RATIONAL SPECTRAL DENSITY MODELS FOR LATTICE DATA

L. IPPOLITI,* R. J. MARTIN
Department of Economics
University G. d'Annunzio, Viale Pindaro 42, 65127 Pescara, Italy

R. J. BHANSALI
Division of Statistics and Probability, Department of Mathematical Sciences
University of Liverpool, Liverpool L69 7ZL, UK

August 25, 2013

ABSTRACT. Conditional autoregressive CAR models, possibly with added noise, unilateral ARMA models, and directly specified correlation DC models, are widely used classes of spatial models. In this paper, we consider their generalization to all models with a rational spectral density function. These models allow a wider range of correlation behaviour, and can provide adequate fits to data with fewer parameters. Some theoretical properties are presented, and comparisons made with CAR correlations. Some methods for estimation are discussed, and fits to some real data are compared.

Dedicated to the memory of J. E. Besag, 1945-2010.

Key Words: Conditional autoregression, Gauss-Markov Random Fields, Image Analysis, Rational Spectral Density Models, Spatial Processes.

1 Introduction

Large amounts of essentially-continuous spatial data are associated with the nodes or interiors of a regular rectangular lattice. Examples include pixellated images which occur in many different applications, regularly-sampled spatial data, and many agricultural field trials. Different types of models have been proposed for analyzing such data. Four main classes are: i) those with a directly specified correlation structure, such as those used in geostatistics; ii) those specified by a generating model involving 'past' (using some site ordering) values and uncorrelated innovations - unilateral (or causal) autoregressive-moving average ARMA models; iii) those specified by a formal equation involving 'past' and 'future' values and uncorrelated errors - simultaneous autoregressions SAR; and iv) those that specify the conditional distribution at each site given

*Corresponding Author: ippoliti@unich.it

the values at all other sites - conditional autoregressive CAR models. An extension to SAR and CAR models adds an independent noise term.

We (essentially) only consider univariate stationary Gaussian models here - i.e. they have finite variance, and the correlation between the observations at two sites only depends on the relative positions of the two sites. The assumption of Normality for analysing a data set is convenient and often reasonable (perhaps after transformation). We assume models are defined on an infinite regular rectangular lattice, and applied to data on a complete finite lattice. Then a CAR, also known as a Gauss-Markov random field, is defined by its conditional means. Every unilateral AR, and every SAR, is equivalent to a CAR (in the sense of having the same correlation structure).

Since many unilateral ARMA models depend on the choice of site ordering, they can appear arbitrary, but they can have some useful properties, especially if the model is separable (i.e. the correlation function is a product of lower-dimensional correlation functions). For example, simulation and likelihood evaluation can be simple. SAR models have some severe logical difficulties (e.g. the errors are correlated with all the observations, and in general the parameters are not uniquely determined).

In general, the correlation structure of a CAR model is hard to determine (except numerically), but the inverse correlations are directly specified. On a planar lattice, this gives most, but not all, elements of the inverse dispersion matrix which is required for Gaussian maximum likelihood estimation - see section 6.1. Another computational difficulty with many CAR models used in practice is that unless the dependence is weak, the parameters are usually very close to the stationarity boundary. Since the correlations of a CAR can decay very slowly from 1, the extension to the noisy CAR has been found useful.

Geostatistical models are defined for continuous space, and are widely used for analyzing data defined on irregular sites or regions - see, for example, Cressie (1993, section 2.3.1). On a regular lattice, they and other models specifying the correlations usually have the drawback that the inverse dispersion matrix does not have a simple form. These lattice models, and moving-average models, in general have an infinite CAR representation.

In time series, the extension of AR and MA models to the ARMA models, which have a rational spectral density, has been extremely useful. The unilateral ARMA models on a lattice have a rational spectral density, but are only a subset of all the possibilities. In this paper, we consider the generalization of finite CARs, unilateral ARMAs, and finite DCs (directly-specified correlation models for which the correlations are 0 outside a neighbourhood of the origin), to all models with a rational spectral density function - RSDs. The RSDs in general have more possible correlation structures, and can give more parsimonious fits to data. Also, they are less likely than CARs to have the estimated parameters very close to the stationarity boundary.

After reviewing the standard lattice models in section 2, the RSD model is defined and some of its properties discussed in section 3. We show how RSDs can arise from operations on CARs in section 4, and compare their correlations with those of CARs in section 5. Section 6 discusses how standard methods for model fitting and identification can be extended to RSDs, and in section 7, fits of models to real data are compared. Note that the results in sections 2 to 5 only depend on second-order properties, and so hold for any distribution. For convenience, we mainly refer here to Guyon (1995), Cressie (1993) and Rue and Held (2005) for known results.

2 Lattice Models

In this section we discuss the usual lattice models. We begin with some definitions in section 2.1, and then review the usual AR, SAR, MA, ARMA, CAR, and DC lattice models in sections 2.2 and 2.3. Some extensions are given in section 2.4.

2.1 Preliminaries

Suppose that $\boldsymbol{t}$, $\boldsymbol{u}$, $\boldsymbol{z}$ and $\boldsymbol{\lambda}$ are $d-$dimensional vectors, and assume that $\{x(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{Z}^d\}$ is a second-order stationary random field on the regular rectangular lattice, with mean zero, autocovariance function $R_x(\boldsymbol{u}) = Cov\{x(\boldsymbol{t}), x(\boldsymbol{t}+\boldsymbol{u})\}$, and autocorrelation function $r_x(\boldsymbol{u}) = R_x(\boldsymbol{u})/\sigma_x^2$, where $\sigma_x^2 = R_x(\boldsymbol{0})$. Provided the sum is finite, the autocovariance generating function - acgf - of x is

$$\Gamma_x(\boldsymbol{z}) = \sum_{\boldsymbol{u}\in\mathbb{Z}^d} R_x(\boldsymbol{u})\boldsymbol{z}^{\boldsymbol{u}}, \quad \boldsymbol{z} \in \mathbb{C}^d$$

where $\boldsymbol{z}^{\boldsymbol{u}} = \prod_{j=1}^d \boldsymbol{z}_j^{u_j}$. Stationarity ensures that the acgf always exists for $|\boldsymbol{z}| = 1$, where $|\boldsymbol{z}| = \sqrt{(\boldsymbol{z}'\boldsymbol{z})}$. Except in section 2.4, we assume that $\sum |R_x(\boldsymbol{u})| < \infty$. The spectral density function - sdf - of x is defined here as $f_x(\boldsymbol{\lambda}) = \Gamma_x(e^{i\boldsymbol{\lambda}}) = \sum R_x(\boldsymbol{u})e^{i\boldsymbol{u}'\boldsymbol{\lambda}} = \sum R_x(\boldsymbol{u})\cos(\boldsymbol{u}'\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} \in (-\pi, \pi]^d$; with its integral being $(2\pi)^d\sigma_x^2$. The inverse relationship is

$$R_x(\boldsymbol{u}) = (2\pi)^{-d} \int e^{i\boldsymbol{u}'\boldsymbol{\lambda}} f_x(\boldsymbol{\lambda}) d\boldsymbol{\lambda}. \tag{1}$$

Provided $f_x(\boldsymbol{\lambda}) > 0$ for all $\boldsymbol{\lambda}$, the inverse covariance and correlation functions of x are, respectively,

$$\tilde{R}_x(\boldsymbol{u}) = (2\pi)^{-d} \int e^{i\boldsymbol{u}'\boldsymbol{\lambda}} \big[f_x(\boldsymbol{\lambda})\big]^{-1} d\boldsymbol{\lambda} \quad \text{and} \quad \tilde{r}_x(\boldsymbol{u}) = \tilde{R}_x(\boldsymbol{u})/\tilde{R}_x(\boldsymbol{0}). \tag{2}$$

For ARMA models we need to define $P(\boldsymbol{z})$ and $Q(\boldsymbol{z})$ as finite Laurent series (about $\boldsymbol{0}$) by

$$P(\boldsymbol{z}) = 1 - \sum_{\boldsymbol{j}\in S_p} \phi_{\boldsymbol{j}} \boldsymbol{z}^{\boldsymbol{j}} \quad \text{and} \quad Q(\boldsymbol{z}) = 1 + \sum_{\boldsymbol{j}\in S_q} \theta_{\boldsymbol{j}} \boldsymbol{z}^{\boldsymbol{j}},$$

where S_m is a finite subset of $\mathbb{Z}^d$ containing neighbours of the origin. Thus $\boldsymbol{t} + S_m$ is the set of neighbours of site $\boldsymbol{t}$. The order of the neighbourhood set is denoted by m, and in the usual way is defined sequentially by the maximum distance between the origin and a point in S_m. Thus, for $d = 2$, the first-order ($m = 1$) neighbours of a site are those 4 sites which are adjacent to it; and the second-order neighbours ($m = 2$) are these plus the 4 diagonally adjacent sites. Let S_m^+ denote $S_m \cup \{\boldsymbol{0}\}$, the origin and its neighbouring sites.
For CAR, DC and RSD models we need

$$A(\boldsymbol{z}) = -\sum_{\boldsymbol{j}\in S_p^+} \alpha_{\boldsymbol{j}} \boldsymbol{z}^{\boldsymbol{j}} \quad \text{and} \quad B(\boldsymbol{z}) = \sum_{\boldsymbol{j}\in S_q^+} \beta_{\boldsymbol{j}} \boldsymbol{z}^{\boldsymbol{j}}$$

to be finite symmetric Laurent series with $\alpha_{\boldsymbol{0}} = -1$, $\beta_{\boldsymbol{0}} = 1$, and satisfying $A(\boldsymbol{z}) = A(\boldsymbol{z}^{-1})$ and $B(\boldsymbol{z}) = B(\boldsymbol{z}^{-1})$. Thus $\alpha_{\boldsymbol{j}} = \alpha_{-\boldsymbol{j}}$ and $\beta_{\boldsymbol{j}} = \beta_{-\boldsymbol{j}}$ for all $\boldsymbol{j}$.
We call $A(\boldsymbol{z})$ reflection-symmetric if the $\alpha_{\boldsymbol{j}}$ are equal for all sign changes on the $\boldsymbol{j}_u$, and completely-symmetric if they are also equal for all permutations of $\boldsymbol{j}$, and similarly for $B(\boldsymbol{z})$. Then, for a model defined by $A(\boldsymbol{z})$ and/or $B(\boldsymbol{z})$, a similar property holds for the correlations $r_x(\boldsymbol{u})$ and inverse

correlations $\tilde{r}_x(\boldsymbol{u})$. We refer to a CAR($p$) as reflection-symmetric, RS-CAR(p), or completely-symmetric, CS-CAR(p), if its $A(\boldsymbol{z})$ has this property; and similarly for a DC with $B(\boldsymbol{z})$, and for a RSD with both $A(\boldsymbol{z})$ and $B(\boldsymbol{z})$.
A White-Noise process, $\epsilon(\boldsymbol{t})$, has uncorrelated components with zero mean, and constant variance σ_ϵ^2. Thus $\Gamma_\epsilon(\boldsymbol{z}) = \sigma_\epsilon^2$.
Let $\boldsymbol{T}$ be a shift operator on an index, such that $\boldsymbol{T}^{\boldsymbol{j}} x(\boldsymbol{t}) = x(\boldsymbol{t}-\boldsymbol{j})$. In general, if $x(\boldsymbol{t}) = H(\boldsymbol{T})y(\boldsymbol{t})$, where y is a stationary process, $H(\boldsymbol{z}) = \sum_{\boldsymbol{j}} h_{\boldsymbol{j}} \boldsymbol{z}^{\boldsymbol{j}}$, and $\sum |h_{\boldsymbol{j}}|$ is finite, standard results show that

$$\Gamma_x(\boldsymbol{z}) = H(\boldsymbol{z})H(\boldsymbol{z}^{-1})\Gamma_y(\boldsymbol{z}), \tag{3}$$

$$\Gamma_{yx}(\boldsymbol{z}) = \sum \mathrm{Cov}\{y(\boldsymbol{t}), x(\boldsymbol{t}+\boldsymbol{u})\}\boldsymbol{z}^{\boldsymbol{u}} = H(\boldsymbol{z})\Gamma_y(\boldsymbol{z}). \tag{4}$$

2.2 AR, MA and ARMA models

Assuming $P(\boldsymbol{z})$ and $Q(\boldsymbol{z})$ have no common factors, the ARMA(p, q) model is

$$P(\boldsymbol{T})x(\boldsymbol{t}) = Q(\boldsymbol{T})\epsilon(\boldsymbol{t}),$$

or

$$x(\boldsymbol{t}) = \sum_{\boldsymbol{j}\in S_p} \phi_{\boldsymbol{j}} x(\boldsymbol{t}-\boldsymbol{j}) + \epsilon(\boldsymbol{t}) + \sum_{\boldsymbol{j}\in S_q} \theta_{\boldsymbol{j}} \epsilon(\boldsymbol{t}-\boldsymbol{j}), \quad \boldsymbol{t} \in \mathbb{Z}^d.$$

In general, stationarity requires that $P(\boldsymbol{z}) \neq \boldsymbol{0}$ for $|\boldsymbol{z}| = 1$, and invertibility (to ensure a valid infinite AR representation) requires that $Q(\boldsymbol{z}) \neq \boldsymbol{0}$ for $|\boldsymbol{z}| = 1$.
From equation (3),

$$\Gamma_x(\boldsymbol{z}) = \sigma_\epsilon^2 \frac{Q(\boldsymbol{z})Q(\boldsymbol{z}^{-1})}{P(\boldsymbol{z})P(\boldsymbol{z}^{-1})},$$

with

$$\Gamma_{\epsilon x}(\boldsymbol{z}) = \sigma_\epsilon^2 \frac{Q(\boldsymbol{z})}{P(\boldsymbol{z})}.$$

An autoregressive AR(p) process has $Q(\boldsymbol{z}) = 1$, and a moving average MA(q) model has $P(\boldsymbol{z}) = 1$. An AR is unilateral if $P(\boldsymbol{z})$ is such that $P(\boldsymbol{T})x(\boldsymbol{t})$ only involves values at $\boldsymbol{t}$ and 'previous' sites. Then stationarity requires $P(\boldsymbol{z}) \neq \boldsymbol{0}$ for $|\boldsymbol{z}| < 1$. The bilateral model is the SAR, which has $x(\boldsymbol{t})$ correlated with all $\epsilon(\boldsymbol{t})$.
Note that $\Gamma_x(\boldsymbol{z})$ for the ARMA is a ratio of finite Laurent series, and that the sdf can be regarded as the ratio of two AR sdf's, or of two MA sdf's.

2.3 CAR and DC models

Let $[x(\boldsymbol{t})|\cdot] = \left[x(\boldsymbol{t})|x(\boldsymbol{t}-\boldsymbol{j}) : \boldsymbol{j} \in \mathbb{Z}^d \backslash \{\boldsymbol{0}\}\right]$ denote $x(\boldsymbol{t})$ conditional on the values at all other sites. Then, under Normality, the conditional autoregression CAR(p) or Gauss-Markov random field is defined by the conditional mean,

$$E\left[x(\boldsymbol{t})|\cdot\right] = \sum_{\boldsymbol{j}\in S_p} \alpha_{\boldsymbol{j}} x(\boldsymbol{t}-\boldsymbol{j}),$$

where $\alpha_{\boldsymbol{j}} = \alpha_{-\boldsymbol{j}} = -\tilde{r}_x(\boldsymbol{j})$ for all $\boldsymbol{j}$, and the constant conditional variance

$$\mathrm{Var}\left[x(\boldsymbol{t})|\cdot\right] = \mathrm{Var}[\eta(\boldsymbol{t})] = \sigma_\eta^2 = \{\tilde{R}_x(\boldsymbol{0})\}^{-1},$$

where

$$\eta(\boldsymbol{t}) = x(\boldsymbol{t}) - E\left[x(\boldsymbol{t})|\cdot\right]$$

is the interpolation error process. Thus a CAR can be regarded as defined by its inverse covariances, whereas a DC (see below) is defined by its covariances. Then

$$A(\boldsymbol{T})x(\boldsymbol{t}) = \eta(\boldsymbol{t}), \quad \boldsymbol{t} \in \mathbb{Z}^d, \tag{5}$$

where $\Gamma_{\eta x}(\boldsymbol{z}) = \sigma_\eta^2$, i.e. the interpolation error $\eta(\boldsymbol{t})$ is uncorrelated with all x's except $x(\boldsymbol{t})$. Provided the stationarity condition $A(\boldsymbol{z}) \neq \boldsymbol{0}$ for $|\boldsymbol{z}| = 1$ is satisfied, from equation (4) the acgf of the CAR is $\Gamma_x(\boldsymbol{z}) = \sigma_\eta^2/A(\boldsymbol{z})$, and the acgf of $\eta(\boldsymbol{t})$ is $\Gamma_\eta(\boldsymbol{z}) = \sigma_\eta^2 A(\boldsymbol{z})$. Thus the interpolation process is correlated, with $r_\eta(\boldsymbol{u}) = \tilde{r}_x(\boldsymbol{u})$. Hence, the interpolation error for a CAR(p) is a DC(p) (see below).

Multiplying equation (5) by $x(\boldsymbol{t}+\boldsymbol{u})$ and taking expectations shows that

$$A(\boldsymbol{T})R_x(\boldsymbol{u}) = \sigma_\eta^2 \delta_{\boldsymbol{u}} \tag{6}$$

where $\delta_{\boldsymbol{0}} = 1$ and $\delta_{\boldsymbol{u}} = 0$ otherwise.

If x is not Gaussian, the conditional mean $E\left[x(\boldsymbol{t})|\cdot\right]$, differs from $\{1 - A(\boldsymbol{T})\}x(\boldsymbol{t})$, but a stationary correlation structure results from the above if $E\left[x(\boldsymbol{t})|\cdot\right]$ is replaced by the linear least-squares interpolator of $x(\boldsymbol{t})$, and $\eta(\boldsymbol{t})$ is the corresponding interpolation error.

Note that all SAR and unilateral AR are equivalent to (i.e. have the same correlation structures as) a CAR, since $P(\boldsymbol{z})P(\boldsymbol{z}^{-1})$ is proportional to an $A(\boldsymbol{z})$. For $d = 1$ both the CAR and SAR models are also equivalent to finite AR models. However, for $d \geq 2$ $A(\boldsymbol{z})$ cannot in general be appropriately factorised, and so a CAR does not in general have a (finite) unilateral AR or SAR representation (see Guyon, 1995, Theorem 1.3.2).

The stationarity conditions are not always intuitively obvious. For example, if $d = 2$, the stationarity condition for a CAR(1) is $|\alpha_{10}| + |\alpha_{01}| < 1/2$, and for the RS-CAR(2), with $\alpha_{11} = \alpha_{1-1}$, the conditions are $|\alpha_{10} + \alpha_{01}| + 2\alpha_{11} < 1/2$ and $|\alpha_{10} - \alpha_{01}| - 2\alpha_{11} < 1/2$. For the CS-CAR(3) it is necessary that $|\alpha_{10}| < 1/2$, $|\alpha_{11}| < 1/2$, $|\alpha_{20}| < 1/4$, $|\alpha_{10}| + \alpha_{11} + \alpha_{20} < 1/4$, $\alpha_{20} - \alpha_{11} < 1/4$, and if $\alpha_{20} + \alpha_{11} < 0$, then we also need $\alpha_{10}^2 < -(\alpha_{11} + 2\alpha_{20})(1 + 4\alpha_{20})$ - see Appendix 1.

In general, no explicit form for the correlation function of a CAR exists. The correlations can be obtained numerically through equation (1) and, provided care is taken over numerical accuracy, recursions in their values using equation (6). An alternative method is to obtain the covariance matrix on a sufficiently large torus lattice (see section 6.1). The CAR correlations decay exponentially - see section 3.1, but possibly very slowly (Besag, 1981).

Let F denote the scaled interpolation variance, σ_η^2/σ_x^2. The measure $1 - F$ has been used as an index of linear determinism, to show how well $E[x(\boldsymbol{t})|.]$ predicts $x(\boldsymbol{t})$ (see, e.g., Bhansali and Ippoliti, 2005). It follows from the recursion (6) that

$$F = 1 - \sum_{\boldsymbol{j} \in S_p} \alpha_{\boldsymbol{j}} r_x(\boldsymbol{j}) = \sum_{\boldsymbol{j} \in S_p^+} r_x(\boldsymbol{j}) \tilde{r}_x(\boldsymbol{j}).$$

The CAR specification also has the unfortunate property that even modest low-lag correlations can require the sum of the parameters to be very close to the stationarity boundary. For example, in two

dimensions, the CS-CAR(1), for which $|\alpha_{10}| < 1/4$, requires α_{10} to be 0.24565 for $r_x(1,0) = 0.50$, and 0.249993 for $r_x(1,0) = 0.75$. See also Table A1 in Bartlett (1975). In fact (see Appendix 2), if $\alpha_{10} = 1/4 - \varsigma$, then as $\varsigma \to 0$, $1 - r_x(1,0) \simeq \pi/\log(2/\varsigma)$, or for $r_x(1,0) = 1 - \tau$, with $\tau > 0$ small, we need $\alpha_{10} \simeq 1/4 - 2\exp(-\pi/\tau)$. For example, for $r_x(1,0) = 0.9$, we have $\alpha_{10} \simeq 1/4 - 4.5 \times 10^{-14}$, and for $r_x(1,0) = 0.98$, we have $\varsigma = 1.2 \times 10^{-68}$. This approximation works very well for $\alpha_{10} > 0.249$ or $r_x(1,0) > 0.59$.
Higher-order CAR fits also tend to have $\sum \hat{\alpha}_{\boldsymbol{j}}$ close to the stationarity limit - see Besag and Kooperberg (1995). This can make parameter estimation very difficult. An approach for overcoming this problem is to use a non-stationary intrinsic CAR model, for which $\sum \alpha_{\boldsymbol{j}}$ equals the stationarity limit - see Besag and Kooperberg (1995) or Rue and Held (2005, Ch. 3). However, to use these models seems unsatisfactory when the low-lag correlations are not all that large.
A finite direct correlation DC(q) model is one for which the correlations are directly specified, and has a finite range - the correlations are all zero outside S_q^+, i.e. $r_x(\boldsymbol{j}) = 0, \forall\, \boldsymbol{j} \notin S_q^+\}$. Let $\beta_{\boldsymbol{j}} = \beta_{-\boldsymbol{j}} = r_x(\boldsymbol{j})$ for all $\boldsymbol{j}$. Then

$$\Gamma_x(\boldsymbol{z}) = \sigma_x^2 B(\boldsymbol{z}).$$

The invertibility condition on $B(\boldsymbol{z})$ corresponds to the above stationarity condition on $A(\boldsymbol{z})$ for a CAR, and similar comments apply to the equivalence or non-equivalence of MA models to DC models. If S_q is small, correlations may need to be small - for example, the CS-DC(1) has $|r_x(1,0)| < 1/(2d)$. Note that many directly-specified correlation models used in geostatistics, such as the Matérn or Gaussian correlation structures, do not have a finite range, and do not have a finite CAR representation. However, the commonly-used spherical model is a DC.

2.4 Some other lattice models

We briefly note here some other proposed lattice models.
Kiiveri and Campbell (1989) suggested a generalisation of the CAR and DC models, which includes symmetric SAR and MA models, by (essentially) specifying an acgf proportional to $1/A(\boldsymbol{z})^{\phi}$, where $A(\boldsymbol{z})$ is as in section 2.1, and $\phi \neq 0$ is a parameter to be estimated. The SAR, CAR, DC and MA models are special cases with $\phi = 2, 1, -1, -2$. The model proposed by Lindgren et al. (2011) is the special case where ϕ is a positive integer and $A(\boldsymbol{z})$ corresponds to a CAR(1).
It was assumed in section 2.1 that $f(\mathbf{0}) < \infty$, but stationary models with long-range dependence have $f(\mathbf{0}) = \infty$, and correlations that decay more slowly than exponentially. Simple examples when $d = 1$ are the self-similar process, fractionally differenced noise, and the harmonic correlation structure - see, e.g., Martin and Walker (1997).

3 RSD models

3.1 Definition and properties

For $d = 1$, ARMA models are a generalisation of AR and MA models, which are characterised by having a rational spectral density (Priestley, 1981, section 4.12.4). That is, the acgf is a rational function: a ratio of directly-specified finite symmetric Laurent series. For $d > 1$, the CAR, unilateral ARMA, and DC models can be extended to models for which the acgf is a rational function. This extension includes the noisy CAR, and allows correlation structures which cannot arise from a (finite) CAR, unilateral ARMA, or DC.

Assume that $A(\boldsymbol{z})$ and $B(\boldsymbol{z})$ are as in section 2.1, have no common factors, and that $A(\boldsymbol{z}) \neq 0$ and $B(\boldsymbol{z}) \neq 0$ for $|\boldsymbol{z}| = 1$.

Definition. A rational spectral density model RSD(p, q) for x has an acgf which is proportional to a ratio of finite symmetric Laurent series, $B(\boldsymbol{z})/A(\boldsymbol{z})$.

The sdf $f_x(\boldsymbol{\lambda})$ can be regarded as the ratio of two CAR sdf's, or of two DC sdf's. Although lattice RSD models have been suggested at least as early as 1980, and are mentioned in Guyon (1995, section 1.4), they do not appear to have been used or discussed further. Continuous-space RSD models were proposed by Vecchia (1985), and have recently been developed by Bolin and Lindgren (2011).

Clearly CAR and DC models are special cases. Note that the correlations for the model with acgf proportional to $B(\boldsymbol{z})/A(\boldsymbol{z})$ are the inverse correlations of the model with acgf proportional to $A(\boldsymbol{z})/B(\boldsymbol{z})$.

Suppose $B(\boldsymbol{z}) \neq 1$ and let $C(\boldsymbol{z}) = \sum_{\boldsymbol{u} \in \mathbb{Z}^d} c_{\boldsymbol{u}} \boldsymbol{z}^{\boldsymbol{u}}$ denote the infinite series for $A(\boldsymbol{z})/B(\boldsymbol{z})$. Then we can use equation (5) to get the infinite CAR representation for x in terms of its interpolation errors η. Since in general $c_0 \neq 1$ and the coefficient of $x(\boldsymbol{t})$ in equation (5) must be 1, the infinite CAR representation for x is

$$c_0^{-1} C(\boldsymbol{T}) x(\boldsymbol{t}) = \eta(\boldsymbol{t}), \tag{7}$$

with acgf $B(\boldsymbol{z})/A(\boldsymbol{z})$ times $c_0 \sigma_\eta^2$. Thus the proportionality constant in the definition can be interpreted as $c_0 \sigma_\eta^2$. We show in Lemma 1 below that $c_0 = 1/\sum_{\boldsymbol{u} \in S_q^+} \beta_{\boldsymbol{u}} \tilde{r}_x(\boldsymbol{u})$.

We can then formally write

$$A(\boldsymbol{T}) x(\boldsymbol{t}) = c_0 B(\boldsymbol{T}) \eta(\boldsymbol{t}), \tag{8}$$

which expresses a finite combination of the x's in terms of a finite combination of the interpolation errors η, and has some similarities to the one-dimensional ARMA representations, but here the η's are correlated. We can then refer to this RSD as a conditional ARMA model.

The conditional ARMA representation (8) of the RSD leads to extended Yule-Walker equations for the covariances and the inverse covariances, generalising equation (6). Multiplying equation (8) by $x(\boldsymbol{t} + \boldsymbol{u})$ and taking expectations gives

$$A(\boldsymbol{T}) R_x(\boldsymbol{u}) = c_0 \sigma_\eta^2 \beta_{\boldsymbol{u}} \quad \forall\, \boldsymbol{u} \in \mathbb{Z}^d. \tag{9}$$

Multiplying equation (8) by $\eta(\boldsymbol{t} + \boldsymbol{u})$, taking expectations, and using $R_\eta(\boldsymbol{u}) = \sigma_\eta^2 r_\eta(\boldsymbol{u})$ and $r_\eta(\boldsymbol{u}) = \tilde{r}_x(\boldsymbol{u})$, gives

$$-\alpha_{\boldsymbol{u}} = c_0 B(\boldsymbol{T}) \tilde{r}_x(\boldsymbol{u}) \quad \forall\, \boldsymbol{u} \in \mathbb{Z}^d. \tag{10}$$

Lemma 1 shows how to evaluate c_0 and the scaled interpolation variance F for a RSD model.

Lemma 1. If $C(\boldsymbol{z}) = A(\boldsymbol{z})/B(\boldsymbol{z})$ and $\Gamma_x(\boldsymbol{z}) = c_0 \sigma_\eta^2 / C(\boldsymbol{z})$, then $c_0 = 1/\sum_{\boldsymbol{u} \in S_q^+} \beta_{\boldsymbol{u}} \tilde{r}_x(\boldsymbol{u})$, and

$$\sigma_\eta^2 / \sigma_x^2 = F = \big\{1 - \sum_{\boldsymbol{u} \in S_p} \alpha_{\boldsymbol{u}} r_x(\boldsymbol{u})\big\} \times \big\{\sum_{\boldsymbol{u} \in S_q^+} \beta_{\boldsymbol{u}} \tilde{r}_x(\boldsymbol{u})\big\}.$$

Proof. From equation (10) with $\boldsymbol{u} = \boldsymbol{0}$, we have

$$\big\{c_0 B(\boldsymbol{T}) \tilde{r}_x(\boldsymbol{u})|_{\boldsymbol{u}=\boldsymbol{0}}\big\} = 1, \text{ so that } c_0 \big\{\sum_{\boldsymbol{u} \in S_q^+} \beta_{\boldsymbol{u}} \tilde{r}_x(\boldsymbol{u})\big\} = 1.$$

From equation (9) with $\boldsymbol{u} = \mathbf{0}$,

$$\big\{A(\boldsymbol{T})r_x(\boldsymbol{u})\sigma_x^2|_{\boldsymbol{u}=\mathbf{0}}\big\} = c_0\sigma_\eta^2, \text{ so that } \sigma_x^2\big\{1 - \sum_{\boldsymbol{u}\in S_p}\alpha_{\boldsymbol{u}}r_x(\boldsymbol{u})\big\} = c_0\sigma_\eta^2. \quad \square$$

Note that if $B(\boldsymbol{z}) = 1$, Lemma 1 gives $c_0 = 1$, and F as in section 2.3.
As with a CAR, the sdf of a RSD is a real analytic function, and hence its correlations and inverse correlations decay exponentially. Let $\bar{u}$ denote $\sum|u_i|$. Then, extending Guyon's (1995, section 1.4) result to the inverse correlations, there exist ψ_1, ψ_2, with $0 < \psi_i < 1$, and m_1, m_2 such that for all $\boldsymbol{u}$

$$|r_x(\boldsymbol{u})| \le m_1\psi_1^{\bar{u}}, \quad |\tilde{r}_x(\boldsymbol{u})| \le m_2\psi_2^{\bar{u}}.$$

3.2 Relationship with CAR correlations

Suppose y is a CAR(p) defined by $A(\boldsymbol{z})$ and x is a RSD(p, q) with acgf proportional to $B(\boldsymbol{z})/A(\boldsymbol{z})$. Then $\Gamma_x(\boldsymbol{z})$ is proportional to

$$B(\boldsymbol{z})\boldsymbol{\Gamma}_y(\boldsymbol{z}) = B(\boldsymbol{z})\left[\sum_{\boldsymbol{u}\in\mathbb{Z}^d} R_y(\boldsymbol{u})z^{\boldsymbol{u}}\right].$$

Thus the correlations of x can be expressed as a finite linear combination of those of y. Although CAR correlations are usually not readily available, this result shows the effect of $B(\boldsymbol{z})$ on these correlations.
Similarly, the inverse correlations $\tilde{r}_x(\boldsymbol{u})$ can be expressed as a finite linear combination of the inverse correlations of the DC defined by $B(\boldsymbol{z})$.

Example 3.1.
Suppose $d = 2$, with x a RSD(p, 1). Then, for a constant $\mathcal{K}$,

$$\boldsymbol{\Gamma}_x(z_1, z_2) = \mathcal{K}\left[1 + \beta_{10}(z_1 + z_1^{-1}) + \beta_{01}(z_2 + z_2^{-1})\right]\sum_{u_1,u_2} r_y(u_1, u_2)z_1^{u_1}z_2^{u_2}.$$

Thus, for example,

$$\begin{aligned} R_x(0,0) &= \mathcal{K}\left[1 + 2\,\beta_{10}\,r_y(1,0) + 2\,\beta_{01}\,r_y(0,1)\right], \\ R_x(1,0) &= \mathcal{K}\left\{r_y(1,0) + \beta_{10}[1 + r_y(2,0)] + \beta_{01}[r_y(1,-1) + r_y(1,1)]\right\}, \end{aligned}$$

and in general $R_x(u_1, u_2) =$

$$\mathcal{K}\big\{r_y(u_1,u_2) + \beta_{10}[r_y(u_1-1,u_2) + r_y(u_1+1,u_2)] + \beta_{01}[r_y(u_1,u_2-1) + r_y(u_1,u_2+1)]\big\}.$$

Example 3.2.
If a White Noise process, ϵ, is added to the CAR(p), y, the result is a special case of a RSD(p, p). This is often referred to as an errors-in-variables CAR model, or a noisy CAR model.
For this, we obtain directly that $\Gamma_x(\boldsymbol{z}) = \Gamma_y(\boldsymbol{z}) + \sigma_\epsilon^2$. Then

$$r_x(\boldsymbol{u}) = r_y(\boldsymbol{u})/(1 + \psi)$$

for $\boldsymbol{u} \neq \mathbf{0}$, with $\psi = \sigma_\epsilon^2/\sigma_y^2$. Thus, the noisy CAR reduces all the (absolute) correlations (except at $\boldsymbol{u} = \mathbf{0}$) by a constant. On the other hand, by taking ψ as negative (essentially using a negative

variance σ_ϵ^2), we can have a RSD model which increases all the (absolute) correlations (except at $\boldsymbol{u} = \mathbf{0}$) by a constant. See also Example 4.1.

Example 3.3.
In general, we can express $B(\boldsymbol{z})/A(\boldsymbol{z})$ as $1 + \big\{B(\boldsymbol{z}) - A(\boldsymbol{z})\big\}/A(\boldsymbol{z})$. Suppose $d = 2$, and consider the RSD with $B(z_1, z_2) = A(z_1, z_2) + \beta_{20}(z_1^2 + z_1^{-2} + z_2^2 + z_2^{-2})$. Let F_C denote σ_η^2/σ_y^2 for the CAR(p) y defined by $A(z_1, z_2)$. Then, $R_x(0,0) = \mathcal{K}\big\{1 + 2\,\beta_{20}\,[r_y(2,0) + r_y(0,2)]/F_C\big\}$, and in general for $(u_1, u_2) \neq (0,0)$,

$$R_x(u_1, u_2) = \mathcal{K}\,\beta_{20}\big\{r_y(u_1 - 2, u_2) + r_y(u_1 + 2, u_2) + r_y(u_1, u_2 - 2) + r_y(u_1, u_2 + 2)\big\}/F_C.$$

Thus, for example,

$$r_x(1,0) = \beta_{20}\frac{r_y(1,0) + r_y(1,-2) + r_y(1,2) + r_y(3,0)}{F_C + 2\,\beta_{20}\,[r_y(2,0) + r_y(0,2)]}.$$

If y is a CS-CAR(1), then $F_C = 1 - 4\,\alpha_{10} r_y(1,0)$, and

$$\frac{r_x(2,0)}{r_x(1,0)} = \frac{1 + 2r_y(2,2) + r_y(4,0)}{r_y(1,0) + 2r_y(2,1) + r_y(3,0)} = \frac{\alpha_{10}\big\{1 + 2r_y(2,2) + r_y(4,0)\big\}}{r_y(2,0)},$$

showing that it is possible for a CS-RSD$(1,3)$ to have $r_x(2,0) > r_x(1,0) > 0$ - see Example 5.2.

Example 3.4.
Suppose $d = 2$, and x is a RS-RSD(2,2) with $\alpha_{11} = -\alpha_{10}\alpha_{01}$, so that $A(z_1, z_2)$ is separable, $A(z_1, z_2) = A_1(z_1)A_2(z_2)$, with $A_1(z_1)$ proportional to $(1 - \alpha_{10}z_1)(1 - \alpha_{10}z_1^{-1})$, and $A_2(z_2)$ proportional to $(1 - \alpha_{01}z_2)(1 - \alpha_{01}z_2^{-1})$. Thus y is an AR(1)×AR(1) with $r_y(u_1, u_2) = \alpha_{10}^{|u_1|}\alpha_{01}^{|u_2|}$. If $\mathcal{K} = 1 + 2\alpha_{10}\beta_{10} + 2\alpha_{01}\beta_{01} + 4\alpha_{10}\alpha_{01}\beta_{11}$, then for $u_1 > 0, u_2 > 0$,

$$\begin{aligned}
\mathcal{K}\, r_x(u_1, 0) &= \{\alpha_{10}(1 + 2\alpha_{01}\beta_{01}) + (\beta_{10} + 2\alpha_{01}\beta_{11})(1 + \alpha_{10}^2)\}\alpha_{10}^{(u_1 - 1)}, \\
\mathcal{K}\, r_x(u_1, u_2) &= \{\alpha_{10}\alpha_{01} + \alpha_{01}\beta_{10}(1 + \alpha_{10}^2) + \alpha_{10}\beta_{01}(1 + \alpha_{01}^2) + \beta_{11}(1 + \alpha_{10}^2)(1 + \alpha_{01}^2)\}\alpha_{10}^{(u_1 - 1)}\alpha_{01}^{(u_2 - 1)}.
\end{aligned}$$

If $B(z_1, z_2)$ is also separable (i.e. $\beta_{11} = \beta_{10}\beta_{01}$), so that x is an ARMA(1,1)×ARMA(1,1), then $\mathcal{K} = (1 + 2\alpha_{10}\beta_{10})(1 + 2\alpha_{01}\beta_{01})$, and for $u_1 > 0, u_2 > 0$, we have

$$\begin{aligned}
\mathcal{K}\, r_x(u_1, 0) &= \{\alpha_{10} + \beta_{10}(1 + \alpha_{10}^2)\}(1 + 2\alpha_{01}\beta_{01})\alpha_{10}^{(u_1 - 1)}, \\
\mathcal{K} r_x(u_1, u_2) &= \{\alpha_{10} + \beta_{10}(1 + \alpha_{10}^2)\}\{\alpha_{01} + \beta_{01}(1 + \alpha_{01}^2)\}\alpha_{10}^{(u_1 - 1)}\alpha_{01}^{(u_2 - 1)}.
\end{aligned}$$

4 Operations on CAR models

We show here that special cases of RSD models arise from various operations on CAR models. The results are simple extensions of those for $d = 1$ which show that the operations on AR models lead to ARMAs. However, even for a CAR(1) with $d = 2$ the manipulations are more complicated - see Examples 4.3 and 4.4. We assume in this section that y is a CAR(p) with acgf $\sigma_\eta^2/A(\boldsymbol{z})$, and use x for the resulting model.

4.1 Addition of CARs

If two uncorrelated CARs of orders p_1 and p_2 are added, the result is in general (if there are no common factors) a RSD(p, q), with $q = max(p_1, p_2)$ and $p \geq q$. This includes the case of adding coloured (i.e. correlated) noise to a CAR, where the noise is modelled by a CAR.

Example 4.1.
A well-known example is the noisy CAR $y + \epsilon$ of Example 3.2. Thus, setting $\nu = \sigma_\eta^2/\sigma_\epsilon^2$, the acgf of the noisy CAR is

$$\Gamma_x(\boldsymbol{z}) = \sigma_\epsilon^2(1+\nu)\frac{B(\boldsymbol{z})}{A(\boldsymbol{z})},$$

where $B(\boldsymbol{z}) = \{A(\boldsymbol{z})+\nu\}/(1+\nu)$, so that $\beta_{\boldsymbol{j}}/\alpha_{\boldsymbol{j}} = -1/(1+\nu)$ for $\boldsymbol{j} \neq \boldsymbol{0}$. Since for $\boldsymbol{j} \neq \boldsymbol{0}$, $\beta_{\boldsymbol{j}}/\alpha_{\boldsymbol{j}}$ is constant and $|\beta_{\boldsymbol{j}}| < |\alpha_{\boldsymbol{j}}|$, the noisy CAR is a specific case of the RSD(p, p). As in Example 3.2, taking $\nu < 0$ (i.e. setting $\sigma_\epsilon^2 < 0$), gives $|r_x(\boldsymbol{u})| > |r_y(\boldsymbol{u})|$ for $\boldsymbol{u} \neq \boldsymbol{0}$.

4.2 Smoothing a CAR

Suppose x is formed by smoothing y, i.e. $x(\boldsymbol{t}) = \sum_{\boldsymbol{j}} h_{\boldsymbol{j}} y(\boldsymbol{t} - \boldsymbol{j})$, for $H(z)$ a finite function of z. Then, equation (3) shows that x is a RSD.

Example 4.2.
A special case in two dimensions ($d = 2$) is summing over adjacent sites for $(k_1 \times k_2)$ blocks, which gives

$$f_x(\lambda_1, \lambda_2) = \frac{[1-\cos(\lambda_1 k_1)]\,[1-\cos(\lambda_2 k_2)]}{[1-\cos(\lambda_1)]\,[1-\cos(\lambda_2)]} f_y(\lambda_1, \lambda_2).$$

4.3 Sampling a CAR

Suppose x is formed by sampling every $\boldsymbol{k}^{th}$ site of a CAR, y, where $\boldsymbol{k} = (k_1, \ldots, k_d)'$. Then

$$f_x(\boldsymbol{\lambda}) = \frac{1}{\prod_{j=1}^d k_j} \sum_{\boldsymbol{j}=\boldsymbol{0}}^{\boldsymbol{k-1}} f_y\left(\frac{\boldsymbol{\lambda} + 2\pi\boldsymbol{j}}{\boldsymbol{k}}\right),$$

where $\boldsymbol{u}/\boldsymbol{v}$ denotes $(u_1/v_1, ...)'$, and x is a RSD.

Example 4.3.
Suppose $d = 2$, $k_1 = k_2 = 2$, and y is a CAR(1). Then, setting $a_j = 1 + \cos(\lambda_j)$, and $b_j = 3 + 4\cos(\lambda_j) + \cos(2\lambda_j)$, for $j = 1, 2$, gives

$$f_x(\lambda_1, \lambda_2) = \sigma_\eta^2 \left[\frac{1 - 2\alpha_{10}^2 a_1 - 2\alpha_{01}^2 a_2}{1 - 4\alpha_{10}^2 a_1 - 4\alpha_{01}^2 a_2 - 8\alpha_{10}^2\alpha_{01}^2 a_1 a_2 + 2\alpha_{10}^4 b_1 + 2\alpha_{01}^4 b_2}\right],$$

which is the sdf of a RSD$(3, 1)$.

4.4 Coarser resolution of a CAR

Suppose a CAR is summed within regular exhaustive disjoint blocks. Then the resulting coarser process is a RSD. The result for the sdf follows by firstly using the result from section 4.2 to get the summed process on overlapping blocks, and then sampling this process using the result in section 4.3. Such RSDs arise in multiresolution image processing whenever a CAR is assumed at the finest level - see, e.g. Lakshmanan and Derin (1993).

Example 4.4.
Suppose $d = 2$, with aggregation of a CAR(1) over (2×2) blocks. Then

$$f_x(\lambda_1, \lambda_2) =$$

$$4\sigma_\eta^2 \left[\frac{1 + (1 - 2\alpha_{10})\alpha_{10}a_1 + (1 - 2\alpha_{01})\alpha_{01}a_2 + 2(1 + \alpha_{10}^2\alpha_{01} + \alpha_{10}\alpha_{01}^2)a_1a_2 - \alpha_{10}^3 b_1 - \alpha_{01}^3 b_2}{1 - 4\alpha_{10}^2 a_1 - 4\alpha_{01}^2 a_2 - 8\alpha_{10}^2\alpha_{01}^2 a_1a_2 + 2\alpha_{10}^4 b_1 + 2\alpha_{01}^4 b_2} \right],$$

which is the sdf of a RSD(3,3).

5 Comparison of RSD and CAR correlations

The representation (7) suggests that it may be possible to approximate a RSD model by a finite order CAR of sufficiently large order. However, due to computational difficulties, CARs of order larger than 5 are not often fitted to lattice data. In Examples 5.1 and 5.2 below, we examine, for $d = 2$, the extent to which CS-CARs of order up to 5 may approximate the correlations of a CS-RSD model. Two different methods are used for choosing the parameters of the approximating CAR model.
If Gaussian maximum likelihood estimation is used for a CAR, the estimates are such that within S_p the estimated correlations from the fitted CAR exactly match the sample correlations (or appropriate averages of them) - see Cressie (1993, section 7.2.2). We therefore use the same procedure, called the ML method, and match the CAR correlations, $r_y(\boldsymbol{u})$, with the RSD ones, $r_x(\boldsymbol{u})$, within S_p. This method should give an indication of the differences that might occur in practice using maximum likelihood estimation. The second method, called the RD method, minimises a weighted sum of squared differences between the correlations at each lag, $\{r_x(\boldsymbol{u}) - r_y(\boldsymbol{u})\}^2$, using the reciprocal lag-distance weights $1/\|\boldsymbol{u}\|$. Rue and Held (2005, section 5.1.2) use this method, with slightly different weights, for approximating geostatistical correlation functions by CARs.
However, these two methods focus solely on matching the correlations of a CAR with those of an RSD, and do not necessarily ensure a good fit to the inverse correlations. We therefore also compare how close the $\tilde{r}_y(\boldsymbol{u})$ are to $\tilde{r}_x(\boldsymbol{u})$, and compare the scaled interpolation error variances F.

Example 5.1.
Consider the CS-RSD(1,1) with $\alpha_{10} = \beta_{10} = 0.248$. This is a special case of Example 4.1 with ν set to -2, and of Example 3.2 with ψ set to $2\alpha_{10}r_y(1,0) - 0.5 \approx -0.226$. The low-lag correlations for this RSD, and for the CS-CAR(1) with the same α_{10}, are given in Table 1. At each lag the ratio is $2/(1+4\alpha_{10}r_y(1,0)) \approx 1.293$. Since $\beta_{10} = \alpha_{10}$, the inverse correlations of this RSD satisfy $\tilde{r}_x(\boldsymbol{u}) = (-1)^{(u_1+u_2)} \times r_x(\boldsymbol{u})$.
The low-lag correlations for the fits are given in Table 1. For the ML fits and $p \leq 3$, the CS-CAR(p) correlations outside S_p^+ do not match the RSD correlations well as the CAR(p)

Models	Lag (1, 0)	(2, 0)	(3, 0)	(1,1)	(2, 1)	(3,1)	(2,2)	(3,2)	(3,3)	F
CS-CAR(1)	0.551	0.358	0.254	0.432	0.320	0.238	0.261	0.207	0.173	-
CS-RSD(1,1)	0.713	0.463	0.328	0.559	0.414	0.308	0.338	0.268	0.223	0.086
CS-CAR(1) ML	0.713	0.583	0.507	0.634	0.557	0.496	0.514	0.471	0.442	0.287
CS-CAR(2) ML	0.713	0.525	0.398	0.559	0.441	0.352	0.368	0.305	0.261	0.190
CS-CAR(3) ML	0.713	0.463	0.295	0.559	0.386	0.255	0.281	0.194	0.138	0.184
CS-CAR(4) ML	0.713	0.463	0.322	0.559	0.414	0.314	0.359	0.300	0.266	0.151
CS-CAR(5) ML	0.713	0.463	0.311	0.559	0.414	0.304	0.338	0.263	0.213	0.139
CS-CAR(1) RD	0.584	0.402	0.300	0.473	0.365	0.285	0.308	0.254	0.218	0.290
CS-CAR(2) RD	0.694	0.495	0.362	0.528	0.404	0.311	0.326	0.262	0.217	0.187
CS-CAR(3) RD	0.704	0.501	0.365	0.542	0.413	0.316	0.331	0.264	0.218	0.185
CS-CAR(4) RD	0.708	0.474	0.331	0.560	0.415	0.309	0.341	0.271	0.228	0.177
CS-CAR(5) RD	0.709	0.471	0.327	0.556	0.413	0.308	0.340	0.270	0.226	0.160

Table 1: Correlations (columns 1 to 9) and F (column 10) for Example 5.1. First row: the CS-CAR(1) with $\alpha_{10} = 0.248$. Second row: the CS-RSD(1,1) with $\alpha_{10} = \beta_{10} = 0.248$. Rows 3 to 12: the CS-CAR(p), $p = 1 \ldots 5$, fitted by ML and RD methods.

correlations are too high for $p = 1, 2$, and too low for $p = 3$. However, the CS-CAR(4) does give a reasonable approximation. As expected, the RD fits of the CS-CAR(p) are better for low p than the ML fits, with the CAR(3) giving a reasonable approximation.
However, the inverse correlations, $\tilde{r}_y(\boldsymbol{u})$, of these CARs can differ substantially from $\tilde{r}_x(\boldsymbol{u})$. For example, the CARs all have $\tilde{r}_y(1, 0) > -0.55$, but $\tilde{r}_x(1, 0) = -0.713$.
The values of F for the CARs are all considerably higher than that of the CS-RSD(1,1). Although the RD method gives better-looking fits for the low-lag correlations outside S_p^+, their values of F are similar to, and mainly slightly worse than, the corresponding ML fits.

Example 5.2.
Consider a case in which $r_x(1, 0)$ is relatively low, as can occur in practice in some agricultural applications. Table 2 shows the low-lag correlations of the 3-parameter CS-RSD(1,3) with $\alpha_{10} = 0.248$, $\beta_{10} = -0.248$, $\beta_{11} = 0$ and $\beta_{20} = 0.2$, and those for the fitted CS-CARs.
Some aspects of the CAR fits are similar to those in Example 5.1. The ML method gave correlations which are too low outside S_p^+ for $p = 1, 2$, and too high for $p \geq 3$. The fit for $p = 4$ is very similar to that for $p = 3$.
The RD fits are less good - only the one for $p = 5$ has $r_y(1, 0) < r_y(2, 0)$. The inverse correlations $\tilde{r}_y(\boldsymbol{u})$ are moderately different from $\tilde{r}_x(\boldsymbol{u})$ for both fits with $p < 5$, but are relatively close for both fits with $p = 5$ (see Table 3). These fits with $p = 5$ also have reasonable values of F (but have 2 more parameters than the CS-RSD). Note that the CS-RSD has $\tilde{r}_x(1, 0) > 0$, but the two CAR(5) fits have $\tilde{r}_y(1, 0) < 0$.

Models	Lag (1, 0)	(2, 0)	(3, 0)	(1,1)	(2, 1)	(3,1)	(2,2)	(3,2)	(3,3)	F
CS-RSD(1,3)	0.391	0.462	0.300	0.363	0.340	0.265	0.282	0.228	0.190	0.407
CS-CAR(1) ML	0.391	0.176	0.086	0.247	0.137	0.074	0.090	0.054	0.036	0.635
CS-CAR(2) ML	0.391	0.254	0.170	0.363	0.232	0.162	0.189	0.139	0.113	0.632
CS-CAR(3) ML	0.391	0.462	0.324	0.363	0.341	0.313	0.361	0.303	0.287	0.561
CS-CAR(4) ML	0.391	0.462	0.325	0.363	0.340	0.314	0.361	0.303	0.287	0.561
CS-CAR(5) ML	0.391	0.462	0.297	0.363	0.340	0.286	0.282	0.265	0.237	0.497
CS-CAR(1) RD	0.558	0.367	0.263	0.441	0.329	0.248	0.271	0.217	0.182	0.612
CS-CAR(2) RD	0.446	0.328	0.244	0.432	0.305	0.236	0.264	0.212	0.183	0.628
CS-CAR(3) RD	0.470	0.342	0.249	0.371	0.295	0.232	0.250	0.206	0.178	0.600
CS-CAR(4) RD	0.462	0.417	0.268	0.393	0.274	0.246	0.267	0.197	0.181	0.545
CS-CAR(5) RD	0.414	0.447	0.276	0.379	0.316	0.256	0.232	0.214	0.177	0.482

Table 2: Correlations (columns 1 to 9) and F (column 10) for Example 5.2. First row: the CS-RSD(1,3) with $\alpha_{10} = 0.248$, $\beta_{10} = -0.248$, $\beta_{11} = 0$ and $\beta_{20} = 0.2$. Rows 2 to 11: the CS-CAR(p), $p = 1 \dots 5$, fitted by ML and RD methods.

Models	Lag (1, 0)	(1,1)	(2, 0)	(2, 1)	(2, 2)
CS-RSD(1,3)	0.049	0.002	-0.359	-0.076	0.184
CS-CAR(5) ML	-0.037	-0.021	-0.264	-0.017	0.108
CS-CAR(5) RD	-0.053	-0.033	-0.277	-0.009	0.132

Table 3: Inverse correlations for Example 5.2. First row: the CS-RSD(1,3) with $\alpha_{10} = 0.248$, $\beta_{10} = -0.248$, $\beta_{11} = 0$ and $\beta_{20} = 0.2$. Rows 2 to 3: the CS-CAR(5) fitted by ML and RD methods.

6 Model Fitting, Simulation and Model Selection

In this section we show that some standard methods for model fitting and identification can be extended to RSDs.

6.1 Model Fitting and Simulation

Suppose the data are observed on an $(n_1 \times \cdots \times n_d)$ lattice $\mathcal{L}$ with $N = \prod_{j=1}^{d} n_j$ sites. Assume $\boldsymbol{x}$ is the N-vector of observations in, say, lexicographic order, and that $\boldsymbol{\theta} = (\boldsymbol{\alpha}' \; \boldsymbol{\beta}')'$, with $\boldsymbol{x} \sim N(\mathrm{G}\boldsymbol{\gamma}, \mathrm{R}_x\sigma^2)$, where G is an $(N \times g)$ matrix and $\boldsymbol{\gamma}$ is a g-vector of parameters. Thus G$\boldsymbol{\gamma}$ represents fixed mean effects, such as trend terms or the effects of different characteristics. There are two common possibilities for the dispersion matrix $\mathrm{R}_x\sigma^2$: either R_x is the correlation matrix and $\sigma^2 = \sigma_x^2$, or R_x^{-1} is the matrix of inverse correlations and $\sigma^2 = \sigma_\eta^2$ (the interpolation variance $\mathrm{Var}[x(\boldsymbol{t})|\cdot]$). Then the deviance, minus twice the log-likelihood, is

$$\mathrm{D}(\boldsymbol{\gamma}, \boldsymbol{\theta}, \sigma^2) = N\log(2\pi\sigma^2) + \log|\mathrm{R}_x| + (\boldsymbol{x} - \mathrm{G}\boldsymbol{\gamma})'\mathrm{R}_x^{-1}(\boldsymbol{x} - \mathrm{G}\boldsymbol{\gamma})/\sigma^2.$$

Assume that R_x or R_x^{-1}, sometimes known as the potential matrix, can be specified and $\mathrm{G}'\mathrm{R}_x^{-1}\mathrm{G}$ is positive definite. Then, in theory, the maximum likelihood fit of a model can be found by minimising the deviance over the valid parameter space. (In practice, optimization can be over R_x positive definite, which for finite N may allow parameter values slightly outside the stationarity limits.) There are however several problems associated with this. If N is very large, deviance evaluation may be time-consuming. If there are many parameters, there may be difficulties in reaching the minimum and rounding errors may make convergence difficult with some routines (or local optima may result from different starting values). Apart from these problems, there are two major difficulties with almost all stationary CAR models. Firstly, the estimates can be extremely close to the stationarity boundary (see section 2.3). Secondly, it is usually not possible to specify R_x explicitly, and on a planar lattice it is usually not possible to specify all of R_x^{-1} explicitly - although the latter is known for all interior sites (those for which $E[x(\boldsymbol{t})|\cdot]$ involves only sites on the lattice $\mathcal{L}$), see section 2.3. The same applies for a RSD with $p > 0$.

Various possibilities have been suggested to overcome this problem with CARs and to complete R_x^{-1}. These essentially specify the conditional mean and variance at border sites (those for which $E[x(\boldsymbol{t})|\cdot]$ involves sites outside $\mathcal{L}$) in terms of sites on the lattice $\mathcal{L}$ in some simple convenient way. Except for the periodic boundary conditions associated with a torus lattice (see below), the result is non-stationary - variances are no longer constant, and correlations at a given lag depend on the sites involved. One option is to set values of $\boldsymbol{x} - \mathrm{G}\boldsymbol{\gamma}$ outside the lattice $\mathcal{L}$ to 0 (sometimes called the Dirichlet conditions). Moura and Balram (1992) discuss two Neumann conditions which use reflected values. Two other possibilities, scaling and including extra sites, are discussed in Besag and Kooperberg (1985).

With Dirichlet or Neumann conditions, there is an exact eigenvalue-eigenvector decomposition of R_x^{-1} for some low-order CARs. In general, Moura and Balram (1992) show that since R_x^{-1} is sparse for CARs, the deviance can efficiently be calculated using the Cholesky decomposition. Boundary conditions are not needed for DCs since R_x is directly specified, and the Cholesky decomposition of Moura and Balram (1992) can be used. However, this method does not seem directly suitable for RSDs.

A possibility for RSDs is to use that $A(\boldsymbol{T})x(\boldsymbol{t})$ is a DC defined by $A(\boldsymbol{z})B(\boldsymbol{z})$. Unless $\min(n_j)$ is small, a good initial estimate can be obtained from the RSD fit on the torus lattice (see below).

Another method frequently used for CARs is (essentially) to map the finite planar lattice $\mathcal{L}$ on to a torus (joining row and column ends). Then R_x^{-1} is highly structured - defined by its first row, whose elements are σ^2/σ_η^2 times the inverse correlations. The standardised eigenvectors of R_x can be taken as $\boldsymbol{w}_{k_1} \otimes \boldsymbol{w}_{k_2} \ldots \otimes \boldsymbol{w}_{k_d}$ for $k_j = 1, ..., n_j$, where $\boldsymbol{w}_{k_j}$ has v^{th} element $n_j^{-1/2} e^{i2\pi k_j v/n_j}$. Its eigenvalues are $\varphi(\boldsymbol{u}) = f(2\pi \boldsymbol{u}/\boldsymbol{n})/\sigma^2$ for $\boldsymbol{u} \in \mathcal{L}$, with $\boldsymbol{n} = (n_1, \cdots, n_d)'$, and $|\mathrm{R}_x| = \prod_{\boldsymbol{u}\in\mathcal{L}} \varphi(\boldsymbol{u})$.
Thus the deviance calculation is very fast and accurate, and the method can also be used for DCs and RSDs. Given R_x^{-1}, R_x can, if required, be quickly found using the fast Fourier transform. This method can give good approximations to the stationary planar correlations.
Provided $\min(n_j)$ is not small, and the estimates are not too close to the parameter boundary, the estimates are very similar whatever boundary conditions are used. Thus, whatever method is most convenient can usually be used for estimation.
The torus assumption gives a stationarity process, but the correlations depend on the size of the lattice, and are unrealistic for sites a long way apart on the planar lattice but close on the torus. Similarly, predicted values for boundary sites using the torus assumption may be unrealistic as they will involve values at distant sites. Provided N is not too large, this problem can largely be overcome by embedding the $(n_1 \times \cdots \times n_d)$ planar lattice $\mathcal{L}$ in a bigger torus lattice, and then using the resulting R_x matrix for the lattice $\mathcal{L}$.
It is simple to simulate a representation if $\boldsymbol{x}$ is Gaussian and R_x or R_x^{-1} can be specified. If $\boldsymbol{\epsilon}$ is a simulation from a $N(0, \sigma^2 \mathrm{I}_N)$, then, using any square root of R_x, $\boldsymbol{x} = \mathrm{G}\boldsymbol{\gamma} + \mathrm{R}_x^{1/2}\boldsymbol{\epsilon}$. In particular, a torus simulation is easy using the known eigenvalues and eigenvectors of R_x.

6.2 Model Selection

Selecting an appropriate RSD(p, q) fit to a data set is a difficult problem, requiring further research. There are some similarities to the problems in choosing an ARMA(p, q) for a time series. One possible approach is to examine the estimated correlations and inverse correlations. In theory, as noted by Garber in 1981 (see Guyon, 1995, Theorem 1.4.1), it is possible to identify the $A(\boldsymbol{z})$ component of the RSD by using the recurrence (9) on the sample correlations for $\boldsymbol{u} \notin S_q^+$, and then using that $A(\boldsymbol{T})x(\boldsymbol{t})$ has acgf proportional to $A(\boldsymbol{z})B(\boldsymbol{z})$ to identify the $B(\boldsymbol{z})$ component. Alternatively, the recurrence (10) on the estimated inverse correlations for $\boldsymbol{u} \notin S_q^+$, could be used for $B(\boldsymbol{z})$. In either case, sampling variation would make implementation imprecise, but it may be possible to select a small number of models for further consideration.
For model checking, the estimated interpolation errors $\hat{\boldsymbol{\eta}}$ can be found using $\hat{\boldsymbol{\eta}} = (\hat{\sigma}_\eta^2/\hat{\sigma}^2)\hat{\mathrm{R}}_x^{-1}(\boldsymbol{x} - \mathrm{G}\hat{\boldsymbol{\gamma}})$, and the predicted values are then $\hat{\boldsymbol{x}} = \boldsymbol{x} - \hat{\boldsymbol{\eta}}$. Other residuals, such as the approximately uncorrelated $\hat{\mathrm{R}}_x^{1/2}\hat{\boldsymbol{\eta}}$, can then be derived.
Assuming checks on the model assumptions that $\boldsymbol{x} \sim N(\mathrm{G}\boldsymbol{\gamma}, \mathrm{R}_x\sigma^2)$ are satisfactory, and the models are fitted by maximum likelihood as in section 6.1, the significance of parameters in a fitted model can be assessed by comparing estimates with their estimated standard errors (see Cressie, 1993, section 7.3.1). Nested models can be compared using the generalized likelihood ratio test GLRT (using the difference in the deviances having an asymptotic χ^2-distribution). These tests are subject to the usual problems associated with multiple testing, the data not conforming perfectly to the model assumptions, and the use of asymptotic theory on finite data.
If there are several models to compare, including ones which are not nested, standard model-selection criteria can be used, such as AIC $=$ D$(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2) + 2\ \mathcal{P}$ and BIC $=$ D$(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2) + \mathcal{P}\ \log(N)$, for a model with $\mathcal{P}$ parameters. At present, little is known about the behaviour of

these criteria for selecting RSD models, or if some other multiplier of $\mathcal{P}$ would be preferable.

7 Real Data Examples

In this section, we use two real data sets to see if RSD models can be useful, and whether their fits may be preferable to CAR fits. The data, represented in Figure 1, are (128×128) portions (rows 200 to 327 and columns 300 to 427) of two texture images, grass (1.1.01) and wool (1.1.05), available from `http://sipi.usc.edu/database/`. The grass image has pixel values ranging from 6 to 238, with sample mean 131.4, and sample standard deviation 49.01. For the wool image, the values range from 64 to 224, with sample mean 147.2, and sample standard deviation 24.47.

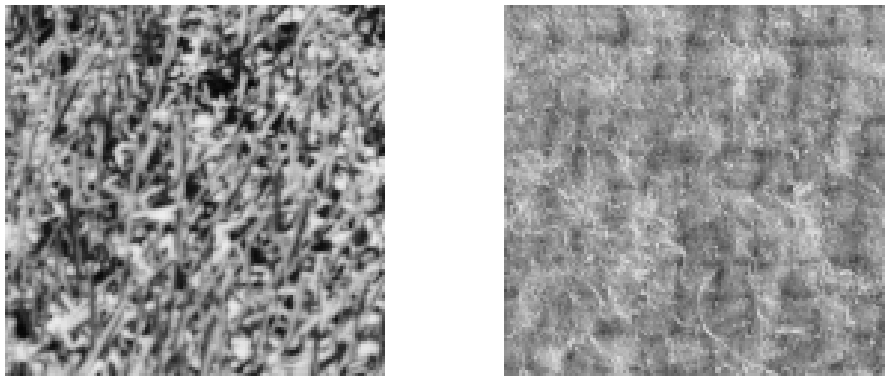

Figure 1: Texture images: *left*) grass (1.1.01); *right*) wool (1.1.05).

A constant mean looks reasonable for both images. The histograms suggest Normality is a plausible working assumption. Since $N = 2^{14}$ is so large, it is reasonable to use a torus lattice for the maximum likelihood fitting (see section 6.1). Predicted values $\hat{\boldsymbol{x}}$ and estimated interpolation values $\hat{\boldsymbol{\eta}}$ were obtained by beginning with the fitted torus values, then setting values of $\boldsymbol{x}$ and $\boldsymbol{\eta}$ outside $\mathcal{L}$ to the sample mean and iteratively getting $\hat{\boldsymbol{x}}$ from equation (8) and $\hat{\boldsymbol{\eta}}$ as $\boldsymbol{x} - \hat{\boldsymbol{x}}$.
The inverse correlations and the scaled interpolation variance F of the data were estimated using equation (2) with the sdf $f_x(\boldsymbol{\lambda})$ estimated by smoothing the periodogram. Using the two-dimensional form of the Daniell window (Priestley, 1981; p. 441) with length (a, a) where a is between 17 and 21, seemed to give an adequate amount of smoothing. The estimated inverse correlations do not

change appreciably over the different a, but the estimated value of the scaled interpolation variance F does increase with a.

Yuan and Subba Rao (1993) gave asymptotic standard errors for the estimated inverse correlations, but did not show asymptotic Normality. To obtain an indication of the distribution when $\tilde{r}_x(\boldsymbol{u}) = 0$, we examined estimated inverse correlations using $a = 19$ from simulated White Noise, and also looked at the simulated distributions outside S_p^+ of the two chosen fitted CARs. These simulations suggest that the estimated inverse correlations are approximately Normally distributed, with standard error around 0.007, so the upper 2.5% point is around 0.014.

The low-lag sample correlations (using a divisor of N) are shown in Tables 4 (grass) and 5 (wool). These are high for neighbouring sites, but they then drop away quite quickly. They are slightly larger between rows than between columns. The grass correlations are higher than the wool ones for adjacent sites, but they then drop away more quickly. The sample correlations do not suggest any symmetries.

u_2									
4	0.091	0.098	0.107	0.132	0.169	0.175	0.146	0.112	0.083
	0.029	0.050	0.063	0.054	0.035	0.047	0.067	0.064	0.038
	0.031	0.051	0.085	0.142	0.211	0.240	0.221	0.127	0.095
3	0.100	0.114	0.146	0.209	0.278	0.278	0.211	0.139	0.090
	0.042	0.083	0.127	0.168	0.207	0.219	0.181	0.105	0.030
	0.045	0.076	0.130	0.224	0.246	0.279	0.239	0.177	0.155
2	0.108	0.138	0.212	0.347	0.482	0.448	0.300	0.177	0.104
	0.046	0.112	0.212	0.347	0.482	0.448	0.300	0.133	0.014
	0.066	0.115	0.200	0.356	0.515	0.461	0.338	0.220	0.133
1	0.114	0.175	0.312	0.573	0.801	0.663	0.386	0.211	0.119
	0.035	0.133	0.312	0.573	0.801	0.663	0.386	0.144	0.005
	0.097	0.173	0.311	0.572	0.808	0.686	0.435	0.235	0.129
0					1.000	0.747	0.399	0.210	0.119
					1.000	0.747	0.399	0.146	0.014
					1.000	0.753	0.421	0.204	0.103
	-4	-3	-2	-1	0	1	2	3	4
					u_1				

Table 4: Grass texture image. Estimated correlations at lags (u_1, u_2), $u_1 = -4, \ldots, 4$; $u_2 = 0, \ldots, 4$. First row: sample correlations. Second row: fitted $\mathrm{CAR}(5)$. Third row: fitted $\mathrm{RSD}(2, 2)$.

The low-lag estimated inverse correlations, using $a = 19$, are given in Tables 6 (grass) and 7 (wool). Taking an informal account of multiple testing, they suggest that for the wool image the only non-zero values may be for $u_1 = 0$, $1 \leq u_2 \leq 3$, plus $u_1 = 1$, $0 \leq u_2 \leq 2$. The grass image has more large values, with most in Table 6 possibly being significantly different from 0. The estimated values of F are 0.0360 ($a = 17$), 0.0374 ($a = 19$), 0.0389 ($a = 21$) for the grass image, and 0.2323 ($a = 17$), 0.2348 ($a = 19$), 0.2374 ($a = 21$) for the wool image.

$\mathrm{CAR}(p)$ models for $1 \leq p \leq 5$, and $\mathrm{RSD}(p, q)$ models for $1 \leq p, q \leq 3$, were fitted to the data.

u_2									
4	0.112	0.172	0.239	0.297	0.318	0.293	0.242	0.184	0.124
3	0.152	0.223	0.304	0.372	0.398	0.358	0.296	0.225	0.156
2	0.190	0.274	0.381	0.480	0.516	0.441	0.355	0.274	0.190
1	0.221	0.321	0.452	0.615	0.741	0.556	0.420	0.314	0.222
0					1.000	0.672	0.474	0.339	0.231
	-4	-3	-2	-1	0	1	2	3	4
					u_1				

Table 5: Wool texture image. Sample correlations at lags $(u_1, u_2),\ u_1 = -4, \ldots, 4,\ u_2 = 0, \ldots, 4$.

u_2									
4	-0.013	0.011	-0.003	-0.015	0.019	0.010	-0.007	-0.003	0.007
	0	0	0	0	0	0	0	0	0
	0.001	-0.004	0.009	-0.021	0.020	0.023	-0.028	0.013	-0.001
3	0.016	-0.014	0.002	0.028	-0.034	-0.034	0.043	-0.023	0.007
	0	0	0	0	0	0	0	0	0
	-0.005	0.011	-0.024	0.055	-0.060	-0.025	0.053	-0.036	0.017
2	-0.003	-0.011	0.048	-0.119	0.152	-0.009	-0.049	0.044	-0.026
	0	0	0.034	-0.104	0.130	-0.052	0.007	0	0
	0.016	-0.033	0.070	-0.147	0.179	-0.020	-0.069	0.069	-0.049
1	-0.040	0.092	-0.202	0.395	-0.531	0.274	-0.087	0.016	0.004
	0	0	-0.118	0.387	-0.555	0.291	-0.068	0	0
	-0.056	0.108	-0.209	0.410	-0.542	0.267	-0.047	-0.025	0.041
0					1.000	-0.696	0.356	-0.166	0.075
					1.000	-0.622	0.170	0	0
					1.000	-0.692	0.342	-0.169	0.084
	-4	-3	-2	-1	0	1	2	3	4
					u_1				

Table 6: Grass texture image. Estimated inverse correlations at lags $(u_1, u_2),\ u_1 = -4, \ldots, 4;\ u_2 = 0, \ldots, 4$. First row: sample inverse correlations. Second row: fitted $\mathrm{CAR}(5)$. Third row: fitted $\mathrm{RSD}(2, 2)$.

u_2									
4	-0.001	0.001	-0.002	-0.003	0.015	-0.012	0.012	-0.006	0.005
3	0.003	-0.004	0.006	0.008	-0.012	0.010	-0.015	0.011	-0.006
2	0.000	0.002	-0.002	0.018	-0.011	0.003	0.004	-0.004	0.005
1	-0.004	0.002	0.014	-0.013	-0.228	0.092	-0.028	0.008	-0.002
0					1.000	-0.446	0.116	-0.040	0.013
	-4	-3	-2	-1	0	1	2	3	4
					u_1				

Table 7: Wool texture image. Estimated inverse correlations at lags (u_1, u_2), $u_1 = -4, \ldots, 4$, $u_2 = 0, \ldots, 4$.

We compare fits here using the GLRT when appropriate, and the AIC and BIC values (see section 6.2). We also compare fitted correlations and inverse correlations with those of the data. The BIC criterion has a much larger penalty for $\mathcal{P}$ parameters than the AIC, since $\log(128^2) \approx 9.70$. Note that $\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2 + 2$ ($\mathcal{P}_1$ from S_p, $\mathcal{P}_2$ from S_q, plus μ and σ^2). The deviance, AIC and BIC values have been reduced by 141000 (grass image), and by 133000 (wool image).

For the grass image, the valid GLRT comparisons do not suggest any model simplification from the general CAR(5) and RSD$(3, 3)$. The best AIC and BIC values among the CARs are for the CAR(5) with $\mathcal{P} = 12 + 2$ which has AIC $= 623.74$ and BIC $= 731.60$. However, the RSD(1,2) model with $\mathcal{P} = 2 + 4 + 2$ has AIC $= 580.10$ and BIC $= 641.73$, and the best values are for the RSD$(2, 2)$ with $\mathcal{P} = 4 + 4 + 2$ which has AIC $= 161.74$ and BIC $= 238.78$. The parameter estimates for the RSD$(2, 2)$ are $\hat{\alpha}_{10} = 0.301$, $\hat{\alpha}_{01} = 0.374$, $\hat{\alpha}_{11} = -0.134$, $\hat{\alpha}_{1-1} = -0.049$, $\hat{\beta}_{10} = 0.407$, $\hat{\beta}_{01} = 0.448$, $\hat{\beta}_{11} = 0.164$, $\hat{\beta}_{1-1} = 0.227$ (the estimates for the CAR(5) are minus the inverse correlations given in Table 6).

Table 4 shows the correlations of the fitted CAR(5) and RSD$(2, 2)$ - see also Figure 2. Although the correlations of the CAR(5) match those of the data within S_5, outside S_5^+ they do not match the sample correlations well as they drop off rapidly. For example, the lag $(0, 4)$ value of 0.035 is well below 0.169. On the other hand, the fitted RSD$(2, 2)$ correlations are reasonably close over $-4 \le u_1 \le 4$ and $0 \le u_2 \le 4$.

Table 6 shows the low-lag inverse correlations for the fitted CAR(5) and RSD$(2, 2)$ models - see also Figure 2. The values for the RSD$(2, 2)$ are mainly much closer to those in Table 6 than those of the CAR(5), e.g. see $u_1 = 0$. The estimated value of F is 0.0329 for the RSD$(2, 2)$, and 0.0464 for the CAR(5).

The $\hat{\eta}(\boldsymbol{t})$ for the two fits do not show any areas where the models fit badly. A comparison of the $\hat{\eta}(\boldsymbol{t})$ for these two fits shows that the range for the RSD$(2, 2)$ is smaller $(-2.37$ to $2.23)$ than that for the CAR(5) $(-2.72$ to $2.32)$, and that the RSD$(2, 2)$ has fewer large values, with eight $|\hat{\eta}(\boldsymbol{t})| > 2$, whereas the CAR$(5)$ has twelve.

For the wool image, the GLRT shows that amongst the CARs $p = 4$ (D $= 605.36$, $\mathcal{P} = 10 + 2$) is sufficient. The AIC and BIC criteria also choose the general CAR(4) (AIC $= 629.36$, BIC $= 721.81$).

Amongst the RSD(p, q), the possible GLRTs suggest that $p = q = 3$ (D $= 507.46$, $\mathcal{P} = 6+6+2$) is necessary. This also has the best AIC (535.46) and BIC (643.32) values. Its AIC and BIC values are much lower than the best CAR values. However, better AIC and BIC values to those of the CAR(4) are obtained by a RSD with fewer parameters; for example we have AIC $= 609.40$ and BIC $= 671.03$ for the RSD(1,2) with $\mathcal{P} = 2 + 4 + 2$, and AIC $= 603.72$ and BIC $= 680.76$

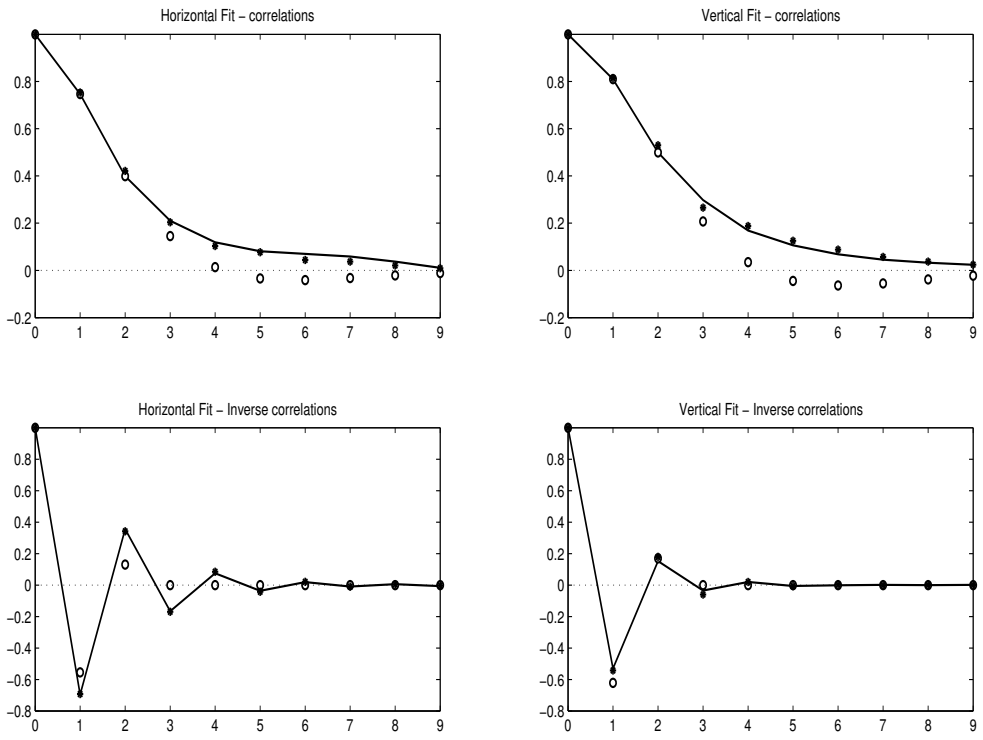

Figure 2: Grass texture image. Sample (solid line) and model-estimated correlations (top) and inverse correlations (bottom) along the horizontal (left) and vertical (right) directions for the $\mathrm{CAR}(5)$ (circles) and $\mathrm{RSD}(2,2)$ (dots)

for the $\mathrm{RSD}(2,2)$ with $\mathcal{P} = 4+4+2$. The fitted $\mathrm{CAR}(4)$ and $\mathrm{RSD}(3,3)$ both have correlations and inverse correlations which match moderately well those of the data. The estimated values of F are 0.2242 for the $\mathrm{RSD}(3,3)$ and 0.2271 for the $\mathrm{CAR}(4)$. The parameter estimates for the $\mathrm{CAR}(4)$ are $\hat{\alpha}_{10} = 0.429$, $\hat{\alpha}_{20} = -0.072$, $\hat{\alpha}_{01} = 0.225$, $\hat{\alpha}_{02} = 0.027$, $\hat{\alpha}_{11} = 0.019$, $\hat{\alpha}_{1-1} = -0.105$, $\hat{\alpha}_{21} = -0.031$, $\hat{\alpha}_{2-1} = -0.014$, $\hat{\alpha}_{12} = -0.013$, $\hat{\alpha}_{1-2} = 0.034$, and for the $\mathrm{RSD}(3,3)$ are $\hat{\alpha}_{10} = 0.391$, $\hat{\alpha}_{20} = -0.036$, $\hat{\alpha}_{01} = 0.438$, $\hat{\alpha}_{02} = -0.044$, $\hat{\alpha}_{11} = -0.127$, $\hat{\alpha}_{1-1} = -0.123$, $\hat{\beta}_{10} = 0.241$, $\hat{\beta}_{20} = 0.002$, $\hat{\beta}_{01} = -0.085$, $\hat{\beta}_{02} = -0.099$, $\hat{\beta}_{11} = 0.053$, $\hat{\beta}_{1-1} = -0.084$.

These two examples show that RSD models with the same number of parameters, or even fewer, can give better and simpler fits than CAR models. For the grass image, the best RSD fit, using the AIC or BIC criterion, also has a better fit of both the correlations and the inverse correlations, and a smaller estimated F.

8 Discussion

RSD models are a natural extension of CAR and DC models, which warrant consideration when fitting spatial models to data. They have a wider range of correlation structures, and for moderate to high correlations they do not usually need the parameters to be so close to the boundary. Thus parameter estimation is computationally easier and more stable. Unless $\mathcal{P}$ is small (up to 6), it can be much faster to fit an RSD with $q > 0$ than a CAR with the same $\mathcal{P}$. For example, fitting the

CAR(5) to the grass data in section 7 took 10 times as long as fitting the RSD$(3, 3)$. RSD models can be useful as simpler fits to data, using fewer parameters, and they can give more accurate predictions. Estimation when there are some missing values, or when the data set contains possible outliers, can be carried out in the usual way.

In this paper we have concentrated on some properties of RSD models. Topics for further work on RSD models include investigating different methods for fitting planar data and obtaining predicted values, seeing if they may be useful as approximations to the correlations or inverse correlations of geostatistical models which have a large or an infinite range, and theoretical investigations of the properties of some approximations to maximum likelihood estimation and methods for model selection and checking model adequacy. Methods for using RSDs in hierarchical modelling also need to be explored. The inverse correlations and the scaled interpolation variance F were estimated in section 7 using an estimate of the sdf. Further research is needed to determine the best estimate of the sdf for these and the index of linear determinism, $1 - F$, and to compare with other possible estimates.

Matlab (The MathWorks, 2010) routines for fitting the model are available from the corresponding author.

Acknowledgement

We are grateful to the reviewers for their useful comments. RJB was supported by a grant from the University of Chieti-Pescara 'G. d'Annunzio' for the project *Statistical Methods for Temporal, Spatial and Spatio-Temporal analysis of environmental phenomena*. RJM is grateful for support from the University of Chieti-Pescara. RJM mentioned the extension to RSDs in a lecture course at what has become known as the 1st Leeds Annual Statistical Research Workshop in 1980, and is grateful to the Department of Statistics, University of Leeds for the invitation.

APPENDIX 1

Stationarity conditions for the CS-CAR(3) when $d = 2$

$f_x(\boldsymbol{\lambda}) > 0$ implies $g(\boldsymbol{\lambda}) < 1/2$, where $g(\boldsymbol{\lambda}) = \alpha_{10}\{\cos(\lambda_1) + \cos(\lambda_2)\} + \alpha_{11}\{\cos(\lambda_1 - \lambda_2) + \cos(\lambda_1 + \lambda_2)\} + \alpha_{20}\{\cos(2\lambda_1) + \cos(2\lambda_2)\}$. We seek maximal values of $g(\boldsymbol{\lambda})$. Now,

$$\frac{\partial g}{\partial \lambda_1} = -\sin(\lambda_1)\{\alpha_{10} + 2\alpha_{11}\cos(\lambda_2) + 4\alpha_{20}\cos(\lambda_1)\}.$$

If $\sin(\lambda_1) = \sin(\lambda_2) = 0$, then $\lambda_j \in \{0, \pi\}$, which implies $\alpha_{11}+\alpha_{20} < 1/4-|\alpha_{10}|$, and $\alpha_{20}-\alpha_{11} < 1/4$. If $\lambda_1 = \lambda_2$, and $\cos(\lambda_1) = -\alpha_{10}/\{2(\alpha_{11} + 2\alpha_{20})\}$, the condition $g(\boldsymbol{\lambda}) < 1/2$ becomes $1 + 4\alpha_{20} + \alpha_{10}^2/(\alpha_{11} + 2\alpha_{20}) > 0$.

APPENDIX 2

Approximate value when $d = 2$ of $r_x(1, 0)$ for a CS-CAR(1) with a large α_{10}

Let $V(\boldsymbol{u})$ denote $R_x(\boldsymbol{u})/\sigma_\eta^2$. Equation (6) with $\boldsymbol{u} = \mathbf{0}$ gives

$$V(0, 0) = 4\alpha_{10}V(1, 0) + 1,$$

so that

$$r_x(1, 0) = \left(1 - \frac{1}{V(0, 0)}\right)/(4\alpha_{10}).$$

From Besag (1981, eq (7)),

$$V(0, 0) = 2K(4\alpha_{10})/\pi,$$

where K is the complete elliptic integral of the first kind. Now, from Abramowitz and Stegun (1965, equation 17.3.26), and correcting for the change in notation - see Besag (1981), we have

$$\lim_{\alpha_{10}\to 1/4}\left[K(4\alpha_{10}) - \frac{1}{2}\log\left(\frac{16}{1 - 16\alpha_{10}^2}\right)\right] = 0.$$

Thus if $\alpha_{10} = 1/4 - \varsigma$ for ς small,

$$r_x(1, 0) \approx \frac{1}{4\alpha_{10}}\left(1 + \frac{\pi}{\log\left(\frac{\varsigma}{2} - \varsigma^2\right)}\right). \tag{11}$$

Since $\varsigma < 0.001$ for $r_x(1, 0) > 0.59$, we can take

$$r_x(1, 0) \approx \frac{1}{4\alpha_{10}}\left(1 - \frac{\pi}{log(2/\varsigma)}\right) \tag{12}$$

or just

$$r_x(1, 0) \approx 1 - \frac{\pi}{log(2/\varsigma)} \tag{13}$$

For example, when $\alpha_{10} = 0.249$, $r_x(1, 0) \approx 0.5898$, and the approximations are (to 4 d.p.) 0.5891 for equation (11), 0.5890 for equation (12) and 0.5867 for equation (13). When $\alpha_{10} = 0.2499$, $r_x(1, 0) \approx 0.6831$, and the approximations are, respectively, 0.6831, 0.6831, and 0.6828.

References

Abramowitz M., Stegun I. A. (Eds.) (1965). *Handbook of mathematical functions*. Dover: New York.

Bartlett, M.S. (1975). *The statistical analysis of spatial pattern*. London: Chapman and Hall.

Besag J. (1981). On a system of two-dimensional recurrence equations. *Journal of the Royal Statistical Society, B*, **43**, 302-309.

Besag J., Kooperberg C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, **82**, 733-746.

Bhansali R.J., Ippoliti L. (2005). Inverse correlations for multiple time series and gaussian random fields and measures of their linear determinism. *Journal of Mathematics and Statistics*, **4**, 287-299.

Bolin D., Lindgren F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *The Annals of Applied Statistics*, **5**, 523-550.

Cressie N. (1993). *Statistics for spatial data*. Revised edition. New York: Wiley.

Guyon X. (1995). *Random fields on a network. Modeling, statistics, and applications*. New York: Springer-Verlag.

Kiiveri H. T., Campbell N. A. (1989). Covariance models for lattice data. *Australian Journal of Statistics*, **31**, 62-77.

Lakshmanan S., Derin H. (1993) Gaussian Markov Random fields at multiple resolutions. In *Markov random fields, theory and applications*, Chellappa, R. and Jain A. Eds., 179-209. Boston: Academic Press.

Lindgren F., Rue H., Lindström J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *Journal of the Royal Statistical Society, B*, **73**, 423-498.

Martin R. J., Walker A.M. (1997). A power-law model and other models for long-range dependence. *Journal of Applied Probability*, **34**, 657-670.

Moura J. M. F., Balram N. (1992). Recursive structure of noncausal Gauss Markov random fields. *IEEE Transactions on Information Theory*, **38**, 334-354.

Priestley M.B. (1981). *Spectral analysis and time series*. London: Academic Press.

Rue H., Held L. (2005). *Gaussian Markov random fields. Theory and applications*. Boca Raton: Chapman & Hall/CRC.

The MathWorks (2010). Matlab, Version 7.10.0.499. Natick, Massachusetts.

Vecchia, A. V. (1985). A general class of models for stationary two-dimensional random processes. *Biometrika*, **72**, 281-291.

Yuan J., Subba Rao T. (1993). Spectral estimation for random fields with applications to Markov modelling and texture classification. In *Markov random fields, theory and applications*, Chellappa, R. and Jain A. Eds., 179-209. Boston: Academic Press.