

# ON THE EIGENVALUE DISTRIBUTION OF ADJACENCY MATRICES FOR CONNECTED PLANAR GRAPHS

DANIEL A. GRIFFITH

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

Manuscript received: February 2, 2015

Revised version: October 10, 2015

GRIFFITH D.A., 2015. On the eigenvalue distribution of adjacency matrices for connected planar graphs. *Quaestiones Geographicae* 34(4), Bogucki Wydawnictwo Naukowe, Poznań, pp. 39–60, 4 tables, 10 figs. DOI 10.1515/quageo-2015-0035, ISSN 0137-477X.

**ABSTRACT:** This paper describes the previously unknown statistical distribution of adjacency matrix spectra for planar graphs, also known as spatial weights matrices, in terms of the following three readily available eigenvalue properties: extremes, rank orderings, and sums of powers. This distribution is governed by at most six parameters that, once known, allow accurate approximations of eigenvalues to be computed without resorting to numerical matrix methods applied on a case-by-case basis. Parameter estimates for illustrative real-world examples are obtained using nonlinear least squares regression techniques. Three conjectures are proposed, and graphical and trend results are reported for a diverse set of planar graph-based matrices.

**KEY WORDS:** adjacency matrix, connected graph, eigenvalue distribution, planar graph, serial structure

*Daniel A. Griffith, School of Economic, Political, and Policy Sciences, University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080, USA; e-mail: dagriffith@utdallas.edu*

Spatial scientists frequently analyze data tagged to a surface tessellation, whose dual graph is used to construct a binary 0–1 adjacency matrix  $C$  which, sometimes together with its eigenfunctions, is employed as a spatial weights matrix in geographical data analyses. This exercise is becoming increasingly more common as geographic information systems (GISs) become more widely used. Accordingly, the matrix is  $n$ -by- $n$  in size, where  $n$  denotes the number of nodes (i.e., areal units in a surface partition, or polygons in a shapefile) and frequently its elements are defined such that  $c_{ij} = 1$  if nodes  $i$  and  $j$  in a graph are connected, and  $c_{ij} = 0$  otherwise; almost always,  $c_{ii} = 0$  by definition. Adjacencies defined in terms of non-zero length common boundaries being shared by areal units partitioning a surface (i.e., contiguity; analogous to the rook's move in chess) result in a planar graph. Frequently matrix  $C$  is converted to its

row-standardized version, say matrix  $W = D^{-1}C$ , where  $D$  is the diagonal matrix whose  $d_{ii}$  element is the  $i$ th row sum of matrix  $C$  (this specification relates to the Laplacian matrix discussed in Chung et al. 2003). Although the foundations of spectral graph theory were laid in the mid-20<sup>th</sup> century, and advances in computer technology continue to expand the numerical applications realm for this theory, calculating eigenvalues when the resulting adjacency matrix is constructed for a very large to massively large number of nodes forming a connected graph<sup>1</sup> (i.e., a graph

<sup>1</sup> If a tessellation comprises disjoint subtessellations, then the eigenvalues for each subtessellation can be calculated (matrices  $C$  and  $W$  become block diagonal), and then combined to construct the full set. This situation simplifies the eigenfunction problem by reducing the size of the matrices involved. It also characterizes a geographic landscape in which indirect effects of spatial dependency cannot percolate through the entire landscape. Most spatial scientists deal with

in which a path can be traced from any given node to any other node in the graph) remains a formidable to impossible task. The degree of difficulty associated with solving this eigenfunction problem in part has motivated scholars to study and derive properties of the distributions of spectra (e.g., Cao, Yuan 1993, 1995; Yong 1999; Liu, Bo 2000; Adler, van Moerbeke 2001; Griffith 2000, 2004; Griffith, Luhanga 2011). Chung (1997: 6) describes this situation by noting that “half of the main problems of spectral theory lie in deriving bounds on the distributions of eigenvalues.” This paper adds to that literature, with special reference to matrix  $\mathbf{W}$ , and emphasis on the rook’s definition of geographic adjacency.

## The state of the problem solution

Determining matrix eigenfunctions continues to be an important and highly relevant area of numerical linear algebra research. Because the sparse matrices most frequently employed in geographical analyses are symmetric, their salient properties (e.g., orthogonal eigensystems, real eigenfunctions) simplify computation of the relevant eigenvalues and eigenvectors. But lack of a patterning of the ones in many of these matrices (i.e., they are for irregular surface partitionings) or the conversion of matrix  $\mathbf{C}$  to  $\mathbf{W}$  complicates this computing. To this end, Golub and van der Vorst (2000: 59) optimistically contend that the eigenfunction problem is solved: “for small [ $n$ -by- $n$ ] matrices [where]  $n \leq 25$  we have the QR method, one of the most elegant numerical techniques produced in the field of numerical analysis; for larger matrices (but smaller than a few thousand), we have a combination of divide and conquer with QR techniques. For the largest matrices, there is the Lanczos method.” But as Hams and de Raedt (2000) point out, large matrices mean those that do not exceed roughly  $n = 10,000$ . Consequently, the Lanczos method was of little use for their  $n = 32,768$  or  $n = 16,777,216$  cases. Furthermore, even for a massively large regular lattice with random linkages, Khorunzhy et al. (2004: 1649) note that “the spectral theory of random graphs (...) is still poorly explored.” In

some ways, their work is reminiscent of that by Barry and Pace (1999) for geographical analyses. One noteworthy difference between findings reported in these latter two papers and results most often needed by spatial scientists is that these studies deal with random graphs or strictly with regular lattices.

Chung et al. (2003) note that the eigenvalues of connected random graphs – which currently are of particular interest because of what these quantities reveal about the internet (see Faloutsos et al. 1999) and other forms of spatial connectivity (see Fefferman, Phong 1980) – follow a power-law distribution. But this is only part of the description. Consider the  $\mathbf{W}$  matrix representation of the surface partitioning of Amazon Brazil into its 323 districts (see Fig. 1). This specific example is selected because its eigenvalue distribution is challenging to describe (specifically, a sizeable gap exists between  $\lambda_{n-1}$  and  $\lambda_n = \lambda_{\min}$ , the second

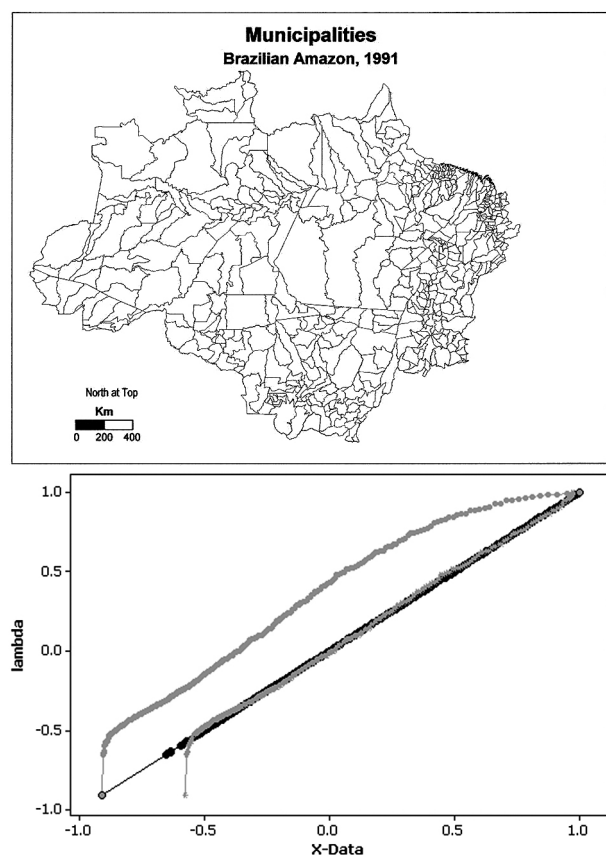


Fig. 1. Results for Amazon Brazil

Top (a): the surface partitioning by municipalities. Bottom (b): eigenvalues (vertical axis: observed; horizontal axis: predicted) for the adjacency graph representation of Amazonian Brazil’s surface partitioning;  $o$  denotes the observed eigenvalues,  $*$  denotes predicted values from equation (1), and  $\bullet$  denotes predicted values from equation (11).

geographic landscapes whose graph theoretical representation is a connected graph.

smallest and smallest eigenvalues). Estimating a sophisticated power-law equation of the form

$$\hat{\lambda} = \alpha + \beta / (r_i^\gamma + \delta)^K, \quad (1)$$

where the estimate of exponent  $K$  often appears to go to 0 (implying a logarithmic transformation denoted by the natural logarithm,  $\text{LN}$ ),  $\lambda_i$  ( $i = 1, 2, \dots, n$ ) are the descending rank-ordered eigenvalues of a graph,  $\hat{\lambda}_i$  are their predicted counterparts,  $r_i$  is the rank (in descending order) of the  $i^{\text{th}}$  eigenvalue, and  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  are parameters of a four-parameter power law (i.e.,  $K = 1$ ). The scatterplot of equation (1) results for Amazon Brazil (denoted in Fig. 1 by asterisks, \*) show that they align reasonably well with the actual eigenvalues, except for the end of the lower tail. However, although the mean and variance of the distribution are matched reasonably well, especially the minimum extreme value fails to be well matched. Often these extreme values play a dominant role in data analyses, and can be estimated exogenously (see Griffith 2000, 2004).

### Latent structure in the eigenvalues for connected irregular planar graphs: an exploratory analysis

A convenience sample was assembled comprising 184 readily available (i.e., found in the literature) empirical surface partitionings constituting connected irregular planar graphs (i.e., the rook's definition of adjacency), ranging in size from 5 to 7,249. Some of these graphs are for regular square tessellations, and some are for Thies-

sen polygon partitioned surfaces, which verge on hexagonal tessellations, whereas the remaining ones are for surface partitionings by administrative geographic units. Statistical descriptions of the eigenvalues for these graphs are useful because the set of eigenvalues for a single graph can be viewed as a population, and hence statistics about them are descriptive rather than inferential. In contrast, all connected planar graphs of a given size  $n$  can be viewed as a population of interest, with a single graph furnishing a sample from that population. As such, statistics about a graph's eigenvalues also may be treated in an inferential context.

### Serial structure in connected irregular planar graphs

The eigenvalues of the adjacency matrix for a connected planar graph may be ordered from smallest to largest,  $\lambda_n \leq \lambda_{n-1} \leq \dots \leq \lambda_2 < \lambda_1$ , imposing a serial structure on them. Time series analysis techniques furnish powerful tools for examining latent structure in such linear series of values. Applying autoregressive-integrated-moving-average (ARIMA) techniques to a set of eigenvalues reveals several properties. Foremost, a first difference needs to be applied to these numerical values, confirming the well-known problem of studying the spacing of eigenvalues. Fig. 2 portrays a typical case, selected on the basis of experience examining geographic weights matrix eigenvalues. The raw eigenvalues (Fig. 2a) suggest a very strong autoregressive structure that all but disappears with first-differencing (Fig. 2b).

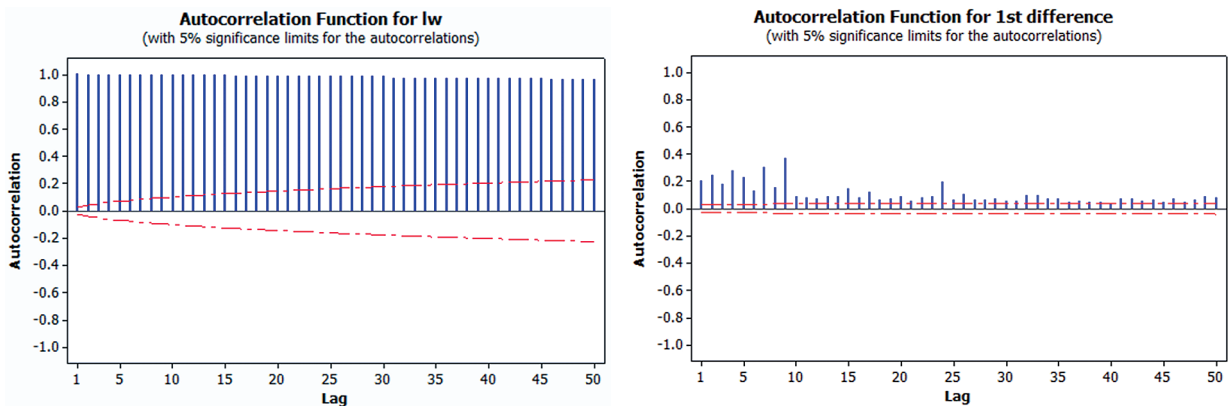


Fig. 2. Correlograms for the eigenvalues of matrix  $W$  for the 1990 Syracuse, NY, census block surface partitioning. Left (a): the raw eigenvalues. Right (b): 1st-differenced eigenvalues.

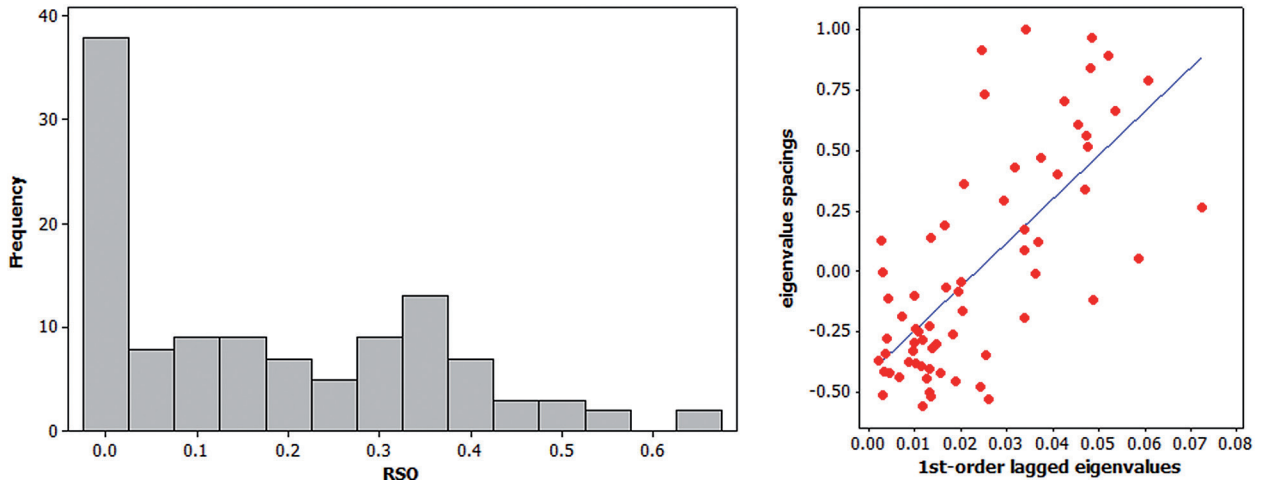


Fig. 3. Latent structure in sequences of matrix  $W$  eigenvalue spacings  
 Left (a): histogram of  $R^2$  values from bivariate regressions predicting eigenvalue spacings from 1st-order lagged eigenvalues. Right (b): scatterplot of the Alabama county tessellation ( $n = 67$ ) results for which  $R^2 = 0.63$ .

Conventional wisdom argues that meaningfully applying time series analysis techniques requires a minimum of 50 sequential values (Box, Jenkins 1976; Tse 1997), with this minimum sample size requirement increasing with the number of parameters to be estimated (e.g., Hyndman, Kostenko 2007). Accordingly, the sample of 184 sets of eigenvalues was reduced to the 115 having at least 50 sequential values. Fig. 3a is the histogram of the  $R^2$  values from a bivariate regression of the eigenvalue spacings regressed on their affiliated lagged values. Eigenvalues for regular square tessellations – even those forming an incomplete rectangular region (e.g., Huffer, Wu 1998) – have an  $R^2$  of 0. Roughly 43 percent of the eigenvalue sets have an  $R^2$  that does not exceed 0.1, indicating that they contain negligible serial correlation in their spacings. One surface partitioning achieves an  $R^2$  of roughly 0.63 (Fig. 3b); it is for an administrative county tessellation. Collectively, these results confirm that eigenvalue spacings tend to display little simple serial structure.

Inspection of a linear landscape helps illustrate why a tendency exists for little simple serial structure in eigenvalue spacings. The matrix  $W$  eigenvalues for this landscape are (Griffith 2000):

$$\{\cos [\pi(k-1)/(n-1)]\}, k = 1, 2, \dots, n \quad (2)$$

The first difference of these consecutive values is

$$\begin{aligned} & \{\cos [\pi(k-1)/(n-1)] - \cos [\pi k/(n-1)]\} = \\ & 2 \sin \left\{ \frac{\pi}{2(n-1)} \right\} \sin \left\{ \frac{\pi(2k-1)}{2(n-1)} \right\}, \\ & k = 1, 2, \dots, n-1 \end{aligned} \quad (3)$$

Fig. 4 furnishes plots of equations (2) and (3) for  $n = 50$ . The spacing is both nonrandom and balanced; it is not simple serial structure. The best that time series analysis tools can do with such data is suggest that analysis should focus on spacings (Fig. 4b), as it almost always does, rather than on raw values (Fig. 4a). One important research question garnering considerable attention addresses the spacing (i.e., gap) between the first and second largest eigenvalues (see Brouwer, Haemers 2012, Chapter 4).

In conclusion, although powerful time series analysis techniques, such as ARIMA models, fail to furnish any insights into the latent serial structure in a sequence of eigenvalues, they do corroborate the focus in the literature on eigenvalue spacings.

### Finite mixture descriptions of connected irregular planar graphs<sup>2</sup>

Statistical descriptions of eigenvalues can take many forms. For matrix  $W$ , the mean and

<sup>2</sup> Earlier work (Martin, Griffith 1998) suggests that the eigenvalue frequency distribution for a connected planar graph can be described as some type of gamma (e.g. truncated) random variable. During research seminars convened at the University of Texas at Dallas in 2006–2008, Dr Michael Tiefelsdorf presented finite beta-mixture descriptions of eigenvalue frequen-



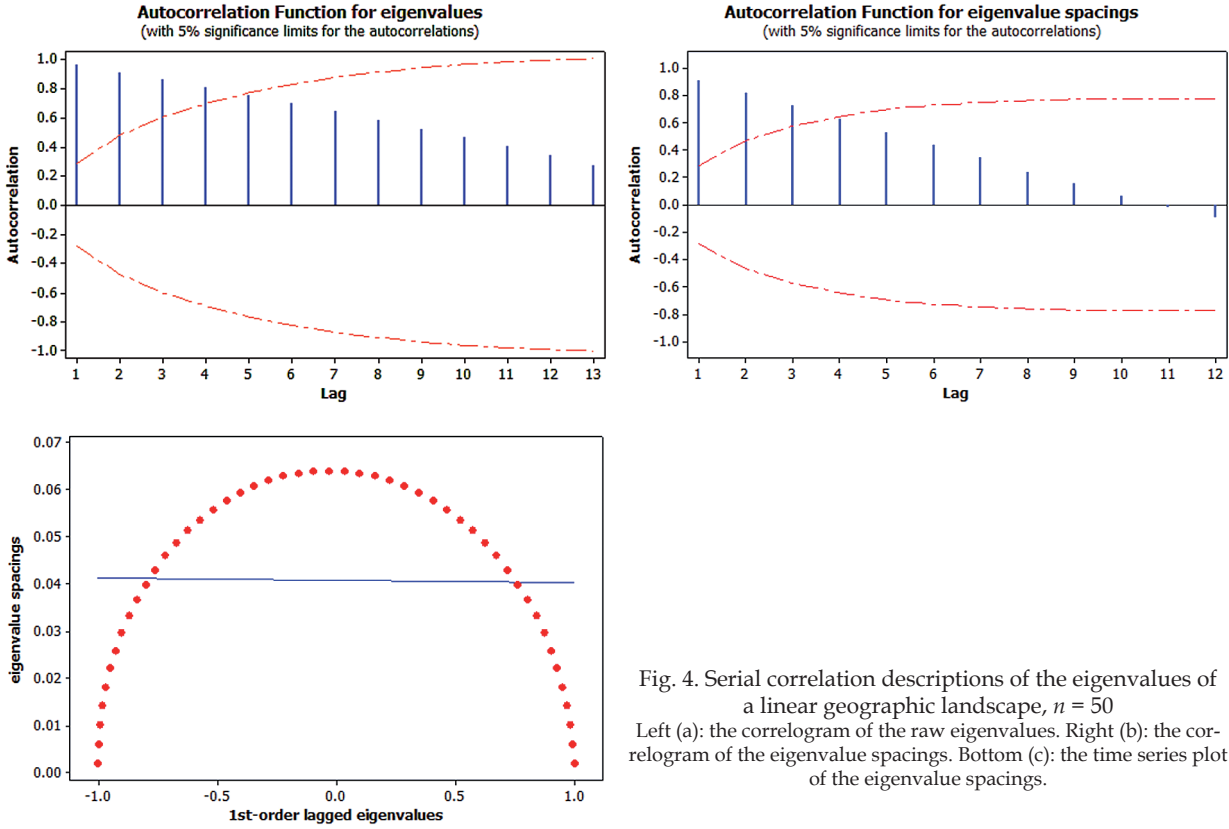


Fig. 4. Serial correlation descriptions of the eigenvalues of a linear geographic landscape,  $n = 50$ . Left (a): the correlogram of the raw eigenvalues. Right (b): the correlogram of the eigenvalue spacings. Bottom (c): the time series plot of the eigenvalue spacings.

variance (i.e., the first and second moments) are known to be, respectively, 0 and  $\mathbf{1}^T \mathbf{D}^{-1} \mathbf{C} \mathbf{D}^{-1} \mathbf{1}$ , where  $\mathbf{1}$  is an  $n$ -by-1 vector of ones, and  $T$  is the matrix transpose operator (Griffith 2000: 102). For symmetric frequency distributions, such as those for a regular square tessellation, all odd moments are known to be 0. Furthermore, for a regular square tessellation forming a complete  $P$ -by- $Q$  rectangular region,  $\mathbf{1}^T \mathbf{D}^{-1} \mathbf{C} \mathbf{D}^{-1} \mathbf{1} = (18PQ + 11P + 11Q + 12)/72$ . Besides moments, a description of the frequency distribution of a set of eigenvalues can be helpful in any effort to approximate them.

Many eigenvalue distributions are positively skewed. Beta, exponential, gamma, inverse Gaussian (i.e., Wald), lognormal, and Weibull are the most popular positively skewed continuous statistical distributions used by applied scientists. Each of these distributions has either a non-negative or a positive support. The beta is perhaps the most flexible of these distributions, and has the added property of having an upper limit; i.e., its support is  $[0, 1]$ . The other distributions need to be truncated to obtain this property.

For example, earlier work attempting to describe the frequency distribution of a set of eigenvalues of connected planar graphs considered the following truncated gamma distribution for random variable  $Y$  as a potential descriptor (Martin, Griffith 1998):

$$f(y; k, y_{\max}, \theta) = \frac{y^{k-1} e^{-y/\theta}}{\theta^k [\Gamma(k) - \Gamma(k, \frac{y_{\max}}{\theta})]}, \quad (4)$$

for  $0 < y \leq y_{\max} \leq \infty$ ,  $k > 0$ , and  $\theta > 0$ ,

where  $f$  is a probability density function,  $y_{\max}$  is an upper bound,  $\Gamma$  is the gamma function, and  $k$  and  $\theta$  are parameters of the gamma distribution. Equation (4) is in keeping with the conceptualization underlying equation (1), succeeding in furnishing an adequate description of the hump but failing to furnish an adequate description of the heavy tail characterizing skewed eigenvalue frequency distributions. This feature of statistical distributions with unbounded support favors use of the beta distribution to describe eigenvalue frequency distributions, which also was considered (Griffith 2003: 49–51).

Because the trace of matrix  $\mathbf{W}$  is 0, the sum of the eigenvalues is 0, and, hence,  $\lambda_1$  for a connected

cy distributions. This section is an extension of those works.

planar graph is positive (by the Perron-Frobenius theorem), implying that at least one eigenvalue must be negative. This property of a set of eigenvalues conflicts with properties of the beta distribution. Fortunately, the extreme eigenvalues of matrix  $\mathbf{W}$  either are known (i.e.,  $\lambda_1 = 1$ ) or can be calculated with numerical techniques (Griffith 2004). Accordingly, the transformed eigenvalues of interest become

$$\frac{\lambda_i - \lambda_n}{1 - \lambda_n}, i = 1, 2, \dots, n. \quad (5)$$

This transformation maps the set of eigenvalues onto the closed interval  $[0, 1]$ , and has a very simple accompanying back-transformation. Because the points 0 and 1 react in a degenerate manner in most software packages, the two modified eigenvalues taking on these specific values can be very slightly perturbed into the interval. Consequently, equation (5) can be treated as if it were a beta random variable.

Using a single beta distribution to describe a skewed eigenvalue frequency distribution

basically is no more successful than using the aforementioned truncated gamma distribution; it fails to furnish an adequate description of the existing heavy tail. This outcome suggests the use of a finite mixture of two or more beta distributions. Just like conventional wisdom suggests a minimum sample size of 50 or more for time series analyses (see the preceding section), it also suggests a minimum sample size of roughly 500 for meaningful finite mixture model results (e.g., Henson et al. 2007). Inspection of such eigenvalue model fitting results (Fig. 5) for sets in the convenience sample under study here having  $n < 535$  corroborates this contention. This restriction further reduces the number of samples available for study to the 35 having at least 500 eigenvalues.

The 1990 Syracuse, NY, census block surface partitioning furnishes the largest of the sets of eigenvalues in the convenience sample, with  $n = 7,249$ . Fitting a 3-beta distributions finite mixture model to it results in one resembling a bell-shaped curve (Fig. 6a) that is not statistically significant. In contrast, both components of a 2-beta distributions finite mixture model (Fig.

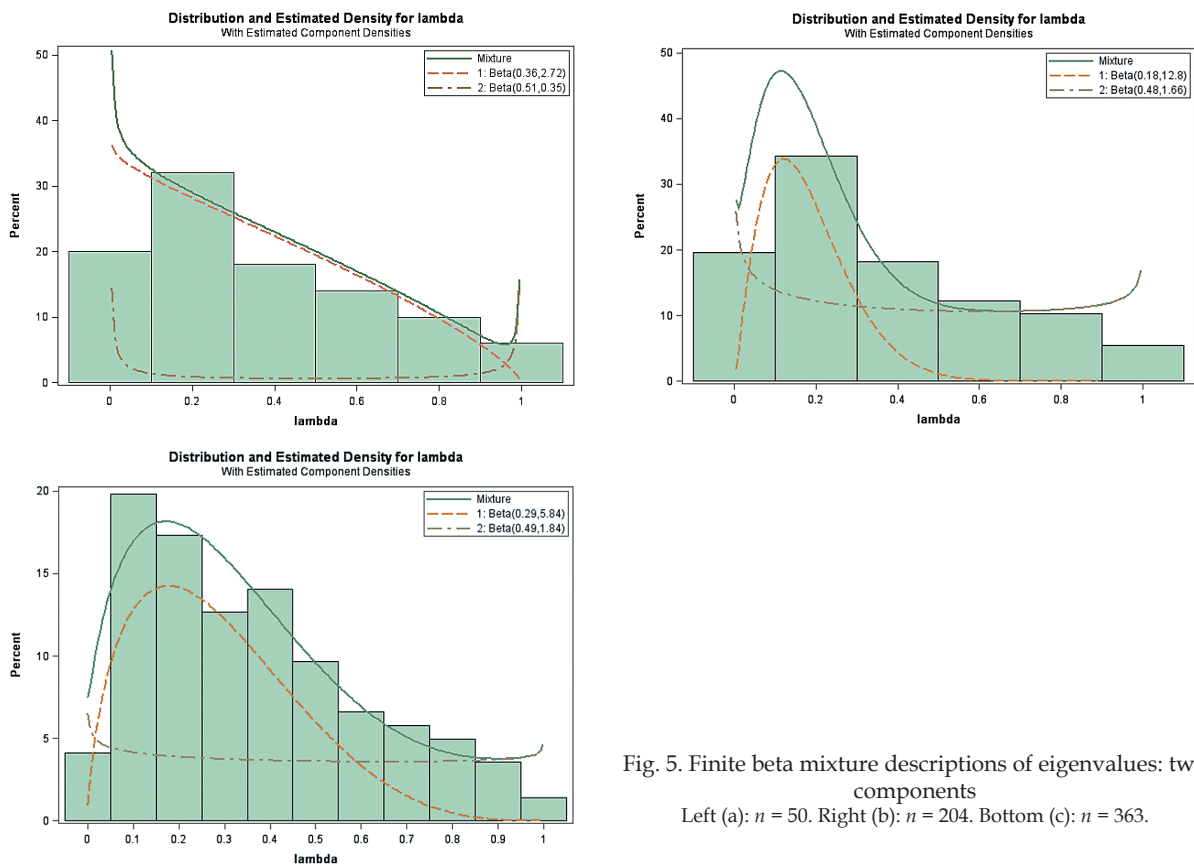


Fig. 5. Finite beta mixture descriptions of eigenvalues: two components  
Left (a):  $n = 50$ . Right (b):  $n = 204$ . Bottom (c):  $n = 363$ .

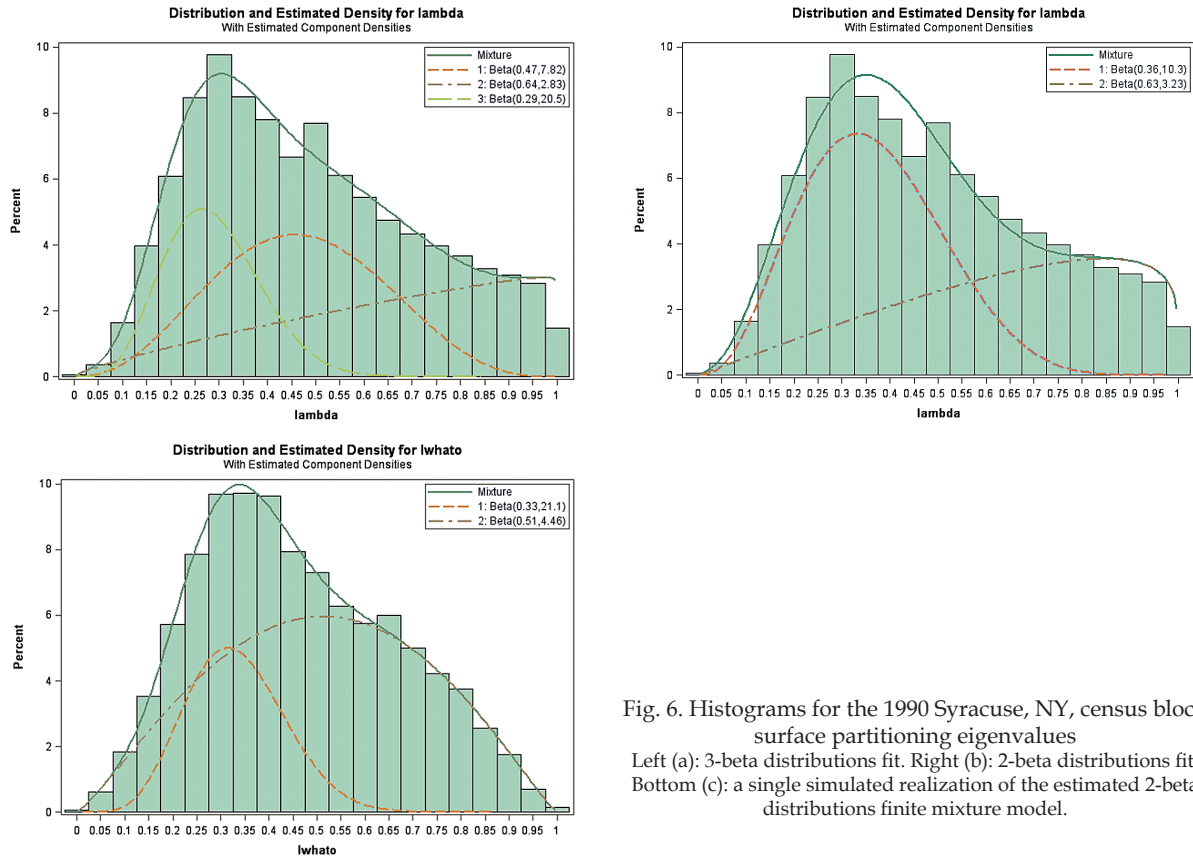


Fig. 6. Histograms for the 1990 Syracuse, NY, census block surface partitioning eigenvalues  
Left (a): 3-beta distributions fit. Right (b): 2-beta distributions fit.  
Bottom (c): a single simulated realization of the estimated 2-beta distributions finite mixture model.

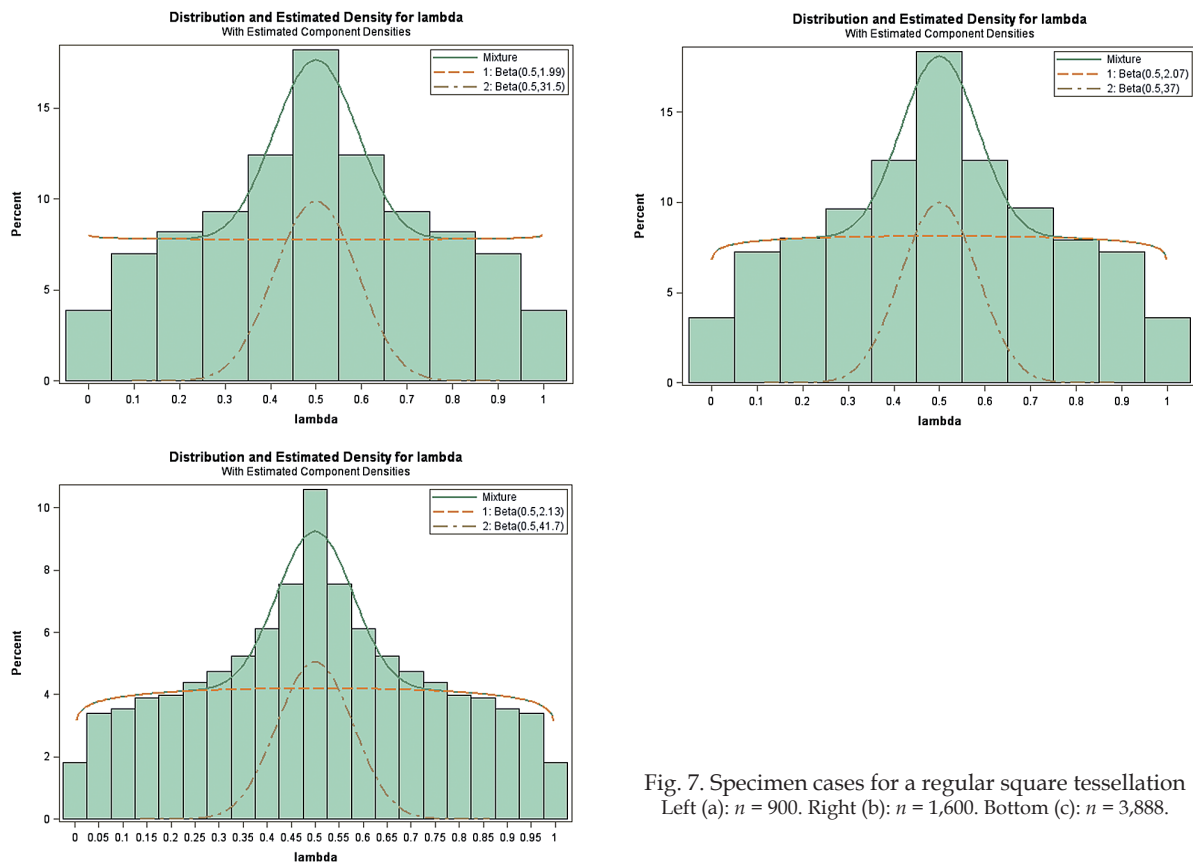


Fig. 7. Specimen cases for a regular square tessellation  
Left (a):  $n = 900$ . Right (b):  $n = 1,600$ . Bottom (c):  $n = 3,888$ .

Table 1. Two-beta mixture estimates for describing eigenvalue frequency distributions for irregular surface partitionings

Dataset	n	P <sub>1</sub>	P <sub>2</sub>	Beta <sub>1</sub>		Beta <sub>2</sub>	
				shape	scale	shape	scale
144	535	0.45109	0.54891	10.5307	33.5618	3.7152	2.7359
82	560	0.38087	0.61913	3.0993	17.3221	2.1084	2.3500
90	584	0.45378	0.54622	15.6297	40.3125	4.4743	3.0444
169	595	0.49913	0.50087	4.3486	16.0993	2.6615	2.2744
4	708	0.50905	0.49095	5.0758	11.1860	3.3501	2.4406
6	731	0.43514	0.56486	5.6709	13.8180	2.6629	2.2355
99	839	0.65643	0.34357	4.4341	6.2253	3.0502	2.0219
94	930	0.46182	0.53818	8.6824	20.9011	3.8749	2.6253
164	1108	0.46276	0.53724	3.4241	14.4060	2.1561	2.1113
153	1122	0.47674	0.52326	3.9052	16.3644	2.4594	2.2667
168	1347	0.52499	0.47501	4.6150	16.3677	3.0422	2.5619
161	1588	0.50370	0.49630	3.9536	15.2695	2.7940	2.4493
21	1754	0.44384	0.55616	6.6828	13.4171	3.3386	2.5488
22	1818	0.46138	0.53862	10.5731	32.7050	4.3801	3.1835
107	2025	0.52389	0.47611	5.2535	8.7615	4.1161	2.6586
142	2246	0.42034	0.57966	2.9406	16.9327	2.5585	2.6743
38	2379	0.43036	0.56964	4.5862	18.5062	3.2821	3.0867
159	2385	0.50491	0.49509	3.8045	15.8300	2.9742	2.5327
96	2489	0.45343	0.54657	3.3624	12.6727	2.8632	2.7542
166	2670	0.51434	0.48566	7.4305	22.9311	4.2678	3.0603
155	2937	0.52692	0.47308	4.4648	16.3087	3.5164	2.7423
156	2966	0.52389	0.47611	4.7289	17.2968	3.2563	2.6262
128	3033	0.45495	0.54505	6.5358	24.0334	3.8493	3.0457
3	3111	0.46062	0.53938	12.7674	32.5233	4.9236	3.3003
141	3324	0.46650	0.53350	19.8799	49.4936	6.2076	3.8820
165	3688	0.54393	0.45607	8.3898	23.4731	4.8717	3.2597
158	4439	0.52535	0.47465	8.0843	24.2288	4.7498	3.2839
160	5164	0.52994	0.47006	5.3690	19.0017	3.9446	2.9731
154	5449	0.54472	0.45528	10.4372	27.1481	5.6887	3.5743
14	7249	0.55297	0.44703	5.9278	10.3392	5.5623	3.2311

NOTE: Beta<sub>1</sub> is the component capturing the hump and relating to  $\lambda_1$ .

NOTE: Beta<sub>2</sub> is the component capturing the heavy tail and relating to  $\lambda_n$ .

6b) are significant. Unfortunately, simulating the eigenvalues with this model fails to produce the desired distribution (Fig. 6c). Nevertheless, their forms suggest that the power law description furnished by equation (1), which links to the largest eigenvalue, needs to be supplemented by a second power law component linking to the smallest eigenvalue.

Table 1 summarizes estimation results for the 2-beta distributions finite mixture models describing the 30 sets of eigenvalues calculated for adjacency matrices representing non-regular square tessellations. The first component (Beta<sub>1</sub>) captures the hump with an approximately bell-shaped beta random variable, and relates to the largest eigenvalue,  $\lambda_1$ . The second component (Beta<sub>2</sub>) captures the heavy tail (i.e., skewness)

and relates to the smallest eigenvalue,  $\lambda_n$ . The mixture weights ( $p_1$  and  $p_2$ ) tend to be approximately equal. The first component is significant for all, and the second component is significant for all but four of the eigenvalue distributions. Histograms for these various sets of eigenvalues resemble Fig. 6b.

The set of eigenvalues for a regular square tessellation are symmetric, and their lack of skewness results in a single mode and only a single beta distribution component. None of the five specimen cases in the convenience sample have a statistically significant second component (see Fig. 7). Each is better characterized by a bell-shaped (i.e., normal) distribution (Fig. 8b). A beta distribution approximation for each of these sets of eigenvalues yields equal shape and scale pa-



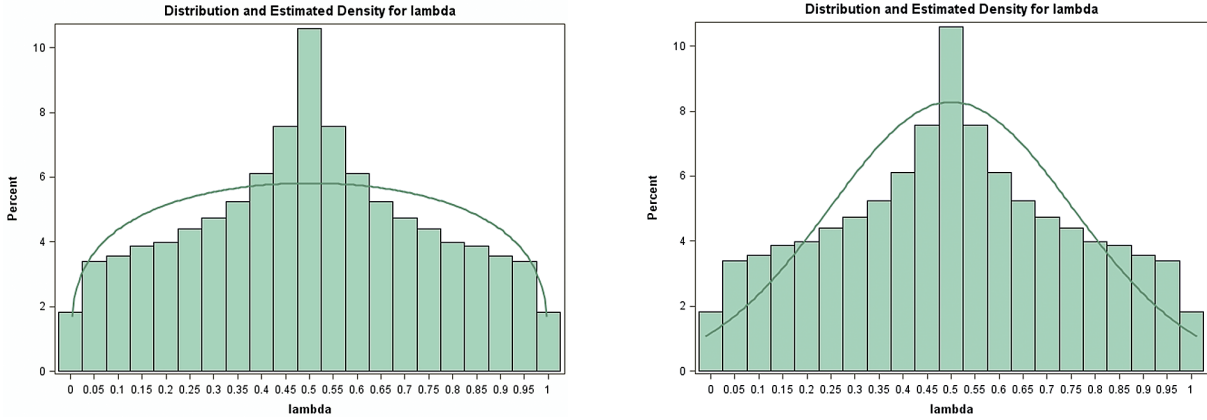


Fig. 8. Approximations to the frequency distribution of the eigenvalues for a regular square tessellation;  $n = 3,888$   
 Left (a): beta approximation (Kolmogorov-Smirnov = 0.04662). Right (b): normal approximation (Kolmogorov-Smirnov = 0.02931).

parameter estimates (due to symmetry), with these estimates ranging from 1.18250 to 1.27612.

In conclusion, although theoretical statistical distribution descriptions yield over-smoothed simulated predicted eigenvalue frequency distributions whose values tend to shrink toward the mean, they do emphasize the need to specify a double power law, one linking to each of the extreme eigenvalues, to describe skewed eigenvalue frequency distributions.

### The special case of a regular square tessellation forming a complete P-by-Q rectangular region

Many remotely sensed images constitute regular square tessellations forming a P-by-Q rectangular region. Fortunately, Griffith (2000, 2004) reports properties for the matrix  $\mathbf{W}$  eigenvalues associated with these surface partitionings that allow them to be approximated with great precision. As mentioned previously, all of the odd moments of these eigenvalues are known to be 0. In addition, the second moment can be calculated from the observed  $\mathbf{W}$  matrix, and also is known (see the preceding discussion). Paralleling the aforementioned variance term for a regular square tessellation forming a complete P-by-Q rectangular region, the fourth moment (relating to kurtosis) is given by:

$$\frac{\sum_{i=1}^{PQ} \lambda_i^4}{PQ} = \frac{1458PQ + 1515(P + Q) + 2584}{10368PQ} \quad (6)$$

,  $p > 2$  and  $Q > 3$

Equation (6) furnishes a precision check tool for the approximated eigenvalues.

The eigenvalue approximation algorithm for this case consists of the following steps:

Step 1: record the sign of each eigenvalue in the variable  $I_{\text{sign}}$  according to its theoretical approximation [based upon Sylvester's (1852) law of inertia]

$$\{\text{COS}[\pi (h-1)/(P-1)] + \text{COS}[\pi (k-1)/Q-1]\}/2, \quad (7)$$

$h = 1, 2, \dots, P$  and  $k = 1, 2, \dots, Q$

Step 2: calibrate  $\gamma$  such that

$$\text{MIN}_{\gamma} \left[ \sum_{h=1}^P \sum_{k=1}^Q \left( I_{\text{sign}} \{ \text{COS}[\pi (h-1)/(P-1)] + \text{COS}[\pi (k-1)/Q-1] \} / 2 \right)^{2\gamma} - \frac{18PQ + 11(P+Q) + 12}{72PQ} \right]^2$$

where  $|x|$  denotes absolute value of  $x$  - this is akin to the method of moments estimation technique;

Step 3: compute

$$\hat{\lambda}_{hk} = I_{\text{sign}} \left\{ \text{COS}[\pi (h-1)/(P-1)] + \text{COS}[\pi (k-1)/Q-1] \right\} / 2 \Big|^\gamma,$$

$h = 1, 2, \dots, P$  and  $k = 1, 2, \dots, Q$ ; and,

Step 4: calculate

$$\frac{\left| \sum_{h=1}^P \sum_{k=1}^Q \hat{\lambda}_{hk}^4 - \frac{1458PQ + 1515(P+Q) + 2584}{10368} \right|}{\frac{145PQ + 1515(P+Q) + 2584}{10368}}$$

the relative error for kurtosis, as a validity check.

One advantage of this approximation is that it preserves both 0 and any duplicate eigenvalues. A simulation experiment implies that  $\hat{\gamma} \approx 1 - 0.14568 [1/(P - 0.02) + 1/(Q - 0.02)] - 0.75917/[(P - 0.37)(Q - 0.37)]$ .

For a 3,000-by-5,000 region,  $\hat{\gamma} \approx 0.99992$ , the first and third moments are 0 to eight decimal places, the relative error for the variance is 0.00761% (this is the minimization result), and the relative error for the validity check (i.e., the fourth moment) is 0.17744%. These results suggest a conjecture in terms of equation (7).

**Conjecture 1.** Let matrix  $\mathbf{W}$  denote an  $n$ -by- $n$  irreducible row-standardized adjacency matrix associated with an undirected connected planar graph  $G$  based upon a regular square tessellation forming a complete  $P$ -by- $Q$  rectangular region and having a symmetric spectrum. Accordingly, its extreme eigenvalues are  $\pm 1$ , with the  $n-2$  ordered intermediate eigenvalues,  $\lambda_k$ , contained in this interval. Then,

$$\hat{\lambda}_{k=(p-1)Q+q} \approx I_{\text{sign}} \left[ \left| \cos\left(\pi \frac{p-1}{P-1}\right) + \cos\left(\pi \frac{q-1}{Q-1}\right) \right| / 2 \right]^\gamma,$$

$$p = 1, 2, \dots, P-1 \text{ and } q = 1, 2, \dots, Q-1,$$

$$(8)$$

where  $\hat{\lambda}_k$  denotes the estimate of  $\lambda_k$  and  $I_{\text{sign}}$  denotes the sign of the sum within the absolute value signs, for a suitable value of parameter  $\gamma$ . If  $P \rightarrow \infty$  and  $Q \rightarrow \infty$ , then  $\gamma \rightarrow 1$ .

The approximations yielded by the algorithm and Conjecture 1 can be improved upon slightly when  $P = Q$  by replacing the estimates for the  $(p, q)$  pairs for which  $\frac{p-1}{P-1} = \frac{q-1}{Q-1}$  and  $\frac{p-1}{P-1} = \frac{Q-p}{Q-1}$  with their exact results given by  $\left[ \cos\left(\frac{p\pi}{P-1}\right) + \cos\left(\frac{q\pi}{Q-1}\right) \right] / 2$ .

This conjecture and the preceding algorithm do not apply to cases where a region is not both

rectangular and complete. In this case, if the associated graph is connected, then neither duplicate nor naturally occurring 0 eigenvalues [i.e., the two cosine terms in equation (7) cancel] exist. Because the variance term is still  $\mathbf{1}^T \mathbf{D}^{-1} \mathbf{C} \mathbf{D}^{-1} \mathbf{1}$ , it can be calculated from known information. Unfortunately, an analytical expression for the fourth moment is no longer available. Considering a ranking of the eigenvalues in descending order, now the approximation equation is

$$I_{\text{sign}} |2(n-i)/(n-1) - 1|^\gamma, i = 1, 2, \dots, n, \quad (9)$$

where  $I_{\text{sign}}$  is the sign of the quantity  $2(n-i)/(n-1) - 1$ . Cressie (1993: 47) furnishes an incomplete square-lattice Pennsylvania coal ash dataset; it involves only 208 of the 368 locations contained in a 16-by-23 regular square tessellation forming a complete rectangular region. Its variance term is  $\mathbf{1}^T \mathbf{D}^{-1} \mathbf{C} \mathbf{D}^{-1} \mathbf{1} = 59.63894$ . Because

$$\sum_{r=1}^n \left[ \frac{r - (n+1)/2}{(n-1)/2} \right]^2 = \frac{PQ(PQ+1)}{3(PQ-1)}$$

$$> \frac{36PQ+11(P+Q)+12}{144}$$

the exponent is a variance deflator, and hence is greater than 1. The minimization solution is  $\hat{\gamma} = 1.26516$ . As a check, the approximate and exact 4<sup>th</sup> moments are calculable, and differ by roughly 3.4%. In other words, the approximation is respectable, and  $n$  is only moderate in size.

Equation (6) also enables asymptotic distribution properties of the eigenvalues to be established. The variance of these eigenvalues is given by:

$$(18PQ + 11P + 11Q + 12)/(72PQ), \quad (10)$$

resulting in the following asymptotic result:

$$\text{Limit}_{Q \rightarrow \infty} \text{Limit}_{P \rightarrow \infty} \frac{[1458PQ + 1515(P+Q) + 2584]/(10358PQ)}{\{[18PQ + 11(P+Q) + 12]/(72PQ)\}^2} =$$

$$\text{Limit}_{Q \rightarrow \infty} \frac{3(505Q + 486Q^2)}{2(11 + 18Q)^2} = 2.25$$

Consequently, although regular square tessellation histograms (e.g., Figs. 7 and 8) suggest the possibility of normality, and the third moment

(i.e., used to measure skewness) is 0, the kurtosis does not converge upon 3. Therefore, the asymptotic distribution of the eigenvalues of matrix  $\mathbf{W}$  for an infinite rectangular tessellation is not normal.

### Approximating eigenvalues for irregular surface partitioning: current practice

To remedy the lack of an eigenvalue approximation equation for connected irregular graphs, and based upon implications from the preceding analyses, Conjecture 1 posits a specification (Griffith 2003: 48) utilizing information provided by the extreme values,  $\lambda_1 = \lambda_{\max}$  and  $\lambda_n = \lambda_{\min}$  (i.e., the maximum and minimum eigenvalues, respectively), the ranking of eigenvalues, and the number of eigenvalues. Converting the ranking of eigenvalues to a relative ranking – such that  $r_i = (n-i)/(n-1)$ ,  $i = 1, 2, \dots, n$  – the rank  $r_i$  indexes a spectrum away from its largest eigenvalue, whereas the rank  $1-r_i$  indexes a spectrum away from its smallest eigenvalue, in keeping with a mixture of two power-law distributions.

**Conjecture 2.** Let matrix  $\mathbf{M}$  denote an  $n$ -by- $n$  irreducible adjacency matrix (either matrix  $\mathbf{C}$  or  $\mathbf{W}$ ) associated with an undirected connected planar graph  $G$ . Let  $\lambda_{\max}$  denote the largest eigenvalue,  $\lambda_{\min}$  the smallest eigenvalue, and  $\lambda_i$  the  $i$ th eigenvalue of matrix  $\mathbf{M}$ . Then

$$\hat{\lambda}_i \approx \frac{\lambda_{\max}}{[\text{LN}(2)+1]^{\gamma_1}} r_i^{\alpha_1 + \beta_1 r_i} [\text{LN}(r_i + 1) + 1]^{\gamma_1} + \frac{\lambda_{\min}}{[\text{LN}(2)+1]^{\gamma_2}} (1 - r_i)^{\alpha_2 + \beta_2(1-r_i)} [\text{LN}(2 - r_i) + 1]^{\gamma_2}, \quad (11)$$

$i = 1, 2, \dots, n$

where  $\hat{\lambda}_i$  denotes the estimate of the  $i$ th eigenvalue,  $\lambda_i$  for suitable values of parameters  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$ ,  $\beta_2$ ,  $\gamma_1$ , and  $\gamma_2$ .

When  $r_i = 1$  (i.e.,  $i = 1$ ), equation (11) reduces to  $\lambda_{\max}$ , and when  $r_i = 0$  (i.e.,  $i = n$ ), equation (11) reduces to  $\lambda_{\min}$ . When equation (11) is estimated with rook case square tessellation eigenvalues, symmetry yields  $\alpha_1 = \alpha_2$ ,  $\beta_1 = \beta_2$ , and  $\gamma_1 = \gamma_2$ .

Results reported in Table 2 reveal that the two extreme values of the distribution are matched with equation (11) but not equation (1), and that

the variance of the equation (11) approximations is much closer to the original than is the variance of the single power law approximations. Table 2 also reports results for the queen's adjacency definition for a square tessellation [rook adjacencies are supplemented with ones also defined in terms of zero length (i.e., point) common boundaries (analogous to the queen's move in chess)]. This popular spatial weights definition (e.g., it is the default in GeoBUGS) is a near-planar graph [i.e., the number of links exceeds that for a planar graph, namely  $6(n-2)$ , by no more than one-third<sup>3</sup>, resulting in matrix  $\mathbf{W}$  still being sparse<sup>4</sup>].

As an aside, the 72-by-79 regular square tessellation with a rook's adjacency definition referred to in Table 2 has  $\hat{\alpha}_1 = \hat{\alpha}_2 = 0.8025$ ,  $\hat{\beta}_1 = \hat{\beta}_2 = -2.0682$ , and  $\hat{\gamma}_1 = \hat{\gamma}_2 = -1.2731$ . But these estimates are calculated with a nonlinear regression whose response variable is the set of eigenvalues. In practice, for extremely large to massively large datasets, these eigenvalues are unknown. The method of moments furnishes one technique for estimating these parameters. But resorting to the method of moments for this estimation can produce unstable results.

### Eigenvalue approximations: lessons from the method of moments

Linear algebra theory states that the sum of the integer powers of the eigenvalues of a matrix equals the trace of that matrix raised to the same integer power. Because the eigenvalues discussed here sum to zero by construction, the sum of these powers divided by  $n$  is equivalent to statistical moments of the eigenvalue distributions.

<sup>3</sup> The Brazilian example has 804 links based on non-zero length shared boundaries, whereas a planar graph of its size has no more than 963.

<sup>4</sup> The number of queen adjacency links in a  $P$ -by- $Q$  regular square tessellation forming a complete rectangular region is  $8PQ - 6P - 6Q + 4$ . The ratio of the difference between this count and the maximum count for a planar graph, to that of the maximum count for a planar graph is  $(PQ - 3P - 3Q + 8)/(3PQ - 6)$ . The limit of this ratio is  $1/3$ .

Table 2. Summary statistics for selected illustrative eigenvalue examples for matrix **W**

Eigenvalue	Mean	Standard deviation	$\lambda_{\min}$	$\lambda_{\max}$
<i>the Brazilian Amazon surface partitioning (v = 0)</i>				
Observed	0	0.45307	-0.91088	1
Equation (1) prediction	0	0.45242	-0.56898	0.98442
Equation (11) prediction	0	0.45298	-0.91088	1
<i>a 72-by-79 regular square tessellation with a rook's adjacency definition</i>				
Observed	0	0.50411	-1	1
Equation (1) prediction	0	0.50034	-0.83725	0.92572
Equation (11) prediction	0	0.50369	-1	1
<i>a 72-by-79 regular square tessellation with a queen's adjacency definition</i>				
Observed	0	0.36347	-0.48577	1
Equation (1) prediction	0	0.36273	-0.35626	0.96156
Equation (11) prediction	0	0.36297	-0.48577	1

NOTE: attempts were unsuccessful to rewrite equation (1) constraining  $\lambda_{\max}$  to equal 1.  
 NOTE: in the queen's case, point connection weights =  $1/\sqrt{2}$ .

**Moments for a hexagonal lattice forming a circular region**

For the regular square and regular hexagonal tessellations forming complete rectangular regions, Griffith (2003: 42-44) reports the first six moments for both matrix **C** and matrix **W**. Results reported in Table 3, which are for a regular hexagonal tessellation that expands in concentric rings from a central hexagon, can be added to this set now. This situation is similar to the regular square tessellation that forms a square region,

but is more similar to irregular tessellations in structure - most irregular planar tessellations can be viewed as being between a square and a hexagonal tessellation in structure.

Provision of Table 3 is for completeness as well as to illuminate skewed eigenvalue distribution features. The limiting eigenvalue distribution here has a skewness statistic of  $\sqrt{2/3} \neq 0$  and a kurtosis statistic of  $5/2 \neq 3$ ; in other words, it does not converge on a normal distribution.

The preceding mixture analysis suggests a 2-component finite beta mixture specification.

Table 3. Analytical moments of eigenvalues extracted from matrices **C** and **W** that depict selected regular hexagonal tessellations expanding in concentric rings ( $r > 2$ ) around a central hexagon

Moment	Matrix C	Matrix W
$\sum_{j=1}^n \lambda_j / n$	0	0
$\sum_{j=1}^n \lambda_j^2 / n$	$6R(3R + 1)/(1+3R+3R^2)$	$(6R^2 + 11R + 6) / [12(1+3R+3R^2)]$
$\sum_{j=1}^n \lambda_j^3 / n$	$36R^2/(1+3R+3R^2)$	$(4R^2 + 7R + 4)/[24(1+3R+3R^2)]$
$\sum_{j=1}^n \lambda_j^4 / n$	$6(45R^2 - 9R - 2)/(1+3R+3R^2)$	$(120R^2 + 289R + 208)/[576(1+3R+3R^2)]$
$\sum_{j=1}^n \lambda_j^5 / n$	$120R(9R - 4)/(1+3R+3R^2)$	$5(288R^2 + 673R + 462)/[10368*(1+3R+3R^2)]$
$\sum_{j=1}^n \lambda_j^6 / n$	$60(102R^2 - 65R + 3)/ (1+3R+3R^2)$	$(16320R^2 + 45151R + 37596)/[124416*(1+3R+3R^2)]$

NOTE: R denotes the number of concentric rings;  $n = 1 + 3R + 3R^2$ .

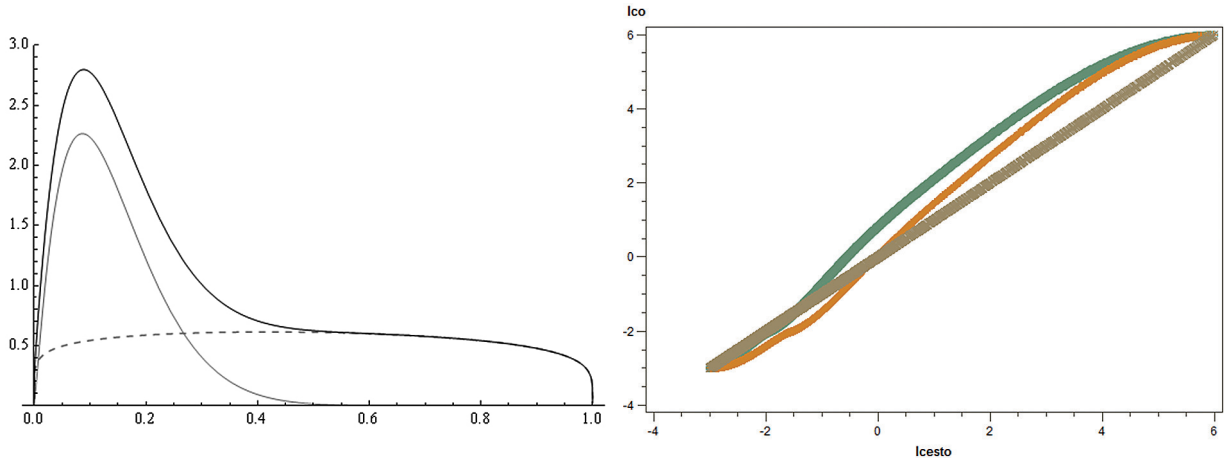


Fig. 9. Asymptotic beta distributions. Left (a): each component and the mixture. Right (b): the predicted eigenvalues, with (brown) and without (green) an inflation adjustment [similar to equation (8),  $\hat{\gamma} = 0.85$ ], for  $n = 5,677$  (i.e.,  $R = 43$ )

But the asymptotic implications for the hexagonal configurations addressed in this section are that the limit is not such a mixture. Estimation for the eigenvalues of matrix  $C$  based upon the method of moments (see Farrell et al. 2011), with the four parameters requiring a matching of the first four moments (i.e.,  $1/3$ ,  $2/27$ ,  $4/243$ , and  $10/729$ ), yields the pair of beta distributions  $B(1.03624, 3.16472)$  and  $B(0.62840, 0.92950)$ ; the probability trajectory through  $R = 43$  concentric rings of hexagons implies  $p = 0.44692$ . These two components fail to produce the correct frequency distribution shape. In contrast, the estimated beta mixture parameter trajectories imply the first component is  $B(2.09525, 12.60373)$ , which combined with the method of moments (only two need to be matched) yields  $B(1.14833, 1.20721)$  as the second component. This solution has the correct form (Fig. 9a), but each of its third and fourth moments deviates from its asymptotic equivalent by roughly 1.75%. Therefore, the finite beta mixture model furnishes only a rough approximation to the distribution of the eigenvalues, as Fig. 9b also illustrates. Its partial success is attributable to its probability density function including  $y^\alpha$ , relating to the maximum eigenvalue, and  $(1 - y)^\beta$ , relating to the minimum eigenvalue [see equation (11)].

As mentioned previously, although the beta mixture conceptualization highlights the presence of two components, one relating to each of the two extreme eigenvalues, the preceding time series conceptualization emphasizes that eigenvalue spacings are extremely important. To this

end, the following equation furnishes a much better eigenvalue estimator than the beta mixtures treated in this section:

$$\hat{\lambda}_i = (\lambda_{\max} - \lambda_{\min})[(r_i - 1)/(n - 1)]^{\beta[(r_i - 1)/(n - 1) + \delta]^{\gamma}} + \lambda_{\min}, \quad i = 1, 2, \dots, n,$$

where  $r_i$  is the ascending  $i$ th rank of a set of ordered eigenvalues. For the case of  $R = 43$  for rings of hexagons,  $\hat{\beta} = 2.49729$ ,  $\hat{\gamma} = 0.31773$ , and  $\hat{\delta} = 0.01583$ . Matching the first three moments with equation (12) yields  $\hat{\beta} = 2.45555$ ,  $\hat{\gamma} = 0.28013$ , and  $\hat{\delta} = 0.00002$ . These sets of estimates render the following summary statistics:

statistic	Table 3	modified beta mixture	equation (12)	equation (12) with moment matching
mean	0	-0.00891	-0.00274	0
variance	5.90805	3.80396	5.86752	5.90805
skewness	0.81650	0.87734	0.81220	0.81650
kurtosis	2.50760	2.93891	2.52330	2.50702

Equation (12), which requires knowing the eigenvalues, and moment matching, which requires knowing only the extreme eigenvalues and the moments, produce substantially better results than the modified (i.e., raised to an exponential power to match the second moment) beta mixture. Because moment matching requires only the first three moments, as mentioned previously, kurtosis furnishes a check, which in this case indicates a slight deviation between the predicted and observed eigenvalues.



### The eigenvalue moments for a connected irregular surface partitioning

Given the first six moments of a set of eigenvalues, with the first being 0 by construction and the second being observable from the matrix itself (i.e.,  $\mathbf{1}^T \mathbf{C} \mathbf{1}$  for matrix  $\mathbf{C}$ , and  $\mathbf{1}^T \mathbf{D}^{-1} \mathbf{C} \mathbf{D}^{-1} \mathbf{1}$  for matrix  $\mathbf{W}$ ) – in other words, the first two moments are known, whereas the remaining four moments need to be calculated, which is impractical when  $n$  is extremely or massively large – the parameters in equation (11) can be estimated without resorting to computing the non-extreme eigenvalues. Equation (8) requires only the first two moments, because symmetry results in all odd moments being 0; preserving symmetry is equivalent to matching these odd moments. Furthermore, exploiting the asymptotic eigenvalue distribution and the theoretical approximation given by equation (7) furnishes even more information. Consequently, previously discussed eigenvalue approximation results for remotely sensed image surface partitions consisting of hundreds of thousands or millions of pixels are extremely good.

Averaged powers of the eigenvalues of a spatial weights matrix furnish the moments of interest here. Because the eigenvalues are known analytically for a binary 0-1 spatial weights matrix representing a regular square tessellation forming a complete rectangular region, their moment approximations can be made for any size geographic landscape. Meanwhile, the moments can be expressed as a function of the various frequencies of neighbors, as follows:

$$a + ao/(PQ) + b\{[(P-2)(Q-2) \times 4^d + 2(P+Q-4) \times 3^d + 4 \times 2^d]/(PQ)\}^c,$$

where the coefficients  $a$ ,  $ao$ ,  $b$ ,  $c$ , and  $d$  need to be estimated.

A sample of 961 such  $P$ -by- $Q$  landscapes, ranging from  $P = 10$  and  $Q = 10$  (i.e.,  $n = 100$ ) to  $P = 100$  and  $Q = 100$  (i.e.,  $n = 10,000$ ) with a rook's definition of adjacency, for which the frequency of neighbors is as follows:

number of neighbors	frequency
4	$(P-2)(Q-2)$
3	$2(P+Q-4)$
2	4

yields

$$\mu_4 \equiv -32 + 24/(PQ) + 17[(P-2)(Q-2) \times 4 + 2(P+Q-4) \times 3 + 4 \times 2]/(PQ), \text{ and}$$

$$\mu_6 \equiv -648 + 600/(PQ) + 262[(P-2)(Q-2) \times 4 + 2(P+Q-4) \times 3 + 4 \times 2]/(PQ),$$

which respectively are identical to results appearing in Griffith (2003: 42);  $c = 1$  and  $d = 1$  are the estimates here. These moment approximations are exact, and furnish the specification for irregular surface partitioning cases.

A sample of 22  $P$ -by- $Q$  landscapes, ranging from 5-by-6 to 75-by-75 (including, for example, 50-by-100) with a rook's definition of adjacency and reflecting the specification  $\mu_2 \equiv \mathbf{1}^T \mathbf{D}^{-1} \mathbf{C} \mathbf{D}^{-1} \mathbf{1}/n$  [i.e., a moment is a function of  $1/(n_i n_j)$ , where  $n_i$  is the row and  $n_j$  is the column sum of matrix  $\mathbf{C}$ ], whose estimation equation is

$$a + ao/(PQ) + b\{[(P-2)(Q-2)/4^d + 2(P+Q-4)/3^d + 4/2^d]/(PQ)\}^c,$$

yields

$$\mu_4 \equiv -0.14063 - 0.04302/(PQ) + 0.24923[(P-2)(Q-2)/4 + 2(P+Q-4)/3 + 4/2]/(PQ), \text{ and}$$

$$\mu_6 \equiv -0.10656 + 0.01684/(PQ) + 0.81685[(P-2)(Q-2)/4 + 2(P+Q-4)/3 + 4/2]/(PQ),$$

which respectively are identical to results appearing in Griffith (2003: 44), except for rounding error<sup>5</sup>;  $c = 1$  and  $d = 1$  are the estimates here. As before, these moment approximations are exact, and furnish the specification for irregular surface partitioning cases.

As noted previously, because of symmetry, the odd moments are 0 for a regular square tessellation coupled with the rook's adjacency.

Asymmetric eigenvalue distributions are positively skewed, and hence have a positive value for the third and fifth moments. The preceding two specimen datasets were merged to explore

<sup>5</sup> An assessment of  $r = 1,000$  ( $n = 3,003,000$ ) yields an approximate set of eigenvalues with a mean of 0.00828, a near-perfect match of the variance, and the following ratios for moments 3-6: 0.98745, 0.98858, 0.98842, and 0.98212. In other words, the approximation is very good, but not perfect.

an appropriate specification for these moments. For the 0-1 binary spatial weights matrix,

$$\mu_3 \approx (\lambda_1 + \lambda_n)^{0.22031} [2.51113 - 1.73877/(PQ) + 0.19128[(f_2 \times 2 + f_3 \times 3 + f_4 \times 4 + f_5 \times 5 + f_6 \times 6)/(PQ)]^{2.00139}, \text{ and}$$

$$\mu_5 \approx (\lambda_1 + \lambda_n)^{1.11521} [67.38840 - 199.54911/(PQ) + 0.59412[(f_2 \times 2 + f_3 \times 3 + f_4 \times 4 + f_5 \times 5 + f_6 \times 6)/(PQ)]^{5.74858}.$$

For the row-standardized spatial weights matrix,

$$\mu_3 \approx (\lambda_1 + \lambda_n)^{0.22438} \{0.03072 - 0.04349/(PQ) + 2.29262[f_{2,3}/(2 \times 3) + f_{2,4}/(2 \times 4) + f_{2,5}/(2 \times 5) + f_{3,3}/(3 \times 3) + f_{3,4}/(3 \times 4) + f_{3,5}/(3 \times 5) + f_{3,6}/(3 \times 6) + f_{4,4}/(4 \times 4) + f_{4,5}/(4 \times 5) + f_{4,6}/(4 \times 6) + f_{5,6}/(5 \times 6) + f_{6,6}/(6 \times 6)]/(PQ)\}^{2.90176}, \text{ and}$$

$$\mu_5 \approx (\lambda_1 + \lambda_n)^{0.22692} \{0.01410 - 0.02316/(PQ) + 1.78297[f_{2,3}/(2 \times 3) + f_{2,4}/(2 \times 4) + f_{2,5}/(2 \times 5) + f_{3,3}/(3 \times 3) + f_{3,4}/(3 \times 4) + f_{3,5}/(3 \times 5) + f_{3,6}/(3 \times 6) + f_{4,4}/(4 \times 4) + f_{4,5}/(4 \times 5) + f_{4,6}/(4 \times 6) + f_{5,6}/(5 \times 6) + f_{6,6}/(6 \times 6)]/(PQ)\}^{2.45354},$$

where  $\lambda_1$  and  $\lambda_n$  are the extreme eigenvalues of matrix C. In all four cases, the pseudo-R<sup>2</sup> is almost exactly 1. A comparison of these equations with those in Griffith (2003: 42-43) for the hexagonal tessellation confirms that these are extremely good approximations. These equations yield exactly 0 for the square tessellation case.

Based on these preceding equations, the moment approximations for irregular tessellations are as follows: matrix C

$$\hat{\mu}_3 \approx (\lambda_1 + \lambda_n)^{0.50159} \left[ 0.44775 + 14.08416/n + 0.14140 \left( \sum_{j=1}^m f_j n_j^{1.42641} \right)^{1.50385} \right], \text{ pseudo-R}^2 = 0.98299$$

$$\hat{\mu}_4 \approx 0.44431 + 37.73972/n + 0.73802 \left( \sum_{j=1}^m f_j n_j^{2.01940} \right)^{1.32278}, \text{ pseudo-R}^2 = 0.99242$$

$$\hat{\mu}_5 \approx (\lambda_1 + \lambda_n)^{0.48873} \left[ -6.35349 + 404.70597/n + 0.69450 \left( \sum_{j=1}^m f_j n_j^{2.22398} \right)^{1.45307} \right], \text{ pseudo-R}^2 = 0.98571$$

$$\hat{\mu}_6 \approx -12.31708 + 2103.96032/n + 1.10621 \left( \sum_{j=1}^m f_j n_j^{2.41215} \right)^{1.76026}, \text{ pseudo-R}^2 = 0.98138$$

matrix W

$$\hat{\mu}_3 \approx (\lambda_1 + \lambda_n)^{0.50997} \left\{ 0.2772 + 0.26157/n + 0.78701 \left[ \sum_{j=1}^m f_j / (n_j n_i) \right]^{3.18904} \right\}, \text{ pseudo-R}^2 = 0.91486$$

$$\hat{\mu}_4 \approx 0.01217 + 0.01829/n + 1.50228$$

$$\left[ \sum_{j=1}^m f_j / (n_j n_i) \right]^{1.81231}, \text{ pseudo-R}^2 = 0.98878$$

$$\hat{\mu}_5 \approx (\lambda_1 + \lambda_n)^{0.38766} \left\{ 0.02463 + 0.36236/n + 0.97646 \left[ \sum_{j=1}^m f_j / (n_j n_i) \right]^{2.92717} \right\}, \text{ pseudo-R}^2 = 0.92774$$

$$\hat{\mu}_6 \approx 0.01486 + 0.12994/n + 1.95422$$

$$\left[ \sum_{j=1}^m f_j / (n_j n_i) \right]^{2.33000}, \text{ pseudo-R}^2 = 0.97659$$

Cross-validation supports the quality of these approximations, which are superior to those appearing in Griffith (2000: 45-46).

### Eigenvalue approximations for connected irregular graphs

Griffith (2000: 48) presents an eigenvalue approximation equation that has six parameters and exploits the rank ordering of a set of eigenvalues. Extensive analysis with the specimen dataset re-

veals that estimation of these six parameters with the method of moments can yield unstable values. The discussion in this section builds upon the preceding features of this paper, namely the calculation of regular square lattice eigenvalues using areal unit coordinates, the pair of extreme eigenvalue power law components, the rank ordering of eigenvalues, and the importance of eigenvalue spacings.

A surface partitioning by a set of polygons can be converted to a geographic distribution of points by calculating each polygon's geometric centroid (a standard GIS function), say  $(u_i, v_i)$ . These are the values used in equation (7) to calculate the eigenvalues of a regular square tessellation forming a complete rectangular region. They also are the values used in equation (8) to calculate eigenvalue approximations. Extending this to an irregular surface partitioning, the coordinate system georeferencing the centroids of these polygons can be standardized so that all centroids lie within the unit square, whose vertex coordinates are, in counterclockwise order,  $\{(0, 0), (1, 0), (1, 1), (0, 1)\}$ . This standardization is achieved by subtracting the smallest coordinate value from each value for a given axis, and then dividing each of the resulting values by the difference between the largest and the smallest original values for each axis:  $([u_i - u_{\min}]/[u_{\max} - u_{\min}], [v_i - v_{\min}]/[v_{\max} - v_{\min}])$ , where  $x_{\min}$  and  $x_{\max}$  respectively are the minimum and maximum values of  $x$ . Next, initial eigenvalue approximations can be calculated with, for eigenvalue  $i$ ,

$$\lambda_{\text{initial},i} = \text{COS} \left( \pi \frac{u_i - u_{\min}}{u_{\max} - u_{\min}} \right) + \text{COS} \left( \pi \frac{v_i - v_{\min}}{v_{\max} - v_{\min}} \right),$$

which also needs to be standardized such that for eigenvalue  $i$ ,

$$\lambda_{\text{temp},i} = \frac{\text{COS} \left( \pi \frac{u_i - u_{\min}}{u_{\max} - u_{\min}} \right)}{\text{MAX}\{\lambda_{\text{initial},i}\} - \text{MIN}\{\lambda_{\text{initial},i}\}} + \frac{\text{COS} \left( \pi \frac{v_i - v_{\min}}{v_{\max} - v_{\min}} \right) - \text{MIN}\{\lambda_{\text{initial},i}\}}{\text{MAX}\{\lambda_{\text{initial},i}\} - \text{MIN}\{\lambda_{\text{initial},i}\}}$$

which produces a set of values ranging from 0 to 1. Accordingly,

**Conjecture 3.** Let matrix  $\mathbf{M}$  denote an  $n$ -by- $n$  irreducible adjacency matrix (either matrix  $\mathbf{C}$  or  $\mathbf{W}$ ) associated with an undirected connected planar graph  $G$ . Let  $\lambda_{\max}$  denote the largest eigenvalue,  $\lambda_{\min}$  the smallest eigenvalue, and  $\lambda_i$  the  $i$ th eigenvalue of matrix  $\mathbf{M}$ . Then

$$\hat{\lambda}_i \approx \lambda_{\max} \lambda_{\text{temp},i}^{\beta \lambda_{\text{temp},i}^{\gamma(\lambda_{\max} + \lambda_{\min})}} + \lambda_{\min} \left( 1 - \lambda_{\text{temp},i}^{\beta \lambda_{\text{temp},i}^{\gamma(\lambda_{\max} + \lambda_{\min})}} \right), \quad (13)$$

$$i = 1, 2, \dots, n.$$

where  $\hat{\lambda}_i$  denotes the estimate of the  $i$ th eigenvalue,  $\lambda_i$ , for suitable values of parameters  $\beta$  and  $\gamma$ .

When  $\lambda_{\text{temp},i} = 1$  (i.e.,  $i = 1$ ), equation (13) reduces to  $\lambda_{\max}$ , and when  $\lambda_{\text{temp},i} = 0$  (i.e.,  $i = n$ ), equation (13) reduces to  $\lambda_{\min}$ . Equation (13) can be rewritten as

$$\hat{\lambda}_i \approx (\lambda_{\max} - \lambda_{\min}) \lambda_{\text{temp},i}^{\beta \lambda_{\text{temp},i}^{\gamma(\lambda_{\max} + \lambda_{\min})}} + \lambda_{\min},$$

which reveals that an exponent is applied to the temporary approximation values contained in the interval  $[0, 1]$ , with this exponent either inflating or deflating these values; the eigenvalue approximation is the eigenvalue range times a standardized rank number in the interval  $[0, 1]$  plus the minimum eigenvalue. Accordingly, the exponent inflation/deflation of the standardized ranks allows matching of the approximate and exact eigenvalue moments. The exponent term  $(\lambda_{\max} + \lambda_{\min})$  ensures that this eigenvalue approximation for regular square tessellations, for example, yields a symmetric set of values. Therefore, equation (13) rivals equation (9), although its fourth moment result deviates by 10.5%.

For a connected irregular planar graph, these approximations can be obtained with the following set of two equations in two unknowns, utilizing the method of moments estimation technique:

$$\sum_{i=1}^n \lambda_i^\omega = \sum_{i=1}^n \hat{\lambda}_i^\omega, \quad \omega = 1, 2.$$

The left-hand side of this equation comprises the values of the first and second moments. The right-hand side is given by the preceding estimated moments for matrix  $\mathbf{C}$  or  $\mathbf{W}$ . The third through sixth moments can function as checks, or be used

to estimate additional parameters included in a refinement of this equation.

**Approximate eigenvalues for matrices C and W: a Poland example**

A full assessment of Conjecture 3 with the specimen sample of eigenvalues is not possible at this time because equation (13) also requires the geometric centroids of polygons, which presently are unavailable for many of the sample surface partitioning. Poland is one exception.

Poland is partitioned into communes ( $n = 2,468$ ), which nest into poviats ( $n = 369$ ), which in turn nest into voivodeships ( $n = 16$ ). Equation (13) furnishes an extremely good approximation for these three sets of eigenvalues, with somewhat better results for matrix C than for matrix W. Fig. 10 indicates that equation (13) needs further development, particularly with regard to negative eigenvalues. But this equation furnishes a remarkably good approximation given that it has only two parameters. It suggests that the coordinates of polygons offer a convenient set of data to approximate eigenvalues, contributing in terms of their spacing property. Considerable future research should be devoted to refining this equation. The preceding moments approximations allow support of up-to-a-six-parameter specification, which presently needs to be identi-

fied. Meanwhile, Table 4 summarizes encouraging preliminary results.

Appendix A presents a detailed step-by-step simple example (i.e., the voivodeships tessellation) of the necessary calculations for computing the eigenvalue approximations. Because  $n$  is so small, the approximations are somewhat poor for this example. As  $n$  increases, the interval defined by the true extreme eigenvalues becomes increasingly more densely filled, with the approximations improving.

Appendix B presents estimates for the same example, not only for polar coordinates, but also for a universal transverse Mercator projection, a Lambert conformal conic projection, and a Lambert azimuthal equal area projection. These map projection results indicate that the eigenvalue approximations seem reasonably insensitive to the coordinate system used. Again, as  $n$  increases, such differences are expected to diminish.

Finally, to help illustrate that the findings summarized in this section are not specific to the Polish example, and help illustrate the generality of this approximation, Appendix C furnishes results based upon a set of random Thiessen polygon surface partitionings. The horizontal and vertical axis coordinates were drawn at random from a uniform distribution. The generated tessellations are for  $n = 500, 1,000, 2,000,$  and  $5,000$  points. Results reported in Appendix C cor-

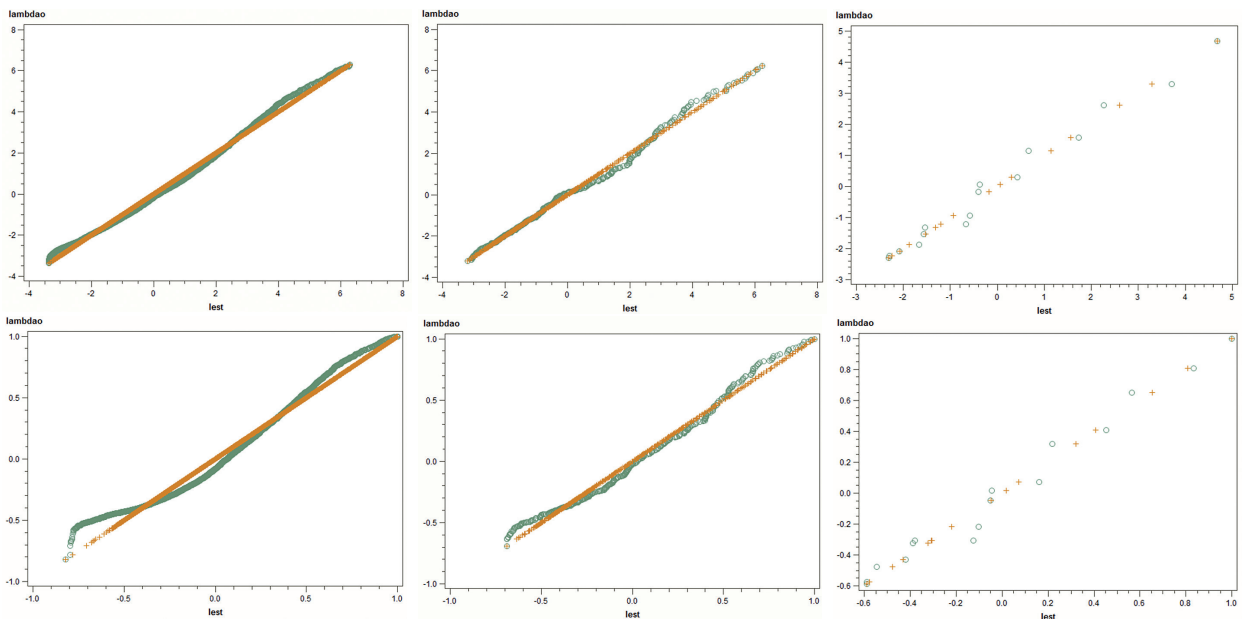


Fig. 10. Scatterplots of approximation versus actual eigenvalues for the surface partitioning of Poland  
 Top row: matrix C. Bottom row: matrix W. Left-hand column: communes. Middle column: poviats. Right-hand column: voivodeships

Table 4. Summary statistics for the nested hierarchy of surface partitions for Poland

Statistic	Communes		Poviats		Voivodeships	
	C	W	C	W	C	W
$\hat{\beta}$	2.00238	1.52497	2.59046	1.82606	2.45731	1.77469
$\hat{\gamma}$	0.00072	0.69815	0.07213	0.16230	-0.00100	-0.66847
mean	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
standard deviation	2.37232	0.41804	2.28773	0.43057	2.12916	0.50222
skewness	0.60478 (0.81512)	0.28047 (0.71272)	0.72695 (0.82560)	0.34867 (0.62738)	0.98429 (0.89987)	0.66728 (0.73192)
excess kurtosis	-0.49848 (-0.27254)	-0.69284 (-0.59217)	-0.31237 (-0.12194)	-0.81155 (-0.66316)	0.16646 (-0.08241)	-0.52344 (-0.59337)
RESS	0.00653	0.03090	0.00528	0.01395	0.01924	0.02494

roborate those summarized here for the Polish tessellations.

## Conclusions and implications

Eigenfunctions, especially eigenvalues, play an important role in parts of spatial analysis. But determining them presents a challenge for large surface partitionings whose sets of polygons have a dual connected planar graph. This paper confirms that the important feature of a set of eigenvalues is their spacing across the range defined by the easily computable extreme values. It also confirms that a two-component finite beta mixture furnishes a good description of the frequency distribution for a set of eigenvalues, but fails to furnish good predictions. This description still is useful because it confirms that an eigenvalue approximation equation needs to include two terms, one that is a function of the largest eigenvalue, and the other that is a function of the smallest eigenvalue.

The first and second statistical moments of a set of eigenvalues always are known from properties of a spatial weights matrix. This paper provides sound approximations for the third through sixth statistical moments, for both the binary spatial weights matrix  $C$ , and its row-standardized counterpart, matrix  $W$ , for the rook's definition of adjacency. These moment approximations can be coupled with the method of moments estimation technique to approximate a set

of eigenvalues, and, in some cases, check the quality of eigenvalue approximations. Approximation results for regular square tessellations for either complete or incomplete rectangular regions are extremely good. Conjecture 1 summarizes this situation.

Conjecture 2 summarizes initial results for irregular surface partitionings. It exploits the rank ordering of a set of eigenvalues, but fails to effectively incorporate their variable spacing property. It also has six parameters to estimate, with the method of moments estimates being unstable for some sets of eigenvalues. Nevertheless, it motivated the establishment of a set of moment approximation equations. Furthermore, its instability motivated the formulation of Conjecture 3, which attempts to better capture the variable spacing property. Equation (13) is an initial specification that, as illustrated by Fig. 10 and Table 4, merits further refinement for proper model identification. This will be the focus of future research, as will be the compilation of areal unit centroids for surface partitionings contained in the specimen dataset, and both the impact of increasing  $n$  and of a wide range of map projections on the quality of the eigenvalue approximations.

## Acknowledgment

Daniel A. Griffith is an Ashbel Smith Professor. This research was supported by the National Science Foundation, research grant # DMS-0611883 (original grant # DMS-0435714).



## References

- Adler M., van Moerbeke P., 2001. Hermitian, symmetric and symplectic random ensembles: PDEs for the distribution of the spectrum. *Annals of Mathematics* 153: 149–189.
- Barry R., Pace R., 1999. Monte Carlo estimates of the log determinant of large sparse matrices. *Linear Algebra and Its Applications* 289: 41–54.
- Box G., Jenkins G., 1976. *Time series analysis: Forecasting, and control*. Holden Day, San Francisco.
- Brouwer A., Haemers W., 2012. *Spectra of graphs*. Springer, New York.
- Cao D., Yuan H., 1993. Graphs characterized by the second eigenvalue. *Journal of Graph Theory* 17: 25–331.
- Cao D., Yuan H., 1995. The distribution of eigenvalues of graphs. *Linear Algebra and its Applications* 216: 211–224.
- Chung F., 1997. *Spectral graph theory*. American Mathematical Society, Providence, RI.
- Chung F., Lu L., Vu V., 2003. Spectra of random graphs with given expected degrees. *Proceedings of the National Academy of Sciences* 100: 6313–6318.
- Cressie N., 1993. *Statistics for spatial data*. Wiley, New York.
- Faloutsos M., Faloutsos P., Faloutsos C., 1999. On power-law relationships of the internet topology. *ACM SIGCOM Computer Communication Review* 29: 251–262.
- Farrell P., Ehsanes Saleh A., Zhang Z., 2011. Methods of moments estimation in finite mixtures. *Sankhyā: The Indian Journal of Statistics* 73-A, Part 2: 218–230.
- Fefferman C., Phong D., 1980. On the asymptotic eigenvalue distribution of a pseudo-differential operation. *Proceedings of the National Academy of Sciences* 77: 5622–5625.
- Golub G., van der Vorst H., 2000. Numerical progress in eigenvalue computation in the 20<sup>th</sup> century. *J. of Computational and Applied Mathematics* 123: 35–65.
- Griffith D., 2000. Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra and its Applications* 321: 95–112.
- Griffith D., 2003. *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization*. Springer, New York.
- Griffith D., 2004. Extreme eigenfunctions of adjacency matrices for planar graphs employed in spatial analyses. *Linear Algebra and Its Applications* 388: 201–219.
- Griffith D., Luhanga U., 2011. Approximating the inertia of the adjacency matrix of a connected planar graph that is the dual of a geographic surface partitioning. *Geographical Analysis* 43: 383–402.
- Hams A., de Raedt H., 2000. Fast algorithm for finding the eigenvalue distribution of very large matrices. *Physical review, E: Statistical physics, plasmas, fluids, and related interdisciplinary topics* 62 (#3): 4365–4377.
- Henson J., Reise S., Kim K., 2007. Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling* 14 (2): 202–226.
- Huffer F., Wu H., 1998. Markov chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics* 54: 509–524.
- Hyndman J., Kostenko A., 2007. Minimum sample size requirements for seasonal forecasting models. *Foresight* 6: 12–15.
- Khorunzhy O., Shcherbina M., Vengerovsky V., 2004. Eigenvalue distribution of large weighted random graphs. *Journal of Mathematical Physics* 45 (#4): 1648–1672.
- Liu B., Bo Z., 2000. On the third largest eigenvalue of a graph. *Linear Algebra and its Applications* 317: 193–200.
- Martin R., Griffith D., 1998. Fast methods for fitting one-parameter spatial models. Department of Geography, Syracuse University Syracuse, NY (unpublished paper).
- Sylvester J., 1852. A demonstration of the theorem that every homogeneous quadratic polynomial is reducible by real orthogonal substitutions to the form of a sum of positive and negative squares. *Philosophical Magazine Series 4*, 4 (23): 138–142.
- Tse R., 1997. An application of the ARIMA model to real estate prices in Hong Kong. *Journal of Property Finance* 8: 152–163.
- Yong X., 1999. On the distribution of eigenvalues of a simple undirected graph. *Linear Algebra and its Applications* 295: 73–80.

## APPENDIX A

Step-by-step calculation of the Polish voivodeships tessellation eigenvalue approximations.

Step 1: obtain the original coordinates (e.g., longitude and latitude)

Step 2: map each coordinate to the unit square with the transformations

$$X = (\text{longitude} - 15.3475) / (22.9318 - 15.3475)$$

$$Y = (\text{latitude} - 49.8613) / (54.1535 - 49.8613)$$

Step 3: calculate the cosine transformation of the unit square coordinates,  $\text{COS}(\pi X)$  and  $\text{COS}(\pi Y)$

Step 4: calculate, and sort in descending order, the initial eigenvalue approximations,

$$\hat{\lambda} = \frac{\text{COS}(\pi X) + \text{COS}(\pi Y) - (-1.7959)}{1.5267 - (-1.7959)}$$

Step 5: compute the true extreme eigenvalues,  $\lambda_{\max}$  and  $\lambda_{\min}$ , which are 1 and -0.5879 for the Polish voivodeships example

Step 6: estimate the exponent  $\hat{\lambda}^{\beta \hat{\lambda}^{\gamma(\lambda_{\max} + \lambda_{\min})}}$  for  $\hat{\lambda}$ , by matching the first two true and approximation moments to calculate the approximate eigenvalues  $\hat{\hat{\lambda}}$ , which are  $\beta = 1.5436$  and  $\gamma = -1.1494$  for the Polish voivodeships example

Step 6 involves setting up a pair of simultaneous equations for estimation purposes, one for each moment.

Table A1 reports the results of these steps for the Polish voivodeships example. Table A2 furnishes comparative statistics for the three sets of eigenvalues.

Table A1. Step-by-step details for geographic weights matrix **W**

Voivodeship	Original coordinates		Unit square coordinates		COS transformation		$\hat{\lambda}$	$\hat{\hat{\lambda}}$	$\lambda$
	Longitude	Latitude	u	v	u	v			
Lower Silesia	16.4127	51.0894	0.1405	0.2861	0.9042	0.6225	1.0000	1.0000	1.0000
Kujavia-Pomerania	18.4895	53.0741	0.4143	0.7485	0.2661	-0.7038	0.9408	0.8345	0.8068
Lublin	22.9047	51.2205	0.9964	0.3167	-0.9999	0.5446	0.8416	0.5640	0.6518
Lubuska Land	15.3475	52.1964	0.0000	0.5440	1.0000	-0.1379	0.8000	0.4542	0.4067
Mazovia	21.0995	52.3446	0.7584	0.5786	-0.7255	-0.2443	0.7062	0.2170	0.3200
Małopolska	20.2650	49.8613	0.6484	0.0000	-0.4495	1.0000	0.6824	0.1595	0.0720
Opole	17.9008	50.6470	0.3367	0.1831	0.4909	0.8392	0.5929	-0.0440	0.0181
Subcarpathia	22.1683	49.9535	0.8993	0.0215	-0.9504	0.9977	0.5899	-0.0504	-0.0470
Podlasie	22.9318	53.2648	1.0000	0.7930	-1.0000	-0.7958	0.5653	-0.1022	-0.2191
Pomerania	17.9781	54.1535	0.3469	1.0000	0.4628	-1.0000	0.5547	-0.1238	-0.3068
Warmia-Mazuria	20.8374	53.8539	0.7239	0.9302	-0.6467	-0.9761	0.4087	-0.3797	-0.3078
Wielkopolska	17.2454	52.3318	0.2502	0.5756	0.7066	-0.2352	0.4035	-0.3872	-0.3217
West Pomerania	15.5665	53.5811	0.0289	0.8666	0.9959	-0.9135	0.3788	-0.4207	-0.4314
Silesia	18.9957	50.3339	0.4810	0.1101	0.0596	0.9408	0.2486	-0.5456	-0.4775
Świętokrzyska Land	20.7726	50.7628	0.7153	0.2100	-0.6260	0.7901	0.0521	-0.5879	-0.5762
Łódź	19.4197	51.6057	0.5369	0.4064	-0.1157	0.2898	0.0000	-0.5879	-0.5879

Table A2. Summary statistics for Polish voivodeship eigenvalues estimation

Eigenvalue	Mean	Standard deviation	Skewness	Excess kurtosis	Minimum	Maximum
$\lambda$	0.0000	0.5022	0.7319	-0.5934	-0.5879	1.0000
$\hat{\lambda}$	0.5478	0.2899	-0.3833	-0.3464	0.0000	1.0000
$\hat{\hat{\lambda}}$	0.0000	0.5022	0.6673	-0.5234	-0.5879	1.0000

## APPENDIX B

Eigenvalue approximations for the Polish voivodeships tessellation and selected map projections.

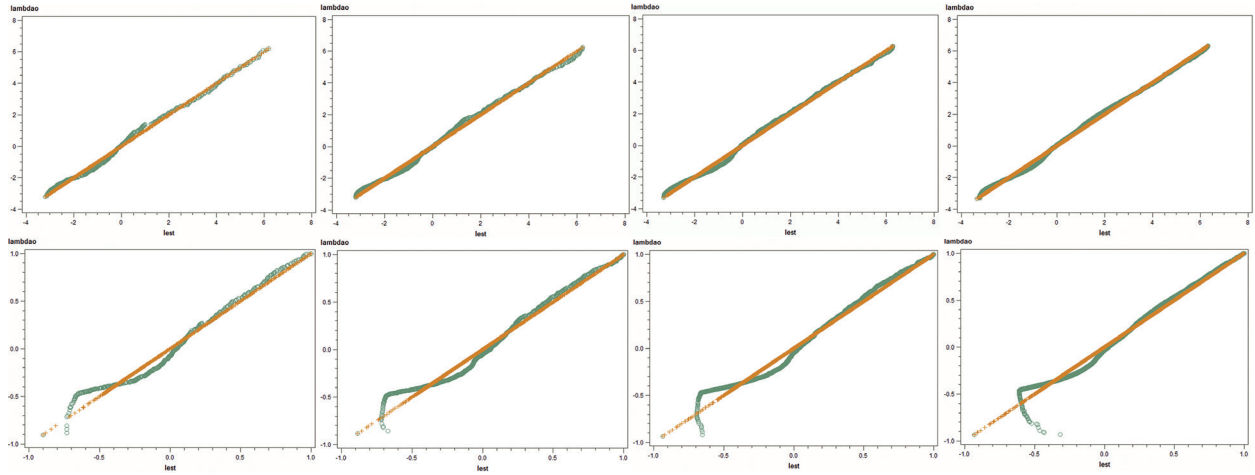
Table B1. Matrix C eigenvalue approximations for selected map projection coordinates

$\lambda(\mathbf{C})$	long/lat	UTM	LAEA	LCC	$\lambda(\mathbf{W})$	long/lat	UTM	LAEA	LCC
4.6811	4.6811	4.6811	4.6811	4.6811	1.0000	1.0000	1.0000	1.0000	1.0000
3.2979	3.7083	3.8973	3.4193	3.4228	0.8068	0.8345	0.8694	0.7806	0.7813
2.6059	2.2666	2.3767	2.3763	2.3699	0.6518	0.5640	0.5907	0.5818	0.5805
1.5601	1.7306	1.3833	1.9291	1.9324	0.4067	0.4542	0.3858	0.4908	0.4915
1.1471	0.6642	0.6623	0.8430	0.8404	0.3200	0.2170	0.2224	0.2534	0.2529
0.2965	0.4235	0.2944	0.3429	0.3444	0.0720	0.1595	0.1333	0.1350	0.1354
0.0641	-0.3749	-0.3516	-0.3149	-0.3115	0.0181	-0.0440	-0.0344	-0.0308	-0.0299
-0.1778	-0.3990	-0.3922	-0.3981	-0.4018	-0.0470	-0.0505	-0.0454	-0.0527	-0.0537
-0.9294	-0.5889	-0.6628	-0.5722	-0.5692	-0.2191	-0.1022	-0.1210	-0.0992	-0.0984
-1.2051	-0.6668	-0.6794	-0.7581	-0.7585	-0.3068	-0.1238	-0.1257	-0.1499	-0.1500
-1.3105	-1.5339	-1.4957	-1.3969	-1.3983	-0.3078	-0.3797	-0.3735	-0.3325	-0.3330
-1.5290	-1.5583	-1.4964	-1.4951	-1.5003	-0.3217	-0.3872	-0.3738	-0.3617	-0.3633
-1.8775	-1.6660	-1.6536	-1.8770	-1.8729	-0.4314	-0.4207	-0.4243	-0.4767	-0.4755
-2.0761	-2.0798	-1.9827	-2.1765	-2.1751	-0.4775	-0.5456	-0.5278	-0.5628	-0.5625
-2.2416	-2.3010	-2.2750	-2.2972	-2.2976	-0.5762	-0.5879	-0.5877	-0.5875	-0.5875
-2.3056	-2.3056	-2.3056	-2.3056	-2.3056	-0.5879	-0.5879	-0.5879	-0.5879	-0.5879
CSS	1.3083	1.3397	0.8395	0.8481		0.0944	0.0763	0.0742	0.0747

Long/lat denotes degrees of longitude and latitude, UTM denotes universal transverse Mercator (Zone 34N) projection, LAEA denotes (North Pole) Lambert azimuthal equal area projection, LCC denotes (Europe) Lambert conformal conic projection, CSS denotes corrected sum of squared differences.

## APPENDIX C

Results for selected Thiessen polygon surface partitionings based upon randomly generated points with each coordinate drawn from the  $[0, 1]$  uniform distribution.



Scatterplots of approximation versus actual eigenvalues for selected Thiessen polygon surface partitionings based upon randomly distributed points. Top row: matrix C. Bottom row: matrix W. Left-hand column:  $n = 500$ . Left-hand middle column:  $n = 1,000$ . Right-hand middle column:  $n = 2,000$ . Right-hand column:  $n = 5,000$

Table C1. Summary statistics for selected Thiessen polygon surface partitionings based on random points

Statistic	n = 500		n = 1,000		n = 2,000		n = 5,000	
	C	W	C	W	C	W	C	W
$\hat{\alpha}$	0.04268	2.02646	0.03139	1.84456	0.03884	3.64997	0.05535	4.34486
$\hat{\gamma}$	1.96495	1.03095	1.98839	1.39045	1.88243	1.29677	2.01190	1.41049
mean	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
standard deviation	2.34841	0.43417	2.37479	0.42670	2.41387	0.41615	2.38785	0.42374
skewness	0.78666 (0.80069)	0.35169 (0.65586)	0.81355 (0.80208)	0.42069 (0.68962)	0.80096 (0.81519)	0.53464 (0.74812)	0.76902 (0.80700)	0.41913 (0.70339)
excess kurtosis	-0.20478 (-0.31393)	-0.60858 (-0.60055)	-0.09850 (-0.34909)	-0.49227 (-0.58849)	-0.18764 (-0.34541)	-0.52266 (-0.53817)	-0.18887 (-0.34248)	-0.55460 (-0.56621)
RESS	0.00625	0.03156	0.00578	0.02983	0.00582	0.03066	0.00454	0.02386

NOTE: exact skewness and excess kurtosis values are in parentheses.