The 3rd International Conference on Mathematics and Statistics (ICoMS-3)
Institut Pertanian Bogor, Indonesia, 5-6 August 2008

MAXIMUM LIKELIHOOD ESTIMATION FOR THE NON-SEPARABLE SPATIAL UNILATERAL AUTOREGRESSIVE MODEL

[1]Norhashidah Awang, [2]Mahendran Shitan

[1] School of Mathematical Sciences, Universiti Sains Malaysia,
11800 USM, Penang, Malaysia

[2] Department of Mathematics, Faculty of Science, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor Darul Ehsan, Malaysia

e-mail : [1] shidah@usm.my, [2] mahen@fsas.upm.edu.my

Abstract. *Several classes of models have been proposed to describe the spatial processes. A special class of non-separable spatial unilateral model for modelling spatial data on two-dimensional rectangular regular grid is the spatial autoregressive model, denoted by AR(* p_1 *,1). In this paper, we establish a procedure to estimate the parameters of this model using the maximum likelihood method. We show that this procedure is practical and easy to be implemented. We illustrate the procedure by fitting the AR(1,1) model to two set of data on the regular rectangular grid. The results show that the model is adequate in capturing the spatial correlation in the data.*

***Keywords**: Parameter estimation; spatial autoregressive model; spatial unilateral model*

1. Introduction

A large amount of research in modelling spatial processes has been conducted and they have covered various applications. Spatial series may be considered as a generalisation of time series, however modelling it is more difficult since the dependence in spatial series spreads in all directions and it encounters larger proportion of edge effects as compared to time series.

In this paper, consideration will be given to modelling spatial process in two-dimensional regular grid where a random variable is defined at each intersection point. Various classes of models have been proposed to describe the process. Among the earliest models of this particular process are the Simultaneous Autoregressive (SAR) model (Whittle, 1954), the Conditional Autoregressive (CAR) model (Besag, 1974) and the Moving Average (MA) model (Haining, 1978).

Many methods and procedures have been developed and proposed to overcome the estimation problems in spatial modelling. Martin (1979, 1990 and 1996) studied a class of models called separable models, while Tjøstheim (1978 and 1983) and Basu and Reinsel (1992, 1993 and 1994) considered the unilateral models. A special characteristic of separable models is that it has a product correlation structure, which in turn simplifies the estimation. However, these models are only applied for data which exhibit separable correlation structure. The unilateral models can be analysed using extensions of time series theory in some special cases and are claimed useful in systems theory and digital filtering (see Tjøstheim, 1978).

A special class of non-separable unilateral model is the spatial autoregressive model, denoted by AR(p_1 ,1). Shitan and Brockwell (1996) have established the procedure to estimate the parameters of this model where the approach is to transform the two-dimensional spatial series to a multiple time series, treating one of the coordinates as a time index and the other as a multivariate index and then performed the multivariate least squares estimation procedures. In this paper, we look at the problem of estimating the parameters of this model from a different perspective. Our approach is to use maximum likelihood method with some modifications at the border to simplify the parameter estimation.

In the next section, the construction of the weight matrices is presented. The derivation of the procedure of estimation is discussed in Section 3. In Section 4, the numerical examples to illustrate the procedure are given and finally the conclusions are presented in Section 5.

2. Construction of the weight matrices

We consider a spatial process in two-dimensional regular grid of size $m \times n$. The non-separable spatial unilateral autoregressive, AR(p_1 ,1) model is defined as,

$$Y_{ij} = \alpha_{10}Y_{i-1,j} + \alpha_{01}Y_{i,j-1} + \alpha_{11}Y_{i-1,j-1} + \dots + \alpha_{p_1 0}Y_{i-p_1,j} + \alpha_{p_1 1}Y_{i-p_1,j-1} + \varepsilon_{ij}, \quad i = 1, 2, \dots, m \text{ and } j = 1, 2, \dots, n, \tag{2.1}$$

where $\{Y_{ij}\}$ is a sequence of two-dimensional random variable with zero mean and the errors ε_{ij} are assumed to be normally distributed with mean 0 and common variance σ^2 at site labelled (i, j), and α_{ij} 's are the parameters to be estimated.

By assuming that the unobserved values to be zeroes, and letting the observation vector,
$\mathbf{Y} = (Y_{11}, Y_{12}, \dots, Y_{1n}, Y_{21}, Y_{22}, \dots, Y_{2n}, \dots, Y_{m1}, Y_{m2}, \dots, Y_{mn})' = (\mathbf{Y_1}, \mathbf{Y_2}, \dots, \mathbf{Y_m})'$, where
$\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in})'$, $i = 1, 2, \dots, m$ and the error vector,
$\boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{12}, \dots, \varepsilon_{1n}, \varepsilon_{21}, \varepsilon_{22}, \dots, \varepsilon_{2n}, \dots, \varepsilon_{m1}, \varepsilon_{m2}, \dots, \varepsilon_{mn})' = (\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \dots, \boldsymbol{\varepsilon_m})'$, where
$\boldsymbol{\varepsilon_i} = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in})'$, $i = 1, 2, \dots, m$, we can rewrite equation (2.1) in the matrix form as,

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Phi}_2 & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \ddots & \ddots & \ddots & \ddots & \cdots & \ddots & \ddots & \cdots & \ddots & \ddots & \ddots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Phi}_{p_1} & \boldsymbol{\Phi}_{p_1-1} & \cdots & \boldsymbol{\Phi}_2 & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{pmatrix}, \tag{2.2}$$

where $\boldsymbol{\Phi}_j$'s are $n \times n$ matrices defined as,

$$\boldsymbol{\Phi}_0 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_{01} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \alpha_{01} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \alpha_{01} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{01} & 0 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Phi}_j = \begin{pmatrix} \alpha_{j0} & 0 & 0 & \cdots & 0 & 0 \\ \alpha_{j1} & \alpha_{j0} & 0 & \cdots & 0 & 0 \\ 0 & \alpha_{j1} & \alpha_{j0} & \cdots & 0 & 0 \\ 0 & 0 & \alpha_{j1} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{j1} & \alpha_{j0} \end{pmatrix}.$$

for $j = 1, 2, \dots, p_1$.

Equation (2.2) can be written more compactly as,

$$\mathbf{Y} = \boldsymbol{\Phi}\mathbf{Y} + \boldsymbol{\varepsilon}, \tag{2.3}$$

where $\boldsymbol{\Phi}$ is $N \times N$ matrix, $N = mn$. It is clear that $\boldsymbol{\Phi}$ is a lower triangular matrix with zeroes on the main diagonal. Then, if we decompose $\boldsymbol{\Phi}$ into $2p_1 + 1$ matrices such that it isolates different parameters, we obtain

$$\mathbf{Y} = (\alpha_{10}\mathbf{W}_{10} + \alpha_{01}\mathbf{W}_{01} + \alpha_{11}\mathbf{W}_{11} + \dots + \alpha_{p_1 0}\mathbf{W}_{p_1 0} + \alpha_{p_1 1}\mathbf{W}_{p_1 1})\mathbf{Y} + \boldsymbol{\varepsilon}, \tag{2.4}$$

where, $\boldsymbol{\Phi} = \alpha_{10}\mathbf{W}_{10} + \alpha_{01}\mathbf{W}_{01} + \alpha_{11}\mathbf{W}_{11} + \dots + \alpha_{p_1 0}\mathbf{W}_{p_1 0} + \alpha_{p_1 1}\mathbf{W}_{p_1 1}$ and $\mathbf{W}_{jk}$, $j = 1, 2, \dots, p_1$; $k = 0, 1$ are the $N \times N$ lower triangular weight matrices with elements ones and zeroes.

3. Maximum likelihood procedure for estimating the parameters

Equation (2.4) can then be written as,

$$\mathbf{Y} = \left(\mathbf{I} - (\alpha_{10}\mathbf{W}_{10} + \alpha_{01}\mathbf{W}_{01} + \alpha_{11}\mathbf{W}_{11} + \ldots + \alpha_{p_1 0}\mathbf{W}_{p_1 0} + \alpha_{p_1 1}\mathbf{W}_{p_1 1})\right)^{-1} \boldsymbol{\varepsilon} \qquad (3.1)$$

or

$$\mathbf{Y} = (\mathbf{I} - \boldsymbol{\Phi})^{-1} \boldsymbol{\varepsilon} \qquad (3.2)$$

where $\mathbf{I}$ is an $N\times N$ identity matrix.
Therefore, the covariance matrix of $\mathbf{Y}$, $\mathbf{V}$ is given as,

$$\mathbf{V} = \sigma^2 (\mathbf{I} - \boldsymbol{\Phi})^{-1}\left[(\mathbf{I} - \boldsymbol{\Phi})^{-1}\right]'. \qquad (3.3)$$

The square root of the determinant of $\mathbf{V}$ is given as,

$$|\mathbf{V}|^{1/2} = \left(\sigma^2\right)^{N/2}\left|(\mathbf{I} - \boldsymbol{\Phi})^{-1}\right|. \qquad (3.4)$$

Since $(\mathbf{I} - \boldsymbol{\Phi})$ is the lower triangular matrix with diagonal elements 1, $\left|(\mathbf{I} - \boldsymbol{\Phi})^{-1}\right| = 1$. This leads to

$$|\mathbf{V}|^{1/2} = \left(\sigma^2\right)^{N/2}. \qquad (3.5)$$

Therefore, the likelihood function l is given as,

$$\begin{aligned} l &= \frac{1}{(2\pi)^{N/2}|\mathbf{V}|^{1/2}}\exp\left\{-\frac{1}{2}\mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}\right\} \\ &= (2\pi)^{-N/2}(\sigma^2)^{-N/2}\exp\left\{-\frac{1}{2\sigma^2}\mathbf{Y}'\left[(\mathbf{I}-\boldsymbol{\Phi})^{-1}(\mathbf{I}-\boldsymbol{\Phi})^{-1}\right]^{-1}\mathbf{Y}\right\} \\ &= (2\pi)^{-N/2}(\sigma^2)^{-N/2}\exp\left\{-\frac{1}{2\sigma^2}\mathbf{Y}'(\mathbf{I}-\boldsymbol{\Phi})(\mathbf{I}-\boldsymbol{\Phi})\mathbf{Y}\right\}. \end{aligned}$$

Thus we obtain the log likelihood, L as

$$L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\mathbf{Y}'(\mathbf{I}-\boldsymbol{\Phi}')(\mathbf{I}-\boldsymbol{\Phi})\mathbf{Y}. \qquad (3.6)$$

The partial derivative of L with respect to α_{jk}, $j = 1, 2, \ldots, p_1$; $k = 0, 1$ is given by

$$\frac{\partial L}{\partial \alpha_{jk}} = -\frac{1}{\sigma^2}\left[-\mathbf{Y}'\mathbf{W}'_{jk}\mathbf{Y} + \alpha_{jk}\mathbf{Y}'\mathbf{W}'_{jk}\mathbf{W}_{jk}\mathbf{Y} + \sum_{\forall r\neq j}\sum_{\forall s\neq k}\alpha_{rs}\mathbf{Y}'\mathbf{W}'_{rs}\mathbf{W}_{jk}\mathbf{Y}\right] \qquad (3.7)$$

for $j = 1, 2, \ldots, 5$.

Equating (3.7) to zero leads to

$$\left[\alpha_{jk}\mathbf{Y}'\mathbf{W}'_{jk}\mathbf{W}_{jk}\mathbf{Y} + \sum_{\forall r\neq j}\sum_{\forall s\neq k}\alpha_{rs}\mathbf{Y}'\mathbf{W}'_{rs}\mathbf{W}_{jk}\mathbf{Y}\right] = \mathbf{Y}'\mathbf{W}'_{jk}\mathbf{Y}. \qquad (3.8)$$

Therefore, denoting $\mathbf{Z}_{jk} = \mathbf{W}_{jk}\mathbf{Y}$, the maximum likelihood for α_{jk}'s can be obtained by solving the equation

$$\begin{pmatrix} \mathbf{Z}'_{10}\mathbf{Z}_{10} & \mathbf{Z}'_{01}\mathbf{Z}_{10} & \cdots & \mathbf{Z}'_{p_1 0}\mathbf{Z}_{10} & \mathbf{Z}'_{p_1 1}\mathbf{Z}_{10} \\ \mathbf{Z}'_{10}\mathbf{Z}_{01} & \mathbf{Z}'_{01}\mathbf{Z}_{01} & \cdots & \mathbf{Z}'_{p_1 0}\mathbf{Z}_{01} & \mathbf{Z}'_{p_1 1}\mathbf{Z}_{01} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{Z}'_{10}\mathbf{Z}_{p_1 0} & \mathbf{Z}'_{01}\mathbf{Z}_{p_1 0} & \cdots & \mathbf{Z}'_{p_1 0}\mathbf{Z}_{p_1 0} & \mathbf{Z}'_{p_1 1}\mathbf{Z}_{p_1 0} \\ \mathbf{Z}'_{10}\mathbf{Z}_{p_1 1} & \mathbf{Z}'_{01}\mathbf{Z}_{p_1 1} & \cdots & \mathbf{Z}'_{p_1 0}\mathbf{Z}_{p_1 1} & \mathbf{Z}'_{p_1 1}\mathbf{Z}_{p_1 1} \end{pmatrix} \begin{pmatrix} \alpha_{10} \\ \alpha_{01} \\ \vdots \\ \alpha_{p_1 0} \\ \alpha_{p_1 1} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}'\mathbf{Z}_{10} \\ \mathbf{Y}'\mathbf{Z}_{01} \\ \vdots \\ \mathbf{Y}'\mathbf{Z}_{p_1 0} \\ \mathbf{Y}'\mathbf{Z}_{p_1 1} \end{pmatrix},$$

or
$$\begin{pmatrix} \alpha_{10} \\ \alpha_{01} \\ \vdots \\ \alpha_{p_10} \\ \alpha_{p_11} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}'_{10}\mathbf{Z}_{10} & \mathbf{Z}'_{01}\mathbf{Z}_{10} & \cdots & \mathbf{Z}'_{p_10}\mathbf{Z}_{10} & \mathbf{Z}'_{p_11}\mathbf{Z}_{10} \\ \mathbf{Z}'_{10}\mathbf{Z}_{01} & \mathbf{Z}'_{01}\mathbf{Z}_{01} & \cdots & \mathbf{Z}'_{p_10}\mathbf{Z}_{01} & \mathbf{Z}'_{p_11}\mathbf{Z}_{01} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{Z}'_{10}\mathbf{Z}_{p_10} & \mathbf{Z}'_{01}\mathbf{Z}_{p_10} & \cdots & \mathbf{Z}'_{p_10}\mathbf{Z}_{p_10} & \mathbf{Z}'_{p_11}\mathbf{Z}_{p_10} \\ \mathbf{Z}'_{10}\mathbf{Z}_{p_11} & \mathbf{Z}'_{01}\mathbf{Z}_{p_11} & \cdots & \mathbf{Z}'_{p_10}\mathbf{Z}_{p_11} & \mathbf{Z}'_{p_11}\mathbf{Z}_{p_11} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}'\mathbf{Z}_{10} \\ \mathbf{Y}'\mathbf{Z}_{01} \\ \vdots \\ \mathbf{Y}'\mathbf{Z}_{p_10} \\ \mathbf{Y}'\mathbf{Z}_{p_11} \end{pmatrix}. \qquad (3.9)$$

4. Numerical examples

We fit the AR(1,1) model to two well-known data sets observed on regular grids to illustrate the procedure mentioned above. The AR(1,1) model is defined as

$$Y_{ij} = \alpha_{10}Y_{i-1,j} + \alpha_{01}Y_{i,j-1} + \alpha_{11}Y_{i-1,j-1} + \varepsilon_{ij}, \qquad (4.1)$$

and hence, equation (2.2) becomes

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix} = \begin{pmatrix} \mathbf{\Phi}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Phi}_1 & \mathbf{\Phi}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_1 & \mathbf{\Phi}_0 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \ddots & \ddots & \ddots & \ddots & \cdots & \ddots & \ddots & \cdots & \ddots & \ddots & \ddots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{\Phi}_1 & \mathbf{\Phi}_0 \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{pmatrix}. \qquad (4.2)$$

From equation (3.9), the estimate of the parameters can be obtained by solving the equation

$$\begin{pmatrix} \hat{\alpha}_{10} \\ \hat{\alpha}_{01} \\ \hat{\alpha}_{11} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}'_{10}\mathbf{Z}_{10} & \mathbf{Z}'_{01}\mathbf{Z}_{10} & \mathbf{Z}'_{11}\mathbf{Z}_{10} \\ \mathbf{Z}'_{10}\mathbf{Z}_{01} & \mathbf{Z}'_{01}\mathbf{Z}_{01} & \mathbf{Z}'_{11}\mathbf{Z}_{01} \\ \mathbf{Z}'_{10}\mathbf{Z}_{11} & \mathbf{Z}'_{01}\mathbf{Z}_{11} & \mathbf{Z}'_{11}\mathbf{Z}_{11} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}'\mathbf{Z}_{10} \\ \mathbf{Y}'\mathbf{Z}_{01} \\ \mathbf{Y}'\mathbf{Z}_{11} \end{pmatrix} \qquad (4.3)$$

where $\mathbf{Z}_{10} = \mathbf{W}_{10}\mathbf{Y}$, $\mathbf{Z}_{01} = \mathbf{W}_{01}\mathbf{Y}$ and $\mathbf{Z}_{11} = \mathbf{W}_{11}\mathbf{Y}$. The matrices $\mathbf{W}_{10}$, $\mathbf{W}_{01}$ and $\mathbf{W}_{11}$ are the $N \times N$ weight matrices given as

$$\mathbf{W}_{10} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \cdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I} & \mathbf{0} \end{pmatrix}, \quad \mathbf{W}_{01} = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \cdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{D} \end{pmatrix} \text{ and } \mathbf{W}_{10} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{D} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \cdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{D} & \mathbf{0} \end{pmatrix}.$$

Here, $\mathbf{I}$ is an $n \times n$ identity matrix and $\mathbf{D}$ is an $n \times n$ matrix defined by

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

The computations are performed by a computer program written in S-Plus.

For the first numerical example, we consider the data set obtained in Cressie (1993) on wheat (yield of grain) where the uniformity trial was conducted by Mercer and Hall (1911). The data are observed on 25×20 grid. Since the data is stationary and fulfils the normality assumption (from the density and the Kolmogorov-Smirnov test), no transformation is needed. Table 1 displays the sample spatial correlations of the mean-corrected data for $s = 0$ to 4 and $t = -4$ to 4, where s is the lag in west-east direction (between columns) and t is the lag in south-north direction (between rows). The table suggests that there exists a considerable amount of spatial correlation in the data, and the correlation is much stronger along the south-north direction (between rows) than along the west-east direction (between columns), for example, $\hat{\rho}_{01} = 0.494$ and $\hat{\rho}_{02} = 0.357$ (correlations along the south-north direction at lags 1 and 2, respectively) whereas $\hat{\rho}_{10} = 0.280$ and $\hat{\rho}_{20} = 0.140$ (correlations along the west-east direction at lags 1 and 2, respectively).

Table 1: Sample spatial correlations, $\hat{\rho}_{st}$ for the wheat (yield of grain) mean-corrected data of size 25×20 obtained in Cressie (1993).

		s				
		0	1	2	3	4
	−4	0.282	0.066	−0.075	0.095	−0.026
	−3	0.303	0.063	−0.011	0.094	−0.006
	−2	0.357	0.114	0.001	0.117	0.022
	−1	0.494	0.167	0.021	0.134	0.056
t	0	1.000	0.280	0.140	0.166	0.065
	1	0.494	0.212	0.112	0.160	0.046
	2	0.357	0.152	0.081	0.192	0.084
	3	0.303	0.096	0.057	0.176	0.045
	4	0.282	0.106	0.062	0.157	0.068

The maximum likelihood estimate of the parameters is obtained using the equation (4.3) and is shown in Table 2, together with the estimates of (−2 ln L) and σ^2. The goodness of the fitted model is examined by the spatial correlations of the residuals. The correlations of the residuals from the fitted model are small and these suggest that the model is adequate in capturing the spatial correlations in the data.

Table 2: Parameters Estimate of AR(1,1) model fit to the wheat (yield of grain) mean-corrected data of size 25×20.

	$\hat{\alpha}_{10}$	$\hat{\alpha}_{01}$	$\hat{\alpha}_{01}$	$\hat{\sigma}^2$	$-2 \ln L$
MLE Estimate	0.226	0.505	-0.093	0.1396	461.8415

In the second numerical example, we analyse the data set on the yield of barley from an agriculture uniformity trial experiment at Plant Breeding Institute, Cambridge, United Kingdom and the data is obtained in Kempton and Howes (1981). The data set are of 7×28. The plot of the data shows that it is not stationary and therefore a transformation is needed to make it stationary. In this analysis, we apply the spatial variate differencing where the first row differencing is performed and the result shows that the resulting series has no apparent trend, suggesting that the transformation is sufficient to make the data stationary. Furthermore, the density plot and the result of the Kolmogorov-Smirnov test suggest that the data satisfy the normality assumption. Table 3 displays the sample spatial correlation of the mean-corrected original data, whereas Table 4 displays the sample spatial correlation of the mean-corrected values of the differenced data. Having satisfied the stationary and normality assumption, we fit the AR(1,1) model to this differenced data and the results are displayed in Table 5. Again, the correlations of the residuals from this model are small and these suggest that the model is adequate in capturing the spatial correlations in the data.

Table 3: Sample spatial correlations, $\hat{\rho}_{st}$ for the yield of barley mean-corrected data of size 7×28 from Kempton and Howes (1981).

		s				
		0	1	2	3	4
	−4	0.470	0.220	0.053	0.035	−0.035
	−3	0.570	0.231	0.037	0.015	−0.065
	−2	0.677	0.241	0.041	0.026	−0.063
	−1	0.796	0.253	0.035	0.015	−0.075
t	0	1.000	0.264	0.013	−0.025	−0.097
	1	0.796	0.190	−0.045	−0.060	−0.101
	2	0.677	0.137	−0.061	−0.064	−0.075
	3	0.570	0.088	−0.063	−0.065	−0.055
	4	0.470	0.044	−0.079	−0.073	−0.051

Table 4: Sample spatial correlations, $\hat{\rho}_{st}$ for the mean-corrected of first row-differenced yields of size 7×27.

		s				
		0	1	2	3	4
	−4	−0.064	0.010	0.079	−0.052	0.021
	−3	−0.030	0.035	−0.014	−0.024	−0.068
	−2	0.005	−0.020	−0.018	0.030	0.014
	−1	−0.206	−0.042	0.020	0.048	0.004
t	0	1.000	0.182	0.077	0.000	0.005
	1	−0.206	−0.052	−0.108	−0.070	−0.094
	2	0.005	0.029	−0.012	−0.022	−0.019
	3	−0.030	−0.060	0.061	0.002	0.043
	4	−0.064	0.006	−0.027	−0.058	−0.034

Table 5: Parameters Estimate of AR(1,1) model fit to the barley mean-corrected differenced data of size 7×27.

	$\hat{\alpha}_{10}$	$\hat{\alpha}_{01}$	$\hat{\alpha}_{01}$	$\hat{\sigma}^2$	$-2 \ln L$
MLE Estimate	0.212	−0.209	0.039	0.0338	-100.3186

5. Conclusion

In this paper, we have proposed a procedure to estimate the parameters of the spatial non-separable unilateral autoregressive models, i.e. the AR(p_1 ,1) model using the maximum likelihood method. The procedure is practical and easy to be implemented. We have illustrated the procedure by fitting the model for $p_1 = 1$ to two set of data on the regular rectangular grid. The results show that the model is adequate in capturing the spatial correlation in the data; hence, we conclude that this class of model and the estimation method proposed can provide as an alternative for modelling spatial data on regular rectangular grid.

6. References

Basu, S. and Reinsel, G. C. (1992), A note on properties of spatial Yule-Walker estimators, *Journal of Statistical Computing and Simulation*, **41**, pp. 243-255.

Basu, S. and Reinsel, G. C. (1993), Properties of the spatial unilateral first order ARMA model, *Advances in Applied Probability*, **25**, pp. 631-648.

Basu, S. and Reinsel, G. C. (1994), Regression models with spatially correlated errors, *Journal of the American Statistical Association: Theory and Method*, **89(425)**, pp. 88-99.

Besag, J. E. (1974), Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society B*, **36**, pp. 192-236.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, Revised Edition, Wiley, New York.

Haining, R. P. (1978b), The moving average model for spatial interaction, *Transactions Institute of British Geographer*, **3**, pp. 202-225.

Kempton, R. A. and Howes, C. W. (1981), The use of neighbouring plot values in the analysis of variety trials, *Applied Statistics*, **30**, pp. 59-70.

Martin, R. J. (1979), A subclass of lattice processes applied to a problem of planar sampling, *Biometrika*, **66**, pp. 209-217.

Martin, R. J. (1990), The use of time series models and methods in the analysis of agricultural field trials, *Communication in Statistics - Theory and Method*, **19(1)**, pp. 55-81.

Martin, R. J. (1996), Some results on unilateral ARMA lattice processes, *Journal of Statistical Planning and Inference*, **50**, pp. 395-411.

Mercer, W. B. and Hall, A. D. (1911), The experimental error of field trials, *Journal of Agricultural Science*, **4**, pp. 107-132.

Shitan, M. and Brockwell, P. J. (1996), An alternative estimation procedure of the spatial AR(p_1,1) model, Research Report No.2, Dept of Statistics and Operations Research, RMIT, Australia.

Tjøstheim, D. (1978), Statistical spatial series modelling, *Advances in Applied Probability*, **10**, pp. 130-154.

Tjøstheim, D. (1983), Statistical spatial series modelling II. Some further results on unilateral lattice processes, *Advances in Applied Probability*, **15**, pp. 562-584.

Whittle, P. (1954), On stationary processes in the plane, *Biometrika*, **41**, pp.434-449.