*Brazilian Journal of Probability and Statistics*
2009, Vol. 23, No. 2, 141–165
DOI: 10.1214/08-BJPS019


# Spatial ARMA models and its applications to image filtering

**Oscar Bustos**[a], **Silvia Ojeda**[a] **and Ronny Vallejos**[b,*]

[a]*Universidad Nacional de Córdoba*
[b]*Universidad Técnica Federico Santa María*

**Abstract.** The objective of this review paper is to summarize the main properties of the spatial ARMA models and describe some of the well-known methods used in image filtering based on estimation of spatial autoregressive models. A new proposal based on robust RA estimation is also presented. Previous studies have shown that under additive outliers the RA estimator is resistant to a small percentage of contamination and behaves better than the LS, M, and GM estimators. A discussion about how well these models fit to a digital image is presented. Some applications using real images are presented to illustrate how an image is filtered in practice.

## 1 Introduction

Image filtering in the presence of noise is a fundamental problem in image processing. The importance of this problem has been recognized for a long time period. Traditionally nonlinear filtering methods such as median filter (Jain (1989)) or $\alpha$-trimmed mean filter are used to remove impulse noise from an image. In the literature there are many other proposed algorithms. For a review, see Katsaggelos (1989) and Banham and Katsaggelos (1997).

In this paper we review some existing recursive algorithms to filter those images that can be well represented by spatial ARMA models on the plane. These algorithms are based on the fact that a contaminated image can be cleaned through robust estimation of the parameters of a suitable model for the image. Robust parameter estimation of spatial ARMA models has been studied in the context of image filtering. For example, Kashyap and Eom (1988) studied the M estimators for the parameters of spatial AR models. Later Allende, Galbiati and Vallejos (2001) proposed the generalized M (GM) estimator. In the experimental results carried out by these authors it was reported that the GM estimator performed better than the M estimator when the process is contaminated with additive outliers. Subsequently, Ojeda, Vallejos and Lucini (2002) proposed the use of the RA estimators as a robust alternative to the M and GM estimators for first order spatial AR models.

[*]Work was performed while author was at Universidad de Valparaíso.

*Key words and phrases.* Spatial autoregressive model, robust estimators, image filtering algorithm.

Received April 2008; accepted December 2008.

As a result of the robust estimation process Kashyap and Eom (1988) suggested an algorithm to filter isolated impulse noise images. They used one extension of the M estimators to attenuate the effect of outliers on the residuals of the model. This algorithm was later developed using the GM estimators. Finally Vallejos and Mardesic (2004) introduced classification techniques as a previous stage in the algorithm developed by Allende, Galbiati and Vallejos (2001). This significantly reduces the number of parameters to estimate, which impact the computational time required to filter an image.

In this article we briefly review the filtering algorithms mentioned above. Also a new filtering algorithm is suggested. A motivation is given in Section 2. Some preliminary notation and the spatial autoregressive models are presented in Section 3. In Section 4 the robust parameter estimation of spatial autoregressive models is discussed. Two examples are developed in Section 5 to illustrate how to fit a spatial AR model in practice. In Section 6 the image filtering algorithms are described. A new algorithm based on RA estimation is also presented. Applications using real images are introduced to illustrate the filtering process. Finally, the main conclusions are summarized in Section 7.

## 2 Motivation

Most of the images of interest, for example, the images of cultivated fields and concentration of population are naturally rich in texture, level of gray, etc. The same thing happens to the images of geographical regions that allow the making of maps and, in general, almost all the images of the earth. In this sense the AR-2D model has two main properties. First, simulation experiments have shown that this model is adequate to represent a diversity of real sceneries. Second, the AR-2D model does not require a large number of parameters to represent different real sceneries (parsimony).

To illustrate the capability of this model to represent different textures, in Figure 1 we show six images generated from AR-2D models with different sets of parameters.

In recent times the spatial autoregressive models have been extensively used to represent images (Bennett and Khotanzad (1999)). In particular, the first-order AR-2D model has been used to represent real scenarios (see Kashyap and Eom (1988)). Theoretical properties of this model were studied by Basu and Reinsel (1993). They derived the correlation structure of the model and the maximum likelihood estimators of the parameters. The first-order AR-2D model is able to represent a number of different textures as is shown in Figure 2.

## 3 Spatial ARMA models

In this section we assume that the underlying image generation process is 2D autoregressive moving average. Thus a pixel in the original image $X(s)$, where $s$

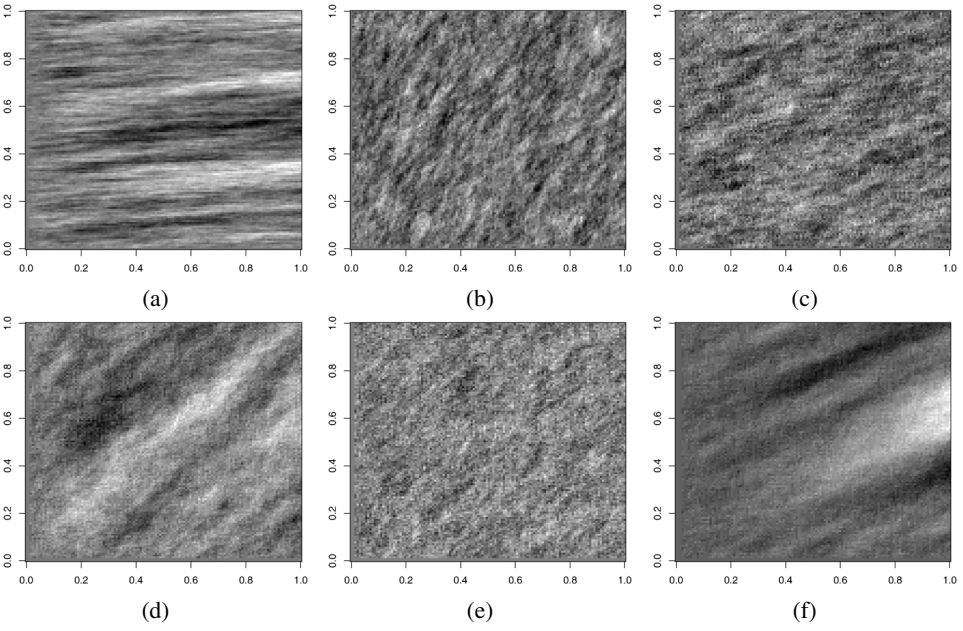



**Figure 1** *Images* (a)–(f) *are realizations of an AR-2D process with 2, 3, 4, 5, 6, and 7 parameters, respectively.*

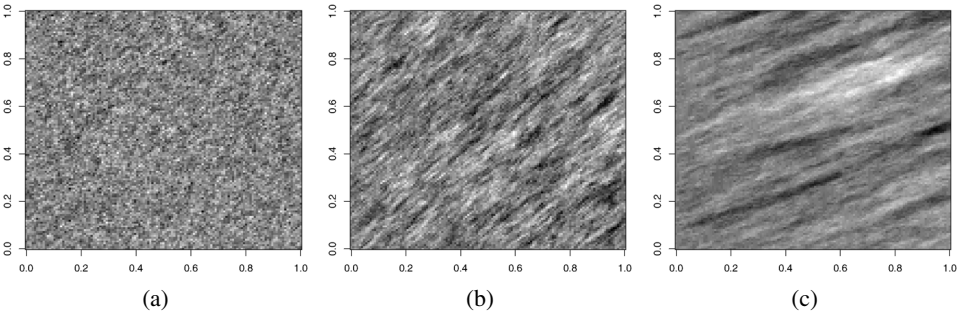



**Figure 2** *Images generated from the model* $X(i, j) = \phi(1, 0)X(i - 1, j) + \phi(0, 1)X(i, j - 1) + \phi(1, 1)X(i - 1, j - 1) + \varepsilon(i, j)$.

represents a spatial position vector, can be modeled as follows:

$$X(s) = \sum_{k=1}^{P} \phi_k(s)X(s + h_k) + \sum_{j=1}^{Q} \theta_j(s)\varepsilon(s + h_j), \tag{3.1}$$

where a particular pixel at site $s$ is predicted as a linear combination of pixels in the current frame plus pixels that are in a neighborhood of a white noise $\varepsilon(s)$, $\phi_k(s)$ and $\theta_j(s)$ are the parameters of the model and the vectors $h_k$, $h_j$ define the support or neighborhood of the model. Note that the model parameters $\phi_k(s)$

and $\theta_j(s)$ are nonstationary since they are a function of position $s$. These models have been considered to extend the models used in time series. However, from a classical point of view the estimation process is extremely difficult because of the large number of unknown parameters to be estimated. Alternatively, Bayesian treatment of models like (3.1) and an algorithm for filling missing gaps in images were suggested by Kokaram (2004).

In order to represent images using models that are statistically treatable, it is common to assume that an image can be represented using a few number of parameters. In this context, three broad categories of spatial models that have been studied are the simultaneous autoregressive (SAR) models (Whittle (1954)), the conditional autoregressive (CAR) models (Besag (1974)), and the moving average (MA) models (Haining (1978)). A spatial ARMA model is described by the equation (Martin (1996))

$$\Phi(B_1, B_2)X(i, j) = \Theta(B_1, B_2)\varepsilon(i, j), \tag{3.2}$$

where the two-dimensional backward operators $\Phi(B_1, B_2)$ and $\Theta(B_1, B_2)$ are given by

$$\Phi(B_1, B_2) = \sum_k \sum_l \phi(k, l) B_1^k B_2^l,$$

$$\Theta(B_1, B_2) = \sum_k \sum_l \theta(k, l) B_1^k B_2^l,$$

with $B_1 X(i, j) = X(i - 1, j)$ and $B_2 X(i, j) = X(i, j - 1)$ and $\varepsilon(i, j)$ are independent random variables with $\mathbb{E}[\varepsilon(i, j)] = 0$ and $\mathrm{Var}[\varepsilon(i, j)] = \sigma^2$.

Similarly to the time series case there are conditions on the two-dimensional polynomials to have stationarity and invertibility. For stationarity it is enough to assume that the complex valued polynomial $\Phi(z_1, z_2)$ is not zero for any $z_1$ and $z_2$ which simultaneously satisfy $|z_1| < 1$ and $|z_2| < 1$.

In practice, the spatial ARMA models have been used in several applications related to agricultural statistics. For example, in Cullis and Glesson (1991) the spatial separable processes were introduced. Later, these processes were used to analyze yield trials in the context of incomplete block designs (Grondona et al. (1996)). Basu and Reinsel (1993) studied the spatial unilateral first-order ARMA model to examine regression models with spatially correlated errors. Martin (1996) presented some results on separable ARMA models. These processes are defined by their correlation function being the product of one-dimensional correlation functions, and were studied first by Quenoulle (1949).

### 3.1 Neighborhoods in $Z^2$

In time series, there is a natural neighbor structure induced by the existing total order of $Z$ (the set of all past values of $t \in Z$ is the set of all integers that are less than $t$). However, for points on the plane, for instance $(m, n) \in Z^2$, there are

several different notions of neighborhood. In general, definitions of neighborhood of a point $(m, n)$ on the plane are motivated by the physical acquisition system of the data. This is the case of images that have been captured by satellites. Here we describe the most frequent definitions of neighborhood found in the literature (see Jain (1989)).

**Definition 1.** For all $(m, n) \in Z^2$:

(a) $S_1(m, n) = \{(k, l) \in Z^2 : (k, l) \neq (m, n)\}$ is called noncausal prediction region at $(m, n)$.

(b) $S_2(m, n) = \{(k, l) \in Z^2 : k < m)\} \cup \{(m, l) \in Z^2 : l \neq n\}$ is called semicausal prediction region at $(m, n)$.

(c) $S_3(m, n) = \{(k, l) \in Z^2 : k < m\} \cup \{(m, l) \in Z^2 : l > n\}$ is called causal prediction region at $(m, n)$ (alternatively called nonsymmetric half plane (NSHP)).

(d) $S_4(m, n) = \{(k, l) \in Z^2 : k \leq m, l \leq n)\} - \{(m, n)\}$ is called strongly causal region at $(m, n)$.

In general, a realization of a stochastic process defined on the plane is observed on one of the prediction regions described above. This motivates the following definition.

**Definition 2.** For all $(m, n) \in Z^2$, and for all $M \in \mathbb{N}$:

(a) $W_{1,M}(m, n) = \{(k, l) \in S_1(m, n) : |k - m| \leq M, |l - n| \leq M\}$ is called noncausal prediction window of order $M$ at $(m, n)$.

(b) $W_{2,M}(m, n) = \{(k, l) \in S_2(m, n) : m - M \leq k < m, |l - n| \leq M\} \cup \{(m, l) : 0 < |l - n| \leq M\}$ is called semicausal prediction window of order $M$ at $(m, n)$.

(c) $W_{3,M}(m, n) = \{(k, l) \in S_3(m, n) : m - M \leq k < m, |l - n| \leq M\} \cup \{(m, l) : n < l \leq n + M\}$ is called causal prediction window of order $M$ at $(m, n)$.

(d) $W_{4,M}(m, n) = \{(k, l) \in S_4(m, n) : m - M \leq k \leq m, n - M \leq l \leq n\}$ is called strongly causal prediction window of order $M$ at $(m, n)$.

### 3.2 Spatial AR processes

Similarly to the one-dimensional case, if $\Phi(B_1, B_2) = 1$, then the process is called moving average. If $\Theta(B_1, B_2) = 1$, the process is called autoregressive.

The simultaneous AR model is defined as

$$\Phi(B_1, B_2)X(i, j) = \varepsilon(i, j). \tag{3.3}$$

Particular cases of these processes are well known in the literature. For example, the isotropic first-order simultaneous autoregressive scheme (Cliff and Ord (1981)) described by the equation

$$X(i, j) = \phi\big(X(i-1, j) + X(i+1, j) + X(i, j-1) + X(i, j+1)\big) + \varepsilon(i, j), \tag{3.4}$$

with $|\phi| < 1/4$ to ensure stationarity. Whittle (1954) has shown that given a set of autocorrelations of a SAR process, there is a unique representation of the process in which $X(i, j)$ can be expressed as an autoregression on $X(i, j)$ $(i = 0$ and $j < 0)$ and $X(i, j)$ $(i < 0,$ and $j$ unrestricted).

A special case of the unilateral AR models where the value at the site $(i, j)$ is a finite autoregression on the strongly causal region $S_4(i, j)$ was examined by Tjostheim (1978). In two dimensions this model becomes

$$X(i, j) = \sum_{k=0}^{p_1} \sum_{l=0}^{p_2} \phi(k, l) X(i - k, j - l) + \varepsilon(i, j), \tag{3.5}$$

with $\phi(0, 0) = 0$. This model leads to the Wold-type representation

$$X(i, j) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \phi(k, l) \varepsilon(i - k, j - l). \tag{3.6}$$

See Tjostheim (1978) for further discussion and properties of related purely nondeterministic processes.

Notice that the study of the invertibility of the complex valued function $\Phi(z_1, z_2)$ is not a simple problem. To get the representation

$$X(i, j) = \Phi(B_1, B_2)^{-1} \varepsilon(i, j), \tag{3.7}$$

the function $\Phi(z_1, z_2)^{-1}$ needs to be written using a Laurent expansion as

$$\Phi(z_1, z_2)^{-1} = \sum_{k,l} \psi_{kl} z_1^k z_2^l. \tag{3.8}$$

Basu and Reinsel (1993) investigated the correlation structure of a general first-order autoregressive process of the form,

$$\begin{aligned} X(i, j) = {} & \phi(1, 0) X(i - 1, j) + \phi(0, 1) X(i, j - 1) \\ & + \phi(1, 1) X(i - 1, j - 1) + \varepsilon(i, j). \end{aligned} \tag{3.9}$$

They showed that the conditions,

$$\begin{aligned} |\phi(k, l)| &< 1, \\ 1 - \phi(0, 1)^2 &> |\phi(1, 0) + \phi(0, 1)\phi(1, 1)|, \\ \big(1 + \phi(1, 0)^2 - \phi(0, 1)^2 - \phi(1, 1)\big)^2 &> 4\big(\phi(1, 0) + \phi(0, 1)\phi(1, 1)\big)^2, \end{aligned}$$

guarantee the stationarity of the process. Because of the uniqueness of representation (3.8), a multinomial expansion for the function $(1 - \phi(1, 0)z_1 - \phi(0, 1)z_2 - \phi(1, 1)z_1 z_2)^{-1}$, can be used to get the convergent representation

$$X(i, j) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{r=0}^{\infty} \frac{(k + l + r)!}{k! l! r!} \phi(1, 0)^k \phi(0, 1)^l \phi(1, 1)^r \varepsilon(i - k - r, j - l - r).$$

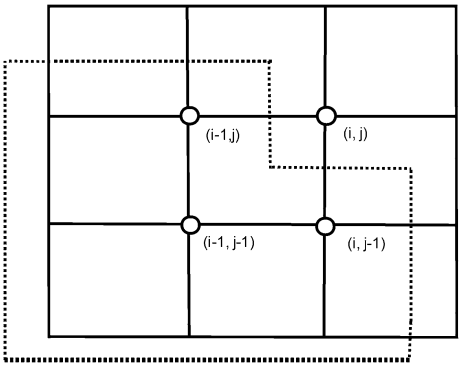



**Figure 3** *Prediction window for the first-order spatial AR model.*

The prediction window in this case is shown in Figure 3.

One way to avoid the analysis of (3.8) is to consider an additional assumption on the parameters of the model. For example, if $\phi(1,1) = -\phi(1,0)\phi(0,1)$, the multiplicative (separable) process simplifies to

$$X(i,j) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\phi(1,0)^k\phi(0,1)^l\varepsilon(i-k, j-l).$$

This model has been investigated by Martin (1979, 1990) who argued that it has many practical uses. Martin (1996) have shown that axially symmetric quadrant ARMA processes are separable.

We also note that one way to avoid expansion (3.8) is to consider more general models as the CAR and SAR models frequently used in spatial statistics. Briefly we can say that a CAR model is described by

$$\big(X(s_i)|X(s_j),\, j \neq i\big) \sim N\left(\mu_i + \rho\sum_{j\neq i} b_{ij}X(s_j),\, \tau_i^2\right), \qquad (3.10)$$

$$i = 1, 2, \ldots, n,$$

where $\rho$ is a parameter that determines the direction (positive or negative) and magnitude of the spatial neighborhood effect, $b_{ij}$ are weights that determine the relative influence of location $j$ on location $i$, and $\mu_i$ and $\tau_i^2$ are the conditional mean and variance, respectively.

Alternatively to CAR specification, the SAR models can be defined. We let $\boldsymbol{\varepsilon}$ induce a distribution for $\mathbf{X}$. Analogous to (3.10), now

$$X_i = \sum_j b_{ij}X_j + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \qquad (3.11)$$

with $\varepsilon_i \sim N(0, \sigma_i^2)$. We notice that CAR and SAR models are not necessarily stationary models. Thus representations (3.10) and (3.11) provides a wider class

of models that can be used even for nonrectangular lattices (see Griffith (1988); Baneerje, Carlin and Gelfand (2004), pages 79–82).

### 3.3 Spatial MA processes

Alternatively to the spatial autoregressive processes, the moving average models are of great theoretical and practical importance, and have been studied in different contexts. The relevance of the MA models is mainly due to the great popularity of the Wold-type decomposition of a regular homogeneous random field. In this decomposition, the purely indeterministic random field component has a unique (but unknown) MA representation. Applications of the Wold-type decomposition include modeling natural textures (Francos, Narasimhan and Woods (1995)), image modeling and retrieval (Liu and Piccard (1996)), and more recently models for space–time adaptive processing radar systems (Francos and Nehorai (2003)). In addition, the MA models have been considered in the context of image segmentation and restoration problems (Krishnamurthy, Woods and Francos (1996)).

It is well known that MA representation models for random fields are less informative than AR or ARMA models. However, from an estimation perspective it is very interesting to know the performance of new proposals for the estimation of the parameters of a model. Alternatively, this problem can be tackled from a Bayesian point of view. In the context of robust methods, most of the robust estimators introduced for AR models in time series do not perform well for MA or ARMA models (Allende and Heiller (1992)).

A linear representation of a spatial moving average model was considered in Francos and Friedlander (1998). In fact, $\mathbf{X} = B\mathbf{e}$, where $\mathbf{X}$ is a vector containing the spatial observations over a completely ordered set; $\mathbf{e}$ is a spatial white noise and $B$ is the matrix containing the parameters of the model according to the moving average structure of the process. The form of $B$ in the context of toroidal approximation of a bidimensional autoregressive model was studied in Kashyap and Chellappa (1983). A complete treatment of the linear representation for multiespectral autoregressive models (MSAR) can be found in Bennet and Khotanzad (1999). Usually, the structure of $B$ has a strong dependency on the prediction window and on the boundary conditions.

Let $\{X(i,j) : (i,j) \in Z^2\}$ be a real-valued random field. Let us assume a total order on the discrete lattice such that $(i,j) \preceq (s,t) \iff (i,j) \in \{\{(k,l) : k = s, l \leq t\} \cup \{(k,l) : k < s, -\infty < l < \infty\}\}$. Then an infinite two-dimensional moving average model can be represented by

$$X(i,j) = \sum_{(0,0) \preceq (k,l)} \theta(k,l) e(i-k, j-l), \tag{3.12}$$

where the sequence $\{e(i,j) : (i,j) \in Z^2\}$ is a white-noise process with respect to the above total order with variance $\sigma_e^2$ and with $\theta(k,l)$ denoting the parameters of the model. As usual, we will assume that $\theta(0,0) = 1$.

In practice, the observed random process is of finite dimensions. For the support $D = \{(i, j) : 0 \leq i \leq S - 1, 0 \leq j \leq T - 1\}$, the observed process is the set $\{X(i, j) : (i, j) \in D\}$. In particular, for $(k, l) \in S_{(N,M)}$, where $S_{(N,M)} = \{(i, j) : 0 \leq i \leq N, 0 \leq j \leq M\}$ and $N, M$ are a priori known natural numbers, model (3.12) can be written in the form

$$X(i, j) = \sum_{(k,l)\in S_{(N,M)}} \theta(k, l)e(i - k, j - l). \tag{3.13}$$

The model obtained by the relation in (3.13) along with the assumption of white noise on the errors is called spatial MA model.

The estimation of the parameters of a model like (3.12) has been widely studied (Whittle (1954)). More recently, this general derivation was presented for the case of noncausal AR models and NSHP AR models (Isaksson (1993)). Since the ML method requires an iterative and computationally intensive procedure, it becomes computationally prohibitive even for moderate size data fields. In Francos and Friedlander (1998) a close form exact expression for the Cramer–Rao lower bound was derived. Using the expressions of the covariance matrix in terms of the moving average model parameters they derived a maximum likelihood algorithm that offers an increasingly attractive alternative to ML estimation. Later, in Vallejos and García-Donato (2006) the authors focused on a Bayesian approach to analyze the MA model. By using simulation, they showed that the correlation structure of MA model is seriously affected when data contains additive contamination. Nevertheless, the posterior distribution of correlations seems to be resistent to the innovation contamination. In that work, they also proposed a more general class of moving average models which deal with contaminated data (CMA model). The main idea is to consider a MA model with $\mathbf{e}$ being a mixture of Normal distributions. The resulting model allows the presence of errors produced under a distribution with a different variance than the usual one. Important characteristics of this new model such as the correlation function were studied. They suggested how to handle this model within the Bayesian framework. First, a prior distribution arising from prior information about the strength of contamination was proposed. Second, some strategies for sampling from a posterior distribution in a MCMC context were discussed. The results obtained in numerical examples have shown the goodness of the CMA model under contaminated data.

Since the models presented above will be useful to represent images let us consider $X(s)$ being the intensity at coordinate location $s$ measured in a finite scale of gray. Precisely, from now on we consider $\mathbf{X} = \{X(s) : s \in S_4(0, 0)\}$ being a two dimensional real random process over the probability space $(\Omega, F, \mu)$.

## 4 Robust parametric estimation

The Yule–Walker and least squares estimators for two-dimensional causal autoregressive models were studied from an asymptotic point of view by Eunho and

Newton (1993). As a result, an explicit expression for the bias of Yule–Walker estimators was obtained. Simulation experiments have shown that the performance of the least squares and asymptotically unbiased Yule–Walker estimators are remarkably similar.

In general, it is not easy to deal with the exact likelihood of spatial autoregressive models. In the literature the exact likelihood estimator has been studied for specific models and neighbor structures. An appealing feature of the first-order spatial ARMA model (3.9) is that the exact likelihood function can be obtained in a convenient computational form. Basu and Reinsel (1993) examined the implementation of the ML estimator. Using a Monte Carlo study they explored the accuracy of the ML estimations.

It is well known that the ML estimators are very sensitive to the outliers (Martin (1980)). This motivated the introduction of several alternative estimators to attenuate the impact of contaminated observations on the estimations. Most of these proposals are natural extensions of the robust estimators studied in time series.

Here we use model (3.9) to describe the well-known robust estimators, however, a more general treatment for AR and MA models can be found in: Kashyap and Eom (1988); Allende, Galbiati and Vallejos (2001); Ojeda, Vallejos and Lucini (2002); Vallejos and Garcia-Donato (2006), and Bustos et al. (2008).

In time series, innovation outliers (IO) and additive outliers (AO) are well known (Fox (1972)). The same notion of data contamination has been studied for spatial processes (Ojeda (1999)). A more recent discussion about types of contamination in the context of time series can be found in Chang, Tiao and Chen (1988), Chen and Lui (1993), and Tsay (1988). The definitions of outliers have been extended to multivariate framework and the effects of multivariate outliers on the joint and marginal models have been examined by Tsay, Peña and Pankratz (2000).

### 4.1 Robust estimators

Several robust estimators have been defined for models containing a finite number of parameters. In order to simplify the notation, in this work we describe the most commonly used robust estimators for a model like (3.9).

Notice that model (3.9) can be rewritten in the linear model form

$$X(i,j) = \boldsymbol{\phi}^T Z(i,j) + \varepsilon(i,j),$$

where $\boldsymbol{\phi}^T = (\phi(1,0), \phi(0,1), \phi(1,1))$ is a parameter vector and

$$Z(i,j)^T = \big(X(i-1,j), X(i,j-1), X(i-1,j-1)\big). \tag{4.1}$$

To obtain the LS estimator of the parameters of model (3.9) we need to minimize the function

$$\sum_{i,j} [X(i,j) - \boldsymbol{\phi}^T Z(i,j)]^2,$$

with respect to $\boldsymbol{\phi}$. The LS estimator is nonrobust when the process is contaminated with innovation or additive outliers (Allende, Galbiati and Vallejos (1998)).

Similarly, the class of M estimators that was proposed for causal autoregressive processes (Kashyap and Eom (1988)), defined by minimizing the function of a finite sample of observations

$$Q(\boldsymbol{\phi}, \sigma) = \sum_{i,j} \left[ \rho\left( \frac{X(i,j) - \boldsymbol{\phi}^T Z(i,j)}{\sigma} \right) + \frac{1}{2} \right] \sigma, \tag{4.2}$$

is robust for innovation outliers, when the function $\rho$ is a differentiable function, convex, symmetric with respect to the origin, with bounded derivative, and such that $\rho(0) = 0$. However, the M estimators are very sensitive when the process is contaminated with additive outliers. This suggested the introduction of robust estimators that are less sensitive to the additive outliers. Allende, Galbiati and Vallejos (2001) developed the GM estimators. A GM estimator of $\boldsymbol{\phi}$ is the solution to the problem of minimizing the nonquadratic function defined by

$$Q(\phi, \sigma) = \sum_{i,j} l_{ij} t_{ij} \left[ \rho\left( \frac{X(i,j) - \boldsymbol{\phi}^T Z(i,j)}{l_{ij}\sigma} \right) + \frac{1}{2} \right] \sigma,$$

where $\rho$ is as in (4.2). Equivalently, the robust GM estimator can be obtained by solving the equation

$$\sum_{ij} t_{ij} \psi \left[ \frac{X(i,j) - \boldsymbol{\phi}^T Z(i,j)}{l_{ij}\sigma} \right] Z^T(i,j) = 0, \tag{4.3}$$

where the influence function $\psi$ is bounded and continuous, and $t_{ij}$ and $l_{ij}$ are weights corresponding to the respective $Z(i,j)$. The principal types of GM estimators are:

(1) Mallows type, where $l_{ij} = 1$ and $t_{ij} = \psi(b_{ij})/b_{ij}$, with $b_{ij} = p^{-1} \times Z(i,j)^T \widehat{C}^{-1} Z(i,j)$, where $\widehat{C}^{-1}$ is a robust estimate of $C^{-1}$ and $C^{-1}$ is a priori unknown covariance matrix for the AR-2D process.

(2) Schweppe type, $l_{ij} = t_{ij} = \psi(b_{ij}/c_r)/(b_{ij}/c_r)$, where $c_r$ is a tuning constant used to yield high efficiencies at the Gaussian AO model.

The estimation of $\sigma$ can be obtained independently using a preliminary robust estimator of the scale parameter.

Alternatively to the GM estimators, Ojeda (1999) and Ojeda, Vallejos and Lucini (2002) introduced the RA estimators for spatial autoregressive processes. This estimator was introduced first in Bustos and Yohai (1986) in the context of ARMA models in time series.

Let $X$ be a zero mean AR-2D process with $\mathrm{Var}(\varepsilon(m,n)) = \sigma^2$. Assume that $X$ is observed on a strongly causal square window of order $M$, $W_M = \{(k,l) \in S : 0 \le k, l \le M\}$, where $S$ is an infinite strongly causal prediction neighborhood

(Guyon (1995)). Let us define $W_M \setminus T = \{(m,n) \in W_M : (m-1, n-1) \in W_M\}$. The residual of order $(m,n)$ in $\boldsymbol{\phi}$ of $X$ is

$$r(m,n) = \begin{cases} -\sum\limits_{(k,l)\in T'} \phi(k,l) X(m-k, n-l), & (m,n) \in (W_m \setminus T), \\ 0, & \text{otherwise}, \end{cases} \tag{4.4}$$

where $T' = T \cup \{(0,0)\}$ and $\phi(0,0) = -1$. Defining the coefficients

$$p_{\boldsymbol{\phi}}(k,l,r) = \frac{(k+l+r)!}{k!\, l!\, r!} \phi(1,0)^k \phi(0,1)^l \phi(1,1)^r,$$

the RA estimator $\widehat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ is defined by the following equations

$$\sum_{k,l,r=0}^{\infty} p_{\widehat{\boldsymbol{\phi}}}(k,l,r) \sum_{(m,n)\in(W_M\setminus T)} \eta\left(\frac{r(m,n)}{\hat{\sigma}}, \frac{r(m-i-k-r, n-j-l-r)}{\hat{\sigma}}\right) = 0, \quad \text{for all } (i,j) \in T, \tag{4.5}$$

$$\sum_{(m,n)\in(W_M\setminus T)} \psi\left(\frac{r(m,n)}{\hat{\sigma}}\right) = 0, \tag{4.6}$$

where $\sigma$ is estimated independently by

$$\hat{\sigma} = Med\big(|r(m,n)| : (m,n) \in (W_M \setminus T)\big)/0.6745, \tag{4.7}$$

$\eta$ is a continuous, bounded, and odd function in two variables and $0.6745 = Med(|Y|)$, where $Y$ is a standard normal random variable. Two possible choices for $\eta$ are $\eta(u,v) = \psi_1(u \cdot v)$, and $\eta(u,v) = \psi_2(u) \cdot \psi_3(v)$, where the most commonly used $\psi_1$, $\psi_2$, and $\psi_3$ functions are the following:

(1) Huber's $\psi$ function

$$\psi_{H,c}(u) = sgn(u) \min(|u|, c)$$

(2) Tuckey's $\psi$ function

$$\psi_{T,k}(u) = \begin{cases} u\left[1 - \left(\frac{u}{k}\right)^2\right]^2, & |u| \le k, \\ 0, & \text{otherwise}. \end{cases}$$

(3) Hampel's $\psi$ function

$$\psi_{HA,a,b,d}(u) = \begin{cases} u, & |u| \le a, \\ a \cdot sgn(u), & a < |u| \le b, \\ a\dfrac{d-|u|}{d-b} sgn(u), & b < |u| \le d, \\ 0, & d < |u|. \end{cases}$$

(4) Andrews' $\psi$ function

$$\psi_{A,n}(u) = \begin{cases} \sin\left(\frac{u}{n}\right), & |u| \le n\pi, \\ 0, & \text{otherwise}, \end{cases}$$

where

$$sgn(u) = \begin{cases} 1, & u > 0, \\ -1, & u < 0, \\ 0, & u = 0. \end{cases}$$

For instance, typical values for the adjusting constant $c$ in $\psi_{H,c}(u)$ are between 1.5 and 2.0, and for $k$ in $\psi_{T,k}(u)$ are between 4.5 and 6. The values of constants $c$ and $k$ should be such that, with the noncontaminated model, the loss of efficiency of the RA estimator with respect to the maximum likelihood estimator does not exceed 5%.

### 4.2 Monte Carlo results

Several simulation experiments have been carried out to explore the performance of the estimators defined above. A first preliminary study by Allende and Heiller (1992) revealed that the GM performs better than the M estimator under additive outliers. Later, the RA estimator was compared with the M, GM, and LS estimators (Ojeda (1999) and Ojeda, Vallejos and Lucini (2002)). In all cases the GM and RA estimators perform much better than the M and LS estimators. The performance of the RA estimator was better than that of the GM estimator. These studies were carried out for a process containing two parameters since the implementation of the RA estimator depends on the dimension of the parameter space. Subsequently, simulation studies were carried out to explore the behavior of the robust estimators in models with more than two parameters (Vallejos, Ojeda and Bustos (2008)). The experiments showed similar patterns as before. The GM and RA estimators were highly superior to the M and LS estimators.

## 5 Model fitting

To gain more insight into the estimation process, in this section we present two examples. The first one illustrates the model selection problem. We suggest the use of cross-validation to find a suitable model for the data. The second example illustrates the local approximation of images by using 2D unilateral AR processes. Real images are compared with the images generated from fitted models. This procedure has been frequently used in the context of image filtering (Kashyap and Eom (1988) and Allende, Galbiati and Vallejos (2001)).

**Example 1.** Consider the image shown in Figure 4. This is a Landsat TM5 image from Sierra Comechingones in the Southwest of Córdoba, Argentina, acquired in December'92, two months after a wildfire. The burned zone corresponds to the darkest area in the middle of the image. Here, we illustrate the model selection

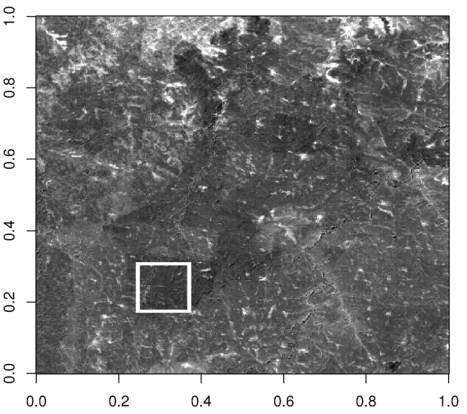



**Figure 4** *Landsat TM5 image from Sierra Comechingones, Córdoba, Argentina.*

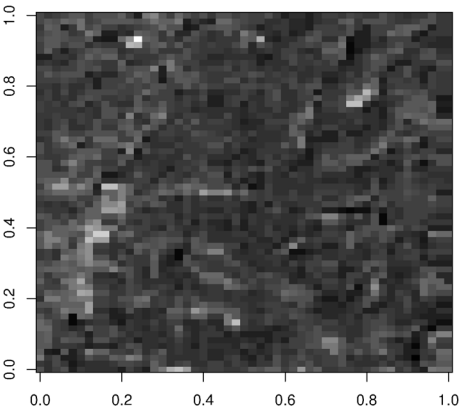



**Figure 5** *Zoom of the image delimited by a white square in Figure 4.*

problem of spatial autoregressive models, following the guidelines given in Rukhin and Vallejos (2008) in the context of flammability of polymers.

Let us consider the following AR models:

$$\text{Model 1: } X(i,j) = \phi_1 X(i-1,j) + \varepsilon(i,j),$$

$$\text{Model 2: } X(i,j) = \phi_2 X(i,j-1) + \varepsilon(i,j),$$

$$\text{Model 3: } X(i,j) = \phi_1 X(i-1,j) + \phi_2 X(i,j-1) + \varepsilon(i,j),$$

$$\text{Model 4: } X(i,j) = \phi_1 X(i-1,j) + \phi_2 X(i,j-1) + \phi_3 (i-1,j-1) + \varepsilon(i,j),$$

where in each case $\{\varepsilon(i,j)\}$ is a white noise. The goal is to fit one of those models to the data located inside of the $50 \times 50$ white square shown in Figure 4. A zoom corresponding to that area is shown in Figure 5.

**Table 1** *LS estimators for models 1–4*

| | $\hat{\phi}_1$ | $\hat{\phi}_2$ | $\hat{\phi}_3$ |
|---|---|---|---|
| Model 1 | 0.9702 | | |
| Model 2 | | 0.9715 | |
| Model 3 | 0.4895 | 0.5007 | |
| Model 4 | 0.4799 | 0.4918 | 0.0186 |

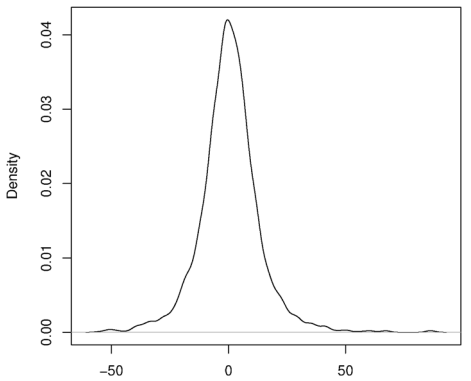



**Figure 6** *Density estimate of the residuals generated from model 4.*

Using (4.4) and the LS estimators for model 4 (see Table 1), the residuals $\hat{r}(i, j)$ were computed for each location $(i, j)$. To gain more insight into the distribution of the residuals, the density estimate of the residuals was plotted (see Figure 6). We observe that the distribution of the residuals looks symmetric about zero and since there is higher kurtosis than in the normal case we fitted a Laplace distribution. The maximum likelihood estimators for the location and scale parameters were $\mu = 0.49$ and $\sigma = 12.56$. Using random numbers from Laplace distribution a new image was created as

$$\hat{X}(i, j) = \hat{\boldsymbol{\phi}}^T Z(i, j) + \hat{r}(i, j), \tag{5.1}$$

where $\hat{r}(i, j)$ represent a random number from the Laplace distribution corresponding to location $(i, j)$ and $Z(i, j)$ is given in (4.1).

The original and estimated images can be compared using a measure of discrepancy. One way to accomplish this is through the mean square error of prediction (MSE) (Vallejos and Garcia-Donato (2006)) given by

$$\text{MSE} = \frac{1}{N} \sum_{i,j} \big(X(i, j) - \hat{X}(i, j)\big)^2, \tag{5.2}$$

where $N$ represents the total number of pixels in each image. In this experiment the MSE was used in the context of cross-validation (Grondona et al. (1996)) to compare models 1–4. Using (5.1) in each case, images of size $30 \times 30$ were generated

**Table 2** *MSE for models 1–4*

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| MSE | 780.1253 | 785.2043 | 793.4743 | 672.1097 |

from models 1–4 and situated in the center of the original ($50 \times 50$) square. These models were used to predict the locations observed in the original image (5.1) but outside of the $30 \times 30$ square region. For each model the MSE was computed over the observations located in the area surrounding the $30 \times 30$ region. These values are shown in Table 2.

The smallest MSE value corresponds to model 4. We will come back to the idea of generating images from equations similar to (5.1) in Section 6.

**Example 2.** We present now a real data example to illustrate the local approximation of images by using 2D unilateral AR processes. The goal is to graphically show that the 2D unilateral AR processes are useful and expressive processes to represent a number of different real scenarios. In this context, the following question arises: is it possible to represent any image by using a 2D AR unilateral process? To heuristically answer this question, here we consider an algorithm that defines what we call a local AR-2D approximated image by using blocks. This algorithm was originally defined by Kashyap and Eom (1988) and later adapted by Bustos et al. (2008).

Suppose that a real image is available. Then divide the original image in blocks, that is, $8 \times 8$ blocks.

**Algorithm.**

For each block:

(1) Compute the least squares estimators of the parameters of the AR-2D model described by

$$X(i, j) = \phi(1, 0)X(i - 1, j) + \phi(0, 1)X(i, j - 1) + \varepsilon(i, j).$$

Let us call the estimators $\widehat{\phi}(1, 0)$ and $\widehat{\phi}(0, 1)$.

(2) Generate the image

$$\widehat{X}(i, j) = \widehat{\phi}(1, 0)X(i - 1, j) + \widehat{\phi}(0, 1)X(i, j - 1),$$

and compute the approximated image $\tilde{X}$ of X as

$$\tilde{X}(m, n) = \widehat{X}(m, n) + \overline{X},$$

where $\overline{X}$ is the mean of the original image $X$.

In order to compare the images produced by the previous algorithm, in Figure 7 we show 3 images. Image (a) is an artificial image ($512 \times 512$) taken from The USC-SIPI Image Database. Image (b) is the AR-2D approximated image using blocks

(a) (b) (c)

**Figure 7** (a) *Original image,* (b) *image generated by the AR-2D models,* (c) *difference between the original and approximated images.*

of size $8 \times 8$, and image (c) is the difference between the original and the AR-2D approximated image. To quantify the discrepancy between the original and fitted image the mean square error given in (5.2) was computed, where $N$ is the total number of pixels excluding the pixels belonging to the boundaries that do not have neighbors on the top or on the left. In this case MSE $= 31.6343$.

In Figure 7(c), we see that the borders have been highlighted. This is a prominent and unexpected result that can be matter of further research. An image segmentation method could be implemented using the information contained in the difference image.

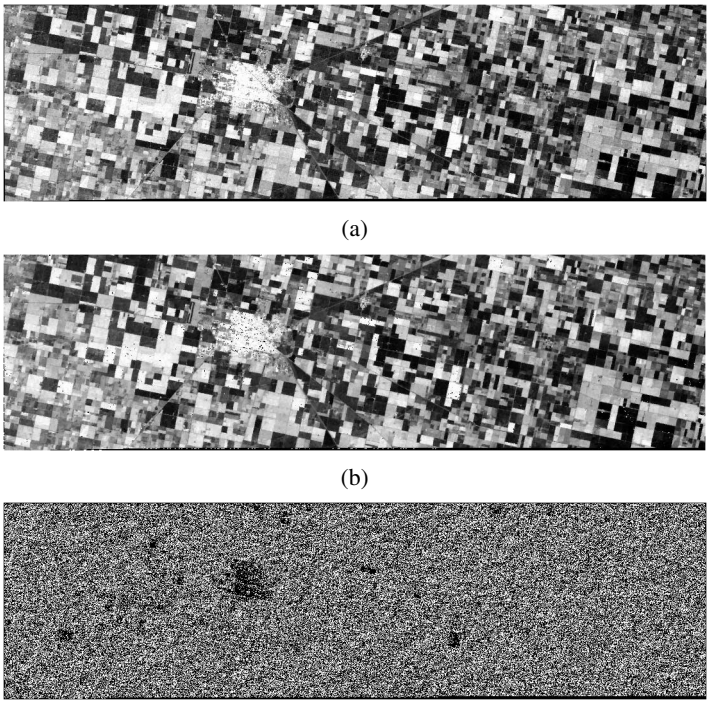
We repeat the same experiment applying the previous algorithm to a LANDSAT image from Colonia Tirolesa, a region near to Córdoba, Argentina. In Figure 8, (a) is the original satellite image, (b) is the approximated image, and (c) is the difference image. The empirical mean square error (MSE) in this case is 220.3293. Notice that the MSE for images shown in Figure 8 is much bigger than the MSE for the images shown in Figure 7. One of the reasons of having a large MSE could be that image (a) shown in Figure 8 is much more heterogeneous than image shown in Figure 7(a). Hence, the block of size $8 \times 8$ may be inadequate to approximate the real image in Figure 8 by an AR-2D process. Also notice that none of the differences found in Figures 7(c) and 8(c) can be visually detected.

This illustration suggests that a wide class of real images can be well represented by the spatial autoregressive models.

## 6 Image filtering algorithms based on robust estimation of spatial AR models

### 6.1 Algorithm 1

The original idea of data cleaning algorithm was given in Kashyap and Eom (1988). They created an algorithm that uses robust parameter estimation to filter



(a)

(b)

(c)

**Figure 8** (a) *Landsat image from Colonia Tirolesa.* (b) *Approximated image of* (a). (c) *Difference between images* (a) *and* (b).

the contamination produced by the outliers. The approach consist in two stages. First, using the contaminated image a robust estimation of the parameters of the autoregressive model is computed. If one observation is contaminated by an outlier the residual will be large. If the normalized residual is large, the observation is modified to reduce the effect of the outlier. Second, a new image is generated using an equation similar to (5.1) but considering a robustified version of the residuals. The procedure is described in the following algorithm.

Given an additively contaminated image $X$, divide the image into windows of size $8 \times 8$ and for each:

1. Initially set $Y^{(0)} = X(i, j)$. Compute the initial estimate $\boldsymbol{\phi}^{(0)}, \sigma^{(0)}$ from the contaminated observations $X(i, j)$ by the least squares algorithm.
2. Consider the $(m+1)$th iteration, where $Y^{(m)}$ and $\boldsymbol{\phi}^{(m)}$ are available. Obtain the updated estimates $r^{(m)}(\cdot)$ and $\hat{r}^{(m)}(\cdot)$ by the following recursive equations.

$$r^{(m)}(i, j) = Y^{(m)} - \boldsymbol{\phi}^{(m)^T} Z^{(m)}(i, j),$$

$$\hat{r}^{(m)}(i, j) = \psi\left[\frac{r^{(m)}(i, j)}{\hat{\sigma}^{(m)}}\right]\hat{\sigma}^{(m)},$$

where $\psi$ is one of the bounded continuous functions as discussed in connection with M estimation, $Z^{(m)}(i,j)^T = (Y^{(m)}(i-1,j), Y^{(m)}(i,j-1), Y^{(m)}(i-1,j-1))$, and $\boldsymbol{\phi}^{(m)^T} = (\phi^{(m)}(1,0), \phi^{(m)}(0,1), \phi^{(m)}(1,1))$.

3. Restore the image $Y^{(m+1)}(i,j)$ using

$$Y^{(m+1)}(i,j) = \boldsymbol{\phi}^{(m)^T} Z^{(m)}(i,j) + \hat{r}^{(m)}(i,j).$$

4. Obtain the M estimators of $\boldsymbol{\phi}^{(m+1)}$ and $\hat{\sigma}^{(m+1)}$ from the cleaned data $Y^{(m+1)}$.
5. Repeat steps 2–4 until $|\sigma^{(m+1)} - \sigma^{(m)}| < \varepsilon_1$ and $\|\boldsymbol{\phi}^{(m+1)} - \boldsymbol{\phi}^{(m)}\| < \varepsilon_2$, where $\varepsilon_1, \varepsilon_2 > 0$ are small (typically $10^{-3}$ or $10^{-6}$), and $\|\cdot\|$ denotes any matrix norm.

Notice that the algorithm can be applied to nonstationary images. The fact that we divide the image in small rectangles allow us to deal with images that have a nonconstant mean. The data cleaning procedure removes outliers without degrading the original image. The robust image model-based method does not produce blurred images after filtering.

### 6.2 Algorithm 2

It is well known that the M estimator is sensitive to the additive outliers. However, the GM estimators attenuate better these outliers. If the same cleaning data process is used for the residuals, Algorithm 1 can be explored using the GM estimator of $\boldsymbol{\phi}$ instead of the M estimator. This approach was investigated by Allende, Galbiati and Vallejos (2001). To facilitate both, the computation of the estimators and the recursive cleaning process, the estimation of $\sigma$ was carried out independently to the estimation of $\boldsymbol{\phi}$. In our experience the estimator of $\sigma$ in (4.7) was enough to reduce the effect of the outliers on the variance component. To visualize the effectiveness of filtering, Allende, Galbiati and Vallejos (2001) applied this algorithm to six images breaking up the original images into small fragments of size $8 \times 8$. In that experiment the results were not sensible to the window size ($8 \times 8$, $12 \times 12$, or $16 \times 16$) for images of size $512 \times 512$. In all cases the effect of the outliers in the filtered images is attenuated by the robust procedure, they have some minor effect, resulting in certain irregularity which can be observed in smooth areas. If a fragment is extremely smooth, with little contamination, the resulting standard deviation is very small, and this can cause numerical instability and an erratic parameter estimation.

### 6.3 Algorithm 3

In the above algorithms, the contaminated image $X$ is divided into windows of size $8 \times 8$ and each of them is assumed to obey a spatial AR model. The subdivision of the image involves only square windows, which is a strong assumption because square regions are not necessarily representative of the patterns in the image. In Vallejos and Mardesic (2004) the authors proposed an algorithm that is a generalization of Algorithm 1 and 2 in two senses. First, it is not required to divide the

image in square windows and, second, the algorithm works with a small number of spatial AR models, which improves the cleaning time of the image.

The algorithm is recursive and incorporates classification as additional information for $X$. The intensity of gray at the coordinate $(i, j)$ will be a combination between past information of the intensity $X(i, j)$ and information contained in the classification stage. The procedure to filter $X$ can be summarized in the following algorithm.

(1) Classify $X$ using a suitable method (in Vallejos and Mardesic (2004), they suggested the thresholding method).
(2) Apply Algorithm 2 to each class provided by the classification technique.

This algorithm does not require the classes to be square subdivisions of the original image. In particular, the classes given by the classification method can be physically separated regions in the original image.

Illustrations presented in Vallejos and Mardesic (2004) have shown that the implementation of the algorithm is faster than previous algorithms due to the fact that this method works with a small number of parameters. Furthermore, when the percentage of contamination is small this method seems to work well in practice. From filtered images it is noticed that there are certain regions where the algorithm is more effective. This is due to two factors. First, the number of iterations that the algorithm performs in each class is different. Second, the texture of the image plays an important role, hence the filtering process is highly affected by the texture of the image.

### 6.4 Algorithm 4

Here we describe a new image filtering algorithm. This algorithm is based on robust estimation of a model defined on a moving window that cover the whole contaminated image. Let us assume that an additively contaminated image $X$ is available. Then the algorithm can be described as follows.

(1) Consider the $7 \times 7$ subimage $Y$ located in the top left corner of $X$.
(2) Compute the robust RA estimator of the parameters of the model

$$Y(i, j) = \phi_1 Y(i-1, j) + \phi_2 Y(i, j-1) + \varepsilon(i, j),$$

where $\varepsilon(i, j)$ are independent and identically distributed variables with mean zero and variance $\sigma^2$. Denote the RA estimations as $\phi_1^{RA}$ and $\phi_2^{RA}$, respectively.
(3) Let $Y(i_0, j_0)$ be the center point of the $7 \times 7$ moving window. Then replace $X(i_0, j_0)$ by

$$X(i_0, j_0) = \widehat{\phi}_1^{RA} Y(i_0 - 1, j_0) + \widehat{\phi}_2^{RA} Y(i_0, j_0 - 1).$$

(4) Move the initial window in step (1) one position to the right (down if one edge of the image is already on the edge of $X$) as is illustrated in Figure 9.

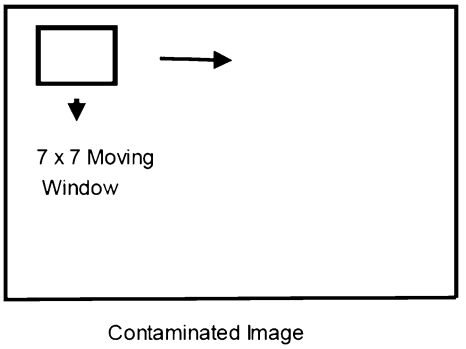



**Figure 9** $7 \times 7$ *moving window in Algorithm 4.*

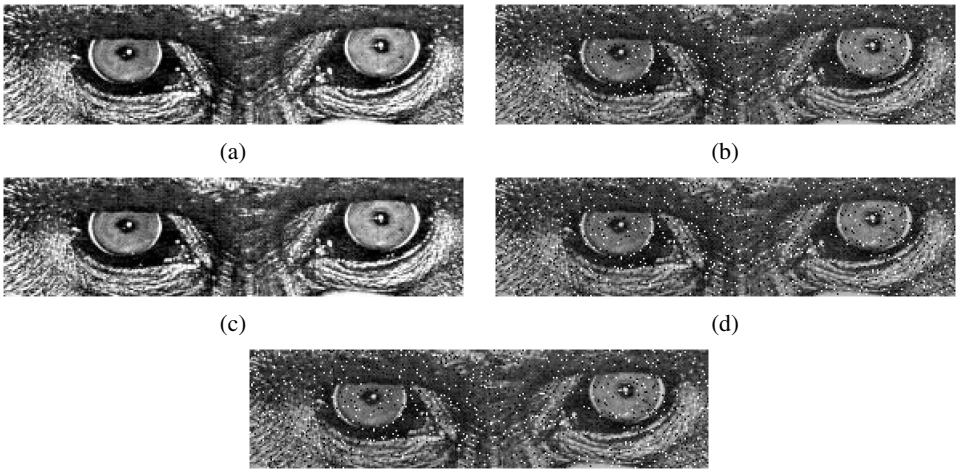



**Figure 10** (a) *Original image.* (b) *5% contaminated image.* (c) *Filtered image using the M estimator.* (d) *Image using the RA estimator.* (e) *Median filter.*

(5) Repeat steps (2) and (4) until the moving window has covered the whole image $X$.

In Figures 10 and 11 we display the images produced by Algorithm 4 (Figure 10(d) and Figure 11(d)). In addition we display the images produced by Algorithm 4 using the M estimator of the parameters instead of the RA estimator (Figure 10(c) and Figure 11(c)). Our results are in agreement with the proposal of Kashyap and Eom (1988). In the filtering process of image shown in Figure 10(a), the median filter introduced distortion. The performance of Algorithm 4 using the RA estimator is the best in terms of recovering the original patterns and textures of the original image. Visually the differences are hard to detect, however, the MSE is smallest when the original and filtered RA images are compared.

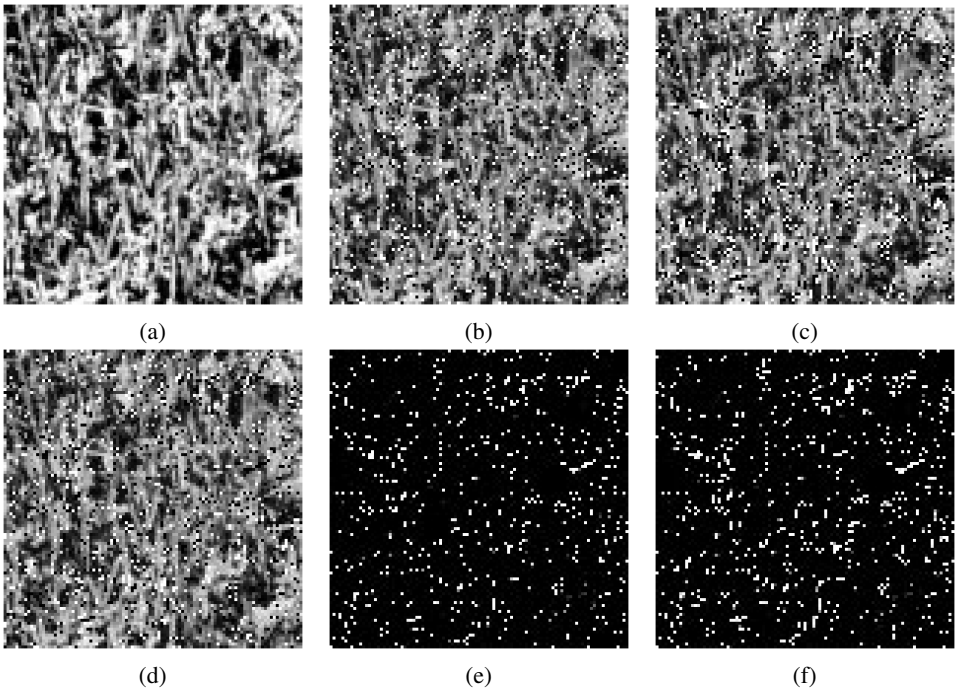



**Figure 11** (a) *Original image.* (b) *10% contaminated image.* (c) *Filtered image using the M estimator.* (d) *Filtered image using the RA estimator.* (e) *Difference* (a)–(c). (f) *Difference* (b)–(d).

Notice that Algorithm 4 is not a new version of the algorithm introduced by Kashyap and Eom. They divided the contaminated image in small $8 \times 8$ windows and then the algorithm is applied to each small window. In our algorithm we defined a moving window, and the filtering is produced just for the point located in the center. That is to say, the number of times that the algorithm needs to compute the robust estimators to filter an image is equal to the number of rows times the number of columns of the original image. In practice some considerations are necessary to treat the edges of the image that are excluded. For example, the grey intensities corresponding to the pixels over the edges can be replaced by the RA estimations using a different structure for the neighbors.

The size of the moving window plays an important role depending on the kind of treatment to be applied to the original image. In our experiments moving windows of size $8 \times 8$, $12 \times 12$, or $16 \times 16$ for images of size $512 \times 512$ produces the same effect on the filtered image. However, we expect to fit a better model for a small window size as mentioned above. If the purpose is to produce segmentation over the original image a large window size will produce a poor fitted model, hence the boundaries and borders will be highlighted. Thus we remark that for filtering contaminated images a small window size for the moving window will be appropriate; on the other hand a large window size will help the segmentation

process to produce differences between the original and fitted images that contain information about the borders and boundaries of the original image.

## 7 Conclusion

This paper reviewed main characteristics and applications of the spatial autoregressive and moving average models. Model fitting and robust estimation allowed the developement of image filtering algorithms. These algorithms originally considered a large number of parameters to estimate, making the filtering process slow and computationally expensive. Later modifications and variants of the original procedures have resulted in algorithms that perform faster. Moreover, the inclusion of these variants in the filtering process have improved the ability of these algorithms to clean different kinds of contaminated images. The performance of these type of algorithms under different kinds of contamination is an interesting open problem to be addressed in the future.

## Acknowledgments

The second author was supported by Secyt project 69/08 (05/B412), Universidad Nacional de Córdoba, Argentina. The third author was partially supported by Fondecyt grant 11075095, Chile. We would like to thank two anonymous referees and the editor for their valuable comments and suggestions.

## References

Allende, H. and Heiller, S. (1992). Recursive generalized M estimates for autoregressive moving-average models. *Journal of Time Series Analysis* **13** 1–18. MR1149267

Allende, H., Galbiati, J. and Vallejos, R. (1998). Digital image restoration using autoregressive time series type models. *Bulletin European Spatial Agency* **434** 53–59.

Allende, H., Galbiati, J. and Vallejos, R. (2001). Robust image modeling on image processing. *Pattern Recognition Letters* **22** 1219–1231.

Banerjee, S., Carlin, B. and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC Press, Florida.

Banham, M. R. and Katsaggelos, A. K. (1997). Digital image restoration. *IEEE Signal Processing Magazine* **14** 24–41.

Basu, S. and Reinsel, G. (1993). Properties of the spatial unilateral first-order ARMA model. *Advances in Applied Probbability* **25** 631–648. MR1234300

Bennet, J. and Khotanzad, A. (1999). Maximum likelihood estimation methods for multispectral random field image models. *IEEE Transaction Pattern Analysis and Machine Intelligence* **21** 537–543.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B* **55** 192–236. MR0373208

Bustos, O. and Yohai, V. (1986). Robust estimates for ARMA models. *Journal of the American Statistical Association* **81** 55–68. MR0830576

Bustos, O., Ruiz, M., Ojeda, S., Vallejos, R. and Frery, A. (2008). Asymptotic behavior of RA-estimates in autoregressive processes. Submitted.

Chang, I., Tiao, G. C. and Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics* **3** 193–204. MR0943602

Chen, C. and Lui, L. (1993). Joint estimation of model parameters and outliers in time series. *Journal of the American Statistical Association* **88** 284–297.

Cliff, A. and Ord, J. (1981). *Spatial Processes: Models and Applications*. Pion Ltd., London. MR0632256

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

Eunho, H. and Newton, H. J. (1993). The bias of estimators of causal spatial autoregressive models. *Biometrika* **80** 242–245. MR1225229

Francos, J., Narasimhan, A. and Woods, J. W. (1995). Maximum likelihood parameter estimation of textures using a Wold-decomposition based model. *IEEE Transactions on Image Processing* **4** 1655–1666.

Francos, J. and Friedlander, B. (1998). Parameter estimation of two-dimensional moving average random fields. *IEEE Transaction Signal Processing* **46** 2157–2165. MR1667322

Francos, J. and Nehorai, A. (2003). Interference mitigation in STAP using the two-dimensional Wold decomposition model. *IEEE Transactions on Signal Processing* **51** 2461–2470.

Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society, Series B* **34** 350–363. MR0331681

Griffith, D. A. (1988). *Advanced Spatial Statistics*. Kluver, Dordrecht, The Netherlands.

Grondona, M. R., Crossa, J., Fox, P. N. and Pfeiffer, W. H. (1996). Analysis of variety yield trials using two-dimensional separable ARIMA processes. *Biometrics* **52** 763–770.

Guyon, X. (1995). *Random Fields on a Network. Modeling, Statistics and Applications*. Springer, Berlin. MR1344683

Haining, R. P. (1978). The moving average model for spatial interaction. *Transactions and Papers, Institute of British Geographers, New Series* **3** 202–225.

Isaksson, A. J. (1993). Analysis of identified 2-D noncausal models. *IEEE Transactions on Information Theory* **39** 525–534.

Jain, A. K. (1989). *Fundamentals of Digital Image Processing*. Prentice Hall, Lugar.

Kashyap, R. L. and Chellappa, R. (1983). Estimation and choice of neighboors in spatial-interaction models of images. *IEEE Transactions on Information Theory* **19** 60–72.

Kashyap, R. and Eom, K. (1988). Robust images techniques with an image restoration application. *IEEE Transactions on Acoustics and Speech Signal Processing* **36** 1313–1325.

Katsaggelos, A. K. (1989). Iterative image restoration algorithms. *Optical Engineering* **28** 735–748.

Krishnamurthy, R., Woods, J. and Francos, J. (1996). Adaptive restoration of textures images with mixed spectra using a generalized Wiener filter. *IEEE Transactions on Image Processing* **5** 648–652.

Kokaram, A. (2004). A statistical framework for picture reconstruction using 2D AR models. *Image and Vision Computing* **22** 165–171.

Liu, F. and Piccard, R. (1996). Periodicity, directionality and randomness: Wold features for image modeling and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18** 722–733.

Martin, R. J. (1979). A subclass of lattice processes applied to a problem in planar sampling. *Biometrika* **66** 209–217. MR0548186

Martin, R. D. (1980). Robust estimation of autoregressive models. In *Direction in Time Series* (D. R. Brillinger and G. C. Tiao, eds.). Institute of Mathematical Statistics, Haywood, CA. MR0624655

Martin, R. J. (1990). The use of time-series models and methods in the analysis of agricultural field trials. *Communications in Statistics Theory Methods* **19** 55–81. MR1060398

Martin, R. J. (1996). Some results on unilateral ARMA lattice processes. *Journal of Statistical Planning and Inference* **50** 395–411. MR1394140

Ojeda, S. M. (1999). Robust RA estimators for bidimensional autoregressive models. Ph.D. dissertation, Facultad de Matemáticas, Astronomía y Física, Universidad Nacional de Córdoba, Argentina.

Ojeda, S. M., Vallejos, R. O. and Lucini, M. (2002). Performance of RA estimator for bidimensional autoregressive models. *Journal of Statistical Simulation and Computation* **72** 47–62. MR1907809

Quenouille, M. H. (1949). Problems in plane sampling. *Annals of Mathematical Statistics* **20** 355–375. MR0032175

Rukhin, A. and Vallejos, R. (2008). Codispersion coefficient for spatial and temporal series. *Statistics and Probability Letters* **78** 1290–1300. MR2444319

Tsay, R. S., Peña, D. and Pankratz, A. E. (2000). Outliers in multivariate time series. *Biometrics* **87** 789–804. MR1813975

Tsay, R. S. (1988). Outliers, level shifts, and variance change in time series. *Journal of Forecasting* **7** 1–20.

Tjostheim, D. (1978). Statistical spatial series modelling. *Advances in Applied Probability* **10** 130–154. MR0471224

Vallejos, R. and Mardesic, T. (2004). A recursive algorithm to restore images based on robust estimation of NSHP autoregressive models. *Journal of Computational and Graphical Statistics* **13** 674–682. MR2087721

Vallejos, R. and Garcia-Donato, G. (2006). Bayesian analysis of contaminated quarter plane moving average models. *Journal of Statistical Computation and Simulation* **76** 131–147. MR2199883

Vallejos, R., Ojeda, S. and Bustos, O. (2008). On RA and GM estimates for spatial autoregressive models. Submitted.

Whittle, P. (1954). On stationary processes on the plane. *Biometrika* **41** 434–449. MR0067450

O. Bustos
S. Ojeda
Facultad de Matemática, Astronomía y Física
Universidad Nacional de Córdoba
Córdoba
Argentina
E-mail: bustos@famaf.unc.edu.ar
ojeda@famaf.unc.edu.ar

R. Vallejos
Departamento de Matemática
Universidad Técnica Federico Santa María
Casilla 110-V Valparaíso
Chile
E-mail: ronny.vallejos@usm.cl
URL: http://www.deuv.cl/vallejos