SPATIAL
STATISTICS

A closer look at the spatial exponential matrix specification

CrossMark

Erica Rodrigues [a,*], Renato Assunção [b], Dipak K. Dey [c]

[a] *Departamento de Estatística, Universidade Federal de Ouro Preto, Diogo de Vasconcelos, 122, Ouro Preto, Brazil*

[b] *Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Avenida Antônio Carlos, 6627, Belo Horizonte, Brazil*

[c] *215 Glenbrook Road, University of Connecticut, Storrs, CT 06269-4098, United States*

A R T I C L E   I N F O

A B S T R A C T

In this paper we analyze the partial and marginal covariance structures of the spatial model with the covariance structure based on a exponential matrix specification. We show that this model presents a puzzling behavior for many types of geographical neighborhood graphs, from the simplest to the most complex. In particular, we show that for this model it is usual to have opposite signs for the marginal and conditional correlations between two areas. We show these results through experimental examples and analytical demonstrations.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Suppose we have a region R partitioned into n disjoint areas $A_1, A_2, \ldots, A_n$ such that $\cup_{i=1}^{n} A_i = R$. The data observed in each of these areas are, typically, counts, sums or some type of aggregated value of characteristics associated with the each of these units. In order to introduce the spatial dependence between them, we need to define a neighborhood structure according to the geographical arrangement of these areas. Once defined the neighborhood structure, we specify models reflecting the geographical location of the data. One recurrent approach relies on ideas from autoregressive time series models. A popular model that include this type of structure is the Simultaneous Autoregressive (SAR) model (Whitle, 1954). The SAR model is specified by a set of regression equations in which the dependent variable is the observation in a particular area and the explanatory variables are

* Corresponding author. Tel.: +55 3188963607.
  E-mail address: ericaa_casti@yahoo.com.br (E. Rodrigues).

the observations on their neighbors. This system of equations is solved simultaneously inducing a multivariate normal distribution.

Let y_i be a value observed in area A_i. The SAR model is determined by the simultaneous solution of the set of equations given by:

$$Y_i = \mu_i + \sum_{j=1}^{n} s_{ij}(Y_j - \mu_j) + \epsilon_i \quad \text{for } i = 1, \ldots, n, \tag{1}$$

where ϵ_i are i.i.d. normally distributed with variance σ^2, $\mu_i = E(Y_i)$, and s_{ij} are parameter-dependent constants with $s_{ii} = 0$. If $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$, then the set of equations (1) leads to a multivariate normal distribution for $\mathbf{Y}$:

$$\mathbf{Y} \sim N\left(\boldsymbol{\mu}, \sigma^2(\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} - \mathbf{S}^T)\right),$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ and the matrix $\mathbf{S}$ is composed by the constants s_{ij} constrained in such a way as to guarantee the existence of $(\mathbf{I} - \mathbf{S})^{-1}$.

The maximum likelihood estimator of the unknown parameters, including the parameter dependent matrix, is typically hard to obtain due to the constrained optimization and the need to deal with the determinant of the covariance matrix. LeSage and Kelley Pace (2007) present a new spatial model that bypass these numerical problems. This is a particular case of the proposal made by Chiu et al. (1996) where general covariance matrices are expressed as the exponential function applied to a certain matrix. A major advantage of this definition is that it guarantees that the covariance matrix is always positive definite. Thus, it becomes unnecessary to put constraints on the parameter space during the estimation procedure to ensure that this property is satisfied. Another advantage is the simple way to obtain the precision matrix.

However, as we show in this paper, the exponential matrix covariance model has a major disadvantage when used in the spatial context. Mainly, this model is not flexible enough to incorporate many covariance patterns that one would expect to find in practice. The model is restricted to covariance patterns that have, in fact, non-intuitive aspects. The main point is that using this approach, in most practical situations, it will lead to opposite signs for the marginal and the partial correlations between pairs of neighboring areas. This limitation implies in an undesirable restrictive behavior for the stochastic structure of the data and therefore requires careful justification for its use. In Section 2, we study the properties of this exponential covariance matrix in detail starting by describing the model proposed by LeSage and Kelley Pace (2007). Consider the partial correlation $\mathrm{Cor}(Y_i, Y_j | \mathbf{Y}_{-ij})$, given all the other variables $\mathbf{Y}_{-ij}$ in the map. We prove in Section 3 that, for even-order neighboring pairs, $\mathrm{Cor}(Y_i, Y_j | \mathbf{Y}_{-ij})$ has the opposite sign as the marginal $\mathrm{Cor}(Y_i, Y_j)$, irrespective of the value of the coefficients in the model. In Section 4, we show that this behavior is non-intuitive and it is not allowed in the usual time series models. In irregular lattices, as those used in geographical data analysis, we showed in Section 5 that the non-intuitive sign change behavior of second-order neighbors is common although not guaranteed. In fact, we show that this happens for certain range of the exponential matrix parameters. Using matrix derivatives, we demonstrate that as the spatial correlation parameter increases, this puzzling behavior is intensified. More disconcerting, we show that, even for first-order neighboring areas in real maps, we can easily generate different signs for marginal and partial correlations. In Section 7, we discuss the consequences of this behavior and present our conclusions.

2. Model definition

To guarantee that the normal distribution induced by the set of simultaneous equations (1) is proper, the matrix $(\mathbf{I} - \mathbf{S})$ must be of full rank. One possibility to ensure this property is to set $\mathbf{S} = \rho\mathbf{D}$ and put restrictions on the space where the parameter ρ is defined. The matrix $\mathbf{D}$ can be specified in various ways. The two most common approaches are the following. In the first one, denoted by $\mathbf{W}$, we define a binary matrix: $\mathbf{W}_{ij}$ receives value 1 if the areas i and j are neighbors, and zero otherwise. Additionally, $\mathbf{W}_{ii} = 0$. Another commonly used alternative is to define the matrix $\mathbf{D}$

as a row standardized version of **W**. That is, $\mathbf{D}_{ij} = \mathbf{W}_{ij}/\mathbf{W}_{i+}$, where $\mathbf{W}_{i+} = \sum_j \mathbf{W}_{ij}$. The parameter ρ represents spatial association between areas. It can be shown that $(\mathbf{I} - \rho\mathbf{D})$ is nonsingular if

$$\rho \in \left(\frac{1}{\lambda_1}, \frac{1}{\lambda_n}\right)$$

where $\lambda_1 < 0 < \lambda_n$ are, respectively, the smallest and largest eigenvalues of the matrix **D**.

Considering the row standardized **D** matrix, we have that $\lambda_n = 1$. In this case the interval where ρ is defined is in the form $\rho \in (1/\lambda_1, 1)$. Under these conditions, it is possible to show that the matrix $(\mathbf{I} - \rho\mathbf{D})^{-1}$ can be expressed as

$$(\mathbf{I} - \rho\mathbf{D})^{-1} = \mathbf{I} + \rho\mathbf{D} + \rho^2\mathbf{D}^2 + \rho^3\mathbf{D}^3 \ldots . \tag{2}$$

The matrices **D** and $\mathbf{D}^k$ can be seen as transition matrices, the first one in one step and the second in k steps. $\mathbf{D}^k$ also represents the k-th order neighborhood. This means that we can go from i to j by passing through k edges, and we cannot do this in less than k steps. Therefore, from Eq. (2), we see that the influence of distant neighborhoods falls geometrically in this model. Assunção and Krainski (2009) analyzed how this decay influences the behavior of marginal correlations induced by the Conditional Autoregressive (CAR) model used as prior distribution in Bayesian models. Martellosio (2012) also addressed this problem focusing on SAR specification.

The idea of LeSage and Kelly Pace is to propose a model in which the influence of more distant neighborhoods falls more quickly. One possibility would be to replace $(\mathbf{I}-\rho\mathbf{D})^{-1}$ in the original version of the model, by the convergent matrix series

$$\mathbf{I} + \alpha\mathbf{D} + \frac{\alpha^2}{2!}\mathbf{D}^2 + \frac{\alpha^3}{3!}\mathbf{D}^3 \ldots . \tag{3}$$

The k-th term in (2) is now divided by $k!$ implying in a faster decrease towards zero and hence in a smaller impact of the k-th neighborhood in the infinite sum (3). This series is called exponential matrix and is denoted by $e^{\alpha\mathbf{D}}$. It has interesting properties that are listed below. Consider two arrays **X** and **Y** of dimensions $n \times n$ and let a and b be two scalars, then

$e^{\mathbf{0}} = \mathbf{I}$;
$e^{a\mathbf{X}}e^{b\mathbf{X}} = e^{(a+b)\mathbf{X}}$;
if $\mathbf{XY} = \mathbf{YX}$ then $e^{\mathbf{Y}}e^{\mathbf{X}} = e^{\mathbf{X}}e^{\mathbf{Y}} = e^{\mathbf{X}+\mathbf{Y}}$;
if **Y** is invertible then $e^{\mathbf{YXY}^{-1}} = \mathbf{Y}e^{\mathbf{X}}\mathbf{Y}^{-1}$.

In the model proposed by LeSage and Kelley Pace (2007) the covariance matrix is defined as

$$\Sigma_\alpha = \sigma^2 \left(e^{-\alpha\mathbf{D}'}e^{-\alpha\mathbf{D}}\right) \tag{4}$$

therefore mimicking the SAR covariance matrix, where **D** is a $n \times n$ matrix of spatial non-negative weights. As previously mentioned, typically it is considered that $\mathbf{D}_{ij} > 0$ if i and j are neighbors and $\mathbf{D}_{ij} = 0$, otherwise.

The covariance model proposed by Chiu et al. (1996) is more general and it is also based on an exponential matrix. They applied their model to analyze longitudinal data but the model could be applied in any regression context where one aims to model the covariance using the design matrix. Although we focus our paper on the spatial model suggested by LeSage and Kelley Pace (2007), the same problems we identified in this model are also present in the more general model of Chiu et al. (1996). In particular, in longitudinal data, it is not reasonable that the marginal and partial correlations have opposite signs. This would mean, for example, that if the measurement of an individual at time t tends to be above (below) the average, the measurement at time $t + 1$ will also tend to stay above (below) the mean. However, conditioned on information obtained at all other instants of time, this relationship assumes an opposite direction. It is easily verified that, for certain design matrices, this actually happens. Note, however, that this type of counterintuitive behavior is not observed in classical time series models such as the autoregressive ones. If their parameters are all positive, the marginal

and partial autocorrelation functions will have both positive signs for any order of neighborhood taken into consideration.

We demonstrate in this paper that the definition proposed by LeSage and Kelley Pace (2007) can induce marginal and conditional correlations with opposite signs. This type of problem has been widely studied when dealing with dependence structure between variables and is known as Simpson's Paradox (Pearl, 2000; Wagner, 1982). This paradox says that the sign of the correlation between two variables can be modified if we condition on a third variable. That is, the marginal correlation, without considering the third variable, has a different sign of the correlation obtained when conditioning in that variable.

This type of non-intuitive behavior for spatial models has been pointed out by Banerjee et al. (2003) (see p. 169) who say that for the SAR and conditional autoregression (CAR) models "transformation to (the covariance matrix) Σ_Y is very complicated and very non-linear. Positive conditional association can become negative unconditional association". However, for SAR and CAR models such kind of behavior is observed only in cases of little practical interest, those in which there is a negative spatial association, or a repulsion, between the areas (Wall, 2004; Assunção and Krainski, 2009; Martellosio, 2012). A simple explanation for this fact can be obtained from a result shown by Jones and West (2005). This result guarantees that if all partial correlations have positive sign, then the marginal correlation will also be positive. In the specification of these models, for positive values of the spatial association parameter, the partial correlations between any pairs of areas are always positive. Thus, the marginal correlation will also have positive sign. Other puzzling results in the spatial case are explained away by Assunção and Krainski (2009) and Martellosio (2012). As we show in the next sections, for the model proposed by LeSage and Kelley Pace (2007), this change on signs occurs even when there is a positive association between the areas.

3. The case of regular lattices

The first case to be analyzed is a simple situation, a regular lattice. This means that the number of neighbors (n_i) of each area is fixed and equal to 4, using the periodic boundary condition which identifies opposite sides of the lattice wrapping it up as a torus. Let us assume that the matrix $\mathbf{D}$ is standardized, i.e., $[\mathbf{D}]_{ij} = 1/n_i$ if areas i and j are neighbors and 0 otherwise. Let us also consider that $\sigma^2 = 1$. As $\mathbf{D}' = \mathbf{D}$ and $\mathbf{D}'\mathbf{D} = \mathbf{D}\mathbf{D}'$ we have that, in the exponential matrix model (4)

$$\Sigma_\alpha = e^{-\alpha \mathbf{D}'} e^{-\alpha \mathbf{D}} = e^{-\alpha \mathbf{D}' - \alpha \mathbf{D}} = e^{-2\alpha \mathbf{D}}.$$

We know that, by definition of the exponential matrix,

$$e^{-2\alpha \mathbf{D}} = \mathbf{I} - (2\alpha)\mathbf{D} + \frac{(2\alpha)^2}{2!}\mathbf{D}^2 - \frac{(2\alpha)^3}{3!}\mathbf{D}^3 + \frac{(2\alpha)^4}{4!}\mathbf{D}^4 - \frac{(2\alpha)^5}{5!}\mathbf{D}^5 + \cdots.$$

The term ij of the matrix $\mathbf{D}^k$ gives the probability that a random walk traveling in the adjacency graph leaves the area i and reaches j in k steps. We know that, for a regular lattice, it is possible to go to an odd-order neighbor (first, third, fifth, etc.) only in an odd number of steps, and to a neighbor of even order (second, fourth, sixth, etc.) in an even number of steps. This means that, if k is odd,

$$\mathbf{D}_{ij}^k \begin{cases} > 0 & \text{if } i \text{ and } j \text{ are neighbor of odd order} \\ = 0 & \text{otherwise} \end{cases}$$

and, if k is even,

$$\mathbf{D}_{ij}^k \begin{cases} > 0 & \text{if } i \text{ and } j \text{ are neighbor of even order} \\ = 0 & \text{otherwise.} \end{cases}$$

So, if i and j are neighbors of first order, the ij-th element of the covariance matrix is given by

$$[e^{-2\alpha \mathbf{D}}]_{ij} = -(2\alpha)[\mathbf{D}]_{ij} - \frac{(2\alpha)^3}{3!}[\mathbf{D}^3]_{ij} - \frac{(2\alpha)^5}{5!}[\mathbf{D}^5]_{ij} + \cdots \tag{5}$$

i.e., the correlation between areas i and j will have the same sign as $-\alpha$.

We also know that the precision matrix (inverse of the covariance matrix) in this case is given by

$$\left(e^{-2\alpha\mathbf{D}}\right)^{-1} = e^{2\alpha\mathbf{D}} = \mathbf{I} + (2\alpha)\mathbf{D} + \frac{(2\alpha)^2}{2!}\mathbf{D}^2 + \frac{(2\alpha)^3}{3!}\mathbf{D}^3 + \frac{(2\alpha)^4}{4!}\mathbf{D}^4 + \frac{(2\alpha)^5}{5!}\mathbf{D}^5 + \cdots.$$

As the sign of the conditional correlation is given by the opposite sign of the entries of the precision matrix, the sign of the conditional correlations between any pairs of areas is the opposite of the sign of α in regular lattices. For example, if i and j are first order neighbors, the ij-th element of the precision matrix $[e^{-2\alpha\mathbf{D}}]^{-1} = e^{2\alpha\mathbf{D}}$ is given by

$$[e^{2\alpha\mathbf{D}}]_{ij} = (2\alpha)[\mathbf{D}]_{ij} + \frac{(2\alpha)^3}{3!}[\mathbf{D}^3]_{ij} + \frac{(2\alpha)^5}{5!}[\mathbf{D}^5]_{ij} + \cdots.$$

By comparing this expression with that presented in (5), we can see that the marginal and partial correlations between i and j have the same sign.

Consider now a pair of second order neighbors, i and j. We have that the ij-th element of the covariance matrix is given by

$$[e^{-2\alpha\mathbf{D}}]_{ij} = \frac{(2\alpha)^2}{2!}[\mathbf{D}^2]_{ij} + \frac{(2\alpha)^4}{4!}[\mathbf{D}^4]_{ij} + \frac{(2\alpha)^6}{6!}[\mathbf{D}^6]_{ij} + \cdots, \tag{6}$$

which is equal to the ij-th element of the precision matrix

$$[e^{2\alpha\mathbf{D}}]_{ij} = \frac{(2\alpha)^2}{2!}[\mathbf{D}^2]_{ij} + \frac{(2\alpha)^4}{4!}[\mathbf{D}^4]_{ij} + \frac{(2\alpha)^6}{6!}[\mathbf{D}^6]_{ij} + \cdots.$$

Comparing this expression with (6) we notice that the marginal correlation between i and j is always positive, but the partial correlation is always negative since it is the opposite sign of the precision matrix. This same behavior can be observed for all neighbors of even order. That is, for any pair of neighbors of this type, the marginal correlation between them will always be positive, but when we condition on other areas of the map this correlation turns out to be negative.

This is not a reasonable result for a model of spatial dependence. Whereas a marginal positive correlation means that if area i has a value above its average, then area j will also tend to have a value higher than expected. If we consider the values of all the other areas in the map to be known, the fact that the area i have a value above its average, leads area j to have a value below its average. This happens only with even-order neighboring pairs and does not depend on the value of the conditioning variables.

For neighbors of even order we observe yet another strong constraint of this model. It only allows the existence of marginal positive correlations between neighbors of this type. That is, neighbors of second, fourth, sixth order can only be positively correlated regardless of the value of α.

Let us now see what are the consequences of these results for the conditional expectations, since the conditional expectation has a slightly more direct interpretation than the conditional correlation. Consider a Gaussian Markov Random Field whose precision matrix is denoted by $\mathbf{Q}$ with entries Q_{ij} and whose mean vector $\boldsymbol{\mu}$ is zero. It is known that the conditional expectation of the random variable at a site i, given the rest of the map ($\mathbf{Y}_{-i}$) can be obtained from the following expression

$$E(Y_i|\mathbf{y}_{-i}) = -\frac{1}{Q_{ii}}\sum_j Q_{ij}y_j. \tag{7}$$

We have seen that, for the case of the regular lattice, if i and j are neighbors of odd order, $Q_{ij} < 0$ and if i and j are neighbors of even order, $Q_{ij} > 0$. From Eq. (7), if we want to predict the value of the variable $\mathbf{Y}$ using the observed values of its neighbors, those from neighbors of odd order will give a positive contribution, while those from neighbors of even order enter with a negative weight in the expectation. That does not seem reasonable, since the marginal correlations of i with each of its neighbors have positive sign, so they should all have a positive weight in the conditional mean.

The regular lattice case induces a very specific type of neighborhood structure, but it is present in many practical situations, such as in image processing. We will see later that this kind of behavior repeats even in cases where we do not consider a regular lattice.
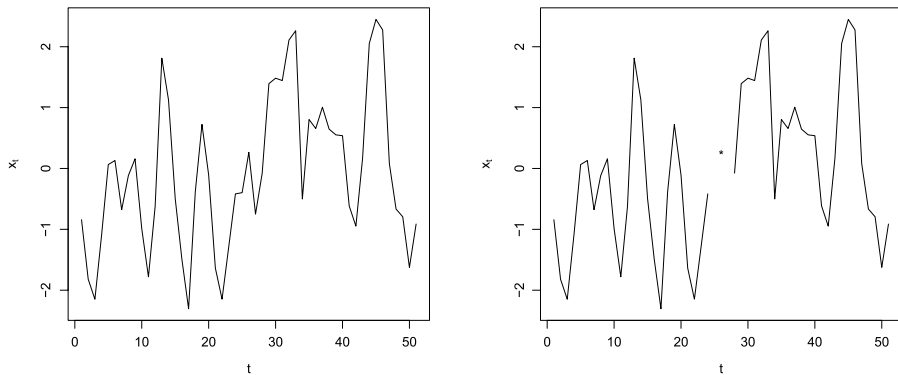
Fig. 1. Time series with 51 observations simulated (left) and simulated series removing observations of orders 25 and 27 (right).

In order to better understand the meaning of this sign change between marginal and conditional correlations, we present in the following section a short simulation study for the one-dimensional case.

4. The case of unidimensional lattices

The goal of this section is to show visually the kind of data that are generated when considering the covariance structure proposed by LeSage and Kelley Pace (2007). To facilitate the understanding and the graphic presentation, let us consider the case in which data is observed over time. We set the parameters $\sigma^2 = 1$ and $\alpha = -1$ and generate a time series with 51 observations. The neighborhood structure is a Markov chain of order one. The series generated is shown on the left of Fig. 1. Two observations were deleted and their values were resampled from the conditional distributions of these two variable points in time, given the rest of the series. The graph on the right panel of Fig. 1 shows the series with the two missing observations. In other words, we have a set of observations $x_1, x_2, \ldots, x_{51}$. We throw away the values of x_{25} and x_{27} and consider these positions as if they were missing observations from our database. We know that the conditional distribution of the vector (X_{25}, X_{27}), given the rest of the series, is a bivariate normal. We generate some pairs of observations with such distribution to visually check their relationship. In Fig. 2, the graph on the left panel shows in solid line the original series and the dashed line shows the values resampled from the conditional distribution. The points marked in black represent the values of the conditional expectation. It is observed that, as expected, if the values of one of the two selected points is below the conditional expectation, the other tends to be above its conditional expectation and vice versa. Fig. 2 also shows the scatter plot generated from various observations for these two variables. We note that there is a negative correlation between the two variables, which was already expected since the two sites are neighbors of second order. In addition, for many points generated, if the observation of one site is below (above) its conditional mean, the observation of the other is above (below).

In order to visually contrast the behavior of the series generated from the analyzed model with the behavior of a classical ARIMA model, we generate 51 observations of an autoregressive process of order 2. The model parameters were set at 0.3 and 0.4 and the variance was fixed at $\sigma^2 = 1$. The data were generated in the same way as was done in the previous case. The graphs on the left panel of Fig. 3 show the original series in solid line and observations generated through conditional distributions in dashed line. The black dots mark the conditional expectation. The right hand side of Fig. 3 shows the scatter plot of observations generated from the conditional distribution. Unlike what happens in the previous case, there is clearly a positive association between the observations in the two sites.

5. Spatial irregular lattice

In the case of real maps, the lattice is typically irregular. The phenomena mentioned earlier is also present in this situation. We first consider a case very close to the regular lattice, which presents
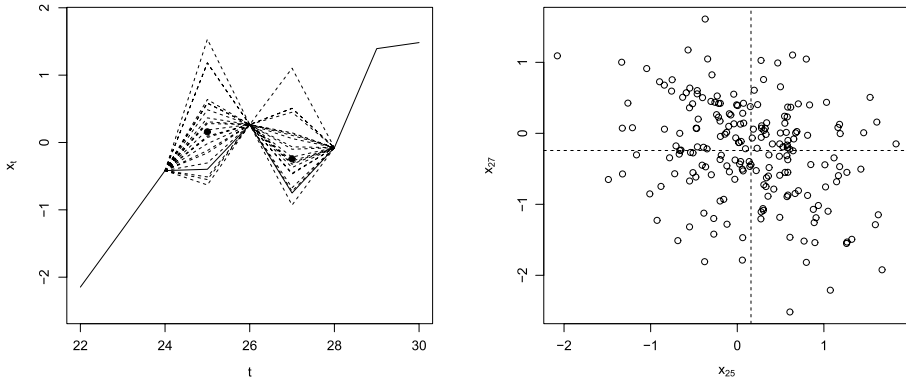
Fig. 2. Series generated, the dashed lines represent 20 observations generated from the conditional distributions and dots mark the conditional expectation (left). Scatter plot of five observations generated from the conditional distribution (right).
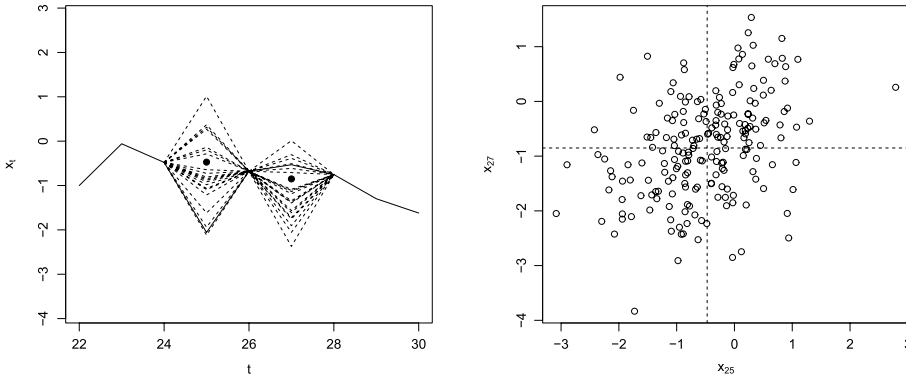
Fig. 3. Time series generated from an autoregressive process of order 2. Dashed lines represent the 20 observations generated from the conditional distributions and the points mark the conditional expectations (left). Scatter plot of 200 observations generated from the conditional distribution (right).

additional counterintuitive results. In this case we observe that the change of signs occurs even between first order neighbors. This situation is illustrated by the map of Iowa.

The map analyzed is shown in Fig. 4. The neighborhood criterion used was the adjacency between areas, i.e., two areas are considered neighbors if, and only if, they share boundaries. The neighborhood matrix is specified as a binary matrix, which takes only values zero and one. We fix $\alpha = -1$, $\sigma^2 = 1$ and hence the covariance matrix is of the form

$$\Sigma_\alpha = e^{\mathbf{W}'} e^{\mathbf{W}},$$

where **W** is a binary matrix such that $[\mathbf{W}]_{ij} = 1$ if i and j are neighbors, and zero otherwise.

We will check now for which pairs of neighbors the marginal and conditional correlations have opposite signs. The first map in Fig. 4 shows the connections between neighbors of the first order. The second map of this same figure shows the connections between neighboring areas that have marginal positive correlation. Note that all pairs highlighted on the map satisfy this condition. The map on the bottom left shows connections between neighboring areas that have positive conditional correlations. Therefore, for all the connections shown in this figure there is no change of signs. The map on the right shows connections between neighboring areas that have negative conditional correlation. This means that, for such linkages, sign change does occur. In this example, even first-order neighbors have marginal and conditional correlations with opposite signs. Despite the layout closely resembling a regular lattice, we have this additional unintuitive behavior.
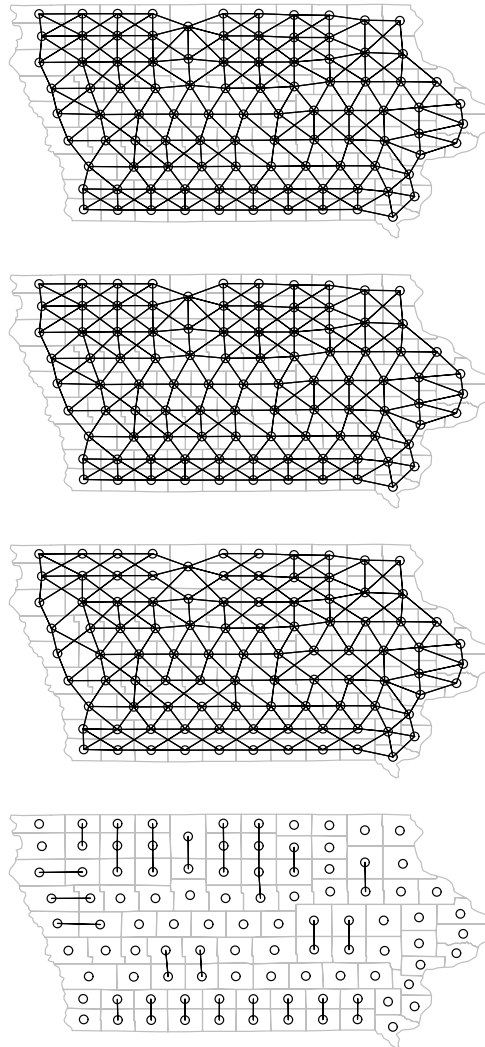
Fig. 4. Iowa map representing marginal and conditional correlations between pairs of neighboring areas. The first figure shows the map of Iowa with links representing neighbors of the first order. The second shows connections between neighbors of first order that have marginal positive correlation. The third figure shows connections between pairs of neighboring areas that have positive conditional correlation. The fourth figure shows connections between pairs of areas that have negative conditional correlations.

Real maps often have areas of very different sizes, inducing very irregular graphs. This situation is more difficult to be analyzed because the topology of the graph becomes more complex.

To illustrate this kind of situation, let us consider the map used in the application presented by LeSage and Kelley Pace (2007). The authors analyzed the participation of Americans in the presidential elections in 2000. The data is aggregated by counties and the entire state of Texas is excluded from the analysis due to irregularity problems in the elections. However, the ensuing discussion does not require any data, but only the spatial structure. The map used is shown in Fig. 5. The authors consider two types of neighborhood: adjacency and *k*-nearest neighbors. For simplicity, we consider here only the adjacency criterion. The links between some neighboring areas according to this criterion are shown in the first graph in Fig. 5. Note that we are expanding only one region of the United States
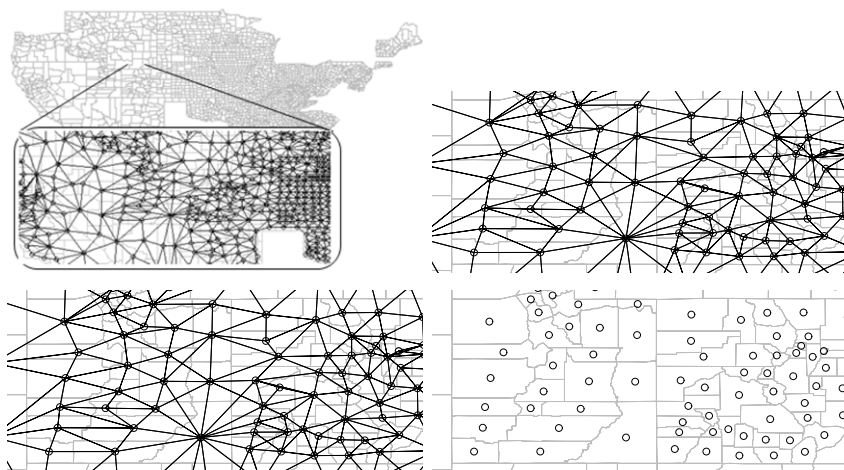
Fig. 5. First-order neighbors with positive marginal correlation, with positive partial correlation and with negative partial correlation.

in order to facilitate visualization. The value estimated for α in the application was -0.74 and this is the value used here. We adopted one of the **D** matrices used by those authors, the row-standardized matrix.

The second graph in Fig. 5 shows some of the pairs of neighbors that have marginal positive correlation. All first-order neighbors have this property. This was expected because for $\alpha < 0$ all the terms of the covariance matrix are positive. The first graph of the second line in the same figure shows the connections between first-order neighbors that have positive partial correlation. These are all neighbors of first order, i.e., in this case there is no change of sign. The last graph of this figure shows connections between first-order neighbors that have negative partial correlation, i.e., no pair of neighbors, which confirms the observation made previously. Differently from Fig. 4, we have no links between adjacent areas here. The reason is that Fig. 4 adopts the binary adjacency matrix for **D** while here we use the row standardized matrix.

The first graph in Fig. 6 shows links between neighbors of second order. The second graph of this figure shows the pairs of second-order neighbors that have positive marginal correlation. As in the first-order neighborhood case, the correlation between any pair of areas is positive. The first graph in the second row of Fig. 6 shows the connections between pairs of neighbors of second order that have positive partial correlation, i.e., no pair of neighbors. The bottom right graph shows the connections between second-order neighbors that have negative partial correlation. For this kind of neighbors, there is change of signs for all pairs. This behavior is repeated in general to neighbors of even order.

The correlations for this map behave in the same way that in the regular graph. Analyzing other maps, with completely different topological structure, with this same value for α, the same behavior was observed. We will see that this occurs because, as the value of α is small, the first order neighborhood pairs dominate the covariance and precision matrices.

To understand why this behavior occurs for any kind of map, we remind that the covariance matrix of the model is given by

$$\sigma^2 e^{-\alpha \mathbf{D}'} e^{-\alpha \mathbf{D}}.$$

Ignoring the variance term σ^2 and using the definition of $e^{-\alpha \mathbf{D}}$, this product can be written as

$$\left(\mathbf{I} - \alpha \mathbf{D}' + \frac{\alpha^2}{2!} \mathbf{D}'^2 - \frac{\alpha^3}{3!} \mathbf{D}'^3 + \cdots \right) \left(\mathbf{I} - \alpha \mathbf{D} + \frac{\alpha^2}{2!} \mathbf{D}^2 - \frac{\alpha^3}{3!} \mathbf{D}^3 + \cdots \right).$$

If $\alpha < 0$, all the terms of the covariance matrix will be positive. That is, the correlation between any pair of areas will be positive. Let us now see what is the form of the precision matrix. This matrix
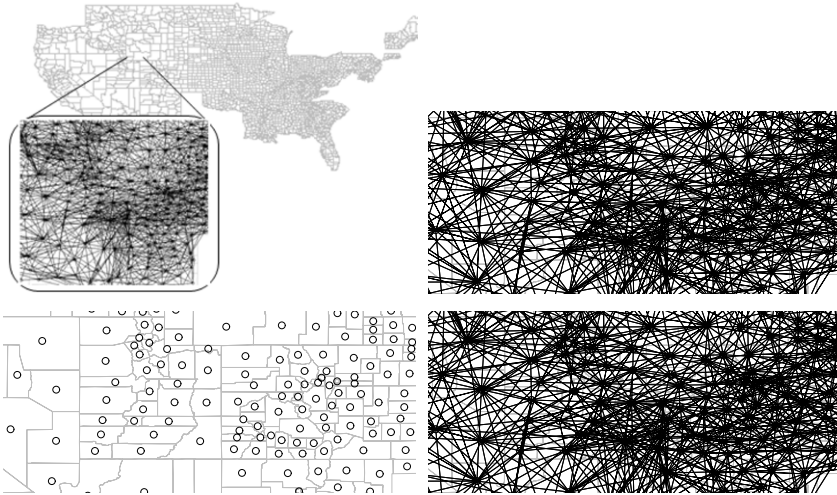
Fig. 6. Second-order neighbors with positive marginal correlation, with positive partial correlation and with negative partial correlation.

is given by

$$(\sigma^2)^{-1} e^{\alpha \mathbf{D}} e^{\alpha \mathbf{D}'}.$$

Again using the definition of $e^{\alpha \mathbf{D}}$ we get

$$\left(\mathbf{I} + \alpha \mathbf{D} + \frac{\alpha^2}{2!}\mathbf{D}^2 + \frac{\alpha^3}{3!}\mathbf{D}^3 + \cdots \right) \left(\mathbf{I} + \alpha \mathbf{D}' + \frac{\alpha^2}{2!}\mathbf{D}'^2 + \frac{\alpha^3}{3!}\mathbf{D}'^3 + \cdots \right).$$

If α is negative, we have the product of two summations involving positive and negative terms. Thus, the terms of the precision matrix can be positive or negative, depending on which array dominate the sum. We can rewrite the product as

$$\mathbf{I} + \alpha \mathbf{D}' + \alpha \mathbf{D} + \frac{\alpha^2}{2!}\mathbf{D}'^2 + \frac{\alpha^2}{2!}\mathbf{D}^2 + \alpha^2 \mathbf{D}'\mathbf{D} + \cdots. \tag{8}$$

Since α is small and $\mathbf{D}$ is limited between [0, 1], as k increases the terms

$$\frac{\alpha^k}{k!}\mathbf{D}^k \quad \text{and} \quad \frac{\alpha^k}{k!}\mathbf{D}'^k$$

go to zero. Therefore, the sum presented in (8) will be dominated only by the smaller exponents. Consider the case where two areas i and j are first-order neighbors. The terms $[\mathbf{D}]_{ij}$ and $[\mathbf{D}']_{ij}$ are both greater than zero and the sum in (8) is dominated by the terms

$$\alpha \mathbf{D}' + \alpha \mathbf{D}.$$

As we are considering $\alpha < 0$, this sum is smaller than zero, which implies a positive correlation conditional. That is, if i and j are first-order neighbors, marginal and conditional correlations have the same sign as we had already observed in the example.

Consider the case in which i and j are second-order neighbors. In this case, we have $[\mathbf{D}]_{ij} = [\mathbf{D}']_{ij} = 0$, but $[\mathbf{D}^2]_{ij}$ and $[\mathbf{D}'^2]_{ij}$ are both greater than zero. Therefore, the sum in Eq. (8) is dominated by the terms

$$\frac{\alpha^2}{2!}\mathbf{D}'^2 + \frac{\alpha^2}{2!}\mathbf{D}^2 + \alpha^2 \mathbf{D}'\mathbf{D}.$$

As all matrices involved in the above sum are positive, the result is a positive number. This implies a negative conditional correlation. That is, if i and j are neighbors of second order, marginal and
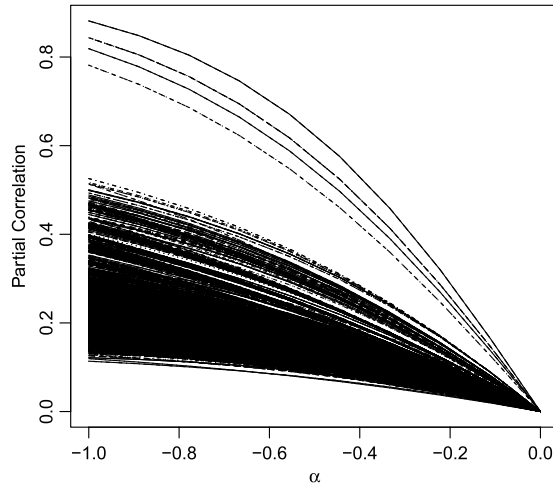
Fig. 7. Correlation between first order neighbors for different values of α.

conditional correlations have opposite signs. Hence, the behavior observed in the previous example is not restricted to the neighborhood structure defined before. For a small value of α, whatever map is considered, there will be change of signs in the correlations of second order neighbors.

6. How the partial correlation varies with α

The way the partial correlation varies with α also shows some non-intuitive aspects. According to LeSage and Kelley Pace (2007) the relationship between the parameter α of the proposed model and the parameter ρ of the SAR model is given by

$$\alpha \approx \ln(1 - \rho).$$

As the values of ρ with practical interest are $\rho > 0$, we observe two important points from this relationship. The first one is that the values of α that have practical interest are $\alpha < 0$. The second one is that α decreases as ρ increases. Taking this into consideration, it would be reasonable that the marginal and conditional correlations have a decreasing behavior with respect to the parameter α. We will see here that, for the case of conditional correlations, this is not true.

Considering again the same map used by LeSage and Kelley Pace (2007), we find the values of the conditional correlations for different values of the parameter α. We will check first how is this relationship for pairs of sites that are neighbors of first order. Fig. 7 presents the values of the correlations between first-order neighbors for α ranging from -1 to 0. Therefore, the conditional correlation decreases with the values of α. As explained earlier, this is a reasonable behavior, because it means that, as the spatial association between areas becomes stronger, the conditional correlation between certain pairs of areas will grow. For example, $\alpha = 0$ is equivalent to $\rho = 0$, and this means independence between the areas. In this case, from the graph, we see that the partial correlation between the neighbors of first order is zero. For $\alpha = -1$ we have $\rho \approx 0.632$, i.e., there should be an increase in the partial correlation between the areas. The graph presented in Fig. 7 shows that this is what happens.

Consider now what happens to the second-order neighbors. Fig. 8 presents the values of the correlations between neighbors of second order for α varying in the interval $[-1, 0]$. The conditional correlation grows with the value of α. In this case, as the value of α decreases, spatial association becomes stronger, the partial correlation between second-order neighbors increases, but it has a negative sign. This is not a reasonable behavior, since as the spatial association increases, it would be expected that, the partial correlation between areas should increase as well, but with a positive sign and not negative, as it is observed.
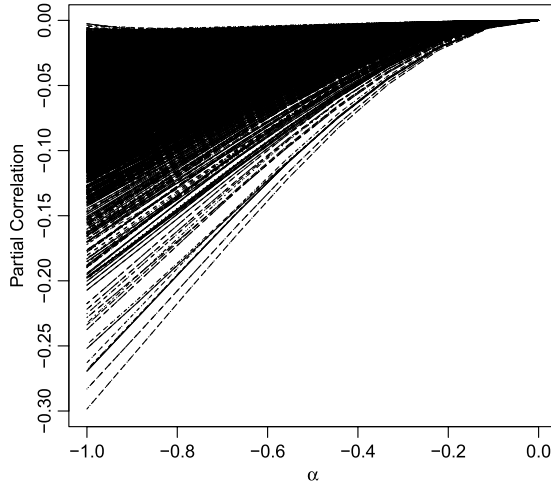
Fig. 8. Correlation between neighbors of second order for different values of α.

We can explain why this non-intuitive behavior happens for second order neighbors. Recall that the precision matrix is given by

$$\sigma^2 e^{\alpha \mathbf{D}} e^{\alpha \mathbf{D}'}.$$

Take derivatives of this product of matrices with respect to the parameter α:

$$\frac{de^{\alpha \mathbf{D}} e^{\alpha \mathbf{D}'}}{d\alpha} = \frac{de^{\alpha \mathbf{D}}}{d\alpha} e^{\alpha \mathbf{D}'} + e^{\alpha \mathbf{D}} \frac{de^{\alpha} \mathbf{D}'}{d\alpha}.$$

Using

$$\frac{de^{\alpha \mathbf{D}}}{d\alpha} = \mathbf{D} e^{\alpha \mathbf{D}}$$

we have

$$\frac{de^{\alpha \mathbf{D}} e^{\alpha \mathbf{D}'}}{\alpha} = \mathbf{D} e^{\alpha \mathbf{D}} e^{\alpha \mathbf{D}'} + e^{\alpha \mathbf{D}} \mathbf{D}' e^{\alpha \mathbf{D}'}.$$

Using the definition of $e^{\alpha \mathbf{D}}$, this derivative becomes

$$\mathbf{D} \left(I + \alpha \mathbf{D} + \alpha^2 \frac{\mathbf{D}^2}{2!} + \cdots \right) \left(I + \alpha \mathbf{D}' + \alpha^2 \frac{\mathbf{D}'^2}{2!} + \cdots \right)$$
$$+ \left(I + \alpha \mathbf{D} + \alpha^2 \frac{\mathbf{D}^2}{2!} + \cdots \right) \mathbf{D}' \left(I + \alpha \mathbf{D}' + \alpha^2 \frac{\mathbf{D}'^2}{2!} + \cdots \right)$$

which can be rewritten as

$$\mathbf{D} + \alpha \mathbf{D}\mathbf{D}' + \alpha \mathbf{D}^2 + \alpha^2 \mathbf{D}^2 \mathbf{D}' + \alpha^2 \frac{\mathbf{D}^2}{2!} + \cdots + \mathbf{D}' + \alpha \mathbf{D}\mathbf{D}' + \alpha \mathbf{D}'^2 + \alpha^2 \mathbf{D}\mathbf{D}'^2 + \alpha^2 \frac{\mathbf{D}'^2}{2!} + \cdots.$$

As this sum converges absolutely, we can change the order of summation. Assuming that α is small, we can ignore the terms with power greater than or equal to three and obtain

$$(\mathbf{D} + \mathbf{D}') + \alpha (\mathbf{D}\mathbf{D}' + \mathbf{D}'\mathbf{D} + \mathbf{D}^2 + \mathbf{D}'^2) + \alpha^2 \left(\mathbf{D}^2\mathbf{D}' + \mathbf{D}\mathbf{D}'^2 + \frac{\mathbf{D}^2}{2!} + \frac{\mathbf{D}'^2}{2!} + \cdots \right).$$

This expression allow us to analyze the sign of this derivative when the areas i and j are neighbors of first or second order. When they are first-order neighbors, the first term dominates the sum and the

derivative is positive. That is, the conditional correlation increases when the parameter α decreases. Remember that the conditional correlation sign has the opposite sign of each term of the precision matrix. Therefore, as the spatial association between the areas increases ($\alpha \to -\infty$), the conditional correlation between all neighbors of first order increases. On the other hand, if i and j are second order neighbors, the first term of the sum is zero and the second term is nonzero and dominates the sum. Since, in general, we are interested in the case where $\alpha < 0$ this sum assumes a negative value. This means that if two areas are second-order neighbors, the conditional correlation between them decreases as the parameter α decreases. That is, as the spatial association between the areas increases, the conditional correlation between the second order neighbors gets stronger. However it gets stronger negatively, which is counterintuitive, since the increase in spatial association is leading to conditional repulsion between the areas.

7. Conclusion

We show in this paper that the spatial model where the covariance structure depends on an exponential matrix has some important non-intuitive aspects. Despite the advantages pointed out by LeSage and Kelley Pace (2007), especially with regard to computational efficiency, the puzzling results shown here should be taken into account when considering this model. The type of non-intuitive behavior that we found in the exponential matrix model can also be observed for other widely used models, such as CAR and SAR models. However, in the case of the latter models, it only occurs when the spatial dependence parameter has a negative sign, i.e., a case with little applicability to real problems. As for the model proposed in LeSage and Kelley Pace (2007), this behavior is usually observed in situations of practical interest. Therefore, the application of this model should be done with caution, since the results may reflect a type of spatial dependence that has difficult practical interpretation.

Acknowledgment

References

Assunção, R., Krainski, E., 2009. Neighborhood dependence in Bayesian spatial models. Biom. J. 51 (5), 851–869.
Banerjee, S., Gelfand, A.E., Carlin, B.P., 2003. Hierarchical Modeling and Analysis for Spatial Data. CRC Press.
Chiu, Y.M.T., Leonard, T., Tsui, K., 1996. The matrix-logarithmic covariance model. J. Amer. Statist. Assoc. 91 (433), 198–210.
Jones, B., West, M., 2005. Covariance decomposition in undirected Gaussian graphical models. Biometrika 92 (4), 779–786.
LeSage, J.P., Kelley Pace, R., 2007. A matrix exponential spatial specification. J. Econometrics 140 (1), 190–214.
Martellosio, F., 2012. The correlation structure of spatial autoregressions. Econometric Theory 28 (6), 1373.
Pearl, J., 2000. Causality: Models, Reasoning and Inference, Vol. 29. Cambridge Univ. Press.
Wagner, C.H., 1982. Simpson's paradox in real life. Amer. Statist. 36 (1), 46–48.
Wall, M., 2004. A close look at the spatial structure implied by the car and sar models. J. Statist. Plann. Inference 121, 311–324.
Whitle, P., 1954. On stationary process in the plane. Biometrika 41 (3–4), 434–449.