Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Multivariate Analysis 97 (2006) 1025 – 1043

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

A likelihood ratio test for separability of covariances

Matthew W. Mitchell[a,*], Marc G. Genton[b], Marcia L. Gumpertz[c]

[a]*Metabolon Inc., RTP, NC, USA*
[b]*Department of Statistics, Texas A&M University, College Station, TX, USA*
[c]*Department of Statistics, North Carolina State University, Raleigh, NC, USA*

Received 23 October 2003
Available online 16 September 2005

Abstract

We propose a formal test of separability of covariance models based on a likelihood ratio statistic. The test is developed in the context of multivariate repeated measures (for example, several variables measured at multiple times on many subjects), but can also apply to a replicated spatio-temporal process and to problems in meteorology, where horizontal and vertical covariances are often assumed to be separable. Separable models are a common way to model spatio-temporal covariances because of the computational benefits resulting from the joint space–time covariance being factored into the product of a covariance function that depends only on space and a covariance function that depends only on time. We show that when the null hypothesis of separability holds, the distribution of the test statistic does not depend on the type of separable model. Thus, it is possible to develop reference distributions of the test statistic under the null hypothesis. These distributions are used to evaluate the power of the test for certain nonseparable models. The test does not require second-order stationarity, isotropy, or specification of a covariance model. We apply the test to a multivariate repeated measures problem.

AMS 1991 subject classification: 62H10; 62H12; 62H15

Keywords: Kronecker product; Multivariate regression; Multivariate repeated measures; Nonstationary; Separable covariance; Spatio-temporal process

* Corresponding author.
E-mail address: mwmitch2@aol.com (M.W. Mitchell).

0047-259X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2005.07.005

1. Introduction

Because of their computational convenience, separable covariances are often applied to multivariate repeated measures or doubly repeated measures [3,6,9] and to spatio-temporal models [10,14,17,20]. For a second-order stationary spatio-temporal process Z, separability is often formulated as the following:

$$C(\boldsymbol{h}, k) = C_1(\boldsymbol{h})C_2(k) \tag{1.1}$$

for all $\boldsymbol{h}, k$, where $C(\boldsymbol{h}, k)$ denotes the covariance between $Z(\boldsymbol{s}+\boldsymbol{h}, t+k)$ and $Z(\boldsymbol{s}, t)$ and C_1 and C_2 are covariance functions of space and time alone, respectively [13].

However, there is no correspondence between stationarity and separability: separable processes need not be stationary, and nonseparable processes may be stationary. A more general definition of separability involves Kronecker products. If $\boldsymbol{U} = (u_{ij})$ is an $s \times s$ matrix and $\boldsymbol{V}$ is a $p \times p$ matrix, the Kronecker product of $\boldsymbol{U}$ and $\boldsymbol{V}$ is the $sp \times sp$ matrix given by

$$\boldsymbol{U} \otimes \boldsymbol{V} = \begin{pmatrix} u_{11}\boldsymbol{V} & \dots & u_{1s}\boldsymbol{V} \\ \vdots & \ddots & \vdots \\ u_{s1}\boldsymbol{V} & \dots & u_{ss}\boldsymbol{V} \end{pmatrix}. \tag{1.2}$$

Let $\boldsymbol{\Sigma}$ be the variance–covariance matrix of the process Z, and let $\boldsymbol{U}$ and $\boldsymbol{V}$ be the covariance matrices for the response variables and time (in the spatio-temporal context, the covariance matrices for space alone and time alone, respectively). The variance-covariance matrix is separable if and only if

$$\boldsymbol{\Sigma} = \boldsymbol{U} \otimes \boldsymbol{V}. \tag{1.3}$$

Note that $\boldsymbol{U}$ and $\boldsymbol{V}$ are not unique since for $a \neq 0$, $a\boldsymbol{U} \otimes (1/a)\boldsymbol{V} = \boldsymbol{U} \otimes \boldsymbol{V}$.

The Kronecker product form of (1.3) provides many computational benefits. Suppose we are modeling a multivariate repeated measures problem with s response variables measured at p times or a spatio-temporal process with s locations and p times. Then the (unstructured) covariance matrix has $sp(sp+1)/2$ parameters, but for a separable process there are $s(s+1)/2 + p(p+1)/2 - 1$ parameters (the -1 is needed in order to identify the model as discussed previously). For prediction and estimation it is necessary to invert the variance–covariance matrix. For example, suppose $p = 100$ and $s = 10$. The nonseparable model requires inversion of a 1000×1000 matrix, while the separable model requires only the inversion of a 10×10 and a 100×100 matrix since the inverse of a Kronecker product is the Kronecker product of the inverses [22, p. 255].

Separability is a convenient property, but there is little written about how to formally test for it. There are a few tests for certain kinds of models. For second-order stationary spatial autoregressive processes, there is an asymptotic chi-square test [23]. In the context of “blur-generated” models, a formal test was not used; but separability was assessed by judging whether the “blurring” parameters that maximized the profile likelihood were close to zero or not [1]. This not only required that the class of models be specified, but also required the fitting of a separable and nonseparable model. Recently, Fuentes [8] proposes a test using properties of the spectral domain.

For multivariate repeated measures, Dutilleul [6] employs the modified likelihood ratio test (LRT) [18, p. 357] for testing separability in the form of (1.3). However, this statistic is used for testing $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma_0}$, where $\boldsymbol{\Sigma_0}$ is a *specified*, *known* matrix (and can be adapted to testing separability if one uses $\boldsymbol{U} \otimes \boldsymbol{V}$ in the role of $\boldsymbol{\Sigma_0}$). The critical values are thus based on the distribution of $r\boldsymbol{S}$ and depend only on r and $m = sp$, not s and p individually (no parameters are estimated under the null and $m(m+1)/2$ parameters are estimated under the alternative). Thus, the critical values do not take into account the variability of $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{V}}$. Furthermore, since the critical values depend only on m, all combinations of s and p that have the same product will have the same critical values. For example, suppose $m = 24$. If $s = 4$ and $p = 6$ or $s = 2$ and $p = 12$, both have the same critical values, but the differences in the number of parameters are 270 and 220, respectively. The test we propose is the ordinary LRT, and the critical values take into account the variability that results from estimating $\boldsymbol{U}$ and $\boldsymbol{V}$. This test does not require the same mean vector across subjects.

In this paper, we propose a LRT of separability for multivariate repeated measures with the same variance–covariance matrix for each subject. The statistic is based on estimating the Kronecker product of two unstructured matrices versus estimating a completely unstructured covariance matrix for multinormal data. We show that the distribution of the test statistic when the null hypothesis is true does not depend on the type of separable model, and hence the distribution can be approximated for any sample size that results in positive definite matrices. This is especially important for small samples where the Type I error is very high if critical values are mistakenly taken from a chi-square distribution. Furthermore, the test does not require the same mean across subjects, isotropy, or second-order stationarity. In addition, the specification of a class of models is not required. However, the test does require the number of replicates, r, to be greater than the product of the dimensions, sp. Hence, the test will be more applicable to multivariate repeated measures than spatio-temporal processes, which often have only one realization.

We will derive the test statistic in Section 2 and show that its distribution does not depend on the type of separable model. We then estimate critical values of the distribution of the test statistic in Section 3 and present simulation studies to illustrate the power of the test in Section 4. We then apply the test to a multivariate repeated measures problem in Section 5.

2. The likelihood ratio test for separability

For testing $H_0 : \boldsymbol{\Sigma} = \boldsymbol{U} \otimes \boldsymbol{V}$ for some positive definite $\boldsymbol{U}$ and $\boldsymbol{V}$ against $H_a : \boldsymbol{\Sigma} \neq \boldsymbol{U} \otimes \boldsymbol{V}$ for any $\boldsymbol{U}$ and $\boldsymbol{V}$, we propose the following LRT: the difference is twice the negative log-likelihood values for the two models. The matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are assumed to be unknown, unpatterned, symmetric matrices with $\boldsymbol{U}$ representing the covariance among response variables (or spatial covariance) and $\boldsymbol{V}$ representing the temporal covariance. Let r be the sample size, s the number of response variables (or number of spatial locations), p the number of times, and $m = sp$.

Let

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E}\boldsymbol{\Sigma}^{1/2}. \tag{2.1}$$

In Eq. (2.1), $\boldsymbol{Y}$ is an $r \times m$ matrix of responses, $\boldsymbol{X}$ is an $r \times q$ matrix of (fixed) covariates (we assume $\boldsymbol{X}$ is full-rank, i.e., $\text{rank}(\boldsymbol{X}) = q$), $\boldsymbol{B} = (\boldsymbol{\beta}_{(1)}, \ldots, \boldsymbol{\beta}_{(m)})$, where each $\boldsymbol{\beta}_{(i)}$ is a $q \times 1$ vector, $\boldsymbol{E}$ is an $r \times m$ matrix with independent rows and each row has a $N_m(\boldsymbol{0}, \boldsymbol{I}_m)$ distribution, and $\boldsymbol{\Sigma}^{1/2}$ is a matrix, such that $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$. The maximum-likelihood estimator of $\boldsymbol{B}$ is

$$\hat{\boldsymbol{B}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{2.2}$$

[15, p. 158]. Note that this solution does not depend on $\boldsymbol{\Sigma}$, so this is the maximum-likelihood estimator for $\boldsymbol{B}$ under either the null or alternative hypothesis.

Under the alternative hypothesis, the maximum-likelihood estimator of $\boldsymbol{\Sigma}$ is $\boldsymbol{S} = (1/r)\boldsymbol{Y}^T(\boldsymbol{I}_r - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Y}$, where $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$, and it is required that $r \geqslant q + m = q + sp$ for $\boldsymbol{S}$ to be invertible. The value of $-2l$, twice the negative log-likelihood function, evaluated at the maximum-likelihood estimators is given by

$$-2l(\hat{\boldsymbol{B}}, \boldsymbol{S}) = r\log|2\pi\boldsymbol{S}| + rsp = rsp\log(2\pi) + r\log|\boldsymbol{S}| + rsp \tag{2.3}$$

[15, p. 159].

We now derive the value of the likelihood function under the null hypothesis that $\boldsymbol{\Sigma} = \boldsymbol{U} \otimes \boldsymbol{V}$. For this derivation, it is necessary to formulate each replicate in matrix form. Let $\boldsymbol{y}_k^T = (y_{11k}, \ldots, y_{1pk}, \ldots, y_{s1k}, \ldots, y_{spk})^T$ be the kth row of $\boldsymbol{Y}$ given in (2.1) and $\boldsymbol{m}_k = E(\boldsymbol{y}_k)$. Let $\boldsymbol{Y}_k$ be the reshaped $s \times p$ matrix

$$\begin{pmatrix} y_{11k} & \cdots & y_{s1k} \\ \vdots & \ddots & \vdots \\ y_{1pk} & \cdots & y_{spk} \end{pmatrix}.$$

Let $\boldsymbol{M}_k = E(\boldsymbol{Y}_k)$, and let $\hat{\boldsymbol{M}}_k$ be the corresponding matrix formed from $\boldsymbol{X}\hat{\boldsymbol{B}}$ (and $\hat{\boldsymbol{m}}_k$ is analogous to $\boldsymbol{m}_k$). Mardia and Goodall [14] and Dutilleul [6] derive the maximum-likelihood estimators for the case when $\boldsymbol{M}_k = \boldsymbol{M}$ for all k. For the more general mean, the derivation of the maximum-likelihood estimators for $\boldsymbol{U}$ and $\boldsymbol{V}$ is nearly identical. The maximum-likelihood estimators for $\boldsymbol{U}$ and $\boldsymbol{V}$ satisfy

$$\hat{\boldsymbol{U}} = \frac{1}{pr}\sum_{k=1}^{r}(\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k)^T\hat{\boldsymbol{V}}^{-1}(\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k),$$
$$\hat{\boldsymbol{V}} = \frac{1}{sr}\sum_{k=1}^{r}(\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k)\hat{\boldsymbol{U}}^{-1}(\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k)^T. \tag{2.4}$$

However, these can be rewritten in terms of the individual elements using the property that $\text{vec}(\boldsymbol{ADC}) = (\boldsymbol{C}^T \otimes \boldsymbol{A})\text{vec}(\boldsymbol{D})$ [15, p. 460]. Thus, we have this alternative formulation of the maximum-likelihood estimators:

$$\text{vec}(\hat{\boldsymbol{U}}) = \left(\frac{1}{pr}\sum_{k=1}^{r}(\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k) \otimes (\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k)\right)^T \text{vec}(\hat{\boldsymbol{V}}^{-1}),$$
$$\text{vec}(\hat{\boldsymbol{V}}) = \left(\frac{1}{sr}\sum_{k=1}^{r}(\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k) \otimes (\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k)\right)\text{vec}(\hat{\boldsymbol{U}}^{-1}). \tag{2.5}$$

In general, the value of twice the negative log-likelihood function $-2l$ [15] for the multivariate normal distribution is given by

$$-2l = rsp\log(2\pi) + r\log|\boldsymbol{\Sigma}| + \sum_{k=1}^{r}(\boldsymbol{y}_k - \boldsymbol{m}_k)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_k - \boldsymbol{m}_k). \tag{2.6}$$

When the null hypothesis is true, (2.6) can be simplified using the following properties of Kronecker products:

$$|\boldsymbol{U}\otimes\boldsymbol{V}| = |\boldsymbol{U}|^p|\boldsymbol{V}|^s \tag{2.7}$$

[22, p. 256], and

$$\begin{aligned}&\sum_{k=1}^{r}(\boldsymbol{y}_k - \boldsymbol{m}_k)^T(\boldsymbol{U}^{-1}\otimes\boldsymbol{V}^{-1})(\boldsymbol{y}_k - \boldsymbol{m}_k)\\ &\quad = \mathrm{trace}\left(\sum_{k=1}^{r}\boldsymbol{U}^{-1}(\boldsymbol{Y}_k - \boldsymbol{M}_k)^T\boldsymbol{V}^{-1}(\boldsymbol{Y}_k - \boldsymbol{M}_k)\right),\end{aligned} \tag{2.8}$$

[3]. Thus, when the null hypothesis is true, (2.6) simplifies to

$$\begin{aligned}-2l =\ & rsp\log(2\pi) + rp\log|\boldsymbol{U}| + rs\log|\boldsymbol{V}|\\ &+\mathrm{trace}\left(\sum_{k=1}^{r}\boldsymbol{U}^{-1}(\boldsymbol{Y}_k - \boldsymbol{M}_k)^T\boldsymbol{V}^{-1}(\boldsymbol{Y}_k - \boldsymbol{M}_k)\right).\end{aligned} \tag{2.9}$$

Eqs. (2.4) imply that

$$\mathrm{trace}\left(\sum_{k=1}^{r}\hat{\boldsymbol{U}}^{-1}(\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k)^T\hat{\boldsymbol{V}}^{-1}(\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k)\right) = rsp. \tag{2.10}$$

Substituting $\hat{\boldsymbol{U}}$, $\hat{\boldsymbol{V}}$, and $\hat{\boldsymbol{M}}_k$ (and hence $\hat{\boldsymbol{B}}$) into Eq. (2.9) yields

$$-2l(\hat{\boldsymbol{U}},\hat{\boldsymbol{V}},\hat{\boldsymbol{B}}) = rsp\log(2\pi) + rp\log|\hat{\boldsymbol{U}}| + rs\log|\hat{\boldsymbol{V}}| + rsp. \tag{2.11}$$

We obtain the following test statistic by taking the difference of (2.11) and (2.3):

$$rp\log|\hat{\boldsymbol{U}}| + rs\log|\hat{\boldsymbol{V}}| - r\log|\boldsymbol{S}|. \tag{2.12}$$

Note that Dutilleul [7] states without derivation that the LRT statistic is $-2\log(\lambda)$, where

$$\lambda = \frac{|\boldsymbol{S}|}{|\boldsymbol{U}|^p|\boldsymbol{V}|^s}. \tag{2.13}$$

This simplifies to

$$2(p\log|\hat{\boldsymbol{U}}| + s\log|\hat{\boldsymbol{V}}| - \log|\boldsymbol{S}|) \tag{2.14}$$

and is similar to (2.12) but should be multiplied by $r/2$.

Theorem 1. *The distribution of the LRT statistic* (2.12) *under the null hypothesis of separability does not depend on* $\boldsymbol{B} = (\boldsymbol{\beta}_{(1)}, \ldots, \boldsymbol{\beta}_{(m)})$, $\boldsymbol{U}$ *or* $\boldsymbol{V}$, *i.e. all separable models yield the same distribution of the LRT statistic for a given* r, s, p, *and* $\boldsymbol{X}$.

The proof is given in the appendix. Since the chi-square distribution of the LRT statistic holds only for large numbers of replicates, this result is especially useful. It allows us to construct a reference distribution of the LRT statistic when the null hypothesis is true since it does not depend on the true values of $\boldsymbol{U}$, $\boldsymbol{V}$, or $\boldsymbol{B}$. Note that although the maximum-likelihood estimates for $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{S}$, and $\boldsymbol{B}$ were derived from the multinormal likelihood function, Theorem 1 applies whether normality is assumed or not. Hence, it is also possible to develop empirical distributions for (2.12) for non-Gaussian response variables by simulating i.i.d. random variables with mean zero and variance one from that distribution. However, the test statistic no longer has the properties that result from maximum-likelihood estimation.

Note that for the test of $\mathrm{H}_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma_0}$, where $\boldsymbol{\Sigma_0}$ is a specified matrix, the ordinary LRT is biased [18, p. 357], which was the reason Dutilleul [6] used the modified LRT for his application. The bias stems from the term trace($r\boldsymbol{S}$) [18, p. 357]. With the LRT for $\mathrm{H}_0 : \boldsymbol{\Sigma} = \boldsymbol{U} \otimes \boldsymbol{V}$, the trace term in the null and alternative likelihood hypotheses for our tests were identical and thus cancelled out. However, in addition to $\boldsymbol{S}$, Eq. (2.12) has $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{V}}$, which makes a power function appear to be intractable. Muirhead observes that the modified LRT for $\mathrm{H}_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma_0}$ is the statistic that results from the Wishart likelihood, rather than the multinormal likelihood. Although the power function is intractable, we derive a "modified" LRT statistic based on the Wishart distribution in order to see if we can improve the power.

Under the conditions in this section, $\boldsymbol{A} = r\boldsymbol{S}$ has a Wishart $\mathrm{W}_{sp}(\boldsymbol{U} \otimes \boldsymbol{V}, r-q)$ distribution when the null hypothesis of separability holds [15, p. 160]. The derivation of the maximum-likelihood estimates is virtually identical to the multinormal case (replace r with $r - q$). Under the alternative hypothesis the maximum-likelihood estimator for $\boldsymbol{\Sigma}$ is $\boldsymbol{S}_W = (1/(r - q))\boldsymbol{A}$, which is unbiased. The maximum-likelihood estimates for $\boldsymbol{U}$ and $\boldsymbol{V}$ from the $\mathrm{W}_{sp}(\boldsymbol{U} \otimes \boldsymbol{V}, r - q)$ distribution are given by

$$\hat{\boldsymbol{U}}_W = \frac{1}{p(r-q)} \sum_{k=1}^{r} (\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k)^T \hat{\boldsymbol{V}}_W^{-1} (\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k),$$

$$\hat{\boldsymbol{V}}_W = \frac{1}{s(r-q)} \sum_{k=1}^{r} (\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k) \hat{\boldsymbol{U}}_W^{-1} (\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k)^T. \tag{2.15}$$

The likelihood ratio statistic based on the Wishart distribution is thus

$$(r-q)p \log|\hat{\boldsymbol{U}}_W| + (r-q)s \log|\hat{\boldsymbol{V}}_W| - (r-q)\log|\boldsymbol{S}_W|. \tag{2.16}$$

This is a linear function of the statistic based on the multinormal distribution. Let L_1 and L_2 be the LRT statistics given by (2.12) and (2.16), respectively. Since $\hat{\boldsymbol{U}}_W = (r/(r-q))\hat{\boldsymbol{U}}$, $\hat{\boldsymbol{V}}_W = (r/(r-q))\hat{\boldsymbol{V}}$, and $\boldsymbol{S}_W = (r/(r-q))\boldsymbol{S}$, we have

$$|(\hat{\boldsymbol{U}}_W \otimes \hat{\boldsymbol{V}}_W)\boldsymbol{S}_W^{-1}| = (r/(r-q))^{sp} |(\hat{\boldsymbol{U}} \otimes \hat{\boldsymbol{V}})\boldsymbol{S}^{-1}|. \tag{2.17}$$

This implies that

$$L_2 = (r-q)(sp)\log(r/(r-q)) + ((r-q)/r)L_1. \tag{2.18}$$

Thus, the "modified" LRT statistic provides no advantage over the ordinary LRT for $H_0 : \Sigma = \boldsymbol{U} \otimes \boldsymbol{V}$ since the power functions are the same. Throughout the remainder of the paper, we focus on the ordinary LRT statistic (2.12).

3. The estimated null distributions and Type I errors for the asymptotic chi-square test

We compute estimated distributions of the LRT statistic under the null hypothesis that $\boldsymbol{\Sigma} = \boldsymbol{U} \otimes \boldsymbol{V}$ for several combinations of r, s and p. Since the distribution of the test statistic does not depend on $\boldsymbol{U}$ and $\boldsymbol{V}$, without loss of generality we chose $\boldsymbol{U} = \boldsymbol{I}_s$ and $\boldsymbol{V} = \boldsymbol{I}_p$. Theorem 1 implies that it is possible to develop distributions for many types of mean models; without loss of generality we chose $\boldsymbol{B} = \boldsymbol{0}$ as the true values of the parameters, and regressed assuming $\boldsymbol{B} = (\mu_1, \ldots, \mu_m)$ ($\boldsymbol{X} = \boldsymbol{1}_r$, where $\boldsymbol{1}_r$ is the $r \times 1$ vector of ones). Estimated distributions can be generated using SAS® [21] *proc mixed*, which uses Newton–Raphson optimization. However, we programmed Dutilleul's algorithm [6] (which is the same as the algorithm given in [14]) for estimating $\boldsymbol{U}$ and $\boldsymbol{V}$ in *R*, and the running time was dramatically reduced. For example, for $r = 50$ and $s = p = 6$, it took SAS® approximately 6 min to run one simulation, but it took *R* approximately 1 s! The *R* code is available upon request to the authors. We generated 10,000 values of the LRT statistic for each combination of r, s, and p, and the critical values are shown in Table 1. The standard errors of the critical values were computed from 10,000 bootstrap samples [4, p. 46].

We can see that the critical values from Table 1 are much different than those using the quantiles from the chi-square distribution with $sp(sp+1)/2 - s(s+1)/2 - p(p+1)/2 + 1$ degrees of freedom, which are shown in Table 2. The Type I errors using the asymptotic chi-square critical values are severe, especially when r is not much larger than ps. For example, the Type I error for the asymptotic chi-square test when $r = 50,\ s = 9$, and $p = 4$ is 1.00 when the nominal level of the test is 0.05! Even for $r = 200,\ s = 9$, and $p = 4$ the Type I error is still approximately 0.38 at the same level. However, as the number of replicates r increases, the critical values from the estimated distributions approach those from the chi-square distributions. The maximum standard error of these Type I error estimates is $\sqrt{0.5(0.5)/10{,}000} \approx 0.005$.

It is very difficult to determine a "small" sample distribution of the LRT (2.12) because $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{V}}$ are interrelated and each is correlated with $\boldsymbol{S}$. However, we can approximate the mean of the LRT (2.12) fairly well, and we can use the ratio of this mean to the asymptotic mean to estimate the critical values of the distribution of the LRT statistic (2.12).

Theorem 2. *For the case when $\boldsymbol{B} = (\mu_1, \ldots, \mu_m)$ and $\boldsymbol{X} = \boldsymbol{1}_r$ (and hence $\hat{\boldsymbol{M}}_k = \bar{\boldsymbol{X}}$ for all k), the expected value of the LRT* (2.12) *is approximately*

$$-r\left(sp\log 2 + \sum_{j=1}^{sp} \psi(0.5(r-j)) - sp\log(r)\right) - (s(s+1)/2 + p(p+1)/2 + sp - 1), \tag{3.1}$$

Table 1
Empirical critical values for testing separability at level $\alpha = 0.05$

s	p	r	Crit value[a] (std. err.)
4	2	25	45.62 (0.11)
4	2	50	40.34 (0.08)
4	2	100	39.00 (0.11)
4	2	200	37.39 (0.12)
4	3	15	159.94 (0.31)
4	3	20	123.68 (0.20)
4	3	25	111.30 (0.18)
4	3	50	94.23 (0.15)
4	3	100	87.92 (0.16)
4	3	200	85.47 (0.12)
4	4	20	257.52 (0.36)
4	4	25	214.92 (0.27)
4	4	50	168.10 (0.21)
4	4	100	154.26 (0.19)
4	4	200	148.70 (0.20)
4	5	25	378.04 (0.45)
4	5	50	268.01 (0.31)
4	5	100	240.20 (0.25)
4	5	200	228.01 (0.22)
4	6	50	395.77 (0.45)
4	6	100	345.63 (0.29)
4	6	200	325.92 (0.27)
6	6	50	1017.41 (0.58)
6	6	100	804.12 (0.43)
6	6	200	736.51 (0.37)
9	2	25	253.13 (0.41)
9	2	50	185.60 (0.23)
9	2	100	165.51 (0.21)
9	2	200	158.20 (0.14)
9	3	30	733.18 (0.83)
9	3	50	494.58 (0.39)
9	3	100	421.84 (0.35)
9	3	200	394.02 (0.33)
9	4	50	1007.10 (0.61)
9	4	100	790.68 (0.46)
9	4	200	725.23 (0.42)

[a]The critical values were computed from 10,000 Monte Carlo runs.

where ψ is the digamma function.

The proof is given in the appendix. We can get improved estimates by mulitplying the second term in (3.1) by $(r-1)/r$, i.e.,

Table 2
Approximate Type I error using the asymptotic chi-square distribution at level $\alpha = 0.05$

s	p	r	df	χ^2 Crit. value	Type I error[a]
4	2	25	24	36.42	0.21
4	2	50	24	36.42	0.11
4	2	100	24	36.42	0.08
4	2	200	24	36.42	0.06
4	3	15	63	82.53	0.96
4	3	20	63	82.53	0.73
4	3	25	63	82.53	0.52
4	3	50	63	82.53	0.20
4	3	100	63	82.53	0.11
4	3	200	63	82.53	0.07
4	4	20	117	143.25	0.99
4	4	25	117	143.25	0.91
4	4	50	117	143.25	0.36
4	4	100	117	143.25	0.15
4	4	200	117	143.25	0.09
4	5	25	186	218.82	1.00
4	5	50	186	218.82	0.64
4	5	100	186	218.82	0.23
4	5	200	186	218.82	0.12
4	6	50	270	309.33	0.89
4	6	100	270	309.33	0.37
4	6	200	270	309.33	0.16
6	6	50	625	684.27	1.00
6	6	100	625	684.27	0.89
6	6	200	625	684.27	0.38
9	2	25	124	150.99	0.98
9	2	50	124	150.99	0.51
9	2	100	124	150.99	0.20
9	2	200	124	150.99	0.11
9	3	30	328	371.23	1.00
9	3	50	328	371.23	0.98
9	3	100	328	371.23	0.50
9	3	200	328	371.23	0.20
9	4	50	612	670.66	1.00
9	4	100	612	670.66	0.89
9	4	200	612	670.66	0.38

[a]The maximum standard errors of these are 0.005.

$$-r\left(sp\log 2 + \sum_{j-1}^{sp} \psi(0.5(r-j)) - sp\log(r)\right)$$
$$-(r/(r-1))(s(s+1)/2 + p(p+1)/2 + sp - 1). \quad (3.2)$$

The approximation in Theorem 2 is based on the asymptotic mean of $r \log |\hat{\boldsymbol{U}} \otimes \hat{\boldsymbol{V}}|$ minus the exact mean of $r \log |\boldsymbol{S}|$. This works well because for $\boldsymbol{S}$ to be positive definite and the test to apply, $r > sp$; however, the maximum-likelihood estimator of $\hat{\boldsymbol{U}} \otimes \hat{\boldsymbol{V}}$ requires only that $r > \max(s/p, p/s)$ [6], so $r \log |\hat{\boldsymbol{U}} \otimes \hat{\boldsymbol{V}}|$ is well approximated by its asymptotic properties when $r > sp$. For example, let $s = 9$, $p = 3$, and $r = 30$, which is close to the minimum number of replicates required, $r = 28$; but $r \geqslant 4$ is required for the maximum-likelihood estimator $\hat{\boldsymbol{U}} \otimes \hat{\boldsymbol{V}}$ to be positive definite. Alternately, the sample size for $\hat{\boldsymbol{U}}$ ignoring $\hat{\boldsymbol{V}}$ is $pr = 90$ and for $\hat{\boldsymbol{V}}$ ignoring $\hat{\boldsymbol{U}}$ the sample size is $sr = 270$. For this example, the empirical mean is 633.87, while the estimate from Theorem 2 gives 638.44. Using the $(r/(r-1))$ adjustment we obtain 635.79. For larger samples, the approximations are closer. A comparison of these estimates of the mean to empirical means are shown in Table 3.

The ratio of the mean given by Theorem 2 to the asymptotic mean gives a good approximation of the critical values of the LRT statistic (2.12). Let $Q_{1-\alpha}$ be the $(1-\alpha)$th quantile from the distribution of the LRT (2.12), and let $\chi^2_{1-\alpha,\xi}$ be the $(1-\alpha)$th quantile from the chi-squared distribution with $\xi = sp(sp+1)/2 - s(s+1)/2 - p(p+1)/2 + 1$ degrees of freedom. Let k be the ratio of the mean given by Theorem 2 with the $r/(r-1)$ adjustment to the asymptotic mean, i.e.,

$$k = \frac{-r\left(sp\log 2 + \sum_{j-1}^{sp} \psi(0.5(r-j)) - sp\log(r)\right) - (r/(r-1))(s(s+1)/2 + p(p+1)/2 + sp - 1)}{sp(sp+1)/2 - s(s+1)/2 - p(p+1)/2 + 1}. \tag{3.3}$$

Then

$$Q_{1-\alpha} \approx k\chi^2_{1-\alpha,\xi}. \tag{3.4}$$

Table 4 has a comparison of the empirical critical values to those obtained from the approximation given by (3.4), and the values are very close. Fig. 1 shows a qq-plot of the quantiles from the empirical distribution versus the quantiles given by (3.4) for $s = 4$, $p = 2$, and $r = 25$. This approximation is very useful when the dimensions of the matrices are very large and determining the empirical distribution may not be computationally feasible. Note that even though Theorem 1 applies whether we sample from normal distributions or not, Theorem 2 depends on the multinormal likelihood. Thus, the approximation given by (3.4) is appropriate only when sampling from Gaussian distributions.

4. Empirical power estimates

We compute the power of the LRT based on the critical values from Table 1. As with the Type I error, the power can be approximated with the chi-square distribution when we have "large" samples from a multinormal distribution. For smaller samples, using the critical values can vastly overstate the power of the test. We will study the power for the following class of models:

$$C[(t+k,i),(t,j)] = \sigma^2(\gamma I(i \neq j) + I(i = j))\frac{\rho_i^k}{1-\rho_i\rho_j}, \tag{4.1}$$

Table 3
Empirical means versus mean estimate from Theorem 2

s	p	r	Emp mean[a]	Est mean 1[b]	Est mean 2[c]
4	2	25	29.93	31.00	30.17
4	2	50	26.60	27.11	26.70
4	2	100	25.29	25.48	25.27
4	2	200	24.53	24.72	24.62
4	3	15	119.39	120.85	118.92
4	3	20	93.78	95.32	93.90
4	3	25	84.49	85.87	84.74
4	3	50	71.75	72.45	71.89
4	3	100	67.13	67.37	67.09
4	3	200	64.97	65.11	64.97
4	4	20	207.62	208.49	206.65
4	4	25	173.80	175.14	173.68
4	4	50	137.50	138.74	138.02
4	4	100	126.07	126.76	126.41
4	4	200	121.36	121.65	121.48
4	5	25	316.58	319.08	317.25
4	5	50	227.72	228.69	227.80
4	5	100	203.85	204.55	204.11
4	5	200	194.56	194.73	194.51
4	6	50	345.17	345.80	344.70
4	6	100	301.19	301.74	301.19
4	6	200	284.14	284.74	284.47
6	6	50	927.14	929.81	928.24
6	6	100	733.81	734.22	733.45
6	6	200	673.37	673.55	673.16
9	2	25	206.70	212.44	209.73
9	2	50	152.28	154.94	153.61
9	2	100	136.57	137.67	137.02
9	2	200	130.10	130.48	130.15
9	3	30	633.87	638.44	635.79
9	3	50	437.37	439.33	437.75
9	3	100	372.21	373.16	372.38
9	3	200	348.13	348.76	348.37
9	4	50	914.73	916.81	914.98
9	4	100	720.36	721.22	720.32
9	4	200	660.33	660.55	660.09
13	2	30	504.78	516.30	512.19
13	2	61	327.08	332.47	330.49

[a]Mean determined from the empirical distribution.
[b]Mean estimated from Theorem 2.
[c]Mean estimated from Theorem 2 with the $(r-1)/r$ adjustment.

Table 4
Comparison of the empirical critical values to the adjusted chi-square values, $\alpha = 0.05$

s	p	r	Empirical crit value[a]	Adj chi-square crit value[b]
4	2	25	45.62	45.78
4	2	50	40.34	40.51
4	2	100	39.00	38.34
4	2	200	37.39	37.36
4	3	15	159.94	155.78
4	3	20	123.68	123.01
4	3	25	111.30	111.01
4	3	50	94.23	94.17
4	3	100	87.92	87.89
4	3	200	85.47	85.11
4	4	20	257.52	253.01
4	4	25	214.92	212.64
4	4	50	168.10	168.98
4	4	100	154.26	154.77
4	4	200	148.70	148.73
4	5	25	378.04	373.23
4	5	50	268.01	268.00
4	5	100	240.20	240.13
4	5	200	228.01	228.83
4	6	50	395.77	394.91
4	6	100	345.63	345.06
4	6	200	325.92	325.90
6	6	50	1017.42	1016.27
6	6	100	804.12	803.00
6	6	200	736.51	737.00
9	2	25	253.13	255.38
9	2	50	185.60	187.04
9	2	100	165.51	166.84
9	2	200	158.20	158.48
9	3	30	733.18	719.60
9	3	50	494.58	495.45
9	3	100	421.84	421.46
9	3	200	394.02	394.29
9	4	50	1007.10	1002.68
9	4	100	790.68	789.36
9	4	200	725.23	723.36

[a]The critical values were computed from 10,000 Monte Carlo runs, which are shown in Table 1.
[b]As determined by (3.3) and (3.4).

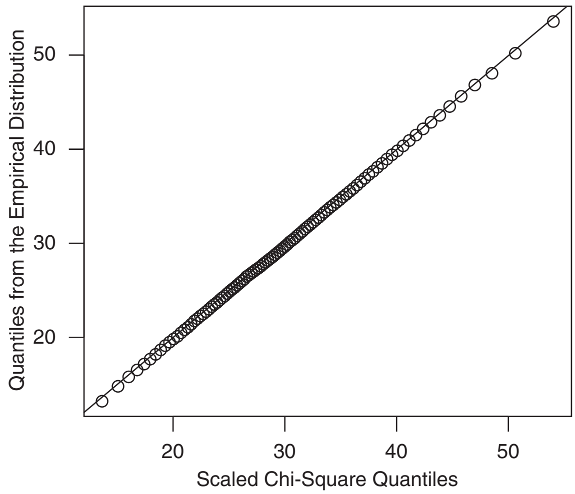

Fig. 1. QQ plot of scaled chi-square versus empirical distribution for $s = 4$, $p = 2$, $r = 25$.

where $I(\cdot)$ is the indicator function and i and j refer to variables i and j. When all the ρ_i are not all equal this covariance is not separable nor stationary. This is a useful model for multivariate repeated measures, where i and j are two different response variables and t and $t + k$ are two time points. This model implies that each variable has a different first-order autoregressive (AR(1)) time series. The separable version of this model is obtained when all the ρ_i are equal and is given by

$$C[(t + k, i), (t, j)] = \sigma^2(\gamma I(i \neq j) + I(i = j))\rho^k, \tag{4.2}$$

which is also stationary. For multivariate repeated measures, this implies that the covariance for the variables has a compound symmetry structure, and the time covariance is an AR(1) matrix.

We examine the power for model (4.1) for four variables and a various number of times. We compute the power among two sets of ρ, (1) $\rho_1 = 0.6$, $\rho_2 = 0.65$, $\rho_3 = 0.7$, and $\rho_4 = 0.75$ and (2) $\rho_1 = 0.9$, $\rho_2 = 0.7$, $\rho_3 = 0.7$, and $\rho_4 = 0.45$. For each case, $\gamma = 0.7$ and $\sigma^2 = 1$ were used. For all the simulations the regression model has $\boldsymbol{X} = \mathbf{1}_r$ and $\boldsymbol{B} = (\mu_1, \ldots, \mu_m)$, and without loss of generality we let $\boldsymbol{B} = \mathbf{0}$ (but are estimated assuming the former structure). We ran 10,000 Monte Carlo simulations in R. The results are shown in Table 5. As expected, the test had a higher power to detect (2) than (1). For both, we had reasonable power for sample sizes of 50 or more.

It is significant that the test we use makes no assumption about stationarity since a test with this assumption would have very low power here. To illustrate, suppose there are four locations with $\rho_1 = 0.5$, $\rho_2 = 0.6$, $\rho_3 = 0.7$, and $\rho_4 = 0.8$. Let $\rho(\cdot, \cdot)$ be the correlation function. Then the estimate of $\rho((t + k, i), (t, i))$ for all i would be the same as that for the model $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0.65$, which is separable. Thus, great care must be exercised in applying a test that assumes stationarity.

Table 5
Empirical power for the nonseparable model at level $\alpha = 0.05$

Dimension			Power for $\rho_1, \rho_2, \rho_3, \rho_4$[a]	
s	p	r	0.6,0.65,0.7,0.75	0.9,0.65,0.75,0.45
4	3	20	0.21	0.31
4	3	25	0.30	0.47
4	3	50	0.80	0.95
4	4	20	0.17	0.27
4	4	25	0.30	0.46
4	4	50	0.86	0.98
4	5	25	0.26	0.41
4	5	50	0.87	0.99

[a] $\gamma = 0.7$ was used for both cases.

5. Application to multivariate repeated measures data

We now demonstrate the test applied to a multivariate repeated data taken on blood counts. The data we use can be found in Rencher [19, p. 261], which were taken from Burdick [2]. Here data were collected on 20 subjects. White blood count, red blood count, and hemoglobin were measured four times each on a blood sample from each patient. On each of these occasions, one of four reagents was used; and the three variables were measured on each occasion. Because subject 14 had a couple of unusual observations, this subject was excluded from the analysis.

Let $\boldsymbol{U}$ ($s = 3$) be the covariance matrix of the responses, and let $\boldsymbol{V}$ ($p = 4$) be the covariance matrix of the repeated measurements of the same variable. Here $r = 19 > 3 \cdot 4 = 12$ so the test is applicable. We model a different mean for each variable and reagent combination: $\boldsymbol{X} = \mathbf{1}_{19}$ and $\boldsymbol{B} = (\mu_{11}, \ldots, \mu_{14}, \mu_{21}, \ldots, \mu_{24}, \mu_{31}, \ldots, \mu_{34})$. Note that if we had two treatment groups of 10 patients each, Theorem 1 still applies.

The test requires that each subject has the same covariance matrix. The variances may differ for the variable and reagent combinations, but each of these must be constant across subjects, as well as the covariances. It is not possible to test this assumption without some grouping of the data. However, we can test whether certain variances are homogeneous or not. First, we test whether the variance for each variable is constant across subjects (across all reagents) with the Levene-Med Test [12]. The p-values for testing the homogeneity of the variances across subjects for variables 1, 2, and 3 are 0.7023, 0.3841, and 1.000, respectively. Next, we test whether the variances for each reagent are homogeneous, and the p-values for each are 1. Thus, even though we cannot test whether the covariance matrix is constant across subjects, we are able to test whether certain variances are constant across subjects, and these tests indicate no deviation from homogeneous variances.

Since we generate the empirical distribution based on normal random variables, we assess the assumption of multivariate normality. Using the statistic E_p proposed by Doornik [5], which does not depend on aymptotic results, we obtained a p-value of 0.938. We also computed the test statistic based on an alternate form of the multivariate Shapiro–Wilk test (denoted W_p^* in [5]) and obtained a p-value of 0.540. Thus, it is reasonable to generate the empirical distribution of the test statistic based on normal random variables.

We empirically determined the critical value for the test statistic at the 0.05 level, which is 127.58 (0.22) based on 10,000 runs using normal random variables (the critical value using (3.4) is 126.82). Computing $\hat{\boldsymbol{U}}$, $\hat{\boldsymbol{V}}$, $\boldsymbol{S}$ and $\hat{\boldsymbol{B}}$, we obtain a test statistic of 153.95. Thus, we have very strong evidence that the covariance is *not* separable (the p-value using the empirical distribution is 0.0041, and the p-value using (3.4) is 0.0020). Let us examine $\hat{\boldsymbol{U}} \otimes \hat{\boldsymbol{V}}$ and $\boldsymbol{S}$ to see where the differences occur (since $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{V}}$ are not unique, we show the Kronecker product, which is unique). We have the following estimates:

$$\hat{\boldsymbol{U}} \otimes \hat{\boldsymbol{V}} = \left(\begin{array}{cccc|cccc|cccc}
2.15 & 2.16 & 2.14 & 2.11 & 0.05 & 0.05 & 0.05 & 0.05 & 0.19 & 0.19 & 0.19 & 0.19 \\
2.16 & 2.18 & 2.16 & 2.13 & 0.05 & 0.05 & 0.05 & 0.05 & 0.19 & 0.20 & 0.19 & 0.19 \\
2.14 & 2.16 & 2.16 & 2.12 & 0.05 & 0.05 & 0.05 & 0.05 & 0.19 & 0.19 & 0.19 & 0.19 \\
2.11 & 2.13 & 2.12 & 2.09 & 0.05 & 0.05 & 0.05 & 0.04 & 0.19 & 0.19 & 0.19 & 0.19 \\
\hline
0.05 & 0.05 & 0.05 & 0.05 & 0.19 & 0.19 & 0.19 & 0.18 & 0.16 & 0.16 & 0.16 & 0.16 \\
0.05 & 0.05 & 0.05 & 0.05 & 0.19 & 0.19 & 0.19 & 0.19 & 0.16 & 0.17 & 0.16 & 0.16 \\
0.05 & 0.05 & 0.05 & 0.05 & 0.19 & 0.19 & 0.19 & 0.19 & 0.16 & 0.16 & 0.16 & 0.16 \\
0.05 & 0.05 & 0.05 & 0.04 & 0.18 & 0.19 & 0.19 & 0.18 & 0.16 & 0.16 & 0.16 & 0.16 \\
\hline
0.19 & 0.19 & 0.19 & 0.19 & 0.16 & 0.16 & 0.16 & 0.16 & 2.08 & 2.09 & 2.07 & 2.04 \\
0.19 & 0.20 & 0.19 & 0.19 & 0.16 & 0.17 & 0.16 & 0.16 & 2.09 & 2.11 & 2.09 & 2.05 \\
0.19 & 0.19 & 0.19 & 0.19 & 0.16 & 0.16 & 0.16 & 0.16 & 2.07 & 2.09 & 2.08 & 2.04 \\
0.19 & 0.19 & 0.19 & 0.19 & 0.16 & 0.16 & 0.16 & 0.16 & 2.04 & 2.05 & 2.04 & 2.02
\end{array}\right) \tag{5.1}$$

and

$$\boldsymbol{S} = \left(\begin{array}{cccc|cccc|cccc}
2.81 & 2.75 & 2.75 & 2.67 & -0.04 & -0.04 & -0.07 & -0.07 & -0.03 & -0.01 & -0.04 & 0.00 \\
2.75 & 2.71 & 2.70 & 2.61 & -0.02 & -0.02 & -0.05 & -0.05 & 0.03 & 0.05 & 0.03 & 0.07 \\
2.75 & 2.70 & 2.70 & 2.61 & -0.02 & -0.02 & -0.05 & -0.05 & 0.04 & 0.07 & 0.04 & 0.08 \\
2.67 & 2.61 & 2.61 & 2.54 & -0.03 & -0.04 & -0.07 & -0.06 & -0.03 & -0.01 & -0.03 & 0.01 \\
\hline
-0.04 & -0.02 & -0.02 & -0.03 & 0.19 & 0.19 & 0.19 & 0.19 & 0.57 & 0.60 & 0.57 & 0.57 \\
-0.04 & -0.02 & -0.02 & -0.04 & 0.19 & 0.19 & 0.19 & 0.19 & 0.57 & 0.60 & 0.57 & 0.57 \\
-0.07 & -0.05 & -0.05 & -0.07 & 0.19 & 0.19 & 0.20 & 0.20 & 0.58 & 0.61 & 0.58 & 0.58 \\
-0.07 & -0.05 & -0.05 & -0.06 & 0.19 & 0.19 & 0.20 & 0.20 & 0.58 & 0.61 & 0.58 & 0.58 \\
\hline
-0.03 & 0.03 & 0.04 & -0.03 & 0.57 & 0.57 & 0.58 & 0.58 & 2.13 & 2.22 & 2.13 & 2.12 \\
-0.01 & 0.05 & 0.07 & -0.01 & 0.60 & 0.60 & 0.61 & 0.61 & 2.22 & 2.33 & 2.23 & 2.22 \\
-0.04 & 0.03 & 0.04 & -0.03 & 0.57 & 0.57 & 0.58 & 0.58 & 2.13 & 2.23 & 2.15 & 2.13 \\
0.00 & 0.07 & 0.08 & 0.01 & 0.57 & 0.57 & 0.58 & 0.58 & 2.12 & 2.22 & 2.13 & 2.13
\end{array}\right). \tag{5.2}$$

We can compare the matrices by examining each 4×4 block. We see that the covariances between white blood count and red blood count on the same patient (rows 1–4, columns 5–8) are higher for the separable model and appear to be uncorrelated under the alternative. The same pattern is seen with the covariances between the white blood count and hemoglobin. However, the covariances between the red blood count and hemoglobin (bottom middle block) are lower for the separable model, while the covariances for the repeated hemoglobin measurements on the same patient are higher for the separable model (the block in the lower right-hand corner).

6. Conclusions

We developed a likelihood ratio test (LRT) for separability of covariances that applies to multivariate repeated measures or a replicated spatio-temporal process. The LRT was derived from the multinormal likelihood functions, and the distribution of this statistic under the null hypothesis $H_0 : \boldsymbol{U} \otimes \boldsymbol{V}$ does not depend on the true values of $\boldsymbol{U}$ or $\boldsymbol{V}$. The distribution also does not depend on the true values of the coefficient matrix $\boldsymbol{B} = (\boldsymbol{\beta}_{(1)}, \ldots, \boldsymbol{\beta}_{(m)})$ in a multivariate regression model. This allows for the distribution to be developed empirically, which is especially important for small sample sizes where using chi-square critical values causes severe Type I errors. Estimated distributions for various dimensions of $\boldsymbol{U}$ and $\boldsymbol{V}$ for different sample sizes were developed. Note that even though the statistic is derived from the multinormal likelihood, the test can be applied to non-Gaussian distributions since we are deriving the distribution empirically. However, if we do sample from a multinormal distribution, we have an approximation of the critical values by scaling a chi-square critical value.

The test may be applied to multivariate repeated measures data and to some spatio-temporal processes. It is directly applicable to data with independent replicates where the number of replicates, r, is greater than the product of the dimensions of $\boldsymbol{U}$ and $\boldsymbol{V}$, such as the multivariate repeated measurements of blood counts demonstrated in Section 5. Some spatio-temporal data, such as electroencephalogram (EEG) data [10], have measurements taken at certain times and locations on a large number of patients and fit without modification into the framework for our proposed test of separability. Many spatio-temporal problems have only one realization. In such cases, when $r = 1$, it may be possible to apply this test by decomposing the data into pseudo-replicates. This idea is explored in [16].

The LRT proposed in this paper does not require a class of models to be specified. Furthermore, the test does not require isotropy or second-order stationarity of the covariances in either dimension. This is especially important since a test assuming stationarity will have very low power to detect certain nonstationary alternatives. However, the test does require a process with the same covariance matrix for each replicate.

Appendix

Proof of Theorem 1. Throughout, we use the same notation as in Section 2. First, we evaluate the determinant of $\boldsymbol{S}$.

$$\begin{aligned} \boldsymbol{S} &= (1/r)\boldsymbol{Y}^T(\boldsymbol{I}_r - \boldsymbol{P_X})\boldsymbol{Y} \\ &= (1/r)(\boldsymbol{U}^{1/2} \otimes \boldsymbol{V}^{1/2})\boldsymbol{E}^T(\boldsymbol{I}_r - \boldsymbol{P})\boldsymbol{E}(\boldsymbol{U}^{1/2} \otimes \boldsymbol{V}^{1/2}). \end{aligned} \tag{A.1}$$

Hence,

$$|\boldsymbol{S}| = |\boldsymbol{U} \otimes \boldsymbol{V}||(1/r)\boldsymbol{E}^T(\boldsymbol{I}_r - \boldsymbol{P_X})\boldsymbol{E}|. \tag{A.2}$$

Before computing the determinants of $\boldsymbol{U}$ and $\boldsymbol{V}$, we will use matrix algebra to re-formulate $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{V}}$. Let $\boldsymbol{E}_k$ and $\boldsymbol{e}_k$ be the analogous quantities to $\boldsymbol{Y}_k$ and $\boldsymbol{y}_k$. Let $\hat{\boldsymbol{E}} = \boldsymbol{P}_X\boldsymbol{E}$ (and $\hat{\boldsymbol{E}}_k$ and $\hat{\boldsymbol{e}}_k$ are analogous to $\hat{\boldsymbol{M}}_k$ and $\hat{\boldsymbol{m}}_k$). First note that

$$\begin{aligned} \boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k &= (\boldsymbol{I}_r - \boldsymbol{P_X})\boldsymbol{X}\hat{\boldsymbol{B}} = (\boldsymbol{I}_r - \boldsymbol{P_X})\boldsymbol{E}(\boldsymbol{U}^{1/2} \otimes \boldsymbol{V}^{1/2}) \\ &= (\boldsymbol{E} - \hat{\boldsymbol{E}})(\boldsymbol{U}^{1/2} \otimes \boldsymbol{V}^{1/2}). \end{aligned} \tag{A.3}$$

Since the rows of $\boldsymbol{Y}$ are $\boldsymbol{y}_k^T$, Eq. (A.3) implies that

$$\boldsymbol{y}_k - \hat{\boldsymbol{m}}_k = (\boldsymbol{U}^{1/2} \otimes \boldsymbol{V}^{1/2})(\boldsymbol{e}_k - \hat{\boldsymbol{e}}_k). \tag{A.4}$$

We will next show that $\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k = \boldsymbol{V}^{1/2}(\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k)\boldsymbol{U}^{1/2}$. In order to do this we need the following relationships:

$$\mathrm{vec}(\boldsymbol{E}_k) = \boldsymbol{e}_k, \tag{A.5}$$

where the vec operator stacks the columns of $\boldsymbol{E}_k$; and

$$\boldsymbol{E}_k = \sum_{i=1}^{s} (\boldsymbol{h}_i^T \otimes \boldsymbol{I}_p)\boldsymbol{e}_k \boldsymbol{h}_i^T, \tag{A.6}$$

where $\boldsymbol{h}_i$ is the $s \times 1$ vector with entry i equal to one and all other entries equal to zero.

Property (VIII) from Mardia et al. [15, p. 460] implies that

$$\mathrm{vec}(\boldsymbol{V}^{1/2}(\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k)\boldsymbol{U}^{1/2}) = (\boldsymbol{U}^{1/2} \otimes \boldsymbol{V}^{1/2})(\boldsymbol{e}_k - \hat{\boldsymbol{e}}_k). \tag{A.7}$$

By Eqs. (A.4) and (A.6), we have

$$\begin{aligned} \boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k &= \sum_{i=1}^{s} (\boldsymbol{h}_i^T \otimes \boldsymbol{I}_p)(\boldsymbol{y}_k - \hat{\boldsymbol{m}}_k)\boldsymbol{h}_i^T \\ &= \sum_{i=1}^{s} (\boldsymbol{h}_i^T \otimes \boldsymbol{I}_p)(\boldsymbol{U}^{1/2} \otimes \boldsymbol{V}^{1/2})(\boldsymbol{e}_k - \hat{\boldsymbol{e}}_k)\boldsymbol{h}_i^T. \end{aligned} \tag{A.8}$$

Substituting (A.7) into (A.8), and then applying (A.6) gives the desired result:

$$\boldsymbol{Y}_k - \hat{\boldsymbol{M}}_k = \boldsymbol{V}^{1/2}(\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k)\boldsymbol{U}^{1/2}. \tag{A.9}$$

As seen by Eqs. (2.4), $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{V}}$ do not have closed forms. However, there is another way to formulate these solutions. We can re-write these in terms of $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{E}_k$ and $\hat{\boldsymbol{E}}$. First let

$$E_1(\cdot) = \left(\frac{1}{pr} \sum_{k=1}^{r} (\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k)^T (\cdot)(\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k) \right)^{-1} \tag{A.10}$$

and

$$E_2(\cdot) = \left(\frac{1}{sr} \sum_{k=1}^{r} (\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k)(\cdot)(\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k)^T \right)^{-1}. \tag{A.11}$$

Now let

$$\hat{\boldsymbol{U}}^* = \boldsymbol{U}^{1/2} \left(\frac{1}{pr} \sum_{k=1}^{r} (\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k)^T (E_2 \circ E_1 \circ E_2 \circ E_1 \ldots)(\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k) \right) \boldsymbol{U}^{1/2} \tag{A.12}$$

and

$$\hat{\boldsymbol{V}}^* = \boldsymbol{V}^{1/2} \left(\frac{1}{sr} \sum_{k=1}^{r} (\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k)(E_1 \circ E_2 \circ E_1 \circ E_2 \ldots)(\boldsymbol{E}_k - \hat{\boldsymbol{E}}_k)^T \right) \boldsymbol{V}^{1/2}, \tag{A.13}$$

where $\circ$ denotes composition. We will show that these satisfy (2.4), and hence are the maximum-likelihood estimates. Substituting $\hat{\boldsymbol{V}}^*$ from (A.13) into the equation for $\hat{\boldsymbol{U}}$ from (2.4) we obtain

$$\begin{aligned}\hat{\boldsymbol{U}} &= \frac{1}{pr}\sum_{k=1}^{r}(\boldsymbol{Y}_k-\hat{\boldsymbol{M}}_k)^T\hat{\boldsymbol{V}}^{*-1}(\boldsymbol{Y}_k-\hat{\boldsymbol{M}}_k)\\ &= \boldsymbol{U}^{1/2}\left(\frac{1}{pr}\sum_{k=1}^{r}(\boldsymbol{E}_k-\hat{\boldsymbol{E}}_k)^T\boldsymbol{V}^{1/2}\hat{\boldsymbol{V}}^{*-1}\boldsymbol{V}^{1/2}(\boldsymbol{E}_k-\hat{\boldsymbol{E}}_k)\boldsymbol{U}^{1/2}\right)\\ &= \boldsymbol{U}^{1/2}\frac{1}{pr}\sum_{k=1}^{r}(\boldsymbol{E}_k-\hat{\boldsymbol{E}}_k)^T\left(\frac{1}{sr}\sum_{k=1}^{r}(\boldsymbol{E}_k-\hat{\boldsymbol{E}}_k)(E_1\circ E_2\ldots)(\boldsymbol{E}_k-\hat{\boldsymbol{E}}_k)^T\right)^{-1}\\ &\quad\times(\boldsymbol{E}_k-\hat{\boldsymbol{E}}_k)\boldsymbol{U}^{1/2}\\ &= \boldsymbol{U}^{1/2}\left(\frac{1}{pr}\sum_{k=1}^{r}(\boldsymbol{E}_k-\hat{\boldsymbol{E}}_k)^T(E_2\circ E_1\circ E_2\circ E_1\ldots)(\boldsymbol{E}_k-\hat{\boldsymbol{E}}_k)\right)\boldsymbol{U}^{1/2}.\end{aligned}\tag{A.14}$$

The same process shows the analogous result for $\hat{\boldsymbol{V}}$. Thus, $\hat{\boldsymbol{U}}=\boldsymbol{U}^{1/2}E_{**}\boldsymbol{U}^{1/2}$, where E_{**} does not depend on $\boldsymbol{U}$ or $\boldsymbol{V}$, and $\hat{\boldsymbol{V}}=\boldsymbol{V}^{1/2}E_{***}\boldsymbol{V}^{1/2}$, where E_{***} does not depend on $\boldsymbol{U}$ or $\boldsymbol{V}$. Substitution of these quantities and (A.2) into (2.12) gives

$$rp\log|E_{**}|+rs\log|E_{***}|-r\log|(1/r)\boldsymbol{E}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{E}|,\tag{A.15}$$

which is independent of $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{B}$. □

Proof of Theorem 2. First,

$$E(r\log|\boldsymbol{S}|)=r\left(sp\log 2+\sum_{j-1}^{sp}\psi(0.5(r-j))-sp\log(r)\right),\tag{A.16}$$

where ψ is the digamma function [11]. The asymptotic mean of $r\log|\hat{\boldsymbol{U}}\otimes\hat{\boldsymbol{V}}|$ can be determined by taking the limit of the LRT statistic for $H_0:\boldsymbol{U}\otimes\boldsymbol{V}=\boldsymbol{I}$ versus $H_1:\boldsymbol{U}\otimes\boldsymbol{V}\neq\boldsymbol{I}$. Negative twice the logarithm of the difference in likelihoods is given by

$$\hat{L}_0=-r\log|\hat{\boldsymbol{U}}\otimes\hat{\boldsymbol{V}}|-rsp+r\,\mathrm{trace}(\boldsymbol{S}).\tag{A.17}$$

Each term in $r\,\mathrm{trace}(\boldsymbol{S})$ has a chi-square distribution with $r-1$ degrees of freedom, hence $r\,\mathrm{trace}(\boldsymbol{S})$ has a chi-square distribution with $sp(r-1)$ degrees of freedom since the trace is sum of sp independent chi-square random variables. The asymptotic mean of $\hat{L}_0$ is $s(s+1)/2+p(p+1)/2-1$. Hence, the asymptotic mean of $r\log|\hat{\boldsymbol{U}}\otimes\hat{\boldsymbol{V}}|$ is $-(s(s+1)/2+p(p+1)/2-1+sp)$. Subtracting (A.16) from this yields the desired result. □

References

[1] P. Brown, P. Diggle, M. Lord, P. Young, Space–time calibration of radar rainfall data, Appl. Statist. 50 (2001) 221–241.

[2] R. Burdick, On the use of canonical analysis to test MANOVA hypotheses, Presented at the Annual Meeting of the American Statistical Association, Washington, DC, August 1979.
[3] N. Chaganty, D. Naik, Analysis of multivariate longitudinal data using quasi-least squares, J. Statist. Plann. Inference 103 (2002) 421–436.
[4] M. Chernick, Bootstrap Methods, Wiley, New York, 1999.
[5] J. Doornik, H. Hansen, An Omnibus Test for Univariate and Multivariate Normality, available at www.doornik.com/research/normal2.pdf, unpublished.
[6] P. Dutilleul, The MLE algorithm for the matrix normal distribution, J. Statist. Comput. Simulation 64 (1999) 105–123.
[7] P. Dutilleul, Maximum likelihood estimation for the matrix normal distribution, Technical Report, Department of Mathematics and Statistics, McGuill University, Montreal, Quebec, Canada, 1994.
[8] M. Fuentes, Testing separability of spatial-temporal covariance functions, J. Statist. Plann. Inference, 2006, in press.
[9] A. Galecki, General class of covariance structures for two or more repeated factors in longitudinal data analysis, Comm. Statist.—Theory Methods 23 (11) (1994) 3105–3119.
[10] H. Huizenga, J. Munck, L. Waldorp, R. Grasman, Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model, IEEE Trans. Biomed. Eng. 49 (6) (2002) 533–539.
[11] S. Kocherlakota, K. Kocherlakota, in: S. Kotz, N. Johnson (Eds.), Generalized Variance in Encyclopedia of Statistical Sciences, vol. 3, Wiley, New York, 1983, pp. 354–357.
[12] R. Kuehl, Statistical Principles of Research Design and Analysis, Wadsworth, Belmont, 1994.
[13] P. Kyriakidis, A.G. Journel, Geostatistical space–time models: a review, Math. Geol. 31 (1999) 651–684.
[14] K. Mardia, C. Goodall, Spatial-temporal analysis of multivariate environmental data, in: G. Patil, C. Rao (Eds.), Multivariate Environmental Statistics, Elsevier, Amsterdam, 1993, pp. 347–386.
[15] K. Mardia, J. Kent, J. Bibby, Multivariate Analysis, ninth ed., Academic Press, London, 1994.
[16] M. Mitchell, M.G. Genton, M. Gumpertz, Testing separability of space–time covariances, Environmetrics, 2005, in press.
[17] M. Mitchell, M. Gumpertz, Spatial variability inside a free-air CO_2 system, J. Agric. Biol. Environ. Statist. (JABES) 8 (3) (2003) 310–327.
[18] R. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.
[19] A. Rencher, Methods of Multivariate Analysis, Wiley, New York, 1995.
[20] I. Rodriguez-Iturbe, J.M. Mejia, The design of networks in time and space, Water Resources Res. 10 (4) (1974) 713–728.
[21] SAS Institute Inc., SAS/IML User's Guide, Version 8, SAS Publishing, Cary, NC, 2000.
[22] J. Schott, Matrix Analysis for Statistics, Wiley, New York, 1997.
[23] M. Shitan, P. Brockwell, An asymptotic test for separability of a spatial autoregressive model, Comm. Statist.—Theory Methods 24 (1995) 2027–2040.