A Multiscale Hierarchical Markov Transition Matrix Model for Generating and Analyzing Thematic Raster Maps

G. P. Patil and C. Taillie

Center for Statistical Ecology and Environmental Statistics
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

Abstract. A model is described for generating hierarchically scaled spatial pattern as represented in a thematic raster map. The model involves a series of Markov transition matrices, one for each level in the scaling hierarchy. In full generality, the model allows the transition matrices to be different at each level, potentially making available a large number of parameters for landscape characterization. The model is self-similar when the transition matrices are all equal. A method is presented for fitting the model to data that take the form of a single-resolution thematic raster map. Explicit analytic solutions are obtained for the fitted parameters. The fitting method is based on a relationship between the hierarchical transitions in the model and spatial transitions at varying distance scales in the data map, a categorical analogy of the geostatistical variogram.

Keywords: Auto-association matrix, Categorical spatial analysis, Detailed balance, Eigenvalues, HMTM model, Landscape characterization, Markov property, Quadtree, Reversibility, Self-similarity, Spatial–Hierarchical duality, Spatial pattern, Spectral theorem.

1 Introduction

With the growth and availability of remote sensing technologies, scientific data are increasingly presented in the form of classified maps. The nominal variables displayed in such a map may be inherently categorical, as in the case of landcover/landuse, or they may represent the grouping of numerical data into several classes, or they may be the result of clustering multiband data. Variogram-based geospatial analysis is a well-developed set of tools for

Prepared with partial support from the National Science Foundation Cooperative Agreement Number DEB-9524722 and 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 Agencies and no official endorsement should be inferred.

studying spatially referenced *numerical* data. But there is no corresponding set of tools for spatial analysis of nominal data. This paper describes a parametric stochastic model for categorical raster maps. The model employs a hierarchy of Markov transition matrices with each level in the hierarchy corresponding to a distance scale in the raster map. The model is accordingly called the hierarchical markov transition matrix (HMTM) model. Different hierarchical levels in the model may have different transition matrices, thereby modeling spatial scaling domains in the raster map. The map (more precisely, the model) is self-similar when all the transition matrices are equal to one another.

Eigen-decompositions of the transition matrix (or matrices) provide a powerful set of parameters for characterizing spatial pattern in the map. Spatial pattern is the result of (i) the relative frequency of the different mapping categories (marginal mapping distribution) and (ii) the spatial arrangement of the mapping categories across the pixels. Although the marginal distribution has no explicit spatial content, it does affect the apparent spatial pattern as perceived by a human studying the map or as measured by many landscape metrics. For example, a dominant category will have to appear in patches even if the spatial arrangement of labeled pixels is a random shuffle. To separate perceptual from real, a model for raster maps should include the marginal mapping distribution as an explicit set of parameters with additional parameters to regulate the actual spatial pattern. This is neatly effected in the HMTM model since the marginal mapping distribution is the principal eigenvector of the transition matrices. The remaining eigenvalues and eigenvectors then serve to characterize spatial pattern and, for simulation purposes, are easily varied while holding the marginal mapping distribution constant. By contrast, in a Gibbs random field model, it is not easy to calculate let alone to control the marginal mapping distribution.

Fitting the HMTM model to a data map employs a link between hierarchical transitions in the model and spatial transitions at varying distances in the map. Statistical dependence of the categorical response on two pixels is characterized by auto-association matrices which depend on the (quadtree) distance between the pixels. These auto-association matrices are analogous to the auto-covariance function (variogram) for numerical responses, but they provide a probability instead of a moment description of the spatial dependence. The auto-association matrices are directly computable by scanning the map. Subject to some mild restrictions described in Section 3, we show that the HMTM model is identifiable and estimable in terms of the auto-association matrices. In fact, the spectral theorem provides explicit formulae for computing the transition matrices from the auto-association matrices.

We envision two types of application of the HMTM model. First, the fitted model parameters may be used for landscape characterization and comparison. Of particular interest in this context is the relationship between model parameters and familiar landscape metrics such as FRAGSTATS (McGarigal and Marks, 1995) and conditional entropy profiles (Johnson and Patil, 1998; Johnson *et al.*, 2000). Second, map simulation with the HMTM model is very fast (Patil and Taillie, 2000a) facilitating Monte Carlo determination of null distributions and confidence statements for map characteristics. For example, Patil and Taillie (2000c) develop statistics for testing self-similarity of categorical raster maps and obtain the null distributions by simulating the fitted HMTM model.

2 The HMTM Model

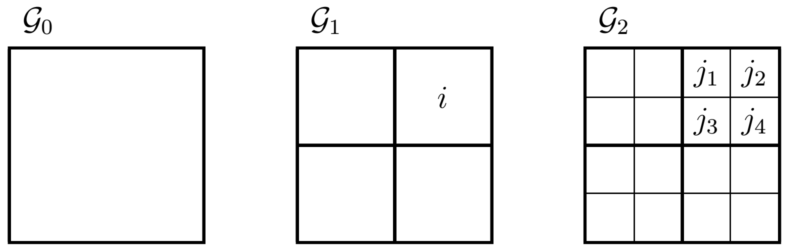

Figure 1: Nested hierarchy of pixels. Each pixel of $\mathcal{G}_n$ subdivides into four subpixels in $\mathcal{G}_{n+1}$. In the diagram, the NE corner of $\mathcal{G}_1$ has color i and its four subpixels in $\mathcal{G}_2$ have colors j_1, j_2, j_3, and j_4.

The HMTM model generates a sequence $\mathcal{G}_0$, $\mathcal{G}_1$, $\cdots$, $\mathcal{G}_L$ of thematic raster maps. For fitting purposes, the final map, $\mathcal{G}_L$, corresponds to the data map and the earlier maps are not observable. Each map in the series covers the same spatial extent, but the successive maps are of increasingly finer resolution. In fact, $\mathcal{G}_0$ consists of a single pixel and the pixels of $\mathcal{G}_n$ are bisected horizontally and vertically to produce the pixels of $\mathcal{G}_{n+1}$ (Figure 2). Each pixel of $\mathcal{G}_n$ contains four subpixels of $\mathcal{G}_{n+1}$. In considering the transition from $\mathcal{G}_n$ to $\mathcal{G}_{n+1}$, it will be convenient to refer to the pixels of $\mathcal{G}_n$ as "mother" pixels and the four subpixels of $\mathcal{G}_{n+1}$ as "daughter" pixels. In particular, $\mathcal{G}_n$ is $2^n \times 2^n$ and contains 4^n pixels. We say that the map $\mathcal{G}_n$ and the transition from $\mathcal{G}_{n-1}$ to $\mathcal{G}_n$ occur at *level n* in the hierarchy.

This hierarchical subdivision by recursive vertical and horizontal bisection is known as a complete quadtree partitioning. There is an enormous literature on quadtrees (see Samet, 1990a,b, and the references therein), but much of this literature is concerned with the data compression potential of incomplete quadtrees in which leaf nodes can occur at any level in the tree. In this paper, quadtrees are always complete and are used to generate the hierarchical pixelation needed for the HMTM model.

Next, we describe how the various mapping categories are assigned to pixels of $\mathcal{G}_n$. This is done using a Markov transition matrix. We suppose that there are K mapping categories or "colors", labeled as 1, 2, $\cdots$, K. At the coarsest scale, the assignment of a color to the single pixel of $\mathcal{G}_0$ is determined by a random draw from an initial stochastic (row) vector

$$\mathbf{p}^{[0]} = [p_1^{[0]}, p_2^{[0]}, \ldots, p_K^{[0]}].$$

Given the assignment of colors to $\mathcal{G}_n$, the assignment to $\mathcal{G}_{n+1}$ is generated by a row stochastic transition matrix,

$$\mathbf{G}^{[n,n+1]} = \left[G_{ij}^{[n,n+1]} \right].$$

The symbol "$\mathbf{G}$" is used for the transition matrices since they **G**enerate the maps. Fix attention on a particular mother pixel of $\mathcal{G}_n$ and let its color be i. The colors of the four daughter pixels are generated by four independent draws according to the distribution specified by the ith row of $\mathbf{G}^{[n,n+1]}$. Thus, given the mother color, the four daughter colors are conditionally independent and conditionally exchangeable. Unconditionally, however, the daughter colors are dependent because the daughters tend to be like their mother. This "genealogical" dependence induces spatial association among neighboring pixels. Marginally, the mapping distribution of $\mathcal{G}_{n+1}$ is obtained from the initial vector $\mathbf{p}^{[0]}$ via the recurrence relation

$$\mathbf{p}^{[n+1]} = \mathbf{p}^{[n]} \mathbf{G}^{[n,n+1]} . \tag{1}$$

We observe two extreme cases. First, when the transition matrices are strongly diagonally dominant (close to an identity matrix), daughter colors are almost always the same as the mother color leading to a contagious map with extensive patches. At the other extreme, when a transition matrix has all its rows the same, daughter colors are assigned at random with the common row as marginal mapping distribution. Since the transition matrix is allowed to vary from transition to transition, the final map $\mathcal{G}_L$ can exhibit different spatial patterns at different scaling levels.

In practice, only the final, floor resolution, map $\mathcal{G}_L$ may be available for analysis. From this single resolution map, we would like to be able to draw inferences on the model parameters by relating spatial scaling levels across $\mathcal{G}_L$ to hierarchical levels in the model.

One fitting method has been briefly described by Johnson *et al.* (2000) and involves computing a sequence of 4-tuple frequency tables from the image $\mathcal{G}_L$ with each table summarizing the spatial pattern at a different spatial scale. Since it is possible to obtain expected frequencies for each table in terms of the transition matrices, the model can be fitted by minimizing a "distance" between observed and expected frequencies. Possible distances are chi-square and Kullback-Liebler. This approach is numerically very complicated because of the sizable number of parameters in the model and also because of the large number of rows in the frequency tables.

The present paper presents a more direct and elegant fitting method which relates the hierarchical transition matrices to spatial transitions at different distances in the map $\mathcal{G}_L$. We present a somewhat simplified version of the model in which the parametric dimension is exactly equal to the number of degrees of freedom in the spatial transition matrices. This permits a perfect matchup between observed and model-predicted versions of the spatial transition matrices. The linkup is actually accomplished analytically by relating the eigen-decomposition of the hierarchical transition matrices to the eigen-decomposition of the spatial transition matrices (after suitable transformation, the first is the square root of the second). Although we accomplish a perfect match between observed and expected spatial transition matrices, the latter are only one set of summarizing characteristics of $\mathcal{G}_L$ and the matchup will generally not be perfect for other summary measures. In fact, comparing observed and predicted values for other summary measures is one way of assessing model adequacy.

We begin with the case where the transition matrices $\mathbf{G}^{[n,n+1]}$ are the same for all hierarchical levels n. This will establish the basic eigen-decomposition which will then be generalized to handle the situation of changing transition matrices.

3 Strongly Reversible Transition Matrices

We will consider row stochastic $K \times K$ transition matrices $\mathbf{G}$ which satisfy the conditions listed below. Such matrices will be called as *strongly reversible*. The conditions are:

(SR1). Each transition matrix $\mathbf{G}$ has a stationary (row) probability vector $\mathbf{p}$. This vector is required to be the initial vector $\mathbf{p}^{[0]}$ in the hierarchical model. According to the recurrence (1), $\mathbf{p}$ is the marginal mapping distribution for each map $\mathcal{G}_n$. We also require that none of the components of $\mathbf{p}$ vanish.

(SR2). For each transition, the pair $(\mathbf{p}, \mathbf{G})$ is *reversible* or, equivalently, satisfies the condition of *detailed balance*. This means that the matrix $\mathbf{H}$ is symmetric, where the components of $\mathbf{H}$ are given by

$$H_{ij} = p_i \, G_{ij} \,, \quad i, j = 1, 2, \cdots, K \,. \tag{2}$$

(SR3). For each transition, the matrix $\mathbf{H}$ is positive definite.

The matrix $\mathbf{H}$ defined above will play an important role in what follows. Note that $\mathbf{H}$ specifies the joint distribution of i and j where i is the color of a mother pixel and j is the color of any of its daughter pixels. Reversibility (SR2) means that the conditional distributions of daughter given mother are the same as the conditional distributions of mother given daughter. Although reversibility is not particularly compelling for hierarchical transitions, it is much more natural for spatial transitions and it does allow us to relate the two types of transition matrices. Reversibility or detailed balance is a refined form of stationarity. Stationarity means that the frequency of transitions out of a given state i are exactly balanced (on average) by transitions into i. Detailed balanced means that transitions from i to any other state j are balanced (on average) by transitions from j to i. Bremaud (1999), Kijima (1997), Lange (1999), Schinazi (1999) and Serfozo (1999) discuss reversible Markov chains.

The matrix $\mathbf{H}$ completely determines the pair $(\mathbf{p}, \mathbf{G})$. In fact, let $\mathbf{H}$ be any symmetric matrix with nonnegative entries and with $H_{\cdot\cdot} = 1$. Define

$$\begin{aligned} p_i &= H_{i\cdot} = H_{\cdot i} \,, \\ G_{ij} &= H_{ij}/p_i \,. \end{aligned}$$

Then, the matrix $\mathbf{G}$ is row stochastic and $\mathbf{p}$ is a stationary vector for $\mathbf{G}$. Accordingly, there is a correspondence

$$\mathbf{H} \iff (\mathbf{p}, \mathbf{G}) \,.$$

From this, we can compute the dimension of the family of strongly reversible transition matrices. Taking into account the symmetry of $\mathbf{H}$ and the normalization to unity, we see that

$$\dim(\mathbf{H}) = \frac{K(K+1)}{2} - 1 = \frac{(K-1)(K+2)}{2} \,.$$

We have been working with Pennsylvania watersheds and K=8 mapping categories. In our context, then, a strongly reversible transition matrix involves 35 parameters. Each transition may have its own transition matrix, but we are requiring $\mathbf{p}$ to be common to all transitions so the total number of parameters across L transitions could be as large as

$$\dim(\mathbf{p}) + L\,(\dim(\mathbf{H}) - \dim(\mathbf{p})) = \frac{(K-1)(KL+2)}{2} \,.$$

With K=8 and L=10 (a 1024×1024 map), this allows as many as 287 parameters compared with 35 parameters for the self-similar model.

Many formulas involving the stationary vector $\mathbf{p}$ can be expressed in matrix form by using the diagonal matrix,

$$\mathbf{P} = \mathrm{diag}(p_1, p_2, \cdots, p_K)\,. \tag{3}$$

For example, the defining equation (2) for $\mathbf{H}$ becomes $\mathbf{H} = \mathbf{PG}$. Since $\mathbf{P}$ is diagonal with positive entries, it can be raised to an arbitrary real power α:

$$\mathbf{P}^\alpha = \mathrm{diag}(p_1^\alpha, p_2^\alpha, \cdots, p_K^\alpha)\,.$$

We write

$$\begin{aligned} \sqrt{\mathbf{P}} &= \mathbf{P}^{1/2} \\ \frac{1}{\sqrt{\mathbf{P}}} &= \mathbf{P}^{-1/2}\,. \end{aligned}$$

We now characterize the condition (SR3) in terms of the eigenvalues of the transition matrix $\mathbf{G}$.

Theorem 1. *Assume that conditions* (SR1) *and* (SR2) *hold. Then,* $\mathbf{H}$ *is positive definite if and only if all the eigenvalues of* $\mathbf{G}$ *are positive.*

Proof. The matrix

$$\widetilde{\mathbf{H}} = \frac{1}{\sqrt{\mathbf{P}}}\mathbf{H}\frac{1}{\sqrt{\mathbf{P}}} = \sqrt{\mathbf{P}}\mathbf{G}\frac{1}{\sqrt{\mathbf{P}}}\,. \tag{4}$$

is symmetric since $\mathbf{H}$ is symmetric. Also, $\mathbf{H}$ is positive definite if and only if $\widetilde{\mathbf{H}}$ is positive definite, i.e., if and only if all the eigenvalues of $\widetilde{\mathbf{H}}$ are positive. But,

$$\begin{aligned} \widetilde{\mathbf{H}} - \lambda\mathbf{I} &= \sqrt{\mathbf{P}}\mathbf{G}\frac{1}{\sqrt{\mathbf{P}}} - \lambda\mathbf{I} \\ &= \sqrt{\mathbf{P}}\left(\mathbf{G} - \lambda\mathbf{I}\right)\frac{1}{\sqrt{\mathbf{P}}}\,. \end{aligned}$$

Taking determinants of both sides of this equation shows that $\mathbf{G}$ and $\widetilde{\mathbf{H}}$ have the same characteristic polynomial and, therefore, the same set of eigenvalues. This completes the proof.

In general, each eigenvalue λ of a row stochastic matrix satisfies $|\lambda| \leq 1$; but the eigenvalues can be negative and can even be complex. However, Bailey (1964) points out that stochastic matrices occurring in applications tend to have positive eigenvalues so that the condition (SR3) is not overly restrictive. In fact, transition matrices in the HMTM model are typically diagonally dominant in order to achieve spatial dependence in $\mathcal{G}_L$. But it is known that the eigenvalues of diagonally dominant transition matrices have positive real parts (Varga, 2000, p. 23). On the other hand, detailed balance implies that the eigenvalues are real (see the proof of Theorem 1 above and note that $\widetilde{\mathbf{H}}$ is symmetric).

The matrix $\widetilde{\mathbf{H}}$ defined by equation (4) also serves to characterize the strongly reversible transition matrix. In fact, let $\widetilde{\mathbf{H}}$ be any $K \times K$ matrix which is symmetric positive definite with nonnegative entries. Also, suppose that $\widetilde{\mathbf{H}}$ has $\lambda = 1$ as an eigenvalue whose (row) eigenvector $\mathbf{q}$ has all its components positive. We may suppose that $\mathbf{q}$ has been scaled to have its Euclidean norm equal to unity. Letting $p_i = q_i^2$, i=1,2,$\cdots$,K, the vector $\mathbf{p}$ is a probability vector. Equation (3) may now be used to define the diagonal matrix $\mathbf{P}$ and then $\mathbf{H}$ can be defined by equation (4). The components of $\mathbf{H}$ are then given by

$$H_{ij} = q_i \widetilde{H}_{ij} q_j \, .$$

Using the fact that $\mathbf{q}$ is an eigenvector of $\widetilde{\mathbf{H}}$ with unit eigenvalue, it follows that $\mathbf{p}$ is the marginal of $\mathbf{H}$ and, by the previous discussion, a stationary vector of $\mathbf{G}$.

We have seen that $\widetilde{\mathbf{H}}$ and $\mathbf{G}$ have the same set of eigenvalues. There is also a useful correspondence between the eigenvectors of the two matrices. Here, $\widetilde{\mathbf{H}}$ is easier to deal with since, by symmetry, its row and column eigenvectors are transposes of one another. Let $\mathbf{y}$ be a row eigenvector of $\widetilde{\mathbf{H}}$ corresponding to an eigenvalue λ. Then, from equation (4), the row eigenvector of $\mathbf{G}$ corresponding to λ is $\mathbf{y}\sqrt{\mathbf{P}}$ and the column eigenvector is $\frac{1}{\sqrt{\mathbf{P}}}\mathbf{y}^T$. In particular, when $\lambda = 1$ and $\mathbf{y} = \mathbf{q}$ with $q_i = \sqrt{p_i}$, the resulting row eigenvector of $\mathbf{G}$ is the stationary vector $\mathbf{p}$ and the resulting column eigenvector is a column of one's (showing $\mathbf{G}$ to be row stochastic).

4 The Self-Similar Case of the HMTM Model

In addition to the requirement of strong reversibility, we now suppose that all of the transition matrices $\mathbf{G}^{[n,n+1]}$ in the HMTM model are equal to one another. The common value will be denoted by $\mathbf{G}$ and the HMTM model is said to be *self-similar*.

We will fit the model by examining its predictions for pairs of adjacent pixels in the floor resolution image $\mathcal{G}_L$. Consider two adjacent pixels in $\mathcal{G}_L$ and regard them as two of the four daughters of a mother pixel in $\mathcal{G}_{L-1}$. Let the mother pixel have color i and let j and k be the colors of the two daughters. By stationarity, i is a draw from the stationary vector $\mathbf{p}$. Therefore, the joint distribution of the daughter colors is given by

$$R_{j\,k} = \sum_{i=1}^{K} p_i G_{i\,j} G_{i\,k} \, , \tag{5}$$

or

$$\mathbf{R} = \mathbf{G}^T \mathbf{P} \mathbf{G} \, . \tag{6}$$

Theorem 2. *The matrix* $\mathbf{R}$ *satisfies the following conditions:*

1. $\mathbf{R}$ *is symmetric and* $R_{\cdot\cdot} = 1$.

2. $R_{\cdot j} = R_{j \cdot} = p_j \, , \quad j = 1, 2, \cdots, K$. *Thus, the marginal vector for* $\mathbf{H}$ *is also the marginal vector for* $\mathbf{R}$.

3. $\mathbf{R}$ *is positive definite.*

Proof. The first two properties are immediate from equation (5). For the third property, let $\mathbf{x}$ be any nonzero K-dimensional column vector. From equation (6),

$$\begin{aligned}\mathbf{x}^T\mathbf{R}\mathbf{x} &= \mathbf{x}^T\mathbf{G}^T\mathbf{P}\mathbf{G}\mathbf{x} \\ &= \mathbf{y}^T\mathbf{P}\mathbf{y} \\ &> 0\end{aligned}$$

since $\mathbf{P}$ is diagonal with positive entries on the diagonal. This completes the proof.

We note that the preceding proof is quite general and does not use strong reversibility of the transition matrices. All that is required is stationarity.

4.1 Hierarchical–Spatial Duality

According to Theorem 2, the matrix $\mathbf{R}$ satisfies the conditions (on $\mathbf{H}$) for strong reversibility. We are thus led to define a *spatial transition matrix* $\mathbf{S}$ with components given by

$$S_{j\,k} = R_{j\,k}/p_j\,, \quad j,k = 1,2,\cdots,K\,.$$

Now, $\mathbf{S}$ is a strongly reversible row stochastic matrix with stationary vector $\mathbf{p}$ and we have a duality between hierarchical and spatial transitions:

$$\begin{aligned}\mathbf{H} &\iff (\mathbf{p},\mathbf{G}) \\ \mathbf{R} &\iff (\mathbf{p},\mathbf{S})\,.\end{aligned}$$

The duality suggests that we consider the spatial analogue of the matrix $\widetilde{\mathbf{H}}$, namely,

$$\widetilde{\mathbf{R}} = \frac{1}{\sqrt{\mathbf{P}}}\mathbf{R}\frac{1}{\sqrt{\mathbf{P}}} = \sqrt{\mathbf{P}}\,\mathbf{S}\frac{1}{\sqrt{\mathbf{P}}}\,. \tag{7}$$

See equation (4). As above, $\widetilde{\mathbf{R}}$ is positive definite and has the same eigenvalues as $\mathbf{S}$. Our main result is that $\mathbf{G}$, $\mathbf{H}$, and $\widetilde{\mathbf{H}}$ are all determined by $\mathbf{R}$; in particular, $\widetilde{\mathbf{H}}$ is the square root of $\widetilde{\mathbf{R}}$.

Theorem 3. *The parameters of the self-similar strongly reversible hierarchical transition model are determined by the matrix* $\mathbf{R}$ *as follows:*

1. $\widetilde{\mathbf{H}} = \sqrt{\widetilde{\mathbf{R}}} = \sqrt{\frac{1}{\sqrt{\mathbf{P}}}\mathbf{R}\frac{1}{\sqrt{\mathbf{P}}}}\,.$
2. $\mathbf{H} = \sqrt{\mathbf{P}}\sqrt{\frac{1}{\sqrt{\mathbf{P}}}\mathbf{R}\frac{1}{\sqrt{\mathbf{P}}}}\sqrt{\mathbf{P}}\,.$
3. $\mathbf{G} = \frac{1}{\sqrt{\mathbf{P}}}\sqrt{\frac{1}{\sqrt{\mathbf{P}}}\mathbf{R}\frac{1}{\sqrt{\mathbf{P}}}}\sqrt{\mathbf{P}}\,.$

Proof. From equation (6), we have

$$\widetilde{\mathbf{R}} = \frac{1}{\sqrt{\mathbf{P}}}\mathbf{R}\frac{1}{\sqrt{\mathbf{P}}}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{\mathbf{P}}}\left(\mathbf{G}^T\mathbf{P}\mathbf{G}\right)\frac{1}{\sqrt{\mathbf{P}}} \\
&= \left(\frac{1}{\sqrt{\mathbf{P}}}\mathbf{G}^T\sqrt{\mathbf{P}}\right)\left(\sqrt{\mathbf{P}}\mathbf{G}\frac{1}{\sqrt{\mathbf{P}}}\right) \\
&= \widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}} \\
&= \widetilde{\mathbf{H}}^2 ,
\end{aligned}
$$

where we have used the symmetry of $\widetilde{\mathbf{H}}$ in the last step. Since $\widetilde{\mathbf{H}}$ is positive definite, we may take square roots of both sides of the last equation to obtain the first part of the theorem. The rest of the theorem then follows directly from the expressions for $\mathbf{H}$ and $\mathbf{G}$ in terms of $\widetilde{\mathbf{H}}$.

There is some intuition behind the Theorem. In order to make a spatial transition from a first pixel to a second adjacent pixel, the model must make two transitions: from the first pixel to the mother pixel and then from the mother pixel back to the second pixel. This may help explain why we have required the hierarchical transition matrices to be reversible and also why the hierarchical matrices are essentially square roots of the corresponding spatial matrices.

4.2 Estimation for the Self-Similar Model

Theorem 3 gives an analytic expression for the model parameters in terms of the matrix $\mathbf{R}$ and allows us to fit the model according to the following procedure:

- Estimate $\mathbf{R}$ from the floor resolution raster map $\mathcal{G}_L$. This is discussed in detail below.
- Estimate $\mathbf{p}$ as the marginal of $\mathbf{R}$, i.e., $\hat{p}_i = \hat{R}_{i\cdot}$.
- Estimate $\widetilde{\mathbf{H}}$, $\mathbf{H}$, and $\mathbf{G}$ by substituting the estimates of $\mathbf{p}$ and $\mathbf{R}$ into the equations of Theorem 3.

Regarding the first step, we may obtain an estimate of $\mathbf{R}$ by scanning the image $\mathcal{G}_L$ and recording the observed frequency of occurrence of colors j, k in adjacent pixels. We may scan from North to South or from West to East, yielding two separate sets of observed frequencies:

$$
O_{jk}^{NS} \quad \text{and} \quad O_{jk}^{WE} .
$$

The model predicts that

$$
R_{j\,k} = R_{k\,j} = E\left[\frac{O_{jk}^{NS}}{O_{\cdot\cdot}^{NS}}\right] = E\left[\frac{O_{jk}^{WE}}{O_{\cdot\cdot}^{WE}}\right] ,
$$

so we can use the estimator

$$
\hat{R}_{jk} = \frac{1}{4}\frac{O_{jk}^{NS}}{O_{\cdot\cdot}^{NS}} + \frac{1}{4}\frac{O_{kj}^{NS}}{O_{\cdot\cdot}^{NS}} + \frac{1}{4}\frac{O_{jk}^{WE}}{O_{\cdot\cdot}^{WE}} + \frac{1}{4}\frac{O_{kj}^{WE}}{O_{\cdot\cdot}^{WE}} , \tag{8}
$$

which symmetrizes across the four compass directions.

Although the hierarchical model predicts that $\mathbf{R}$ is positive definite, there is no apparent mathematical reason why the sample value must be positive definite. (We have yet to encounter an example in which positive definiteness fails, however.) Serious failure, i.e., one or more large negative eigenvalues, would be indicative of model inadequacy. If there are only a few slightly negative eigenvalues, then one might regularize by replacing the negative eigenvalues by zero or by small positive values in the spectral decomposition of the estimated $\mathbf{R}$.

5 The Multiscale HMTM Model

The previous section considered the case in which all of the transition matrices are equal. We now allow the each hierarchical transition to have its own transition matrix. We do require that (i) each transition matrix $\mathbf{G}^{[n,n+1]}$ is strongly reversible and (ii) all the transition matrices have the same stationary vector $\mathbf{p}$ which is also the initial vector in the model. This model will be referred to as the *multiscale* HMTM model. On the basis of the floor resolution map $\mathcal{G}_L$, we will develop procedures for estimating the transition matrices at the different scales.

The analysis in the previous section really obtained an estimator for the last transition matrix. Equality of all the transition matrices was used only to ensure stationarity. The estimate of this last transition matrix was based on the joint distribution of colors in adjoining, or nearest neighbor, pixels. So it is natural to estimate the transition matrix at other hierarchical levels by the joint distribution of colors in a pair of more widely separated pixels. Specifically, we say that two pixels are neighbors at spatial scaling level n, $n = 1, \cdots, L$, if the pixels are in the same row or the same column of the raster map and if there are exactly $2^{n-1} - 1$ pixels lying between them. Thus, two pixels are neighbors at spatial scaling level 1 if they are adjacent, and this means that they can be regarded as having the same mother pixel in the hierarchical sense. Similarly, two pixels are neighbors at spatial scaling level 2 when there is exactly one pixel between them; in this case, they can be regarded as having the same grandmother in the hierarchical sense. This correspondence between spatial and hierarchical levels continues: when two pixels are neighbors at spatial scaling level n, then it is necessary to go back n generations in the hierarchical model to arrive at a common ancestor.

Our notation for the transition matrices $\mathbf{G}^{[n,n+1]}$ is a bit elaborate, so we are going to change the notation for typographical simplicity. The new notation will also reflect the order reversal between the spatial scaling levels and the hierarchical levels. We write

$$\begin{aligned}
\mathbf{G}_1 &= \mathbf{G}^{[L-1,L]} \\
\mathbf{G}_2 &= \mathbf{G}^{[L-2,L-1]} \\
&\cdots \\
\mathbf{G}_L &= \mathbf{G}^{[0,1]}\,.
\end{aligned}$$

The components of $\mathbf{G}_n$ will be denoted by G_{njk}. Similar notation will be employed for the matrices $\mathbf{H}$ and $\widetilde{\mathbf{H}}$.

Let $\mathbf{R}_n = [R_{njk}]$ specify the joint distribution of colors i, j in pairs of pixels which are neighbors at spatial scaling level n. Thus, $\mathbf{R}_1$ is the same as the matrix $\mathbf{R}$ from the previous section. The equation (5) can be extended to broader spatial scales. For example,

$$R_{2jk} = \sum_{i,j_1,k_1=1}^{K} p_i \cdot G_{2ij_1} G_{1j_1j} \cdot G_{2ik_1} G_{1k_1k}\,,$$

or, as a matrix equation,

$$\mathbf{R}_2 = \mathbf{G}_1^T \mathbf{G}_2^T \mathbf{P} \mathbf{G}_2 \mathbf{G}_1\,. \tag{9}$$

It is evident that this last equation extends to broader spatial scales, e.g.,

$$\begin{aligned} \mathbf{R}_3 &= \mathbf{G}_1^T \mathbf{G}_2^T \mathbf{G}_3^T \mathbf{P} \mathbf{G}_3 \mathbf{G}_2 \mathbf{G}_1\,. \\ \mathbf{R}_4 &= \mathbf{G}_1^T \mathbf{G}_2^T \mathbf{G}_3^T \mathbf{G}_4^T \mathbf{P} \mathbf{G}_4 \mathbf{G}_3 \mathbf{G}_2 \mathbf{G}_1\,. \\ &\text{etc.} \end{aligned} \tag{10}$$

It follows immediately that the analogue of Theorem 2 holds for the matrices $\mathbf{R}_n$: they are all symmetric and positive definite with $\mathbf{p}$ as their marginal probability vector.

Recalling that $\widetilde{\mathbf{H}} = \sqrt{\mathbf{P}}\mathbf{G}\frac{1}{\sqrt{\mathbf{P}}}$, we may insert factors of $\frac{1}{\sqrt{\mathbf{P}}}\sqrt{\mathbf{P}}$ into the equations (11) to obtain the corresponding equations for the $\widetilde{\mathbf{H}}$ and $\widetilde{\mathbf{R}}$ matrices:

$$\begin{aligned} \widetilde{\mathbf{R}}_1 &= \widetilde{\mathbf{H}}_1^2 \\ \widetilde{\mathbf{R}}_2 &= \widetilde{\mathbf{H}}_1 \widetilde{\mathbf{H}}_2^2 \widetilde{\mathbf{H}}_1 \\ \widetilde{\mathbf{R}}_3 &= \widetilde{\mathbf{H}}_1 \widetilde{\mathbf{H}}_2 \widetilde{\mathbf{H}}_3^2 \widetilde{\mathbf{H}}_2 \widetilde{\mathbf{H}}_1 \\ \widetilde{\mathbf{R}}_4 &= \widetilde{\mathbf{H}}_1 \widetilde{\mathbf{H}}_2 \widetilde{\mathbf{H}}_3 \widetilde{\mathbf{H}}_4^2 \widetilde{\mathbf{H}}_3 \widetilde{\mathbf{H}}_2 \widetilde{\mathbf{H}}_1 \\ &\text{etc.} \end{aligned} \tag{11}$$

But we can solve these equations recursively for the $\widetilde{\mathbf{H}}$ matrices in terms of the $\widetilde{\mathbf{R}}$ matrices and earlier $\widetilde{\mathbf{H}}$ matrices, as follows:

$$\begin{aligned} \widetilde{\mathbf{H}}_1^2 &= \widetilde{\mathbf{R}}_1 \\ \widetilde{\mathbf{H}}_2^2 &= \widetilde{\mathbf{H}}_1^{-1} \widetilde{\mathbf{R}}_2 \widetilde{\mathbf{H}}_1^{-1} \\ \widetilde{\mathbf{H}}_3^2 &= \widetilde{\mathbf{H}}_2^{-1} \widetilde{\mathbf{H}}_1^{-1} \widetilde{\mathbf{R}}_3 \widetilde{\mathbf{H}}_1^{-1} \widetilde{\mathbf{H}}_2^{-1} \\ \widetilde{\mathbf{H}}_4^2 &= \widetilde{\mathbf{H}}_3^{-1} \widetilde{\mathbf{H}}_2^{-1} \widetilde{\mathbf{H}}_1^{-1} \widetilde{\mathbf{R}}_4 \widetilde{\mathbf{H}}_1^{-1} \widetilde{\mathbf{H}}_2^{-1} \widetilde{\mathbf{H}}_3^{-1} \\ &\text{etc.} \end{aligned} \tag{12}$$

Both sides of each equation in the system (12) are positive definite so we can take square roots to obtain explicit recursive solutions:

$$\begin{aligned} \widetilde{\mathbf{H}}_1 &= \sqrt{\widetilde{\mathbf{R}}_1} \\ \widetilde{\mathbf{H}}_2 &= \sqrt{\widetilde{\mathbf{H}}_1^{-1} \widetilde{\mathbf{R}}_2 \widetilde{\mathbf{H}}_1^{-1}} \end{aligned}$$

$$\begin{aligned}
\widetilde{\mathbf{H}}_3 &= \sqrt{\widetilde{\mathbf{H}}_2^{-1}\widetilde{\mathbf{H}}_1^{-1}\widetilde{\mathbf{R}}_3\widetilde{\mathbf{H}}_1^{-1}\widetilde{\mathbf{H}}_2^{-1}} \\
\widetilde{\mathbf{H}}_4 &= \sqrt{\widetilde{\mathbf{H}}_3^{-1}\widetilde{\mathbf{H}}_2^{-1}\widetilde{\mathbf{H}}_1^{-1}\widetilde{\mathbf{R}}_4\widetilde{\mathbf{H}}_1^{-1}\widetilde{\mathbf{H}}_2^{-1}\widetilde{\mathbf{H}}_3^{-1}} \\
&\quad \text{etc.}
\end{aligned} \tag{13}$$

It should be clear how the estimation is carried out. The image is scanned to obtain estimates of the matrices $\mathbf{R}_n$ and of the corresponding matrices $\widetilde{\mathbf{R}}_n$, which are then substituted into the equations (13).

6 HMTM Model for Binary Maps

We now elaborate on the case of the binary HMTM model since 2×2 transition matrices have a simple and explicit eigenstructure. Let A and B stand for the two categories in the binary map. An arbitrary 2×2 row stochastic matrix has the form

$$\mathbf{G} = \begin{bmatrix} 1-\delta_A & \delta_A \\ \delta_B & 1-\delta_B \end{bmatrix}, \tag{14}$$

where $0 \leq \delta_A, \delta_B \leq 1$. Here, δ_A is the probability of a transition from A to B while δ_B is the probability of a transition from B to A. In general, the non-unit eigenvalue of a row stochastic 2×2 matrix is the difference between the $(2,2)$ and $(1,2)$ entries. This gives

$$\lambda = 1 - \delta_A - \delta_B$$

as the non-unit eigenvalue of $\mathbf{G}$ which is subject to the constraint $-1 \leq \lambda \leq 1$. As long as $\lambda \neq 1$, $\mathbf{G}$ has a unique stationary vector whose components are given by

$$\begin{aligned}
p_A &= \frac{\delta_B}{1-\lambda} \\
p_B &= \frac{\delta_A}{1-\lambda}\,.
\end{aligned}$$

When $\lambda = 1$, $\mathbf{G}$ is the identity matrix and every two-component probability vector is stationary for $\mathbf{G}$. In either case, we have

$$\begin{aligned}
\delta_A &= p_B(1-\lambda) \\
\delta_B &= (1-p_B)(1-\lambda)\,,
\end{aligned} \tag{15}$$

so that $\mathbf{G}$ can be parametrized by p_B and λ instead of δ_A and δ_B. For given p_B in the range $0 < p_B < 1$, the eigenvalue λ must satisfy

$$1 - \min\left\{\frac{1}{1-p_B}, \frac{1}{p_B}\right\} \leq \lambda \leq 1\,. \tag{16}$$

But because we require nonnegative eigenvalues in the HMTM model, the lower bound on λ is automatically satisfied. Thus, the constraints in our parametrization are

$$0 < p_B < 1 \qquad \text{and} \qquad 0 \leq \lambda \leq 1\,.$$

Two by two transition matrices always satisfy the condition of detailed balance; but one can check this by noting that the matrix $\mathbf{H}$ is given by

$$\begin{aligned}\mathbf{H} &= \begin{bmatrix} (1-p_B)(1-\delta_A) & (1-p_B)\delta_A \\ p_B\delta_B & p_B(1-\delta_B) \end{bmatrix} \\ &= \begin{bmatrix} 1-p_B-\Delta & \Delta \\ \Delta & p_B-\Delta \end{bmatrix},\end{aligned}$$

where $\Delta = p_B(1-p_B)(1-\lambda)$. Since this last matrix expression is manifestly symmetric, reversibility is established.

The binary HMTM model is therefore parametrized by

$$p_B \qquad \text{and} \qquad \lambda_1, \lambda_2, \cdots, \lambda_L\,,$$

where λ_n is the non-unit eigenvalue of the transition matrix $\mathbf{G}_n = \mathbf{G}^{[L-n,L-n+1]}$. Note that the λ parameters are numbered from the floor resolution upward: λ_1 is the eigenvalue for the last transition while λ_L is for the first transition. The model is self-similar when all the λ_n are equal.

The parameters λ_n control the distance-lagged autocorrelation in the floor resolution map $\mathcal{G}_L$. Fix two pixels s and t in $\mathcal{G}_L$ and let $1_B(s)$ and $1_B(t)$ be the indicators of color B on these pixels. Suppose it is necessary to go up h levels in the quadtree hierarchy to reach a common ancestor of s and t. We call h the quadtree or genealogical distance between s and t. It is a true distance function and satisfies a strong form of the triangle inequality known as the ultrametric condition, specifically,

$$\mathrm{dist}(x,z) \le \max\{\mathrm{dist}(x,y),\, \mathrm{dist}(y,z)\} \le \mathrm{dist}(x,y) + \mathrm{dist}(y,z)\,.$$

Ultrametrics are employed in the dendrograms of classification theory (Legendre and Legendre, 1998, p. 313) and the valuations of $\wp$-adic analysis (Lang, 1965, chap. XII; Schikhof, 1984, chap. 1). Patil and Taillie (2000b) show that the correlation between $1_B(s)$ and $1_B(t)$ is determined by their quadtree distance and is given by

$$\mathrm{Corr}\,(1_B(s)\,, 1_B(t)) = \lambda_L^2 \lambda_{L-1}^2 \cdots \lambda_h^2\,.$$

In particular, for the self-similar model, this becomes

$$\mathrm{Corr}\,(1_B(s),\, 1_B(t)) = \lambda^{2h}\,,$$

where h is the quadtree distance between the pixels. Note that the λ parameters in these last two equations are squared. The sign of λ cannot be determined from the floor resolution map alone. This is why we have imposed positive definiteness of $\mathbf{H}$ in order to obtain identifiability of the HMTM model from the floor resolution map.

References

[1] Bailey, N. T. J. (1964). *The elements of Stochastic Processes*. Wiley, New York.

[2] Bremaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues.* Springer, New York.

[3] 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.

[4] Johnson, G. D., Myers, W. L., Patil, G. P., and Taillie, C. (2000). Fragmentation profiles for real and simulated landscapes. *Environmental and Ecological Statistics* (to appear).

[5] Kijima, M. (1997). *Markov Processes for Stochastic Modeling*. Chapman & Hall, London.

[6] Lang, S. (1965). *Algebra.* Addison-Wesley, Reading, MA.

[7] Lange, K. (1999). *Numerical Analysis for Statisticians.* Springer, New York.

[8] Legendre, P. and Legendre, L. (1998). *Numerical Ecology*, second English edition. Elsevier, New York.

[9] 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.

[10] Patil, G. P. and Taillie, C. (2000a). 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.

[11] Patil, G. P. and Taillie, C. (2000b). Properties of Binary Raster Maps Generated from the HMTM Model. Technical Report Number 2000-0702, Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University, University Park, PA 16802.

[12] Patil, G. P. and Taillie, C. (2000c). Testing for Self-similarity of Thematic Raster Maps Using the HMTM Model. Technical Report Number 2000-0703, Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University, University Park, PA 16802.

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

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

[15] Schikhof, W. H. (1984). *Ultrametric Calculus: An Introduction to $\wp$-adic Analysis*. Cambridge University Press, Cambridge.

[16] Schinazi, R. B. (1999). *Classical and Spatial Stochastic Processes.* Birkhauser, Boston.

[17] Serfozo, R. (1999). *Introduction to Stochastic Networks*. Springer, New York.

[18] Varga, R. S. (2000) *Matrix Iterative Analysis*, second edition. Springer, New York.