Conditional Autoregressive Moving Average Models for Lattice Data

L. **Ippoliti**[a] R.J. **Martin**[a] R.J. **Bhansali**[b]

- a) **Department of Economics**
University G. d'Annunzio - Chieti-Pescara

- b) **Division of Statistics and Probability, Department of Mathematical Sciences**
University of Liverpool

Spatial Processes on a d-dimensional regular lattice

Assume that t, u, z and λ are $d-$dimensional vectors and that $\{x(t), t \in \mathbb{Z}^d\}$ is a Gaussian random field on the regular rectangular lattice.

Assume also that the random field is and second-order stationary with constant mean, μ, autocovariance function $R_x(u) = Cov\{x(t), x(t+u)\}$, and autocorrelation function $r_x(u) = R_x(u)/\sigma_x^2$, where $\sigma_x^2 = R_x(0)$.

The autocovariance generating function - acgf - is

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

and, provided $\sum_{u \in \mathbb{Z}^d} |R_x(u)| < \infty$, the spectral density function -sdf- is defined here as $f_x(\lambda) = \Gamma_x(e^{i\lambda})$ where $\lambda \in (-\pi, \pi]^d$. Note that we are using a form of the sdf for which its integral is $(2\pi)^d \sigma_x^2$.

Spatial Processes on a d-dimensional regular lattice

The inverse equation giving $R_x(u)$ is

$$R_x(u) = (2\pi)^{-d} \int \exp(iu'\lambda) f_x(\lambda) d\lambda.$$

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

$$\tilde{R}_x(u) = (2\pi)^{-d} \int \exp(iu'\lambda) \left[f_x(\lambda)\right]^{-1} \, d\lambda \quad \text{and} \quad \tilde{r}_x(u) = \tilde{R}_x(u)/\tilde{R}_x(0).$$

For convenience, we mainly refer here to Guyon (1995), Cressie (1993) and Rue and Held (2005) for known results.

Processes with Rational Spectra

Let

$$A(z) = 1 - \sum_{j \in S_p} \alpha_j z^j \quad \text{and} \quad B(z) = 1 + \sum_{j \in S_q} \beta_j z^j$$

be finite symmetric Laurent series satisfying $A(z) = A(z^{-1})$ and $B(z) = B(z^{-1})$, where $z^j = (z^{j_1} \ldots, z^{j_d})$, and S_m is a finite subset of $\mathbb{Z}^d$ containing neighbours of the origin. Thus $\alpha_j = \alpha_{-j}$ and $\beta_j = \beta_{-j}$ for all j. Thus $t + S_m$ is the set of neighbours of site t.

The order of the neighbourhood set is denoted by m, and 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.

Modelling approaches for lattice data

Geostatistical models are defined for continuous space, and are widely used for analyzing data defined on irregular regions - see, for example, Cressie (1993, sec 2.3.1). **On a regular lattice, they and other DC models usually have the defect that R_x^{-1} usually has no simple form**.

UARMA models can appear arbitrary, as they may depend on the choice of site ordering, but can have some useful properties, especially if the model is separable. **SAR models have some severe logical difficulties** (e.g. the independent errors are correlated with all the observations).

Every UAR, and every SAR, is a special case of a CAR, also known as a **Gauss-Markov random field** - GMRF. **These models have nice properties but also suffer from some disadvantages!**

Why Conditional ARMA processes

The **correlation structure of a CAR model is usually hard to determine** (except numerically), though the inverse correlations are directly specified and give the inverse dispersion matrix, required for Gaussian ML estimation.

CAR does not apply in non-Markovian situations in which the conditional dependence structure is not confined to a finite subset of the lattice but decays exponentially over an infinite set.

No appreciable correlations are generated by the model unless the parameters are very close to the stationary boundary and which can render the parameter estimation difficult. The use of non-stationary intrinsic CAR models has been suggested (Besag and Kooperberg, 1995) for overcoming this problem. However, this proposal may seem contrived and unsatisfactory, especially for data with modest low-lag correlations.

CAR(p) Processes

A compact representation of a CAR(p) is:

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

where $A(z) = 1 - \sum_{j \in S_p} \alpha_j z^j$, satisfies $A(z) = A(z^{-1})$ and $A(e^{i\lambda}) > 0$ for all λ, and

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

denotes the **interpolation error process** with variance $\sigma_\eta^2 = \{\tilde{R}_x(0)\}^{-1}$ and sdf $\sigma_\eta^2 A(e^{i\lambda})$.

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 $r_x(1,0) = 0.75$.

Conditional ARMA processes

In this work, we consider the generalization of CARs and finite DCs to processes with a rational spectral density function -RSDs.

1. RSD models are a natural extension of CAR and FDC models, which warrant consideration when fitting spatial models to data.
2. They have a wider range of correlation structures, and for moderate to high correlations they do not need the parameters to be so close to the boundary.
3. They can be useful as simpler fits to data, using fewer parameters, and they can give more accurate predictions.

RSD(p,q) - CARMA(p,q) - Processes

Assume that $A(z)$ and $B(z)$ have no common factors, and that $A(z) \neq 0$ and $B(z) \neq 0$ for $|z| = 1$.

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

The sdf $f_x(\lambda)$ can be regarded as the ratio of two CAR sdf's, or two DC sdf's. RSD models are mentioned in Guyon (1995, sec 1.4) and little appears to have been done on them.

Clearly CAR and DC are special cases.

RSD(p,q) - CARMA(p,q) - Processes

We can then formally write

$$A(T)x(t) = c_0 B(T)\eta(t),$$

which expresses a finite combination of the x's in terms of a finite combination of the η's, and has some similarities to the one-dimensional ARMA representations, but now the η's are correlated.

We can then refer to this $\mathrm{RSD}(p, q)$ as a **Conditional ARMA model - CARMA(p, q)**.

Note that the correlations for the model with acgf proportional to $B(z)/A(z)$ are the inverse correlations of the model with acgf proportional to $A(z)/B(z)$.

RSD(p,q) - CARMA(p,q) - Processes

As with a CAR, the sdf of a RSD is a real analytic function, and hence its **correlations and inverse correlations decay exponentially**. The **covariances and the inverse covariances satisfy extended Yule-Walker equations**. Multiplying the RSD model equation by $x(t+u)$ or $\eta(t+u)$, respectively, and taking expectations gives

$$A(T)R_x(u) = c_0\sigma_\eta^2\beta_u \quad \forall\ u \in \mathbb{Z}^d;$$

$$\sigma_\eta^2\alpha_u = c_0\sigma_\eta^2 B(T)\tilde{r}_x(u) \quad \forall\ u \in \mathbb{Z}^d.$$

Scaled interpolation variance, F, for a RSD model: if $C(z) = A(z)/B(z)$, then $c_0 = 1/\sum \beta_u \tilde{r}_x(u)$, and

$$\sigma_\eta^2/\sigma_x^2 = F = \{1 - \sum_{u\neq 0} \alpha_u r_x(u)\} \times \{\sum_u \beta_u \tilde{r}_x(u)\}.$$

Operations on CAR processes: Addition of CARs

We note here that special cases of RSD processes arise from various operations on CAR processes. It is also of course true that operations on RSD processes result in RSD models. In the following we give some examples. **We assume in this section that y is a CAR(p) with acgf $\sigma_\eta^2/A(z)$, and use x for the resulting model.**

If two independent CARs of orders p_1 and p_2 are added, the result is in general (if there are no common factors) a $\mathrm{RSD}(p, q)$, with $q = max(p_1, p_2)$ and $p \geq q$.

A well-known example is if a White Noise process, ξ, is added to a CAR y, which gives a special case of a $\mathrm{RSD}(p, p)$. Thus, if a $\mathrm{CAR}(p)$ has conditional variance σ_η^2, and setting $\nu = \sigma_\eta^2/\sigma_\xi^2$, the sdf of the NCAR, is

$$f_x(\lambda) = \sigma_\xi^2(1+\nu)\frac{B(z)}{A(z)}$$

where $\beta_j/\alpha_j = -1/(1+\nu)$. Since β_j/α_j is constant and $|\beta_j| < |\alpha_j|$, the Noisy CAR is only a specific case of the general $\mathrm{CARMA}(p, p)$.

Operations on CAR processes: Smoothing a CAR

Suppose $x(t)$ is formed by smoothing $y(t)$, i.e. $x(t) = \sum_j h(j)y(t-j)$, for h a finite function of z.

Then $f_x(\lambda) = h(z)h(z^{-1})f_y(\lambda) = \sigma_\eta^2 B(z)/A(z)$, where $B(z) = h(z)h(z^{-1})$.

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

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

Operations on CAR processes: Sampling a CAR

Suppose x is formed by sampling every kth site of a CAR, y, where $k = (k_1, \ldots, k_d)$. Then the sdf of x is

$$f_x(\lambda) = \frac{1}{|k|} \sum_{j=0}^{k-1} f_y\left(\frac{\lambda + 2\pi j}{k}\right)$$

where $|k| = \prod_{i=1}^{d} k_i$, and $u/v = (u_1/v_1, \ldots)$. Then x is a RSD.

Suppose $d = 2$, $k_1 = k_2 = 1$ and consider a $\mathrm{CAR}(1)$. Then, setting $a_1 = 1 + \cos(\lambda_1)$, $a_2 = 1 + \cos(\lambda_2)$, $a_3 = 3 + 4\cos(\lambda_2) + \cos(2\lambda_2)$ and $a_4 = 3 + 4\cos(\lambda_1) + \cos(2\lambda_1)$, the sdf for x is

$$f_x(\lambda_1, \lambda_2) = \sigma_\eta^2 \left[\frac{1 - 2\alpha_{10}^2 a_1 - 2\alpha_{01}^2 a_2}{1 - 4\alpha_{10}^2 a_1 - 4\alpha_{01}^2 a_2 - 8\alpha_{10}^2\alpha_{01}^2 a_1 a_2 + 2\alpha_{10}^4 a_3 + 2\alpha_{01}^4 a_4}\right]$$

which is the spectral density function of a $\mathrm{CARMA}(1,3)$.

Operations on CAR processes: Coarser resolution of a CAR

If a CAR is summed within blocks, and the result for exhaustive disjoint blocks is used, then the coarser process is a RSD. The result for f_x follows by firstly using the result from smoothing to get the summed process on overlapping blocks, and then sampling this process using the result above.

For example, if $d = 2$, and aggregating x over (2×2) blocks, setting $a_5 = 1 - 2\alpha_{10}$, $a_6 = 1 - 2\alpha_{01}$ and $a_7 = 1 + \alpha_{10} + \alpha_{01}$, gives

$$f_x(\lambda_1, \lambda_2) = 4\sigma_\eta^2 \left[\frac{1 + \alpha_{10} a_5 a_1 + \alpha_{01} a_6 a_2 + 2\alpha_{10}\alpha_{01} a_7 a_1 a_2 - \alpha_{10}^3 a_3 - \alpha_{01}^3 a_4}{1 - 4\alpha_{10}^2 a_1 - 4\alpha_{01}^2 a_2 - 8\alpha_{10}^2\alpha_{01}^2 a_1 a_2 + 2\alpha_{10}^4 a_3 + 2\alpha_{01}^4 a_4} \right],$$

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

Some Properties of a RSD Model: Relating RSD to CAR correlations

Suppose y is a CAR defined by $A(z)$. Then the acgf for x is

$$\begin{aligned} \mathbf{\Gamma}_x(z) &\propto B(z)\mathbf{\Gamma}_y(z) \\ &\propto B(z)\left[\sum_{u=-\infty}^{u=\infty} R_y(u)z^u\right]. \end{aligned}$$

Thus the correlations of x can be expressed in terms of those of y. Although CAR correlations are usually not readily available, this result shows the effect of $B(z)$ on these correlations. In particular, correlation structures can be obtained which cannot arise from a CAR.

Suppose $d = 2$, and consider an RSD$(p, 1)$. Then, for a constant $\mathcal{K}$

$$R_x(u_1, u_2) = \mathcal{K}\Big(r_y(u_1, u_2) + \beta_{10}[r_y(u_1 - 1, u2) + r_y(u_1 + 1, u2)]$$

$$+\beta_{01}[r_y(u_1, u_2 + 1) + r_y(u_1, u_2 - 1)]\Big).$$

Comparison of RSD and CAR correlations

There are two ways to match CAR correlations with the RSD correlations.

1. Firstly, we note that if Gaussian ML 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 - see Cressie (1993, section 7.2.2).
2. An alternative method (see Rue and Held, 2005, sec. 5.1.2), is to match all the estimated and sample correlations as closely as possible. Here, we minimize the sum of the weighted squared differences between the correlations at each lag,

$$\{r_x(u) - r_y(u)\}^2,$$

using inverse lag-distance weights $1/\|u\|$.

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 $\tilde{r}_y(u)$ are to $\tilde{r}_x(u)$, and compare the scaled interpolation error variances F.

Comparison of CS-RSD and CS-CAR correlations

Example 1: Consider the $\mathrm{CS\text{-}RSD}(1,1)$ with $\alpha_{10} = \beta_{10} = 0.248$. The inverse correlations of this RSD satisfy $\tilde{r}_x(u) = (-1)^{(u_1+u_2)} \times r_x(u)$.

Models	Lag (1, 0)	(2, 0)	(3, 0)	(1,1)	(2, 1)	(3,1)	(2,2)	(3,2)	F
$\mathrm{CAR}(1)$	0.551	0.358	0.254	0.432	0.320	0.238	0.261	0.207	-
$\mathrm{RSD}(1,1)$	0.713	0.463	0.328	0.559	0.414	0.308	0.338	0.268	0.086
$\mathrm{CAR}(1)_1$	0.713	0.583	0.507	0.634	0.557	0.496	0.514	0.471	0.287
$\mathrm{CAR}(2)_1$	0.713	0.525	0.398	0.559	0.441	0.352	0.368	0.305	0.190
$\mathrm{CAR}(3)_1$	0.713	0.463	0.295	0.559	0.386	0.255	0.281	0.194	0.184
$\mathrm{CAR}(4)_1$	0.713	0.463	0.322	0.559	0.414	0.314	0.359	0.300	0.151
$\mathrm{CAR}(5)_1$	0.713	0.463	0.311	0.559	0.414	0.304	0.338	0.263	0.139
$\mathrm{CAR}(1)_2$	0.584	0.402	0.300	0.473	0.365	0.285	0.308	0.254	0.290
$\mathrm{CAR}(2)_2$	0.694	0.495	0.362	0.528	0.404	0.311	0.326	0.262	0.187
$\mathrm{CAR}(3)_2$	0.704	0.501	0.365	0.542	0.413	0.316	0.331	0.264	0.185
$\mathrm{CAR}(4)_2$	0.708	0.474	0.331	0.560	0.415	0.309	0.341	0.271	0.177
$\mathrm{CAR}(5)_2$	0.709	0.471	0.327	0.556	0.413	0.308	0.340	0.270	0.160

Table: $\mathrm{CAR(p)}_1$: ML FIT; $\mathrm{CAR(p)}_2$: LS FIT, $p = 1, \ldots, 5$

The inverse correlations, $\tilde{r}_y(u)$, of these CARs can differ substantially from $\tilde{r}_x(u)$. For example, the CARs all have $\tilde{r}_y(1,0) > -0.55$, but $\tilde{r}_x(1,0) = -0.713$.

Comparison of CS-RSD and CS-CAR correlations

Example 2: In this example correlations and inverse correlations are generated by a 3-parameter CS-RSD(1,3) with $\alpha_{10} = 0.248$, $\beta_{10} = -0.248$, $\beta_{11} = 0$ and $\beta_{20} = 0.200$, and those for the fitted CS-CARs.

Some aspects of the CAR fits are similar to those in Example 1. **The least square fits (shown below) are less good** - only the one for $p = 5$ has $r_y(1,0) < r_y(2,0)$. These fits with $p = 5$ have reasonable values of F (but have 2 more parameters than the CS-RSD).

Models	Lag (1, 0)	(2, 0)	(3, 0)	(1,1)	(2, 1)	(3,1)	(2,2)	(3,2)	F
RSD(1,3)	0.391	0.462	0.300	0.363	0.340	0.265	0.282	0.228	0.407
CAR(1)	0.558	0.367	0.263	0.441	0.329	0.248	0.271	0.217	0.612
CAR(2)	0.446	0.328	0.244	0.432	0.305	0.236	0.264	0.212	0.628
CAR(3)	0.470	0.342	0.249	0.371	0.295	0.232	0.250	0.206	0.600
CAR(4)	0.462	0.417	0.268	0.393	0.274	0.246	0.267	0.197	0.545
CAR(5)	0.414	0.447	0.276	0.379	0.316	0.256	0.232	0.214	0.482

Comparison of CS-RSD and CS-CAR correlations

Example 2: the inverse correlations $\tilde{r}_y(u)$ are moderately different from $\tilde{r}_x(u)$ for both (ML/LS) fits with $p < 5$, but are relatively close for both fits with $p = 5$ (see Table below). 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)	(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) LS	-0.053	-0.033	-0.277	-0.009	0.132

Table: First row: inverse correlations of a CS-RSD(1,3) with $\alpha_{10} = 0.248$, $\beta_{10} = -0.248$, $\beta_{11} = 0$ and $\beta_{20} = 0.200$. Rows 2 to 3: inverse correlations of the CS-CAR(5) fitted by the ML and LS methods.

Model Fitting

Suppose the data are observed on an $(n_1 \times n_2)$ lattice $\mathcal{L}$ with $n = n_1\ n_2$ sites. Assume x is the n-vector of observations in, say, lexicographic order, and that $\theta = (\alpha'\ \beta')'$, with $x \sim N(\mu 1_n, \mathrm{R}_x\sigma^2)$, where 1_n is an $n-$vector of ones. Then the deviance, minus twice the log-likelihood is

$$\mathrm{D}(\theta, \sigma^2) = n \log(2\pi\sigma^2) + \log|\mathrm{R}_x| + (x - \mu 1_n)^T \mathrm{R}_x^{-1}(x - \mu 1_n)/\sigma^2$$

Then in theory models can be fitted by minimising the deviance over the valid parameter space. (In practice, optimization can be over R_x positive definite.)
A method frequently used is (essentially) to map the finite planar lattice $\mathcal{L}$ on to a torus (joining row and column ends), sometimes called periodic boundary conditions.

Model Identification and simulation

Model identification: Assuming model checks are satisfactory, fitted models can then be compared by their deviance values, by the generalized likelihood ratio test GLRT for nested models (using the difference in the deviances as having an asymptotic χ^2-distribution), and by using standard model-selection criteria such as AIC $= \mathrm{D} + 2\ N$ and BIC $= \mathrm{D} + N\ \log(n)$, for a model with N parameters (including the variance).

Simulation: If x is Gaussian and R_x or R_x^{-1} can be specified, it is simple to simulate a representation using any square root of R_x by $x = E[x] + \mathrm{R}_x^{1/2}\epsilon$ for ϵ a simulation from a $N(0, \sigma^2\mathrm{I})$. In particular, a torus simulation is easy using the known eigenvalues and eigenvectors of R_x.

Real Data Example: texture analysis

The data set is a (128×128) portion of a texture image, grass (1.1.01), available from `http://sipi.usc.edu/database/`.

Before modelling the spatial dependence, the planar lattice was mapped on to a (128×128) torus lattice. A constant mean looks reasonable. The histograms suggested Normality is a plausible working assumption, and the fitting used maximum likelihood.

The inverse correlations and the scaled interpolation variance F of the data were estimated with the sdf $f_x(\lambda)$ estimated by smoothing the periodogram. Using the two-dimensional form of the Daniell window with length (a, a) where a is between 17 and 21, gives an adequate amount of smoothing.

The standard errors for the estimated inverse correlations are obtained by Bootstrap. 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.

Real Data Example: texture analysis

The low-lag sample correlations (using a divisor of n) are shown below. These are high for neighbouring sites, but they then drop away quite quickly. They are slightly larger between rows than between columns. 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
3	0.100	0.114	0.146	0.209	0.278	0.278	0.211	0.139	0.090
2	0.108	0.138	0.212	0.347	0.482	0.448	0.300	0.177	0.104
1	0.114	0.175	0.312	0.573	0.801	0.663	0.386	0.211	0.119
0					1.000	0.747	0.399	0.210	0.119
	-4	-3	-2	-1	0	1	2	3	4
					u_1				

Real Data Example: texture analysis

The low-lag **estimated inverse correlations**, using $a = 19$, are given below. Taking some account of multiple testing, they suggest that most are significantly different from 0.

u_2									
4	-0.013	0.011	-0.003	-0.015	0.019	0.010	-0.007	-0.003	0.007
3	0.016	-0.014	0.002	0.028	-0.034	-0.034	0.043	-0.023	0.007
2	-0.003	-0.011	0.048	-0.119	0.152	-0.009	-0.049	0.044	-0.026
1	-0.040	0.092	-0.202	0.395	-0.531	0.274	-0.087	0.016	0.004
0					1.000	-0.696	0.356	-0.166	0.075
	-4	-3	-2	-1	0	1	2	3	4
					u_1				

Real Data Example: summary of the results

Amongst the fitted CAR's, both AIC $= 623.74$ and BIC $= 731.60$ criteria choose the general $\mathrm{CAR}(5)$ with $N = 12 + 2$ parameters (including the mean and the variance).

However, the RSD(1,2) model with $N = 2 + 4 + 2$ has AIC $= 580.10$ and BIC $= 641.73$, and the best values are for the $\mathrm{RSD}(2, 2)$ with $N = 4 + 4 + 2$ which has AIC $= 161.74$ and BIC $= 238.78$.

Although the correlations of the CAR(5) match those of the data within S_5, outside $S_5 \cup \{0\}$ 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 $\mathrm{RSD}(2, 2)$ correlations are reasonably close over $-4 \leq u_1 \leq 4$ and $0 \leq u_2 \leq 4$.

The values of the low-lag inverse correlations for the $\mathrm{RSD}(2, 2)$ are mainly much closer to those of the data than those of the CAR(5). The estimated value of F is 0.0329 for the $\mathrm{RSD}(2, 2)$, and 0.0464 for the CAR(5) - reference values for F are 0.0360 ($a = 17$), 0.0374 ($a = 19$), 0.0389 ($a = 21$).