Sankhyā : The Indian Journal of Statistics
San Antonio Conference: selected articles
2002, Volume 64, Series A, Pt.2, pp 344-363

MULTISCALE FREQUENCY TABLE ANALYSIS OF LANDSCAPE FRAGMENTATION IN THEMATIC RASTER MAPS

By G.P. PATIL
and
C. TAILLIE
Penn State University, USA

SUMMARY. This paper is concerned with extracting and summarizing the spatial pattern that may be present in landcover and other thematic raster maps. A sequence of rather large frequency tables is calculated by scanning the map and recording the landcover in configurations of four pixels whose inter-pixel distances become increasingly large. Perceived spatial pattern in a map is determined in part by dominance in the marginal landcover distribution. Likewise, the calculated frequency tables are much affected by the marginal landcover distribution. In order to adjust for this effect and extract a measure of intrinsic spatial pattern, we compare the actual frequency tables to what we would expect the tables to be if landcover were assigned to pixels as independent random draws from the marginal landcover distribution, i.e., for a pattern-free map. Discrepancy between the two sets of frequency tables is assessed using Kullback-Liebler distance, leading to several profiles of fragmentation in the landscape. Methods are illustrated with landcover maps for three Pennsylvania watersheds chosen to represent a range of human impact and disturbance.

1. **Introduction**

We wish to extract and summarize the spatial pattern in thematic raster maps. *Thematic* means that each pixel carries a categorical rather than a numerical value. A typical example is that of a landcover map derived from supervised classification of satellite imagery. In fact, the methods of this paper have been motivated by our study of the landcover of 102 watersheds of the Commonwealth of Pennsylvania as delineated for the State Water

Paper received October 2000; revised January 2002.
AMS (2000) subject classification. Primary 62M40; secondary 62P12.
Keywords and phrases. Collapsing frequency tables, diversity, dominance, entropy, fragmentation profile, intrinsic spatial pattern, Kullback-Liebler distance, marginal landcover distribution, pattern-free maps.

Management Plan. See Johnson and Patil (1998) and Johnson *et al.* (1999, 2001a,b,c). Table 1 shows the landcover distribution for three of these watersheds. There are K=8 landcover categories: water; three forest categories; three vegetation categories; and an unvegetated category. Pennsylvania is heavily forested with relatively little surface water and this is reflected by the landcover percentages in the table. The annual herbaceous category is associated with agriculture while perennial herbaceous includes grassland and pasture and is frequently found along the edge of annual herbaceous regions. The unvegetated category includes buildings, roads, quarries, etc., and (in Pennsylvania) is indicative of human activity and of urbanization.

These three watersheds were chosen to lie along a gradient of human impact. Sinnemahoning Creek in the north central portion of the state lies in the Allegheny National Forest region and is least impacted by human activity. The Jordan Creek (Lehigh) watershed is near the Delaware River in eastern Pennsylvania and represents a transitional landscape that barely maintains a connected forest matrix which is encroached by agriculture and suburban landuse. Finally, the Conestoga Creek watershed (near the city of Philadelphia) is dominated by open agricultural land and urban/suburban aggregate. Remaining forest in the Conestoga watershed is badly fragmented.

The gradient of human impact across the three watersheds is clearly captured by the landcover distributions in Table 1. There is a steady decline in the forest cover with corresponding increases in the urban/suburban and agriculture related coverages. What these marginal landcover distributions do not tell us is how the landcover is spatially distributed across the map. There are two facets to this issue: (i) for a given watershed, how does the spatial dependence change with the distance scale?, and (ii) How does the multiscale spatial pattern vary from watershed to watershed. In particular, does varying spatial pattern differentiate the watersheds in ways that cannot already be accounted for by the marginal landcover distributions?

Our approach to these questions involves scanning the map and calculating a sequence,

$$\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \cdots ,$$

of relative frequency tables. The rows of the tables are indexed by the 4-tuples of landcover categories in ordered sets of four pixels. For the table $\mathcal{F}_0$, the four pixels are adjacent in a 2×2 window. Thus, $\mathcal{F}_0$ is intended to express the spatial pattern at the finest scale of resolution. For $\mathcal{F}_t$, $t > 0$, the four pixels are allowed to be more widely separated in order to capture spatial pattern at broader scales.

When one looks at a thematic map, the perceived spatial pattern is determined in part by the marginal landcover distribution. Thus, if forest is a dominant category, the map would have large patches of forest and a strong empirical forest-to-forest association in neighboring pixels. This would be so, even if the pixels in the map were randomly shuffled. We propose to counteract this effect by asking what we would expect the tables $\mathcal{F}_t$ to be for a randomly shuffled map with the same marginal landcover distribution. Specifically, we calculate the expectation,

$$\mathcal{N}_t = E[\mathcal{F}_t], \tag{1}$$

when landcover categories are assigned to pixels as independent draws (with replacement) from the marginal landcover distribution. The draws are with replacement as a mathematical convenience. The symbol $\mathcal{N}$ is intended to connote a **n**ull or pattern-free map. The difference (however measured) between $\mathcal{F}_t$ and $\mathcal{N}_t$ is then an intrinsic measure of spatial pattern for a fixed marginal landcover distribution and at a given distance scale t. We often use a Kullback-Liebler distance to assess the difference between $\mathcal{F}_t$ and $\mathcal{N}_t$.

Table 1. Marginal landcover distributions for three Pennsylvania watersheds. Sinnemahoning Creek watershed is mostly forested; Jordan Creek is transitional with suburban impact; and Conestoga Creek is an agricultural-urban-suburban mosaic.

	Sinnemahoning		Jordan		Conestoga	
Landcover	Pixels	Pct	Pixels	Pct	Pixels	Pct
Water	18,921	1	8,229	1	3,767	0
Conifer Forest	331,266	12	134,852	10	85,480	7
Mixed Forest	548,505	20	233,213	18	99,375	8
Broadleaf Forest	1,505,149	54	284,397	22	304,960	24
Mixed Vegetation	211,335	8	241,399	19	160,944	13
Perennial Herbaceous	126,495	5	143,501	11	106,050	8
Annual Herbaceous	47,305	2	212,182	16	386,018	30
Unvegetated	986	0	36,001	3	127,657	10
Total:	2,789,962		1,293,774		1,274,251	

2. 4-Tuple Frequency Tables

Now we describe how the frequency tables $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \cdots$ are obtained. First some notation. We suppose that the maps are square with $2^L \times 2^L$ pixels. The Pennsylvania watersheds are not square (and in fact have ragged

boundaries), so it was necessary to embed each into a square image with a minimal value of L and then to mask pixels outside the watershed with an artificial "no data" category. Effects of the embedding and masking would be a distraction here and we simply refer to Johnson *et al.* (2000a). For purposes of this paper, then, the map is $2^L \times 2^L$ with no artificial masking category. Let K be the number of landcover categories which are labeled as $j = 1, \cdots, K$. However, there is no numerical or ordinal significance to the labels.

Next, we slice the map into window panes, each of size 2×2. There are 4^{L-1} such panes and each pane contains four pixels. We arrange these four pixels in the order: northwest corner, northeast corner, southwest corner, southeast corner. This gives a 4-tuple (j_1, j_2, j_3, j_4) of landcover categories where j_1 is the landcover category in the northwest corner, etc. Scanning the map and recording the relative frequency of the various 4-tuples (j_1, j_2, j_3, j_4) gives the finest resolution 4-tuple table $\mathcal{F}_0$. The rows of this table are indexed by the 4-tuples (j_1, j_2, j_3, j_4), so there are K^4 rows in all. With 8 categories, this gives $8^4 = 4096$ rows, which is large but manageable. Unfortunately, the number of rows grows rapidly with K, so that about 20 landcover categories is the maximum that can be handled in this approach.

The marginal landcover distribution

$$\mathbf{p} = [p_1, p_2, \cdots, p_K] \tag{2}$$

can be computed from the frequency table $\mathcal{F}_0$.

THEOREM 1. *Let the relative frequency of row* (j_1, j_2, j_3, j_4) *in the* $\mathcal{F}_0$ *table be denoted by*

$$\pi_{j_1 j_2 j_3 j_4} = \pi^{[0]}_{j_1 j_2 j_3 j_4} .$$

Then, the components of the marginal landcover distribution are given by

$$p_j = \frac{1}{4} \left(\pi_{j\cdots} + \pi_{\cdot j\cdot\cdot} + \pi_{\cdot\cdot j\cdot} + \pi_{\cdots j} \right) . \tag{3}$$

Equation (3) is fairly obvious and is generalized in Theorem 2 below.

In practice, of course, the marginal landcover distribution is actually obtained directly from the map and in much the same manner as $\mathcal{F}_0$: slice the map into individual pixels and record the relative frequency of occurrence of the different landcover categories across the pixels. In fact, this idea is worth pursuing a bit further. Suppose we have sliced the map into individual pixels and they are now spread out helter-skelter across the desktop with each pixel bearing its landcover label. Our job is to put the map back together.

We might start by trying to paste pixels together four at a time into 2×2 blocks. Notice that the 4-tuple distribution $\mathcal{F}_0$ tells us exactly how to do this. At this point, our desktop is covered with 2×2 blocks of pixels. The blocks are oriented in the sense that each has a north, south, east, and west edge. The next step in reconstructing the map is to paste blocks together, four at a time, into 4×4 superblocks of pixels. The frequency table $\mathcal{F}_1'$ needed to accomplish this step would have its rows indexed by 4-tuples of 4-tuples of pixels, i.e., 16-tuples. Notice that $\mathcal{F}_0$ is computable from $\mathcal{F}_1'$ just as the marginal landcover distribution is computable from $\mathcal{F}_0$ by slicing (Theorem 1). But there is a problem: the table $\mathcal{F}_1'$ is simply too big. It has K^{16} rows, which gives $8^{16} \approx 10^{14}$ rows for $K = 8$ landcover categories and over 40 million rows for K as small as 3. In most cases, $\mathcal{F}_1'$ has more rows than there are pixels in the map, so most of the rows would have a frequency of zero.

Our strategy is to collapse the table $\mathcal{F}_1'$ down to a table $\mathcal{F}_1$ whose rows, like those of $\mathcal{F}_0$, are indexed by 4-tuples of landcover categories:

$$\mathcal{F}_1' \implies \mathcal{F}_1 .$$

Traditionally, one collapses a large frequency table by grouping multiple rows from the large table into a single row of the smaller table, i.e., by a many-to-one mapping. Instead, we want to smear each row of $\mathcal{F}_1'$ across multiple rows of $\mathcal{F}_1$. Fix a row of $\mathcal{F}_1'$ with relative frequency f and think of the index for that row as a 4-tuple of 4-tuples. Choose a pixel from the NW corner pane (4 possibilities), a pixel from the NE corner pane (4 possibilities), a pixel from the SW corner pane (4 possibilities), and a pixel from the SE corner pane (4 possibilities). This gives $4 \times 4 \times 4 \times 4 = 256$ possible selections and each selection determines a 4-tuple of pixels and associated 4-tuple (j_1, j_2, j_3, j_4) of landcover categories. The row of our collapsed table $\mathcal{F}_1$ indexed by (j_1, j_2, j_3, j_4) has its frequency incremented by $f/256$. Thus, the frequency f is smeared across 256 rows of $\mathcal{F}_1$ (these 256 rows need not be distinct, however). We refer to the foregoing construction as the *stratified random collapse* of $\mathcal{F}_1'$ to $\mathcal{F}_1$. An alternative collapse (which will not be investigated in this paper) is the *systematic random collapse* defined as follows. Select one pixel from the NW corner (4-possibilities) and then select the corresponding pixel from each of the remaining three corners. This determines four rows of $\mathcal{F}_1$, whose frequencies are incremented by $f/4$. The systematic collapse may offer advantages in that the inter-pixel distances are held constant. We plan to investigate this in a forthcoming paper.

The collapse from $\mathcal{F}_1'$ to $\mathcal{F}_1$ can be described from another point of view, which may help explain the terms "stratified random" and "systematic ran-

dom." Let the rows of $\mathcal{F}_1'$ be arranged along a horizontal (X) axis and the rows of $\mathcal{F}_1$ along a vertical (Y) axis. We are going to define a joint probability distribution on (X, Y) pairs so that the Y-marginal determines the $\mathcal{F}_1$ table. We define the joint distribution by taking the X-marginal to be $\mathcal{F}_1'$. Next, given an X-value, randomly select a pixel from each of its four corners (strata) thereby determining a Y-value. This chance mechanism defines a family of conditional distributions on Y given X and thus the needed joint distribution.

2.1. *The coarser resolution tables.* We have now defined the tables $\mathcal{F}_0$ and $\mathcal{F}_1$ and will describe how the coarser resolution tables $\mathcal{F}_t$, $t \geq 2$, are obtained. In the process, we also explain how $\mathcal{F}_1$ is actually computed without recourse to the conceptual table $\mathcal{F}_1'$. Fix an integer t in the range $0 \leq t \leq L-1$. Recall that the map is of size $2^L \times 2^L$. Slice the map into panes of size $2^{t+1} \times 2^{t+1}$, giving

$$\frac{4^L}{4^{t+1}} = 4^{L-t-1} \tag{4}$$

as the number of panes. Bisect each pane horizontally and vertically to obtain four corners each of size $2^t \times 2^t$. From each corner, select a pixel (4^t possibilities per corner). Each selection of four pixels determines a row (j_1, j_2, j_3, j_4) of $\mathcal{F}_t$ whose frequency is incremented by

$$\frac{f}{4^t 4^t 4^t 4^t} = \frac{f}{256^t}\,, \tag{5}$$

where f is the relative frequency of each pane, i.e., the reciprocal of the number of panes as given by equation (4).

Efficient algorithms have been developed for computing the tables $\mathcal{F}_t$, $0 \leq t \leq L-1$. The algorithms are based on a complete quadtree decomposition (Samet, 1990a,b) of the map, followed by postorder traversal (Knuth, 1973) of the tree. See Johnson *et al.* (2000a) and Patil and Taillie (2000b) for details.

The following generalization of Theorem 1 will be needed in Section 4.

THEOREM 2. *Let the relative frequency of row (j_1, j_2, j_3, j_4) in the $\mathcal{F}_t$ table be denoted by*

$$\pi_{j_1 j_2 j_3 j_4} = \pi^{[t]}_{j_1 j_2 j_3 j_4}\,.$$

Then, the components of the marginal landcover distribution are given by

$$p_j = \frac{1}{4}\left(\pi_{j\cdots} + \pi_{\cdot j\cdot\cdot} + \pi_{\cdot\cdot j\cdot} + \pi_{\cdots j}\right)\,. \tag{6}$$

PROOF. The idea behind the proof is essentially the same as that of Theorem 1, but the notation becomes a bit elaborate. Fix attention on the NW corner of some pane. This corner contains 4^t pixels whose landcover categories comprise a 4^t-dimensional vector

$$\mathbf{k}_1 = [k_{11}, k_{12}, k_{13}, \cdots, k_{1,4^t}]\,. \tag{7}$$

Components of this vector will be written as k_{1a} with the letter a as a generic subscript. The notation $\mathbf{k}_2$, $\mathbf{k}_3$, $\mathbf{k}_4$ with generic subscripts b, c, d will be used for the other three corners. The landcover in all 4^{t+1} pixels of the pane is then determined by the list $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$, $\mathbf{k}_4$, and we write

$$\pi(\mathbf{k}_1; \mathbf{k}_2; \mathbf{k}_3; \mathbf{k}_4)$$

for the relative frequency of occurrence of this 4^{t+1}-tuple of landcovers across all the panes. We need notation for various marginals of this distribution. When we sum over all the components of $\mathbf{k}_1$, the resulting marginal is denoted by $\pi(\cdot; \mathbf{k}_2; \mathbf{k}_3; \mathbf{k}_4)$. The notation $\pi(\langle a{:}j\rangle; \mathbf{k}_2; \mathbf{k}_3; \mathbf{k}_4)$ means sum over all components of $\mathbf{k}_1$ except for component a which is set equal to j. Analogous notation is employed for the other three arguments of π. The definition of $\mathcal{F}_t$ can now be expressed formally as

$$\pi_{j_1 j_2 j_3 j_4} = \frac{1}{256^t} \sum_{\mathbf{k}_1;\mathbf{k}_2;\mathbf{k}_3;\mathbf{k}_4} \sum_{a;b;c;d} \delta(j_1, k_{1a})\delta(j_2, k_{2b})\delta(j_3, k_{3c})\delta(j_4, k_{4d}) \\ \pi(\mathbf{k}_1; \mathbf{k}_2; \mathbf{k}_3; \mathbf{k}_4)$$

where $\delta(\cdot, \cdot)$ is the Kronecker delta. Changing the order of summation and using the definition of the Kronecker delta gives

$$\pi_{j_1 j_2 j_3 j_4} = \frac{1}{256^t} \sum_{a;b;c;d} \pi(\langle a{:}j_1\rangle; \langle b{:}j_2\rangle; \langle c{:}j_3\rangle; \langle d{:}j_4\rangle)\,.$$

Sum over j_2, j_3, j_4 and replace j_1 by j to obtain

$$\begin{aligned} \pi_{j\cdots} &= \frac{1}{256^t} \sum_{a;b;c;d} \pi(\langle a{:}j\rangle; \cdot; \cdot; \cdot) \\ &= \frac{1}{4^t} \sum_a \frac{1}{4^t} \sum_b \frac{1}{4^t} \sum_c \frac{1}{4^t} \sum_d \pi(\langle a{:}j\rangle; \cdot; \cdot; \cdot) \\ &= \frac{1}{4^t} \sum_a \pi(\langle a{:}j\rangle; \cdot; \cdot; \cdot)\,. \end{aligned}$$

As in the proof of Theorem 1, the summation on the right is a ratio whose denominator is the number of panes (given by expression 4) and whose numerator is the number of pixels carrying landcover category j and which occur in NW corners of panes. By symmetry, the right hand side of equation (6) is a ratio whose numerator is the total number of pixels of category j in the map and whose denominator is

$$4 \cdot 4^t \cdot 4^{L-t-1} = 4^L \,,$$

which is the total number of pixels in the map.

2.2. *4-tuple frequencies for pattern-free maps.* Recall that the null or pattern-free map results when landcover categories are assigned to pixels as random draws from the marginal landcover distribution (2).

The null frequency table $\mathcal{N}_t$ is $E[\mathcal{F}_t]$ under this null model. But under the model, each selection from a pane corner is itself a random draw from $\mathbf{p}$. Accordingly, we obtain the following result.

THEOREM 3. *The null tables $\mathcal{N}_0, \mathcal{N}_1, \cdots, \mathcal{N}_{L-1}$ are all equal and have their frequencies given by*

$$\nu_{j_1 j_2 j_3 j_4} = p_{j_1} p_{j_2} p_{j_3} p_{j_4} \,. \tag{8}$$

In view of this result, we drop the subscript t from $\mathcal{N}_t$ and denote the null table as $\mathcal{N}$.

2.3. *The coarsest scale table $\mathcal{F}_{L-1}$.* We can obtain an explicit representation for the last table in the series $\mathcal{F}_0, \mathcal{F}_1, \cdots, \mathcal{F}_{L-1}$. According to the definition of $\mathcal{F}_{L-1}$, we slice the map into panes of size $2^L \times 2^L$, quarter each pane, and select a pixel from each quarter. Here there is only one pane consisting of the entire map which is then quartered. Let

$$\begin{aligned} &\mathbf{p}^{\mathrm{NW}}, \qquad \mathbf{p}^{\mathrm{NE}}, \\ &\mathbf{p}^{\mathrm{SW}}, \qquad \mathbf{p}^{\mathrm{SE}} \end{aligned} \tag{9}$$

be the marginal landcover distributions in the four quarters. By the very definition of $\mathcal{F}_{L-1}$, we obtain the following.

THEOREM 4. *The table $\mathcal{F}_{L-1}$ has its frequencies given by*

$$\pi^{[L-1]}_{j_1 j_2 j_3 j_4} = p^{\mathrm{NW}}_{j_1} p^{\mathrm{NE}}_{j_2} p^{\mathrm{SW}}_{j_3} p^{\mathrm{SE}}_{j_4} \,. \tag{10}$$

We say that the map is *homogeneous in the large* if the four corner marginal landcover distributions in (9) are equal. In this case, they are also equal to the marginal landcover distribution for the map as a whole. While we do not expect actual maps to be homogeneous in the large, the four corner distributions are approximately equal in many instances and it is useful to explore the consequences of homogeneity in the large.

THEOREM 5. *If the map is homogeneous in the large, then $\mathcal{F}_{L-1}$ depends only on the marginal landcover distribution* $\mathbf{p}$ *and has its 4-tuple frequencies given by*

$$\pi^{[L-1]}_{j_1 j_2 j_3 j_4} = p_{j_1} p_{j_2} p_{j_3} p_{j_4} .$$

In particular, $\mathcal{F}_{L-1} = \mathcal{N}$.

The theorem says that under homogeneity in the large, the sequence $\mathcal{F}_0, \mathcal{F}_1, \cdots, \mathcal{F}_{L-1}$ "converges" to the null table $\mathcal{N}$.

3. **Fragmentation-Type Distributions**

The tables $\mathcal{F}_t$ are too large for visual inspection, so we need to calculate summarizing parameters $\theta(\mathcal{F}_t)$ for each table. If θ is numerical we can then plot $\theta(\mathcal{F}_t)$ versus the scale t to produce a *multiscale profile*. Under homogeneity in the large, $\mathcal{F}_{L-1} = \mathcal{N}$, so that $\theta(\mathcal{F}_{L-1}) = \theta(\mathcal{N})$.

In this section, $\theta(\mathcal{F}_t)$ is another frequency table computed from $\mathcal{F}_t$ but having only five rows. The rows of $\mathcal{F}_t$ are indexed by 4-tuples (j_1, j_2, j_3, j_4) and can be classified into four "fragmentation" types depending upon the number of distinct landcover categories among j_1, j_2, j_3, j_4. The 4-tuples of type 2 (i.e., those with exactly two distinct landcover categories) can be further classified into two subtypes depending on whether the two landcover categories are distributed among the four pixels in a 50–50 split or a $\frac{1}{4}$–$\frac{3}{4}$ split. Table 2 lists the five different fragmentation types and gives the number of rows of $\mathcal{F}_t$ which map to each type. The latter is expressed as the product of a descending factorial expression giving the number of ways of selecting the distinct landcovers multiplied by a small integer giving the number of ways of distributing the landcovers among the four pixels. Care has to be taken to avoid double counting in determining this integer. Thus, for type 2b, the integer coefficient is 3 rather than $\binom{4}{2}$=6 because an interchange of A and B is already accounted for in the descending factorial. Similarly, the coefficient for type 3 is 6 instead of 12 because interchange of B and C is already accounted for.

TABLE 2. THE FIVE FRAGMENTATION TYPES OF 4-TUPLES (j_1, j_2, j_3, j_4). THE SECOND COLUMN GIVES A CANONICAL PERMUTATION OF THE COMPONENTS OF (j_1, j_2, j_3, j_4) WHERE A, B, C, D STAND FOR DISTINCT LANDCOVER CATEGORIES. THE THIRD COLUMN GIVES THE NUMBER OF ROWS OF $\mathcal{F}_t$ THAT MAP TO EACH FRAGMENTATION TYPE. THE LAST COLUMN GIVES THE RELATIVE FREQUENCY OF EACH TYPE FOR THE NULL TABLE $\mathcal{N}$. THE QUANTITIES D_1, D_2, D_3 IN THE LAST COLUMN ARE DEFINED IN EQUATION (13).

Type	$j_1\, j_2\, j_3\, j_4$	# Rows	Rel Freq under $\mathcal{N}$
1	AAAA	K	$1 - D_3$
2a	AAAB	$4K(K-1)$	$-4D_2 + 4D_3$
2b	AABB	$3K(K-1)$	$-6D_1 + 3D_1^2 + 3D_3$
3	AABC	$6K(K-1)(K-2)$	$6D_1 - 6D_1^2 + 12D_2 - 12D_3$
4	ABCD	$K(K-1)(K-2)(K-3)$	$3D_1^2 - 8D_2 + 6D_3$
	Total:	K^4	1

The five types are associated with different fragmentation patterns. Type 1 tends to occur in the midst of a patch, type 2 at the boundary between two patches, and type 3 where three patches come together. We call type 4 the "busy" type. These interpretations depend upon how patch size relates to the scaling parameter t.

The last column of Table 2 is called the *FragType* distribution at scale t and will be denoted by $\mathcal{T}_t$. In general, explicit expressions are not available for the five type frequencies; it is necessary to sum the appropriate rows of $\mathcal{F}_t$ to obtain these frequencies. However, type frequencies are available for the null table $\mathcal{N}$ and these turn out to be functions of certain diversity indices evaluated on the marginal landcover distribution. Patil and Taillie (1979, 1982) define a one-parameter family of diversity indices by

$$\Delta_\alpha = \frac{1 - \sum_{j=1}^{K} p_j^{\alpha+1}}{\alpha}\,, \quad \alpha \geq -1\,.$$

The divisor α is not important in the present context, and we write $D_\alpha = \alpha\Delta_\alpha$. Explicitly,

$$\begin{aligned} D_1 &= 1 - \sum p_j^2 = 1 - S_2\,, \\ D_2 &= 1 - \sum p_j^3 = 1 - S_3\,, \\ D_3 &= 1 - \sum p_j^4 = 1 - S_4\,. \end{aligned} \tag{11}$$

These are all diversity indices with D_1 as the so-called Simpson index.

THEOREM 6. *For the null model $\mathcal{N}$, the five type frequencies are functions of D_1, D_2, D_3 and are given in the last column of Table 2.*

PROOF. Under the null model, landcover categories are assigned to pixels as random draws (with replacement) from the marginal landcover distribution $\mathbf{p}$. The relative frequency of type 1 (AAAA) under this model is therefore

$$\sum_{\mathrm{A}} p_{\mathrm{A}}^4 = 1 - D_3 \,.$$

For type 2a (AAAB), the relative frequency is

$$\begin{aligned} 4 \sum_{\mathrm{A}} \sum_{\mathrm{B} \neq \mathrm{A}} p_{\mathrm{A}}^3 p_{\mathrm{B}} &= 4 \sum_{\mathrm{A}} p_{\mathrm{A}}^3 \sum_{\mathrm{B} \neq \mathrm{A}} p_{\mathrm{B}} = 4 \sum_{\mathrm{A}} p_{\mathrm{A}}^3 \left(1 - p_{\mathrm{A}}\right) \\ &= 4 \left(\sum_{\mathrm{A}} p_{\mathrm{A}}^3 - \sum_{\mathrm{A}} p_{\mathrm{A}}^4 \right) = 4 \left(D_3 - D_2\right) \,. \end{aligned}$$

For type 2b (AABB), the relative frequency is

$$\begin{aligned} 3 \sum_{\mathrm{A}} \sum_{\mathrm{B} \neq \mathrm{A}} p_{\mathrm{A}}^2 p_{\mathrm{B}}^2 &= 3 \sum_{\mathrm{A}} p_{\mathrm{A}}^2 \sum_{\mathrm{B} \neq \mathrm{A}} p_{\mathrm{B}}^2 = 3 \sum_{\mathrm{A}} p_{\mathrm{A}}^2 \left(S_2 - p_{\mathrm{A}}^2\right) = 3 \left(S_2^2 - S_4\right) \\ &= 3 \left(1 - 2D_1 + D_1^2 - 1 + D_3\right) = 3 \left(-2D_1 + D_1^2 + D_3\right) \,. \end{aligned}$$

For type 3 (AABC), the relative frequency is

$$\begin{aligned} & 3 \sum_{\mathrm{A}} \sum_{\mathrm{B} \neq \mathrm{A}} \sum_{\mathrm{C} \neq \mathrm{A},\mathrm{B}} p_{\mathrm{A}}^2 p_{\mathrm{B}} p_{\mathrm{C}} \\ &= 6 \sum_{\mathrm{A}} p_{\mathrm{A}}^2 \sum_{\mathrm{B} \neq \mathrm{A}} p_{\mathrm{B}} \left(1 - p_{\mathrm{A}} - p_{\mathrm{B}}\right) = 6 \sum_{\mathrm{A}} p_{\mathrm{A}}^2 \left(\left(1 - p_{\mathrm{A}}\right) \sum_{\mathrm{B} \neq \mathrm{A}} p_{\mathrm{B}} - \sum_{\mathrm{B} \neq \mathrm{A}} p_{\mathrm{B}}^2 \right) \\ &= 6 \sum_{\mathrm{A}} p_{\mathrm{A}}^2 \left(\left(1 - p_{\mathrm{A}}\right)^2 - \left(S_2 - p_{\mathrm{A}}^2\right) \right) = 6 \sum_{\mathrm{A}} p_{\mathrm{A}}^2 \left(1 - 2p_{\mathrm{A}} + p_{\mathrm{A}}^2 - S_2 + p_{\mathrm{A}}^2\right) \\ &= 6 \sum_{\mathrm{A}} p_{\mathrm{A}}^2 \left(D_1 - 2p_{\mathrm{A}} + 2p_{\mathrm{A}}^2\right) = 6 \left(D_1 S_2 - 2S_3 + 2S_4\right) \\ &= 6 \left(D_1 - D_1^2 + 2D_2 - 2D_3\right) \,. \end{aligned}$$

Fortunately, we do not have to explicitly compute the type 4 frequency because the five type frequencies must sum to unity.

3.1. *FragType distributions for three watersheds.* For each of the three watersheds described in Section 1, we have calculated the FragType distributions $\mathcal{T}_t$ from the 4-tuple tables $\mathcal{F}_t$. The latter were obtained for scales t in the range $0 \leq t \leq 7$. Larger values of t would have been possible for particular watersheds, depending upon watershed size, but the upper bound

of t=7 works for all 102 watersheds. The resulting FragType distributions appear in Table 3. The null FragType distributions, $\mathcal{T}_{\text{null}}$, are also shown for each watershed. These were calculated from the expressions appearing in the last column of Table 2. As expected from Theorem 5, the empirical FragType distribution converges toward the null FragType distribution for each watershed as the scale t increases. In assessing convergence, keep in mind that the largest tabulated scale of 7 is less than $L - 1$ for each watershed. The type 1 frequencies decrease monotonically toward their limit, while the type 3 and type 4 frequencies increase monotonically toward their limits. This results in a steady shift in the occurrence of type 1 4-tuples to that of types 3 and 4 with increasing scale.

TABLE 3. FRAGTYPE DISTRIBUTIONS FOR THE THREE WATERSHEDS OF TABLE 1. THE ROW LABELED "NULL" IS THE FRAGTYPE DISTRIBUTION FOR THE NULL TABLE $\mathcal{N}$ WITH THE SAME MARGINAL LANDCOVER DISTRIBUTION AS THE WATERSHED IN QUESTION.

	Scale	Relative Frequency				
Watershed	t	Type 1	Type 2a	Type 2b	Type 3	Type 4
Sinnemahoning	0	0.549	0.263	0.132	0.053	0.003
	1	0.407	0.323	0.146	0.116	0.008
	2	0.286	0.347	0.152	0.197	0.019
	3	0.194	0.344	0.147	0.280	0.034
	4	0.145	0.333	0.136	0.336	0.050
	5	0.121	0.330	0.130	0.361	0.058
	6	0.108	0.328	0.123	0.379	0.063
	7	0.096	0.322	0.120	0.394	0.068
	null	0.086	0.322	0.112	0.404	0.076
Jordan	0	0.290	0.303	0.174	0.209	0.025
	1	0.168	0.298	0.153	0.324	0.058
	2	0.093	0.261	0.133	0.413	0.100
	3	0.051	0.218	0.115	0.473	0.143
	4	0.029	0.182	0.102	0.508	0.180
	5	0.018	0.156	0.090	0.527	0.209
	6	0.012	0.137	0.083	0.536	0.232
	7	0.008	0.113	0.073	0.539	0.266
	null	0.006	0.097	0.066	0.541	0.290
Conestoga	0	0.366	0.285	0.162	0.169	0.018
	1	0.226	0.311	0.145	0.274	0.044
	2	0.142	0.300	0.131	0.355	0.072
	3	0.093	0.273	0.119	0.414	0.101
	4	0.068	0.249	0.108	0.449	0.125
	5	0.050	0.226	0.100	0.476	0.149
	6	0.035	0.201	0.090	0.500	0.175
	7	0.030	0.193	0.087	0.507	0.183
	null	0.012	0.135	0.075	0.528	0.250

3.2. *FragType profiles.* Direct comparison of the FragType tables for different watersheds is difficult because the magnitude of the tabulated entries is very much influenced by the marginal landcover distribution. To adjust for this effect and arrive at a measure of intrinsic spatial pattern, we propose to use the Kullback-Liebler distance,

$$d_t = D_{\mathrm{KL}}(\mathcal{T}_t, \mathcal{T}_{\mathrm{null}}) \,,$$

between the empirical FragType distribution $\mathcal{T}_t$ and its null counterpart $\mathcal{T}_{\mathrm{null}}$. Large values of d_t indicate a strong spatial dependence at distance scale t since the actual map is then very different from the pattern-free map at that scaling.

Recall that the Kullback-Liebler distance between two discrete probability distributions $f(x)$ and $g(x)$ is given by

$$D_{\mathrm{KL}}(f, g) = \sum_x f(x) \ln \left(\frac{f(x)}{g(x)} \right) . \tag{12}$$

See Kullback and Liebler (1951) and Kulhavy (1996). The Kullback-Liebler distance is nonnegative and it vanishes if and only if f and g define the same distribution. The expression (12) is not symmetric in f and g. As general guidance one should take f as the empirical-based and g as the model-based distribution. (For nested models, take g belonging to the more restricted model.) Accordingly, we take f to be the empirical FragType distribution $\mathcal{T}_t$ and g to be the null FragType distribution $\mathcal{T}_{\mathrm{null}}$.

We call the plot of the Kullback-Liebler distance d_t versus t as the *FragType profile* for the map. The profile vanishes identically for the pattern-free map with given marginal landcover distribution, and, in general, height of the profile above the horizontal axis indicates the strength of the intrinsic spatial pattern at varying distance scales. The profiles typically decrease towards zero with increasing t and are often well approximated by a curve of the form

$$d_t = a \exp(-bt) \,,$$

for suitable choice of the parameters a and b. When this approximation is applicable, the profile is nicely summarized by the pair (a, b) which facilitates inter-watershed comparisons. Validity of the approximation is empirical; we know of no theoretical justification.

FragType profiles for the three example watersheds are tabulated in Table 4 and are plotted in Figure 1. The Sinnemahoning profile is clearly differentiated from the profiles for the other two watersheds and exhibits

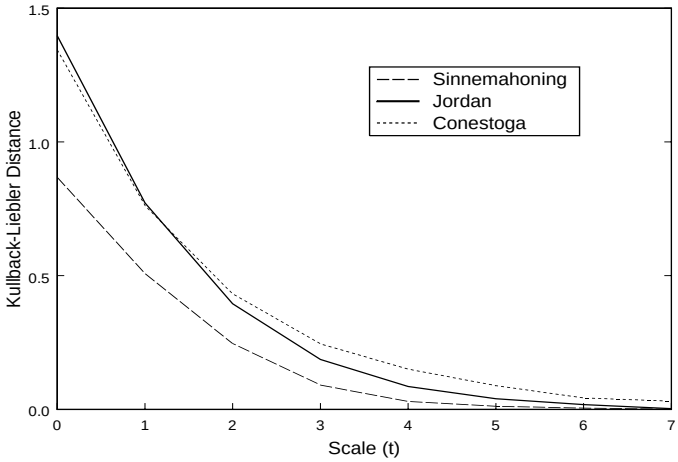

Figure 1. FragType profiles for the three watersheds of Table 1.

a relatively weak spatial pattern which dies rapidly with increasing scale. FragType profiles for the Jordan and Conestoga watersheds are similar but the Conestoga spatial pattern is a bit more persistent with increasing scale. This may be typical of agricultural regions.

Table 4. FragType profiles for the three watersheds of Table 1.

Scale (t)	Sinnemahoning	Jordan	Conestoga
0	0.868	1.397	1.344
1	0.508	0.772	0.762
2	0.246	0.395	0.432
3	0.091	0.187	0.245
4	0.030	0.086	0.151
5	0.012	0.040	0.089
6	0.005	0.018	0.043
7	0.001	0.003	0.031

4. **Absoute Entropy Profiles**

The entropy of a discrete distribution $f(x)$ is

$$H(f) = -\sum_x f(x)\ln(f(x))\,.$$

With a fixed finite set of atoms x, the entropy is a maximum for the uniform or even distribution and is a minimum (in the limiting sense) when the distribution becomes concentrated on a single atom.

Consider a thematic raster map with 4-tuple frequency tables $\mathcal{F}_0, \mathcal{F}_1, \cdots, \mathcal{F}_{L-1}$. We define the absolute entropy of the map at scaling level t to be

$$e_t = H(\mathcal{F}_t)\,.$$

The idea is that with strong spatial dependence, the $\mathcal{F}_t$ table will concentrate its frequencies on particular 4-tuples with a resulting *drop* in entropy. A plot of e_t versus t is called the *absolute entropy profile* (AEP). The qualifier "absolute" distinguishes this profile from one defined in Johnson *et al.* (2000a) but based on conditional entropy. A comparison of the two profiles is made by Burnicki (2001). Typically the absolute entropy profiles are monotone increasing, concave downward, and level off to a horizontal asymptote. These features are illustrated by the AEPs for our three example watersheds; see Table 5 and Figure 2.

Table 5. Absolute entropy profiles for the three watersheds of Table 1. The last row, labeled as "null" is the null entropy $e_{\text{NULL}} = H(\mathcal{N})$. For the Jordan watershed, the profile rises to a value slightly above e_{NULL}, contrary to theory. This is a consequence of the masking of pixels outside the watershed.

Scale (t)	Sinnemahoning	Jordan	Conestoga
0	3.46	5.43	5.01
1	4.11	6.20	5.81
2	4.67	6.73	6.29
3	5.07	7.07	6.59
4	5.25	7.24	6.77
5	5.31	7.34	6.90
6	5.34	7.41	7.01
7	5.36	7.48	6.93
null	5.39	7.45	7.21

The three profiles in Figure 2 have similar shapes but differ in their vertical heights, i.e., their horizontal asymptotes. Since we cannot let $t \to \infty$, the horizontal asymptote is not unambiguously defined but we can assess it in two different ways: (i) as the final entropy value $e_{L-1} = H(\mathcal{F}_{L-1})$ or (ii) as the null entropy $e_{\text{null}} = H(\mathcal{N})$ corresponding to the marginal landcover distribution $\mathbf{p}$. The two assessments are equal when the map is homogeneous in the large and are nearly equal in most cases.

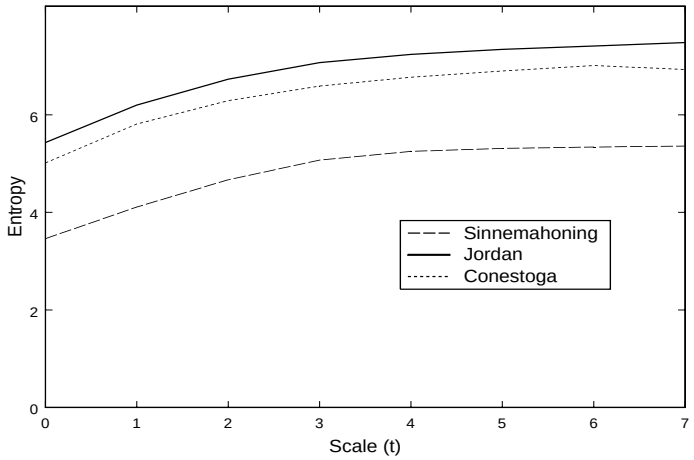

Figure 2. Absolute entropy profiles for the three watersheds of Table 1.

The entropy of a cartesian product is the sum of the entropies of the factors, so Theorem 4 and Theorem 5 tell us that

$$e_{L-1} = H(\mathcal{F}_{L-1}) = H(\mathbf{p}^{\mathrm{NW}}) + H(\mathbf{p}^{\mathrm{NE}}) + H(\mathbf{p}^{\mathrm{SW}}) + H(\mathbf{p}^{\mathrm{SE}}) \tag{13}$$

$$H(\mathcal{N}) = H(\mathbf{p} \times \mathbf{p} \times \mathbf{p} \times \mathbf{p}) = 4H(\mathbf{p})\,. \tag{14}$$

The $H(\mathcal{N})$ assessment of the horizontal asymptote is mathematically more convenient and we use it throughout. The horizontal asymptote is then determined entirely by the marginal landcover distribution $\mathbf{p}$. In fact, $H(\mathbf{p})$ is the Shannon index of diversity of the marginal landcover distribution. The profile asymptote is high when the marginal landcover distribution is nearly uniform across the landcover categories and is low when the landcover is dominated by one or two categories.

We have observed that most of the differences among the three profiles of Figure 2 is due to variation in their horizontal asymptotes and is therefore explained by differences among the marginal landcover distributions instead of differences in spatial pattern. Variation in spatial pattern among the watersheds would be reflected in different profile shapes. An obvious way of removing the effect of varying $\mathbf{p}$ and to compare profile shape is to look at the difference between the horizontal asymptote and the profile:

$$e_t^{\mathrm{c}} = H(\mathcal{N}) - H(\mathcal{F}_t), \tag{15}$$

which we call the complemented AEP.

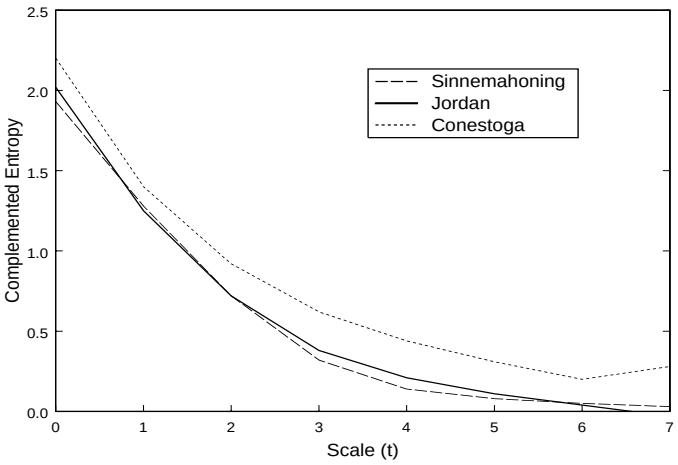

Figure 3. Complemented absolute entropy profiles for the three watersheds of Table 1.

Complemented AEPs for the three watersheds are plotted in Figure 3. Here, the Conestoga watershed is differentiated from the other two watersheds, whose complemented AEPs are nearly identical. This is in contrast to the FragType profile (Figure 1) which differentiated the Sinnemahoning watershed from the other two. Intrinsic spatial pattern is multidimensional in nature and if the three watersheds are envisioned as lying in a high dimensional space, then two measures are needed here to successfully discriminate the three watersheds on the basis of intrinsic spatial pattern (i.e., in directions "orthogonal" to the directions of variation associated with the marginal landcover distribution). As Figure 2 emphasizes, however, the differences among the three watersheds are primarily due to differences in their marginal landcover distributions rather than their intrinsic spatial pattern.

Another way of removing the effect of the marginal landcover distribution on the profiles is to compute the Kullback-Liebler distance between $\mathcal{F}_t$ and $\mathcal{N}$, in analogy with the previous section. The two ways of making the adjustment turn out to be equivalent as shown by the following theorem.

THEOREM 7. $D_{\mathrm{KL}}(\mathcal{F}_t, \mathcal{N}) = H(\mathcal{N}) - H(\mathcal{F}_t) = 4H(\mathbf{p}) - H(\mathcal{F}_t)\,.$

PROOF. Let the relative frequency of row (j_1, j_2, j_3, j_4) in the $\mathcal{F}_t$ table be denoted by

$$\pi_{j_1 j_2 j_3 j_4} = \pi^{[t]}_{j_1 j_2 j_3 j_4}\,.$$

Then, applying the definition of the Kullback-Liebler distance,

$$
\begin{aligned}
D_{\mathrm{KL}}(\mathcal{F}_t, \mathcal{N}) &= \sum_{j_1;j_2;j_3;j_4} \pi_{j_1 j_2 j_3 j_4} \ln\left(\frac{\pi_{j_1 j_2 j_3 j_4}}{\nu_{j_1 j_2 j_3 j_4}}\right) \\
&= -H(\mathcal{F}_t) - \sum_{j_1;j_2;j_3;j_4} \pi_{j_1 j_2 j_3 j_4} \ln\left(\nu_{j_1 j_2 j_3 j_4}\right) \\
&= -H(\mathcal{F}_t) - \sum_{j_1;j_2;j_3;j_4} \pi_{j_1 j_2 j_3 j_4} \ln\left(p_{j_1} p_{j_2} p_{j_3} p_{j_4}\right) \qquad \text{by Theorem 3} \\
&= -H(\mathcal{F}_t) - \sum_{j_1;j_2;j_3;j_4} \pi_{j_1 j_2 j_3 j_4} \ln(p_{j_1}) - \sum_{j_1;j_2;j_3;j_4} \pi_{j_1 j_2 j_3 j_4} \ln(p_{j_2}) \\
&\qquad - \sum_{j_1;j_2;j_3;j_4} \pi_{j_1 j_2 j_3 j_4} \ln(p_{j_3}) - \sum_{j_1;j_2;j_3;j_4} \pi_{j_1 j_2 j_3 j_4} \ln(p_{j_4}) \\
&= -H(\mathcal{F}_t) - \sum_{j} \left(\pi_{j\ldots} + \pi_{\cdot j\cdot\cdot} + \pi_{\cdot\cdot j\cdot} + \pi_{\ldots j}\right) \ln(p_j) \\
&= -H(\mathcal{F}_t) - 4 \sum_{j} p_j \ln(p_j) \qquad \text{by Theorem 2} \\
&= 4H(\mathbf{p}) - H(\mathcal{F}_t)\,.
\end{aligned}
$$

This completes the proof.

5. **Discussion**

A central theme of the paper is that perceived spatial pattern is partly due to dominance/evenness of the marginal landcover distribution. Numerous measures of landscape fragmentation have been proposed (McGarigal and Marks, 1995), but fail to recognize or adjust for this effect. We have proposed that spatial pattern be analyzed using a sequence of 4-tuple frequency tables. Because of their large size, the tables must be further summarized. We have suggested two summarizations (the FragType profile and the complemented AEP) which employ Kullback-Liebler distance to adjust for the effect of the marginal landcover distribution. A limitation of both profiles is that they discard landcover labels. Thus, interchanging two landcover categories has no effect on the profiles but may result in very different landscapes. The 4-tuple frequency tables do retain the labels and the challenge is to find useful but label dependent summarizations of the tables.

One promising line of attack is model-based (Patil and Taillie, 2000a) and employs Markov transition matrices to generate thematic raster maps.

The eigenvectors of the transition matrices retain the labeling and the principal eigenvector is precisely the marginal landcover distribution. The non-principal eigenvectors are indicative of spatial pattern. The model, and therefore the eigenvectors, can be estimated from the 4-tuple frequency tables by minimizing the Kullback-Liebler distance between the observed tables $\mathcal{F}_t$ and their model-predicted counterparts.

Examination of Figure 1 and Figure 3 raises the question of whether the profile differences in these figures are "statistically significant." Any answer to this question has to be model-based (since the map is a "sample" of size 1) and will probably require simulation. The model described in the preceding paragraph can be employed for this purpose since both the fitting and the simulation are quite rapid (Patil and Taillie, 2000b).

Acknowledgement. This article was prepared with partial support from the United States Environmental Protection Agency Cooperative Agreement Number CR-825506. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

References

Burnicki, A. (2001). Comparison of absolute and conditional entropy profiles for assessing landscape fragmentation. MS Thesis, Department of Statistics, Penn State University, University Park, PA.

Johnson, G.D., Myers, W.L. and Patil, G.P. (1999). Stochastic generating models for simulating hierarchically structured multi-cover landscapes. *Landscape Ecology*, **14**, 413-421.

Johnson, G.D., Myers, W.L., Patil, G.P. and Taillie, C. (2001a). Fragmentation profiles for real and simulated landscapes. *Environ. Ecol. Stat.*, **8(1)**, 5-20.

Johnson, G.D., Myers, W.L., Patil, G.P. and Taillie, C. (2001b). Predictability of surface water pollution loading in Pennsylvania using watershed-based landscape measurements. *J. Amer. Water Resources Assoc.*, **37(4)**, 821-835.

Johnson, G.D., Myers, W.L., Patil, G.P. and Taillie, C. (2001c). Characterizing watershed-delineated landscapes in Pennsylvania using conditional entropy profiles. *Landscape Ecology* (revised submission).

Johnson, G.D. and Patil, G.P. (1998). Quantitative multiresolution characterization of landscape patterns for assessing the status of ecosystem health in watershed management areas. *Ecosystem Health*, **4**, 177-187.

Knuth, D.E. (1973). *The Art of Computer Programming, vol.* 1: *Fundamental Algorithms.* Addison-Wesley, Reading, Mass.

Kulhavy, R. (1996). *Recursive Nonlinear Estimation: A Geometric Approach.* Springer, New York.

Kullback, S. and Liebler, R.A. (1951). On information and sufficiency. *Ann. Math. Statist.*, **22**, 79-86.

McGarigal, K. and Marks, B. (1995). FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure. General Technical Report PNW-GTR-351. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.

Patil, G.P. and Taillie, C. (1979). An overview of diversity. In Grassle, J.F., Patil, G.P., Smith, W. and Taillie, C. (eds.). *Ecological Diversity in Theory and Practice*, *Statistical Ecology*, Vol. **6**. International Co-operative Publishing House, Fairland, Maryland.

Patil, G.P. and Taillie, C. (1982). Diversity as a concept and its measurement. *J. Amer. Statist. Assoc.*, **77**, 548-567.

Patil, G.P. and Taillie, C. (2000a). A Multiscale Hierarchical Markov Transition Matrix Model for Generating and Analyzing Thematic Raster Maps. Technical Report Number 2000-0603, Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University, University Park, PA 16802.

Patil, G.P. and Taillie, C. (2000b). A Computer Program for Simulating Thematic Raster Maps with the HMTM model. Technical Report Number 2000-0604, Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University, University Park, PA 16802.

Samet, H. (1990a). *Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS*. Addison-Wesley, Reading, MA.

Samet, H. (1990b). *The Design and Analysis of Spatial Data Structures*. Addison-Wesley, Reading, MA.

G.P. Patil and C. Taillie
Center for Statistical Ecology and
Environmental Statistics
Department of Statistics
Penn State University
University Park, PA 16802 USA
E-mail: gpp@stat.psu.edu
taillie@stat.psu.edu