Positive spatial autocorrelation impacts on attribute variable frequency distributions

Daniel Griffith[*]

School of Economic, Political, and Policy Sciences, University of Texas at Dallas, Dallas, USA

Abstract

Researchers commonly inspect histograms as a first step in data analysis, often finding that these graphs fail to closely align with any of the several hundred ideal frequency distributions. The purpose of this paper is to address how positive spatial autocorrelation –the most frequently encountered in practice– can distort histograms when they are constructed with georeferenced data. Normal, Poisson, and binomial random variables –three widely applicable ones– are studied after establishing appropriate moment generating functions, and are illustrated with selected simulations. The simulations were designed with an ideal surface partitioning, and with the irregular China county surface partitioning. Results show that even moderate levels of positive spatial autocorrelation, while not affecting means, not only inflate variance, but also modify the probabilities of extreme and/or central values, and can alter skewness and kurtosis. A methodology is outlined for recovering the underlying unautocorrelated frequency distributions.

Keywords: Beta random variable · Binomial random variable · Histogram · Normal random variable · Poisson random variable · spatial autocorrelation

Mathematics Subject Classification: Primary 62F99 · Secondary 60E99.

1. Introduction

Often statistical as well as other quantitative geospatial data analysis begins with an inspection of attribute variable histograms. Most spatial statistical research to date addresses impacts of spatial autocorrelation (SA) on parameter estimates, with the general conclusion that positive SA tends to have little or no impact on first moment types of parameter estimates, while inflating their respective standard errors (i.e., impacts are on variance estimates). This tendency implies that as SA in a random variable (RV) increases, its tails should become heavier and its center should become flatter. Dutilleul and Legendre (1992) appear to be the only researchers to systematically investigate this topic, although they do so in a rather artificial geographic context and only for a normal RV.

---

[*]Corresponding author. Daniel A. Griffith. School of Economic, Political and Policy Sciences University of Texas at Dallas 800 W. Campbell Rd Richardson, Texas 75080-3021. Email: dagriffith@utdallas.edu

Positive SA is widely acknowledged as a source of variance inflation for normal RVs, and a source of overdispersion (i.e., excess variance) for Poisson and binomial RVs. But how does this increased variation impact upon a variable's histogram? The purpose of this paper is to address this question. Intuitively speaking, variance increases as increasingly extreme values (i.e., outliers) appear in a histogram. SA-generated heavy tails in a normal distribution are consistent with this data feature. But although a Poisson RV can have extreme large counts, its extreme small counts can only become excessive numbers of zeroes. Meanwhile, a binomial RV cannot have radically extreme values, because its counts are constrained to be in the closed interval $[0, N_{tr}]$, where N_{tr} denotes the number of trials in the experiment under consideration. In other words, is some of the quite bothersome noise in or potential dirtiness of data geospatial researchers routinely encounter simply a manifestation of SA?

The paper is organized as follows. Section 1 discusses about the effect of the spatial autocorrelation in the specification of model. Section 2 presents a standard benchmark based on normal curve theory. Section 3 provides a heuristic overview of the beta distribution. Sections 4 and 5 introduce results about the Poisson and Binomial distributions. Section 6 is devoted to methodological aspects for spatial scientists. Finally, some discussion and implications are presented in Section 7.

2. Accounting for SA in a RV Model Specification

Besag (1974) introduced and summarized the notion of auto-models for spatial data. His specifications contain a response variable, Y_i, on the left-hand side of an equation for location (i.e., observation) i, and some linear combination of Y_{js} $(i \neq j)$ on the right-hand side of the same equation for the $n-1$ other locations j. This is the model specification addressed by Cliff and Ord (1973) and Ord (1975) for Gaussian RVs; their early work more fully develops the auto-normal model. Weaknesses of this general auto-model specification include: an auto-Poisson and auto-negative binomial probability model can accommodate only negative SA; because of the intractability of normalizing constants, non-normal RV parameter estimation requires Markov chain Monte Carlo (MCMC) techniques; and, attention has been restricted to the natural exponential family of statistical distributions (e.g., auto-normal, auto-logistic, auto-binomial, auto-Poisson, and auto-negative binomial).

Because Besag's (1974) auto-Poisson and auto-negative binomial models, for example, suffer from an inability to account for positive SA, by far the most common nature of SA to manifest in empirical data, he and his collaborators (see, e.g., Besag et al., 1991; Besag and Kooperberg, 1995; Besag et al., 1995) introduced a random effects term into a hierarchical Bayesian model specification to compensate for this drawback. This conceptualization represents SA as a feature of model parameters, rather than correlated response variable values. It casts a model intercept as an observation-specific surrogate for unobserved variables by expressing it as a random deviation from some global intercept. For georeferenced data, a random effects term comprises two components: a spatially structured component accounting for spatial dependence, and a spatially unstructured component accounting for overdispersion. Software such as WinBUGS implements this specification with a conditional autoregressive (CAR) normal prior distribution. The intrinsic CAR (ICAR) version, which sets the autoregressive parameter to its maximum value and is an improper prior but a limiting case of the CAR version, is the most common specification in practice. A convolution prior distribution can be constructed for a random intercept by summing a spatially structured and a spatially unstructured parameter. The set of exchangeable prior distributions for this spatially unstructured intercept almost always are assumed to be univariate normal with mean zero (0). The formulation extends to a frequentist case by analyzing repeated measures: the random effects term becomes an observation-specific constant across

repeated measures. A spatial filter can be used to capture the spatially structured random effects, removing the estimation necessity to include repeated measures. This type of model specification posits that empirical probabilities are correct, while simple model parameters are not. In contrast, an auto-model posits that simple model parameters are correct, while empirical probabilities are conditional on other observations. Consequently, in the development of Bayesian mapping analysis (see, e.g., GeoBUGS in WinBUGS), this direct dependency between values of a response is replaced by the incorporation of SA into prior parameter distributions; see Clayton and Kaldor (1987) and Besag et al. (1991).

A modified version of this latter approach is adopted in this paper; i.e., parameters in RV probability model specifications are modified to account for SA. In other words, "[t]he data model is one of conditionally independent [RVs], conditional on parameters that are distributed [over some geographic landscape] according to a spatial process"; see Kaiser et al. (2002, p. 450). Accounting for SA in this way allows individual observations to be treated as being independent.

2.1 Eigenvector spatial filtering

The eigenvector spatial filtering approach is a methodology that accounts for SA in RVs by incorporating heterogeneity into parameters in order to model non-homogeneous populations. It renders a mixture of distributions that can be used to model observed georeferenced data whose various characteristics differ from those that are consistent with a single, simple underlying distribution with constant parameters across all observations. The aim of this technique is to capture SA effects with a set of spatial proxy variables-namely, eigenvectors-rather than to identify a global SA parameter governing average direct pairwise correlations between selected observed values. As such, it utilizes the misspecification interpretation of SA, which assumes that SA is induced by missing exogenous variables, which themselves are spatially autocorrelated, and hence relates to heterogeneity.

Paralleling the Hammersley-Clifford theorem (see Besag, 1974), SA exists between subsets of neighbors in a set of n locations: a location j is said to be a geographic neighbor of location i if and only if the specification of the functional form of a probability density/mass (pdf/pmf) function for location i is dependent upon location j. In eigenvector spatial filtering, this dependency is captured by a common factor (see, e.g., Burridge, 1980, 1981) that is a linear combination of synthetic variates summarizing distinct features of the neighbors' geographic configuration structure for a given georeferenced data set. The synthetic variates may be the eigenvectors of the matrix

$$\left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{n}\right) \mathbf{C} \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{n}\right), \tag{1}$$

the matrix appearing in the numerator of the Moran Coefficient (MC) index of SA[1], where $\mathbf{C}$ often is an n-by-n binary 0-1 geographic connectivity matrix (i.e., $c_{ij} = 1$ when locations i and j are neighbors, and zero (0) otherwise[2]), $\mathbf{1}$ is an n-by-1 vector of ones, and superscript T denotes matrix transpose. Cliff and Ord (1973, p. 34) and Clifford et al. (1989, p. 47) used this matrix to construct the moment generating function (MGF) of the MC under an assumption of normality. de Jong et al. (1984) showed that the extreme eigenfunctions of this matrix define the most extreme levels possible of SA for a given surface partitioning, a result in combination with Tiefelsdorf and Boots (1995)

---

[1]The Geary ratio counterpart to matrix the Equation (1) also could be used.
[2]An eigenfunction spatial filtering comparison of this topological-based definition of geographic neighbor with one employing inter-point distances appears in Griffith and Peres-Neto (2006).

and Griffith (1996) that attaches conceptual meaning to the extracted synthetic variates. These variates summarize distinct map pattern features because they are both orthogonal and uncorrelated; see Griffith (2000a).

The eigenfunction problem solution is similar to that obtained with principal components analysis in which a covariance matrix is given by $[\mathbf{I}+k\,(\mathbf{I}-\mathbf{1}\,\mathbf{1}^\top/n)\,\mathbf{C}\,(\mathbf{I}-\mathbf{1}\,\mathbf{1}^\top/n)]$, for some suitable value of k; sequential, rather than simultaneous, variance extraction is desired in order to preserve interpretation of the extremes; see de Jong et al. (1984). This solution relates to the following decomposition theorem (after Tatsuoka, 1988, p. 41): the first eigenvector, say $\mathbf{E}_1$, is the set of numerical values that has the largest MC achievable by any set of real numbers for the spatial arrangement defined by the geographic connectivity matrix $\mathbf{C}$; the second eigenvector is the set of real numbers that has the largest achievable MC by any set that is uncorrelated with $\mathbf{E}_1$; the third eigenvector is the third such set of values; and so on through $\mathbf{E}_n$, the set of values that has the largest negative MC achievable by any set that is uncorrelated with the preceding $(n-1)$ eigenvectors.

The corresponding eigenvalues index these levels of SA: $\mathrm{MC} = n\mathbf{E}^\top\mathbf{C}\mathbf{E}/\mathbf{1}^\top\mathbf{C}\mathbf{1}$. But, in contrast to principal components analysis, rather than using the resulting eigenvectors to construct linear combinations of attribute variables (which would be the n 0-1 binary indicator variables forming matrix $\mathbf{C}$), the eigenvectors themselves (instead of principal components scores) are the desired synthetic variates, each containing n elements, one for each areal unit (i.e., location). For a given geographic landscape surface partitioning, the eigenvectors represent a statistical fixed effect in that matrix $(\mathbf{I}-\mathbf{1}\,\mathbf{1}^\top/n)\,\mathbf{C}\,(\mathbf{I}-\mathbf{1}\,\mathbf{1}^\top/n)$ does not, and hence they do not change from one attribute variable to another. Frequently, the linear combination of eigenvectors describing SA is what changes between attribute variables.

As with spatial autoregression analysis, eigenvector spatial filters depend upon the specification of matrix $\mathbf{C}$. A large number of the extracted eigenvectors can be classified into three qualitatively different positive SA groups: global, regional, and local (Figure 1). The first two members of this first group tend to be eigenvectors that strongly correlate with the underlying Cartesian coordinate system used to geocode locations (relating them to trend surface or gradient analysis), while the third eigenvector tends to portray a more centrally positioned hill/mound pattern; in other words, this group comprises gradients and dispersed sets of large-sized clusters of similar values. This second group comprises eigenvectors that portray dispersed sets of moderate-sized clusters of similar values. And, this third group comprises eigenvectors that portray dispersed sets of small-sized clusters of similar values. These latter two classes relate to local indices of SA; see, e.g., Anselin (1995). These three types of map pattern respectively reflect strong, moderate, and weak positive SA. Spatial filtering utilizes linear combinations of these eigenvectors-$\boldsymbol{E\beta}$, where $\boldsymbol{\beta}$ is a vector of regression coefficients-to describe SA latent in a given georeferenced response variable, capturing local spatial associations whose aggregate sum is measured by some global index value (e.g., the MC).

2.2   Relationships between spatial filtering and spatial autoregression

Consider, for example, the following nonlinear auto-normal model specification (the SAR specification; see Ord, 1975), written in matrix notation:

$$\mathbf{Y} = \mu\mathbf{1} + (\mathbf{I} - \rho\,\mathbf{C})^{-1}\,\epsilon, \quad \epsilon \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}),$$
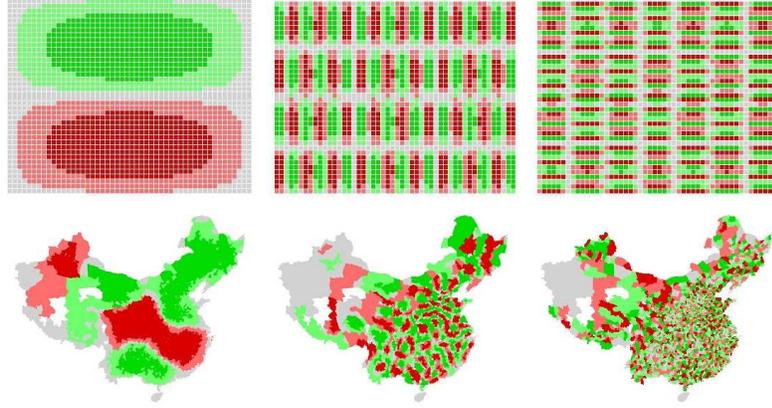
Figure 1. Spatial filter map patterns for a regular square tessellation (top) and the China county surface partitioning (bottom); quintile eigenvector value classes (which are relative, to a factor of -1) range from dark green to dark red. Top left (a): global map pattern. Top middle (b): regional map pattern. Top right (c): local map pattern. Bottom left (d): global map pattern. Bottom middle (e): regional map pattern. Bottom right (f): local map pattern.

where $\mathbf{C}$ is a geographic weights matrix and ρ is a SA parameter. This equation has no covariates because it is being used to construct a histogram for random variable $\mathbf{Y}$. Now

$$(\mathbf{I} - \rho\mathbf{C})\mathbf{Y} = \mu(\mathbf{I} - \rho\mathbf{C})1 + \epsilon$$
$$\mathbf{Y} = \rho\mathbf{C}\mathbf{Y} + \mu\mathbf{1} - \mu\rho\mathbf{C}\mathbf{1} + \epsilon$$
$$= \rho\mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^\top + \mu\mathbf{1} - \mu\rho\mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^\top 1 + \epsilon$$
$$= \mathbf{E}\left(\boldsymbol{\Lambda}\mathbf{E}^\top \rho\mathbf{Y}\right) + \mu\mathbf{1} - \mu\rho\mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^\top\mathbf{1} + \epsilon$$
$$= \mathbf{E}\beta^* + \mu\mathbf{1} - \mu\rho\mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^\top\mathbf{1} + \epsilon,$$

where $\mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^\top$ is the spectral decomposition of matrix $\mathbf{C}$. Standard principal components regression replaces covariate matrix $\mathbf{X}$ with its orthogonal and uncorrelated component scores. Here matrix $\mathbf{E}$ is orthogonal but not uncorrelated (because the mean of each eigenvector is not necessarily zero). It can be transformed to an orthogonal and uncorrelated matrix by replacing matrix $\mathbf{C}$ with matrix $(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\,\mathbf{C}\,(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)$; in other words, the initial model becomes $\mathbf{Y} = \mu\mathbf{1} + [\mathbf{I} - \rho(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\,\mathbf{C}\,(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)]^{-1}\epsilon^1$. Consequently,

$$Y = \rho(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\,\mathbf{C}\,(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\mathbf{Y} + \mu\mathbf{1} - \mu\rho(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\,\mathbf{C}\,(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\mathbf{1} + \epsilon$$
$$= \mathbf{E}^*(\boldsymbol{\Omega}\mathbf{E}^{*\top}\rho\mathbf{Y}) + \mu\mathbf{1} + \epsilon$$
$$= \mathbf{E}^*\boldsymbol{\beta} + \mu\mathbf{1} + \epsilon.$$

Because $\rho \in (1/\omega_{\min}, 1/\omega_{\max})$, where $\omega_{\min}$ and $\omega_{\max}$ are the extreme eigenvalues of matrix $(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\,\mathbf{C}\,(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)$, $\rho w_j < 1$.

Now consider the reduced linear specification

$$\mathbf{Y} = \mathbf{E}_k^*\boldsymbol{\beta} + \mu\mathbf{1} + \epsilon,$$

for which the standard regression result is $\hat{\boldsymbol{\beta}}_k = (\mathbf{E}_k^{*\top}\mathbf{E}_k^*)^{-1}\mathbf{E}_k^{*\top}\mathbf{Y}$, all $\mathbf{E}_j^*$ for which $\mathbf{E}_j^{*\top}Y \approx 0$ have been removed (hence, $k << n$) from the model specification, and recalling that $\mathbf{E}_j^{*\top}\mathbf{1} = 0$ for all but the eigenvector proportional to vector $\mathbf{1}$ (which has an

---

[1] Pre- and post-multiplying by the projection matrix results in the 1st eigenvalue of matrix C being replaced by zero, and all of the other eigenvalues being asymptotically the same as those for C.

eigenvalue of 0). Because $\rho w_j < 1$ for the nonlinear case, its regression coefficients are less than those for the linear case ($\rho w_j \mathbf{E}_j^{*\top} \mathbf{Y}$ rather than $\mathbf{E}_j^{*\top} \mathbf{Y}$). This is one trade-off between the nonlinear and linear specifications.

This result extends to the auto-binomial model through its standard log-linear specification:

$$< \ln \left[\frac{\mathrm{E}(p)}{1 - \mathrm{E}(p)} \right] >= \alpha \mathbf{1} + \rho \, \mathbf{C} \left\{ < \ln \left[\frac{\mathrm{E}(p)}{1 - \mathrm{E}(p)} \right] > - \alpha \mathbf{1} \right\},$$

where $< \cdot >$ denotes a vector, and $\mathrm{E}(p) = 1/(1 + e^{\alpha})$, when $\rho = 0$. From the preceding derivations, this equation implies

$$< \ln \left[\frac{\mathrm{E}(p)}{1 - \mathrm{E}(p)} \right] >= \alpha \mathbf{1} + \mathbf{E}_k^* \boldsymbol{\beta}_k.$$

For the auto-Poisson model, following Cressie,

$$< \ln \mathrm{E}(\mathbf{Y}) >= \alpha \mathbf{1} + \rho \mathbf{C}(\mathbf{N} - \alpha \mathbf{1})$$

for the vector of counts, $\mathbf{N}$. As a Poisson regression, this equation also can be rewritten as before:

$$< \ln \mathrm{E}(\mathbf{Y}) >= \alpha \mathbf{1} + \mathbf{E}_k^* \boldsymbol{\beta}_k.$$

As such, it is a respecification of Besag's auto-model that is in keeping with his random effects model specification.

The preceding description furnishes equivalencies between the selected auto-models and the spatial filter models. Many of the same histogram results can be simulated with the auto-models (using the Kaiser-Cressie winsoring for the auto-Poisson), employing MCMC techniques especially for the auto-Poisson case. But these auto-models are unable to achieve the extreme levels of SA that a spatial filter specification can achieve, because it involves a reduced form in that all eigenvectors for which $\mathbf{E}_j^{*\top} Y \approx 0$ are not included in the model specification. These extreme cases are the ones Besag once noted are of great interest.

2.3   A FUNCTIONAL FORM OF SELECTED EIGENVECTORS

A regular square tessellation (which is associated with a remotely sensed image) can be used to illustrate selected features of the eigenvectors used to construct spatial filters. One advantage of using matrix $\mathbf{C}$ based upon this surface partitioning is that its eigenvectors are known analytically to be $2/(\sqrt{(P+1)(Q+1)}) \sin[hp\pi/(P+1)] \sin[kq\pi/(Q+1)]$, for $p = 1, \ldots, P$ and $q = 1, \ldots, Q$, constituting the elements of an eigenvector with specified values h and k, for $h = 1, \ldots, P$ and $k = 1, \ldots, Q$, being the n different eigenvectors. Those vectors for which h and/or k are even integers also are eigenvectors of Equation (1). Furthermore, after replacing its first eigenvector with $(1/\sqrt{n})\mathbf{1}$ (which has a corresponding eigenvalue of 0), the remaining roughly $PQ/4$ eigenvectors for which both h and k are odd integers converge on their Equation (1) counterparts as n increases; see Griffith (2000a). A noteworthy eigenvector solution property is that to secure uniqueness, software packages produce normalized eigenvectors, meaning that $\mathbf{E}_j^{\top} \mathbf{E}_j = 1$. But uniqueness is to a multiplicative factor of -1, because $\mathbf{E}_j^{\top} \mathbf{E}_j$ produces a sum of squared values. Consequently, the sign of each regression coefficient in vector $\boldsymbol{\beta}$ obtained by regressing some response variable $\mathbf{Y}$ on a subset of K eigenvectors $\mathbf{E}_k$ becomes relative.

A second property associated with normalizing eigenvector $\mathbf{E}_j$ is that its elements e_{ij} tend to become smaller as n increases. This property is illustrated by taking the limit of $(1/\sqrt{n})\mathbf{1}$, which is zero. The presence of this particular eigenvector means that the projection matrix $(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)$ centers all of the non-principal eigenvalues. In addition, this particular eigenvector suggests that multiplying all of the normalized eigenvectors by $\sqrt{n}$ should result in their converging to a constant other than 0 as n increases. Consequently,

LEMMA 2.1 All of the elements of the normalized non-principal eigenvectors of matrix $\mathbf{C}$ based upon a regular square tessellation forming a complete rectangular region go to 0 as n goes to ∞.

PROOF $\lim_{P \to \infty} \lim_{Q \to \infty} 2/(\sqrt{(P+1)(Q+1)}) \sin[g_1 p\pi/(P+1)] \sin[g_2 q\pi/(Q+1)] = 0$, for $0 < h < 1$ and $0 < k < 1$, $hP = p$ and $kQ = q$, $g_1 = 1, \ldots, P$ and $g_2 = 1, \ldots, Q$, where g_1 and g_2 cannot both be 1. ∎

Multiplying the eigenvectors by $\sqrt{n}\sqrt{PQ}$ results in each eigenvector element converging to a constant.

LEMMA 2.2 When a normalized non-principal eigenvector of matrix $\mathbf{C}$ based upon a regular square tessellation forming a complete rectangular region is multiplied by $\sqrt{n} = \sqrt{PQ}$, the elements of this eigenvector converge on constants as n goes to ∞.

PROOF $\lim_{P \to \infty} \lim_{Q \to \infty} 2\sqrt{PQ}/(\sqrt{(P+1)(Q+1)}) \sin[g_1 p\pi/(P+1)] \sin[g_2 q\pi/(Q+1)] = 2\sin(h\pi)\sin(k\pi) \in [-2, 2]$, for $0 < h < 1$ and $0 < k < 1$, $hP = p$ and $kQ = q$, $g_1 = 1, \ldots, P$ and $g_2 = 1, \ldots, Q$, where g_1 and g_2 cannot both be 1. ∎

Therefore, premultiplying a vector by $\sqrt{n}$ results in at least some of its elements not going to 0 with increasing n, and the sum of the elements of each adjusted eigenvector remaining 0, whereas the sum of its squared values now equals n. Accordingly, its variance becomes standardized to 1. In subsequent sections, eigenvectors are the normalized vectors pre-multiplied by $\sqrt{n}$.

2.4 ASSUMPTIONS UNDERLYING AND MAJOR IMPLICATIONS OF THE SPATIAL MODEL SPECIFICATIONS

The principal assumptions for eigenvector spatial filter models are as follows:

ASSUMPTION 1. The direct SA structure of a system only depends upon cliques containing no more than two locations,

ASSUMPTION 2. The probability distribution associated with each location accounts for SA across the parent map through the parameters of this distribution,

ASSUMPTION 3. The n RVs are (conditionally) independent.

Assumption 1 supports the construction of the geographic neighbors matrix $\mathbf{C}$. In addition, by construction, the eigenvectors are orthogonal and uncorrelated, each has a mean equal to 0 [induced by the projection matrix $(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)$] and a sum of squares equal to n (because they are normalized eigenvectors premultiplied by $\sqrt{n}$), and each is unique except for a multiplicative factor or -1.

The auto-models literature reveals that the mean of a RV tends to be unbiased (one notable exception is the autoregressive response, or spatial lag, specification when covariates are present), whereas the variance tends to be altered by the presence of non-zero SA. Eigenvector spatial filtering specifications preserve these two features of RVs. Although Kaiser and Cressie (1997) established a methodology that allows some degree of positive SA to be accounted for in Poisson RVs through winsorization, the attainable levels are modest at best. Eigenvector spatial filtering specifications allow marked levels of positive

SA to be accounted for in a Poisson or negative binomial RV. Meanwhile, the auto-normal model has heterogeneous variance (because of the matrix inversion involved), and becomes unstable as its autoregressive parameter approaches the boundary of its feasible SA parameter space. And, both the winsorized auto-Poisson and the auto-binomial models tend to enter phase transitions even when SA reaches only moderate levels. These are features not retained by the eigenvector spatial filtering specifications.

Eigenvector spatial filters involve one of two kinds of specification. The first one simply is an additive term, $\mathbf{E}_i\boldsymbol{\beta}$, where $\mathbf{E}_i$ is a 1-by-K row vector of the ith elements of K eigenvectors, that converts a constant, α, to an observation-specific, variable intercept term, α_i, for observation i. This intercept term is expressed as

$$\alpha_i = \alpha + \mathbf{E}_i\boldsymbol{\beta}, \tag{2}$$

and is suitable for linear models. Because $\mathbf{1}^\top\mathbf{E}\boldsymbol{\beta} = 0$, $\bar{\alpha}_i = \alpha$. If $\boldsymbol{\beta} = \mathbf{0}$ (i.e., zero SA), $\alpha_i = \alpha, \forall i$. The second specification is a multiplicative one, k_i, for observation i. Now an intercept term for a nonlinear model, such as that for a Poisson RV, is expressed as

$$e^{\alpha_i} = e^\alpha \left(\frac{ne^{\mathbf{E}_i\boldsymbol{\beta}}}{\sum_{i=1}^n e^{\mathbf{E}_i\boldsymbol{\beta}}} \right),$$

where $k_i = ne^{\mathbf{E}_i\boldsymbol{\beta}}/\sum_{i=1}^n e^{\mathbf{E}_i\boldsymbol{\beta}}$. In other words, SA either inflates, does not modify, or deflates a location-specific intercept term, depending upon whether or not the observation-specific quantity $e^{\mathbf{E}_i\boldsymbol{\beta}}$ is greater than , equal to, or less than the average spatial filter value $(\sum_{i=1}^n e^{\mathbf{E}_i\boldsymbol{\beta}})/n$. Because $\sum_{i=1}^n (ne^{\mathbf{E}_i\boldsymbol{\beta}}/\sum_{i=1}^n e^{\mathbf{E}_i\boldsymbol{\beta}})/n = 1$, $\overline{e^{\alpha_i}} = e^\alpha$. If $\boldsymbol{\beta} = \mathbf{0}$ (i.e., zero SA), $e^{\alpha_i} = e^\alpha, \forall i$. A generalized version of this multiplicative specification is given by $e^{\mathbf{E}_i\boldsymbol{\beta}}(\sum_{i=1}^n e^{\mathbf{E}_i\boldsymbol{\beta}}/n)^\eta$, where η may be other than -1. Selection of one or the other of these two specifications is based upon a mean response exhibiting a geographic trend while its intercept term remains unbiased. Of note is that the appropriate use of additive or multiplicative SA terms also characterizes spatial autoregressive model specifications (e.g., the auto-normal versus auto-Poisson model).

3. A STANDARD BENCHMARK: NORMAL CURVE THEORY

Historically, variable transformations, such as Box-Cox and Manly, were devised so that well-developed normal curve theory could be used to analyze appropriately transformed RVs of almost any type. Griffith (2000b) presented a respecification of the auto-normal pdf as one for n independent normal RVs, $Y_i = 1, \ldots, n$ for which the expression of each mean is defined by Equation (2), in keeping with linear model theory (i.e., a standard linear regression specification):

$$P(Y_1, \ldots, Y_n) = (2\pi\sigma^2)^{-n/2} e^{-\sum_{i=1}^n [y_i - (\mu_i + \mathbf{E}_i\boldsymbol{\beta})]^2/(2\sigma^2)},$$

where μ denotes the unautocorrelated RV population mean and σ^2 denotes the unautocorrelated RV population variance, both of which are constant across a map. This joint specification modifies a parameter by rewriting the non-constant mean of an individual observation i as $\mu_i = \mu + \mathbf{E}_i\boldsymbol{\beta}$, for which, as an aside, $\mathbf{E}_i\boldsymbol{\beta}$ becomes the spatially structure random effects term in a mixed model specification. Accordingly, for each location i,

$$P(Y_i) = (2\pi\sigma^2)^{-1/2} e^{-[y_i - (\mu + \mathbf{E}_i\boldsymbol{\beta})]^2/(2\sigma^2)} = \text{pdf}_i. \tag{3}$$

The joint MGF of Equation (3) for n independent RVs is

$$M(t_1, \ldots, t_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{y_1 t_1} \mathrm{pdf}_1 \ldots e^{y_n t_n} \mathrm{pdf}_n \, dy_1 \ldots dy_n,$$

with the marginal distribution for observation i being given by

$$M(\mathbf{0}_{m_1}^{\top}, t_i, \mathbf{0}_{m_2}^{\top}) = \int_{-\infty}^{\infty} e^{y_i t_i} \mathrm{pdf}_i \, dy_i = e^{(\mu + \mathbf{E}_i \boldsymbol{\beta})t + \sigma^2 t^2/2}, \tag{4}$$

where $\mathbf{0}_{m_j}$ is an m_j-by-1 vector of 0s, with $m_1 + m_2 = n - 1$.

The mean of a set of n independent map values has the MGF

$$M_{(\sum_{i=1}^{n} Y_i)/n}(t) = \prod_{i=1}^{n} e^{(\mu + \mathbf{E}_i \beta)t/n + \sigma^2 (t/n)^2/2},$$

whose first derivative with respect to t, when $t = 0$, is μ, the mean of the normal RV in the absence of SA. In other words, the mean is unbiased, and a sample histogram tends to be distributed around the constant part of the population map mean, except for sampling error.

The variance of a set of n independent map values is given by the average across all locations of the second derivative of Equation (4) with respect to t_i, when $t_i = 0$, of Equation (4), adjusted for its mean, yielding

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^2\right) = \frac{1}{n}\sum_{i=1}^{n}[(\mathbf{E}_i \boldsymbol{\beta})^2 + \sigma^2] = \sum_{j=1}^{K}\beta_j^2 + \sigma^2 = \sigma^2\left[1 + \frac{1}{\sigma^2}\sum_{j=1}^{K}\beta_j^2\right],$$

where K denotes the number of eigenvectors in a spatial filter, and $(\sum_{i=1}^{n}\beta_j^2)/\sigma^2$ a SA-signal-to-independent-RV-noise ratio. The variance inflation factor (VIF) is $[1 + (\sum_{j=1}^{K}\beta_j^2)/\sigma^2]$. This term reduces to 1 when SA is zero (i.e., $\beta_j = 0$, $\forall j$).

The skewness of a set of n independent map values is given by the average across all locations of the third derivative of Equation (4) with respect to t_i, when $t_i = 0$, of Equation (4), adjusted for its mean, yielding

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^3\right) = \frac{1}{n}\sum_{i=1}^{n}[(\mathbf{E}_i \boldsymbol{\beta})^3 + 3(\mathbf{E}_i \boldsymbol{\beta})\sigma^2] = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{E}_i \boldsymbol{\beta})^3$$

$$= \sigma^3 \frac{1}{n\sigma^3}\sum_{i=1}^{n}\left(\sum_{j=1}^{K} e_{ij}\beta_j\right)^3,$$

where e_{ij} denotes the ith element of the jth eigenvector, which is the numerator of the skewness measure. Expansion of the cubic expression reveals that this quantity is not 0 if a spatial filter comprises at least three eigenvectors for which $\beta_j \neq 0$, or if the surface partitioning in question is other than a regular square tessellation (which lacks symmetry in the set of elements contained in each of its eigenvectors). Therefore, skewness is given

by

$$\alpha_3 = \frac{\frac{1}{n\sigma^3} \sum_{i=1}^{n} \left(\sum_{j=1}^{K} e_{ij}\beta_j \right)^3}{\left(1 + \sum_{i=1}^{K} \frac{\beta_j^2}{\sigma^2} \right)^{3/2}},$$

indicating that SA tends to distort the skewness of a normal RV. When zero SA is present, $\alpha_3 = 0$. Finally, the kurtosis (not the excess kurtosis) of a set of n independent map values is given by the average across all locations of the fourth derivative with respect to t_i, when $t_i = 0$, of Equation (4), adjusted for its mean, yielding

$$E\left(\frac{1}{n} \sum_{i=1}^{n} \{Y_i - \mu\}^4 \right) = \frac{1}{n} \sum_{i=1}^{n} [(\mathbf{E}_i\boldsymbol{\beta})^4 + 6(\mathbf{E}_i\boldsymbol{\beta})^2\sigma^2 + 3\sigma^4]$$

$$= \sigma^4 \left[\frac{1}{n\sigma^4} \sum_{i=1}^{n} \left(\sum_{j=1}^{K} e_{ij}\beta_j \right)^4 + \frac{6}{\sigma^2} \sum_{j=1}^{K} \beta_j^2 + 3 \right],$$

which is the numerator of the kurtosis measure. If at least a single $\beta_j \neq 0$, this quantity does not equal $3\sigma^4$. Therefore, kurtosis is given by

$$\alpha_4 = \frac{\frac{1}{n\sigma^4} \sum_{i=1}^{n} \left(\sum_{j=1}^{K} e_{ij}\beta_j \right)^4 + \frac{6}{\sigma^2} \sum_{j=1}^{K} \beta_j^2 + 3}{\left(1 + \frac{1}{n\sigma^4} \sum_{j=1}^{K} \frac{\beta_j^2}{\sigma^2} \right)^2},$$

indicating that SA also tends to distort the kurtosis of a normal RV. When zero SA is present, $\alpha_4 = 3$; otherwise, it does not.

In conclusion, besides variance inflation, SA impacts upon the skewness and the kurtosis of normal RVs, a finding essentially ignored in the literature. This impact is more acute for surface partitioning that do not form a square tessellation. As an example of regular square tessellation data, consider such a tessellation forming a complete 50-by-40 rectangular landscape ($n = 2000$). Suppose the spatial filter is given by SF $= 4E_1 + 2E_2 + E_3$, the three eigenvectors of matrix expression give in Equation (1) portraying the highest levels of positive SA, and consider a standard normal RV. Figure 2a portrays a simulated outcome for this case, with its affiliated summary statistics appearing in Table 1. As an example of irregular surface partitioning data, consider the partitioning of China into counties ($n = 2,379$). Using its first three eigenvectors of matrix expression give in Equation (1) portraying the highest levels of positive SA produces a simulated outcome portrayed in Figure 3a. Its affiliated summary statistics appear in Table 2.

4. A Heuristic Overview: The Beta RV

The beta distribution is very flexible, being able to mimic a sinusoidal RV, a uniform RV, a skewed (e.g., exponential form) RV, and a normal RV, and as such furnishes a wide range of insights into the impacts of SA on histograms. 1985 and McKenzie demonstrated that an auto- version of the beta probability model is feasible at least for time series data. More recently, Kaiser et al. (2002) furnished the first spatial version of a beta probability model that accounts for SA. One issue about which a spatial filter version of the beta probability model informs concerns selection between a multiplicative and an additive

Table 1. Simulation examples for a 50-by-40 regular square tessellation.

Statistic	Standard normal RV	Beta RV				Poisson RV	Binomial RV
		$\alpha = \gamma = 12$	$\alpha = \gamma = 250$	$\alpha = \gamma = 12$; SF/15	$\alpha = 36$ $\gamma = 12$		
				Theoretical results			
Map mean	0	0.50	0.50	0.50	0.75	20	5
VIF	22.0	4.50	1.17	1.19	3.33	76.3	4.3
Skewness	0.42	0.29	0.02	0.02	0.02	3.05	0.09
Kurtosis	2.59	2.39	2.98	1.69	2.55	12.52	1.73
				Simulation results			
MC	0.95	0.81	0.15	0.22	0.71	0.98	0.85
% variance accounted for by SF	95.4	80.7	14.5	23.5	71.0	98.7	84.9
Goodness-of-fit[1]	Reject H_0	Reject H_0	Fail to Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0

Table 2. Simulation examples for the China county surface partitioning.

Statistic	Standard normal RV	Beta RV				Poisson RV; SF/20	Binomial RV SF/2
		$\alpha = \gamma = 12$; SF/2	$\alpha = \gamma = 3$; SF/8	$\alpha = \gamma = 1.1$; SF/22	$\alpha = 36, \gamma = 12$; SF/2		
				Theoretical results			
Map mean	0	0.5	0.5	0.5	0.5	20	5
VIF	22.0	1.88	1.21	1.08	1.58	19.93	3.89
Skewness	5.77	0.02	0.00	0.00	-0.29	32.98	0.01
Kurtosis	55.29	3.27	2.32	1.81	3.35	1369.89	1.92
				Simulation results			
MC	1.12	0.54	0.23	0.11	0.42	0.51	0.90
% variance accounted for by SF	95.5	48.9	19.3	11.1	37.3	95.4	82.6
Goodness-of-fit[2]	Reject H_0	Mixed	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0

parameter adjustment to account for latent SA. Let $k_i = e^{\mathbf{E}_i \boldsymbol{\beta}}/(\sum_{i=1}^{n} e^{\mathbf{E}_i \boldsymbol{\beta}}/n)$ be the multiplicative SA factor; if $\boldsymbol{\beta} = \mathbf{0}$, then $k_i = 1$. The respecified marginal beta pdf$_i$ may be written as

$$\frac{\Gamma(\alpha k_i, \gamma k_i)}{\Gamma(\alpha k_i)\Gamma(\gamma k_i)} y_i^{\alpha k_i - 1}(1 - y_i)^{\gamma k_i - 1},$$

where $0 < y_i < 1$, $\alpha > 0$ and $\gamma > 0$, are shape parameters, $\alpha k_i > 0$ and $\gamma k_i > 0$, and Γ denotes the gamma function. The marginal distribution for observation i derived from the joint MGF is

$$M(\mathbf{0}^\top, t_i, \mathbf{0}^\top) = \int_0^1 e^{y_i t_i} \text{pdf}_i dy_i = 1 + \sum_{h=1}^{\infty} \left[\prod_{r=0}^{h-1} \frac{\alpha k_i + r}{\alpha k_i + \gamma k_i + r} \right] \frac{t_i^h}{h!}.$$

The expected value of the mean of n independent map values has the MGF

$$M_{(\sum_{i=1}^{n} Y_i)/n}(t) = \prod_{i=1}^{n} \left\{ 1 + \sum_{h=1}^{\infty} \left[\prod_{r=0}^{h-1} \frac{\alpha k_i + r}{\alpha k_i + \gamma k_i + r} \right] \frac{t_i^h}{n^h h!} \right\},$$

whose first derivative with respect to t, when $t = 0$, is $\alpha/(\alpha + \gamma)$, which is the mean of the beta RV in the absence of SA. In other words, the mean is unbiased as well as a constant
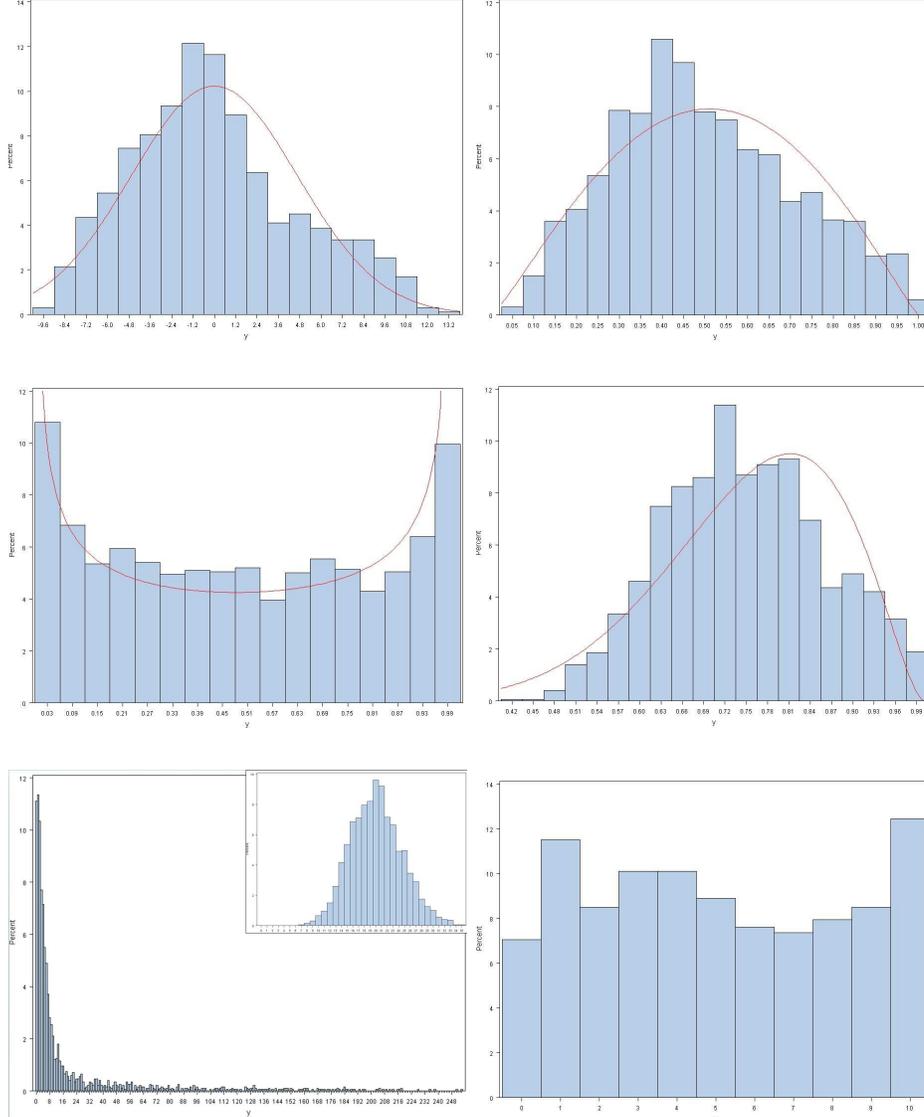
Figure 2.   Histograms for simulated positive SA outcomes based on a square tessellation. Top left (a): normal RV. Top right (b): bell-shaped beta RV. Middle left (c): uniform-shaped beta RV. Middle right (d): negative skewed beta RV. Bottom left (e): Poisson RV. Bottom right (f): binomial RV.

across a map, and a sample histogram tends to be distributed around the population map mean, except for sampling error.

A conspicuous problem with this multiplicative specification is that because $\mu_i = \alpha/(\alpha + \gamma)$, $\forall i$, no SA can be detected in a mean response map pattern employing this multiplicative respecification, even though it creates variance inflation and impacts skewness and kurtosis. For the case where $\alpha = \gamma$, as $\alpha = \gamma$ increases, the shape of the beta distribution increasingly is indistinguishable from that of a normal distribution. Therefore, in keeping with Equation (3), this feature suggests that the appropriate specification should yield $\mu_i = \alpha/(\alpha + \gamma) + f(\mathbf{E}_i\boldsymbol{\beta})$. Consequently, the multiplicative specification is inappropriate for a beta RV.

In contrast, the additive respecification, such that the parameters become $\alpha + \mathbf{E}_i\boldsymbol{\beta}$ and $\gamma - \mathbf{E}_i\boldsymbol{\beta}$, results in the marginal beta pdf$_i$ being written as

$$\frac{\Gamma(\alpha + \mathbf{E}_i\boldsymbol{\beta}, \gamma - \mathbf{E}_i\boldsymbol{\beta})}{\Gamma(\alpha + \mathbf{E}_i\boldsymbol{\beta})\Gamma(\gamma - \mathbf{E}_i\boldsymbol{\beta})} y_i^{\alpha + \mathbf{E}_i\boldsymbol{\beta_i} - 1}(1 - y_i)^{\gamma - \mathbf{E}_i\boldsymbol{\beta_i} - 1},$$
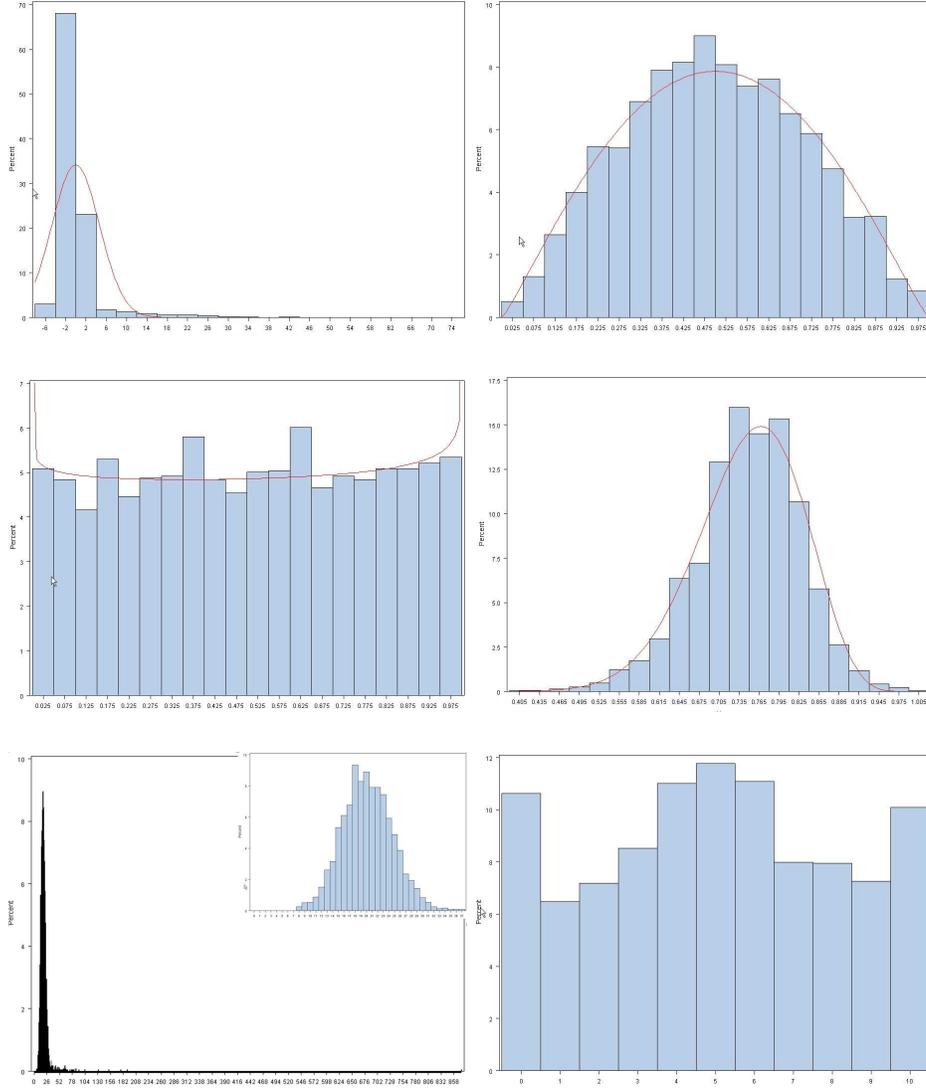
Figure 3. Histograms for simulated positive SA outcomes based on a square tessellation. Top left (a): normal RV. Top right (b): bell-shaped beta RV. Middle left (c): uniform-shaped beta RV. Middle right (d): negative skewed beta RV. Bottom left (e): Poisson RV. Bottom right (f): binomial RV.

where, $\alpha + \mathbf{E}_i\boldsymbol{\beta} > 0$ and $\gamma - \mathbf{E}_i\boldsymbol{\beta} > 0$. These restrictions need to be carefully assessed because each eigenvector always contains both positive and negative elements (i.e., its elements sum to 0, by construction). This specification is in keeping with that formulated by Hardouin and Yao (2008). Now the marginal MGF for location i is

$$1 + \sum_{h=1}^{\infty} \left(\prod_{r=0}^{h-1} \frac{\alpha + \mathbf{E}_i\boldsymbol{\beta} + r}{\alpha + \gamma + r} \right) \frac{t_i^h}{[} h!]. \tag{5}$$

With this additive specification, the expected value of a map mean becomes

$$\mathrm{E}\left(\frac{1}{n} \sum_{i=1}^{n} Y_i \right) = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\alpha}{\alpha + \gamma} + \mathbf{E}_i \frac{\boldsymbol{\beta}}{\alpha + \gamma} \right] = \frac{\alpha}{\alpha + \gamma},$$

which is the mean of the beta RV in the absence of SA. In addition, the individual location i means, $\alpha/(\alpha + \gamma) + \mathbf{E}_i\boldsymbol{\beta}/(\alpha + \gamma)$, display SA. Again, a sample histogram tends to be

distributed around the population map mean, except for sampling error.

The variance of a set of n independent map values is given by the average across all locations of the expected value of the second marginal moment about the mean, or

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^2\right) = \frac{\alpha\gamma}{(\alpha+\gamma)^2(\alpha+\gamma+1)}\left\{1 + \frac{\alpha+\gamma}{\alpha\gamma}\sum_{j=1}^{K}\beta_j^2\right\}.$$

The VIF is $\{1 + [(\alpha+\gamma)/(\alpha\gamma)]\sum_{j=1}^{K}\beta_j^2\}$. This term reduces to 1 when SA is zero (i.e., $\beta_j = 0$, $\forall j$).

The numerator of skewness calculated with a set of n independent map values is given by the average across all locations of the expected value of the third marginal moment about the mean, or

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^3\right) = \frac{(\gamma+\alpha)\left[3(\gamma-\alpha)\sum_{j=1}^{K}\beta_j^2 + (\alpha+\gamma)\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{K}E_{ij}\beta_j\right)\right]}{(\alpha+\gamma)^3(\alpha+\gamma+1)(\alpha+\gamma+2)}$$

$$+\frac{2\alpha\gamma(\gamma-\alpha)}{(\alpha+\gamma)^3(\alpha+\gamma+1)(\alpha+\gamma+2)}.$$

Expansion of its cubic expression reveals that this average moment is not 0 when $\alpha = \gamma$ if a spatial filter comprises at least three eigenvectors for which $\beta_j \neq 0$, or if the surface partitioning in question is other than a regular square tessellation. Here, skewness is given by

$$\alpha_3 = \frac{\sqrt{\alpha+\gamma+1}\left\{2\gamma\alpha(\gamma-\alpha)+(\alpha+\gamma)\right\}\left[3\gamma\alpha\sum_{j=1}^{K}\beta_j^2 + (\gamma+\alpha)\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{K}E_{ij}\beta_j\right)^3\right]}{(\gamma\alpha)^{\frac{3}{2}}(\alpha+\gamma+2)\left(1+\frac{\alpha+\gamma}{\alpha\gamma}\sum_{j=1}^{K}\beta_j^2\right)^{\frac{3}{2}}},$$

indicating that SA tends to distort the skewness of a beta RV. When zero SA is present, $\alpha_3 = 0$ when $\alpha = \gamma$, although $\alpha_3 = 0$ is converged upon as $\alpha = \gamma \to \infty$, holding the included SF constant, with non-zero SA affecting this rate of convergence. Finally, the numerator of kurtosis calculated with a set of n independent map values is given by the average across all locations of the expected value of the 4th marginal moment about the mean, or

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^4\right) = \left\{3\alpha\gamma(2\gamma^2 + \gamma^2\alpha - 2\gamma\alpha + \gamma\alpha^2 + 2\alpha^2)\right.$$

$$+(\gamma+\alpha)(11\gamma^2 + 6\gamma^2\alpha - 14\gamma\alpha + 6\gamma\alpha^2 + 11\alpha^2)$$

$$\times\sum_{j=1}^{K}\beta_j^2 + 6(\gamma+\alpha)^2(\gamma-\alpha)\left[\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{K}E_{ij}\beta_j\right)^3\right]$$

$$\left.+(\gamma+\alpha)^3\left[\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{K}E_{ij}\beta_j\right)^4\right]\right\}\Bigg/$$

$$\left\{(\alpha+\gamma)^4(\alpha+\gamma+1)(\alpha+\gamma+2)(\alpha+\gamma+3)\right\}.$$

If at least a single $\beta_j \neq 0$, the second, third, and fourth terms on the right-hand-side of this equation do not equal zero. Therefore, kurtosis is given by

$$
\alpha_4 = \left\{ 3\alpha\gamma(\alpha + \gamma + 1)(2\gamma^2 + \gamma^2\alpha - 2\gamma\alpha + \gamma\alpha^2 + 2\alpha^2) \right.
$$

$$
+ (\gamma + \alpha)(\alpha + \gamma + 1)(11\gamma^2 + 6\gamma^2\alpha - 14\gamma\alpha + 6\gamma\alpha^2 + 11\alpha^2)\left(\sum_{j=1}^{K} \beta_j^2 \right)
$$

$$
+ 6(\gamma + \alpha)^2(\gamma - \alpha)(\alpha + \gamma + 1)\left[\frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{K} E_{ij}\beta_j \right)^3 \right]
$$

$$
\left. + (\gamma + \alpha)^3(\alpha + \gamma + 1)\left[\frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{K} E_{ij}\beta_j \right)^4 \right] \right\} \Bigg/
$$

$$
\left\{ \left[(\alpha\gamma)^2(\alpha + \gamma + 2)(\alpha + \gamma + 3)\left(1 + \frac{\alpha + \gamma}{\alpha\gamma} \sum_{j=1}^{K} \beta_j^2 \right)^2 \right] \right\},
$$

indicating that SA also tends to distort the kurtosis of a beta RV. Whether or not zero SA is present, α_4 converges on 3 as $\alpha = \gamma \to \infty$, holding the included SF constant; but, non-zero SA affects this rate of convergence.

In conclusion, besides variance inflation, SA impacts upon the skewness and the kurtosis of beta RVs. The type of georeferenced RV of interest here is illustrated by the set of digital numbers measuring radiance in a remotely sensed image, which commonly range from 0 to 255, with these two extremes indicating minimum and maximum radiance. As with normal RVs, SA impacts are more acute for surface partitioning that do not form a square tessellation. Continuing the preceding regular square tessellation example, but for a beta rather than a normal RV, suppose $\alpha = \gamma = 12$ (note: these values were selected because the extremes of the spatial filter are roughly 8 and 11), which yields the estimates $\widehat{\alpha} = 2.2$ and $\widehat{\gamma} = 2.1$ when non-zero SA is overlooked. Figure 2b portrays a simulated outcome for this case. A simulated outcome for the case of increasing the parameter values to $\alpha = \gamma = 250$ (a value far greater than what is needed in the absence of SA in order to mimic a bell-shaped curve) illustrates the preservation of a bell-shaped curve when a SF is held constant and $\alpha = \gamma$ increases. Now $\widehat{\alpha} = 215.6$, $\widehat{\gamma} = 215.3$. Thus, although a bell-shape is preserved, the parameter estimates are grossly wrong. Next, Figure 2c portrays a simulated outcome for a uniform distribution (i.e., $\alpha = \gamma = 1$, and the spatial filter is set to SF/15), which yields the estimates $\widehat{\alpha} = 0.70$ and $\widehat{\gamma} = 0.69$. Finally, Figure 2d portrays a simulated outcome for a negatively skewed distribution with $\alpha = 36$ and $\gamma = 12$, which yields the estimates $\widehat{\alpha} = 8.2$ and $\widehat{\gamma} = 2.7$. Affiliated summary statistics for each of these four simulations appear in Table 1.

Meanwhile, as an example of irregular surface partitioning data, again consider the county partitioning of China. The spatial filter $\mathrm{SF} = 4\mathrm{E}_4 + 2\mathrm{E}_9 + \mathrm{E}_{10}$, which renders results that are more comparable to those obtained with the regular square tessellation than does $\mathrm{SF} = 4\mathrm{E}_1 + 2\mathrm{E}_2 + \mathrm{E}_3$, has a range of roughly -11 to 12. This spatial filter is used here for illustrative purposes because dramatic asymmetries in the elements of the original spatial filter involving $\mathrm{E}_1, \mathrm{E}_2$ and E_3 virtually always result in a histogram that resembles a near-bell-shaped curve (α and γ have to be at least 78) with noticeable positive skewness. This new spatial filter produces a slightly lower level of positive SA, but furnishes results that parallel those for a regular square tessellation, allowing counterparts to Figures 2b-2d to be constructed (Figures 3b-3d). Affiliated summary statistics appear in Table 2.

5. The Poisson RV

Griffith (2002) also presented a respecification of the auto-Poisson model, one that can account for a wide range of positive SA, that expresses each log-mean as the sum of a constant and an eigenvector spatial filter, yielding the pmf$_i$

$$\frac{e^{-e^{[\ln(\mu)+\mathbf{E}_i\beta-\ln(\sum_{i=1}^n e^{\mathbf{E}_i\beta}/n)]}}\left\{e^{[\ln(\mu)+\mathbf{E}_i\beta-\ln(\sum_{i=1}^n e^{\mathbf{E}_i\beta}/n)]}\right\}^{y_i}}{y_i!}=$$

$$\frac{e^{\mu e^{\mathbf{E}_i\beta}/(\sum_{i=1}^n e^{\mathbf{E}_i\beta}/n)}\left[\frac{\mu e^{\mathbf{E}_i\beta}}{(\sum_{i=1}^n e^{\mathbf{E}_i\beta}/n)}\right]^{y_i}}{y_i!}, \quad (6)$$

where ln denotes the natural logarithm, and $\mu > 0$. This specification allows the mean of an individual observation i to be rewritten as $\mu_i = \mu e^{\mathbf{E}_i\beta}/\left(\sum_{i=1}^n e^{\mathbf{E}_i\beta}/n\right)$ -the multiplicative spatial filter specification-which results in a constant mean being inflated or deflated, depending upon whether the eigenvector spatial filter term is greater than, or less than, the average spatial filter term.

The marginal distribution for observation i derived from the joint MGF is

$$M(\mathbf{0}^\top, t_i, \mathbf{0}^\top) = \sum_{y_i=0}^\infty e^{y_i t_i}\mathrm{pmf_i} = e^{\mu[e^{\mathbf{E}_i\beta}/(\sum_{i=1}^n e^{\mathbf{E}_i\beta}/n)](e^t-1)}.$$

The expected value of the mean of n independent map values has the MGF

$$M_{(\sum_{i=1}^n Y_i)/n}(t) = \prod_{i=1}^n e^{\mu[e^{\mathbf{E}_i\beta}/(\sum_{i=1}^n e^{\mathbf{E}_i\beta}/n)](e^t-1)},$$

whose first derivative with respect to t, when $t = 0$, is μ, which is the mean of the Poisson RV in the absence of SA. In other words, the mean is unbiased, and a sample histogram tends to be distributed around the population map mean, except for sampling error.

The variance of a set of n independent map values is given by the average across all locations of the expected value of the second marginal moment about the mean, or

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n\{Y_i-\mu\}^2\right) = \mu\left[1-\mu+\frac{\frac{\mu}{n}\sum_{i=1}^n e^{2E_i\beta}}{\left(\frac{1}{n}\sum_{i=1}^n e^{E_i\beta}\right)^2}\right].$$

The VIF is $[1-\mu+(\mu/n)\sum_{i=1}^n e^{2E_i\beta}/(\sum_{i=1}^n e^{E_i\beta}/n)^2]$. This term reduces to 1 when SA is zero (i.e., $\beta_j = 0, \forall j$).

The numerator of skewness calculated with a set of n independent map values is given by the average across all locations of the expected value of the third marginal moment about the mean, or

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n\{Y_i-\mu\}^3\right) = \mu\left[1-3\mu+2\mu^2+\frac{\frac{3(\mu-\mu^2)}{n}\sum_{i=1}^n e^{2E_i\beta}}{\left(\frac{1}{n}\sum_{i=1}^n e^{E_i\beta}\right)^2}+\frac{\frac{\mu^2}{n}\sum_{i=1}^n e^{3E_i\beta}}{\left(\frac{1}{n}\sum_{i=1}^n e^{E_i\beta}\right)^3}\right].$$

Here, skewness is given by

$$
\alpha_3 = \frac{\mu^{-1/2}\left[1 - 3\mu + 2\mu^2 + \frac{\frac{3(\mu-\mu^2)}{n}\sum_{i=1}^{n}e^{2E_i\beta}}{\left(\frac{1}{n}\sum_{i=1}^{n}e^{E_i\beta}\right)^2} + \frac{\frac{\mu^2}{n}\sum_{i=1}^{n}e^{3E_i\beta}}{\left(\frac{1}{n}\sum_{i=1}^{n}e^{E_i\beta}\right)^3}\right]}{\left[1 - \mu + \frac{\frac{\mu}{n}\sum_{i=1}^{n}e^{2E_i\beta}}{\left(\frac{\sum_{i=1}^{n}e^{E_i\beta}}{n}\right)^2}\right]^{\frac{3}{2}}},
$$

indicating that SA tends to distort the skewness of a Poisson RV. When zero SA is present, $\alpha_3 = 0$ when $\mu \to \infty$, a property not preserved in the presence of non-zero SA. Finally, the numerator of kurtosis calculated with a set of n independent map values is given by the average across all locations of the expected value of the fourth marginal moment about the mean, or

$$
E\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^4\right) = \mu\left\{1 - 4\mu + 6\mu^2 - 3\mu^3 + (7\mu - 12\mu^2 + 6\mu^3)\left[\frac{\frac{1}{n}\sum_{i=1}^{n}e^{2E_i\beta}}{\left(\sum_{i=1}^{n}\frac{1}{n}e^{E_i\beta}\right)^2}\right]\right.
$$
$$
\left. + (6\mu^2 - 4\mu^3)\left[\frac{\frac{1}{n}\sum_{i=1}^{n}e^{3E_i\beta}}{\left(\frac{1}{n}\sum_{i=1}^{n}e^{E_i\beta}\right)^3}\right] + \mu^3\left[\frac{\frac{1}{n}\sum_{i=1}^{n}e^{4E_i\beta}}{\left(\frac{1}{n}\sum_{i=1}^{n}e^{E_i\beta}\right)^4}\right]\right\}.
$$

Therefore, kurtosis is given by

$$
\alpha_4 = \frac{1}{\mu}\left\{1 - 4\mu + 6\mu^2 - 3\mu^3 + (7\mu - 12\mu^2 + 6\mu^3)\left[\frac{\frac{1}{n}\sum_{i=1}^{n}e^{2E_i\beta}}{\left(\frac{1}{n}\sum_{i=1}^{n}e^{E_i\beta}\right)^2}\right]\right.
$$
$$
\left. + \frac{\frac{(6\mu^2 - 4\mu^3)}{n}\sum_{i=1}^{n}e^{3E_i\beta}}{\left(\frac{\sum_{i=1}^{n}e^{E_i\beta}}{n}\right)^3} + \frac{\frac{\mu^3}{n}\sum_{i=1}^{n}e^{4E_i\beta}}{\left(\frac{1}{n}\sum_{i=1}^{n}e^{E_i\beta}\right)^4}\right\} \Big/ \left\{\left[1 - \mu + \frac{\frac{\mu}{n}\sum_{i=1}^{n}e^{2E_i\beta}}{\left(\frac{1}{n}\sum_{i=1}^{n}e^{E_i\beta}\right)^2}\right]^2\right\},
$$

indicating that SA also distorts the kurtosis of a Poisson RV. In addition, α_4 no longer converges on 3 as $\mu \to 0$ (i.e., the right-hand term in braces does not reduce to $3\mu + 1$).

In conclusion, besides variance inflation, SA impacts upon the skewness and the kurtosis of Poisson RVs. As with normal and beta RVs, SA impacts are more acute for surface partitioning that do not form a square tessellation. Continuing the preceding regular square tessellation example, but for a Poisson RV and the spatial filter set to SF/3, suppose $\mu = 20$ (the minimum value at which a normal approximation is considered very good). Figure 2e portrays a simulated outcome for this case; the insert depicts the unautocorrelated distribution. In this case, SA dramatically increases the number of 0s, reducing the minimum count from 7 to 0, while increasing the maximum count from 35 to 254, and transforms the shape to one that looks more like an exponential RV; the resulting generalized linear model deviance statistic is roughly 44 (indicating excessive extra-Poisson variation). Affiliated summary statistics for this simulation appear in Table 1. Meanwhile, yet again as an example of irregular surface partitioning data, consider the county partitioning of China. Figure 3e portrays a simulated outcome for this case; the insert depicts the unautocorrelated distribution. The deviance statistic quantifying extra-Poisson variation is approximately 5. As with the square tessellation, SA dramatically increases the number of 0s, reducing the minimum count from 7 to 0, while increasing the maximum count from 37 to 873. Affiliated summary statistics for this simulation appear in Table 2.

6. THE BINOMIAL RV

Griffith (2004) presented a respecification of the auto-logistic/binomial model that expresses the expected value of the log-odds ratio as the sum of a constant and an eigenvector spatial filter. This specification allows the probability of an individual observation i to be rewritten as

$$p_i = \frac{1}{1 + e^{\ln[(1-p)/p]}/\theta_i}, \tag{7}$$

where $\theta_i = e^{\mathbf{E}_i \beta + \eta \ln(\sum_{i=1}^{n} e^{\mathbf{E}_i \beta}/n)}$ -the generalized multiplicative specification-and p is the average population probability across a map. The parameter η ensures that $\sum_{i=1}^{n} p_i/n = p$. At least for simulation purposes, this parameter can be estimated with software such as SAS PROC NLMIXED. In data analysis situations, it becomes part of the estimated intercept term. Meanwhile, the marginal MGF affiliated with Equation (7) is

$$M(\mathbf{0}^\top, t_i, \mathbf{0}^\top) = \sum_{y_i=1}^{N_{tr}} e^{y_i t_i} \mathrm{pmf_i} = \left[\frac{p\,e^{t_i}}{p + \frac{1-p}{\theta_i}} + \frac{1-p}{\theta_i^2} + \frac{1-p}{\theta_i} \right]^{N_{tr}}.$$

With this multiplicative specification, the expected value of a map mean becomes

$$\mathrm{E}\left(\frac{1}{n} \sum_{i=1}^{n} Y_i \right) = \frac{N_{tr}p}{n} \frac{\sum_{i=1}^{n} \theta_i}{(1-p) + p\theta_i} = N_{tr}p,$$

where the value of η is selected to ensure that this equality holds. Once more, a sample histogram tends to be distributed around the population map mean, except for sampling error.

The variance of a set of n independent map values is given by the average across all locations of the expected value of the second marginal moment about the mean, or

$$\mathrm{E}\left(\frac{1}{n} \sum_{i=1}^{n} \{Y_i - N_{tr}p\}^2 \right) = N_{tr}p(1-p) \left[\frac{1}{n} \sum_{i=1}^{n} \frac{N_{tr}p(1-p)(1+\theta_i^2) + [1 - 2N_{tr}p(1-p)]\,\theta_i}{[(1-p) + p\theta_i]^2} \right].$$

The VIF is $\left[(1/n) \sum_{i=1}^{n} \left\{ N_{tr}p(1-p)(1+\theta_i^2) + [1 - 2N_{tr}p(1-p)]\,\theta_i \right\} / [(1-p) + p\theta_i]^2 \right]$. This term reduces to 1 when SA is zero (i.e., $\beta_j = 0, \forall j$).

Based upon the third marginal moment about the mean, here skewness is given by

$$\alpha_3 = [N_{tr}p(1-p)]^{-1/2}$$
$$\left\{ \sqrt{n} \sum_{i=1}^{n} \frac{N_{tr}^2 p^2 (1-p)^2 (\theta^3 - 1) - p\left[1 - 3N_{tr}(1-p) + 3N_{tr}^2 p(1-p)^2 \right] \theta_i^2}{[(1-p) + p\theta_i]^3} \right.$$
$$\left. + \sqrt{n} \sum_{i=1}^{n} \frac{(1-p)\left[1 - 3N_{tr}p + 3N_{tr}^2 p^2 (1-p) \right] \theta_i}{[(1-p) + p\theta_i]^3} \right\} \Bigg/$$
$$\left\{ \left[\sum_{i=1}^{n} \frac{N_{tr}p(1-p)(1+\theta_i^2) + [1 - 2N_{tr}p(1-p)]\,\theta_i}{[(1-p) + p\theta_i]^2} \right]^{3/2} \right\},$$

indicating that SA tends to distort the skewness of a binomial RV. When zero SA is present, the right-hand term reduces to $(1-2p)$, and $\alpha_3 = 0$ when $p = 1/2$ or as $N_{tr} \to \infty$,

properties that are not preserved in the presence of non-zero SA. Finally, kurtosis is

$$
\begin{aligned}
\alpha_4 = \Bigg\{ & n \sum_{i=1}^{n} \frac{N_{tr}^3 p^3 (1-p)^3 (1+\theta_i^4)}{\left[(1-p)+p\theta_i\right]^4} \\
& + n \sum_{i=1}^{n} \frac{1 - 4N_{tr} + 6N_{tr}^2 + 4N_{tr}(1 - 3N_{tr} - N_{tr}^2)p + 6N_{tr}^2(1 + 2N_{tr})p^2 - 12N_{tr}^3 p^3 + 4N_{tr}^3 p^4}{p^{-2}\theta_i^{-3}\left[(1-p)+p\theta_i\right]^4} \\
& - n \sum_{i=1}^{n} \frac{p(1-p)\left[4 - 7N_{tr} + 12N_{tr}^2 p - 6N_{tr}^2(N_{tr}+2)p^2 + 12N_{tr}^3 p^3 - 6N_{tr}^3 p^4\right]\theta_i^2}{\left[(1-p)+p\theta_i\right]^4} \\
& + n \sum_{i=1}^{n} \frac{(1-p)^2\left[1 - 4N_{tr}p + 6N_{tr}^2 p^2 - 4N_{tr}^3 p^3 + 4N_{tr}^3 p^4\right]\theta_i}{\left[(1-p)+p\theta_i\right]^4} \Bigg\} \Bigg/ \\
& \left\{ N_{tr}p(1-p)\left[\sum_{i=1}^{n} \frac{N_{tr}p(1-p)(1+\theta_i^2) + \left[1 - 2N_{tr}p(1-p)\right]\theta_i}{\left[(1-p)+p\theta_i\right]^2}\right]^2 \right\},
\end{aligned}
$$

indicating that SA also distorts the kurtosis of a binomial RV. In addition, α_4 no longer converges on 3 as $N_{tr} \to \infty$.

Therefore, in the presence of non-zero SA, a binomial histogram behaves similar to what is portrayed by both the beta RV (although for discrete values) and the Poisson RV. If $p = 1/2$, the binomial histogram tends to: (1) increasingly mimic a bell-shaped curve as N_{tr} increases, for low levels of SA; mimic a discrete uniform distribution for intermediate levels of SA; and, mimic a sinusoidal RV for high levels of SA (i.e., mostly 0 and N_{tr} values). As p moves toward 0, the binomial histogram behaves more like that for a Poisson RV, except that it is constrained to have a maximum count value, N_{tr} (i.e., extremes tend to be smaller).

In conclusion, besides variance inflation, SA impacts upon the skewness and the kurtosis of binomial RVs. As with normal, beta, and Poisson RVs, SA impacts are more acute for surface partitioning that do not form a regular square tessellation. Continuing the preceding regular square tessellation example, but for a binomial RV, for $p = 1/2$, the original spatial filter has $\eta = 0.06405$; for $p = 1/10$, $\eta = 0.54660$. The former histogram resembles Figure 2c, whereas the latter histogram resembles Figure 2e. For $p = 1/2$ and SF/2.5, the histogram for a simulated outcome resembles a discrete uniform distribution (Figure 2f); the resulting generalized linear model deviance statistic is roughly 5 (i.e., substantial extra-binomial variation). Affiliated summary statistics for this simulation appear in Table 1. Meanwhile, as an example of irregular surface partitioning data, once more consider the county partitioning of China, and consider the new spatial filter, SF $= 4E_4 + 2E_9 + E_{10}$; asymmetries in the original spatial filter eigenvector elements cause marked distortion only in one tail of the resulting histogram. For $p = 1/2$, this spatial filter has $\eta = 0.00108$. The resulting generalized linear model deviance statistic for a simulated outcome is roughly 5 (extra-binomial variation). The counterpart to Figure 2f is Figure 3f. And, affiliated summary statistics for this simulation appear in Table 2.

7. Methodology for Spatial Scientists

The preceding discussion implies that a spatial scientist should be interested in the nature of a RV uncontaminated by SA. Regression analysis furnishes a tool for identifying an underlying spatially independent frequency distribution. The approach outlined here parallels that for removing trends with classical linear models to obtain normally distributed residuals.

7.1 Recovering the underling unautocorrelated histogram

Recovering the underling histogram for a normal RV is easy. The spatial filter is a linear combination of eigenvectors constituting a non-constant part of the mean response. The underlying histogram can be constructed by regressing the response variable Y on the appropriate set of eigenvectors, and then constructing a histogram with the residuals plus the intercept term. A synthetic positive spatially autocorrelated normal RV, with a mean of 0 and a variance of 1, was generated with the SAS pseudo-random number generator and a spatial filter for a 50-by-40 regular square tessellation. Figure 4 shows the way the realization changes from its initial normal distribution (Figure 4a), to its positively spatially autocorrelated distribution (Figure 4b), and then to its recovered underlying normal distribution (Figure 4c). Figure 4d is a scatterplot constructed with the original and recovered value pairs, and reveals a very close correspondence between the two (the corresponding bivariate linear regression has nearly 100% of the variation in the initial values accounted for by the recovered values, yields an intercept of 0, and a slope of 1). Summary statistics confirm that the recovered histogram (the estimated spatial filter accounts for roughly 96% of the variance) for this simulation example more closely conforms to a normal frequency distribution; see Table 3.
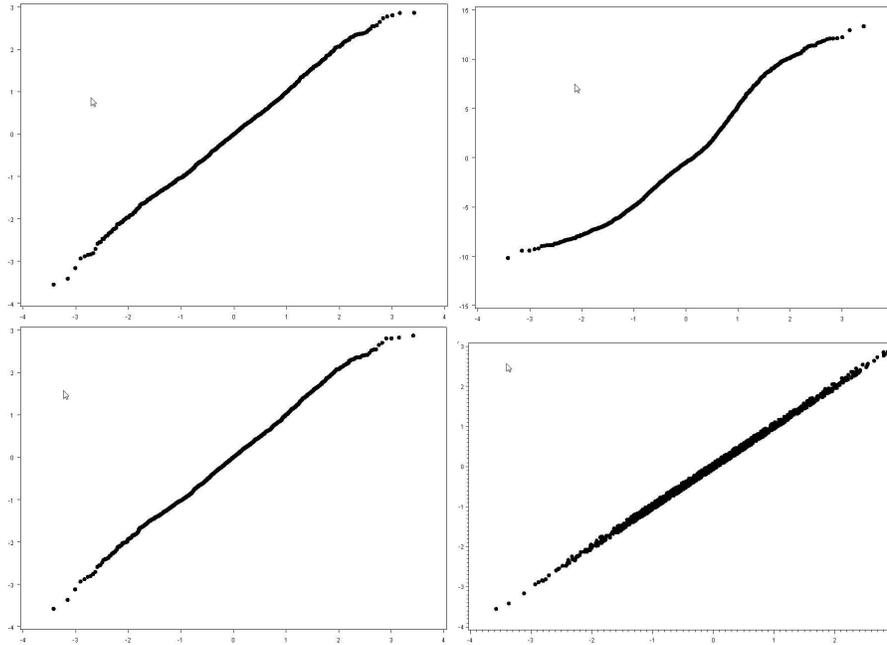

Figure 4. Left (a): normal quantile plot for original values. Left-middle (b): normal quantile plot after strong positive spatial autocorrelation embedded into original values. Right-middle (c): normal quantile plot for recovered values. Right (d): scatterplot of original and recovered value pairs.

Table 3. Summary statistics the simulation example.

Statistic	Original values	Spatially autocorrelated values	Recovered values
Shapiro-Wilk probability	0.4078	< 0.0001	0.4686
Skewness	0.0320	0.4270	0.0251
Kurtosis	2.9768	2.5868	2.9827

Because this simulation employed a regular square tessellation, the increase in skewness is surprisingly high; its theoretical value is 0.4196. The change decrease kurtosis is as anticipated; its theoretical value is 2.5950. In addition, the variance inflation attributed to the embedded positive spatial autocorrelation is roughly 22.

Recovering the underling histogram for a Poisson RV is more difficult, in part because one SA impact on this RV is a tendency to create excess zeroes, and in part because it involves a discrete RV. Fortunately, this former complication can be handled with a zero-inflated Poisson regression model. This latter complication introduces some distortion in the recovered results. The recovered histogram can be constructed with the following data analysis steps:

STEP 1. Calculate expected values with a zero-inflated Poisson regression of a georeferenced RV on a candidate set of eigenvectors,

STEP 2. Estimate Poisson probabilities for these expected values,

STEP 3. Use the estimated probabilities and adjusted intercept term alone to compute the corresponding counts with a CDF, which then are used to construct the histogram.

Figure 5 portrays the results for a simulated example. Although the recovered histogram has an inflated mean, its variance is correct, and it is much better behaved vis-à-vis a Poisson RV.
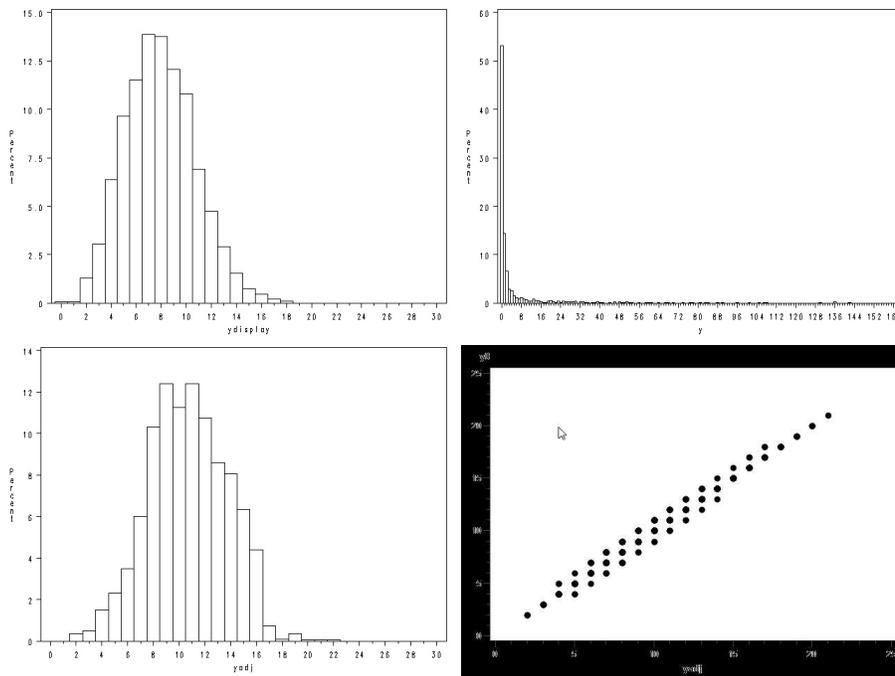

Figure 5. Left (a): simulated Poisson counts histogram. Left-middle (b): positive spatially autocorrelated Poisson counts histogram. Right-middle (c): spatial filter recovered Poisson counts histogram. Right (d): scatterplot of original and recovered value pairs.

Finally, recovering the underling histogram for a binomial RV is even more difficult, in part because a correction factor that needs to be estimated is involved, and in part because both 0 and N_{tr} can become inflated. Fortunately, this correction factor appears to converge on 1 in the limit, although this convergence seems to be hampered by irregularities in surface partitioning. Unfortunately, N_{tr}-inflated binomial model estimation techniques do not exist. The recovered histogram for a simulated binomial RV with $N_{tr} = 50$ and $p = 0.26801$ more closely conform to a binomial distribution (Figure 6). The recovered histogram can be constructed with the following data analysis steps:

STEP 1. Standardize to $N_{tr} = 1,000$ (exploits asymptotics for correction factor),

STEP 2. Use beta distribution to redistribute $y = N_{tr}$ inflated values,

STEP 3. Utilize the estimated probabilities and adjusted intercept term alone to compute the corresponding counts with a CDF, which then are used to construct the histogram.

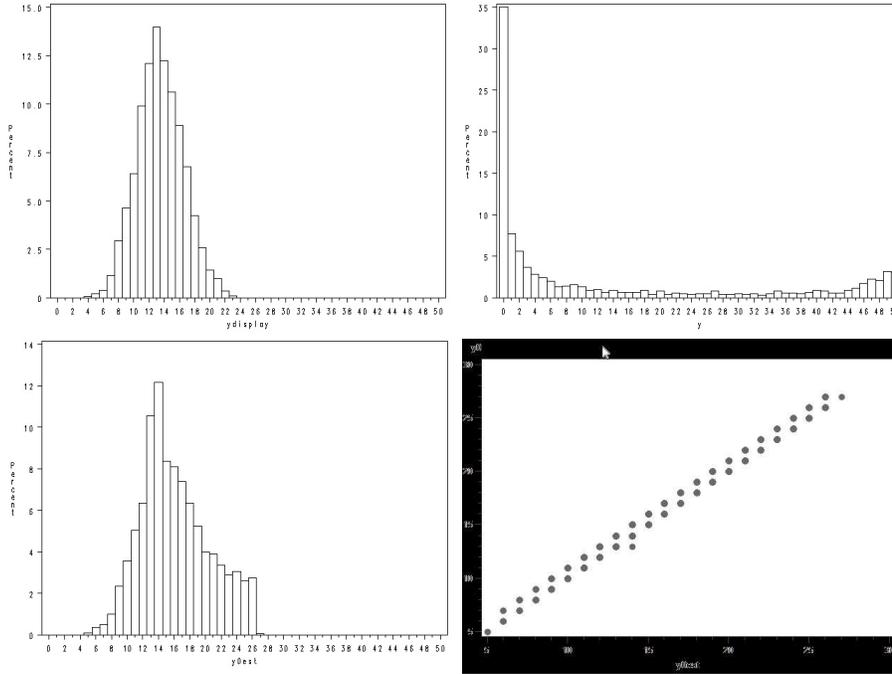Again the recovered histogram is much better behaved vis-à-vis a binomial RV.

Figure 6. Left (a): simulated binomial counts histogram. Left-middle (b): positive spatially autocorrelated binomial counts histogram. Right-middle (c): spatial filter recovered binomial counts histogram. Right (d): scatterplot of original and recovered value pairs.

8. Discussion and Implications

Based upon the eigenvector spatial filter probability model specification, the preceding analyses suggest the following conjecture:

Conjecture: Even if positive SA does not introduce bias into a map mean for a georeferenced RV, it inflates the RV's map variance, and it tends to distort the RV's skewness and kurtosis, rendering a misleading histogram constructed with observed georeferenced data when non-zero SA is overlooked.

Four specific instances of this conjecture are the following theorems.

Theorem 8.1 Positive SA does not introduce bias into a map mean for a georeferenced normal RV, but inflates its variance, and alters its skewness and kurtosis, rendering a misleading histogram constructed with observed georeferenced data when non-zero SA is overlooked.

Proof

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right)=\mu;\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\mu\right)^2\right)=\sigma^2\mathrm{VIF},\quad\mathrm{VIF}\neq 1;$$

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i-\mu\}^3\right)\neq 0;\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i-\mu\}^4\right)\neq 3.$$

This is the most straightforward RV result. ∎

THEOREM 8.2 *Positive SA does not introduce bias into a map mean for a georeferenced beta RV with parameters $\alpha > 0$ and $\gamma > 0$, but inflates its variance, and alters its skewness and kurtosis, rendering a misleading histogram constructed with observed georeferenced data when non-zero SA is overlooked.*

PROOF

$$E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{\alpha}{\alpha+\gamma};$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^2\right) = \left\{\frac{\alpha\gamma}{(\alpha+\gamma)^2(\alpha+\gamma+1)}\right\}\text{VIF}, \quad \text{VIF} \neq 1;$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^3\right) \neq \frac{2(\gamma-\alpha)\sqrt{\alpha+\gamma+1}}{(\alpha+\gamma+2)\sqrt{\alpha\gamma}};$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^4\right) \neq \frac{(\alpha+2)(\alpha+3)(\alpha+\gamma)(\alpha+\gamma+1)}{\alpha(\alpha+1)(\alpha+\gamma+2)(\alpha+\gamma+3)}.$$

∎

This beta RV treatment appears to be one of the few attempts to formulate a beta RV specification that accounts for SA; see also Kaiser et al. (2002); Hardouin and Yao (2008). Its principal advantage is that it mimics many other RV histograms, making it an important but overlooked spatial model supporting the preceding conjecture, and as such furnishes a heuristic tool for exploring the impacts of SA on numerous RVs. In addition, although an asymptotic normal approximation remains a property, convergence is slowed by positive SA.

THEOREM 8.3 *Positive SA does not introduce bias into a map mean for a georeferenced Poisson RV, but inflates its variance, and alters its skewness and kurtosis, rendering a misleading histogram constructed with observed georeferenced data when non-zero SA is overlooked.*

PROOF

$$E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \mu; E\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^2\right) = \mu\text{VIF}, \text{VIF} \neq 1;$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^3\right) \neq \frac{1}{\sqrt{\mu}}; E\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^4\right) \neq 3 + \frac{1}{\mu}.$$

∎

This result is particularly useful because the Poisson, exponential, and gamma auto-models tend to be of little practical interest because they can be applied only to spatial competition situations. Kaiser and Cressie (1997) offered a specification to circumvent this restriction. In contrast, the eigenvector spatial filter specification furnishes probability models for a wide range of spatial cooperation situations involving count data, and as such appears to be superior to this alternative specification. Finally, the following theorem is established.

THEOREM 8.4 *Positive SA does not introduce bias into a map mean for a georeferenced binomial RV with parameters p and N_{tr} (i.e., the number of trials), but inflates its variance,*

and alters its skewness and kurtosis, rendering a misleading histogram constructed with observed georeferenced data when non-zero SA is overlooked.

PROOF

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right) = N_{tr}p;\, \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^2\right) = N_{tr}p(1-p)\mathrm{VIF}, \quad \mathrm{VIF} \neq 1;$$

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^3\right) \neq (1-2p)/\sqrt{N_{tr}p(1-p)};$$

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \mu\}^4\right) \neq \left[1 + 3p(1-p)(N_{tr} - 2)\right]/\left[N_{tr}p(1-p)\right].$$

∎

For Theorems 3 and 4, the asymptotic normal approximation property also is lost in the presence of positive SA.

In contrast, the central limit theorem for $\overline{y}$ still holds, although the rate of convergence is slowed by the presence of positive SA. This finding is in keeping with Mardia and Marshall (1984).

THEOREM 8.5 Let Y_i, $i = 1,\ldots,n$, be n real values tagged to n regular square lattice locations forming a rectangular region on a map, each with a finite mean, μ, and variance, and with the map containing positive SA (i.e., the Y_i display nearby location correlation). Given an eigenvector spatial filter specification of the n marginal pdf$_i$/pmf$_i$s, as the number of locations increases (either infill or increasing domain asymptotics), the distribution of the sample map mean approaches a normal distribution with mean μ and variance $\sigma^2\mathrm{VIF}/n$ regardless of whether the original Y_i are normal, beta, Poisson or binomial RVs.

PROOF

$$\mathrm{E}\left(\overline{y}\right) = \mu;\, \mathrm{E}\left(\{\overline{y} - \mu\}^2\right) = \sigma^2\mathrm{VIF}/n,\ \mathrm{VIF} \neq 1;$$

$$\lim_{n\to\infty}\mathrm{E}\left(\{\overline{y} - \mu\}^3\right) = \frac{\lim_{n\to\infty}\mu_3^*}{\sqrt{n}\left(\sigma^2\mathrm{VIF}\right)^{3/2}} = 0;\, \frac{\lim_{n\to\infty}\mathrm{E}\left(\{\overline{y} - \mu\}^4\right)}{\left(\sigma^2\mathrm{VIF}\right)^2} = 3,$$

given Lemma 2, where μ_k^* are the kth moments about the mean when positive SA is present. This Lemma 2 also implies that the sums of the third and of the fourth powers of a spatial filter across a map converge to finite values in the limit; see Appendix. ∎

The variance of the sampling distribution decreases as n increases, but more slowly than if VIF = 1. Skewness converges on 0, but more slowly than if VIF = 1 and $\mu_3^* = \mu_3$. And, convergence on a normal distribution is slower in the presence of positive SA. This result relates to the reduction in degrees of freedom from n (i.e., the effective degrees of freedom) for a sampling distribution when positive SA is present in data; see, e.g., Clifford et al. (1989).

In conclusion, the eigenvector spatial filter mixture model specification outlined here furnishes a relatively simple way to account for SA in georeferenced data for any valid marginal distribution [1]. It also reveals that constructing histograms with such data can

---

[1] Griffith and Peres-Neto (2006) showed that this eigenvector formulation generalizes from the binary matrices used in this paper to distance-based neighbor matrices. The conceptualization parallels that presented in Hodges and Reich (2010).

yield misleading suggestions about the RVs under study when SA is overlooked. Consequently, spatial scientists need to initiate descriptive statistical analyses based upon recovered frequency distributions that remove impacts of latent SA in their georeferenced data. In addition, future research needs to improve the independent and identically distributed (i.e., iid) RV recovery technique, which for non-normal RVs is slightly distorted in the right-hand tails (see Figures 5 and 6), and yields a slightly inflated mean.

APPENDIX: CONVERGENCE VALUES FOR THE ILLUSTRATIVE ADDITIVE SPATIAL FILTER, SF $= 4E_1 + 2E_2 + E_3$

(A1) $\lim_{n \to \infty} \sum_{j=1}^{P} \sum_{k=1}^{Q} SF_{jk} = 0,$

(A2) $\lim_{n \to \infty} \sum_{j=1}^{P} \sum_{k=1}^{Q} SF_{jk}^2 = \sum_{h=1}^{3} \beta_h^2 = 4^2 + 2^2 + 1^2 = 21,$

(A3) $\lim_{n \to \infty} \sum_{j=1}^{P} \sum_{k=1}^{Q} SF_{jk}^3 = 6 \sum_{j=1}^{P} \sum_{k=1}^{Q} \beta_f \beta_g \beta_h e_{jk,f} e_{jk,g} e_{jk,h} = \frac{6(4)(2)(1)(2048)}{(255\pi^2)} = 44.2679,$

(A4) $\lim_{n \to \infty} \sum_{j=1}^{P} \sum_{k=1}^{Q} SF_{jk}^4 = \frac{9}{4} \sum_{h=1}^{3} \beta_h^4 + 6 \sum_{g=1}^{2} \sum_{h=g+1}^{3} \beta_g^2 \beta_h^2$

$$= \frac{9}{4}(4^4 + 2^4 + 1^4) + 6(4^2 2^2 + 4^2 1^2 + 2^2 1^2) = 1126.97.$$

The respective values for the 50-by-40 square tessellation are as follows:

(A1): 0,
(A2): 21,
(A3): 43.29,
(A4): 1118.25.

(A3) and (A4) in this numerical example are very close to their asymptotic values.

REFERENCES

Anselin L., 1995. Local indicators of spatial association - LISA. Geographical Analysis, 27, 93-115.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of The Royal Statistical Society Series B - Statistical Methodology, 36, 192-326.

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

Besag, J., York, J., Molli, A., 1991. Bayesian image restoration, with two applications in spatial statistics, with discussion. Annals of the Institute of Statistical Mathematics, 43, 1-59.

Besag, J., Green, P., Higdon, D., Mengersen, K., 1995. Bayesian computation and stochastic systems. Statistical Science, 10, 3-41.

Burridge, P., 1980. On the Cliff-Ord test for spatial autocorrelation. Journal of The Royal Statistical Society Series B - Statistical Methodology, 42, 107-8.

Burridge, P., 1981. Testing for a common factor in a spatial autoregressive model. Environment and Planning A, 13, 795-800.

Clayton, D., Kaldor, J., 1987. Empirical Bayes estimates of age-standardized relative risk for use in disease mapping. Biometrics, 43, 671-681.

Cliff, A., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.

Clifford, P., Richardson, S., Hmon, D., 1989. Assessing the significance of the correlation between two spatial processes. Biometrics, 45, 123-134.

de Jong, P., Sprenger, C., van Veen, F., 1984. On extreme values of Moran's I and Geary's c. Geographical Analysis, 16, 17-24

Dutilleul, P., Legendre, P., 1992. Lack of robustness in two tests of normality against autocorrelation in sample data. Journal of Statistical Computation and Simulation, 42, 79-91.

Griffith, D., 1996. Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. Canadian Geographer, 40, 351-367

Griffith, D., 2000a. Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. Linear Algebra and its Applications, 321, 95-112

Griffith, D., 2000b. A linear regression solution to the spatial autocorrelation problem. Journal of Geographical Systems, 2, 141-156

Griffith, D., 2002. A spatial filtering specification for the auto-Poisson model. Statistics and Probability Letters, 58, 245-251

Griffith, D., 2004. A spatial filtering specification for the auto-logistic model. Environment and Planning A, 36, 1791-1811.

Griffith, D., Peres-Neto, P., 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. Ecology, 87, 2603-13.

Hodges, J., Reich, B., 2010. Adding spatially-correlated errors can mess up the fixed effect you love. The American Statistician, 64, 325-339.

Hardouin, C., Yao, J-F., 2008. Multi-parameter automodels and their applications, Biometrika. 95, 335-349.

Kaiser, M., Cressie, N., 1997. Modeling Poisson variables with positive spatial dependence. Statistics and Probability Letters, 35, 423-432.

Kaiser, M., Cressie, N., Lee, J., 2002. Spatial mixture models based on exponential family conditional distributions. Statistica Sinica, 12, 449-474.

McKenzie, E., 1985. An autoregressive process for beta random variables. Management Science, 31, 988-997.

Mardia, K., Marshall, R., 1984. Maximum likelihood estimation for models of residual covariance in spatial regression. Biometrics, 41, 287-294.

Ord, J.K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association, 70, 120-126.

Tatsuoka, M., 1988. Multivariate Analysis, 2nd Edition, with contributions by P., Lohnes. Mcmillan, New York.

Tiefelsdorf, M., Boots, B., 1995. The exact distribution of Moran's I. Environment and Planning A, 27, 985-999.