

Testing for separability of space–time covariances

Matthew W. Mitchell^{1,*†}, Marc G. Genton² and Marcia L. Gumpertz³

¹*Metabolon, Inc., Durham, NC27713, U.S.A.*

²*Department of Statistics, Texas A&M University, College Station, TX, U.S.A.*

³*Department of Statistics, North Carolina State University, Raleigh, NC, U.S.A.*

SUMMARY

Separable space–time covariance models are often used for modeling in environmental sciences because of their computational benefits. Unfortunately, there are few formal statistical tests for separability. We adapt a likelihood ratio test based on multivariate repeated measures to the spatio–temporal context. We apply this test to an environmental monitoring data set. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: elevated CO₂; FACE; Kronecker product; separable space–time covariance

1. INTRODUCTION

Separable models are often used in spatio–temporal models or multivariate repeated-measures because of their computational simplicity compared to non-separable models (Rodriguez-Iturbe and Mejia, 1974; Dutilleul, 1999; Huizenga *et al.*, 2002). The definition of separability is often formulated in the context of second-order stationary models and is given by the following (Kyriakidis and Journel, 1999):

$$C(\mathbf{h}, k) = C_1(\mathbf{h})C_2(k) \quad (1)$$

for all spatial and temporal lags \mathbf{h} and k , respectively, where C represents the joint covariance function, and C_1 and C_2 represent spatial and temporal covariance functions, respectively. With this definition, the overall joint process can be seen as the product of two independent processes, one that occurs in space and another that occurs in time. However, we do not observe realizations of the two separate processes, only the joint process. In terms of the joint process, separability can be formulated as

$$C(\mathbf{h}, k) = (1/\sigma^2)C(\mathbf{h}, 0)C(\mathbf{0}, k) \quad (2)$$

where $\sigma^2 = C(\mathbf{0}, 0)$ is the variance of the process.

*Correspondence to: Matthew W. Mitchell, Metabolon, Inc., 800 Capitola Dr., Durham, NC 27713, U.S.A.

†E-mail: mwmitch2@aol.com

However, separable processes need not be stationary, and non-separable processes may be stationary such as those proposed by Cressie and Huang (1999) and Gneiting (2002). We use a more general definition of separability that involves Kronecker products. Let \mathbf{U} and \mathbf{V} be the s -dimensional and p -dimensional variance–covariance matrices for space and time, respectively, and let Σ denote the sp by sp variance–covariance matrix of the joint-process. The covariance is separable if and only if

$$\Sigma = \mathbf{U} \otimes \mathbf{V} \quad (3)$$

Note that \mathbf{U} and \mathbf{V} are unique only up to constant multiples since $a\mathbf{U} \otimes (1/a)\mathbf{V} = \mathbf{U} \otimes \mathbf{V}$ for $a \neq 0$.

Because the determinant and inverse of Kronecker products are computed from the determinants and inverses of \mathbf{U} and \mathbf{V} , separability provides great computational convenience. For example, inversion of the variance–covariance matrix is required for linear prediction. For the data in our application, a non-separable covariance requires inversion of a 1586×1586 matrix. However, a separable covariance requires only the inversion of a 122×122 matrix and a 13×13 matrix.

Separability is a convenient property, but there are few formal statistical tests available. Many applications test separability for particular classes of models such as second-order stationary spatial autoregressive models (Shitan and Brockwell, 1995) and ‘blur-generated’ models (Brown *et al.*, 2001). One class of models recently proposed by Gneiting (2002) has a parameter that controls the degree of non-separability. The function is given by

$$C(\mathbf{h}, k) = \frac{\sigma^2}{(a|k| + 1)} \exp\left(-\frac{c\|\mathbf{h}\|}{(a|k| + 1)^{\beta/2}}\right) \quad (4)$$

where $a \geq 0$, $c \geq 0$, and $0 \leq \beta \leq 1$. This model will be referred to henceforth as the ‘Gneiting model’. When $\beta = 0$ the model is separable, and the most extreme non-separable model occurs when $\beta = 1$. Thus, one could test separability by testing whether $\beta = 0$ or not.

Cressie and Huang (1999) proposed a class of non-separable second-order stationary isotropic models given by

$$C(\mathbf{h}, k) = \frac{\sigma^2(a|k| + 1)}{((a|k| + 1)^2 + b^2\|\mathbf{h}\|^2)^{3/2}} \quad (5)$$

where $a \geq 0$ and $b \geq 0$. This model will be henceforth referred to as the ‘Cressie and Huang non-separable model’. It is possible to form a corresponding separable model from (5) by forming $(1/\sigma^2)C(\mathbf{h}, 0)C(\mathbf{0}, k)$, which is given by

$$C(\mathbf{h}, k) = \frac{\sigma^2}{(a|k| + 1)^2(b^2\|\mathbf{h}\|^2 + 1)^{3/2}} \quad (6)$$

This model will henceforth be referred to as the ‘Cressie and Huang separable model’. One could test separability by testing which gives the better fit (note that both of these have the same number of parameters).

For the Cressie and Huang and the Gneiting models, we created contour plots of the *correlations* for a strongly correlated process ($\rho(0, 1) = \rho(1, 0) = 0.85$, where $\rho(\|\mathbf{h}\|, |k|)$ is the correlation function)

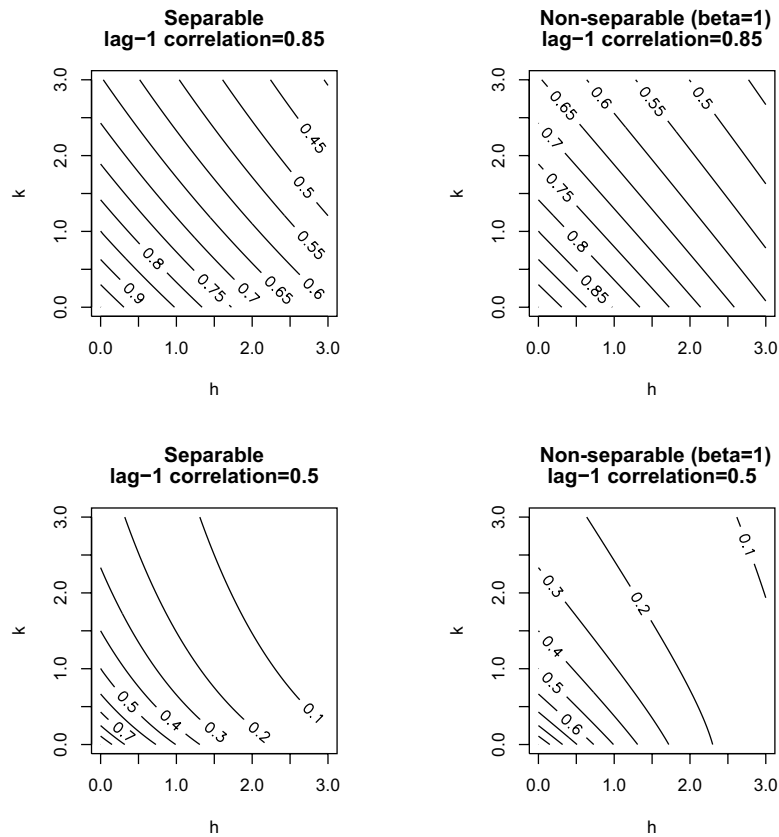


Figure 1. Contour plots of Gneiting correlation models (4)

and a moderately correlated process ($\rho(1, 0) = \rho(0, 1) = 0.5$). These plots are shown in Figures 1 and 2. For the Gneiting models, we compare the separable case ($\beta = 0$) to the most extreme non-separable case ($\beta = 1$).

For the Cressie and Huang models, we see for the strongly correlated processes that the non-separable and separable models are very similar for $h \leq 2$ and $k \leq 2$. For the moderately correlated processes, the two models are very similar for $h \leq 1.5$ and $k \leq 1.5$. However, beyond these lags the correlations are lower than 0.2. A similar pattern occurs for both cases with the Gneiting models. So there is not much difference for strongly correlated processes for the separable Gneiting model and the most extreme non-separable model. The differences for the Gneiting or Cressie and Huang models occur for larger lags or when the correlations are lower. Hence, even if a test for separability shows that the non-separable model is better within that class of models, there may be little important difference between the separable and non-separable models to justify the loss in computational benefits. Furthermore, if we decide, for example, that the non-separable Cressie and Huang model is better than the separable one, there may be a different separable model that fits better than the non-separable Cressie and Huang model.

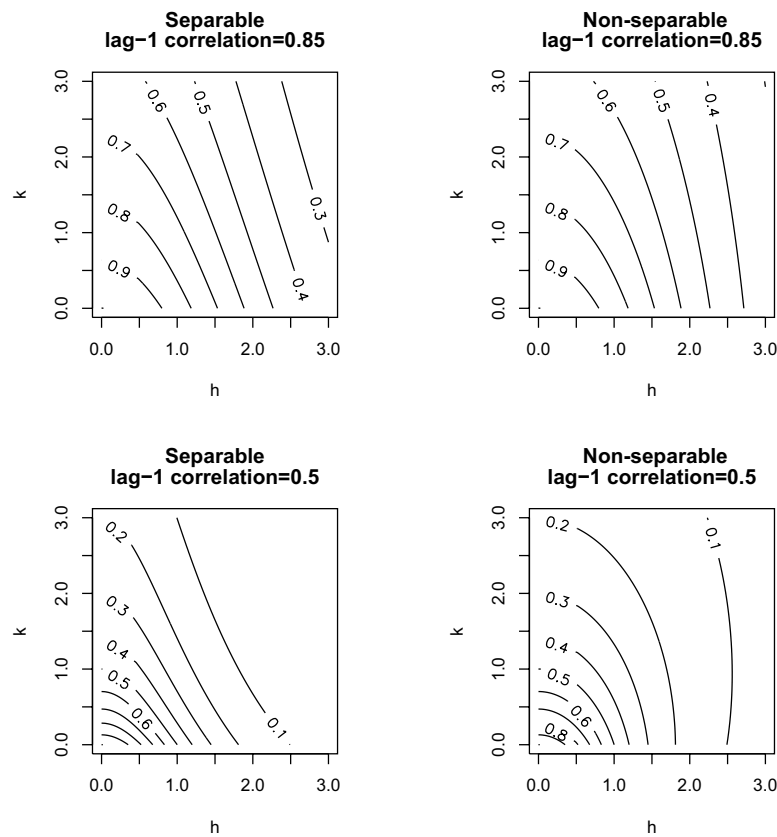


Figure 2. Contour plots of Cressie and Huang correlation models (5), (6)

As we see, testing separability within a class of models has its limitations. There are a few general tests. Fuentes (2005) proposes a test of separability for spatio-temporal processes by using properties of the spectral representation. Dutilleul (1999) tests separability in the context of independent identically distributed (i.i.d.) multivariate repeated measures, but there are some limitations to this application (Mitchell *et al.*, 2004). Recently, Mitchell *et al.* (2004) proposed a likelihood ratio test (LRT) for independent identically distributed (i.i.d.) multivariate repeated measures. This test is simple to implement and does not require stationarity in time or space, but does require the number of subjects to be greater than the product of the number of locations and the number of times, which is not realistic for most spatio-temporal applications and will need to be addressed (see Section 3).

First in Section 2 we discuss testing separability for multivariate repeated measures, and we give an important theorem concerning the distribution of that test statistic. Then, in Section 3, we discuss adapting the test to the case of one replicate. For this case type I error and power analyses are conducted. In Section 4 we apply the test to the RiceFace data, which is an environmental monitoring data set.

2. TESTING SEPARABILITY: MULTIVARIATE REPEATED MEASURES

Let $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_r$ be i.i.d. normal random vectors with $E(\mathbf{y}_k) = \mathbf{m}$ (each time–location combination may have a different mean, but each replicate has the same mean vector) and variance–covariance matrix Σ , where $\mathbf{y}_k = (y_{11k}, \dots, y_{1pk}, \dots, y_{s1k}, \dots, y_{spk})^T$, with s and p being the numbers of spatial locations and time points, respectively. Let \mathbf{Y}_k be the $p \times s$ matrix such that $\text{vec}(\mathbf{Y}_k) = \mathbf{y}_k$, where vec is the vectorization operator, and similarly let $\bar{\mathbf{Y}}$ be the corresponding matrix such that $\text{vec}(\bar{\mathbf{Y}}) = \bar{\mathbf{y}}$. When the null hypothesis of separability holds, the variance–covariance can be written as $\mathbf{U} \otimes \mathbf{V}$. The maximum likelihood estimators for \mathbf{U} and \mathbf{V} satisfy (Dutilleul, 1999):

$$\begin{aligned}\hat{\mathbf{U}} &= \frac{1}{pr} \sum_{k=1}^r (\mathbf{Y}_k - \bar{\mathbf{Y}})^T \hat{\mathbf{V}}^{-1} (\mathbf{Y}_k - \bar{\mathbf{Y}}) \\ \hat{\mathbf{V}} &= \frac{1}{sr} \sum_{k=1}^r (\mathbf{Y}_k - \bar{\mathbf{Y}}) \hat{\mathbf{U}}^{-1} (\mathbf{Y}_k - \bar{\mathbf{Y}})^T\end{aligned}\quad (7)$$

Mitchell *et al.* (2004) derived negative twice the difference in the log-likelihoods as

$$rs \log|\hat{\mathbf{V}}| + rp \log|\hat{\mathbf{U}}| - r \log|S| \quad (8)$$

where S is the usual maximum likelihood estimator of an unpatterned, symmetric, positive-definite matrix. Mitchell *et al.* (2004) proved the following theorem.

Theorem 1: The distribution of the LRT statistic (8) under the null hypothesis of separability does not depend on the true values of \mathbf{m} , \mathbf{U} or \mathbf{V} .

Theorem 1 allows one to develop empirical distributions, and examples of these for various combinations of s , p and r are given in Mitchell *et al.* (2004). Since the choices of \mathbf{m} , \mathbf{U} and \mathbf{V} do not affect the distribution of the statistic, one may simulate the distribution by choosing $\mathbf{m} = \mathbf{0}$, $\mathbf{U} = \mathbf{I}_s$ and $\mathbf{V} = \mathbf{I}_p$. In other words, simply generate $N(0, 1)$ random variables, fit the mean (a mean for every space–time combination), compute $\hat{\mathbf{U}}$, $\hat{\mathbf{V}}$ and S and compute the LRT statistic (8). Repeat this process a large number of times in order to develop the empirical distribution. The necessary statistics can be computed with SAS[®] PROC MIXED (SAS Institute, 2000), but we are able to compute the necessary statistics much faster with *R* using Dutilleul’s algorithm (Dutilleul, 1999) for estimating $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$. Note that it is required that $r > sp$ in order for S to be positive-definite. In the next section, we discuss how to apply this test to the case when $r = 1$.

3. TESTING SEPARABILITY IN THE SPATIO–TEMPORAL CONTEXT

Unlike multivariate repeated measures, spatio–temporal processes typically have $r = 1$ (a notable exception can be found in Huizenga *et al.*, 2002). The LRT statistic (8) in the previous section can be adapted to the case when $r = 1$. Let \mathbf{Y} be the vector of all the observations with $E(\mathbf{Y}) = \mathbf{X}\mathbf{B}$, where the dimensions of \mathbf{Y} , \mathbf{X} and \mathbf{B} are ps by 1, ps by q , and q by 1, respectively. The test can be adapted as follows: (i) create bootstrap replicates by fitting \mathbf{U} , \mathbf{V} , \mathbf{B} , and sampling from the residuals, $(\hat{\mathbf{U}} \otimes \hat{\mathbf{V}})^{-1/2}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})$, to compute the test statistic, or (ii) create pseudo-replicates from the data by partitioning the data into approximately independent ‘chunks’. The first approach is practical for moderately sized data sets. For example, suppose $s = 10$ and $p = 10$. Create 200 ‘reps’ (or some other number above $100 = 10 \times 10$.) and compute the test statistic. Then generate an empirical distribution

for $s = 10, p = 10, r = 200$. However, this approach is limited by the somewhat arbitrary choice of r and is not practical for larger data sets (the empirical distribution is not computationally feasible). The second approach is practical for data sets that are rich in one dimension relative to the other, such as the RiceFACE data (see Section 4). We now discuss this technique in more detail.

Suppose there are nine locations and 200 times. One could decompose the data into pseudo-replicates as follows: let times 1–4 be ‘rep’ 1, times 5–8 be ‘rep’ 2, etc. This yields 50 ‘replicates’ with nine locations and four times. For the replicates to be approximately i.i.d., the mean must be constant across ‘reps’ and each ‘rep’ must have the same temporal covariance (which requires a condition close to stationarity in time for the original process) with low correlations between ‘reps’. We use s, p and r as before, and we denote the number of times within a pseudo-replicate by p^* , and we denote the number of pseudo-replicates by r^* .

First we performed simulations to estimate the type I error for various space–time combinations using this approach. We assess the effects of the dimensions of the pseudo-replicates and the temporal correlation of the original process on the type I error. A zero mean model was used, and no mean was fitted. We performed 2500 simulation runs for each combination of s, p^*, r^* and compared these to critical values computed from 10 000 runs when $s = s, p = p^*$ and $r = r^*$. All these computations were performed in R , and the code may be obtained from the authors.

We modeled the spatial correlation with an exponential correlation function with lag-1 correlation approximately equal to 0.75. The spatial locations used in the simulations were positioned on a regular square grid with unit spacing. The simulated process had 200 time points with four and nine spatial locations. For each combination we chose an AR(1) model with temporal correlation parameters $\rho = 0.1, 0.2, \dots, 0.8, 0.9$. We set $\sigma^2 = 1$ as the variance of the joint-process. We created pseudo-replicates with $p^* = 2, 3$ or 4, which gives 100, 66 and 50 pseudo-replicates, respectively (for $p^* = 3$ we omitted the last 2 points).

For $p^* = 2$ and $p^* = 3$ with $s = 4$ or $s = 9$, the type I errors are approximately the nominal levels α (0.10, 0.05, 0.01) when $\rho = 0.7$ or less. For $p^* = 4$ with $s = 4$ or $s = 9$, the approximate type I errors are equal to the nominal levels when $\rho = 0.8$ or less. For $\alpha = 0.05$, the type I errors for $\rho = 0.8$ and $p^* = 2$ are approximately 0.13 and 0.21 for $s = 4$ and $s = 9$, respectively. A similar increase in type I error from $s = 4$ to $s = 9$ occurs with the other combinations as well. In general, the type I error is inversely proportional to p^* , but directly proportional to s . However, with smaller p^* , more replicates are possible, which increases the power. The results are presented in Table 1.

Table 1. Estimating type I error with AR(1) temporal autocorrelation

ρ^a	p^*	Type I error ^b $s = 4$	Type I error ^b $s = 9$
0.7	2	0.05	0.04
0.8	2	0.13	0.21
0.9	2	0.44	0.88
0.7	3	0.04	0.04
0.8	3	0.07	0.09
0.9	3	0.25	0.46
0.7	4	0.04	0.04
0.8	4	0.05	0.06
0.9	4	0.11	0.18

^a Temporal autocorrelation.

^b Level is 0.05.

We also performed simulations with another temporal correlation function:

$$\rho(k) = \frac{1}{(ak + 1)^2} \quad (9)$$

which is the function used in example 3 in Cressie and Huang (1999) when the spatial lag $\|\mathbf{h}\|$ is zero. The parameter a was chosen to yield lag-1 temporal correlations of $\rho(1) = 0.1, 0.2, \dots, 0.8, 0.9$. This temporal correlation decays more slowly than the AR(1) models, so we expect higher type I errors. The same spatial correlation matrix was used as before. For $p^* = 2$ with $s = 4$ or $s = 9$, the type I errors are approximately the nominal levels for ρ equal to 0.5 or less. For $p^* = 3$ with $s = 4$ or $s = 9$, the type I errors are approximately the nominal levels for ρ equal to 0.6 or less, and for $p^* = 4$ it is 0.7 or less. The results are presented in Table 2. Thus, as expected there are higher type I errors than with the AR(1) models, but the pattern is similar (the type I error is inversely proportional to p^* , but directly proportional to s).

For highly correlated data, it is possible to reduce the type I error by creating ‘gaps’. For example, with $p = 200$ omit $t = 3, 6, \dots, 195, 198$, create the pseudo-replicates with $t = 1, 2, t = 4, 5$, etc. For the two temporal correlation functions discussed earlier with $s = 4$, we computed the type I error when omitting every third data point (67 pseudo-replicates with $p^* = 2$), every fourth data point (50 pseudo-replicates with $p^* = 3$), and every fifth data point (40 pseudo-replicates with $p^* = 4$). The estimated type I errors based on 2500 simulation runs are shown in Tables 3 and 4. We see that the type I errors are much lower than when there are no gaps. However, many still have type I errors greater than the nominal levels.

To further reduce the type I error, larger ‘gaps’ can be created. We did this by omitting every other pair of times (rep 1 has $t = 1, 2$, rep 2 has $t = 5, 6$, etc.); which for $t = 200$ creates 50 pseudo-replicates with $p^* = 2$. To avoid discarding data, it is also possible to compute the average statistic across both subsets—i.e. subset one has $t = 1, 2, t = 5, 6$, etc. and subset two has $t = 3, 4, t = 7, 8$, etc.—compute the LRT for each, and then compute the average LRT value. This further reduces the

Table 2. Estimating type I error with Cressie–Huang temporal autocorrelation

ρ^a	p^*	Type I error ^b $s = 4$	Type I error ^b $s = 9$
0.5	2	0.05	0.05
0.6	2	0.06	0.07
0.7	2	0.09	0.14
0.8	2	0.22	0.41
0.9	2	0.56	0.95
0.5	3	0.05	0.05
0.6	3	0.06	0.05
0.7	3	0.06	0.07
0.8	3	0.13	0.17
0.9	3	0.33	0.62
0.5	4	0.04	0.04
0.6	4	0.05	0.05
0.7	4	0.05	0.05
0.8	4	0.07	0.09
0.9	4	0.16	0.27

^a Lag-1 temporal autocorrelation.

^b Level is 0.05.

Table 3. Estimating type I error with AR(1) temporal autocorrelation with gaps, $s = 4$

ρ^a	p^*	Type I error ^b with gaps	Type I error ^b , no gaps
0.8	2	0.08	0.13
0.9	2	0.26	0.44
0.8	3	0.06	0.07
0.9	3	0.14	0.25
0.8	4	0.05	0.05
0.9	4	0.09	0.11

^a Lag-1 temporal autocorrelation.^b Level is 0.05.Table 4. Estimating type I error with Cressie–Huang temporal autocorrelation with gaps, $s = 4$

ρ^a	p^*	Type I error ^b with gaps	Type I error ^b , no gaps
0.8	2	0.14	0.22
0.9	2	0.37	0.56
0.8	3	0.09	0.13
0.9	3	0.20	0.33
0.8	4	0.05	0.07
0.9	4	0.11	0.16

^a Lag-1 temporal autocorrelation.^b Level is 0.05.

type I error, although for many cases the estimated type I error from using the averaged statistics is lower than the nominal levels as shown in Tables 5 and 6. Overall, we see that increasing the ‘gaps’ reduces the type I error as expected; however, it also reduces the power since the number of pseudo-replicates is decreased.

Now we examine the power of our test. We estimate the power for the following class of models:

$$C[(s_i, t + k), (s_j, t)] = \sigma^2 \exp(-b\|\mathbf{h}\|) \frac{\rho_i^k}{1 - \rho_i \rho_j} \quad (10)$$

where $\mathbf{h} = s_i - s_j$. This covariance is neither separable nor (second-order) stationary when all the ρ_i are not equal. A process with this covariance has a different first-order autoregressive time series for each location. When the ρ_i are all equal, then the covariance is separable and stationary.

Table 5. Estimating type I error with AR(1) temporal autocorrelation omitting alternate pairs, $s = 4$

ρ^a	Type I error ^b with gaps	Type I error ^b with gaps, averaged	Type I error ^b with no gaps
0.8	0.06	0.02	0.13
0.9	0.17	0.10	0.44

^a Lag-1 temporal autocorrelation.^b Level is 0.05.

Table 6. Estimating type I error with Cressie–Huang temporal autocorrelation omitting alternate pairs, $s = 4$

ρ^a	Type I error ^b with gaps	Type I error ^b with gaps, averaged	Type I error ^b with no gaps
0.7	0.06	0.02	0.09
0.8	0.09	0.03	0.22
0.9	0.26	0.18	0.56

^a Lag-1 temporal autocorrelation.

^b Level is 0.05.

We test the power for three scenarios: (i) four locations on a regular grid with unit spacing, 200 time points, 100 pseudo-replicates, no gaps ($p^* = 2$); (ii) four locations on a regular grid with unit spacing, 200 time points, no gaps, 50 pseudo-replicates, no gaps ($p^* = 4$); and (iii) nine locations on an irregular grid, 100 time points, 50 pseudo-replicates ($p^* = 2$). For (i) and (ii), we set $b = 0.37$ (which gives an approximate lag-1 spatial correlation of 0.7), $\rho_1 = 0.8$, $\rho_2 = 0.6$, $\rho_3 = 0.8$ and $\rho_4 = 0.3$ (average lag-1 temporal correlation is 0.625). For (iii) we set $b = 0.37$ and $\rho_1 = 0.8$, $\rho_2 = 0.6$, $\rho_3 = 0.8$, $\rho_4 = 0.5$, $\rho_5 = 0.5$, $\rho_6 = 0.7$, $\rho_7 = 0.85$, $\rho_8 = 0.4$ and $\rho_9 = 0.75$ (average is approximately 0.63). The nine spatial locations for (3) are shown in Figure 3.

Simulations were performed similarly as with the type I error computations with 10 000 simulation runs for each. From Table 1 we see that the type I error is approximately the nominal level for these temporal correlations. We use a nominal level of 0.05 for all tests.

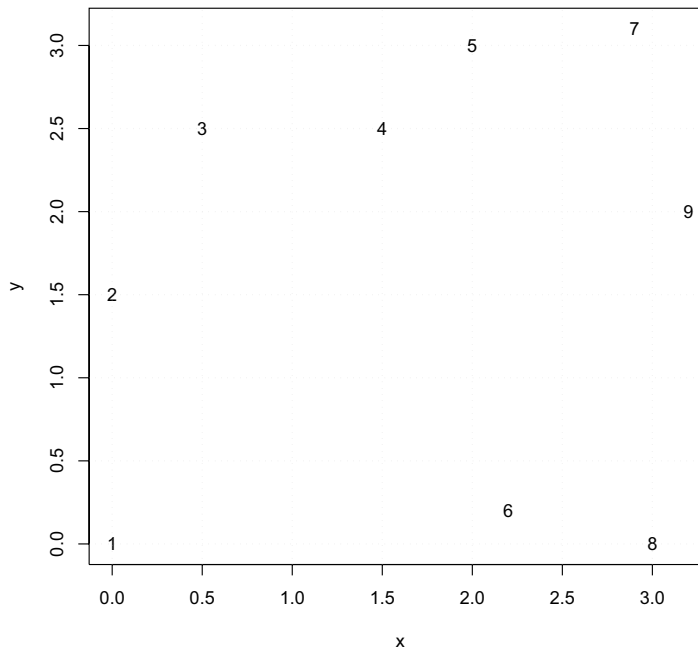


Figure 3. Grid used for power simulation (3)

The test had very high power for each scenario. The empirical powers for the three cases are the following: (i) 1.0000, (ii) 0.9999 and (iii) 0.9618. Scenario (iii) has similar dimensions to the RiceFACE data, which we analyze in Section 4, so our test should have sufficient power to detect such a departure from separability. Note that a test that assumes stationarity has no power to detect this covariance since it assumes that the covariance does not depend on (x, y, t) , and the time correlation depends on the location.

4. TESTING SEPARABILITY OF FACE DATA

The Japanese RiceFACE (Free-Air CO₂ Exchange) project was a study of the effects of elevated levels of CO₂ on rice. The project spanned 1998–2000 and is one of many FACE projects around the world, which investigate the effects of global change on various ecosystems. The RiceFACE system emits CO₂ directly into the air in a field, and the CO₂ is dispersed by the wind. The experiment was performed in eight plots in a set of rice fields in Shizukuishi, Iwate, Japan. Four plots received no additional CO₂ above the ambient levels, while the other four received CO₂ from rings that consisted of polyethylene tubes. We refer to the area enclosed by the ring as a ‘FACE plot’. More details of the RiceFACE experiment are found in Okada *et al.* (2001).

Because the wind disseminated the CO₂, the CO₂ was not uniformly distributed throughout the FACE plots. One of our responsibilities was to estimate the concentration of CO₂ in various sub-regions in the plots across the whole season and for various periods during the season. To do this, we developed a spatio-temporal model, which employed meteorological covariates for the mean and had a separable covariance. More details of our spatio-temporal model and the RiceFACE project are given in Mitchell and Gumpertz (2003).

For testing separability we use data from the year 2000, which had 13 locations and 122 time points. The analysis is performed separately for each FACE plot. We fit the mean first and then partition the residuals into pseudo-replicates.

For the mean we fitted the target level (*tar*), the height of the plants (*H*), the aperture of the solenoid valve (*mv*), the wind speed (*ws*) and the square of the wind speed (*ws*²) for temporal covariates, and fit different coefficients for each location, i.e.

$$m(i, t) = \beta_{0i} + \beta_{1i}tar(t) + \beta_{2i}H(t) + \beta_{3i}mv(t) + \beta_{4i}ws(t) + \beta_{5i}ws(t)^2 \quad (11)$$

for $t = 1, 2, \dots, 121, 122$ and locations $i = 1, 2, \dots, 13$.

We now show that the covariance of the residuals is separable for this mean model when the covariance of the original process is separable. This is because all the regressor variables are time-dependent only and not spatially varying. Let the data be arranged so that we have time = 1, loc = 1, ..., loc = 13, time = 2, loc = 1, ..., loc = 13, ..., time = 122, loc = 1, ..., loc = 13, with variance-covariance matrix $V \otimes U$ for the joint space-time process. Let X be the matrix of covariates. Here

$$X = T \otimes I_{13} \quad (12)$$

where T is the matrix of temporal covariates. Let P_X be the projection matrix of X , which is $X(X^T X)^{-1} X^T$. It is straightforward to show that

$$P_X = P_T \otimes I_{13} \quad (13)$$

where P_T is the projection matrix of T . The covariance of the residuals for the RiceFACE data is given by

$$(\mathbf{I}_{1586} - \mathbf{P}_X)(\mathbf{V} \otimes \mathbf{U})(\mathbf{I}_{1586} - \mathbf{P}_X) \quad (14)$$

Since $\mathbf{I}_{1586} = \mathbf{I}_{122} \otimes \mathbf{I}_{13}$, the covariance of the residuals for this application can be written as

$$\text{Var}(\hat{\mathbf{E}}) = (\mathbf{V} - \mathbf{V}\mathbf{P}_T - \mathbf{P}_T\mathbf{V} + \mathbf{P}_T\mathbf{V}\mathbf{P}_T) \otimes \mathbf{U} \quad (15)$$

which is separable as claimed.

After computing the residuals, we decomposed the data into 61 pseudo-replicates of size $s = 13$ by $p^* = 2$. To estimate a possible type I error with this model, we generated 1000 multinormal samples with the same mean as fitted above for FACE plot 'A' in the year 2000. For the temporal *correlation* we used the model given in Mitchell and Gumpertz (2003), which was an AR(1) process with $\hat{\rho} = 0.299$. We modeled the spatial *covariance* with an unstructured covariance matrix. To estimate the critical values, we generated $N(0, 1)$ random variables as before, then estimated the mean with the covariates used above, then estimated $\hat{\mathbf{U}}$, $\hat{\mathbf{V}}$ and $\hat{\mathbf{S}}$ for $s = 13$, $p = 2$ and $r = 61$ based on the residuals, and then computed the LRT (8). We obtained the following critical values: 360.67, 371.33 and 395.00 for levels 0.10, 0.05 and 0.01, respectively. Based on the 1000 simulations, the estimated type I errors are 0.098, 0.050 and 0.010, respectively, indicating that the non-stationary mean and temporal correlation of the process have virtually no effect on the size of the test.

Since for the application in Mitchell and Gumpertz (2003) an AR(1) model was used, another possibility for testing the separability of the FACE data is to compute negative twice the difference of the log-likelihood values for a model with the same AR(1) correlation parameter for each location versus a model with a different AR(1) correlation parameter for each location. This addresses whether the particular separable model used is better than a particular alternative. We used the chi-square critical values for 12 degrees of freedom. We modeled the spatial covariance with an unstructured matrix. Since these critical values rely on the asymptotic distribution of the test statistic, we examined the type I error. Based on 5000 simulation runs, the estimated type I errors are approximately 0.212, 0.138, 0.035 for levels 0.10, 0.05 and 0.01. Hence the p -values from this test are too small.

We tested separability with both methods, and the results are shown in Table 7. Overall, the p -values for the test comparing the two particular models are larger than those obtained from the test involving pseudo-replication. Considering that the p -values with the former method are too small, this is especially remarkable. This indicates that deviations from separability include different classes of models than the one with different AR(1) parameters for each location.

Table 7. P -values for testing separability of the RiceFACE data

Ring	LRT1 ^a	LRT2 ^b
A	0.1133	0.3137
B	0.1222	0.2307
C	0.0010	0.1694
D	0.4755	0.7199

^a Test using pseudo-replicates.

^b Test comparing two AR(1) models.

Judging from Table 7, we see that, overall, the model with the same AR(1) parameter for each location is sufficient compared to that with a different correlation parameter for each location. The test using pseudo-replication shows more evidence of non-separability than the test based on different AR(1) parameters for each location, but most results are not significant at the 0.10 level. Overall, it appears that a separable space–time covariance is sufficient for the RiceFACE data.

5. CONCLUSIONS

There are very few formal tests available for testing separability of spatio–temporal covariances. However, it is possible to successfully test separability in many applications by adapting a likelihood ratio test for the multivariate repeated measures context. For moderately sized data sets, this test can be adapted using a bootstrap method. For data sets, such as the RiceFACE data, which are much richer in time than in space, it is possible to partition the data into approximately i.i.d. pseudo-replicates.

The type I errors for the test involving pseudo-replication depend on the dimensions of the pseudo-replicates. The larger the value of p^* (the ‘time’ dimension in a pseudo-replicate) and the smaller the value of s (spatial dimension), the lower the type I error. For a process with nine locations and 200 times with an AR(1) temporal correlation of 0.7, the type I errors are approximately equal to the nominal levels for $p^* = 2$. For more highly correlated data, the type I error can be reduced by creating ‘gaps’. For Theorem 1 to strictly apply, the mean is required to be constant across the pseudo-replicates. However, our simulation study showed that this had virtually no effect on the application involving the RiceFACE data.

The test in the multivariate repeated measures context requires no stationarity in space or time, but the pseudo-replication method requires *close* to stationarity in time since the pseudo-replicates are required to have the same temporal covariances. This is slightly more general than stationarity in time since the variances within a pseudo-replicate need not be equal (and hence all the time variances of the original process will not be equal). This reduction is not surprising since we have only one true replicate. Because of this restriction, one may consider assuming stationarity in time and fitting a Toeplitz or banded Toeplitz temporal structure. If we do not assume stationarity in space, then the non-separable covariance matrix for a zero mean process could be estimated with

$$\hat{C}(Z(s_i, t), Z(s_j, t + k)) = \sum_{l=1}^{p-k} (y_{i,t+l} * y_{j,t+l+k}) / (p - k) \quad (16)$$

For the separable case, the matrices \hat{U} and \hat{V} are interdependent, so one could estimate these similar to the algorithm given in Dutilleul (1999) by estimating \hat{V} with the usual Toeplitz (or banded Toeplitz) structure instead of an unstructured matrix. However, these computations are significantly more computationally intensive than those with the pseudo-replicates since the operations must be done on the *entire* covariance matrix (for the RiceFACE data, a 1586 by 1586 matrix for the non-separable case), not ‘chunks’ of it. This severely hampers the ability to generate the empirical distribution. Furthermore, there is no guarantee that the sample covariance matrix will be positive definite. For example, consider the following Toeplitz matrix:

$$\begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix} \quad (17)$$

The determinant of the matrix is $1 - 2\rho^2$. Thus, if $\rho^2 > 0.5$ the matrix will not be positive definite.

Testing separability in the pseudo-replicates gives more general results than comparing a class of models, and this test is adaptable to many situations. It is especially useful because the spatial structure need not be stationary or isotropic. However, the results of the test do not indicate what type of separable or non-separable model should be used. For future work, it is worth studying how much effect non-separability has on applications such as kriging.

ACKNOWLEDGEMENTS

We thank the associate editor and the referees for helpful comments. Our work on the RiceFACE project was supported by the CREST (Core Research for Evolutional Science and Technology) program of the Japan Science and Technology Corporation (JST). We would especially like to thank Kazuhiko Kobayashi, Masumi Okada, Mark Lieffering, Mayumi Yoshimoto, Hirofumi Nakamura and Han-Yong Kim.

REFERENCES

- Brown P, Diggle P, Lord M, Young P. 2001. Space–time calibration of radar rainfall data. *Applied Statistics* **50**: 221–241.
- Cressie N, Huang HC. 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* **94**: 1330–1340.
- Dutilleul P. 1999. The mle algorithm for the matrix normal distribution. *Journal of Statistical Computation and Simulation* **64**: 105–123.
- Fuentes M. 2005. Testing separability of spatial-temporal covariance functions. *Journal of Statistical Planning and Inference*, to appear.
- Gneiting T. 2002. Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association* **97**: 590–600.
- Huizenga H, Munck J, Waldorp L, Grasman R. 2002. Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model. *IEEE Transactions on Biomedical Engineering* **49**(6): 533–539.
- Kyriakidis PC, Journel AG. 1999. Geostatistical space–time models: a review. *Mathematical Geology* **31**: 651–684.
- Mitchell M, Gumpertz M. 2003. Spatial variability inside a free-air CO₂ system. *JABES* **8**(3): 310–327.
- Mitchell M, Genton MG, Gumpertz M. 2004. A likelihood ratio test for separability of covariances, revised. *Journal of Multivariate Analysis*.
- Okada M, Lieffering M, Nakamura H, Yoshimoto M, Kim HY, Kobayashi K. 2001. Free-air CO₂ enrichment (FACE) using pure CO₂ injection: system design. *New Phytologist* **150**: 251–260.
- Rodriguez-Iturbe I, Mejia JM. 1974. The design of networks in time and space. *Water Resources Research* **10**(4): 713–728.
- SAS Institute Inc. 2000. *SAS/IML User's Guide, Version 8*. SAS Publishing: Cary, NC, U.S.A.
- Shitan M, Brockwell P. 1995. An asymptotic test for separability of a spatial autoregressive model. *Communications in Statistics—Theory and Methods* **24**: 2027–2040.