Statistics and Computing **12**: 125–134, 2002

Simulation-based inference for simultaneous processes on regular lattices

XAVIER DE LUNA* and MARC G. GENTON†

**Department of Economics, Umeå University, S-90187 Umeå, Sweden*
xavier.deluna@econ.umu.se
†*Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA*
genton@stat.ncsu.edu

Received August 2000 and accepted February 2001

The article proposes a simulation-based inferential method for simultaneous processes defined on a regular lattice. The focus is on spatio-temporal processes with a simultaneous component, that is such that contemporaneous spatial neighbors are potential explanatory variables in the model. The new method has the advantage of being simpler to implement than maximum likelihood and allows us to propose a robust estimator. We give asymptotic properties, present a Monte Carlo study and an illustrative example.

Keywords: indirect inference, quadrant process, robust estimation, spatio-temporal process, Yule-Walker estimator

1. Introduction

In this paper we propose and study the use of a simulation-based inference technique for simultaneous autoregression models (see e.g. Cressie 1993, Sections 6.3 and 6.8). These are models for data typically observed on a two or three dimensional lattice, depending on whether the time dimension is present or not, although higher dimensional grids could also be considered. The simultaneity arises when contemporaneous spatial neighbors are explanatory variables in the model. These models are useful whenever the existence of spatial and temporal interdependencies are of interest and need to be assessed. Areas of application include agricultural trials, socio-economic studies, and pollution analysis (e.g. Haining 1990). The solely spatial version was studied in the seminal paper of Whittle (1954). When the time dimension is present, simultaneous models are most useful for descriptive purposes (for instance to uncover the actual dependencies existing between spatially and temporally distributed observations), while conditional autoregression (see, e.g, Cressie 1993, Section 6.3) models should be used when a forecast-minded interpretation is sought. Furthermore, simultaneous models are also useful for modeling spatio-temporal dependencies present in the residuals of a regression model which is then the actual interest of the study. A simultaneous modeling of the residual process has then the advantage of being more parsimonious than a corresponding conditional model.

Maximum likelihood estimation for simultaneous models is to date only possible for Gaussian processes (Ali 1979, Guyon 1982, Kiiveri 1992), and is notoriously not robust to the presence of outliers. On the other hand, least-squares and Yule-Walker type estimators which often are simpler to implement and asymptotically equivalent to maximum likelihood, happen not to be consistent for simultaneous models (Whittle 1954). All these problems arise from the simultaneity. The simulation-based estimator that we introduce avoids the estimation of a simultaneous model, by considering an auxiliary unilateral model which is fitted to the observed data as well as to simulated data from the simultaneous model of interest. A calibration of these two fits leads to an estimation of the latter model. The price of this simplicity is a loss in efficiency. On the other hand, we are able to perform consistent inference for more complex simultaneous models. Thus, for instance, the proposed estimator can be robustified for the presence of outliers. Simulation-based estimators based on auxiliary models were first proposed in Gouriéroux, Monfort and Renault (1993). In Genton and de Luna (2000) these estimators were developed as a general robust inference technique, and their applicability to the robust estimation of autoregressive and moving average time series models was studied in de Luna and Genton (2001).

The paper starts by describing in Section 2 the simultaneous autoregression model, the proposed inferential method and its asymptotic theory. In Section 3 a robust estimator is developed.

Section 4 presents some small sample properties through a Monte Carlo study. An illustrative example is provided in Section 5 and the paper is concluded in Section 6.

2. Model and inference

2.1. *Simultaneous processes*

Let $\{Z(x, y, t)\}$, where $(x,y) \in \mathbb{Z}^2$ indexes the spatial location, and $t \in \mathbb{Z}$ the time, be a spatio-temporal stochastic process admitting a STAR (simultaneous spatio-temporal autoregression) representation

$$Z(x, y, t)-\mu = \sum_{i,j,k} \theta_{ijk}(Z(x-i, y-j, t-k)-\mu)+\epsilon(x, y, t), \tag{1}$$

where the summations are done over $|i|, |j| \leq p$ and $0 \leq k \leq q$; here $\theta_{000} = 0$ ($k = 0$ corresponds to a SAR, simultaneous spatial autoregression, representation). The process $\{\epsilon(x, y, t)\}$ is independently and identically distributed with mean zero and finite variance σ^2. The model is fully described when a distributional assumption on the noise $\{\epsilon(x, y, t)\}$ is made. To obtain the likelihood a Gaussian noise is commonly assumed. For the inferential method described below, we only need the distribution to be such that realizations can be generated with a computer program. The process $\{Z(x, y, t)\}$ is here assumed to be covariance-stationary under translation of (x, y, t). This can be shown to be the case if the polynomial $1 - \sum_{i,j,k} \theta_{ijk} z_1^i z_2^j z_3^k$ in complex variables z_1, z_2, and z_3 is non-zero when simultaneously we have $|z_1| = |z_2| = 1$ and $|z_3| \leq 1$ (Whittle 1954, Ali 1979). Note that the two processes with $\theta_{i,j,k}$ and $\theta_{-i,-j,k}$ interchanged have the same covariance structure. Moreover, we assume in the sequel that $\theta_{ij0} \neq 0$ for at least one (i, j) pair (simultaneous process). Simultaneity is justified, for instance, when model (1) is fitted on time-aggregated data for descriptive purposes, see Cressie (1993, p. 450). Figure 1 illustrates the dependence structure of two simple SAR models. Both are such that $p = 1$, although one of them has non-zero coefficients only for the nearest neighbors: In the sequel we distinguish such truncated structures with the notation $p = 1$ (T). We have assumed for clarity of notation a constant mean μ, although the case $\mu := \mu(x, y, t)$, a so-called large scale variation, does not add further conceptual complications.

For non-simultaneous processes ($\theta_{ij0} = 0$, for all i, j), least squares and Yule-Walker provide consistent estimators, but this is not true anymore when the process is simultaneous (Whittle 1954). Whittle also showed how approximate maximum Gaussian likelihood estimation could be tackled when the time dimension was not present (SAR models). This was later improved on by Guyon (1982). Several generalizations of the SAR model have also been considered such as the introduction of a moving average component (Cressie 1993, Chapters 6 and 7). The consideration of the time dimension leads to further complications, and the maximum likelihood inference has been studied by Ali (1979) for Gaussian processes.

Fig. 1. *Graphical representation of the spatial dependencies of two SAR models: Equation (1) with $k = 0$, $p = 1$ and $\theta_{110} = \theta_{-110} = \theta_{1-10} = \theta_{-1-10} = 0$ (left diagram); $k = 0$, $p = 1$ (right diagram). The black dot depends on 4 and 8 neighbors respectively (white dots)*

2.2. *Auxiliary representation and simulation-based estimation*

The simulation-based inferential method introduced below is based on a unilateral (non-simultaneous) auxiliary representation of the STAR process $Z(x, y, t)$, namely a finite order QAR (quadrant autoregression) model

$$\sum_{0 \leq i,j,k \leq r} \pi_{ijk}(Z(x-i, y-j, t-k)-\mu) = \nu(x, y, t), \tag{2}$$

where $\pi_{000} = 1$ and $\{\nu(x, y, t)\}$ is a process with mean zero and finite variance τ^2, such that $E(\nu(x, y, t)\mathbf{Z}) = \mathbf{0}$, where $\mathbf{Z}$ is the vector formed of $Z(x-i, y-j, t-k)$, for $0 \leq i, j, k \leq r$, except for $Z(x, y, t)$. Representation (2) is uniquely defined, and can be consistently fitted with the Yule-Walker estimator (see Section 2.3). Note that the process $\{\nu(x, y, t)\}$ will generally be correlated unless r is large enough (often infinity is necessary, see Tjøstheim (1978)). We therefore allow the noise $\{\nu(x, y, t)\}$ to be correlated in the sequel. Finally, seven other representations may be obtained by letting i, j or/and k be negative instead of positive, e.g., by summing over $0 \geq i \geq -r$ in (2). In the sequel, it is assumed that one of the unilateral representations has been chosen.

Because there is no trivial way of analytically retrieving the parameterization of the model of interest (1) from the auxiliary model (2), simulated data is used in the spirit of the indirect inference procedure introduced by Gouriéroux, Monfort and Renault (1993); see also Brännäs and de Luna (1998) and de Luna and Genton (2001). Let $\boldsymbol{\theta}$ be the vector formed by all the parameters in (1), i.e. θ_{ijk} for $|i|, |j| \leq p$ and $0 \leq k \leq q$, and σ^2. We assume without loss of generality that $\mu = 0$. Denote also by $\boldsymbol{\pi}$ the vector formed by all the parameters of (2), i.e. π_{ijk}, for $0 \leq i, j, k \leq r$ and τ^2. Assume we have observed $Z(x, y, t)$ on a regular three-dimensional grid with dimensions $n_x \times n_y \times n_t$. The indirect estimator for $\boldsymbol{\theta}$ is then obtained as follows:

(i) Fit model (2) with the Yule-Walker estimator (Tjøstheim 1978, Ha and Newton 1993) using the available data set, thus yielding $\hat{\boldsymbol{\pi}}$, an estimate of $\boldsymbol{\pi}$.
(ii) Simulate $s \geq 1$ data sets (on grids of dimension $(n_x \times n_y \times n_t)$) from the model of interest (1). To these s sets

of observations fit (2), again with the Yule-Walker estimator, thus yielding s estimates $\boldsymbol{\pi}_1^*, \ldots, \boldsymbol{\pi}_s^*$, with which the average $\boldsymbol{\pi}^* = 1/s \sum_{i=1}^{s} \boldsymbol{\pi}_i^*$, an estimate of $\boldsymbol{\pi}$, is computed.

The estimator $\boldsymbol{\pi}^*$ is a function of the simulated data and therefore $\boldsymbol{\theta}$. By varying $\boldsymbol{\theta}$ we try to bring $\boldsymbol{\pi}^*$ close to $\hat{\boldsymbol{\pi}}$ where close is defined by a generalized least squares distance. In other words:

(iii) An estimator for $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} (\boldsymbol{\pi}^*(\boldsymbol{\theta}) - \hat{\boldsymbol{\pi}})' \Omega (\boldsymbol{\pi}^*(\boldsymbol{\theta}) - \hat{\boldsymbol{\pi}}). \quad (3)$$

Note that in (ii) the same s realizations for $\{\epsilon(x, y, t)\}$ must be used throughout the algorithm in order to ensure a successful minimization in (iii) (we used the downhill simplex method) and valid asymptotic properties for $\hat{\boldsymbol{\theta}}$, see Gouriéroux, Monfort and Renault (1993).

Implicit to this estimation algorithm is the choice of s and the dimension r of the auxiliary parameter $\boldsymbol{\pi}$. These are discussed in connection to the asymptotic properties below.

2.3. *Asymptotics*

Let $h(\boldsymbol{\theta}) = \boldsymbol{\pi}$ be the function linking the parameter of interest $\boldsymbol{\theta}$ and the auxiliary parameter $\boldsymbol{\pi}$. If h is locally injective around the true value of $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\pi}}$ is a consistent and asymptotically normal estimator, then we have (Gouriéroux and Monfort, 1996, Proposition 4.2), for $n_x, n_y, n_t \to \infty$,

$$\sqrt{n_x n_y n_t}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(\mathbf{0}, W), \quad (4)$$

where $W = (1 + 1/s)\{(D'\Omega D)^{-1} D'\Omega V \Omega D (D'\Omega D)^{-1}\}$, with D the matrix of the first partial derivatives (Jacobian) $D = \partial h(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$ evaluated at the true value of $\boldsymbol{\theta}$, and V is the asymptotic covariance matrix of $\hat{\boldsymbol{\pi}}$. The matrix Ω may be chosen so as to maximize the efficiency (Gouriéroux and Monfort 1996). The above results state that $\hat{\boldsymbol{\theta}}$ is consistent for any $s \geq 1$, and that efficiency can be improved by increasing s. However, for $s > 3$ the gain in efficiency is marginal and generally not worth the computational cost implied.

For (4) to hold for the estimator defined by (3), we need to show that (a) $h(\boldsymbol{\theta}) = \boldsymbol{\pi}$ defined by (2) is injective and (b) the Yule-Walker estimator is a consistent and asymptotically normal estimator of $\boldsymbol{\pi}$. We address these two issues in turn.

(a) *Injectivity of $h(\boldsymbol{\theta}) = \boldsymbol{\pi}$ defined by* (2): Assuming $\theta_{i,j,k} = \theta_{-i,-j,k}$ the STAR process is identifiable and the vector $\boldsymbol{\gamma}$ formed of the infinite sequence of autocovariances is uniquely defined (Whittle 1954). Furthermore, the Yule-Walker equations (Tjøstheim 1978, Section 6) define a one-to-one mapping between $\boldsymbol{\pi}$ and a finite number of autocovariances, say forming the vector $\boldsymbol{\gamma}_r$, with $\boldsymbol{\pi} = g(\boldsymbol{\gamma}_r)$. By the uniqueness of $\boldsymbol{\gamma}$, there exists a finite number of autocovariances $\boldsymbol{\gamma}_r$, such that $\boldsymbol{\gamma}_r = k(\boldsymbol{\theta})$ is an injective function, thereby guaranteeing the injectivity of $h(\cdot) = g(k(\cdot))$. Thus, when r is large enough—the dimension of $\boldsymbol{\pi}$ must be at least as large as the dimension of $\boldsymbol{\theta}$—the function h is locally injective around the true value of $\boldsymbol{\theta}$. Note that a sufficient lower bound for r for the injectivity of h to hold is not straightforward to deduce because it is not possible to characterize $\boldsymbol{\gamma}$ as a function of $\boldsymbol{\theta}$ (Whittle 1954). This does not imply a real practical hinder since it suffices to use r large enough.

(b) *Asymptotics for the Yule-Walker estimator*: Consistency and asymptotic normality of the Yule-Walker estimator $\hat{\boldsymbol{\pi}}$ of $\boldsymbol{\pi}$ has been proved when the QAR representation is such that the process $\{\nu_t\}$ is independently distributed, see Tjøstheim (1978) and Ha and Newton (1993). In our context, $\{\nu_t\}$ is typically correlated and more general results are needed. Guyon (1982) shows consistency of the sample autocovariances of a stationary spatial process. Rosenblatt (1986, Theorem 2) establishes the asymptotic normality and gives the asymptotic covariance matrix of an arbitrary fixed number of estimated autocovariances. The consistency and asymptotic normality of $\hat{\boldsymbol{\pi}}$, a function of a finite number of autocovariances, is thus a direct consequence of these results.

Thus, $\hat{\boldsymbol{\theta}}$ defined by (3) has asymptotic distribution given by (4) under the following assumptions.

(A1) The process $\{Z(x, y, t)\}$ admits representation (1) with $\theta_{i,j,k} = \theta_{-i,-j,k}$, and the polynomial $1 - \sum_{i,j,k} \theta_{ijk} z_1^i z_2^j z_3^k$ in complex variables z_1, z_2, and z_3 is non-zero when simultaneously we have $|z_1| = |z_2| = 1$ and $|z_3| \leq 1$.

(A2) (Assumptions of Theorem 2 in Rosenblatt 1986) $\{Z(x, y, t)\}$ is a strongly mixing and strictly stationary stochastic process with mean zero and bounded non-centered moments up to order eight. Finally, cumulants of $\{Z(x, y, t)\}$ up to eighth order are absolutely summable.

Note that (A2) holds under (A1) if $\epsilon(x, y, t)$ is Gaussian.

To conclude this section let us discuss the tuning parameter r. The estimation method proposed is not as dependent on r as it may appear. For instance, consistency is ensured as soon as r is large enough so that the dimension of $\boldsymbol{\pi}$ is at least equal to the dimension of $\boldsymbol{\theta}$. A heuristic approach to the choice of r is to start with some value (for instance chosen with an information criterion such as AIC or BIC, (respectively Akaike and Bayesian information criterion, see e.g. Cressie (1993)), ensuring that as little structure as possible remains in the process $\nu(x, y, t)$), estimate $\boldsymbol{\theta}$, and repeat the procedure for a few other values of r to ensure that the estimation of $\boldsymbol{\theta}$ is not sensitive to this parameter.

2.4. *Inference*

To perform inference on the indirect estimator of $\boldsymbol{\theta}$ we need to compute the matrix W. Denote by R the covariance matrix of the sample autocovariances given in Rosenblatt (1986). Then, $V = CRC'$, where C is the Jacobian of the function $g(\cdot)$ as function of γ_r. However, when r is large enough for the process $\{\nu_t\}$ to be nearly uncorrelated, the asymptotic variance of $\hat{\boldsymbol{\pi}}$ provided in Tjøstheim (1978) is simpler to compute and should constitute a good approximation of V.

Due to the complexity of a bilateral model such as (1), the function $h(\boldsymbol{\theta}) = \boldsymbol{\pi}$, and therefore its Jacobian $D(\boldsymbol{\theta})$, cannot be

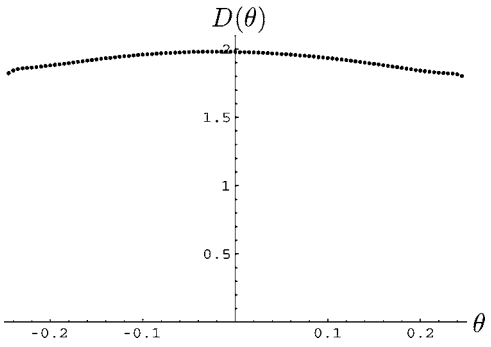

Fig. 2. *Approximation of D(θ) for the simple SAR model described by equation (5)*

retrieved analytically (Whittle 1954). It is, however, possible to obtain a consistent estimator of W, see Gouriéroux, Monfort and Renault (1993, Appendix 2) as soon as a consistent estimator of V is available. Assume for illustration that θ is of dimension one. Then, if $h(\cdot)$ is known, a numerical approximation of the scalar $D(\hat{\theta})$ is then, for instance, given by

$$\frac{h(\hat{\theta} + \varepsilon/2) - h(\hat{\theta} - \varepsilon/2)}{\varepsilon},$$

with ε small. The unknown function $h(\cdot)$ can then be replaced by a consistent estimator based on simulated data as in step (ii) in the previous section. Thus, we obtain $\pi^*(\hat{\theta} + \varepsilon/2)$ and $\pi^*(\hat{\theta} - \varepsilon/2)$ which we substitute for the corresponding $h(\cdot)$ values in the above ratio.

For illustration we display in Fig. 2, such approximations of D(θ) for different θ values corresponding to the simple SAR model (whose dependence structure is illustrated in Fig. 1, left diagram)

$$Z(x, y) = \theta\{Z(x, y-1) + Z(x, y+1) + Z(x-1, y) + Z(x+1, y)\} + \epsilon(x, y), \quad (5)$$

where $\{\epsilon(x, y)\}$ has standard normal distribution, i.e. $\sigma^2 = 1$, and θ is estimated based on the auxiliary quadrant autoregression representation (special case of (2)):

$$Z(x, y) = \pi\{Z(x, y-1) + Z(x-1, y)\} + \nu(x, y).$$

Thus, inference associated to $\hat{\boldsymbol{\theta}}$, as for instance Wald tests, can be carried out as usual by using W where D and V are computed as indicated above.

3. Robust inference

3.1. *Robustness for simultaneous processes*

In this section, the setting of Huber (1981) and Hampel *et al.* (1986) for the theoretical study of robust properties of an inferential method is followed. The main tool to describe such properties is the influence function (Hampel 1974). Hampel's original definition was aimed at the independent and identically distributed case. The definition of the influence function can, however, be generalized in order to be applicable to the spatio-temporal stochastic processes dealt with in this paper: For a multidimensional statistical functional T at a distribution F, the influence function $IF(\nu, T, F)$ at the probability measure ν (defining the contamination process) is defined by the limit:

$$IF(\nu, T, F) = \lim_{\varepsilon \to 0+} \frac{T(F_\varepsilon) - T(F)}{\varepsilon}, \quad (6)$$

in those ν where this limit exists. As $\varepsilon \to 0$, the distribution of the contaminated process $F_\varepsilon \to F$. Note that Künsch's (1984) and Martin and Yohai's (1986) influence functions are particular cases of (6), see Genton and de Luna (2000). The importance of the influence function lies in its appealing heuristic interpretation: It measures the asymptotic bias caused by an infinitesimal contamination of the observations. A robust estimator is one such that this bias is bounded.

The influence function can be derived for the simulation-based estimator of $\boldsymbol{\theta}$ from the influence function of the auxiliary estimator of $\boldsymbol{\pi}$. Denote by Π and Θ the statistical functionals corresponding respectively to the estimators $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\theta}}$. In this case, F is the marginal distribution (corresponding to all the variables included in the auxiliary QAR representation) of the joint distribution of the process. Let $IF_{QAR}(\nu, \Pi, F)$ be the vector influence function of the auxiliary estimator of $\boldsymbol{\pi}$, and $IF_{STAR}(\nu, \Theta, F)$ be the vector influence function of the simulation-based estimator of $\boldsymbol{\theta}$. The function h defined in the previous section sends Θ on $\Pi = h(\Theta)$. Then, from Genton and de Luna (2000, Theorem 1):

$$IF_{STAR}(\nu, \Theta, F) = P(\Theta(F))\, IF_{QAR}(\nu, \Pi, F), \quad (7)$$

where $P(\Theta(F)) = [D(\Theta(F))'\Omega D(\Theta(F))]^{-1} D(\Theta(F))'\Omega$. Equation (7) holds when h is locally injective around $\Theta(F)$ and the Jacobian matrix $D(\Theta(F))$ exists. Therefore, $IF_{STAR}(\nu, \Theta, F)$ is bounded if $IF_{QAR}(\nu, \Pi, F)$ is bounded. In particular, we can replace $IF_{QAR}(\nu, \Pi, F)$ by either Künsch's (1984) or Martin and Yohai's (1986) influence function, and obtain $IF_{STAR}(\nu, \Theta, F)$ for both definitions respectively.

The breakdown point is another important feature of reliability of an estimator (Huber 1981, 1984, Hampel *et al.* 1986). It indicates how many data points need to be replaced by arbitrary values to destroy the estimator, i.e. to bring the estimator to the boundaries of the parameter space. This concept can be applied to a dependent data setting; see, e.g. Lucas (1997), Genton (1998) and Ma and Genton (2000). The simulation-based estimator discussed here has the same breakdown point than the auxiliary estimator used on the observed data, because the simulated data is outlier-free.

A consequence of these results is that a robust auxiliary estimator on the data yields a robust estimator of the parameters of the simultaneous model. On the other hand, one may use an efficient (non-robust) estimator of the auxiliary parameter on the simulated data, the latter being outlier-free.

3.2. *Robust Yule-Walker estimator*

Robustness can be achieved by using robust Yule-Walker equations, i.e. by using robust estimators of the autocovariances. Ma and Genton (2000) have proposed a highly robust estimator $\hat{\gamma}_Q$ of the autocovariance function defined by:

$$\hat{\gamma}_Q(\mathbf{h}, \mathbf{z}) = \frac{1}{4}\big[\hat{Q}^2(\mathbf{u}+\mathbf{v}) - \hat{Q}^2(\mathbf{u}-\mathbf{v})\big], \tag{8}$$

where $\mathbf{z}$ is the data vector, and $\mathbf{u}$ and $\mathbf{v}$ are vectors of data from $\mathbf{z}$, such that the data in $\mathbf{v}$ are apart from the data in $\mathbf{u}$ by a vector $\mathbf{h}$. This estimator is based on the identity $\mathrm{Cov}(X, Y) = \frac{1}{4\alpha\beta}[\mathrm{Var}(\alpha X + \beta Y) - \mathrm{Var}(\alpha X - \beta Y)]$, $\forall \alpha, \beta \in \mathbb{R}$. We have set $\alpha = \beta = 1$ and used the highly robust estimator of scale $\hat{Q}(\mathbf{y}) = c\,\{|Y_i - Y_j|; i < j\}_{(m)}$ proposed by Rousseeuw and Croux (1992, 1993), where $\mathbf{y} = (Y_1, \ldots, Y_n)^T$ is the sample, $m = \mathrm{int}[\binom{n}{2} + 2)/4] + 1$ and $\mathrm{int}[\cdot]$ denotes the integer part. The factor c is for consistency: at the Gaussian distribution $c = 2.2191$. This means that we sort the set of all absolute differences $|Y_i - Y_j|$ in increasing order for $i < j$ and then compute its m-th order statistic (approximately the $1/4$ quantile for large n). This value is multiplied by c, thus yielding $\hat{Q}$. Note that $\hat{Q}$ can be computed using no more than $O(n \log n)$ time and $O(n)$ storage, by means of the fast algorithm described in Croux and Rousseeuw (1992).

The estimator (8) turns out to be a highly robust estimator of autocovariance. It has a bounded influence function, and therefore the influence function $IF_{QAR}(\nu, \Pi, F)$ of the auxiliary parameters, obtained by solving the robust Yule-Walker equations, is also bounded. As a consequence, the influence function $IF_{STAR}(\nu, \Theta, F)$ of the simulation-based estimator is also bounded by (7). The breakdown point of the estimator (8) is 25% relatively to $\mathbf{z}$, because the breakdown point of $\hat{Q}$ is 50% relatively to $\mathbf{u}+\mathbf{v}$ and $\mathbf{u}-\mathbf{v}$ (see Genton 1998, Ma and Genton 2000). Therefore, the breakdown point of the simulation-based estimator is also 25%. Note that the Gaussian asymptotic efficiency of $\hat{Q}$ is 82%, thus yielding very good efficiency properties for the simulation-based estimator.

4. Monte Carlo study

In order to illustrate the performances that may be obtained with simulation-based estimators, we present a Monte Carlo experiment based on a simple SAR and STAR model. Robustness properties towards outliers are also analyzed.

Simulated data on a $n_x \times n_y$ grid are obtained by generating a vector $\mathbf{z} \in \mathbb{R}^{n_x n_y}$ from a multivariate normal distribution with mean vector zero and variance-covariance matrix $\Sigma = \sigma^2[(I - N)'(I-N)]^{-1}$, where N is the weighted neighbor matrix and I is the identity matrix. Typically, the Cholesky decomposition of Σ is used, but faster simulation is obtained by noting that $\Sigma = GG'$ with $G = \sigma(I - N)^{-1}$. Thus, $\mathbf{z}$ is the solution of the (sparse) linear system $(I - N)\mathbf{z} = \sigma\mathbf{e}$, with $\mathbf{e}$ a vector of independent and identically distributed standard normal components. Note that an iterative simulation method is also available (Heuvelink 1998, p. 83).

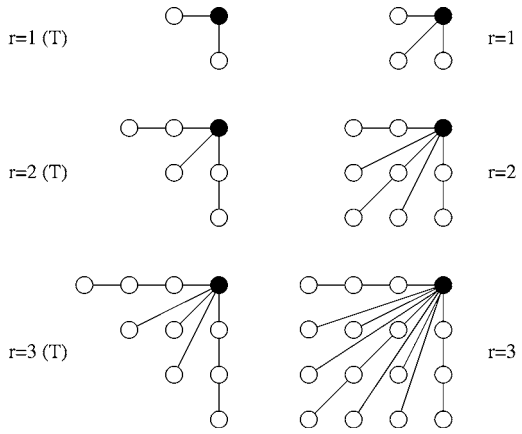

Fig. 3. *Graphical representation of a QAR model for $r = 1$ (T), $r = 2$ (T), $r = 3$ (T), as well as their nontruncated versions. The black dot depends on neighbors (white dots) in the lower left quadrant only*

4.1. *A simple SAR model*

Consider the simple SAR model of order $p = 1$ (T) with the same dependence θ in all four directions, described by equation (5), where $\{\epsilon(x, y)\}$ has standard normal distribution, i.e. $\sigma^2 = 1$. The process $Z(x, y)$ is covariance-stationary if $|\theta| < 1/4$, and we study the situations $\theta = -0.2, 0.1, 0.2$. We use simulated data on 10×10 and 15×15 grids. Observations on the edges are simulated by letting the unavailable Z's be equal to zero. We use the `Splus` library *S+SpatialStats* to obtain maximum likelihood estimators (ML) for θ and σ^2, and a *Pascal* program to obtain simulation-based estimators based on QAR auxiliary models, as well as truncated versions, i.e. QAR models where the sum in (2) is over $0 \le i, j, k, i+j+k \le r$, see Fig. 3. We use the orders $r = 1, 2$, and 3 in both cases. QAR parameters are estimated using the biased corrected Yule-Walker estimator of Ha and Newton (1993). We use $s = 1$, $s = 3$, and the identity matrix for Ω. Results for 200 replicates are summarized in Tables 1–4.

The results show that the simulation-based estimators may perform well but are in general less efficient than ML for the parameter θ. On the other hand, for the estimation of σ^2, ML is often outperformed by our estimator. The best, in terms of bias and efficiency, simulation-based estimators were obtained with small orders r. Typically $r = 1$ seems to be a good choice in our examples, although increasing the order r by one or two did not imply significant losses of efficiency. This confirms our previous remark, given at the end of Section 2.2, that the simulation-based estimator is not very sensitive to the choice of the structure of the auxiliary representation. Increasing s from 1 to 3 improved the efficiency of the simulation-based estimators. However, in unreported experiments we used larger values of s and did not observe significant efficiency improvement to the experiments reported here. As expected, estimators improved when increasing

Table 1. *Results of estimating θ and σ^2 from 200 simulated realizations of a SAR model on a 10×10 grid with $\theta = 0.1$ and $\sigma^2 = 1$ using maximum likelihood and the simulation-based method proposed in this paper*

		dim ($\boldsymbol{\pi}$)	mean ($\hat{\theta}$)	s.d. ($\hat{\theta}$)	mean ($\hat{\sigma}^2$)	s.d. ($\hat{\sigma}^2$)
	ML		0.093	0.035	1.007	0.268
$s = 1$	$r = 1$ (T)	3	0.096	0.049	1.006	0.222
	$r = 1$	4	0.097	0.050	1.006	0.219
	$r = 2$ (T)	6	0.099	0.051	1.006	0.222
	$r = 2$	9	0.096	0.054	0.999	0.245
	$r = 3$ (T)	10	0.097	0.053	1.006	0.229
	$r = 3$	16	0.080	0.060	1.009	0.249
$s = 3$	$r = 1$ (T)	3	0.101	0.039	1.008	0.153
	$r = 1$	4	0.100	0.041	1.012	0.154
	$r = 2$ (T)	6	0.100	0.042	1.008	0.156
	$r = 2$	9	0.100	0.042	1.003	0.160
	$r = 3$ (T)	10	0.100	0.042	1.006	0.162
	$r = 3$	16	0.094	0.045	1.001	0.217

Table 2. *Results of estimating θ and σ^2 from 200 simulated realizations of a SAR model on a 10×10 grid with $\theta = 0.2$ and $\sigma^2 = 1$ using maximum likelihood and the simulation-based method proposed in this paper*

		dim ($\boldsymbol{\pi}$)	mean ($\hat{\theta}$)	s.d. ($\hat{\theta}$)	mean ($\hat{\sigma}^2$)	s.d. ($\hat{\sigma}^2$)
	ML		0.193	0.024	1.018	0.316
$s = 1$	$r = 1$ (T)	3	0.192	0.035	1.019	0.245
	$r = 1$	4	0.194	0.037	1.021	0.245
	$r = 2$ (T)	6	0.192	0.034	1.024	0.237
	$r = 2$	9	0.185	0.040	1.027	0.247
	$r = 3$ (T)	10	0.182	0.041	1.025	0.259
	$r = 3$	16	0.158	0.052	1.026	0.308
$s = 3$	$r = 1$ (T)	3	0.198	0.031	1.029	0.187
	$r = 1$	4	0.200	0.030	1.018	0.173
	$r = 2$ (T)	6	0.199	0.027	1.022	0.189
	$r = 2$	9	0.195	0.030	1.027	0.228
	$r = 3$ (T)	10	0.193	0.031	1.023	0.188
	$r = 3$	16	0.179	0.041	1.015	0.247

Table 3. *Results of estimating θ and σ^2 from 200 simulated realizations of a SAR model on a 10×10 grid with $\theta = -0.2$ and $\sigma^2 = 1$ using maximum likelihood and the simulation-based method proposed in this paper*

		dim ($\boldsymbol{\pi}$)	mean ($\hat{\theta}$)	s.d. ($\hat{\theta}$)	mean ($\hat{\sigma}^2$)	s.d. ($\hat{\sigma}^2$)
	ML		−0.198	0.020	0.999	0.300
$s = 1$	$r = 1$ (T)	3	−0.197	0.032	1.019	0.257
	$r = 1$	4	−0.195	0.037	1.013	0.250
	$r = 2$ (T)	6	−0.197	0.037	1.017	0.254
	$r = 2$	9	−0.190	0.041	1.015	0.258
	$r = 3$ (T)	10	−0.190	0.036	1.011	0.264
	$r = 3$	16	−0.166	0.052	1.021	0.312
$s = 3$	$r = 1$ (T)	3	−0.200	0.030	1.024	0.194
	$r = 1$	4	−0.200	0.032	1.017	0.189
	$r = 2$ (T)	6	−0.198	0.032	1.019	0.188
	$r = 2$	9	−0.194	0.036	1.016	0.195
	$r = 3$ (T)	10	−0.192	0.033	1.013	0.205
	$r = 3$	16	−0.183	0.041	1.036	0.368

Table 4. *Results of estimating θ and σ^2 from 200 simulated realizations of a SAR model on a 15×15 grid with $\theta = 0.2$ and $\sigma^2 = 1$ using maximum likelihood and the simulation-based method proposed in this paper*

		dim ($\boldsymbol{\pi}$)	mean ($\hat{\theta}$)	s.d. ($\hat{\theta}$)	mean ($\hat{\sigma}^2$)	s.d. ($\hat{\sigma}^2$)
	ML		0.196	0.013	1.034	0.195
$s = 1$	$r = 1$ (T)	3	0.200	0.021	1.014	0.155
	$r = 1$	4	0.198	0.024	1.022	0.147
	$r = 2$ (T)	6	0.197	0.023	1.025	0.149
	$r = 2$	9	0.191	0.026	1.032	0.150
	$r = 3$ (T)	10	0.192	0.027	1.034	0.152
	$r = 3$	16	0.184	0.030	1.041	0.161
$s = 3$	$r = 1$ (T)	3	0.201	0.020	1.002	0.112
	$r = 1$	4	0.198	0.020	1.012	0.111
	$r = 2$ (T)	6	0.197	0.018	1.013	0.114
	$r = 2$	9	0.195	0.022	1.021	0.116
	$r = 3$ (T)	10	0.194	0.020	1.020	0.115
	$r = 3$	16	0.191	0.022	1.024	0.116

Table 5. *Results of estimating θ_1, θ_2, and σ^2 from 200 simulated realizations of a STAR model on a 10×10 grid with $\theta_1 = \theta_2 = \theta = 0.2$ and $\sigma^2 = 1$ using maximum likelihood and the simulation-based method proposed in this paper*

		dim ($\boldsymbol{\pi}$)	mean ($\hat{\theta}$)	s.d. ($\hat{\theta}$)	mean ($\hat{\sigma}^2$)	s.d. ($\hat{\sigma}^2$)
	ML		0.188	0.042	1.015	0.273
$s = 1$	$r = 1$ (T)	3	0.195	0.061	1.012	0.231
	$r = 1$	4	0.196	0.065	1.013	0.228
	$r = 2$ (T)	6	0.198	0.067	1.016	0.231
	$r = 2$	9	0.190	0.068	1.016	0.240
	$r = 3$ (T)	10	0.191	0.069	1.017	0.238
	$r = 3$	16	0.162	0.081	1.018	0.263
$s = 3$	$r = 1$ (T)	3	0.199	0.048	1.013	0.156
	$r = 1$	4	0.200	0.053	1.017	0.164
	$r = 2$ (T)	6	0.200	0.057	1.015	0.159
	$r = 2$	9	0.196	0.057	1.012	0.167
	$r = 3$ (T)	10	0.195	0.057	1.013	0.165
	$r = 3$	16	0.185	0.063	1.011	0.194

the sample size, from 10×10 to 15×15 grid, as can be seen in Table 4.

4.2. *A toy STAR model*

Consider the STAR model of order $p = 1$ and $q = 1$, with dependence θ_1 in time and dependence θ_2 in space:

$$Z(x,t) = \theta_1 Z(x,t-1) + \theta_2\{Z(x-1,t) + Z(x+1,t)\} + \epsilon(x,t), \quad (9)$$

where $\{\epsilon(x,t)\}$ has standard normal distribution, i.e. $\sigma^2 = 1$. This model is fairly unrealistic since the space component is of dimension one. However, potential applications can be, for instance, observations taken on a road or a river.

We start with the particular case of $\theta_1 = \theta_2 = \theta$. The process $Z(x,t)$ is then covariance-stationary if $|\theta| < 1/3$, and we study the situations $\theta = 0.2$. We use again simulated data on 10×10 grids. Table 5 provides the results for 200 replicates with the same auxiliary models as in the previous section. Results are similar to those in Table 2. We note, however, that the standard deviation of $\hat{\theta}$ in Table 5 is approximately 4/3 of the one in Table 2, probably due to the fact that we have now information about θ only in 3 directions instead of 4.

Next, we consider the case of $\theta_1 \neq \theta_2$. The process $Z(x,t)$ is then covariance-stationary if $|\theta_1| < 1$, $|\theta_2| < 1/2$, and the pair (θ_1, θ_2) lies in the rhombus with vertices $(1,0)$, $(0,1/2)$, $(-1,0)$, and $(0,-1/2)$. We study the situation $\theta_1 = 0.1$ and $\theta_2 = 0.2$. Note that maximum likelihood estimation is more complex to implement for situations $\theta_1 \neq \theta_2$ because no common term θ can be factorized out of the variance-covariance matrix of the data. Table 6 provides results for 200 replicates with the same auxiliary models as in the previous section, on simulated data on 10×10 grids. We note again that the standard deviation of $\hat{\theta}_1$ in Table 6 is approximately twice the one

Table 6. *Results of estimating θ_1, θ_2, and σ^2 from 200 simulated realizations of a STAR model on a 10 × 10 grid with $\theta_1 = 0.1$, $\theta_2 = 0.2$, and $\sigma^2 = 1$ using maximum likelihood and the simulation-based method proposed in this paper*

		dim ($\boldsymbol{\pi}$)	mean ($\hat{\theta}_1$)	s.d. ($\hat{\theta}_1$)	mean ($\hat{\theta}_2$)	s.d. ($\hat{\theta}_2$)	mean ($\hat{\sigma}^2$)	s.d. ($\hat{\sigma}^2$)
	ML		0.079	0.097	0.190	0.044	0.961	0.295
$s = 1$	$r = 1$ (T)	3	0.094	0.152	0.202	0.061	0.989	0.218
	$r = 1$	4	0.093	0.149	0.201	0.062	0.985	0.217
	$r = 2$ (T)	6	0.096	0.173	0.201	0.067	0.992	0.219
	$r = 2$	9	0.089	0.180	0.191	0.069	0.997	0.221
	$r = 3$ (T)	10	0.098	0.183	0.191	0.073	1.000	0.220
	$r = 3$	16	0.091	0.215	0.165	0.089	1.020	0.252
$s = 3$	$r = 1$ (T)	3	0.105	0.120	0.201	0.056	0.992	0.181
	$r = 1$	4	0.104	0.119	0.202	0.054	0.991	0.181
	$r = 2$ (T)	6	0.107	0.123	0.201	0.059	0.996	0.184
	$r = 2$	9	0.110	0.138	0.196	0.062	1.007	0.193
	$r = 3$ (T)	10	0.113	0.134	0.197	0.062	1.005	0.192
	$r = 3$	16	0.096	0.151	0.186	0.063	1.018	0.224

of $\hat{\theta}_2$, probably due to the fact that we have now information about θ_1 only in one direction, instead of two directions for θ_2.

4.3. *Robustness analysis*

In this section, we test the robustness performances of two simulation-based estimators and compare them with the maximum likelihood estimator. We consider replacement outliers (RO), i.e. although a SAR model is fitted to the data, the generating process is $Y(x, y) = (1 - B(x, y))Z(x, y) + B(x, y)H(x, y)$, where Z admits a simple SAR model as described in Section 4.1 with $\theta = 0.2$ and $\sigma^2 = 1$. The Bernoulli process is such that $P(B = 1) = \varepsilon = 5\%$, 10%, and the process H is normal with mean zero and variance Var $(H) = \eta^2$Var (Z) with $\eta^2 = 9$, 100. Results based on 200 replicates on 10 × 10 grids are presented in Table 7 for the maximum likelihood estimator (ML), the simulation-based estimator based on Yule-Walker (IYW for indirect Yule-Walker), and the estimator based on highly robust Yule-Walker (IYWR).

First, we see that without contaminations, i.e. $\varepsilon = 0$, the bias of the two simulation-based estimators is similar to the one of the maximum likelihood estimator. The variability of IYWR is slightly higher than IYW, as expected. However, this loss of efficiency is quite small due to the high efficiency of the robust autocovariance estimator (8). When $\varepsilon > 0$, we gain a great deal of robustness with IYWR. The ML and IYW estimators of θ are both attracted towards zero, i.e. the center of the parameter space. This effect has already been noted by Lucas (1997) in the context of time series, when defining a breakdown point towards zero. A similar effect can be seen in de Luna and Genton (2001) when studying the robustness of indirect estimators of ARMA processes. We note also that IYW performs slightly better than ML in the presence of RO, especially for the estimation of σ^2.

Table 7. *Results of estimating θ and σ^2 from 200 simulated realizations of a SAR model on a 10 × 10 grid with $\theta = 0.2$, and $\sigma^2 = 1$ using maximum likelihood and the simulation-based method proposed in this paper. Each realization is contaminated by ε% replacement outliers, normal with mean zero and variance η^2*

	mean ($\hat{\theta}$)	s.d. ($\hat{\theta}$)	mean ($\hat{\sigma}^2$)	s.d. ($\hat{\sigma}^2$)
$\varepsilon = 0\%$				
ML	0.193	0.024	1.018	0.316
IYW	0.200	0.030	1.018	0.173
IYWR	0.195	0.041	1.034	0.270
$\varepsilon = 5\%$, $\eta^2 = 9$				
ML	0.151	0.039	2.671	1.533
IYW	0.153	0.046	1.616	0.449
IYWR	0.179	0.046	1.275	0.358
$\varepsilon = 5\%$, $\eta^2 = 100$				
ML	0.067	0.052	52.026	59.264
IYW	0.078	0.072	6.267	3.658
IYWR	0.201	0.046	1.022	0.380
$\varepsilon = 10\%$, $\eta^2 = 9$				
ML	0.119	0.042	4.949	2.908
IYW	0.120	0.047	2.184	0.622
IYWR	0.165	0.049	1.527	0.452
$\varepsilon = 10\%$, $\eta^2 = 100$				
ML	0.030	0.042	162.751	144.949
IYW	0.046	0.075	11.602	5.501
IYWR	0.206	0.055	1.036	0.769

5. Illustrative example: Wheat data

In this section, a data set on wheat yields analyzed previously by several authors, among which Whittle (1954) and Cressie (1993), is used for illustration. The 500 observations (yields obtained on different plots) are arranged in a regular 20 × 25

Table 8. *Wheat yields estimation results of different models (first column) with different inferential methods (second column)*

Model	Inference	Location	Scale
QAR	YW	$\hat{\pi}_1 = 0.502, \hat{\pi}_2 = 0.193$	0.151
QAR	YWR	$\hat{\pi}_1 = 0.579, \hat{\pi}_2 = 0.196$	0.131
QAR	WhML	$\hat{\pi}_1 = 0.488, \hat{\pi}_2 = 0.202$	–
SAR	IYW	$\hat{\theta} = 0.159$	0.142
SAR	IYWR	$\hat{\theta} = 0.174$	0.120
SAR	WhML	$\hat{\theta} = 0.159$	–

Notes: SAR corresponds to Model (5). The inferential methods are labeled YW for Yule-Walker (only QAR models), WhML for ML estimators reported in Whittle (1954), IYW for the indirect estimator and IYWR for its robustified version, both as described in the text.

rectangular lattice, see Cressie (1993) for a detailed description. Although, fitting directly to the data a stationary SAR model has been argued not to be justified without first adjusting for a large scale variation, we fit a SAR model as Whittle did to have grounds for comparison. Thus, to the mean corrected data we fit

$$Z(x, y) = \theta\{Z(x, y-1) + Z(x, y+1) + Z(x-1, y) + Z(x+1, y)\} + \epsilon(x, y),$$

where $\{\epsilon(x, y)\}$ is assumed Normally distributed with mean zero and variance σ^2. For the indirect estimator the auxiliary QAR model was chosen through BIC. The model QAR 1(T), see Fig. 3, was picked. The use of BIC gives a consistent identification of the dimension of the QAR structure when the process admits a QAR representation (2) with $\{\nu_t\}$ an independently distributed process. Hence, in the present context we expect BIC to choose a QAR structure such that as little dependence as possible is left out in the process $\{\nu_t\}$. A robust indirect estimator was also used. Results are reported in Table 8. The results we obtain with the non-robustified simulation-based estimator are not very different from the results reported by Whittle (no estimate of scale was provided then). This is less true with the robustified estimator which provides a larger value for θ. This difference between robust and non-robust estimators may be due here either to the misspecification of the model noted above (need for a large scale component) or to the presence of outliers.

6. Conclusion

We have presented a simulation-based inferential method for simultaneous spatio-temporal stochastic processes. The parameters of such simultaneous models are complex to estimate and, for instance, least squares is not consistent. Gaussian maximum likelihood estimation is feasible and implemented in the module *S+SpatialStats* of *Splus*. When the model is known to be Gaussian then ML should be used since it has been proven to be most efficient, at least, for the location parameters. However, the simulation-based estimator enables us to perform inference for non-Gaussian simultaneous models, and, for instance, we have presented a version robust to the presence of outliers. The Monte Carlo study conducted shows that such a robustification of the estimator largely compensates the loss of efficiency under the uncontaminated model.

The stochastic processes considered in this work are defined on regular lattices. Although for the time dimension this regularity covers most practical applications, this is not the case for the spatial component. Thus, in order to give a wider range of application to the methodology herewith developed, we are planning to generalize it to irregular lattices.

Acknowledgments

We would like to thank the participants of the 18th Leeds Annual Statistics Research Workshop on spatial-temporal modeling and its applications for their helpful comments on preliminary results to those reported in this paper. We are also grateful to Sara Sjöstedt-de Luna and two anonymous referees for valuable comments that have improved the paper. The first author is grateful to the Wikström Foundation for its financial support.

References

Ali M.M. 1979. Analysis of stationary spatial-temporal processes: Estimation and prediction. Biometrika 66: 513–518.

Brännäs K. and de Luna X. 1998. Generalized method of moment and indirect estimation of the ARasMA model. Computational Statistics 13: 485–494.

Cressie N. 1993. Statistics for Spatial Data. John Wiley & Sons, New York.

Croux C. and Rousseeuw P.J. 1992. Time-efficient algorithms for two highly robust estimators of scale. Computational Statistics 1: 411–428.

de Luna X. and Genton M.G. 2001. Robust simulation-based estimation of ARMA models. Journal of Computational and Graphical Statistics 10: 370–387.

Genton M.G. and de Luna X. 2000. Robust simulation-based estimation. Statistics and Probability Letters 48: 253–259.

Gouriéroux C. and Monfort A. 1996. Simulation-Based Econometric Methods. Oxford University Press, Oxford.

Gouriéroux C., Monfort A., and Renault A.E. 1993. Indirect inference. Journal of Applied Econometrics 8: 85–118.

Guyon X. 1982. Parameter estimation for a stationary process on a d-dimensional lattice. Biometrika 69: 95–105.

Ha E. and Newton H.J. 1993. The bias of estimators of causal spatial autoregressive processes. Biometrika 80: 242–245.

Haining R. 1990. Spatial Data Analysis in the Social and Environmental Sciences. Cambridge University Press, Cambridge.

Hampel F.R. 1974. The influence curve and its role in robust estimation. Journal of the American Statistical Association 69: 383–393.

Hampel F.R., Ronchetti E.M., Rousseeuw P.J., and Stahel W.A. 1986. Robust Statistics, the Approach Based on Influence Functions. John Wiley & Sons, New-York.

Heuvelink G.B.M. 1998. Error Propagation in Environmental Modelling with GIS. Taylor & Francis Inc. Bristol.

Huber P.J. 1981. Robust Statistics. John Wiley & Sons, New-York.

Huber P.J. 1984. Finite sample breakdown point of M- and P-estimators. The Annals of Statistics 12: 119–126.

Kiiveri H.T. 1992. Fitting spatial correlation models: Approximating a likelihood approximation. Australian Journal of Statistics 34: 497–512.

Künsch H. 1984. Infinitesimal robustness for autoregressive processes. Annals of Statistics 12: 843–863.

Lucas A. 1997. Asymptotic robustness of least median of squares for autoregressions with additive outliers. Communication in Statistics, Theory and Methods 26: 2363–2380.

Ma Y. and Genton M.G. 2000. Highly robust estimation of the autocovariance function. Journal of Time Series Analysis 21: 663–684.

Martin R.D. and Yohai V.J. 1986. Influence functionals for time series (with discussion). The Annals of Statistics 14: 781–855.

Rosenblatt M. 1986. Parameter estimation for finite-parameter stationary random fields. Journal of Applied Probability A23: 311–318.

Rousseeuw P.J. and Croux C. 1992. Explicit scale estimators with high breakdown point. In: Dodge Y. (Ed.), L_1 Statistical Analyses and Related Methods. pp. 77–92.

Rousseeuw P.J. and Croux C. 1993. Alternatives to the median absolute deviation. Journal of the American Statistical Association 88: 1273–1283.

Tjøstheim D. 1978. Statistical spatial series modeling. Advances in Applied Probability 10: 130–154.

Whittle P. 1954. On stationary processes in the plane. Biometrika 41: 434–449.