Robust Estimators in a Spatial Unilateral Autoregressive Model

Eric A. Grau[+] and Jye-Chyi Lu[*]

+: Research Triangle Institute (TRI) International, RTP, NC 27709;
*: Contacting person. The School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205.

ABSTRACT

Data organized in spatial lattices (e.g., semiconductor die grids, agricultural field trials) are commonly modeled using spatial autoregressive processes. When outliers are present in the data, classical parameter estimation techniques for these models, such as least squares and maximum likelihood, can be inefficient and potentially mislead the analyst. This article proposes robust methods for estimating parameters in spatial autoregressive AR(1) × AR(1) model that accommodate outliers in the data. While the methods sacrifice a minimal amount of efficiency when outliers are not present, Monte Carlo simulations show that these methods have smaller bias and variance in the autocorrelation estimates than either the least squares or Gaussian maximum likelihood. These methods also give improved estimates of the mean function parameters when outliers are present. Cases studied include (1) outliers are present when the mean function is zero; (2) the data are corrected for a given mean function; or (3) the mean function is estimated simultaneously with variance function in Generalized Least Squares. A sketch of derivations of the asymptotic distributions of the robust estimates is given. Three examples from the literature illustrate the advantages of proposed methods.

Key words: Additive outliers; Generalized least squares; M-estimates; Median polish; Random field; Spatial Statistics.

1. INTRODUCTION

Modeling spatially correlated lattice data is important in many applications. For example, when analyzing data from semiconductor die grid, electronic circuit layouts, medical image patterns, or agricultural field trials (e.g., Kempton and Howes, 1981, Besag and Kempton, 1986), data are often collected in two-dimensional lattices or grids. Martin (1990) and Cullis and Gleeson (1991) used the separable autoregressive moving average (ARMA) models, mainly of first order, for modeling two-dimensional (2D) lattice field trial data. Zimmerman and Harville (1991) examined special 2D random field models, including the separable first-order autoregressive (AR) correlation model. Basu and Reinsel (1993) studied the properties of the spatial unilateral first-order ARMA models. Martin (1979, 1990) and Basu and Reinsel (1993, 1994) showed that the separable spatial unilateral ARMA models are capable of representing a wide range of general spatial correlation patterns in two dimensions including both isotropic and anisotropic correlation patterns, and have wide use in practical modeling. A special case of the unilateral ARMA models, the AR(1) × AR(1) model, will be discussed in this paper. This model can be written as

$$y_{ij} = \alpha_1 y_{i-1,j} + \alpha_2 y_{i,j-1} + \alpha_3 y_{i-1,j-1} + e_{ij}, \quad \alpha_3 = -\alpha_1\alpha_2, \quad \varepsilon_r \sim i.i.d.\ N(0, \sigma^2), \tag{1}$$

where $i = 1, \ldots, m$ (rows), $j = 1, \ldots, n$ (columns), $y_{ij} = 0$ for $i \leq 0$ and $j \leq 0$, and *i.i.d.* stands for independent and identically distributed.

Spatial models usually require that the observations homogeneously follow a Gaussian distribution, as is the case with the AR(1) × AR(1) model above. The assumption that observations are homogeneously generated from the same Gaussian distribution is, in practice, often violated, either due to the contamination of the distribution by a few errant observations, or because the generating process was not Gaussian at all. The discordant observations are usually referred to as outliers, which are caused either by measurement errors, recording mistakes, or by some event which makes an observation or set of observations unusual. These outliers will have significant impact to the analysis of data, especially in precision engineering work such as Nortel's electronics manufacturing, where the outliers can easily make the treatment effects insignificant and thus it is difficult to show that the new product innovation is effective.

Outliers can occur randomly or in clusters. For example, particle contamination in semiconductor manufacturing will create random outliers; whereas problems in the soldering process might lead to a cluster of errant observations in printed circuit board production. These randomly distributed or clustered outliers may or may not influence surrounding observations depending on the physical process of generating these outliers. Figure 1 provides a visualization of the impact of these outliers with simulated data on 2D grids mimicking experience learned from electronics manufacturing processes, where the presence of outliers made the experimental study of treatment effects very difficult in developing new products at Nortel. This motivates the research of robust estimation methods in this article. See Section 4.1 for the details of data simulation.

The following presents another example with real-life data taken from the literature. To challenge the often held assumption of uniformity within and between growth chambers, from a series of well planned experiments Lee and Rawlings (1982) analyzed data from a uniformity trial comparing several indices of soybean plant growth after twenty days in growth chambers of various sizes, under various temperature conditions, using five indicators of plant growth. They found that growth was not uniform within and between growth chambers. Data are presented for one of the chambers is provided in Figure 2, using one of the indicators of plant growth, total leaf area of each plant Y_{ij}, under the daytime/nighttime temperature regime of 22/18° C. Here i and j represent the position of the plant within the growth chamber. Two apparent outliers at positions (6, 3) and (3, 10) could present problems when estimating the mean function, the spatial autocorrelation, or both. See Sections 2.3 and 5.3 for further analyses to understand this data set. Note that in the semiconductor manufacturing the uniformity of measurements from material deposition processes has similar problems as described above. Both clustered and randomly distributed "outliers" could occur.

(Place Figures 1 and 2 here.)

In this paper, we consider alternative methods of estimating parameters in the AR(1) × AR(1) model that can accommodate outliers more readily than classical estimators. These methods will be both *resistant* (to the existence of outliers) and *robust* (to deviations from the model assumptions, which in our case include an assumption of Gaussianity). Section 2 presents the underlying model and the usual estimation techniques. In Section 3, robust estimators are developed. Section 4 compares the efficiencies of the proposed robust estimators with the maximum likelihood (ML) and least squares (LS) estimators through Monte Carlo studies. Section 5 illustrates the proposed methods with examples from the literature. Section 6 concludes this article and discusses potential future work.

2. MODELING SPATIALLY CORRELATED GRID DATA

2.1 Description of AR(1) × AR(1) Model

Suppose $\boldsymbol{Z}(\boldsymbol{s})$ is a random variable measured at some generic spatial location $\boldsymbol{s}$ in d-dimensional Euclidean space. If we let $\boldsymbol{s}$ vary over the index set $D \subset \mathbf{R}^{d,}$ then $\{\boldsymbol{Z}(\boldsymbol{s})\text{: } \boldsymbol{s} \in D\}$ defines a (random) spatial process. For an $m \times n$ rectangular grid that is part of a lattice spatial process, the random variable $\boldsymbol{Z}$ measured at site s is labeled as $Z_{i,j}$, where the spatial location s is labeled as site (i, j); there are a total of mn sites. Then $Z_{i,j}$ can be decomposed into the following components

$$Z_{i,j} = \mu_{i,j} + y_{i,j}, \quad i = 1, \dots, m; \quad j = 1, \dots n,$$

where $\mu_{i,j}$ represents the mean function, and $y_{i,j}$ represents the stochastic variation and dependence in the data. If the points on the grid are independent, then a common assumption is that $y_{i,j}$'s are random $N(0, \sigma^2)$ errors. However, if spatial dependence is in the data, it can be represented by geostatistical models or spatial unilateral models, including the spatial unilateral first-order ARMA models. Under the assumption that the stochastic variation can be represented by a first-order AR model, an alternative representation of this decomposition is given as

$$Z_{i,j} = \mu_{i,j} + \alpha_1 y_{i-1,j} + \alpha_2 y_{i,j-1} + \alpha_3 y_{i-1,j-1} + \varepsilon_{ij}, \; i = 1, \ldots, m; \; j = 1, \ldots n,$$

where the mean function represents large-scale variation, and the stochastic dependence terms represent small-scale variation, as modeled by the first-order AR model. This AR model is a special case of more general unilateral ARMA models, the properties of which have been discussed by Tjostheim (1978, 1981, 1983), Martin (1990), and Basu and Reinsel (1993, 1994).

The procedures in this paper are applied to the AR(1) × AR(1) model with the restriction $\alpha_3 = -\alpha_1\alpha_2$. This restriction considerably simplifies the inferential procedures, including estimation of parameters using maximum likelihood, and has been found practically useful. See Martin (1979, 1990) for a more thorough discussion of the simplifying properties of this restriction, and the applications of the resulting model. Martin also discusses possible checks on whether such a restriction is reasonable through the singular value decomposition of a matrix of sample correlations or a Monte Carlo test. If the mean function can be represented by a linear regression such that $(X\beta)_{i,j} = \mu_{i,j}$, then the large-scale variation parameters (β) and the small-scale variation parameters (α_1 and α_2), can be estimated using Generalized Least Squares (GLS). See Basu and Reinsel (1994) for details on using GLS on the AR(1) × AR(1) model.

2.2 Classical Methods of Estimation of Autocorrelation Parameters

The likelihood function for the AR(1) × AR(1) model is given by Basu (1990) and Basu and Reinsel (1993) as

$$\mathrm{L}(\alpha, \sigma^2 \mid y) = (\sigma^2)^{-mn/2} (1 - \alpha_1^2)^{n/2} (1 - \alpha_2^2)^{m/2} \exp\{-\mathrm{S}(\alpha) / (2\sigma^2)\},$$

where

$$\mathrm{S}(\alpha) = (1 - \alpha_2^2) \sum_{i=1}^{m} (y_{i1} - \alpha_1 y_{i-1,1})^2 + (1 - \alpha_1^2) \sum_{j=1}^{n} (y_{1j} - \alpha_2 y_{1,j-1})^2 +$$

$$(1 - \alpha_1^2)(1 - \alpha_2^2)\, y_{11}^2.$$

ML estimates can be obtained using Newton-Raphson or Fisher scoring method. Details are available in Basu and Reinsel (1993). Consistency and asymptotic normality of the ML estimates were established by Basu (1990) with the following asymptotic covariance matrix C_{ML}^{-1}, where

$$C_{ML} = \mathrm{Diag}[(1 - \alpha_1^2)^{-1}, (1 - \alpha_2^2)^{-1}, (2\sigma^4)^{-1}]. \tag{2}$$

Basu and Reinsel (1993) also confirmed (by simulation) that the estimated standard errors, as obtained from the finite sample information matrix, are asymptotically equivalent to (2). Basu (1990) and Khalil (1991) investigated linear and nonlinear LS estimators, respectively, of the autocorrelation parameters in the AR(1) × AR(1) model. Khalil showed that both estimators are consistent and asymptotically normal with the same covariance matrix C_{LS}^{-1}, where $C_{LS} = \mathrm{Diag}[(1 - \alpha_1^2)^{-1}, (1 - \alpha_2^2)^{-1}]$. Thus, the least squares and maximum likelihood estimators of the autocorrelation parameters are asymptotically equivalent.

(Please place Table 1 here.)

2.3 Application of Classical Estimators to Lee-Rawlings Example

In modeling spatial lattice data, mean functions can be used to capture global trends (large-scale variation) exhibited in the data. If interest is only in estimating the spatial dependence in the data, we can estimate the autocorrelation parameters after accounting for, or "correcting for", the mean function. For the Lee-Rawlings data presented in the introduction, we corrected the data with the overall mean, and considered a mean model that accounted for the trend in the data, where the design matrix included a column for a constant, row number (position i), column number (position j), and an interaction. One possible representation of the large- and small-scale variation is given by

$$Z_{i,j} = \text{overall mean} + \text{row effects} + \text{column effects} + \text{residual}.$$

Here the overall mean + row effects + column effects represents the mean function, and the residual is assumed to follow an AR(1) × AR(1) model. The overall mean, row effects, and column effects can be removed by subtracting row means, then subtracting column means (also referred to as "mean polish"). The estimates of the row and column effects obtained using mean polish are equivalent to Ordinary Least Squares estimates of those effects. A robust alternative to mean polish is median polish (Tukey, 1977). Instead of arriving at an overall mean, row effects, column effects, and residuals in a single iteration, median polish requires several iterations to arrive at the decomposition. If interest is in the mean function parameters themselves, or in both the mean function and autocorrelation parameters, then the preferred strategy is to use GLS methods, which were briefly described in Section 2.1, and in detail in Basu and Reinsel (1994).

Estimates of the autocorrelation parameters for residuals from the Lee-Rawlings data using mean-corrected residuals, and trend-corrected residuals are presented in Table 1. A note of caution with these estimates: Cressie (1993, pages 48-50) shows that subtracting a mean function and analyzing the residuals results in a negative bias in the estimation of the covariance function and therefore the autocorrelation parameter estimates (since the covariances are usually positive). Additional bias may also be present due the outliers apparent in the data.

The two classical estimators, already indicated to be asymptotically equivalent, do not differ substantially in Table 1, even with the small grid size in this example. Some of the spatial autocorrelation that is apparent in the mean-corrected model is simply a reflection of an inappropriate mean model. It results in less autocorrelation in the residuals using a trend model to represent the mean model. Regardless of the model used to estimate the mean, we would expect that outliers would severely impact these estimates, given the small grid size.

The conclusions that applied to mean-corrected residuals would also apply to GLS estimators; namely, that the classical estimators of the autocorrelation parameters may be severely impacted if outliers are present. Comparisons using estimates from GLS are not provided here; however, since the analyses using residuals after correcting for the mean model are just the first iteration of GLS, it should not be surprising that conclusions are consistent whether GLS was used or not.

3. ROBUST ESTIMATION METHODS

3.1 Contamination of AR(1) × AR(1) Model

Because Model (1) is a 2D generalization of the one-dimensional AR(1) model, we can consider contamination of this model as a generalization of the contamination of an AR(1) model. Two types of outliers in time series, identified by Fox (1972), are additive outliers and innovative outliers. The additive outlier, as defined by Fox, is a "gross error of observation or recording error" which only affects a single observation. An innovations outlier is one that not only affects the single observation, but also subsequent observations. The extension of the additive outlier to two dimensions is straightforward. It is also possible to mathematically describe a spatial model with innovative outliers, however the practical application is not as clear. The nature of the AR(1) × AR(1) would require the innovations to affect observations in exactly two of the four directions. Since real-life incidence of the innovation outlier in space is questionable, it is not discussed in this paper. The extension of the definition of the additive outlier to two dimensions follows below:

Spatial AO Model: Additive Outliers

The random variables y_{ij} can be expressed as $W_{ij} + V_{ij}$ where W_{ij} follows a separable AR(1) × AR(1) Model with *i.i.d.* $N(0, \sigma^2)$ errors. The random variables V_{ij} are independent of W_{ij}, and are *i.i.d.* following the distribution H:

$$H = (1 - \xi)\, \delta_0 + \xi\, N(0, \tau^2), \tag{3}$$

where δ_0 is the distribution that assigns probability one to the origin. Therefore, with probability $1 - \xi$, the AR(1) × AR(1) model is observed, and with probability ξ the response at site (i, j) follows the AR(1) × AR(1) model plus an error with distribution $N(0, \tau^2)$.

Remark. For the data generated from the outlier model given above, we recommend using an extension of the method proposed in Chang, Tiao and Chen (1988) (for the one-dimensional time series data) to test the presence of outliers and identify their locations. However, to keep our presentation focused, only the studies relevant to Model (3) are discussed here. Define

$$\eta_{i,j}^{(L)} = 1 \text{ if } (i, j) = \boldsymbol{L};\ \ 0 \text{ otherwise},$$

where $\boldsymbol{L}$ is a 2D location random variable describing where the outlier occurred. Using the spatial unilateral AR(1) model $Y_{i,j}$ defined in Model (1) as an example, a spatial AO mean-contamination model is defined as follows:

$$Z_{i,j} = Y_{i,j} + \lambda\ \eta_{i,j}^{(\boldsymbol{L})},$$

where λ is an unknown fixed parameter to estimate. In this model, at the location $\boldsymbol{L}$, the outlier only changes the mean from zero to an unknown non-zero quantity λ. As in Chang, Tiao, and Chen, an iterative procedure could be used for outlier detection and parameter estimation.

The methods for estimating parameters in the AR(1) × AR(1) model presented in Section 2 are not resistant to outliers. We therefore propose two new estimators that are resistant to outliers, and evaluate the effectiveness of each estimator when the model is a contaminated AR(1) × AR(1).

3.2 Robust M-Estimator

Denote $w_{i,j} = (z_{i-1,j},\ z_{i,j-1},\ z_{i-1,j-1})'$, $\theta = (\alpha_1, \alpha_2)'$ and $\pi = (\alpha_1, \alpha_2, \alpha_3)'$. Under the constraint where $\alpha_3 = -\alpha_1 \alpha_2$, the robust M-estimate θ_M is found by minimizing

$$\min_\theta \sum_{i=2}^{m} \sum_{j=2}^{n} \rho\,(z_{i,j} - \boldsymbol{w}'_{\boldsymbol{i,j}}\, \pi),$$

where $\rho(\bullet)$ is a symmetric robustifying loss function. See Equations (4) and (5) below for examples. Taking the first derivatives of this function with respect to the parameters π, we obtain the robust M-estimator $\hat{\theta}_M$ as a solution of the following estimating equations.

$$\sum_{i=2}^{m}\sum_{j=2}^{n} \psi(z_{i,j} - \mathbf{w'_{i,j}}\pi)(z_{i-1,j} - \alpha_2 z_{i-1,j-1}) = 0,$$

$$\sum_{i=2}^{m}\sum_{j=2}^{n} \psi(z_{i,j} - \mathbf{w'_{i,j}}\pi)(z_{i,j-1} - \alpha_1 z_{i-1,j-1}) = 0,$$

where $\psi(\bullet) = \rho'(\bullet)$ is a bounded function with $t\psi(t) \geq 0$ and usually $\psi'(0) = 1$. The ψ–functions commonly used are either from the Huber family given by

$$\psi_{\mathrm{H},\, c_1}(x) = \mathrm{sgn}(x)\min(|x| / \hat{\sigma}, c_1)\, \hat{\sigma}, \tag{4}$$

where sgn(x) is the sign function, or from the redescending bisquare family:

$$\psi_{\mathrm{B},\, c_2}(x) = x\,(1 - x^2 / c_2^2)^2, \tag{5}$$

developed by Tukey (1977), and $\hat{\sigma}$ can be computed simultaneously as a robust estimate of the scale parameter. For example, the following estimator is commonly used in the one-dimensional time-series robust estimation literature (e.g., Denby and Martin (1979) and Bustos and Yohai (1986)):

$$\hat{\sigma} = \mathrm{med}(\,|r_{i,j}|\,) / 0.6745, \quad i = 2, 3, \ldots, m\,, \quad j = 2, 3, \ldots, n.$$

Here $r_{i,j} = z_{i,j} - w'_{i,j}\hat{\pi}$. The values of the tuning constants c_1 and c_2 depend upon how we wish to define "outlier". Smaller values of c_1 and c_2 neglect the effect of borderline outliers, and larger values negate the effect of only the most serious outliers. Experience indicates that appropriate values of c_1 would be between 1.0 and 1.5; appropriate values of c_2 lie in the range from 4.5 to 7.5.

Following a technique suggested by Beaton and Tukey (1974), Denby and Martin (1979) used iterative weighted least squares (IWLS) to calculate parameter estimates to determine M-estimates and GM-estimates; without the weighting term, IWLS is just a modified form of Gauss-Newton nonlinear least squares, which is what we used in the 2D case. Following the approach of Gleeson and McGilchrist (1980), edge sites are set aside in the estimation procedure, used only as neighbors for interior sites.

3.3 *RA*-Estimator

The separable spatial unilateral AR(1) model can be rewritten in terms of the existing data as $(1 - \alpha_1 \mathrm{B}_1)(1 - \alpha_2 \mathrm{B}_2)\, y_{i,j} = \varepsilon_{i,j}$, where B_1 and B_2 are backshift operators defined as $\mathrm{B}_1\, yij = y_{i-1,j}$ and $\mathrm{B}_2\, y_{ij} = y_{i,j-1}$, or

$$y_{ij} = (1 - \alpha_1 B_1)^{-1}(1 - \alpha_2 B_2)^{-1}\ \varepsilon_{i,j}. \tag{6}$$

The least squares estimate $\hat{\theta}_{LS}$ is determined in the usual way as the minimum of $\sum_{i=1}^{m}\sum_{j=1}^{n} \varepsilon_{i,j}^{2}$. Differentiating with respect to the autocorrelation parameters, and replacing y_{ij} where it appears by (6), the least squares estimates are determined as the solutions to the following equations:

$$\sum_{s=0}^{\infty}\hat{\alpha}_1^{\,s}\sum_{i=s+2}^{m}\sum_{j=2}^{n} r_{i,j}\, r_{i-s-1,j} = 0, \quad \sum_{s=0}^{\infty}\hat{\alpha}_2^{\,s}\sum_{i=2}^{m}\sum_{j=s+2}^{n} r_{i,j}\, r_{i,j-s-1} = 0.$$

The *RA*-estimator is formed by making the residual covariances, given by

$$\sum_{i=s+2}^{m}\sum_{j=2}^{n} r_{i,j}\, r_{i-s-1,j} \text{ and } \sum_{i=2}^{m}\sum_{j=s+2}^{n} r_{i,j}\, r_{i,j-s-1},$$

robust, by replacing $r_{i,j}\, r_{i-s-1,j}$ and $r_{i,j}\, r_{i,j-s-1}$ by their robustified counterparts, $\eta(r_{i,j}, r_{i-s-1,j})$ and $\eta(r_{i,j}, r_{i,j-s-1})$, where η is a robustifying function. One example of a robustifying function : η: $\mathbf{R}^2 \Rightarrow \mathbf{R}$ is the Mallows function, $\eta(\mu,\upsilon)=\psi(\mu)\,\psi(\upsilon)$, where ψ: $\mathbf{R} \Rightarrow \mathbf{R}$ is the Huber or Tukey's bisquare function defined as before. Another candidate robustifying function, the Hampel function, was introduced in the one-dimensional case by Bustos and Yohai (1986), but will not be considered here.

The *RA*-estimate $\hat{\theta}_{RA}$ is a solution of the equations (see Appendix I for details)

$$\sum_{s=0}^{\infty}\hat{\alpha}_1^{\,s}\sum_{i=s+2}^{m}\sum_{j=2}^{n} \eta(r_{i,j}, r_{i-s-1,j}) = 0; \quad \sum_{s=0}^{\infty}\hat{\alpha}_2^{\,s}\sum_{i=2}^{m}\sum_{j=s+2}^{n} \eta(r_{i,j}, r_{i,j-s-1}) = 0. \tag{7}$$

If we use the Mallows function, computation of the *RA*-estimate can be done using the least squares procedure iteratively in a procedure extended from Bustos and Yohai (1986) in one dimension. Let $r^*_{i,j} = \psi(r_{i,j}/\hat{\sigma})\hat{\sigma}$. Then

$$z^*_{i,j} = (1 - \hat{\sigma}_1 B_1)(1 - \hat{\sigma}_2 B_2)\, r^*_{i,j}, \tag{8}$$

which could be regarded as a Winsorization of the AR(1) × AR(1) process. Bustos and Yohai describe the process, which is paraphrased here with adjustments for the 2D case:

> Start with initial estimates $\hat{\alpha}_{10}$, $\hat{\alpha}_{20}$, and $\hat{\sigma}$. Given the estimate $\hat{\alpha}_{1n}$ and $\hat{\alpha}_{2n}$ corresponding to the n^{th} iteration, define $z_{i,j}^{*(n)}$ by (8) with $\hat{\alpha}_{1n}$ and $\hat{\alpha}_{2n}$ instead of $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Then $\hat{\alpha}_{1n+1}$ and $\hat{\alpha}_{2n+1}$ is the LS estimate of the series $z_{i,j}^{*(n)}$. Simultaneous iteration on $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\sigma}$ is also possible and desirable.

When the tuning constant (defined in Section 3.2) is very large, the estimates provided by this technique are equivalent to least squares.

In this paper, we compare estimation of the autocorrelation parameters in Model (1) using three forms of the *RA*-estimator, in addition to the ML estimator, where the three forms are determined by the robustifying function η. The first is not "robustifying" at all, but is simply the LS estimator, where $\eta(\mu,\upsilon) = \mu\upsilon$, which, as indicated earlier, is asymptotically equivalent to the ML estimator. In Appendix I, we show that the robust *M*-estimator is a version of the *RA*-estimator with $\eta(\mu,\upsilon) = \psi(\mu)\upsilon$. When the tuning constant approaches ∞ and $\psi(\mu)$ is appropriately rescaled, $\psi(\mu)\upsilon$ approaches $\mu\upsilon$ (i.e., the Robust *M*-estimator becomes equivalent to the LS-estimator).

3.4 Asymptotic Properties of the *M*- and *RA*-estimators

The asymptotic properties of the *M*- and *RA*-estimators can be derived in a manner similar to the one-dimensional case (Bustos, Fraiman and Yohai, 1984). Suppose there are N nodes in the lattice grid. Suppose y_{ij} follows an AR(1) × AR(1) model, with errors that have symmetric distribution F. Let the autocorrelation parameters in Model (1) be defined as $\boldsymbol{\theta} = (\alpha_1, \alpha_2)$. Let $\hat{\theta}_{RA}$ be the solution of the *RA*-estimating equations (7), and let θ_0 be the true vector parameter. The following theorem presents its asymptotic variance under the normal error assumption.

Theorem 1. Suppose that the error distribution of F is normal. Let η and ψ be bounded and continuously differentiable functions (When using the Mallow's η function and Tukey's bisquare, these conditions will hold.). Assume that

$$E[\eta_1(\varepsilon_{i,j}, \varepsilon_{i\text{-}1,j})\, \varepsilon_{i\text{-}1,j}] \neq 0 \quad \text{and} \quad E[\psi'(\varepsilon_{i,j})] \neq 0,$$

where $\eta_1(\mu, \upsilon) = \partial\eta(\mu,\upsilon) / \partial\mu$. Then the variance of the asymptotic normal distribution of $(N)^{-1/2}(\hat{\theta}_{\mathrm{RA}} - \theta_0)$ is given by

$$\sigma_{11} = \sum_s \alpha_1^{2s}\, E(\eta^2(\varepsilon_{i,j}, \varepsilon_{i\text{-}1,j}))\{\sum_s \alpha_1^{2s}\, E(\eta_1(\varepsilon_{i,j}, \varepsilon_{i\text{-}1,j})\, \varepsilon_{i\text{-}1,j})\}^{-2},$$

$$\sigma_{22} = \sum_s \alpha_2^{2s}\, E(\eta^2(\varepsilon_{i,j}, \varepsilon_{i,j\text{-}1}))\{\sum_s \alpha_2^{2s}\, E(\eta_2(\varepsilon_{i,j}, \varepsilon_{i,j\text{-}1})\, \varepsilon_{i,j\text{-}1})\}^{-2},$$

and

$$\sigma_{12} = \sigma_{21} = 0,\ \eta_1(\mu, \upsilon) = \partial\eta(\mu,\upsilon) / \partial\mu..$$

A brief outline of the proof is given in Appendix II. More details are provided in Grau (2000).

Remark. The assumptions that η and ψ are bounded do not hold for the non-robust LS estimators described in Section 2; however, if we substitute $\mu\upsilon$ for $\eta(\mu,\upsilon)$ in the theorem, it is easy to show that we arrive at the variance-covariance matrix (2).

4. SIMULATION STUDIES

4.1 Experimental Study Setup

In order to ascertain the efficacy of these methods, Monte Carlo simulation studies with 500 replications were conducted comparing the relative efficiency of the proposed robust estimators against the LS and ML estimators of α_1 and α_2 described in Section 2. Six scenarios were used to compare autocorrelation parameter estimates: (1) no attempt to estimate the mean function; (2) extracting a single mean estimate; (3) detrending in the manner described in Section 2.3; (4) applying mean polish; (5) applying median polish; and (6) generalized least squares, with mean function parameters corresponding to a single mean and the trend parameters described in Section 2.3. In the case of generalized least squares, the estimates of the mean function parameters were also evaluated.

To generate a grid of data following Model (1), we first need to generate the border values. This was done by following the procedure described by Etchison (1993, page 67), who also simulated an AR(1) × AR(1) model. For the remainder of values, the $\varepsilon_{i,j}$ were generated as standard normal deviates, from which the $y_{i,j}$ were generated following Model (1). Standard normal errors were generated using the subroutine GASDEV from *Numerical Recipes in Fortran* (Press *et al*. 1992), with adjustments to the variances made for the border cases.

The simulations were conducted using both the bisquare and the Huber ψ function. However, since there were difficulties with convergence using the bisquare function in about 20% of the data sets, results presented here are limited to the Huber ψ function. Simulations were conducted for square lattices with 7, 10, 15, 25 and 40 nodes on a side, as well as for the tuning constant set to $c = 1.0$ and $c = 1.5$; for the 7 × 7 case (with $c = 1.5$), the Monte Carlo simulation error is about 0.005. Results are presented here for the grid with 7 on a side and a tuning constant of 1.5; conclusions that follow would be more pronounced with larger sample sizes and more severe Winsorization.

The autocorrelation parameter values were fixed as $\alpha_1 = 0.7$ and $\alpha_2 = 0.4$; other values were investigated including negative correlations, and conclusions were not affected. We used the Gauss-Newton algorithm to obtain the LS parameter estimates, which, as a nonlinear estimator, is asymptotically equivalent to linear LS under the separable model (Khalil 1991). For the ML estimates, we used the likelihood presented in Section 2.2 as determined by Basu and Reinsel (1993). The performance of the robust techniques was compared to LS and ML under four scenarios: (1) no outliers present; (2)

a single AO, at a fixed point found near the center; (3) 10% of the observations are AO's, found in a cluster at a fixed location near the center; and (4) 10% of the observations are AO's, randomly dispersed throughout the grid.

In the case of random AO outliers, the w_{ij} were generated as a grid of uncontaminated AR(1) × AR(1) data points following the description given above. A uniform(0,1) deviate u was randomly selected using the RAN2 subroutine from *Numerical Recipes in Fortran* (Press *et al*., 1992), and if $u > r$, then $y_{ij} = w_{ij}$. If $u \leq r$, then v_{ij} were generated as $N(0, \tau^2)$ random deviates, and $y_{ij} = w_{ij} + v_{ij}$. Simulations were conducted for $\tau^2 = 9$, $\tau^2 = 36$ and $\tau^2 = 100$. Results presented here are for $\tau^2 = 100$; conclusions presented here are consistent with those for less severe outliers. See Grau (2000) for details of results from other configurations.

4.2 Evaluating Estimators of α_1 and α_2

Tables 2-5 summarize the mean, mean square error, and relative efficiencies, of LS, ML, robust *M*- and *RA*-estimates of the AR(1) × AR(1) autocorrelation parameters from 500 simulations of 7 × 7 grids under all four outlier configurations presented in Section 4.1, for four of the six methods of extracting the mean discussed in Section 4.1. The results from the second method (single mean extracted) and third method (detrended data, using the trend function presented in Section 2.3) are not presented here. These methods are just the first iteration of the GLS estimates with the same mean functions estimated.

(Please place Tables 2-5 here.)

In these tables, under the heading "Mean", we find the empirical means calculated from the 500 simulated data sets. "MSE" is the Mean Squared Error, which is a combination of the bias of the empirical mean when compared to the true autocorrelation and the empirical variance. Relative efficiencies, labeled as "Eff." in the tables, were determined by calculating the ratio of the mean square errors of the LS estimator and the ML, robust *M*- and *RA*-estimators. When the relative efficiency is larger than one, the MSE in the LS estimate is larger than the MSE in the other estimates. (Since the ML and LS estimators are asymptotically equivalent, we would expect the ratio of their MSE's to be close to 1, particularly for larger grids.)

To get some sense of the level of error associated with the relative efficiency, Fieller's theorem (Fieller, 1954) was used to create confidence intervals, within which the true relative efficiency would be found with 95% confidence, labeled as "95% C.I." in the tables. The width of the confidence bands depends upon the simulation error, and would therefore be narrower (wider) if a greater (lesser) number of simulations had been used.

Estimates of σ^2 were also compared. A loss function that penalizes underestimation and overestimation of the variance equally is due to Stein (1964), and is defined as $L(\sigma^2, V) = V / \sigma^2 - 1 - \log(V / \sigma^2)$, where V is the estimator of σ^2 that is being evaluated. As is pointed out in Casella and Berger (1990, pages 465-467), MSE is not a good statistic to use to evaluate estimators of σ^2. In the MSE, "underestimation has only a finite penalty while overestimation has an infinite penalty," (Casella and Berger, 1990, page 467). The Stein loss function was used instead of MSE in the comparisons.

4.3 Results: Monte Carlo Study to Evaluate Estimators

In general, regardless of whether or not a mean function is estimated, the robust estimators are adequate when no outliers are present, although the maximum likelihood estimator is significantly superior to the other techniques, particularly with small grids, mainly due to a smaller empirical variance. However, the LS and ML techniques are asymptotically equivalent, and this was borne out by the investigation with larger grids (not presented here). The robust estimators provide a definite advantage when outliers are present. The *RA*-estimator (i.e., the *RA*-estimator that uses the Hampel and Huber functions) is clearly superior to the other techniques in the presence of severe outliers, particularly when the outliers are clustered.

Tables 2-5 indicate that the biases of the LS, ML, and robust *M*-estimates are much worse than that of the *RA*-estimate in the presence of AO's. The bias of LS and ML estimates are comparable, but lower empirical variances with ML estimates give rise to superior estimates with the smaller grids. With increased grid size, the differences between the ML and LS estimators diminish.

Using the residuals from detrended data to represent the errors will introduce bias into the estimation of the variance function, due mostly to the linear constraints on the residuals. This bias, derived by Cressie (1993, page 49), is negative if the covariances are positive, which is usually the case. Cressie and Glonek (1984) showed that when trying to estimate the variance function with Gaussian data points equally spaced on a transect, the bias when working with median polish residuals is smaller than the bias that exists when the effects are estimated using a mean. Cressie (1993, page 50) conjectures that the same relationship will hold with larger grids.

The negative bias in the estimation of the variance function should translate into a negative bias for some or all of the parameters to be estimated. We would also expect this bias to be less with median polish residuals as opposed to mean polish residuals. Having no outliers present translates to an analog of the Gaussian process discussed by Cressie and Glonek. Tables 2-5 show that the bias is indeed more negative with mean polish residuals than with median polish residuals, though the difference is far greater with the classical estimators (LS and ML) than with the robust estimators. Moreover, the difference in bias diminishes considerably with the larger grid size.

This conclusion remains the same for the autocorrelation parameter estimates when outliers are introduced. Even though the robust estimators have less bias than their nonrobust counterparts, the improvement obtained by using median polish residuals versus mean polish residuals is far less than the improvement obtained with the classical estimators. As a corollary to this, the gain in efficiency obtained by using robust estimators instead of nonrobust estimators on median polish detrended data is also less (often significantly so) than the gain obtained by using robust estimators on mean polish detrended data.

Results comparing estimates of σ^2 are not presented here in the interests of space; conclusions with these estimates are consistent with those presented for the autocorrelation parameters.

5. REAL-LIFE EXAMPLES

The proposed robust estimation techniques are applied to three data sets taken from the literature, including the Lee-Rawlings data presented in the introduction. Standard errors of the parameter estimates for *RA*-estimation and *M*-estimation techniques were calculated using a parametric bootstrap, where the model depended on the type of outliers apparently present in the residuals. Autocorrelation parameter estimates are determined for residuals from using models consisting of a single mean, and a trend model described in Section 2.3. The estimates shown are equivalent to the first iteration of the GLS, so naturally these estimates were very similar to the autocorrelation parameter estimates obtained using GLS.

5.1 Data from Kempton and Howes (1981)

This data consist of uniformity trials on a 28 × 7 grid of spring barley plot yields (Kempton and Howes, 1981). No obvious outliers (or additional trends) are visually apparent in a plot of the data in Figure 3. Since no trends were apparent, only the mean was considered as a mean function. Estimates shown in Table 6 show little difference between the estimates from the various methods.

(Please place Figure 3 and Tables 6 and 7 here.)

We saw in the simulation section that when median polish is used to detrend the data, the spatial correlation structure is most like that of the original data. We see the same result here in Table 7: with mean polish, the spatial correlation in the east-west direction (α_2) goes away, whereas with median polish there is still a significant (or nearly significant) spatial autocorrelation in that direction. For both methods, the spatial autocorrelation in the both directions have decreased; this could be due to the negative bias associated with estimating so many effects in the mean function.

Both methods do bring out a curious anomaly: whereas the LS and the two robust estimates of α_1 do not differ by much, the ML estimate is quite a bit larger than the other estimates (though the ML estimate 95% confidence interval would cross the 95% confidence interval for the other estimators). For this data set, it is possible that the negative bias alluded to earlier do not affect the ML estimates severely as the other estimators.

5.2 Data from Cullis, Lill, Fisher, Read and Gleeson (1989)

Cullis *et al*. (1989) analyzed data from an experiment assessing the yield potential of 525 test lines from the southern New South Wales wheat breeding programs (mean-corrected data shown in Figure 4). Two missing values were interpolated using the means of nearest neighbors.

Estimates of parameters from residuals of the mean-alone model (Table 8) show some differences between the *RA*-estimates and the LS and *M*-estimation methods. Although no obvious outliers or trends are seen from the three-dimensional data plot, the difference in parameter estimates indicates the possible presence of additive outliers. After considering a more complicated mean model, outliers still seem to be present in the spatially correlated residuals, as indicated by the difference in parameter estimates from residuals of the trend surface model (also in Table 8). It is apparent that these data have trends in them, as shown by the significant trend parameter corresponding to the column number. The trend parameter estimates are virtually identical between the estimation techniques, as are the residual mean square values.

For both mean polish and median polish residuals, the spatial correlation in the east-west direction (α_2) seems to disappear compared to the data with just a mean or trend surface removed, as shown in Table 9. There are no major differences between the median polish residual estimates of the autocorrelation parameters and the mean polish residual estimates, which is probably due to the large grid size.

(Place Figure 4 and Tables 8 and 9 here.)

5.3 Data from Lee and Rawlings (1982)

In the introduction, possible difficulties associated with using LS or ML estimation techniques were pointed out with the Lee-Rawlings data. That outliers are present is evident from these results, where both robust estimators give estimates quite different from the standard techniques, as shown in Table 10. Two apparent outliers at positions (6, 3) and (3, 10) have created havoc for the LS and ML estimators of the autocorrelation parameters, regardless of what mean function is used; replacing these values with the overall mean gives estimates that are close to those given by the *RA*-estimator. One of the trend parameter estimates is arguably significant. The trend parameter estimates obtained using *RA* do differ somewhat from those of ML and LS, which are the result of using differing autocorrelation parameter estimates. It is clear that median polish and mean polish have removed most of the spatial autocorrelation in both directions (Table 11).

(Please place Tables 10 and 11 here.)

6. CONCLUSION AND FURTHER STUDIES

We have studied the stationary AR(1) × AR(1) model, and introduced robust alternatives to the classical estimators of the autocorrelation parameters in this model. The AR(1) × AR(1) model is a special case of the more general spatial ARMA models, which have been studied extensively in the literature (without treatment of possible outlier(s)), and shown to have wide use, particularly with two-dimensional arrays of data. Even the AR(1) × AR(1) model, which is the focus of this article, has been shown to have wide use, as cited by Martin (1990) and Basu and Reinsel (1993).

We have attempted to show where robust estimation of the autocorrelation parameters is advantageous in two contexts. First, we looked at the error process itself (assuming a mean of zero, and imposing no treatment structure in the model). Then, we looked at situations where treatment structure exists, both when prediction was the final goal, and the mean function must be removed from the data, and when estimating the mean function itself was the final goal. The two robust techniques are generalizations of methods developed in the time-series context: the robust *M*-estimation technique and Residual Autocovariance (*RA*) technique. Monte Carlo simulations suggest that the *RA* technique would be more appropriate than least squares or maximum likelihood for the efficient estimation of parameters when outliers are present.

We applied the robust estimators to three examples, one where robust estimation was not necessary, but the robust estimators performed as well as standard techniques, and two where the *RA*-estimator gave us somewhat different estimates than the LS, ML, or *M*-estimators.

When outliers contaminate data, techniques for estimating parameters are generally badly affected, and it is in the practitioner's interest to find robust alternatives. For data that is spatially correlated, outliers not only affect the estimation of mean function, but also of the variance function. This research presents alternative estimators of the autocorrelation parameters in a model that has seen a wide array of uses with regular grid data. Not only will the estimators improve the estimation of the autocorrelation parameters, but in some cases will improve the efficiency of the mean parameters in certain cases, though the scale of the improvement is much less than what was witnessed with the autocorrelation parameters. In particular, a small but statistically significant improvement was observed with a single outlier and with clustered outliers, with most of the improvement resulting from a decrease in the variance of the estimate.

The AR(1) × AR(1) model can be extended by adding autocorrelation terms (at longer lags), moving average terms, and by removing the assumption of separability. The definitions of these robust estimators can be extended to the first-order nonseparable model, and can probably be defined for the other extensions as well.

APPENDIX

A. Robust *M*-estimator – A Special Case of the *RA*-estimator

Let the autocorrelation parameters in Model (1) be defined as $\theta = (\alpha_1, \alpha_2)$. Let the error be defined as $\varepsilon_{i,j} = (1 - \alpha_1 B_1)(1 - \alpha_2 B_2)\, z_{i,j}$, and the residual as $r_{i,j} = (1 - \hat{\alpha}_1 B_1)(1 - \hat{\alpha}_2 B_2)\, z_{i,j}$, where B_1 and B_2 are defined in Section 3.3. The least squares estimate $\hat{\theta}_{LS}$ is determined in the usual way as the minimum of $\sum_{i=1}^{m}\sum_{j=1}^{n} \varepsilon^2_{i,j}$. Differentiating with respect to the autocorrelation parameters to obtain the minimum, we arrive at the usual normal equations:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} r_{i,j}\, \partial r_{i,j} / \partial\alpha_1 = 0 \quad\text{and}\quad \sum_{i=1}^{m}\sum_{j=1}^{n} r_{i,j}\, \partial r_{i,j} / \partial\alpha_2 = 0,$$

where $\partial r_{i,j} / \partial\alpha_1 = \partial r_{i,j} / \partial\alpha_1 \big|_{\alpha_1=\hat{\alpha}_1}$ and $\partial r_{i,j} / \partial\alpha_2 = \partial r_{i,j} / \partial\alpha_2 \big|_{\alpha_2=\hat{\alpha}_2}$. Since $\partial r_{i,j} / \partial\alpha_1 =$

$-B_1(1 - \alpha_2 B_2)\, z_{i,j} = -B_1(1 - \alpha_1 B_1)^{-1}\, r_{i,j} = -B_1 \sum_s \alpha_1^{\,s} B_1^{\,s} \varepsilon_{i,j} = -\sum_s \alpha_1^{\,s} \varepsilon_{i-1-s,j}$, and

$\partial r_{i,j}/\partial\alpha_2 = -\sum_s \alpha_2^{\,s} \varepsilon_{i,j-1-s}$, we obtain the following equivalent normal equations:

$$\sum_s \hat{\alpha}_1^{\,s} \sum\sum r_{i,j}\, r_{i-s-1,j} = 0, \quad \sum_s \hat{\alpha}_2^{\,s} \sum\sum r_{i,j}\, r_{i,j-s-1} = 0.$$

The RA-estimate is defined by replacing $\sum\sum r_{i,j}\ r_{i\text{-}s\text{-}1,j}$ and $\sum\sum r_{i,j}\ r_{i,j\text{-}s\text{-}1}$, respectively, with $\eta(r_{i,j}, r_{i\text{-}s\text{-}1,j})$ and $\sum\sum \eta(r_{i,j}, r_{i,j\text{-}s\text{-}1})$.

When the robustifying product function is $\eta(\mu,\upsilon) = \psi(\mu)\upsilon$, we have

$$\sum_s \hat{\alpha}_1^{\,s} \sum\sum \eta(r_{i,j}, r_{i\text{-}s\text{-}1,j}) = \sum_s \hat{\alpha}_2^{\,s} \sum\sum \eta(r_{i,j}, r_{i,j\text{-}s\text{-}1}) = 0.$$

But, $\sum_s \alpha_1^{\,s} \varepsilon_{i\text{-}1\text{-}s,j} = -\,\partial\,\varepsilon_{i,j} / \partial\alpha_1 = \mathrm{B}_1(1\text{-}\ \alpha_2\mathrm{B}_2)\, z_{i,j.}$ Thus, we have

$$\sum_s \hat{\alpha}_1^{\,s} \sum\sum \eta(r_{i,j}, r_{i\text{-}s\text{-}1,j}) = \sum\sum \psi(r_{i,j})\, \mathrm{B}_1(1 - \hat{\alpha}_2\mathrm{B}_2)\, z_{i,j}$$

$$= \sum\sum \psi(r_{i,j})\, (z_{i\text{-}1,j} - \hat{\alpha}_2\, z_{i\text{-}1,j\text{-}1}) = 0.$$

Using the same idea, we also get

$$\sum_s \hat{\alpha}_2^{\,s} \sum\sum \eta(r_{i,j}, r_{i,j\text{-}s\text{-}1}) = \sum\sum \psi(r_{i,j})\, \mathrm{B}_2(1 - \hat{\alpha}_1\mathrm{B}_1)\, z_{i,j}$$

$$= \sum\sum \psi(r_{i,j})\, (z_{i,j\text{-}1} - \hat{\alpha}_1\, z_{i\text{-}1,j\text{-}1}) = 0.$$

These are the robust M-estimating equations. Hence the robust M-estimator is just a special case of the RA-estimator.

B. Asymptotic Distribution of Robust *M*- and *RA*-estimates

To show asymptotic normality, define the M-estimating equations as $G_N(\theta)$ which is defined as

$$G_N(\theta) = (1/\,N)\,[\sum_{s=0}^{\infty} \alpha_1^{\,s} \sum_{i=s+2}^{m}\sum_{j=2}^{n} \eta(\varepsilon_{i,j}, \varepsilon_{i\text{-}s\text{-}1,j}),\ \sum_{s=0}^{\infty} \alpha_2^{\,s} \sum_{i=2}^{m}\sum_{j=s+2}^{n} \eta(\varepsilon_{i,j}, \varepsilon_{i,j\text{-}s\text{-}1})]'.$$

Applying the Taylor series expansion to $G_N(\theta)$ and evaluating the results at $\hat{\theta}$, we obtain

$$G_N(\hat{\theta}) = G_N(\theta_0) + (\hat{\theta} - \theta_0)\, \dot{G}_N(\theta_0) + \boldsymbol{R}_N = \mathbf{0},$$

From this it can be shown that

$$\sqrt{N}\,(\hat{\theta} - \theta_0) = -(\dot{G}_N(\theta_0))^{-1} \sqrt{N}\, G_N(\theta_0) + \sqrt{N}\, \boldsymbol{R}_N,$$

where $\dot{G}_N(\theta_0)$ is the 2×2 matrix of partial derivatives. Ergodicity allows us to assert that

$$\dot{G}_N(\boldsymbol{\theta}_0) \xrightarrow{p} \mathrm{E}(\partial/\partial\boldsymbol{\theta}[\sum_{s=0}^{\infty}\alpha_{10}{}^{s}\eta(\varepsilon_{i+s,j},\ \varepsilon_{i-1,j}),\ \sum_{t=0}^{\infty}\alpha_{20}{}^{t}\eta(\ \varepsilon_{i,j+t},\ \varepsilon_{i,j-1})]\ ) = \mathrm{A}(\boldsymbol{\theta}_0).$$

Stationarity and ergodicity allows us to invoke the Central Limit Theorem, such that

$$G_N(\boldsymbol{\theta}_0) \xrightarrow{d} N_2[\mathbf{0}, \mathbf{B}(\boldsymbol{\theta}_0)].$$

Slutsky's Theorem leads us to

$$\sqrt{N}\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_2[\mathbf{0}, \mathbf{A}^{-1}(\boldsymbol{\theta}_0)\mathbf{B}(\boldsymbol{\theta}_0)[\ \mathbf{A}^{-1}(\boldsymbol{\theta}_0)]^t],$$

where A and B are very complicated and thus evaluated in Grau (2000).

REFERENCES

Basu, S. (1990). "Analysis of first order spatial bilateral ARMA models," Unpublished PhD Thesis, University of Wisconsin, Madison, WI.

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

Basu, S., and Reinsel, G. C. (1994). "Regression Models with spatially correlated errors," *Journal of the American Statistical Association*, **89**, 88-99.

Beaton, A. E. and Tukey, J. W. (1974). "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data," *Technometrics*, **16**, 147-185.

Besag, J. E., and Kempton, R. A. (1986). "Statistical analysis of field experiments using neighboring plots," *Biometrics*, **42**, 231-251.

Bustos, D. H., Fraiman, R., and Yohai, V.J. (1984), "Asymptotic behavior of the estimates based on residual autocovariances for ARMA models," in *Lecture Notes in Statistics* **26**, 1984, eds. J. Franke, W. Hardle, and D. Martin, New York: Springer-Verlag.

Bustos, D. H. and Yohai, V. J. (1986), "Robust estimates for ARMA models," *Journal of the American Statistical Association*, **81**, 155-168.

Casella, G., and Berger, R. L. (1990). *Statistical Inference*. Belmont, CA: Duxbury.

Chang, I., Tiao, G. C., and Chen, C. (1988), "Estimation of time series parameters in the presence of outliers," *Technometrics*, **30**, 193-204.

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

Cressie, N. A. C. and Glonek, G. (1984). "Median based covariogram estimators reduce bias," *Statistics and Probability Letters*, **2**, 299-304.

Cullis, B. R., and Gleeson, A. C. (1991). "Spatial analysis of field experiments--an extension to two dimensions," *Biometrics*, **47**, 1449-1460.

Cullis, B. R., Lill, W. J., Fisher, J. A., Read, B. J., and Gleeson, A. C. (1989). "A new procedure for the analysis of early generation variety trials," *Applied Statistics*, **38**, 361-375

Denby, L., and Martin, R. D. (1979), "Robust estimation of the first order autoregressive parameter," *Journal of the American Statistical Association*, **74**, 140-146

Etchison, T. L. (1993). "Model Identification and Selection Techniques for Stationary ARMA(p_1, q_1) × ARMA(p_2, q_2) Processes," Unpublished PhD Thesis, Department of Statistics, North Carolina State University, Raleigh.

Fieller, E. C. (1954). "Some problems in Interval Estimation," *Journal of the Royal Statistical Society B*, **16**, 175-185.

Fox, A. J. (1972). "Outliers in Time Series," *Journal of the Royal Statistical Society B*, **43**, 350-363.

Grau, E. A. (2000). "Robust estimation of autocorrelation parameters in the AR(1) × AR(1) model," Unpublished PhD Thesis, North Carolina State University, Raleigh, NC.

Kempton, R. A., and Howes, C. W. (1981), "The Use of Neighboring Plot Values in the Analysis of Variety Trials," *Applied Statistics*, **30**, 59-70.

Khalil, T. M. (1991), "A study of the doubly geometric process for stationary and a non-stationary case," Unpublished PhD Thesis, North Carolina State University, Raleigh, NC.

Lee, C. S., and Rawlings, J. O. (1982), "Design of experiments in growth chambers - uniformity trials in the North Carolina State University phytotron," *Crop Science*, **22**, 551-558.

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

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

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992), *Numerical Recipes in Fortran*, 2nd ed. Cambridge: Cambridge University Press.

Stein, C. (1964). "Inadmissibility of the Usual Estimator of the Variance of a Normal Distribution with Unknown Mean," *Annals of the Institute of Statistical Mathematics*, **16**, 155-160.

Tjostheim, D. (1978). "Statistical spatial series modeling," *Advances in Applied Probability*, **10**, 130-154.

Tjostheim, D. (1981). "Autoregressive modeling and spectral analysis of array data in the plane," *IEEE Trans. Geosci. Rem. Sensing*, **19**, 15-24.

Tjostheim, D. (1983). "Statistical spatial series modeling II: Some further results on unilateral lattice processes," Advances in Applied Probability, **15**, 562-584.

Tukey, J. W. (1977). Exploratory Data Analysis. Reading, MA: Addison-Wesley.

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

Zimmerman, D. L., and Harville, D. A. (1991). "A random field approach to the analysis of field-plot experiments and other spatial experiments," *Biometrics*, **47**, 223-239.

Figure 1. Plots of Clustered and Randomized Outliers in a 7 × 7 grid

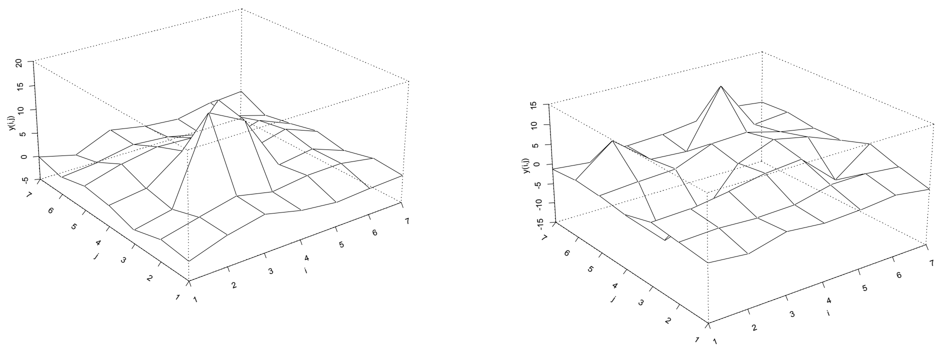

Figure 2. Plot and Contour of Lee and Rawlings (1982) Data (Mean Corrected).

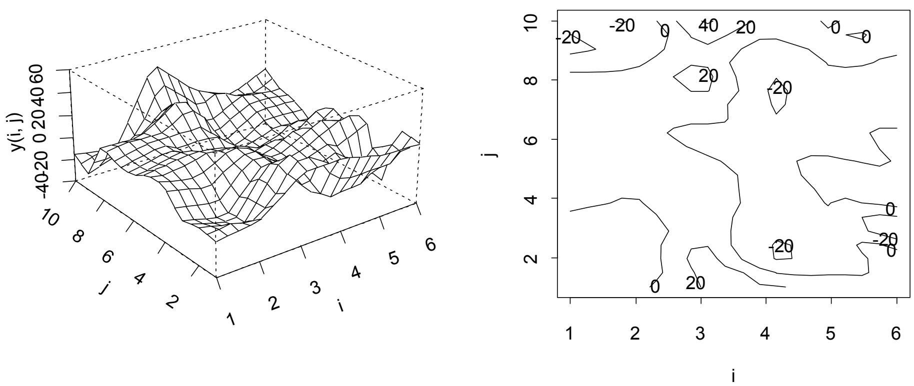

Figure 3. Plot and Contour of Kempton-Howes (1981) Data (Mean Corrected).

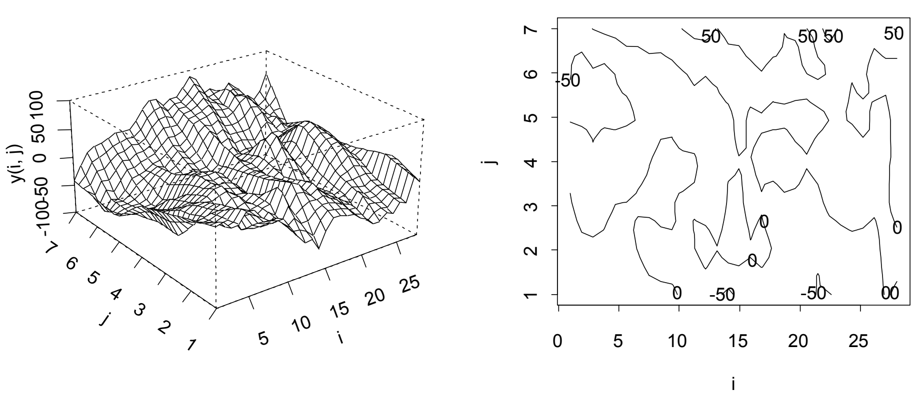

Figure 4. Plot and Contour of Cullis *et al.* (1989) Data (Mean Corrected)

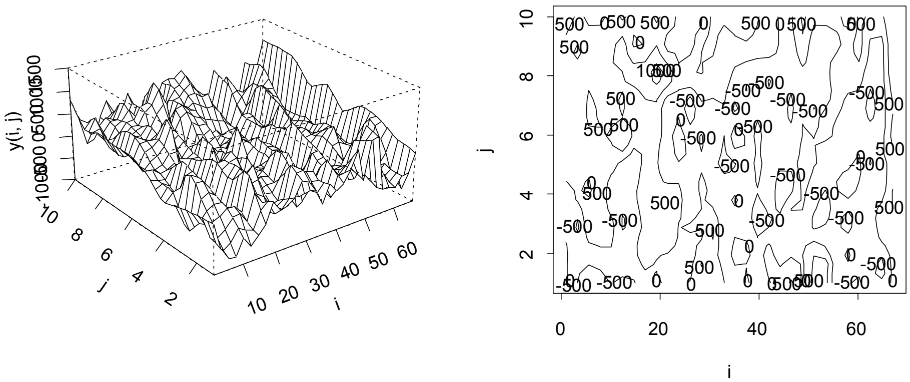

Table 1. Lee-Rawlings Data, Mean-corrected and Trend-corrected Residuals

	Mean model		Trend-surface model	
Estimator	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
Least squares	0.384 (0.146)	0.650 (0.125)	-0.045 (0.151)	0.306 (0.147)
Max. Likelihood	0.395 (0.131)	0.592 (0.110)	0.034 (0.142)	0.274 (0.129)

Table 2. Performance of Estimators of Autocorrelation Parameters under Model with No Outliers

	Estimates of α_1				**Estimates of α_2**			
Estimator	Mean	MSE	Eff.	95% C. I.	Mean	MSE	Eff.	95% C. I.
No mean est.								
LS	0.67	0.017	1.00		0.37	0.027	1.00	
MLE	0.67	0.012	1.37	1.20, 1.56	0.37	0.022	1.24	1.15, 1.35
Robust M	0.67	0.018	0.97	0.93, 1.00	0.37	0.027	0.98	0.95, 1.01
RA	0.67	0.020	0.86	0.82, 0.91	0.37	0.029	0.94	0.90, 0.99
Mean polish								
LS	0.30	0.191	1.00		0.11	0.118	1.00	
MLE	0.32	0.178	1.07	1.03, 1.10	0.11	0.120	0.99	0.95, 1.03
Robust M	0.30	0.190	1.00	0.98, 1.01	0.11	0.117	1.00	0.99, 1.02
RA	0.31	0.190	1.00	0.98, 1.02	0.11	0.119	0.99	0.97, 1.02
Median pol.								
LS	0.35	0.155	1.00		0.18	0.076	1.00	
MLE	0.37	0.138	1.10	1.06, 1.15	0.18	0.076	0.99	0.94, 1.05
Robust M	0.32	0.175	0.89	0.86, 0.91	0.16	0.084	0.89	0.86, 0.92
RA	0.29	0.199	0.77	0.74, 0.80	0.14	0.099	0.76	0.72, 0.80
GLS, Mean								
LS	0.65	0.021	1.00		0.35	0.029	1.00	
MLE	0.65	0.017	1.23	1.11, 1.37	0.34	0.026	1.14	1.06, 1.22
Robust M	0.65	0.021	1.01	0.92, 1.10	0.35	0.030	0.99	0.96, 1.01
RA	0.64	0.023	0.87	0.83, 0.91	0.35	0.031	0.93	0.90, 0.99
GLS, Trend								
LS	0.61	0.028	1.00		0.30	0.039	1.00	
MLE	0.58	0.034	0.85	0.79, 0.90	0.26	0.046	0.84	0.78, 0.90
Robust M	0.61	0.029	0.98	0.95, 1.01	0.30	0.039	0.99	0.96, 1.02
RA	0.61	0.031	0.92	0.88, 0.97	0.30	0.040	0.97	0.92, 1.02

Table 3. Performance of Estimators of Autocorrelation Parameters under Model with One Outlier at Position at Position (3, 3)

	Estimates of α_1				**Estimates of α_2**			
Estimator	Mean	MSE	Eff.	95% C. I.	Mean	MSE	Eff.	95% C. I.
No mean est.								
LS	0.39	0.153	1.00		0.20	0.069	1.00	
MLE	0.40	0.139	1.08	1.06, 1.10	0.21	0.063	1.10	1.07, 1.13
Robust M	0.45	0.117	1.29	1.23, 1.36	0.19	0.072	0.96	0.91, 1.00
RA	0.61	0.033	4.61	4.00, 5.39	0.29	0.044	1.59	1.44, 1.76
Mean polish								
LS	0.04	0.474	1.00		-0.04	0.225	1.00	
MLE	0.07	0.444	1.07	1.06, 1.08	-0.02	0.206	1.09	1.07, 1.11
Robust M	0.09	0.403	1.17	1.16, 1.19	-0.00	0.186	1.20	1.18, 1.23
RA	0.22	0.257	1.84	1.77, 1.93	0.10	0.122	1.84	1.74, 1.97
Median pol.								
LS	0.13	0.357	1.00		0.06	0.137	1.00	
MLE	0.15	0.338	1.06	1.04, 1.07	0.07	0.129	1.06	1.03, 1.08
Robust M	0.12	0.356	1.00	0.98, 1.02	0.05	0.135	1.01	0.98, 1.04
RA	0.24	0.239	1.49	1.41, 1.57	0.09	0.122	1.12	1.05, 1.21
GLS, Mean								
LS	0.36	0.172	1.00		0.16	0.087	1.00	
MLE	0.37	0.160	1.08	1.06, 1.09	0.16	0.082	1.06	1.04, 1.09
Robust M	0.42	0.132	1.31	1.25, 1.37	0.16	0.083	1.05	1.01, 1.09
RA	0.59	0.035	4.96	4.38, 5.67	0.27	0.043	2.01	1.82, 2.23
GLS, Trend								
LS	0.30	0.219	1.00		0.10	0.120	1.00	
MLE	0.30	0.213	1.03	1.01, 1.04	0.09	0.121	1.00	0.98, 1.01
Robust M	0.35	0.180	1.21	1.17, 1.26	0.11	0.110	1.10	1.06, 1.13
RA	0.55	0.052	4.22	3.78, 4.74	0.23	0.059	2.06	1.88, 2.26

Table 4. Performance of Estimators of Autocorrelation Parameters under Model with 10% Clustered Outliers

	Estimates of α_1				**Estimates of α_2**			
Estimator	Mean	MSE	Eff.	95% C. I.	Mean	MSE	Eff.	95% C. I.
No mean est.								
LS	0.11	0.397	1.00		0.07	0.182	1.00	
MLE	0.12	0.382	1.04	1.03, 1.05	0.07	0.163	1.12	1.10, 1.13
Robust M	0.16	0.332	1.20	1.15, 1.24	0.10	0.136	1.34	1.26, 1.41
RA	0.45	0.096	4.16	3.77, 4.61	0.22	0.068	2.68	2.40, 3.01
Mean polish								
LS	-0.13	0.724	1.00		-0.14	0.344	1.00	
MLE	-0.11	0.688	1.05	1.05, 1.06	-0.12	0.313	1.10	1.09, 1.11
Robust M	-0.05	0.590	1.23	1.21, 1.25	-0.07	0.271	1.27	1.24, 1.31
RA	0.11	0.372	1.94	1.88, 2.02	0.03	0.181	1.90	1.80, 2.02
Median pol.								
LS	-0.00	0.533	1.00		-0.01	0.234	1.00	
MLE	0.00	0.518	1.03	1.02, 1.03	-0.01	0.214	1.09	1.08, 1.10
Robust M	0.02	0.482	1.10	1.07, 1.14	0.01	0.184	1.27	1.20, 1.34
RA	0.15	0.333	1.60	1.51, 1.70	0.07	0.139	1.69	1.53, 1.86
GLS, Mean								
LS	0.08	0.438	1.00		0.03	0.202	1.00	
MLE	0.09	0.419	1.05	1.04, 1.06	0.04	0.182	1.11	1.10, 1.12
Robust M	0.13	0.363	1.21	0.92, 1.10	0.07	0.149	1.36	1.28, 1.43
RA	0.42	0.108	4.06	3.72, 4.47	0.21	0.068	2.97	2.67, 3.31
GLS, Trend								
LS	0.04	0.493	1.00		-0.00	0.235	1.00	
MLE	0.04	0.475	1.04	1.03, 1.04	0.00	0.214	1.10	1.09, 1.11
Robust M	0.08	0.413	1.19	1.15, 1.23	0.02	0.184	1.28	1.21, 1.34
RA	0.36	0.157	3.14	2.90, 3.41	0.14	0.100	2.34	2.14, 2.56

Table 5. Performance of Estimators of Autocorrelation Parameters under Model with 10% Random Outliers

	Estimates of α_1				**Estimates of α_2**			
Estimator	Mean	MSE	Eff.	95% C. I.	Mean	MSE	Eff.	95% C. I.
No mean est.								
LS	0.20	0.297	1.00		0.08	0.141	1.00	
MLE	0.20	0.288	1.03	0.99, 1.06	0.08	0.133	1.06	1.00, 1.10
Robust M	0.22	0.276	1.08	1.04, 1.11	0.09	0.123	1.15	1.08, 1.20
RA	0.47	0.093	3.19	2.88, 3.56	0.20	0.070	2.00	1.79, 2.22
Mean polish								
LS	-0.08	0.651	1.00		-0.13	0.313	1.00	
MLE	-0.08	0.648	1.00	0.99, 1.02	-0.13	0.304	1.03	1.00, 1.06
Robust M	-0.02	0.556	1.17	1.15, 1.19	-0.07	0.244	1.29	1.25, 1.32
RA	0.11	0.394	1.65	1.59, 1.73	0.04	0.165	1.90	1.80, 2.01
Median pol.								
LS	0.04	0.468	1.00		-0.02	0.202	1.00	
MLE	0.03	0.468	1.00	0.98, 1.03	-0.02	0.198	1.02	0.97, 1.08
Robust M	0.04	0.454	1.03	1.00, 1.06	-0.00	0.177	1.14	1.09, 1.19
RA	0.13	0.348	1.34	1.26, 1.50	0.05	0.147	1.37	1.26, 1.50
GLS, Mean								
LS	0.16	0.333	1.00		0.05	0.159	1.00	
MLE	0.16	0.327	1.02	0.99, 1.04	0.04	0.157	1.01	0.97, 1.06
Robust M	0.19	0.301	1.10	1.07, 1.13	0.07	0.155	1.17	1.11, 1.23
RA	0.46	0.097	3.44	3.14, 3.79	0.20	0.068	2.35	2.13, 2.59
GLS, Trend								
LS	0.08	0.434	1.00		-0.02	0.224	1.00	
MLE	0.07	0.438	0.99	0.96, 1.02	-0.05	0.164	0.97	0.94, 1.03
Robust M	0.15	0.351	1.24	1.20, 1.28	0.03	0.167	1.00	0.97, 1.02
RA	0.42	0.122	3.57	3.26, 3.93	0.18	0.080	2.80	2.54, 3.11

Table 6. Kempton-Howes Data, Mean-corrected Residuals

Estimator	$\hat{\alpha}_1$	$\hat{\alpha}_2$
Least squares	0.811 (0.050)	0.213 (0.078)
Max. Likelihood	0.812 (0.043)	0.241 (0.076)
Robust M	0.811 (0.052)	0.220 (0.080)
RA	0.812 (0.053)	0.239 (0.083)

Table 7. Kempton-Howes Data, Mean-polish and Median-polish Residuals

	Mean polish		Median polish	
Estimator	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
Least squares	0.696	-0.012	0.629	0.164
Max. Likelihood	0.784	0.015	0.745	0.209
Robust M	0.694	0.008	0.617	0.169
RA	0.694	0.056	0.646	0.214

Table 8. Cullis *et al*. Data, Mean-corrected and Trend-corrected Residuals

	Mean-corrected		Trend-corrected	
Estimator	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
Least squares	0.356 (0.039)	0.303 (0.039)	0.308 (0.040)	0.273 (0.039)
Max. Likelihood	0.359 (0.037)	0.307 (0.038)	0.312 (0.038)	0.274 (0.038)
Robust M	0.357 (0.039)	0.311 (0.039)	0.317 (0.040)	0.281 (0.040)
RA	0.363 (0.040)	0.338 (0.040)	0.326 (0.041)	0.310 (0.041)

Table 9. Cullis *et al*. Data, Mean-polish and Median-polish Residuals

	Mean polish		Median polish	
Estimator	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
Least squares	0.339	0.009	0.317	0.048
Max. Likelihood	0.370	0.004	0.362	0.042
Robust M	0.356	-0.013	0.331	0.016
RA	0.377	-0.011	0.366	0.018

Table 10. Lee-Rawlings Data, Mean-corrected and Trend-corrected Residuals

	Mean-corrected		Trend-corrected	
Estimator	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
Least squares	0.384 (0.146)	0.650 (0.125)	-0.045 (0.151)	0.306 (0.147)
Max. Likelihood	0.395 (0.131)	0.592 (0.110)	0.034 (0.142)	0.274 (0.129)
Robust M	0.410 (0.132)	0.649 (0.113)	0.033 (0.145)	0.394 (0.137)
RA	0.300 (0.151)	0.737 (0.115)	-0.025 (0.155)	0.455 (0.144)

Table 11. Lee-Rawlings Data, Mean-polish and Median-polish Residuals

	Mean polish		Median polish	
Estimator	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
Least squares	-0.037	0.005	-0.001	-0.007
Max. Likelihood	-0.054	0.106	0.011	0.121
Robust M	-0.040	0.009	-0.074	-0.020
RA	-0.071	0.044	-0.195	0.002