Spatial Statistics 25 (2018) 68–85

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models

Jay M. Ver Hoef [a,*], Ephraim M. Hanks [b], Mevin B. Hooten [c,d]

[a] *Marine Mammal Laboratory, NOAA Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, United States*
[b] *Department of Statistics, The Pennsylvania State University, United States*
[c] *U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology, Colorado State University, United States*
[d] *Department of Statistics, Colorado State University, United States*

ARTICLE INFO

Article history:
Received 26 October 2017
Accepted 22 April 2018
Available online 1 May 2018

Keywords:
Lattice models
Areal models
Spatial statistics
Covariance matrix

ABSTRACT

We clarify relationships between conditional (CAR) and simultaneous (SAR) autoregressive models. We review the literature on this topic and find that it is mostly incomplete. Our main result is that a SAR model can be written as a unique CAR model, and while a CAR model can be written as a SAR model, it is not unique. In fact, we show how any multivariate Gaussian distribution on a finite set of points with a positive-definite covariance matrix can be written as either a CAR or a SAR model. We illustrate how to obtain any number of SAR covariance matrices from a single CAR covariance matrix by using Givens rotation matrices on a simulated example. We also discuss sparseness in the original CAR construction, and for the resulting SAR weights matrix. For a real example, we use crime data in 49 neighborhoods from Columbus, Ohio, and show that a geostatistical model optimizes the likelihood much better than typical first-order CAR models. We then use the implied weights from the geostatistical model to estimate CAR model parameters that provides the best overall optimization.

Published by Elsevier B.V.

1. Introduction

Cressie (1993, p. 8) divides statistical models for data collected at spatial locations into two broad classes: (1) geostatistical models with continuous spatial support, and (2) lattice models, also called

* Corresponding author.
E-mail address: jay.verhoef@noaa.gov (J.M. Ver Hoef).

https://doi.org/10.1016/j.spasta.2018.04.006
2211-6753/Published by Elsevier B.V.

areal models (Banerjee et al., 2004), where data occur on a (possibly irregular) grid, or lattice, with a countable set of nodes or locations. The two most common lattice models are the conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models, both notable for sparseness of their precision matrices. These autoregressive models are ubiquitous in many fields, including disease mapping (e.g., Clayton and Kaldor, 1987; Cressie and Chan, 1989; Lawson, 2013), agriculture (Cullis and Gleeson, 1991; Besag and Higdon, 1999), econometrics (Anselin, 1988; LeSage and Pace, 2009), ecology (Lichstein et al., 2002; Kissling and Carl, 2008), and image analysis (Besag, 1986; Li, 2009). CAR models form the basis for Gaussian Markov random fields (Rue and Held, 2005) and the popular integrated nested Laplace approximation methods (INLA, Rue et al., 2009), and SAR models are popular in geographic information systems (GIS) with the GeoDa software (Anselin et al., 2006). Hence, both CAR and SAR models serve as the basis for countless scientific conclusions. Because these are the two most common classes of models for lattice data, it is natural to compare and contrast them. There has been sporadic interest in studying the relationships between CAR and SAR models (e.g., Wall, 2004), and how one model might or might not be expressed in terms of the other (Haining, 1990; Cressie, 1993; Martin, 1987; Waller and Gotway, 2004), but there is little clarity in the existing literature on the relationships between these two classes of autoregressive models.

Historically, CAR and SAR models were obtained constructively, which naturally led to results on conditions of the constructions that yielded positive-definite covariance matrices. However, our goal is the opposite. We investigate how to obtain the properties of CAR and SAR models from a positive definite covariance matrix. We aim to clarify, and add to, the existing literature on the relationships between CAR and SAR covariance matrices. Cressie and Wikle (2011, p. 185) show how to obtain a CAR covariance matrix from a geostatistical covariance matrix, and, by extension, from any valid covariance matrix. We add to this by showing that any positive-definite covariance matrix for a multivariate Gaussian distribution on a finite set of points can be written as either a CAR or a SAR covariance matrix, and hence any valid SAR covariance matrix can be expressed as a valid CAR covariance matrix, and vice versa. This result shows that on a finite dimensional space, both SAR and CAR models are completely general models for spatial covariance, able to capture any positive-definite covariance. While CAR and SAR models are among the most commonly-used spatial statistical models, this correspondence between them, and the generality of both models, has not been fully described before now. These results also shed light on some previous literature. CAR and SAR models are often developed with sparseness in mind, where sparseness is the notion that the precision matrix has many zeros, allowing for the use of compact computer storage and fast computing algorithms for sparse matrices. Our results do not necessarily lead to sparse precision matrices for the SAR or CAR specifications, which is a desirable property for these models, so we spend some time investigating this with examples and discussion.

This paper is organized as follows: In Section 2, we review SAR and CAR models and lay out necessary conditions for these models. In Section 3, we provide theorems that show how to obtain SAR and CAR covariance matrices from any positive definite covariance matrix, which also establishes the relationship between CAR and SAR covariance matrices. In Section 4, we provide examples of obtaining SAR covariance matrices from a CAR covariance matrix on fabricated data, and a real example for obtaining a CAR covariance matrix from a geostatistical covariance matrix. Finally, in Section 5, we conclude with a detailed discussion of the incomplete results of previous literature.

2. Review of SAR and CAR models

In what follows, we denote matrices with bold capital letters, and their ith row and jth column with small case letters with subscripts i, j; for example, the i, jth element of $\mathbf{C}$ is $c_{i,j}$. Vectors of fixed values are denoted as lower case bold letters while vectors (or matrices) of random variables are bold, capital, and italic; let $\boldsymbol{Z} \equiv (Z_1, Z_2, \ldots, Z_n)^T$ be a vector of n random variables at the nodes of a graph (or junctions of a lattice). The edges in the graph, or connections in the lattice, define neighbors, which are used to model spatial dependency. Broad reviews of SAR and CAR can be found in Besag (1974), Wall (2004), and Ver Hoef et al. (2018), and in many books (e.g., Anselin, 1988; Haining, 1990; Cressie, 1993; Schabenberger and Gotway, 2005; Cressie and Wikle, 2011; Banerjee et al., 2014).