ABSTRACT: The need to compute a normalizing constant hampers parameter estimations of spatial auto-Gaussian models because that constant is a function of an n-by-n matrix, which in turn is a function of some spatial autoregressive parameter(s). Approximations to this normalizing constant that yield estimates indistinguishable from exact maximum likelihood ones are evaluated for examples involving \( n \leq 3,888 \). Also with these approximations, estimates that appear sensible are computed for examples involving \( n \) as large as 250,000. Moreover, these approximations allow virtually any size georeferenced data set to be described with a spatial autoregressive model, and are shown to have negligible impact upon the standard error estimates for the regression coefficients. Consequently, from now on a quantitative spatial analyst should find implementation of spatial auto-Gaussian models quite simple and straightforward.

KEY WORDS: Jacobian; spatial autocorrelation; spatial autoregression; auto-Gaussian; eigenvalues

1. Background

Positive spatial autocorrelation is a tendency for similar values of a single variable \( Y \) to be present in nearby locations on a map; it is displayed when observations contained in data sets are locationally tagged to the earth's surface (i.e., georeferenced data sets). The prevailing nature
and degree of spatial autocorrelation may be denoted by $\rho$, while the self-covariation of $n$ geographically neighboring values within a variable may be represented with the $n$-by-$n$ matrix $V^{-1}\sigma^2$, which is a function of $\rho$. This geographic dependency feature of georeferenced data is captured by the auto-Gaussian log-likelihood function

$$\text{constant} - (n/2)\ln(\sigma^2) + \ln[\det(V)] - (Y - XB)^T V (Y - XB)/(2\sigma^2),$$

(1)

where $\det(V)$, superscript $T$, and $LN$ respectively denote the matrix determinant and transpose operations and the natural logarithm, $Y$ is an $n$-by-1 vector of georeferenced values, $X$ is an $n$-by-($p+1$) matrix of $p$ corresponding predictor variables coupled with a vector of ones, and vector $\beta$ and scalar $\sigma$ respectively denote the standard nonconstant mean and constant variance. The parameters of equation (1) most often are estimated using maximum likelihood (ML) techniques.

The determinant, $\det(V)$, is a normalizing constant because it ensures that the probability density function integrates to 1; it is the Jacobian of the transformation from a spatially autocorrelated mathematical space to a spatially unautocorrelated one. For spatial scientists wishing to implement spatial statistical models, this normalizing constant is problematic (Ripley, 1990; Cressie, 1991) because it (a) is a function of unknown spatial autocorrelation parameters, (b) is unwieldy, (c) fails to have a closed-form expression, and (d) almost always defies a numerical solution for sufficiently large $n$ since it involves an $n$-by-$n$ matrix. Solutions to these problems are the focus of this paper and derive from extending the work of Griffith and Sone (1995), Barry and Pace (e.g., 1999), Smirnov and Anselin (2001), and Pace and LeSage (2002). They exploit the smoothness of the curve traced by $\det(V)$, as illustrated in Figure 1.
The complete term involving \( \det(V) \), \( -\frac{2}{n} \ln[\det(V)] \), may be accurately approximated with

\[
\alpha_0 + \alpha_1 \ln(\delta_1 + \rho) - \alpha_2 \ln(\delta_2 - \rho)
\]

(Griffith and Sone, 1995). The five coefficients \( \alpha_0, \alpha_1, \delta_1, \alpha_2 \) and \( \delta_2 \) are calibrated using the eigenvalues of matrix \( V \). Equation (2) not only accurately describes the term \( -\frac{1}{n} \ln[\det(V)] \) based upon the square root of \( \det(V) \)—but it also fails to introduce any biases into the parameter estimates of equation (1).

In the literature, matrix \( V \) often is expressed as a function of both a binary (0-1) connectivity matrix \( C \) (whose \( c_{ij} \) element value is 1 if locations i and j are nearby, and 0 otherwise) and the spatial autocorrelation parameter \( \rho \). Two popular specifications of this matrix are \( V = (I - \rho C)^{-1} \) and \( V = [(I - \rho C)^T(I - \rho C)]^{-1} \). A more appealing specification of the second expression for application purposes converts matrix \( C \) to its stochastic version, denoted here by matrix \( W \), where \( w_{ij} = \frac{c_{ij}}{n \sum_{j=1}^n c_{ij}} \). This conversion process, which casts \( y_i \) in equation (1) as a function of the average of its surrounding geographically nearby values, entails calculating vector \( \mathbf{C1} \), converting this vector to a diagonal matrix, say \( D \), and then constructing \( W = D^{-1/2} \mathbf{C} \).

\( W \) is an asymmetric matrix whose eigenvalues are the same as those of symmetric matrix \( D^{-1/2} \mathbf{C} D^{-1/2} \). Accordingly, the specification of matrix \( V \) for the second of the two popular versions

\(^1\)The search for equation (2) and then its calibration using the eigenvalues of matrix \( V \) was motivated by Ord (1975), who promotes use of the eigenvalue form of \( \det(V) \).
becomes \([ (I - \rho W)^T (I - \rho W) ]^{-1} \), with all of the eigenvalues of matrix \(W\) being real numbers.

2. The normalizing constant approximation: history, description, and generalization

To help in understanding the nature and consequences of problems with the normalizing constant when estimating parameters of equation (1), a brief overview of the history, description, and general specifications of the auto-Gaussian normalizing constant is furnished. While addressing these problems dates back to Whittle (1954), the first simplified equation describing the normalizing constant, based upon the eigenvalues of matrix \(V\), appears in Griffith and Sone (1995). Although alternative approximations appear in the more recent literature, an understanding of the Griffith-Sone solution furnishes a conceptual basis for better understanding the nature and role of this normalizing constant. The Griffith-Sone solution involves establishing approximation equation coefficients that no longer require the full set of \(n\) eigenvalues of matrix \(V\) to be calculated, thus allowing estimation of parameters contained in equation (1) for extremely large georeferenced data sets. This methodology is extended in §4 to show how equation (2) can be used to calculate asymptotic standard errors for large georeferenced data sets.

2.1. History

The impediment to implementing spatial statistics because of the normalizing constant term contained in equation (1) was recognized early on by Whittle (1954). He proposed an approximation relating to the integral of the log-spectral density function for a stationary spatial autoregressive process operating over a regular square tessellation (i.e., surface partitioning) of
areal units. Unsuccessful attempts to extend Whittle's work using the Maple software package, in a quest for a closed-form solution, led to equation (2). The subsequent search for a functional form of equation (2) involved a series of numerical experiments with increasingly larger regular square tessellations, conducted on a Cray supercomputer. This experimental work resulted in the identification of the expression

\[ 2\alpha_n LN(\delta_n) - \alpha_n LN(\delta_n + \rho) - \alpha_n LN(\delta_n - \rho) \]  

(Griﬃth, 1990, p. 187). The ﬁve coefﬁcients were found to be functions of \( n \), and expression (3) then was extended to irregular tessellations (Griﬃth, 1992) and to regular hexagonal tessellations (Griﬃth, 1993), resulting in expression

\[ \alpha_{1,n} LN(\delta_{1,n}) + \alpha_{2,n} LN(\delta_{2,n}) - \alpha_{1,n} LN(\delta_{1,n} + \rho) - \alpha_{2,n} LN(\delta_{2,n} - \rho) \]  

Expression (4) reduces to expression (3) because of the symmetry of the eigenvalue frequency distribution for a regular square tessellation.

Feedback on these numerical experiments's results included (a) graphical identiﬁcation of relationships between the coefﬁcients of expression (4) and the extreme eigenvalues of matrix \( C \) or \( W \), whichever was being employed, and (b) Martin's (1993) discussions of relationships between simple quantities easily calculable for these matrices and approximation of the Jacobian term. Griﬃth and Sone (1995) conducted a numerical experiment and discovered that maximum likelihood estimates (MLEs) based upon expression (4) are indistinguishable from those based upon maximizing expression (1); savings in computational time and computer memory can be
Substantial when expressions (3) and (4) are utilized.

Spatial statistics is not the only part of statistics seeking to approximate this type of normalizing constant. The spatial autocorrelation result is more similar to that for a classical bivariate correlation situation than that for a traditional nonlinear regression specification for a quadratic design problem, proposed by Hamilton and Watts (1985). The bivariate log-likelihood function has a Jacobian term described by expression (3), with $\delta_n = \alpha_n = 1$. Figure 2a depicts the plot of this Jacobian, as well as the Hamilton-Watts quadratic approximation to it, over the feasible parameter space. Clearly the quadratic approximation supplies a poor descriptor [e.g., its RESS, the relative error sum of squares (i.e., the error sum of squares divided by the total corrected sum of squares), is 0.6210 for the interval defined by $\pm 0.999$, and 0.1590 for the interval defined by $\pm 0.96$], even though both curves are concave upwards. Although the quadratic curve closely follows the Jacobian across all but the extreme portions of the feasible parameter space, it tends to moderately underpredict in the interval (-0.4, 0.4) and moderately overpredict in intervals (-0.96, -0.4) and (0.4, 0.96). In comparison, Figure 2b depicts the plot of the Jacobian term for a linear geographic landscape (the spatial autocorrelation situation), as well as its expression (3) approximation. Clearly expression (3) furnishes an exceptionally good descriptor. The symmetry of the plot, attributable to the symmetry of the eigenvalue frequency distribution, can be contrasted with sample asymmetric plots associated with the more typical

---

In their case, a quadratic approximation to the Jacobian term is posited—but for the experimental design supporting inferences when estimating parameters of nonlinear statistical models.

Expression (3) does contain specification error. Considering matrix $W$ for a linear geographic landscape, whose eigenvalues are given by $\cos\left(\frac{k}{n-1}\pi\right)$; asymptotically the Jacobian term converges upon $-\ln\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 - \rho^2}\right)$. For $n = 1,000,000$ the coefficients of expression (3) are $\beta_n = 0.23326$ and $\delta_n = 1.02664$, with a RESS equaling $4.2 \times 10^{-3}$, when this expression is equated to the asymptotic exact Jacobian expression; corresponding unconstrained results are $\beta_n = 0.24202$ and $\delta_n = 1.03521$, with a RESS equaling $6.7 \times 10^{-4}$. Hence specification error is present, though it is rather trivial and is partially compensated for by allowing $\beta_n$ and $\delta_n$ to freely vary.
skewed frequency distributions appearing in Figure 1.

2.2. Derivation of Griffith and Sone's approximation specification

The single parameter auto-Gaussian model Jacobian term may be denoted by \( J(\rho) \). As it stands, this Jacobian term is for either a simultaneous autoregressive (SAR) or an autoregressive response (AR) spatial statistical model, for which matrix \( V \) usually is given by \([ (I - \rho W)^T (I - \rho W) ]^{-1} \). Therefore, \( \det(V) \) is defined in terms of either the eigenvalues of matrix \( C \) or matrix \( W \). The specification selected for matrix \( V \) alters the coefficient values but not the form of the approximation equation. If a conditional autoregressive (CAR) spatial statistical model is posited, then only matrix \( C \) is used in expression (1), the square root of this term is needed, and matrix \( V \) is given by \( (I - \rho C)^{-1} \).

The algebraic derivation of approximation equation (1) appears in ***** (2002b) and supplies a conceptual basis for equations (2)-(4). The logical development begins with a separation of positive and negative eigenvalues. Next, applying the trapezoidal rule of calculus for approximating the area under a curve with polygons, the sum of the eigenvalues of \( LN[\det(V)] \) is approximated by the sum of a pair of integrals, one spanning the negative and one spanning the positive eigenvalue range. These ranges reveal the importance of the extreme eigenvalues. Solving the pair of integrals renders the following equation as a possible approximation, which contains the spatial autocorrelation parameter \( \rho \) as a denominator term:

\[
\hat{J}_1(\rho) = 0, \text{ if } \rho = 0, \text{ and }
\]
\[ \hat{J}_1(\rho) = \alpha_{1,n}\{\frac{LN(1-\rho \lambda_n)}{\rho \lambda_n} + 1\} - \gamma_{1,n} LN(1-\rho \lambda_n) \]

\[ \hat{J}_2(\rho) = \rho \frac{\alpha_{2,n} \lambda_n^2 - \alpha_{1,n} \lambda_1^2}{2} + \alpha_{2,n} |\lambda_n| LN\left(\frac{\delta_{2,n}}{\lambda_n}\right) + \alpha_{1,n} \lambda_1 LN\left(\frac{\delta_{1,n}}{\lambda_1}\right) \]

\[ - \alpha_{2,n} |\lambda_n| LN\left(\frac{\delta_{2,n}}{\lambda_n}\right) - \alpha_{1,n} \lambda_1 LN\left(\frac{\delta_{1,n}}{\lambda_1}\right) - \rho. \]

The extreme eigenvalues of matrix \( W \) are denoted by \( \lambda_1 \) (the principal eigenvalue) and \( \lambda_n \) (the most negative eigenvalue). Because \( \rho \) can equal 0, terms from a truncated log-series expansion can be substituted to remove \( \rho \) as a divisor in equation (5), rendering equation (6):

While experiments by Griffith and Sone (1995) show that equation (6) furnishes a very good approximation of the normalizing constant, subsequent experiments suggest that equation (5) may be preferable to equation (6) because it tracks modestly better the curve of the Jacobian term across the feasible values of \( \rho \).

The derivation of equations (5) and (6) discloses why the coefficients of equations (3) and (4) are a function of \( n \) and the extreme eigenvalues, why the intercept term of equation (2) takes on the forms it does in equations (3) and (4), why equation (5) should perform better than equation (6), and why the form of equation (2) does not change for the different popular specifications of matrix \( V \).
2.3. Extensions of Griffith and Sone's approximation

Equation (6) is continuous across the feasible parameter space; it is, however, subject to additional deviations from the exact Jacobian term because of the introduction of a truncated log-series expansion. Equation (5) does not suffer from this drawback, but it is discontinuous at \( \rho = 0 \). For both equations (5) and (6), though, the set of \( n \) eigenvalues must be calculated before the coefficients \( \alpha_{j,n} \), \( \delta_{j,n} \), and \( \gamma_{j,n} \) \((j = 1, 2)\) can be calibrated, severely restricting their utility. To remedy this situation, the coefficient estimators (7a)-(7d) are proposed for when matrix \( \mathbf{W} \) is used in equation (1) and \( J(\rho) \) is approximated with equation (5). These estimation equations were obtained by first calibrating equation (5) for 148 conveniently available surface partitionings, ranging in size from \( n = 7 \) to \( n = 7,249 \). The feasible parameter space \([0.999 \lambda_n, 0.999 \lambda_1]\) was divided into a set of 200 regularly spaced values, and then the coefficients were estimated by setting \( J(\rho)^{1/2} = -\frac{1}{n} LN[\det(\mathbf{V})] \) as the dependent variable in a nonlinear regression.

Next these estimated coefficients were used as dependent variables in nonlinear regression equation versions of (7a)-(7d), where \( n_i \) denotes the number of ones in row \( i \) of matrix \( \mathbf{C} \).

\[
\hat{\alpha}_{1,n} = -0.0473 - 94.2089\left(\frac{1}{n+5}\right)^{7/3} + 1.1987\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} n_i n_j}{n} - 0.2927 LN(|\lambda_n| - 0.5) \quad (7a)
\]

\[
\hat{\alpha}_{2,n} = 0.0714 - 94.2089\left(\frac{1}{n+5}\right)^{7/3} + 1.1987\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} n_i n_j}{n}, \quad (7b)
\]
The coefficient pairs were estimated simultaneously to ensure their equality in the regular square tessellation case. Since the maximum eigenvalue of matrix $W$ always is one, equations (7b) and (7d) simplify, because they involve $\lambda_1 = 1$ rather than $\lambda_n$. The term $\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}}{n}$ equals the average sum of the squared eigenvalues and can be calculated quickly because matrix $C$ is sparse, often having fewer than $6(n-2)$ ones. This quantity equals $\frac{n+1}{2}$ for a linear surface partitioning, $\frac{18PQ + 11P + 11Q + 12}{72}$ for a regular square surface partitioning forming a P-by-Q rectangular region ($n = PQ$), $\frac{30PQ + 25P + 24Q + 23}{180}$ for a P-by-Q rectangular region partitioned into regular hexagons, and $\frac{3n^3 + 11n^2 - 39n + 73}{24n^2 - 48n + 24}$ for a maximally connected surface partitioning. When divided by $n$, these four quantities respectively asymptotically converge on $1/2$, $1/4$, $1/6$, and $1/8$. Finally, ***** (2002b) presents an algorithm for estimating the minimum

\[
\hat{Y}_{1,n} = 0.9837 + 0.0246 \left( \frac{20.0202}{n-2} \right)^{(l+1,6l_n)} , \quad \text{and} \tag{7c}
\]

\[
\hat{Y}_{2,n} = 0.9837 + 0.0246 \left( 1 + \frac{20.0202}{n-2} \right)^{2.6} . \tag{7d}
\]
The respective pseudo-$R^2$ values for equations (7a)-(7d) are 0.894, 0.755, 0.685 and 0.865. These four equations produced Shapiro-Wilk statistics for their residuals, suggesting non-normality. A common set of anomalous configurations for these equations, revealed by outlier detection diagnostic statistics, contained only 2 of the 148 surfaces. Any deficiencies uncovered by diagnostics, however, are more than offset by not having to compute all $n$ eigenvalues. Only the minimum eigenvalue needs to be computed, since theoretically $\lambda_1 = 1$ because matrix $V$ is specified in terms of matrix $W$.

Results reported in Table 1 suggest that equations (7a)-(7d) should perform reasonably well for massively large geographic landscapes upon which typical irregular surface partitionings have been superimposed. In practice these partitionings appear to be close to some mixture of the regular square and hexagon surface partitionings. For example, in the 148 surfaces in the database explored here, 26.2% of the areal units have 4 neighbors, 21.4% have 5 neighbors, and 15.5% have 6 neighbors. In addition, the set of $\frac{n}{n_{ij}}$ values contained in that database scatters about a negative exponentially declining trajectory that converges on 0.1695, which falls between the limits of $1/4$ for a regular square and $1/6$ for a regular hexagonal surface partitioning. The asymptotic Jacobian term, $-LN[\frac{1}{2} + \frac{1}{2}\sqrt{1 - \rho^2}]$, was used to compute the coefficients for the linear surface partitioning. The approximate eigenvalues for a 3500-by-3000...
rectangular region\(^5\), specified in \(*\)\(\ast\)\(\ast\)\(\ast\)\(\ast\) (2002a), were used to compute the coefficients for the regular square and for the regular hexagonal surface\(^6\) partitioning. Finally, Griffith and Sone (1995) note that two of the eigenvalues for a maximally connected surface partitioning are 1 and \(\frac{1}{n-1}\); the remaining n-2 eigenvalues may be approximated with \(\frac{1}{2}\cos\left(\frac{n-k-2}{n-3}\pi\right)\), \(k = 0, 1, \ldots, n-2\). These approximate eigenvalues were used, for the case of \(n = 1,000,000\), to compute the coefficients for the maximally connected surface partitioning. Results for the linear and maximally connected cases are informative because they furnish the lower and upper limits for a geographic landscape.

| Table 1. Coefficients for equation (5) for very large benchmark surface partitionings |
|--------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                | linear          | square          | hexagon         | maximally connected |
| \(\alpha_1\)                  | 0.6643          | 0.3745          | 0.5567          | 0.4707           |
| Eq. (7a)                       | 0.6707          | 0.3711          | 0.5705          | 0.6270           |
| \(\alpha_2\)                  | 0.6643          | 0.3745          | 0.3179          | 0.3220           |
| Eq. (7b)                       | 0.6707          | 0.3711          | 0.2712          | 0.2212           |
| \(\delta_1\)                  | 1.0645          | 0.9952          | 0.9957          | 1.1295           |
| Eq. (7c)                       | 1.0083          | 1.0083          | 0.9922          | 0.9908           |
| \(\delta_2\)                  | 1.0645          | 0.9952          | 0.9945          | 0.9266           |
| Eq. (7d)                       | 1.0083          | 1.0083          | 1.0083          | 1.0083           |

\(^5\)These dimensions were selected because this is the size of region analyzed in \(*\)\(\ast\)\(\ast\)\(\ast\)\(\ast\) (2002a), who reports that the exponent for approximating the eigenvalues is \(\gamma = 0.99991\).

\(^6\)The coefficients were adjusted to satisfy the variance constraint, yielding
To summarize, the derivation of equations (5) and (6) reveals the role \( n \) plays in approximating the Jacobian term, which itself is governed by \( n \) because its logarithmic form is an arithmetic average. The derivation also reveals the role played by the extreme eigenvalues of a geographic weights matrix, eigenvalues that can be determined even for very large numbers of geographic units. Findings about quantities easily calculated for the geographic weights matrix (e.g., \( W \)) should be helpful in establishing better analytical statistical distribution properties for the autoregressive parameter \( \rho \). And the functional forms of the four coefficients appearing in equation (5) exhibit a converging trajectory for very large surface partitionings.

2.4. Alternatives to the Griffith-Sone Jacobian approximation

Several alternative approximations to the Gaussian normalizing constant appear in the recent literature. These approximations are reviewed as well as extended here, and in the next section they are used to estimate the autoregressive parameter \( \rho \), for comparative purposes.

Barry and Pace (1999) outline a Monte Carlo simulation method for estimating the normalizing constant. This procedure generates \( P \) n-by-1 pseudo-random variables \( x_p \) from a normal distribution with mean 0 and variance 1, and then computes cross-products between them and powers of the geographic weights matrix times them:

\[
J(\rho)^{1/2} \approx -\frac{\sum_{j=1}^{P} \sum_{k=1}^{m} x_p^T W^k x_p \cdot \rho^k}{P}.
\]

\[
\hat{\lambda}_{hk} = [0.389 \cos\left(\frac{h \pi}{3499}\right) + 0.389 \cos\left(\frac{k \pi}{2999}\right) + (1-2\times0.389) \cos\left(\frac{h \pi}{3499}\right) \cos\left(\frac{k \pi}{2999}\right)]^{0.974798}.
\]
Barry and Pace suggest setting $P$ to 20 and $m$ to 20. Analyses undertaken for this paper used $P = 30$ (based upon conventional minimum sample size considerations) and $m = 20$, and frequently encountered the drawing of samples rendering poor approximations. When the estimate of $\rho$ noticeably deviated from those estimates obtained using other approximations, new samples were drawn. Regardless, one of the deviant estimates obtained with equation (5) was obtained with this particular approximation. A second drawback of this approximation is its restricting values of $\rho$ to the interval $(-1, 1)$. Because the true lower bound of this interval is $1/\lambda_n$—which equals -1 in general only for regular square tessellations and selected surface partitionings involving triangulations, and approaches -2 in many irregular tessellation cases as well as for the hexagonal and maximally connected surface partitionings—the artificial restriction imposed by $(-1, 1)$ misrepresents the role of the Jacobian term in the estimation process.

Smirnov and Anselin (2001) employ a characteristic polynomial approach to approximating the Jacobian term, noting that

$$\det(I - \rho W) = \sum_{k=0}^{n} q_k \rho^k \approx \sum_{k=0}^{m} q_k \rho^k, \ m << n.$$ 

Two important features of the right-hand side truncated series are that (a) $q_0 = 1$, since the diagonal entries of the matrix are unity, and (b) $q_1 = 0$, since the sum of the eigenvalues of matrix $W$ are 0. In addition, the symmetry of the set of eigenvalues about zero for a regular square tessellation results in all odd values of $k$ having $q_k = 0$. The important implementation task left by Smirnov and Anselin (2001) is the determination and estimation of salient $q_k$. 
(2002c) shows that for regular square tessellations, only the coefficients $q_2$, $q_4$, and $q_{20}$ play a dominant role in determining $\det(I - \rho W)$. Exploration of the 148-surfaces database reveals that, for irregular tessellations, the coefficients $q_2$, $q_3$, $q_4$, and $q_5$ play a dominant role in determining $\det(I - \rho W)$, yielding

$$\det(I - \rho W) = e^{I(\rho)^{1/2}} \approx e^{J_3(\rho)^{1/2}} = 1 + q_2 \rho^2 + q_3 \rho^3 + q_4 \rho^4 + q_5 \rho^5 .$$

The following are estimator equations for these quantities:

$$\hat{q}_2 = 0.0411 - 0.0245 \ln(|\lambda_n| - 0.45) + 0.0276 \text{ISQ} ,$$

(9a)

$$\hat{q}_3 = 0.1894 - 0.2087 |\lambda_n| + 0.0193 \text{ISQ} ,$$

(9b)

$$\hat{q}_4 = 0.0257 + 2.5923 \left( \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}}{n} - 0.125 \right)^{1.5} + 0.0037 \text{ISQ} ,$$

(9c)

$$\hat{q}_5 = -0.0919 + 0.1645 |\lambda_n| - 0.0726 \text{ISQ} .$$

(9d)

$\text{ISQ}$ is a binary 0-1 indicator variable, taking on a value of 1 when results are for a regular square tessellation surface partitioning. Coefficient estimates $\hat{q}_3$ and $\hat{q}_5$ are 0 when $\text{ISQ} = 1$. The respective pseudo-$R^2$ values for these equations are 0.548, 0.532, 0.864, and 0.451. These four equations produced Shapiro-Wilk statistics that indicate their residuals deviate from normality. These estimators appear to improve as $n$ increases, with the Jacobian approximation RESS being
described by

\[ 0.0066 + 0.0991e^{-0.0548n}. \]

The pseudo-\( R^2 \) value for this equation is 0.839.

Pace and LeSage (2002) show that

\[
\frac{\rho^2}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{c_{ij}}{n_i n_j} \leq -\frac{1}{n} LN[det(V)] \leq \left[ \rho + LN(1- \rho) \right] \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{c_{ij}}{n_i n_j}. 
\]

These lower and upper bounds define an interval containing the Jacobian term, and suggest that an approximation for this term could be written as a linear combination of them:

\[
J(\rho)^{\frac{1}{2}} \approx J_4(\rho)^{\frac{1}{2}} = -\{ \alpha [\rho + LN(1- \rho)] - \beta \frac{\rho^2}{2} \} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{c_{ij}}{n_i n_j}. 
\]

The following are estimation equations for \( \alpha \) and \( \beta \):

\[
\hat{\alpha} = 0.1656, \quad \text{and} \quad \hat{\beta} = 0.6078 + 2.1889 \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}}{n_i n_j}. 
\]
The pseudo-R^2 value for this second equation is 0.583. The Shapiro-Wilk statistic calculated for its residuals indicates non-normality. These estimators also appear to improve as n increases, with the Jacobian approximation RESS being described by

\[ 0.0180 + 0.0827e^{-0.0333n}. \]

The pseudo-R^2 value for this equation is 0.637.

These various ways of approximating the Jacobian term allow very quick computations of spatial autoregressive parameter estimates, even for massively large georeferenced data sets. Based upon the RESS, assessments summarized in Table 2 reveal that these approximations work well for the square and hexagonal surface partitions, a mixture of which essentially characterizes most observed irregular surface partitionings. When the generalized coefficients are used, the best approximation for practical analyses appears to be furnished by equation (5); equation (10) provides the worst approximation across the range of surface partitionings. When a surface partitioning for some massively large georeferenced data set involves a regular square tessellation, such as a remotely sensed image, a spatial scientist may prefer to improve the precision of the Jacobian approximation by calibrating the coefficients of equation (5) using the approximate eigenvalues reported in Griffith (1999) or using the asymptotic square tessellation specific coefficients for equation (8) reported in ***** (2002c): \( \hat{q}_2 \approx 0.1174, \hat{q}_4 \approx 0.0742, \) and \( \hat{q}_{20} \approx 0.0552. \)
<table>
<thead>
<tr>
<th></th>
<th>linear</th>
<th>square</th>
<th>hexagon</th>
<th>maximally connected</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RESS eq. (5)</strong></td>
<td>0.0009</td>
<td>0.0001</td>
<td>0.00003</td>
<td>0.0051</td>
</tr>
<tr>
<td><strong>RESS eq. (7a)-&amp;7d)</strong></td>
<td>0.0844</td>
<td>0.0075</td>
<td>0.0457</td>
<td>0.0601</td>
</tr>
<tr>
<td>q2</td>
<td>0.0780</td>
<td>0.0932</td>
<td>0.0834</td>
<td>0.1044</td>
</tr>
<tr>
<td><strong>Eq. (9a)</strong></td>
<td>0.0833</td>
<td>0.0833</td>
<td>0.0924</td>
<td>0.1145</td>
</tr>
<tr>
<td>q3</td>
<td>0</td>
<td>0</td>
<td>0.0400</td>
<td>0.0306</td>
</tr>
<tr>
<td><strong>Eq. (9b)</strong></td>
<td>0</td>
<td>0</td>
<td>0.0698</td>
<td>0.0851</td>
</tr>
<tr>
<td>q4</td>
<td>0.5636</td>
<td>0.1266</td>
<td>0.0340</td>
<td>-0.0461</td>
</tr>
<tr>
<td><strong>Eq. (9c)</strong></td>
<td>0.6247</td>
<td>0.1440</td>
<td>0.0477</td>
<td>0.0257</td>
</tr>
<tr>
<td>q5</td>
<td>0</td>
<td>0</td>
<td>-0.0019</td>
<td>-0.0397</td>
</tr>
<tr>
<td><strong>Eq. (9d)</strong></td>
<td>0</td>
<td>0</td>
<td>0.0024</td>
<td>-0.0097</td>
</tr>
<tr>
<td><strong>RESS eq. (8)</strong></td>
<td>0.0264</td>
<td>0.0040</td>
<td>0.0036</td>
<td>0.0102</td>
</tr>
<tr>
<td><strong>RESS (9a)-(9d)</strong></td>
<td>0.0390</td>
<td>0.0050</td>
<td>0.0666</td>
<td>0.1309</td>
</tr>
<tr>
<td><strong>α</strong></td>
<td>0.0802</td>
<td>0.0180</td>
<td>0.0368</td>
<td>-0.0303</td>
</tr>
<tr>
<td><strong>Eq. (11a)</strong></td>
<td>0.1656</td>
<td>0.1656</td>
<td>0.1656</td>
<td>0.1656</td>
</tr>
<tr>
<td><strong>β</strong></td>
<td>1.4501</td>
<td>0.6485</td>
<td>0.3605</td>
<td>0.4196</td>
</tr>
<tr>
<td><strong>Eq. (11b)</strong></td>
<td>1.7023</td>
<td>1.1550</td>
<td>0.9726</td>
<td>0.8814</td>
</tr>
<tr>
<td><strong>RESS eq. (10)</strong></td>
<td>0.0711</td>
<td>0.0272</td>
<td>0.0217</td>
<td>0.0754</td>
</tr>
<tr>
<td><strong>RESS (11a)-(11b)</strong></td>
<td>0.2195</td>
<td>0.1364</td>
<td>0.0320</td>
<td>0.3191</td>
</tr>
</tbody>
</table>
3. Implementation of the Jacobian approximation

A Jacobian approximation can be implemented with various degrees of knowledge about the eigenvalues of a geographic connectivity or weights matrix. The simplest but most numerically intensive situation is when all of the eigenvalues are known. A second situation occurs when the eigenvalues are approximately rather than exactly known. A third case occurs when virtually nothing is known about the eigenvalues beyond \( \lambda_1 \). Ten examples were selected from the 148 surface partitionings database, involving \( n \) ranging from 35 to 3,888, to construct Figures 1 and 3 and to serve as illustrative examples in this section.

3.1. The Jacobian approximation when all of the eigenvalues are known

This case occurs when \( n \) is small enough that all of the eigenvalues can be calculated numerically, or in special cases for which eigenvalues are theoretically known. The eigenvalues are known for a regular square tessellation when equation (1) employs matrix \( C \). In general, the eigenvalues of matrix \( W \) can be computed for geographic partitionings that do not have too large of \( n \) (i.e., several thousand with today's computer technology), and can be very accurately approximated for a regular square tessellation (Griffith, 1999). But as \( n \) approaches 10,000+ for irregular tessellations, supercomputing resources, which spatial social scientists tend to shy away from (Kiernan, 1998), are required.

Table 3 contains selected descriptive statistics for the RESSs calculated for the 148 surface partitionings in the database being explored here. Only equations (5) and (6) are calibrated using the entire set of eigenvalues. They furnish extremely good descriptions of the
The Jacobian term across the feasible parameter space, approximations that are superior to those furnished by equations (8) and (10). In addition, equation (5) approximates the Jacobian term marginally better than equation (6) does. To summarize, when all of the eigenvalues are known, the Jacobian term approximation should tend to deviate from the actual Jacobian term by about 0.2%, on average, with a worse case scenario of about 1%. Major advantages in using a Jacobian approximation here are a reduction in computational requirements of repeated calculations of the Jacobian term during each nonlinear least squares iteration, and an improved ability to establish analytical properties of $\hat{\rho}$. The approximation always is a relatively simple expression, whereas the Jacobian term itself involves a polynomial of degree $n$.

<table>
<thead>
<tr>
<th>Equation</th>
<th>Mean RESS</th>
<th>Standard deviation RESS</th>
<th>Minimum RESS</th>
<th>Maximum RESS</th>
<th>Mean squared error for selected $\rho$ estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation (5)</td>
<td>0.00172</td>
<td>0.00245</td>
<td>0.00000</td>
<td>0.01015</td>
<td>0.00003</td>
</tr>
<tr>
<td>Equation (6)</td>
<td>0.00376</td>
<td>0.00285</td>
<td>0.00020</td>
<td>0.01182</td>
<td>0.00007</td>
</tr>
<tr>
<td>Equation (8)</td>
<td>0.01919</td>
<td>0.01757</td>
<td>0.00240</td>
<td>0.09133</td>
<td>0.00041</td>
</tr>
<tr>
<td>Equation (10)</td>
<td>0.03585</td>
<td>0.02359</td>
<td>0.01026</td>
<td>0.11197</td>
<td>0.00058</td>
</tr>
</tbody>
</table>

To illustrate estimation of the autoregressive parameter $\rho$, results for the ten selected database examples are reported in Table 3 and portrayed in Figure 3a. For each example the georeferenced quantities were subjected to a Box-Cox power transformation to have the data frequency distributions better conform to a normal curve. Constant mean (i.e., $X\beta = \mu I$) spatial autoregressive models were estimated using these data, with the sole purpose of comparing parameter estimates; certainly better statistical descriptions of these data are possible.
Estimation results reported in Table 3 reveal that any of these Jacobian approximations render very good parameter estimates, equation (5) furnishing the best estimates. The plots appearing in Figure 3a also show that equations (5) and (6) align best with the actual values of \( \rho \), that equation (8) goes from over- to underestimation as \( \rho \) increases, and that equation (1) goes from under- to overestimation. The mean squared error (MSE) for their average is 0.00012, which is a considerable improvement over the MSE for each estimator alone but is still inferior to the MSE obtained with equation (5). Little seems to be gained by computing the n eigenvalues so that the exact Jacobian term can be employed for parameter estimation.

3.2. The Jacobian approximation when the n-1 nonprincipal eigenvalues are unknown but can be approximated

Matrix \( W \) rather than \( C \) often is employed to specify the inverse covariance matrix \( V \) in equation (1). The analytical eigenvalues for matrix \( W \), however, are not known, except for a linear surface partitioning. Griffith (1999) proposes very good approximations for the eigenvalues of both a regular square and a regular hexagonal surface partitioning, few of which are actually known analytically. An approximation for the maximally connected case appears in §2.3. The approximations by Griffith (1999) were used to calculate the results reported in Tables 1 and 2. They exploit the following properties: (1) the maximum eigenvalue is 1; (2) the sum of the eigenvalues is 0; (3) the sum of the squared eigenvalues is \( \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{c_{ij}}{n_i n_j} \); (4) eigenvalues for the linear geographic landscape are \( \cos(\frac{k \pi}{n-1}) \), \( k = 0, 1, \ldots, n-1 \); and, (5) the minimum eigenvalue either is known or can be calculated or estimated. Results for actual and approximated eigenvalues for relatively large geographic surfaces are summarized in Table 4, with those
obtained using the approximate eigenvalues comparing very favorably with those obtained using the actual eigenvalues. For example, estimation results reported in Table 4 indicate that equations (8) and (10) should perform as well, or nearly as well, as equation (5). The regular square tessellation empirical examples (e.g., Montana agricultural field plot and High Peak remotely sensed data) are in the upper right-hand corner of Figure 3a. These clusters of points exhibit considerably less dispersion than the other eight clusters appearing on the graph; for all practical purposes, the estimates from equations (8) and (10) estimates are the same as those from equations (5) and (6).

<table>
<thead>
<tr>
<th>coefficients</th>
<th>75-by-75 square</th>
<th>75-by-75 hexagon</th>
<th>n = 4500 maximally connected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>actual</td>
<td>approximate</td>
<td>actual</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.37889</td>
<td>0.37887</td>
<td>0.56706</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.37889</td>
<td>0.37887</td>
<td>0.32199</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.99750</td>
<td>0.99745</td>
<td>0.99537</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.99750</td>
<td>0.99745</td>
<td>0.99772</td>
</tr>
<tr>
<td>RESS</td>
<td>0.00004</td>
<td>0.00004</td>
<td>0.00004</td>
</tr>
<tr>
<td>$q_2$</td>
<td>0.09300</td>
<td>0.09304</td>
<td>0.08449</td>
</tr>
<tr>
<td>$q_3$</td>
<td>0</td>
<td>0</td>
<td>0.04149</td>
</tr>
<tr>
<td>$q_4$</td>
<td>0.13316</td>
<td>0.13302</td>
<td>0.03664</td>
</tr>
<tr>
<td>$q_5$</td>
<td>0</td>
<td>0</td>
<td>-0.00111</td>
</tr>
<tr>
<td>RESS</td>
<td>0.00452</td>
<td>0.00451</td>
<td>0.00383</td>
</tr>
</tbody>
</table>
Findings for these approximations are of interest beyond benchmark comparisons. The approximation for a square tessellation can be employed in real-world remotely sensed data analyses. The approximation for a hexagonal tessellation helps analyze georeferenced data collected through projects motivated by the now defunct US EPA. Environmental Monitoring and Assessment Program (EMAP). The national EMAP coverage involves nearly 13,000 hexagons that have been superimposed upon the coterminous US.

3.3. The Jacobian approximation when the n-1 nonprincipal eigenvalues are unknown and lack a known approximation

A set of eigenvalues furnishes important information about its associated surface partitioning and in turn about the underlying geographic configuration giving rise to spatial autocorrelation. Most irregular surface partitionings defy analytical or approximate eigenvalue determination, but the extreme eigenvalues always can be either computed or estimated. While the coefficients of equation (5) can be estimated with equations (7a)-(7d) for any surface partitioning with a very large number of areal units, equations (8) and (10) do not require knowledge of the eigenvalues beyond $\lambda_1 = 1$; however, the coefficients of the latter two capture summary information about the eigenvalues. In addition, Barry and Pace (1999) outline a simulation method that also does not require explicit knowledge of the eigenvalues.
The basic information knowable about any set of eigenvalues includes the properties (1), (2), (3), and (5) outlined at the beginning of §3.2. Equations (7a)-(7c), (9a)-(9d), and (11a)-(11b) utilize these quantities to operationalize equations (5), (8) and (10) without calculating the full set of eigenvalues. Better estimates of these coefficients can be obtained by calibrating equations (5), (8) and (10) to the set of approximated eigenvalues for a regular square or hexagonal tessellation. In addition, ***** (2002c) reports better coefficient estimators for equations (9a)-(9d), supplementing these coefficients with q20. The improved estimators for the square tessellation case helps prevent the MLE of \( \hat{\rho} \) from becoming the upper bound of the feasible parameter space strictly due to specification error introduced by the estimator. For example, a biomass measure for the High Peak remotely sensed image data set produces \( \hat{\rho} = 1 \), whereas the special version reported in ***** (2002c) produces \( \hat{\rho} = 0.9920 \), a value much closer to the actual value of 0.9886. This parameter space boundary problem may be common for remotely sensed data, which tend to contain extremely high levels of positive spatial autocorrelation; the latest high resolution, hyperspectral data probably will display even higher spatial autocorrelation levels.

Improved estimators for equations (11a)-(11b) also are available for regular square tessellations:

\[
\hat{\alpha} = -0.09 + 0.51 \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{c_{ij}}{n_in_j}; \quad \text{and} \\
\hat{\beta} = 1.00 + 1.17 \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{c_{ij}}{n_in_j}. \tag{12a}
\]

\[
\hat{\beta} = 1.00 + 1.17 \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{c_{ij}}{n_in_j}. \tag{12b}
\]
For the High Peak data example, these estimators improve the MLE of $\rho$ over that obtained using equations (11a)-(11b), reducing it from 0.9909 to 0.9889, which again is closer to the actual value of 0.9886. In addition, these special estimators for remotely sensed data suggest that the asymptotic coefficients for a regular square tessellation are $\hat{\alpha} \sim 0.0375$ and $\hat{\beta} \sim 1.2925$, which differ from the general asymptotic values of 0.1656 and 1.1550 suggested by equations (11a)-(11b). In other words, paying particular attention to the special features of a regular square tessellation furnishes operational knowledge about its eigenvalues.

The Barry and Pace (1999) Monte Carlo simulation method for approximating the log-Jacobian term through the traces of powers of matrix $W$ requires no knowledge of the eigenvalues. One of its major drawbacks, noted earlier, is the possibility of obtaining an inferior MLE of $\rho$ solely because a poor random sample is drawn. The poorest estimates appearing in Figures 3a-3c are those found with this method. Summary assessment results for the approximations based upon this simulation method appear in Table 5. Comparing them with results reported in Table 3 further illustrates the inferiority of this simulation method. Although the MSEs are greater here, the average increase in the deviation of $\hat{\rho}$ from $\rho$ is of little practical consequence.

<table>
<thead>
<tr>
<th>Table 5. Goodness-of-fit assessment of selected database Jacobian term approximations</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean RESS</td>
</tr>
<tr>
<td>calibrated with the Barry-Pace simulated Jacobian term</td>
</tr>
<tr>
<td>equation (5)</td>
</tr>
</tbody>
</table>

25
A second drawback of the Barry-Pace simulation method is its truncation of the feasible parameter space at -1. Irregular surface partitionings can have a substantial gap between $1/\lambda_n$ and -1 (see Figure 1). Smirnov and Anselin (2001) note that in practice this truncation is not a serious problem since negative, let alone markedly negative, spatial autocorrelation is rarely observed. The log-Jacobian term bounds established by Barry and Pace are based upon a power series expansion of $\ln(I - \rho W)$. Convergence of this power series is governed by $\rho \lambda_1$ and $\rho \lambda_n$, requiring $1/\lambda_n < \rho < 1$. But Barry and Pace separate $\rho$ and $\lambda_n$, rendering their requirement of $-1 < \rho$ to attain convergence. Calibration results summarized in Table 5 illustrate that this truncation is more problematic than any of these researchers suggest, but only for calibrating a Jacobian approximation. Overall the estimates of $\rho$ still are very good (see Figure 3c).

4. Implications for standard error estimates

The Jacobian approximation proposed here obtains parameter estimates that essentially are
indistinguishable from their exact Jacobian term counterparts. It also yields some reliable standard error estimates. First consider the regression coefficient estimates for the more general nonconstant mean case, which are given by

\[ b = (X^\top \hat{\hat{\V}} X)^{-1} X^\top \hat{\hat{\V}} Y, \]  

(13)

where \( \hat{\hat{\V}} \) is an estimate because it uses \( \hat{\rho} \). Substituting the MLE of \( \sigma \) into equation (1) reveals that the more general spatial autoregression problem is equivalent to a standard regression problem of the form

\[ \frac{Y}{e^{-J(\rho)^{1/2}}} = \rho W \frac{Y}{e^{-J(\rho)^{1/2}}} + (I - \rho W) \frac{X}{e^{-J(\rho)^{1/2}}} + \rho W \frac{\varepsilon}{e^{-J(\rho)^{1/2}}}, \]

(14)

for the SAR specification. An AR specification replaces matrix \((I - \rho W)\), premultiplying matrix \(X\) in equation (14) with \(I\); a CAR specification is slightly more complicated to describe, but follows the same principle. In other words, one can estimate the parameters of a spatial autoregressive model simply by sequentially substituting a systematic sample of values of \( \rho \) from across the feasible parameter space\(^7\) \( \left( \frac{1}{\lambda_n}, 1 \right) \) performing an ordinary least squares (OLS) regression with equation (14) after each substitution, and then equating \( \hat{\rho} \) to that value of \( \rho \) achieving the minimum MSE.\(^8\)

A comparison of equations (13) and (14) indicates that the Jacobian term does not

\(^7\)In most empirical cases spatial autocorrelation is positive (verified by computing the Moran Coefficient), and accordingly the search could be restricted to the narrower interval \((0, 1)\).
directly impact upon the regression parameter estimates. Because $e^{-J(\rho)\frac{1}{2}}$ is a multiplicative constant that would appear twice in $(X^T \hat{V} X)^{-1}$ and twice in $X^T \hat{V} Y$, it would divide out of the estimator. Replacing $J(\hat{\rho})$ with $\hat{J}(\hat{\rho})$ does not change this situation; however, it results in a small deviation in the estimate $\hat{\rho}$, thus slightly altering the estimates but indirectly through $\hat{V}$ (as noted in the empirical analyses reported in this paper). The same is true for the standard error term for the $j^{th}$ predictor variable in a regression analysis, denoted by $s_{bj}$. Therefore, the proposed Jacobian approximation has little effect on either the regression coefficient estimates or their accompanying standard error estimates. Consequently, once $\hat{\rho}$ is obtained, a researcher can regress $(I - \hat{\rho} W)Y$ on $(I - \hat{\rho} W)X$ using the OLS technique to calculate and/or verify the regression coefficients and their corresponding standard errors.

The asymptotic standard error of $\hat{\rho}$ is given by entering $-E[\frac{\partial^2 LN(L)}{\partial \rho^2}]$ into the inverse information matrix, and does not fair as well as $s_{bj}$:

$$-E[\frac{\partial^2 LN(L)}{\partial \rho^2}] = n \frac{\partial^2 J(\rho)}{\partial \rho^2} + \text{TR} \{[W(I - \rho W)^{-1}]^T [W(I - \rho W)^{-1}] \}/\sigma^2,$$  \hspace{1cm} (15)

where $L$ denotes the likelihood function form of equation (1), $E$ the calculus of expectations, and $\text{TR}$ the matrix trace operator. The numerator of the second term in the right-hand side of expression (15) can be rewritten as

---

8 As an aside, a nonlinear regression routine automates this very procedure.
9 The inverse information matrix from which a measure of the variation of the MLE can be obtained is presented in, among others, Ord (1975) for the SAR model, Upton and Fingleton (1985) for the AR model, and Griffith (1988) for the CAR model.
\[ \text{TR}\{D^{1/2}(I - \rho D^{1/2}CD^{1/2})^{-1}[D^{1/2}CD^{1/2}]D^{-1}[D^{1/2}CD^{1/2}](I - \rho D^{1/2}CD^{1/2})^{-1}D^{1/2}\} . \quad (16) \]

This form of the expression reveals that when each areal unit has the same number of nearby areal units (e.g., and infinite square or hexagonal tessellation), say \(k\), then the diagonal matrix \(D\) becomes \(\frac{1}{k}I\), resulting in \(D^{-1}\) and the two \(D^{1/2}\) matrices dividing out of the expression. This reduction results in expression (16) equating to \(n\frac{\partial^2 J(\rho)}{\partial \rho^2} = \sum_{i=1}^{n} \frac{\lambda_i^2}{i(1 - \rho \lambda_i)^2}\). When \(\rho = 0\), the first term in the right-hand side of expression (15) equals \(\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} \), while the second term in the right-hand side equals \(W^TW = \sum_{i=1}^{n} \frac{1}{n_i}\). Thus, a reasonable approximation for expression (16) is

\[ \frac{\sum_{i=1}^{n} \frac{1}{n_i} \frac{\partial^2 J(\rho)}{\partial \rho^2}}{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}} = \frac{\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n} \frac{\lambda_i^2}{1(1 - \rho \lambda_i)^2}}{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}} . \quad (17) \]

Increasing the size of the surface upon which a regular square or hexagonal tessellation is superimposed results in expression (17) converging to \(n\frac{\partial^2 J(\rho)}{\partial \rho^2}; D^{-1}\) and the two \(D^{1/2}\) matrices in expression (16) asymptotically reduce to the identity matrix. For the Murray superfund site database example, expression (16) equals 101.3280 whereas expression (17) equals \(\frac{47.0194}{44.4722} \times 95.9559 = 101.4542\); for the Chicago census tracts example, expression (16) equals \(\frac{47.0194}{44.4722} \times 95.9559 = 101.4542\); for the Chicago census tracts example, expression (16) equals \(\frac{47.0194}{44.4722} \times 95.9559 = 101.4542\).
2329.7246 whereas expression (17) equals \( \frac{373.9392}{349.5418} \times 2192.3121 = 2345.3314 \); for the Montana agricultural field plots example, expression (16) equals 22482.6845 whereas expression (17) equals \( \frac{996.3333}{994.1663} \times 22385.9402 = 22434.7367 \). In each of these database examples the approximate value differs from the actual value by less than 1%.

Variation of the MLE \( \hat{\rho} \) is indirectly a function of the variance estimate \( \hat{\sigma} \), a parameter estimate with which it covaries, for both the SAR and CAR models, and with both \( \hat{\sigma} \) and \( b \) for the AR model. Thus, this asymptotic variance is directly a function of \( \frac{\partial^2 J(\rho)}{\partial \rho^2} \) and indirectly a function of \( \frac{\partial J(\rho)}{\partial \rho} \). A question remains to be answered: how do these quantities compare with their respective \( \frac{\partial^2 J_1(\rho)}{\partial \rho^2} \) and \( \frac{\partial J_1(\rho)}{\partial \rho} \) counterparts? In other words, how does the negligible specification error latent in equation (5) propagate?

The answer is alluded to upon inspection of the three database examples just discussed for which \( \hat{\rho} \) has been calculated: the Murray superfund site, the Chicago census tracts, and the Montana agricultural field plots georeferenced data sets. Their numerical results for the Jacobian term and its first and second derivatives, calculated using \( \hat{\rho} \) and reported in Table 6, suggest that \( \hat{J}_1(\rho) \) and \( \frac{\partial^2 J_1(\rho)}{\partial \rho^2} \) yield reasonably accurate approximation values, but \( \frac{\partial \hat{J}_1(\rho)}{\partial \rho} \), which is contained in off-diagonal cells of the information matrix, can yield a rather poor approximation.
Table 6. Quantities associated with the computation of standard errors.

<table>
<thead>
<tr>
<th>geographic landscape</th>
<th>(-\sum_{i=1}^{n} LN(1 - \rho \lambda_i))</th>
<th>(\sum_{i=1}^{n} \frac{\lambda_i}{1 - \rho \lambda_i})</th>
<th>(\sum_{i=1}^{n} \frac{\lambda_i^2}{(1 - \rho \lambda_i)^2})</th>
<th>(n \hat{J}_1(\rho))</th>
<th>(n \frac{\partial \hat{J}_1(\rho)}{\partial \rho})</th>
<th>(n \frac{\partial^2 \hat{J}_1(\rho)}{\partial \rho^2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murray</td>
<td>7.51</td>
<td>32.44</td>
<td>94.30</td>
<td>7.18</td>
<td>26.27</td>
<td>93.88</td>
</tr>
<tr>
<td>Chicago</td>
<td>169.42</td>
<td>587.04</td>
<td>2192.31</td>
<td>166.45</td>
<td>422.97</td>
<td>2208.73</td>
</tr>
<tr>
<td>Montana</td>
<td>682.86</td>
<td>2757.84</td>
<td>22385.94</td>
<td>683.01</td>
<td>2771.50</td>
<td>21753.82</td>
</tr>
</tbody>
</table>

Results obtained with equations (7a)-(7d), however, are less encouraging, often rendering substantial deviations from values like those appearing in Table 6. Except for the special version pertaining to a regular square tessellation, the \(J_3(\rho)\) and \(J_4(\rho)\) Jacobian approximations fail to furnish accurate first and second derivative results. Summary comparative results appearing in Table 7 for the Montana example illustrate this point. Even the generalized coefficients furnished by equations (7a)-(7d) outperform the \(\hat{J}_3(\rho)\) and \(\hat{J}_4(\rho)\) approximations. The result obtained with the approximate eigenvalues is extremely good; using this result, the coefficient for expression (17) is \(\frac{996.3333}{994.1667} \times 22443 = 22492\), which is roughly 0.25% greater than the result obtained using the exact eigenvalues and roughly 3% greater than its counterpart obtained with the result appearing in Table 7.

Table 7. Quantities associated with the computation of standard errors for the example Montana data (\(\hat{\rho} = 0.9477\)).

<table>
<thead>
<tr>
<th></th>
<th>(J_1(\rho))</th>
<th>(J_3(\rho))</th>
<th>(J_4(\rho))</th>
</tr>
</thead>
</table>

31
5. Discussion and future directions

Four Jacobian approximations are evaluated in this paper. As Figures 3a, 3b, and 3c indicate, by employing any of these approximations, virtually any size of georeferenced data set can now be adequately described with a spatial autoregressive model. For example, \( \hat{\rho} \) was estimated for the 45,994 enumeration areas of Canada, with generalized coefficient Jacobian approximations rendering values ranging from 0.8380 to 0.8691; \( \hat{\rho} \) also was estimated for the 224,753 1990 US block groups, rendering values ranging from 0.9068 to 0.9255. The approach also can be extended to space-time model specifications, as shown in Griffith (1996). The single overarching advantage of these approximations making them preferable to employing the actual Jacobian term, is that they support more direct data analysis by always substituting a relatively simple function for an \( n^{th} \)-order polynomial.
Findings reported here indicate that equation (5) should be the preferred approximation, because it produces superior results for both $\hat{\varrho}$ and the standard error of $\hat{\varrho}$. Future research needs to focus on modifying equations (7a)-(7c) to improve $\hat{\varrho}$ and $\frac{\partial^2 \hat{J}_1(\varrho)}{\partial \varrho^2}$ obtained with them. But the measures to achieve this improvement must be easily and quickly calculable. The Barry-Pace simulation procedure is deemed to be the most problematic approximation method because poorer results are possible strictly due to the chance of generating poor random numbers.

Equation (8) illustrates that improved results for special surface partitionings may be obtainable. Replacing equations (9a)-(9d) with findings reported in ***** (2002c) substantially improves $\hat{J}_3(\varrho)$ results for the Montana database example (see Table 7). Another way to improve results would be to establish a good eigenvalue approximator. Again, comparative results reported in Table 7 illustrate this in the Montana database example. Although the eigenvalue approximator for a regular hexagonal surface partitioning is quite good, approximation of its negative eigenvalues needs improvement.

Repeatedly, observed irregular surface partitionings seem to be a mixture of square, hexagonal, and maximal connectivity surface partitionings. Comparisons in Tables 1, 2 and 4, for example, establish benchmark and limiting partitionings results. The square and hexagonal tessellation cases furnish insights about observed partitionings. Future research is needed that establishes the mixture of these two partitionings related to an observed irregular surface partitioning. The linear and maximally connected tessellation cases respectively furnish a lower and an upper bound for observed partitionings.
Finally, findings reported here need to be extended to higher-order models. The extremely high levels of spatial autocorrelation exhibited by high-resolution hyperspectral remotely sensed data, for instance, demand this. If the spatial autocorrelation specifications relate to the SAR model studied here, then the associated Jacobian approximations will not change.

6. References


Figure 1. Ten specimen Jacobian term plots from the 148 surface partitionings data set. Asterisk denotes Thiessen polygons for Huancavelica (Peru); plus denotes 1990 Chicago census tracts; solid circle denotes High Peak remotely sensed pixels; diamond denotes kreisunits for a major part of Germany; square denotes kreisunits for a small, disjoint part of Germany; cross denotes counties of the former German Democratic Republic; triangle denotes 1980 Houston census tracts; left triangle denotes Thiessen polygons for Murray superfund site; right triangle denotes Montana agricultural field plots; and, upside down triangle denotes US counties.
Figure 2a. Dot portrays the Jacobian term for a traditional bivariate situation, with r varying between ±1. Asterisk portrays the Hamilton-Watts quadratic approximation: \(0.6355 + 0r + 2.7130r^2\).
Figure 2b. Dot portrays the Jacobian term for a linear geographic landscape. Asterisk portrays the equation (3) approximation: 

$$0.2527[2 \ln(1.0537) - \ln(1.0537 - \rho) - \ln(1.0537 + \rho)] .$$
Figure 3. Top: (a) estimates calculated with the full parameter space. Middle: (b) estimates calculated with simulated data and the parameter space truncated at -1. Bottom: (c) estimates calculated with the parameter space truncated at -1. Solid dot denotes equation (5) estimates; asterisk denotes equation (6) estimates; solid square
denotes equation (8) estimates; solid diamond denotes equation (1) estimates; dotted line denotes equality points.