{m(n)}$, and $T^{m(n)}$ denote the union of all $T^b_{m(n)}$. Specifically, $|T^{m(n)}| = k_n m(n)$ and $|D^{m(n)}| = |S||T^{m(n)}|$. Observe that

$$a_n = k_n^{-1/2} \sum_{b=1}^{k_n} a^b_n = |T^{m(n)}|^{1/2}(\widehat{C}_{ij,D^{m(n)}} - C_{ij}),$$

and $|T_n|/|T^{m(n)}| \to 1, |D_n|/|D^{m(n)}| \to 1$ as $n \to \infty$. Thus we get $\text{cov}(A_n, a_n) \to \sigma^2$ and $\text{var}(A_n - a_n) \to 0$ as $n \to \infty$.

Proof of S2. We employ a telescope argument here. Let ι denote the imaginary number. We define $U_b = \exp(\iota x a^b_n k_n^{-1/2})$, $X_r = \prod_{b=1}^{r} U_b$ and $Y_r = U_{r+1}$. Extending Theorem 17·2·1 (p. 306) and the telescope argument (p. 338) in Ibragimov & Linnik (1971) to our space-time context, we have

$$\mathrm{cov}(X_r, Y_r) \leqslant 16\alpha(n^{-\eta}) = O(n^{-\epsilon\eta})$$

by equation (5), and

$$|\phi_n'(x) - \phi_n(x)| \leqslant 16k_n O(n^{-\epsilon\eta}) = O(n^{1-\alpha-\epsilon\eta}).$$

The last equality in the above expression follows from $O(k_n)$=$O(\frac{n}{n^\alpha})$=$O(n^{1-\alpha})$. Since $1/(1+\epsilon) < \eta < \alpha < 1$, $1-\alpha-\epsilon\eta < 1-\eta-\epsilon\eta < 0$. Then $|\phi_n'(x) - \phi_n(x)| \to 0$ as $n \to \infty$.

Proof of S3. Observe by equation (6) that $\mathrm{E}\{|(a_n^b)'|^{2+\delta}\} < C_\delta$ for some constant C_δ. Since $(a_n^b)'$ are independent and identically distributed,

$$\mathrm{var}\left\{\sum_{i=1}^{k_n} \left(a_n^b\right)'\right\} = k_n \mathrm{var}\{\left(a_n^b\right)'\}.$$

Defining $\sigma_n^2 = \mathrm{var}\{(a_n^b)'\}$, we have $\sigma_n^2 \to \sigma^2$ from the proof of S1. Thus

$$\lim_{n\to\infty} \sum_{b=1}^{k_n} \frac{\mathrm{E}\{\left|\left(a_n^b\right)'\right|^{2+\delta}\}}{\left[\mathrm{var}\{\sum_{b=1}^{k_n}\left(a_n^b\right)'\}\right]^{(2+\delta)/2}} \leqslant \lim_{n\to\infty} C_\delta \frac{k_n}{\left(k_n\sigma_n^2\right)^{(2+\delta)/2}} = 0.$$

Thus applying Lyapounov's theorem, we have

$$k_n^{-1/2} \sum_{b=1}^{k_n} \left(a_n^b\right)' \longrightarrow N(0, \sigma^2)$$

in distribution as $n \to \infty$. The Cramér–Wold device proves the joint normality. □

References

Berliner, L. M. (2000). Hierarchical Bayesian modeling in the environmental sciences. *Allg. Stat. Arch.* **84**, 141–53.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14**, 1171–9.

Dagan, G. (1985). Stochastic modeling of groundwater flow by unconditional and conditional probabilities: The inverse problem. *Water Resour. Res.* **21**, 65–72.

de Luna, X. & Genton, M. G. (2005). Predictive spatio-temporal models for spatially sparse environmental data. *Statist. Sinica* **15**, 547–68.

Dhoukan, P. (1994). *Mixing: Properties and Examples*. New York: Springer.

Fuentes, M. (2006). Testing for separability of spatial–temporal covariance functions. *J. Statist. Plan. Infer.* **136**, 447–66.

Gelfand, A. E., Schmidt, A. M., Banerjee, S. & Sirmans, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test* **13**, 263–312.

Genton, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics* **18**, 681–95.

Gneiting, T., Genton, M. G. & Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In *Statistics of Spatio-Temporal Systems*, Ed. B. Finkenstaedt, L. Held, and V. Isham, pp. 151–75, Boca Raton, FL: Chapman & Hall, CRC Press.

Grzebyk, M. & Wackernagel, H. (1994). Multivariate analysis and spatial/temporal scales: Real and complex models. In *Proceedings of the XVIIth International Biometrics Conference*, 1993. Hamilton, Ontario.

Guan, Y., Sherman, M. & Calvin, J. A. (2004). A nonparametric test for spatial isotropy using subsampling. *J. Am. Statist. Assoc.* **99**, 810–21.

Haslett, J. & Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind-power resource (with discussion). *Appl. Statist.* **38**, 1–50.

Ibragimov, I. A. & Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen, The Netherlands: Wolters–Noordhoff Publishing.

Li, B., Genton, M. G. & Sherman, M. (2007). A nonparametric assessment of properties of space-time covariance functions. *J. Am. Statist. Assoc.* **102**, 736–44.

Li, B., Genton, M. G. & Sherman, M. (2008). On the asymptotic joint distribution of sample space-time covariance estimators. *Bernoulli* **14**, 228–48.

MAJUMDAR, A. & GELFAND, A. E. (2007). Multivariate spatial modeling using convolved covariance functions. *Math. Geol.* **39**, 225–45.

MARDIA, K. V. & GOODALL, C. R. (1993). Spatial–temporal analysis of multivariate environmental monitoring data. In *Multivariate Environmental Statistics*, 347–86. North-Holland Ser. Statist. Probab. **6**. Amsterdam: North-Holland.

MARDIA, K. V., KENT, J. T. & BIBBY, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

MITCHELL, M. W., GENTON, M. G. & GUMPERTZ, M. L. (2005). Testing for separability of space-time covariances. *Environmetrics* **16**, 819–31.

MITCHELL, M. W., GENTON, M. G. & GUMPERTZ, M. L. (2006). A likelihood ratio test for separability of covariances. *J. Mult. Anal.* **97**, 1025–43.

ROYLE, J. A. & BERLINER, L. M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *J. Agric. Biol. Envir. Statist.* **4**, 29–56.

SCACCIA, L. & MARTIN, R. J. (2002). Testing for simplification in spatial models. In *COMPSTAT 2002*, pp. 581–6. Heidelberg: Physica.

SCACCIA, L. & MARTIN, R. J. (2005). Testing axial symmetry and separability of lattice processes. *J. Statist. Plan. Infer.* **131**, 19–39.

SAIN, S. R. & CRESSIE, N. (2007). A spatial model for multivariate lattice data. *J. Economet.* **140**, 226–59.

SCHMIDT, A. M. & GELFAND, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *J. Geophy. Res.* **108**, STS 10, 1–9.

VER HOEF, J. M. & BARRY, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *J. Statist. Plan. Infer.* **69**, 275–94.

VER HOEF, J. M. & CRESSIE, N. (1992). Multivariate spatial prediction. *Math. Geol.* **25**, 219–40.

WACKERNAGEL, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*, 2nd ed. Berlin: Springer.

[*Received August* 2007. *Revised May* 2008]