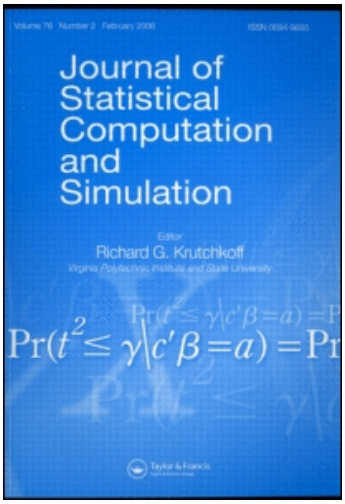

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713650378

Model-based tests for simplification of lattice processes

L. Scaccia[a]; R. J. Martin
[a] Dipartimento di Istituzioni Economiche e Finanziarie, Università di Macerata, MC, Italy

First published on: 07 December 2009

To cite this Article Scaccia, L. and Martin, R. J.(2011) 'Model-based tests for simplification of lattice processes', Journal of Statistical Computation and Simulation, 81: 1, 89 — 107, First published on: 07 December 2009 (iFirst)
To link to this Article: DOI: 10.1080/00949650903188452
URL: http://dx.doi.org/10.1080/00949650903188452

Journal of Statistical Computation and Simulation
Vol. 81, No. 1, January 2011, 89–107

Taylor & Francis
Taylor & Francis Group

Model-based tests for simplification of lattice processes

L. Scaccia[a]* and R.J. Martin[b]

[a]*Dipartimento di Istituzioni Economiche e Finanziarie, Università di Macerata, 62100 MC, Italy;*
[b]*Wirksworth, Derbyshire, DE4 4EB, UK*

(*Received 4 January 2009; final version received 16 July 2009*)

Separable processes represent a convenient class of models for data collected on a regular rectangular lattice. Three model-based tests, for testing separability and testing axial symmetry and separability together, are presented. These are shown to be much more powerful than existing model-free tests using the sample periodogram, provided the model assumptions hold. A simulation study also suggests that these tests are not very sensitive to small departures from the assumed process.

Keywords: autoregressive process; axial symmetry; doubly geometric process; lattice process; separability; spatial process

AMS Subject Classification: 62M30; 62F03; 62F05; 11K45

1. Introduction

Many of the problems of spatial modelling can be overcome by using separable processes. This subclass of spatial processes has several advantages, including rapid fitting and simple extensions of many techniques developed and successfully used in time series analysis. In particular, a major advantage of these processes is that the covariance matrix for realization on a rectangular lattice can be expressed as the Kronecker product of two smaller matrices that arise from two underlying one-dimensional processes, and hence its determinant and inverse are easily determinable. These separable models are, though, not always appropriate, and therefore formal testing of the hypothesis of separability is desirable.

However, few methods for testing separability have been proposed—(see the discussion in Scaccia and Martin [1]). For Gaussian data, Guo and Billard [2] suggested the Wald test comparing the fit of an AR(1)·AR(1) process and the more general unilateral autoregressive Pickard process, although this actually tests for axial symmetry and separability together. Recently, Genton and Koul [3] presented a test comparing these processes, based on the least absolute deviation residuals. This test seems to have higher power when the innovations have heavier tails than the Gaussian. Scaccia and Martin [4,5] investigated tests using the sample covariances, but found these tests

*Corresponding author. Email: scaccia@unimc.it

ISSN 0094-9655 print/ISSN 1563-5163 online

DOI: 10.1080/00949650903188452
http://www.informaworld.com

difficult to implement because the covariances are highly intercorrelated. Scaccia and Martin [1,4,5] developed tests based on the sample periodogram that can be applied to a large range of stationary processes, and test separability in two stages. First, axial symmetry (which is a necessary condition for a process to be separable) is tested with four possible statistics (two of which do not rely on normality of the data), and then separability is tested. Lu and Zimmerman [6] also proposed tests using the sample periodogram for axial or reflection symmetry (and complete symmetry). Their tests perform similarly to those in Scaccia and Martin [1].

Methods have also been proposed for testing separability between space and time in spatio-temporal processes. Fuentes [7] developed a simple two-factor analysis of variance procedure using spectral methods to test spatio-temporal separability that does not assume stationarity, but does need a large sample. Li *et al.* [8] provided a unified framework to assess the assumptions of full symmetry, spatial isotropy and separability, which is based on the asymptotic joint normality of sample space–time covariance estimators. However, their method is applicable to data collected on a random field with a fixed spatial domain and an increasing temporal domain, with replications over time being required to estimate the variance–covariance matrix of the multivariate normal distribution assumed for the sample space–time covariance estimators, and thus, is not adequate for purely spatial data.

In this article, we examine model-based tests for separability on its own and for axial symmetry and separability together. The proposal is to fit a particular model (non-symmetric or non-separable) to the data, then to restrict it to be axially symmetric or separable by imposing constraints on its parameters, and to test if the reduction is statistically justified. Any such reduction could be tested by the generalised likelihood ratio test (GLRT) or tests that are asymptotically equivalent, such as the Wald test or the Score test. Although the asymptotic theory for these test statistics is well established, their distributions for finite samples need to be investigated. For this, we propose Monte Carlo studies for a given finite sample size, model and parameter restriction. Here, we concentrate on the fit to the data of the separable AR(1)·AR(1) process and two non-separable processes that have the AR(1)·AR(1) as a special case. The reason for choosing the AR(1)·AR(1) process is its relevance in real applications. For example, Jain [9] used it in the study of image processing, Martin [10], Cullis and Gleeson [11], Basu and Reinsel [12], Genton and Koul [3] in agricultural trials, whereas Tjøstheim [13] used it in digital filtering. The AR(1)·AR(1) process has also been found to fit well remotely-sensed data [14], where the point spread function is usually taken as the product of the E–W and N–S scan components (see Section 6 for details).

The advantage of the tests proposed here, compared with those in Scaccia and Martin [1], is a larger power in detecting departures from separability when the specified model is correct. Our approach differs from that in Guo and Billard [2] as we also consider, in addition to the Wald test, the asymptotically equivalent GLRT and the Score test (which is simpler to implement) and compare their performances. We also compare different processes to the AR(1)·AR(1) so that we can test separability on its own. Finally, by analogy with the Durbin–Watson test for serial correlation in regression models, we carry out a brief sensitivity analysis, in order to see whether these model-based tests can be used as more general tests.

This article is organised as follows. Section 2 gives some basic notation and definitions. Section 3 describes the spatial models, with particular regard to the computation of their dispersion matrices, which are used in Section 4, where formal tests are developed. In Section 5, we describe a simulation study to evaluate the correspondence between the observed distribution of the tests and their theoretical distribution for different lattice sizes, under the null and the alternative hypotheses and for different processes. In Section 6, we apply the test statistics for axial symmetry and separability to some radar remote sensing data. Finally, conclusions are given in Section 7.

2. Notation and definitions

We assume that data occur on an n_1 by n_2 rectangular lattice, with rows indexed by $i_1 = 1, \ldots, n_1$ and columns by $i_2 = 1, \ldots, n_2$. Row and column lags are g_1 and g_2, with $g_j = -(n_j - 1), \ldots, 0, \ldots, (n_j - 1)$, for $j = 1, 2$. The sites are ordered lexicographically, so that (i_1, i_2) precedes $(i_1, i_2 + 1)$ for $i_2 < n_2$, and (i_1, n_2) precedes $(i_1 + 1, 1)$. Data can be considered as a realization of random variables Y_{i_1,i_2}. Assuming that the vector Y contains the Y_{i_1,i_2} in site order, we can consider, as a model:

$$E(Y) = A\vartheta \quad \text{and} \quad \text{var}(Y) = V\sigma^2,$$

where A is an $n \times p$ matrix (with $n = n_1 n_2$), ϑ is a p-vector of parameters, σ^2 is a scale parameter and the dispersion matrix $V = V(\delta)$ is an $n \times n$ positive-definite matrix depending on the q-vector of parameters δ. We assume here that the process is Gaussian. We also assume that the study region considered is homogeneous or that the data have had obvious trends or other effects removed to permit us to make the assumption of stationarity with $E(Y) = \mu 1_n$ so that $Y \sim N(\mu 1_n, V\sigma^2)$, where 1_n is a column vector of ones.

DEFINITION 2.1 *If $y \sim N(\mu 1_n, V\sigma^2)$, then the loglikelihood is*

$$l(\delta, \mu, \sigma^2; y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log \mid V \mid -\frac{1}{2\sigma^2}(y - \mu 1_n)' V^{-1}(y - \mu 1_n).$$

If a hat denotes the maximum likelihood estimate, $e = y - \hat{\mu} 1_n$ and $\hat{V} = V(\hat{\delta})$, then

$$\hat{\sigma}^2 = \frac{1}{n} e' \hat{V}^{-1} e, \qquad \hat{\mu} = (1_n' \hat{V}^{-1} 1_n)^{-1} 1_n' \hat{V}^{-1} y$$

and $\hat{\delta}$ is obtained by minimising $(e' V^{-1} e)^n |V|$.

DEFINITION 2.2 *For a second-order stationary spatial process in two dimensions, the covariance function (or covariogram) at lags g_1 and g_2 is*

$$C(g_1, g_2) = \text{cov}(Y_{i_1,i_2}, Y_{i_1+g_1,i_2+g_2}) = E[(Y_{i_1,i_2} - \mu)(Y_{i_1+g_1,i_2+g_2} - \mu)].$$

Then, the theoretical autocorrelation at lags g_1, g_2 is $\rho(g_1, g_2) = C(g_1, g_2)/C(0, 0)$, where $C(0, 0) = \text{var}(Y_{i_1,i_2}) = \sigma_Y^2$, and $C(g_1, g_2) = C(-g_1, -g_2)\ \forall g_1, g_2$.

DEFINITION 2.3 *A second-order stationary two-dimensional process is axial or reflection symmetric if the following equivalent conditions are satisfied:*

$$\rho(g_1, g_2) = \rho(g_1, -g_2) \quad \text{or} \quad C(g_1, g_2) = C(g_1, -g_2)\ \forall g_1, g_2.$$

If, in addition,

$$\rho(g_1, g_2) = \rho(g_1, 0)\rho(0, g_2) \text{ or, equivalently,} \quad C(g_1, g_2) \propto C(g_1, 0)C(0, g_2)\ \forall g_1, g_2$$

the process is also separable.

A separable process is obviously also an axially symmetric process.

3. Models compared

The test statistics used in this article are based on the comparison between the fits of the data to the separable AR(1)·AR(1) and to two non-separable processes that have the AR(1)·AR(1) as a special case. The test for separability uses the axially-symmetric $\text{CAR}(2)_{\text{SD}}$, and the test for axial symmetry and separability together uses the Pickard process. In this section, the three models used are briefly described, with particular regard to their inverse dispersion matrices V^{-1}, sometimes called the potential matrices, or to obtaining $e'V^{-1}e$, which is needed for the likelihood.

3.1. *The AR(1)·AR(1) processes*

The simplest non-trivial unilateral autoregressive separable process is the AR(1)·AR(1) process, also called the doubly-geometric process [15]. It can be specified as

$$Y_{i_1,i_2} = \alpha_1 Y_{i_1-1,i_2} + \alpha_2 Y_{i_1,i_2-1} - \alpha_1\alpha_2 Y_{i_1-1,i_2-1} + \epsilon_{i_1,i_2},$$

where the ϵ_{i_1,i_2} are assumed to be independently distributed as $N(0, \sigma_\epsilon^2)$. For stationarity, the parameters need to satisfy $|\alpha_1| < 1$ and $|\alpha_2| < 1$. The correlations are $\rho(g_1, g_2) = \alpha_1^{|g_1|}\alpha_2^{|g_2|}$, showing that the process is separable.

The dispersion matrix for a separable process on a rectangular lattice can be decomposed into two dispersion matrices arising from one-dimensional AR(1) processes. Taking $\sigma^2 = \sigma_\epsilon^2$ gives

$$V = V_1 \otimes V_2,$$

where V_j, for $j = 1, 2$, are dispersion matrices of the one-dimensional processes, taken here to have (i, k)th element equal to $\alpha_j^{|i-k|}/(1-\alpha_j^2)$, for $i = 1, \ldots, n_j$ and $k = 1, \ldots, n_j$. The inverse and the determinant of the dispersion matrix can, thus, be found from the known V_j^{-1} and $|V_j|$.

3.2. *The $\text{CAR}(2)_{\text{SD}}$*

The axially symmetric, non-separable model considered is a particular case of a second-order conditional autoregressive process, CAR(2) [16]. Here, the $\text{CAR}(2)_{\text{SD}}$, with equal diagonal parameters, will be considered. It can be expressed as

$$\begin{aligned} E[Y_{i_1,i_2}|Y_{l_1,l_2} : (l_1, l_2) \neq (i_1, i_2)] = {} & \beta_1(Y_{i_1-1,i_2} + Y_{i_1+1,i_2}) + \beta_2(Y_{i_1,i_2-1} + Y_{i_1,i_2+1}) \\ & + \beta_3(Y_{i_1-1,i_2-1} + Y_{i_1+1,i_2+1} + Y_{i_1-1,i_2+1} + Y_{i_1+1,i_2-1}) \end{aligned}$$

with constant conditional variance $\sigma^2 = \sigma_\kappa^2$, and stationarity requiring $|\beta_1 + \beta_2| + 2\beta_3 < 1/2$ and $|\beta_1 - \beta_2| - 2\beta_3 < 1/2$, which implies $|\beta_1| < 1/2$, $|\beta_2| < 1/2$, $|\beta_3| < 1/4$.

In spite of their usage in a variety of disciplines including image processing and economics, there are some difficulties with conditional models. In general, explicit spatial correlation properties are not easily determined, and also the elements of the dispersion matrix corresponding to sites on the boundary of a finite lattice are very difficult to obtain. For a stationary $\text{CAR}(2)_{\text{SD}}$ on an infinite lattice, or on a finite torus lattice,

$$V^{-1} = I_{n,n} - \sum_{i=1}^{3} \beta_i A_i \tag{1}$$

with $A_1 = T_{n_1,n_1} \otimes I_{n_2,n_2}$, $A_2 = I_{n_1,n_1} \otimes T_{n_2,n_2}$ and $A_3 = T_{n_1,n_1} \otimes T_{n_2,n_2}$, where I_{n_j,n_j}, for $j = 1, 2$, is the n_j by n_j identity matrix and T_{n_j,n_j} is a neighbour incidence matrix, with elements equal

to 1 if sites are row (for $j = 1$) or column (for $j = 2$) neighbours, and zero otherwise. In this article, we use Equation (1) on a finite planar lattice. These Dirichlet boundary conditions imply the model is, for n_1, n_2 not small, mildly non-stationary. The determinant of this dispersion matrix can be calculated from its known eigenvalues, $1/|V| = \prod_{i=1}^{n_1}\prod_{j=1}^{n_2}(1 - 2\beta_1 c_{1i} - 2\beta_2 c_{2j} - 2\beta_3 c_{1i}c_{2j})$, where $c_{rs} = \cos(2\pi s/(n_r + 1))$.

The AR(1)·AR(1) with parameters α_1, α_2 is the special case of a stationary CAR(2)$_{\mathrm{SD}}$ with $\beta_3 = -\beta_1\beta_2$, where $\beta_j = \alpha_j/(1 + \alpha_j^2)$, $j = 1, 2$.

3.3. *The Pickard process*

The non-separable, non-symmetric model considered in this article, and referred to as Pickard process, was considered by Pickard [17], Tory and Pickard [18] and Basu and Reinsel [12]. It is a first-order unilateral autoregressive model and can be specified as follows:

$$Y_{i_1,i_2} = \alpha_1 Y_{i_1-1,i_2} + \alpha_2 Y_{i_1,i_2-1} + \alpha_3 Y_{i_1-1,i_2-1} + \epsilon_{i_1,i_2}.$$

Stationarity requires $|\alpha_1 + \alpha_2| < 1 - \alpha_3$ and $|\alpha_1 - \alpha_2| < 1 + \alpha_3$. It is a special case of a CAR(2) with unequal diagonal parameters. Neither the correlations nor the inverse of the dispersion matrix are simple. Some ways to obtain V^{-1} are discussed in Martin [19] – note that there are some misprints in the specified equation given there. Taking $\sigma^2 = \sigma_\epsilon^2$, expressions for $e'V^{-1}e$ can be obtained directly from the exact simulation equations, and using symmetry, or from the conditional distributions [19].

Let $\Delta = \sigma_\epsilon^2/\sigma_Y^2$, with $\Delta^2 = D_1D_2D_3D_4$, where $D_1 = 1 - \alpha_1 - \alpha_2 - \alpha_3$, $D_2 = 1 + \alpha_1 + \alpha_2 - \alpha_3$, $D_3 = 1 + \alpha_1 - \alpha_2 + \alpha_3$ and $D_4 = 1 - \alpha_1 + \alpha_2 + \alpha_3$. It is shown in Basu and Reinsel [12] and Martin [19] that the determinant of the dispersion matrix for a Pickard process is given by

$$|V| = \Delta^{-(n_1+n_2-1)}[1 - \rho^2(1,0)]^{n_1-1}[1 - \rho^2(0,1)]^{n_2-1},$$

where $\rho(1,0) = (1 + \alpha_1^2 - \alpha_2^2 - \alpha_3^2 - \Delta)/[2(\alpha_1 + \alpha_2\alpha_3)]$ and $\rho(0,1) = (1 - \alpha_1^2 + \alpha_2^2 - \alpha_3^2 - \Delta)/[2(\alpha_2 + \alpha_1\alpha_3)]$. Note also that $\rho(1,1) = \alpha_1\rho(0,1) + \alpha_2\rho(1,0) + \alpha_3$ and $\rho(1,-1) = \rho(1,0)\rho(0,1)$.

The AR(1)·AR(1) is a special case of a Pickard process with $\alpha_3 = -\alpha_1\alpha_2$ and is the only axially symmetric (and separable) Pickard process [15].

4. Testing axial symmetry and separability

The tests here compare the fit of an AR(1)·AR(1) to that of a more general process – the Pickard for testing axial symmetry and separability together, and the CAR(2)$_{\mathrm{SD}}$ for testing separability on its own. In both cases, the null hypothesis expresses a nonlinear restriction on the parameter vector. Let η_δ denote $\delta_3 + \delta_1\delta_2$. Then, the hypotheses can be expressed as

$$\begin{cases} H_0 : \eta_\delta = 0, \\ H_1 : \eta_\delta \neq 0, \end{cases}$$

where, as appropriate, $\delta = (\delta_1\ \delta_2\ \delta_3)'$ is the parameter vector α of the Pickard model or β of the CAR(2)$_{\mathrm{SD}}$ model.

We consider the three asymptotically equivalent tests: the GLRT, the Wald test and the Score test [20, pp. 314–315].

The GRLT statistic to compare an AR(1)·AR(1) with a Pickard or a $\text{CAR}(2)_{\text{SD}}$ can be expressed as

$$W(\hat{\delta}, \tilde{\delta}) = -2[l(\tilde{\delta}, \tilde{\mu}, \tilde{\sigma}^2; y) - l(\hat{\delta}, \hat{\mu}, \hat{\sigma}^2; y)],$$

where $l(\delta, \mu, \sigma^2; y)$ denotes the loglikelihood, with a hat and a tilde on a parameter denoting, respectively, the parameter estimates under H_1 and H_0.

Since the asymptotic variance of $\hat{\eta}_\delta$ is $h(\delta)$, the Wald test statistic can be written as

$$W_e(\hat{\delta}) = \frac{\hat{\eta}_\delta^2}{h(\hat{\delta})},$$

where $h(\delta) = (\delta_2\ \delta_1\ 1) I^{-1}_{\delta|\sigma^2} (\delta_2\ \delta_1\ 1)'$ and $I_{\delta|\sigma^2}$ is the conditional Fisher information matrix on δ, given σ^2:

$$I_{\delta|\sigma^2} = I_{\delta\delta} - I_{\delta\sigma^2} I^{-1}_{\sigma^2\sigma^2} I_{\sigma^2\delta}$$

with $I_{\delta\delta} = -E\left[\frac{\partial^2}{\partial\delta\partial\delta'} l(\delta, \mu, \sigma^2; y)\right]$ and $I_{\delta\sigma^2}$, $I_{\sigma^2\sigma^2}$ and $I_{\sigma^2\delta}$ similarly defined.

The score function $U(\delta)$ is the derivative of the loglikelihood with respect to the parameter vector δ. If we write $\eta = (\delta_1\ \delta_2\ \eta_\delta)'$, then the Score test statistic is $U_{\eta_\delta}(\eta)^2 \text{var}(\hat{\eta}_\delta)$ evaluated at $\tilde{\eta}$ [20, p. 324]. Expressing $U(\eta)$ in terms of $U(\delta)$ shows that $U_{\eta_\delta}(\eta) = U_3(\delta)$ where $U_3(\delta)$ is the third component of $U(\delta)$. Hence, the Score test statistic becomes

$$W_u(\tilde{\delta}) = U_3(\tilde{\delta})^2 h(\tilde{\delta}).$$

These three test statistics are all asymptotically distributed, under the null hypothesis and under general conditions of regularity, as a χ^2_1 distribution. Under the alternative hypothesis, the distribution of the three tests is asymptotically a χ^2_1 with non-centrality parameter $\nu_\delta = \eta_\delta^2 / h(\delta)$. The three test statistics are all asymptotically consistent, which means that their power tends to 1 as $n \to \infty$, at least for local alternatives.

Notice that the GLRT test needs parameter estimates under both the restricted and the unrestricted models, whereas the Wald test requires estimates only under the unrestricted model and the Score test only under the restricted one.

4.1. *Evaluation and maximization of the loglikelihood*

For the evaluation and the maximization of the loglikelihood under the different models, $|V|$ and $e'V^{-1}e$ are required. These can be obtained from results in Sections 3.1, 3.2 and 3.3, respectively, for the AR(1)·AR(1), the $\text{CAR}(2)_{\text{SD}}$ and the Pickard model. The following expressions for $e'V^{-1}e$ can be obtained:

- For an AR(1)·AR(1) model

$$\begin{aligned} e'V^{-1}e = {} & a_1 + a_2 + (1+\alpha_1^2)a_3 + (1+\alpha_2^2)a_4 + (1+\alpha_1^2)(1+\alpha_2^2)a_5 - 2\alpha_1 a_6 \\ & - 2\alpha_2 a_7 - 2\alpha_1(1+\alpha_2^2)a_8 - 2\alpha_2(1+\alpha_1^2)a_9 + 2\alpha_1\alpha_2(a_{10} + a_{11}), \end{aligned}$$

where

$$a_1 = e_{1,1}^2 + e_{n_1,n_2}^2, \qquad a_7 = \sum_{j=1}^{n_2-1}(e_{1,j}e_{1,j+1} + e_{n_1,j}e_{n_1,j+1}),$$

$$a_2 = e_{1,n_2}^2 + e_{n_1,1}^2, \qquad a_8 = \sum_{i=1}^{n_1-1}\sum_{j=2}^{n_2-1} e_{i,j}e_{i+1,j},$$

$$a_3 = \sum_{i=2}^{n_1-1}(e_{i,1}^2 + e_{i,n_2}^2), \qquad a_9 = \sum_{i=2}^{n_1-1}\sum_{j=1}^{n_2-1} e_{i,j}e_{i,j+1},$$

$$a_4 = \sum_{j=2}^{n_2-1}(e_{1,j}^2 + e_{n_1,j}^2), \qquad a_{10} = \sum_{i=1}^{n_1-1}\sum_{j=1}^{n_2-1} e_{i,j}e_{i+1,j+1},$$

$$a_5 = \sum_{i=2}^{n_1-1}\sum_{j=2}^{n_2-1} e_{i,j}^2, \qquad a_{11} = \sum_{i=1}^{n_1-1}\sum_{j=1}^{n_2-1} e_{i+1,j}e_{i,j+1}.$$

$$a_6 = \sum_{i=1}^{n_1-1}(e_{i,1}e_{i+1,1} + e_{i,n_2}e_{i+1,n_2}),$$

- For a $\mathrm{CAR}(2)_{\mathrm{SD}}$ model

$$e'V^{-1}e = b_1 - 2\beta_1 b_2 - 2\beta_2 b_3 - 2\beta_3 b_4,$$

where $b_1 = a_1 + a_2 + a_3 + a_4 + a_5$, $b_2 = a_6 + a_8$, $b_3 = a_7 + a_9$ and $b_4 = a_{10} + a_{11}$.

- For a Pickard model

$$e'V^{-1}e = a_1 + \psi a_2 + (1+\alpha_1^2)a_3 + (1+\alpha_2^2)a_4 + (1+\alpha_1^2+\alpha_2^2+\alpha_3^2)a_5 - 2\alpha_1 a_6 - 2\alpha_2 a_7 - 2(\alpha_1 - \alpha_2\alpha_3)a_8 - 2(\alpha_2 - \alpha_1\alpha_3)a_9 - 2\alpha_3 a_{10} + 2\alpha_1\alpha_2 a_{11}, \qquad (2)$$

where

$$\psi = \alpha_2^2 + \frac{\Delta}{1-\rho^2(1,0)} = \alpha_1^2 + \frac{\Delta}{1-\rho^2(0,1)} = \frac{1+\alpha_1^2+\alpha_2^2-\alpha_3^2+\Delta}{2}.$$

Since we use a (mildly) non-stationary $\mathrm{CAR}(2)_{\mathrm{SD}}$ process, when implementing the tests we fit the non-stationary AR(1)·AR(1), which results by constraining η_β to 0. In this way, the restricted model is nested in the more general one.

As a final remark, we stress the importance of an adequate choice of the starting values for the maximization of the likelihood. Otherwise, convergence may fail. Different methods were considered to overcome this problem. These included, using as starting values, the true values and simple (least-squares) estimates.

4.2. *Computation of the score vector*

In order to implement the Score test, it is necessary to evaluate $U_3(\delta)$ under the null hypothesis. Consider the first derivative of the loglikelihood with respect to the ith component of the

parameter vector:

$$\frac{\partial l(\delta, \mu, \sigma^2; y)}{\partial \delta_i} = -\frac{1}{2}\frac{\partial \log |V|}{\partial \delta_i} - \frac{1}{2\sigma^2} e' \frac{\partial V^{-1}}{\partial \delta_i} e \quad \text{for } i = 1, 2, 3.$$

Since [20, p. 110]

$$E\left(\frac{\partial l(\delta, \mu, \sigma^2; y)}{\partial \delta_i}\right) = 0, \tag{3}$$

then

$$\frac{\partial l(\delta, \mu, \sigma^2; y)}{\partial \delta_i} = \frac{1}{2}\mathrm{tr}\left(\frac{\partial V^{-1}}{\partial \delta_i} V\right) - \frac{1}{2\sigma^2} e' \frac{\partial V^{-1}}{\partial \delta_i} e$$

and for a $\mathrm{CAR}(2)_{\mathrm{SD}}$ specified by Equation (1), the score vector is simply obtained from

$$\frac{\partial V^{-1}}{\partial \beta_i} = -A_i. \tag{4}$$

The determination of the score vector for a Pickard process is a bit more complicated. From Equation (3), it follows that

$$\frac{\partial l(\delta, \mu, \sigma^2; y)}{\partial \delta_i} = \frac{1}{2\sigma^2}\left[E\left(e' \frac{\partial V^{-1}}{\partial \delta_i} e\right) - e' \frac{\partial V^{-1}}{\partial \delta_i} e\right].$$

Now, recalling Equation (2), the derivatives of $e'V^{-1}e$ for a Pickard process are

$$\begin{aligned}
e' \frac{\partial V^{-1}}{\partial \alpha_1} e &= \psi_1 a_2 + 2\alpha_1(a_3 + a_5) - 2(a_6 + a_8) + 2\alpha_3 a_9 + 2\alpha_2 a_{11}, \\
e' \frac{\partial V^{-1}}{\partial \alpha_2} e &= \psi_2 a_2 + 2\alpha_2(a_4 + a_5) - 2(a_7 + a_9) + 2\alpha_3 a_8 + 2\alpha_1 a_{11}, \\
e' \frac{\partial V^{-1}}{\partial \alpha_3} e &= \psi_3 a_2 + 2\alpha_3 a_5 + 2\alpha_2 a_8 + 2\alpha_1 a_9 - 2a_{10},
\end{aligned}$$

where $\psi_i = \partial\psi/\partial\alpha_i = c_i\alpha_i + \Delta_i/2$, with $c_1 = c_2 = 1$, $c_3 = -1$ and $\Delta_i = \partial\Delta/\partial\alpha_i = (\Delta/2)\sum_{j=1}^{4}(1/D_j)(\partial D_j/\partial\alpha_i)$, with $\partial D_j/\partial\alpha_i$ equal to 1 or -1. Then, the expected value of $e'(\partial V^{-1}/\partial\alpha_i)e$ can be easily obtained using

$$\begin{aligned}
E(a_2) &= 2\sigma_Y^2, & E(a_7) &= 2\sigma_Y^2(n_2 - 1)\rho(0, 1), \\
E(a_3) &= 2\sigma_Y^2(n_1 - 2), & E(a_8) &= \sigma_Y^2(n_1 - 1)(n_2 - 2)\rho(1, 0), \\
E(a_4) &= 2\sigma_Y^2(n_2 - 2), & E(a_9) &= \sigma_Y^2(n_1 - 2)(n_2 - 1)\rho(0, 1), \\
E(a_5) &= \sigma_Y^2(n_1 - 2)(n_2 - 2), & E(a_{10}) &= \sigma_Y^2(n_1 - 1)(n_2 - 1)\rho(1, 1), \\
E(a_6) &= 2\sigma_Y^2(n_1 - 1)\rho(1, 0), & E(a_{11}) &= \sigma_Y^2(n_1 - 1)(n_2 - 1)\rho(1, -1).
\end{aligned}$$

Notice that the score vector for the Pickard process can be expressed in a compact form, which will be useful in the next section. Let $e'(\partial V^{-1}/\partial\delta_i)e = \sum_r f_{ir}a_{ir}$, where f_{ir} is a function of δ and a_{ir} is a function of e. Then,

$$\frac{\partial l(\delta, \mu, \sigma^2; y)}{\partial \delta_i} = -\frac{1}{2\sigma^2}\sum_r f_{ir}[a_{ir} - E(a_{ir})].$$

4.3. *Computation of the information matrix*

Both the Wald test and the Score test require the computation of the inverse of the conditional information matrix on δ given σ^2. Using Mardia and Marshall [21]:

$$I_{\delta|\sigma^2}(i,j) = \frac{1}{2}\mathrm{tr}\left(\frac{\partial V^{-1}}{\partial \delta_i} V \frac{\partial V^{-1}}{\partial \delta_j} V\right) - \frac{1}{2}\mathrm{tr}\left(\frac{\partial V^{-1}}{\partial \delta_i} V\right)\mathrm{tr}\left(\frac{\partial V^{-1}}{\partial \delta_j} V\right)\Big/ n, \tag{5}$$

where the first term is $I_{\delta\delta}$ and the second one is $-I_{\delta\sigma^2} I^{-1}_{\sigma^2\sigma^2} I_{\sigma^2\delta}$. Putting the appropriate V matrix and the appropriate derivatives in Equation (5), the information matrix for the different processes can be found.

The information matrix for a $\mathrm{CAR}(2)_{\mathrm{SD}}$ is simply obtained from Equation (5) using the derivatives of V^{-1} given in Equation (4).

For a Pickard process, the second term of Equation (5) can be calculated using Equation (3) and results of the previous section. The first term is

$$-E\left(\frac{\partial^2 l(\delta,\mu,\sigma_\epsilon^2;y)}{\partial\alpha_i\partial\alpha_j}\right) = E\left(\frac{1}{2\sigma_\epsilon^2}\left\{\sum_r \frac{\partial f_{ir}}{\partial\alpha_j}[a_{ir} - E(a_{ir})] + \sum_r f_{ir}\frac{\partial[a_{ir} - E(a_{ir})]}{\partial\alpha_j}\right\}\right). \tag{6}$$

Since $E[a_{ir} - E(a_{ir})] = 0$ and $\partial a_{ir}/\partial\alpha_j = 0$, expression (6) simply reduces to

$$-E\left(\frac{\partial^2 l(\delta,\mu,\sigma_\epsilon^2;y)}{\partial\alpha_i\partial\alpha_j}\right) = -\frac{1}{2\sigma_\epsilon^2}\sum_r f_{ir}\frac{\partial E(a_{ir})}{\partial\alpha_j}. \tag{7}$$

Notice that in Equation (7) we need derivatives, with respect to α_i, of $\rho(1,0)$, $\rho(0,1)$, $\rho(1,1)$ and $\rho(1,-1)$. Let $\rho_i(g_1,g_2) = \partial\rho(g_1,g_2)/\partial\alpha_i$, then, for $i = 1,2,3$,

$$\begin{aligned}
\rho_i(1,0) &= \frac{d_i\alpha_i - e_i\rho(1,0) - \Delta_i/2}{\alpha_1 + \alpha_2\alpha_3},\\
\rho_i(0,1) &= \frac{h_i\alpha_i - l_i\rho(0,1) - \Delta_i/2}{\alpha_2 + \alpha_1\alpha_3},\\
\rho_i(1,1) &= \alpha_1\rho_i(0,1) + \alpha_2\rho_i(1,0) + m_i,\\
\rho_i(1,-1) &= \rho_i(1,0)\rho(0,1) + \rho(1,0)\rho_i(0,1),
\end{aligned}$$

where $d_1 = 1$, $d_2 = d_3 = -1$, $e_1 = 1$, $e_2 = \alpha_3$, $e_3 = \alpha_2$, $h_1 = h_3 = -1$, $h_2 = 1$, $l_1 = \alpha_3$, $l_2 = 1$, $l_3 = \alpha_1$, $m_1 = \rho(0,1)$, $m_2 = \rho(1,0)$ and $m_3 = 1$. Using the expressions above immediately gives, for $i = 1,2,3$:

$$\begin{aligned}
I_{\delta\delta}(1,i) = \frac{1}{\Delta^2}\{&\Delta_i[\psi_1 + n_2(n_1-2)\alpha_1] - n_2(n_1-1)[\rho(1,0)\Delta_i - \rho_i(1,0)\Delta]\\
&+ (n_1-2)(n_2-1)\alpha_3[\rho(0,1)\Delta_i - \rho_i(0,1)\Delta]\\
&+ (n_1-1)(n_2-1)\alpha_2[\rho(1,-1)\Delta_i - \rho_i(1,-1)\Delta]\},
\end{aligned}$$

$$
\begin{aligned}
I_{\delta\delta}(2,i) = \frac{1}{\Delta^2}\{&\Delta_i[\psi_2 + n_1(n_2-2)\alpha_2] - n_1(n_2-1)[\rho(0,1)\Delta_i - \rho_i(0,1)\Delta] \\
&+ (n_1-1)(n_2-2)\alpha_3[\rho(1,0)\Delta_i - \rho_i(1,0)\Delta] \\
&+ (n_1-1)(n_2-1)\alpha_1[\rho(1,-1)\Delta_i - \rho_i(1,-1)\Delta]\}, \\
I_{\delta\delta}(3,i) = \frac{1}{\Delta^2}\{&\Delta_i[\psi_3 + (n_1-2)(n_2-2)\alpha_3] - (n_1-1)(n_2-1)[\rho(1,1)\Delta_i - \rho_i(1,1)\Delta] \\
&+ (n_1-1)(n_2-2)\alpha_2[\rho(1,0)\Delta_i - \rho_i(1,0)\Delta] \\
&+ (n_1-2)(n_2-1)\alpha_1[\rho(0,1)\Delta_i - \rho_i(0,1)\Delta]\}.
\end{aligned}
$$

5. Simulation study

Simulations were used to evaluate how large a lattice needs to be for the null asymptotic χ^2_1 distribution of the test statistics to be reasonable. Lattices of increasing dimension were considered, starting from an 11×11 lattice, and the comparison between the empirical and the asymptotic distribution was made on the basis of their means and standard deviations, as well as using the X^2 test for goodness-of-fit. However, principally we considered the probabilities of exceeding the 95th and 99th percentile of the theoretical distribution, as these are of most concern for using the tests.

Each setting was simulated 1000 times to estimate the distribution of the test statistics. The vector of observations y was simulated as $y = T\epsilon$, where ϵ is a random vector of n independent $N(0,1)$ observations, and T is a matrix such that $V = TT'$. The random vector ϵ was generated using the computer software Matlab [22], and randomly permuting the result to guard against possible serial correlation. The matrix T can be chosen in any convenient way. We usually used the Cholesky decomposition that specifies T as a lower triangular matrix, and if the observations are ordered lexicographically, the method corresponds to a finite unilateral moving average representation of the process. For the Pickard process, a slightly different method, corresponding to a finite unilateral autoregressive representation, was used [19].

Under the null hypothesis, five different separable $\text{CAR}(2)_{\text{SD}}$ (i.e. AR(1)·AR(1), indicated as A1–A5 in the appendix) and five different separable Pickard (i.e. AR(1)·AR(1), A6–A10 in the appendix) were simulated, respectively, for the test statistics for separability and those for axial symmetry and separability. Moreover, to evaluate the sensitivity of the test statistics proposed to misspecification of the model, five $\text{ARMA}(p_1, q_1)\cdot \text{ARMA}(p_2, q_2)$ processes (AM1–AM5) were also simulated. Details on these models can be found in Martin [19].

For comparing the power of the test statistics developed in this article with those in Scaccia and Martin [1], the same models used in there are simulated here, under the alternative hypotheses of no separability (C1–C5) and no axial symmetry (P1–P5).

Notice that if the estimated probability of rejection, in %, is p, then its estimated standard error is $\sqrt{p(100-p)/1000}\%$. Thus, the estimated standard errors corresponding to a 5% and a 1% rejection probability are, respectively, approximately equal to 0.7% and 0.3%.

5.1. *Test statistics for separability*

The estimated probabilities of rejecting the null hypothesis of separability when true for each of the three test statistics are given in Table 1. For an 11×11 lattice, these are quite close to the nominal levels and generally the 95% confidence intervals for the rejection probabilities include the 5% and 1% nominal levels. The only exceptions are for process A5, whose parameters

Table 1. Observed rejection probabilities of the GLRT, Wald and Score tests under the null hypothesis of separability.

		11×11 lattice		15×15 lattice	
Process	Test	5% level	1% level	5% level	1% level
A1	GLRT	4.2	0.7	5.6	1.4
	Wald	6.3	2.1	7.1	2.0
	Score	4.2	0.7	5.6	1.4
A2	GLRT	3.6	0.8	4.5	0.9
	Wald	5.6	1.5	5.8	1.2
	Score	3.5	0.8	4.4	0.9
A3	GLRT	4.3	0.3	3.5	0.6
	Wald	5.0	1.1	3.9	0.9
	Score	4.2	0.3	3.4	0.6
A4	GLRT	3.9	0.5	3.4	0.9
	Wald	4.2	0.4	3.0	0.9
	Score	3.8	0.2	3.2	1.0
A5	GLRT	2.9	0.0	3.3	1.0
	Wald	1.3	0.0	3.5	0.6
	Score	1.0	0.0	2.3	0.2

Note: Processes simulated: separable $\text{CAR}(2)_{\text{SD}}$.

are closer to the boundary. However, this feature is not particularly worrying, since the rejection probabilities are actually smaller than they should be, leading to more conservative tests. Moreover, the approximation to the null distribution seems to improve rapidly as the lattice size increases. For the Wald test, it seems that the rejection probabilities are generally higher than those of the other two tests and decrease as the parameters get closer to the boundary.

Table 2 shows the observed power (π_o) of the three different test statistics, under five non-separable $\text{CAR}(2)_{\text{SD}}$ processes. The knowledge of the asymptotic distribution of the tests under

Table 2. Simulated power of the tests for separability.

		11×11 lattice						15×15 lattice					
				5% level		1% level				5% level		1% level	
Process	Test	$\nu_\beta^{(e)}$	$\nu_\beta^{(o)}$	π_e%	π_o%	π_e%	π_o%	$\nu_\beta^{(e)}$	$\nu_\beta^{(o)}$	π_e%	π_o%	π_e%	π_o%
C1	GLRT	1887.5	38.5	100.0	99.9	100.0	99.7	4640.6	89.7	100.0	100.0	100.0	100.0
	Wald		344.3		99.9		99.8		1225.0		100.0		100.0
	Score		30.6		99.9		99.8		64.0		100.0		100.0
C2	GLRT	20.9	9.5	99.6	86.1	97.7	67.6	41.7	24.8	100.0	99.4	100.0	98.3
	Wald		11.9		87.6		73.0		28.1		99.5		98.6
	Score		11.0		88.0		71.3		28.9		99.5		98.7
C3	GLRT	8.3	3.6	82.2	52.0	62.0	24.2	16.4	11.5	98.2	91.0	93.0	78.8
	Wald		3.3		53.8		19.1		9.5		90.9		77.1
	Score		4.8		57.7		33.3		14.9		92.6		83.9
C4	GLRT	6.3	2.5	70.9	37.8	47.4	13.5	12.6	8.3	94.4	82.7	83.5	61.9
	Wald		2.3		40.5		6.3		6.8		82.7		62.1
	Score		3.6		45.1		24.7		11.5		86.0		70.8
C5	GLRT	0.6	0.1	12.1	5.7	3.6	1.3	1.3	0.7	20.7	12.4	7.6	3.9
	Wald		0.1		6.1		1.3		0.7		11.8		3.5
	Score		0.1		6.1		1.6		0.8		14.2		4.5

Note: Processes simulated: nonseparable $\text{CAR}(2)_{\text{SD}}$.

the alternative hypothesis has been exploited to calculate also the expected power (referred to as π_e) of the tests. The expected power at the 5% and 1% level of the tests was obtained as the probability that a χ^2_1 with non-centrality parameter ν_β exceeds, respectively, the 95th and the 99th percentile of the χ^2_1 distribution. In Table 2, the observed and expected values of ν_β, indicated, respectively, as $\nu_\beta^{(o)}$ and $\nu_\beta^{(e)}$, are also given. The value of $\nu_\beta^{(e)}$ was calculated from the true information matrix, $I(\beta)$. Note that when the parameters are close to the boundary, their estimates must be inside the boundary so that $\nu_\beta^{(o)}$ can have a large negative bias for $\nu_\beta^{(e)}$.

The power of the three tests is clearly linked to the parameters of the process and increases as ν_β does (with $|\eta_\beta|$ being a large component of ν_β). Notice that the processes in Table 2 are in decreasing order with respect to $|\eta_\beta|$, with process C5 being very close to separability. In this case, the power of the tests is nearly the same as the significance level. For the three tests, the observed power seems to be smaller than expected for small values of $|\eta_\beta|$. This is probably because the lattices are too small for a good approximation to the expected non-central χ^2. The discrepancy between the expected and the observed power seems to decrease as the lattice size increases. It also seems that the power is generally slightly higher for the Wald and the Score test than for the GLRT.

Comparing Table 2 with the power of the test for separability based on the sample periodogram, given in Table 2 of Scaccia and Martin [1], it can be seen that the GLRT, the Wald test and the Score test are clearly much more powerful. For example, for process C1, the power of the test based on the periodogram (using the maximum range of $[0,\ 10\pi/11]$ for the frequencies) was 27.4% and 10.3% at the 5% and the 1% level, respectively. However, these model-based tests clearly have the drawback that they depend on the specification of the model, whereas the test based on the sample periodogram only requires stationarity of the process.

Table 3 illustrates the effects of misspecification of the model on the level of the three tests, under the null hypothesis of separability. The observed rejection probabilities of the Wald test are all low or very low. Those of the GLRT and Score test are high for some processes and low for others. The only reasonable results seem to be for the GLRT with AM5 (both lattice sizes), and

Table 3. Effects of model misspecification on the level of the tests for separability, when the null hypothesis is true.

		11×11 lattice		15×15 lattice	
Process	Test	5% level	1% level	5% level	1% level
AM1	GLRT	7.3	1.6	14.4	4.9
	Wald	0.5	0.0	3.7	0.4
	Score	7.8	0.9	11.4	2.4
AM2	GLRT	0.6	0.0	2.2	0.1
	Wald	0.0	0.0	0.1	0.0
	Score	4.3	0.8	3.7	0.8
AM3	GLRT	0.7	0.0	0.4	0.0
	Wald	0.2	0.0	0.3	0.0
	Score	0.3	0.0	0.3	0.0
AM4	GLRT	8.2	2.2	12.1	4.1
	Wald	0.9	0.0	3.5	0.0
	Score	1.9	0.2	4.1	0.2
AM5	GLRT	4.0	0.5	4.2	0.5
	Wald	0.4	0.0	0.4	0.0
	Score	0.8	0.0	0.1	0.0

Note: Processes simulated: ARMA(p_1, q_1)·ARMA(p_2, q_2).

AM1, AM4 (11 × 11 lattice); and the Score test with AM2 (both lattice sizes) and AM1 (11 × 11 lattice).

5.2. *Test statistics for axial symmetry and separability*

Table 4 shows the observed probabilities of rejecting the null hypothesis of axial symmetry and separability for each of the three tests. Most values seem to be reasonable, although there is a suggestion (particularly for A10) that those for the Wald test are a little high, and those for the Score test a little low.

Table 5 shows the observed power of the three different test statistics, together with the expected power and the observed and expected non-centrality parameter, under five non-symmetric, non-separable Pickard processes. The power of the three tests increases with ν_α, with the observed power being a bit smaller, in some cases, than the expected one. This discrepancy seems, however, to decrease as the lattice size increases. It also seems that the power is generally slightly higher for the Wald test and slightly lower for the Score test (although this may be related to the sizes of the tests having the same effect – see Table 4).

Considering Tables 2 and 5, it is evident that the GLRT, the Wald test and the Score test can be much more powerful for testing an AR(1)·AR(1) versus a Pickard, than for testing an AR(1)·AR(1) versus a $\text{CAR(2)}_{\text{SD}}$. This is due to the fact that in the first case we are testing for both axial symmetry and separability at the same time, while in the second case we are only testing for separability.

Effects of model misspecification on the level of the tests, when the null hypothesis is true, can be observed in Table 6. On 11×11 lattices, the rejection probabilities are higher than they should be for most of the processes. The performance of the GLRT statistic seems, though, promising: its observed rejection probabilities are generally closer to the expected ones than those of the other test statistics. In particular, the usage of the GLRT seems perfectly reasonable for two out of the five ARMA(p_1, q_1)·ARMA(p_2, q_2) processes considered.

Table 4. Observed rejection probabilities of the GLRT, Wald and Score tests under the null hypothesis of axial symmetry and separability.

		11 × 11 lattice		15 × 15 lattice	
Process	Test	5% level	1% level	5% level	1% level
A6	GLRT	4.1	0.4	4.4	1.4
	Wald	4.3	0.7	4.3	1.4
	Score	4.1	0.3	4.1	1.4
A7	GLRT	5.4	0.6	5.1	0.9
	Wald	6.4	0.5	5.8	1.2
	Score	4.9	0.6	4.1	0.7
A8	GLRT	5.0	0.5	5.1	0.9
	Wald	6.1	1.1	5.4	0.8
	Score	4.6	0.3	5.3	0.5
A9	GLRT	5.4	1.0	5.7	1.0
	Wald	6.2	1.0	5.9	1.5
	Score	3.2	0.3	4.4	0.6
A10	GLRT	6.7	1.7	7.1	1.6
	Wald	8.8	2.1	7.8	2.0
	Score	1.5	0.1	2.4	0.3

Note: Processes simulated: separable Pickard.

Table 5. Simulated power of the tests for axial symmetry and separability.

		11×11 lattice						15×15 lattice					
				5% level		1% level				5% level		1% level	
Process	Test	$\nu_\alpha^{(e)}$	$\nu_\alpha^{(o)}$	$\pi_e\%$	$\pi_o\%$	$\pi_e\%$	$\pi_o\%$	$\nu_\alpha^{(e)}$	$\nu_\alpha^{(o)}$	$\pi_e\%$	$\pi_o\%$	$\pi_e\%$	$\pi_o\%$
P1	GLRT	91.2	53.3	100.0	100.0	100.0	100.0	171.0	106.9	100.0	100.0	100.0	100.0
	Wald		85.6		100.0		100.0		162.0		100.0		100.0
	Score		39.3		100.0		100.0		79.8		100.0		100.0
P2	GLRT	64.1	45.7	100.0	100.0	100.0	100.0	122.6	87.3	100.0	100.0	100.0	100.0
	Wald		65.9		100.0		100.0		125.0		100.0		100.0
	Score		13.6		99.8		95.9		24.2		100.0		100.0
P3	GLRT	35.0	24.7	100.0	99.9	100.0	98.1	67.4	52.0	100.0	100.0	100.0	100.0
	Wald		28.9		99.9		98.8		59.2		100.0		100.0
	Score		21.3		99.9		97.5		46.8		100.0		100.0
P4	GLRT	15.4	11.1	97.5	88.9	91.1	75.4	29.5	23.7	100.0	99.4	99.8	98.2
	Wald		12.6		90.6		77.7		26.1		99.6		98.4
	Score		9.7		87.0		71.4		21.2		99.4		97.7
P5	GLRT	0.2	0.0	7.0	4.7	1.7	0.8	0.3	0.2	8.8	7.8	2.3	1.9
	Wald		0.0		5.5		1.2		0.3		8.6		2.1
	Score		0.0		4.0		0.6		0.2		7.5		1.9

Note: Processes simulated: non-symmetric, non-separable Pickard.

Table 6. Effects of model misspecification on the level of the tests for axial symmetry and separability, when the null hypothesis is true.

		11×11 lattice		15×15 lattice	
Process	Test	5% level	1% level	5% level	1% level
AM1	GLRT	10.9	3.2	10.9	3.0
	Wald	12.9	4.4	12.4	4.1
	Score	5.1	0.6	7.2	0.7
AM2	GLRT	5.8	1.1	4.0	1.0
	Wald	6.9	1.3	4.5	0.9
	Score	3.8	0.3	3.1	0.5
AM3	GLRT	7.6	1.8	7.0	1.5
	Wald	8.0	2.0	7.1	1.8
	Score	7.6	1.5	7.0	1.3
AM4	GLRT	8.3	1.8	10.1	3.1
	Wald	9.6	3.5	12.3	4.4
	Score	2.5	0.1	5.0	0.5
AM5	GLRT	5.4	1.5	7.0	1.4
	Wald	6.4	1.9	7.5	2.4
	Score	3.0	0.4	4.2	0.9

Note: Processes simulated: ARMA(p_1, q_1)·ARMA(p_2, q_2).

6. An application

In this section, the tests for separability and for axial symmetry and separability together, investigated in Section 5, are implemented on a real data set from radar remote sensing.

The data were collected during the AgriSAR 86 remote sensing campaign organized by the Joint Research Centre of the European Communities, Ispra. During this campaign, several sites over

Europe were imaged from an airplane using a particular Synthetic Aperture Radar (SAR) with the aim of providing a calibrated multi-temporal and multi-polarization data set over agricultural test sites, in order to investigate the backscatter behaviour of growing crops. The SAR data were augmented by ground data collected in order to understand the phenomena observed in the SAR data. For this reason, the data set has been analysed by several researchers with different objectives [23–27]. Here, a small amount of the data from the Feltwell (Cambridgeshire, UK) test site are considered with the aim of verifying the separability of the *point spread function*, which is briefly described below.

The interaction of the radar signal with the Earth's surface causes a change in both the amplitude and the phase of the signal itself. The amplitude and the phase of the backscattered signal represent the specific reflectivity of the point on the ground which returned the signal. The output, $y(r)$, from the SAR system summarizes the information from a large number of scatterers within an area surrounding the point r and is given by a convolution of the reflectivity with the SAR system function, called the point spread function. When a region is imaged, a sample of the $y(r)$ is taken and this sample is denoted by $\{y_{ij} : i = 1, \ldots, n_1;\ j = 1, \ldots, n_2\}$. For each pixel (i, j), two measurements are produced, which can be regarded as the real and imaginary parts of the complex number y_{ij}. To produce an image, either the amplitudes $|y_{ij}|$ or the intensities $|y_{ij}|^2$ are calculated and transformed into a grey scale (integers 0 to $2^k - 1$ for some k).

The point spread function causes the SAR data to be correlated even when the imaged surface is homogeneous and no texture is present. Thus, it leads to obvious problems when analysing the data. In some cases, researchers are interested in studying the texture of a given region of the image. However the correlation structure of the data will be confounded by the presence of the correlation structure due to the point spread function. In other cases, the surface imaged will be divided into regions considered small enough to be homogeneous, and the interest will be in discriminating between different features (e.g. crops, land usage), on the basis, for example, of the mean reflectivity of each of the regions. The dependence of the data in each region, due to the point spread function, obviously affects any tests to detect the existence of such a difference in the mean reflectivity. Thus, removal of, or at least allowance for, the effect of the point spread function is one of the fundamental problems when analysing the SAR remote sensing data.

The idealized SAR theory states that the point spread function can be modelled as a separable spatial process. Here, we consider a portion from each of eight different fields [28], small enough (20×10 pixels) to be considered homogeneous, so that the observable correlation structure can be assumed to be due to the point spread function, and test whether the SAR theory is valid. The data, consisting of the intensities $|y_{ij}|^2$, which varied from 0 to 255, were transformed taking their cubic root to reduce their non-normality. Then, a graphical analysis of the sample autocorrelation and partial autocorrelation structure of the transformed data was performed. The analysis, not reported here [28], revealed that for all the fields considered, the one-dimensional autocorrelations and partial autocorrelations were compatible with those of an AR(1) process. Moreover, the two-dimensional autocorrelations at lags $(1, 1)$ and $(1, -1)$ looked very similar in accordance with the hypothesis of axial symmetry. Thus, an AR(1)·AR(1) model was fitted to the data and compared with a Pickard model to test for axial symmetry and separability together. Since the sample autocorrelation structure suggested that axial symmetry could be reasonable, the AR(1)·AR(1) model was also compared with a $\mathrm{CAR}(2)_{\mathrm{SD}}$ to test just for separability. Tables 7 and 8 show how the fit of an AR(1)·AR(1) to the SAR data compares with that of a Pickard and $\mathrm{CAR}(2)_{\mathrm{SD}}$, respectively. The restricted model seems to fit the data well in comparison with the unrestricted models: in both cases, the p-values associated with all the asymptotically equivalent test statistics used are larger than 10%, providing no evidence against the null hypotheses. Thus, the data turned out to be consistent with both axial symmetry and separability, as suggested by the SAR theory.

Table 7. Tests for axial symmetry and separability.

Field	GLRT	p-value	Wald test	p-value	Score test	p-value
217	0.323	0.570	0.335	0.563	0.311	0.577
240	2.307	0.129	2.617	0.106	1.997	0.158
246	0.737	0.391	0.789	0.375	0.700	0.403
266	0.836	0.360	0.833	0.361	0.851	0.356
120	0.004	0.948	0.004	0.948	0.004	0.948
131	0.112	0.738	0.114	0.735	0.109	0.741
98	0.093	0.761	0.083	0.774	0.104	0.748
155	0.011	0.916	0.012	0.913	0.010	0.919

Note: Processes fitted: AR(1)·AR(1) versus Pickard.

Table 8. Tests for separability.

Field	GLRT	p-value	Wald test	p-value	Score test	p-value
217	0.655	0.418	0.647	0.421	0.662	0.416
240	2.450	0.118	2.282	0.131	2.373	0.123
246	0.010	0.920	0.010	0.920	0.010	0.920
266	1.459	0.227	1.428	0.232	1.503	0.220
120	0.028	0.867	0.029	0.865	0.028	0.867
131	0.424	0.515	0.426	0.514	0.425	0.514
98	0.609	0.435	0.612	0.434	0.625	0.429
155	0.484	0.487	0.485	0.486	0.491	0.483

Note: Processes fitted: AR(1)·AR(1) versus $\text{CAR}(2)_{\text{SD}}$.

7. Discussion

Model-based tests for testing the hypothesis of separability and that of axial symmetry and separability have been proposed. The test for separability is based on the comparison between the fit of a $\text{CAR}(2)_{\text{SD}}$ and the fit of an AR(1)·AR(1), while the test for axial symmetry and separability compares the fit of a Pickard and that of an AR(1)·AR(1). The GLRT, the Wald and the Score test statistics were used for the comparison. Although the three test statistics are asymptotically equivalent, their properties differ for finite samples, and so it is necessary to investigate which one is best in the present situation.

The Score test uses estimates under the null hypothesis and thus the loglikelihood is maximized with respect to just two parameters. For the AR(1)·AR(1), the estimates are easily found by iteratively solving two cubic equations [15]. The Wald test uses instead estimates under the alternative hypothesis, while the GLRT uses both estimates and is, therefore, computationally more intensive than the other two tests. On the other hand, the Wald and the Score tests need the second derivatives of the loglikelihood, which are not easily found. When computing the three test statistics to compare the fit of a Pickard with that of an AR(1)·AR(1), we observed that the time required to compute the Wald test and the GLRT is, respectively, approximately four and five times than that required to compute the Score test. On the other hand, computing the Fisher information matrix and the score vector turned out to be more computationally expensive for a $\text{CAR}(2)_{\text{SD}}$ than for a Pickard process. Thus, when computing the three test statistics to compare the fit of a $\text{CAR}(2)_{\text{SD}}$ with that of an AR(1)·AR(1), the GLRT required the least computing time, at least on an 11×11 lattice, with the Wald test and the Score test being, respectively, approximately more than three and more than six times slower to compute. Increasing the lattice size, however, makes the time demanded for the maximization of the likelihood increase faster than that for computing the Fisher information matrix and the score vector: on a 15×15 lattice, we observed

much smaller differences between the computing time for the three test statistics. On larger lattice sizes, the Score test is thus expected to be the least computationally demanding also when testing separability on its own.

Apart from a computational point of view, the simulations show that, generally, the GLRT is preferable as it converges faster to its asymptotic distribution than the other two tests. We therefore suggest using this test, rather than the other two, especially when the lattice size is not large.

The advantage of the tests proposed here, compared with those in Scaccia and Martin [1], is that they have a much larger power for detecting departures from separability. The drawback is that they depend on a specified model. However, a brief sensitivity analysis, using separable processes that are different from the specified AR(1)·AR(1), suggests that it may be reasonable to use the GLRT, provided that the true process is not too different from the one fitted under the alternative hypothesis.

Finally, notice that the GLRT, Wald and Score tests could be extended to test axial symmetry and separability for processes other than those considered here, provided these hypotheses can be expressed as restrictions on the parameters of the process. Analogously, they can be extended to test for simplification of processes defined on a three-dimensional space or space–time processes.

Acknowledgements

We are grateful for the helpful comments of an anonymous referee.

References

[1] L. Scaccia and R.J. Martin, *Testing axial symmetry and separability of lattice processes*, J. Statist. Plann. Inf. 131 (2005), pp. 19–39.
[2] J.H. Guo and L. Billard, *Some inference results for causal autoregressive processes on a plane*, J. Time Ser. Anal. 19 (1998), pp. 681–691.
[3] M.G. Genton and H.L. Koul, *Minimum distance inference in unilateral autoregressive lattice processes*, Stat. Sin. 18 (2008), pp. 617–631.
[4] L. Scaccia and R.J. Martin, *Testing axial symmetry and separability of lattice processes*, Res. Rep. no. 530/02, School of Mathematics and Statistics, Sheffield University, Sheffield, UK, 2002.
[5] L. Scaccia and R.J. Martin, *Testing for simplification in spatial models*, in COMPSTAT 2002, Proceedings in Computational Statistics, W. Hardle and B. Ronz, eds., Physica-Verlag, Heidelberg, 2002, pp. 581–586.
[6] N. Lu and D.L. Zimmerman, *Testing for directional symmetry in spatial dependence using the periodogram*, J. Statist. Plann. Inf. 129 (2005), pp. 369–385.
[7] M. Fuentes, *Testing for separability of spatial–temporal covariance functions*, J. Statist. Plann. Inf. 136 (2006), pp. 447–466.
[8] B. Li, M.G. Genton, and M. Sherman, *A nonparametric assessment of properties of space-time covariance functions*, J. Am. Stat. Assoc. 102 (2007), pp. 736–744.
[9] A.K. Jain, *Advances in mathematical models for image processing*, Proc. IEEE 69 (1981), pp. 502–528.
[10] R.J. Martin, *The use of time-series models and methods in the analysis of agricultural field trials*, Commun. Statist. Theory Meth. 19 (1990), pp. 55–81.
[11] B.R. Cullis and A.C. Gleeson, *Spatial analysis of field experiments: an extension to two dimensions*, Biometrics 47 (1991), pp. 1449–1460.
[12] S. Basu and G.C. Reinsel, *Properties of the spatial unilateral first-order ARMA model*, Adv. Appl. Probab. 25(3) (1993), pp. 631–648.
[13] D. Tjøstheim, *Autoregressive modelling and spectral analysis of array data in the plane*, IEEE Trans. Geosci. Remote Sens. 19 (1981), pp. 15–24.
[14] N.A. Campbell, *Remote sensing in practice*, Paper presented at the Open Forum on Statistics and Pattern Recognition, Edinburgh, July, 1985.
[15] R.J. Martin, *A subclass of lattice processes applied to a problem in planar sampling*, Biometrika 66 (1979), pp. 209–217.
[16] J. Besag, *Spatial interaction and the statistical analysis of lattice systems (with discussion)*, J. Roy. Statist. Soc. B 36 (1974), pp. 192–236.
[17] D.K. Pickard, *Unilateral Markov fields*, Adv. Appl. Probab. 12 (1980), pp. 655–671.
[18] E.M. Tory and D.K. Pickard, *Unilateral Gaussian fields*, Adv. Appl. Probab. 24 (1992), pp. 95–112.
[19] R.J. Martin, *Some results on unilateral ARMA lattice processes*, J. Statist. Plann. Inf. 50 (1996), pp. 395–411.
[20] D.R. Cox and D.V. Hinkley, *Theoretical statistics*, Chapman and Hall, London, 1974.

[21] K.V. Mardia and R.J. Marshall, *Maximum likelihood estimation of models for residual covariance in spatial regression*, Biometrika 71 (1984), pp. 135–146.
[22] MathWorks, *Matlab version 6.0. Release 12*, Natick, MA, 2000.
[23] T. Le Toan, H. Laur, E. Mougin and A. Lopes, *Multitemporal and dual polarisation observations of agricultural vegetation covers By X-Band SAR images*, IEEE Trans. Geosci. Remote Sens. 27 (1989), pp. 709–717.
[24] A. Fiumara, N. Pierdicca, and M. Ricottilli, *Crops radar responses analysis based on Agrisar '86 data*, in *Proceedings of IGRASS '88 Symposium*, Vol. 2, Edinburgh, UK, 13–16 September 1988, pp. 1131–1132.
[25] G.M. Foody, P.J. Curran, G.B. Groom and D.C. Munro, *Crop classification with multi-temporal X-Band SAR data*, in *Proceedings of IGRASS '88 Symposium*, Vol. 1, Edinburgh, UK, 13–16 September 1988, pp. 217–220.
[26] C.C.F. Yanasse, S. Quegan, and R.J. Martin, *Inferences on spatial and temporal variability of the backscatter from growing crops using AgriSAR data*, Intern. J. Remote Sens. 13 (1992), pp. 493–507.
[27] S. Quegan, C. Yanasse, H. de Groof, P.N. Churchill and A.J. Sieber, *The radiometric quality of AgriSAR data*, Intern. J. Remote Sens. 12 (1991), pp. 277–302.
[28] L. Scaccia, *Testing for simplification in spatial models*, Ph.D. thesis, Università di Perugia, Italy, 2000.

Appendix

The five separable $\mathrm{CAR}(2)_{\mathrm{SD}}$ models simulated are as follows:

Label	β	$\rho(1,0)$	$\rho(0,1)$	$\rho(1,1)$	$\rho(1,-1)$
A1	(0, 0, 0)	0	0	0	0
A2	(0.15, 0.1,−0.015)	0.154	0.101	0.016	0.016
A3	(0.25, 0.2, −0.05)	0.268	0.209	0.056	0.056
A4	(0.35, 0.3, −0.105)	0.408	0.333	0.136	0.136
A5	(0.45, 0.4, −0.18)	0.627	0.500	0.313	0.313

The five separable Pickard models simulated are as follows:

Label	α	$\rho(1,0)$	$\rho(0,1)$	$\rho(1,1)$	$\rho(1,-1)$
A6	(0, 0)	0	0	0	0
A7	(0.2, 0.3)	0.2	0.3	0.06	0.06
A8	(0.4, 0.5)	0.4	0.5	0.20	0.20
A9	(0.6, 0.7)	0.6	0.7	0.42	0.42
A10	(0.8, 0.9)	0.8	0.9	0.72	0.72

The five separable $\mathrm{ARMA}(p_1, q_1)\cdot\mathrm{ARMA}(p_2, q_2)$ models simulated are as follows:

Label	θ_1	φ_1	θ_2	φ_2	$\rho(1,0)$	$\rho(0,1)$	$\rho(1,1)$	$\rho(1,-1)$
AM1	(0.5, 0.4)	0	(0.3, 0.2)	0	0.833	0.375	0.313	0.313
AM2	(0, 0)	0.3	(0.7, 0.2)	0	0.275	0.875	0.241	0.241
AM3	(0, 0)	0.4	(0, 0)	0.6	0.345	0.441	0.152	0.152
AM4	(0.3, 0.2)	0.2	(0.5, 0.4)	0	0.156	0.833	0.130	0.130
AM5	(0.4, 0.1)	0.6	(0.3, 0.4)	0.3	0.259	0.220	0.057	0.057

where θ_1 and φ_1 are, respectively, the parameters of the autoregressive part and those of the moving average part of process $\mathrm{ARMA}(p_1, q_1)$, while θ_2 and φ_2 are, respectively, the parameters of the autoregressive part and those of the moving average part of process $\mathrm{ARMA}(p_2, q_2)$.

The five non-separable $\mathrm{CAR}(2)_{\mathrm{SD}}$ models simulated are as follows:

Label	β	$\lvert\eta_\beta\rvert$	$\rho(1,0)$	$\rho(0,1)$	$\rho(1,1)$	$\rho(1,-1)$
C1	(0, 0, 0.25)	0.25	0	0	0.599	0.599
C2	(0.11, 0.07, 0.16)	0.17	0.491	0.476	0.488	0.488
C3	(0.2, 0.2, 0.05)	0.09	0.593	0.593	0.514	0.514
C4	(0.25, 0.25, 0)	0.06	0.631	0.631	0.538	0.538
C5	(0.3, 0.2, −0.03)	0.03	0.391	0.290	0.165	0.165

The five non-symmetric, non-separable Pickard models simulated are as follows:

Label	α	$\lvert\eta_\alpha\rvert$	$\rho(1,0)$	$\rho(0,1)$	$\rho(1,1)$	$\rho(1,-1)$
P1	(0.1, 0.2, 0.6)	0.62	0.426	0.476	0.733	0.203
P2	(0.6, 0.7, −0.8)	0.38	0.180	0.624	−0.300	0.112
P3	(0.3, 0.4, 0.26)	0.38	0.694	0.733	0.757	0.509
P4	(0.1, 0.6, 0.2)	0.26	0.426	0.716	0.527	0.305
P5	(0.3, 0.6, −0.15)	0.03	0.329	0.611	0.231	0.201