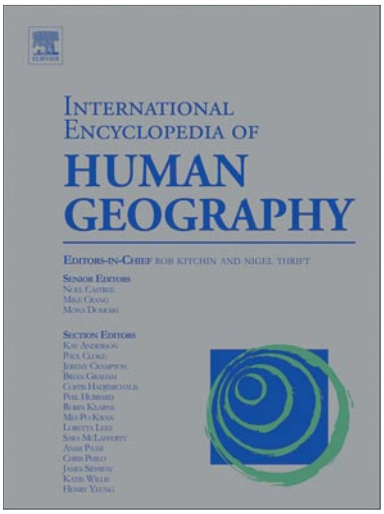

Li W, Zhang C. 2009. Markov Chain Analysis. In Kitchin R, Thrift N (eds) International Encyclopedia of Human Geography, Volume 6, pp. 455–460. Oxford: Elsevier.
ISBN: 978-0-08-044911-1

Markov Chain Analysis

W. Li and C. Zhang, University of Connecticut, Storrs, CT, USA

Glossary

Algorithm A step-by-step problem-solving procedure, especially, an established, recursive, computational procedure for solving a problem.

Correlation Statistical techniques or measures which show whether, and how strongly, pairs of variables are related.

Geostatistics A collection of statistical methods which were traditionally used in geosciences, describing spatial correlation among sample data and using it in various types of spatial models.

Markov Chain Random Field A single Markov chain that moves or jumps in a space and at any location, interacts with its nearest known neighbors in different directions, and decides its state by the interactions.

Nearest Known Neighbor A location with known value which is the nearest along one direction to the location to be estimated and which may be a sampled location or a previously estimated location.

Nonlinear Property of a kind of system whose behavior is not expressible as a sum of the behaviors of its descriptors.

Range The distance in which the difference of the variogram or transiogram from the sill gets neglectable.

Simulation An imitation of some real thing or process, which generally entails representing certain key characteristics or behaviors of a selected physical or abstract system.

Sill The limit of the variogram or transiogram tending to infinite lag distances.

Transiogram It refers to a transition probability diagram. Theoretically, it is defined as a two-point conditional probability function $p_{ij}(h)$ over the distance lag h.

Introduction

Probabilistic models are appropriate in human geography for obvious reasons: geographical concepts and measurements tend to be inexact, geographical relationships are often complicated and poorly understood, and geographical manifestations of human behavior remain unpredictable despite much research. It is convenient to treat the processes of changes as random in aggregate and to describe them in a form of some stochastic process describing the changes in the locational pattern. The Markov chain theory is the simplest process of this kind.

Since Brown's analysis of innovation diffusion in 1963, Markov chains entered the geographical literature. Then one-dimensional Markov chains became popular tools in the late 1960s and early 1970s in human geography for describing and modeling social mobility, such as migration, city growth, and changes in population distribution, residential structure, transport networks, industrial structure and pattern, and land use. Collins *et al.* provided a review on applications of Markov models in geography in 1974. However, the applications of Markov chains in human geography faded relatively since the late 1970s. The main reason may be related with the limitations of one-dimensional first-order stationary Markov chains and the difficulties in constructing high-order or multidimensional Markov chain models.

Recently, a Markov chain random field (MCRF) theory was suggested. This new Markov chain theory extended a single Markov chain for multidimensional modeling. The measure for MCRFs is called 'transiogram' (i.e., *transi*tion probability dia*gram*), rather than the conventionally used 'transition probability matrix' (TPM). Transiogram solves the major technical problem in estimation of transition probabilities in many geographical applications of Markov models. It overcomes the limitation of TPM that it can only be used to provide estimates for time periods equal to the time interval of the input time series data or lag distances equal to the sampling interval of the input spatial data. For example, a 1981–85 land-use change matrix can only provide quinquennial estimates. However, transiogram provides a way for estimating continuous transition probabilities (with any time or space separation) from samples. Transiogram also provides a visual spatial relationship measure for categorical data.

One-Dimensional Markov Chains

A Markov chain represents a sequence of random variables X_1, X_2, X_3, … with the Markov property. The Markov property means that given the present state, the future state is independent of past states. Here a 'state' refers to a category defined by users, such as a type of land use or a status (e.g., move, stay) of a resident. For a first-order Markov chain, we have

$$\Pr(X_{n+1} = x | X_n = x_n, \ldots, X_1 = x_1) = \Pr(X_{n+1} = x | X_n = x_n) \quad [1]$$

The possible values of X_i form a finite set of states S, called the state space of the chain. The description tool of a Markov chain is its TPM. For a stationary Markov chain, its TPM does not change with time (or space). Below is a TPM of a discrete-time (or space), two-state stationary first-order Markov chain defined on the state space $S=(1, 2)$:

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \quad [2]$$

This means that if the present state of the Markov chain is 1, it will have a probability of 0.7 moving to the same state (i.e., state 1) and a probability of 0.3 moving to state 2 at the next time (or space) step. If the present state of the Markov chain is 2, it will have a probability of 0.6 moving to the same state and a probabilty of 0.4 moving to state 1 at the next time (or space) step.

Stationary Markov chains have a property – after sufficient transition steps, their transition probabilities reach an unchangeable state and the different rows become the same. Such a row of transition probabilities is called an equilibrium vector, where each element w_i represents the 'proportion' of the corresponding state i. The equilibrium vector can be obtained by multiplying the TPM with itself sufficient times until the resultant TPM has no changes in its element values.

The above TPM in eqn. [2] has diagonal entries (i.e., p_{11} and p_{22}), which means the Markov chain may move from a state to the same state (i.e., self-transition). If the self-transition probability of a Markov chain on state i is p_{ii}, the probability that the number of consecutive self-transitions of the Markov chain on state i is k, which obeys the geometric distribution

$$P(T_i = k) = \theta_i(1-\theta_i)^{k-1}, k = 1, 2, \ldots \quad [3]$$

where $\theta_i=(1-p_{ii})$. The geometric distribution is a discrete analog of an exponential distribution. When using such a Markov chain model to simulate the spatial sequence of a phenomenon, the sequences of states, such as *AAABBBBAABBBBB*..., are generated, and the numbers of consecutive self-transitions of a state such as *B* obey the geometric distribution.

The number of self-transitions on a state represents the 'wait' time (or distance) of the Markov chain on the state. For studying land-use change, this 'wait' time refers to the time (e.g., years) of the land to be continuously used in one way (e.g., cropping). In many cases, 'wait' time may not follow a geometric distribution. Thus, the above Markov chain model may not be suitable for all cases. An appropriate Markov chain model for cases where the geometric distribution cannot be applied may be an embedded Markov chain, which considers only transitions between different states in its TPM and deals with 'wait' time in a user-defined way.

For a three-state, stationary, embedded Markov chain, its TPM can be expressed as

$$Q = \begin{bmatrix} - & q_{12} & q_{13} \\ q_{21} & - & q_{23} \\ q_{31} & q_{32} & - \end{bmatrix} \quad [4]$$

In the above TPM, only transitions between different states are counted and self-transitions are ignored. When using such a Markov chain model to simulate the spatial sequence of a phenomenon, only the sequences of different states, such as *ABCBACBACA*..., are generated. To describe the 'wait' time of each state, an appropriate probability distribution chosen by users, such as a lognormal distribution, may be used. We found no literature in human geography that used the embedded Markov chain approach. But such a method may find its usefulness in the future because 'wait' time in some phenomena in human geography may not fit a geometric or exponential distribution.

Transiograms

A transiogram refers to a transition probability diagram (a curve or a series of points, of transition probabilities with increasing separate distance) for characterizing the autocorrelation of a class (or state or category) or the cross-correlation from a class to another class. Limitations in data collection make it impossible to estimate a continuous measure directly from sparse sample data. Transiogram overcomes the data limitation and can be used to detect and model spatial autocorrelation or cross-correlation at different scales from sample data through inferring transition probability models. A transiogram can be represented as a transition (or conditional) probability function on a continuous lag $\mathbf{h}$:

$$p_{ij}(\mathbf{h}) = \Pr(z(\mathbf{x}+\mathbf{h}) = j | z(\mathbf{x}) = i) \quad [5]$$

where z is a reliazation of the random z at a specific location $\mathbf{x}$. Here Z may be one-dimensional or multidimensional. The second order stationarity assumption is applicable here so that $p_{ij}(\mathbf{h})$ is dependent only on the lag $\mathbf{h}$, not on the location $\mathbf{x}$. An auto-transiogram $p_{ij}(\mathbf{h})$ represents the self-dependence (i.e., auto-correlation) of a single class i and a cross-transiogram $p_{ij}(\mathbf{h})$ $(i \neq j)$ represents the cross-dependence of class j on class i. Here class i is called a head class and class j is called a tail classes. i and j are not interchangeable here because cross-transiogram is normally asymmetric.

Transiograms directly estimated from sample data are called 'experimental transiograms'. Continuous transiogram models can be acquired by using mathematical models to fit experimental transiograms. Basic mathematical models for modeling experimental transiograms include the exponential model and the spherical model. Expert knowledge may be incorporated in the estimation of transiogram models. Here expert knowledge refers to the knowledge of experts in parameter estimation of transiogram models, which typically include sills, ranges, and model types (e.g., exponential, spherical). Figure 1 shows that an experimental autotransiogram and an experimental cross-transiogram are approximately fitted by basic mathematical models. Here, the sill of a transiogram refers to the stable height that the transiogram gradually approaches. Theoretically, the sill of a transiogram should be equal to the proportion of the corresponding tail class. The range of a transiogram refers to the lag distance where the transiogram approaches the sill. When the separate distance between two data is less than the range, they are considered spatially dependent (or correlated); otherwise, they may be considered independent.

Transiograms have several uses. One use is to estimate transition probabilities, particularly multistep transition probabilities, from insufficient data. When data are insufficient, transition probabilities estimated from data are inaccurate. Under this situation, reasonable transition probabilities can be obtained by using mathematical models to fit experimental transiograms; thus characteristics of a large set of geographic data (e.g., population) can be inferred using a small sample set. For example, the racial migration probabilities in a large city may be inferred from a small sample set by using transiograms.

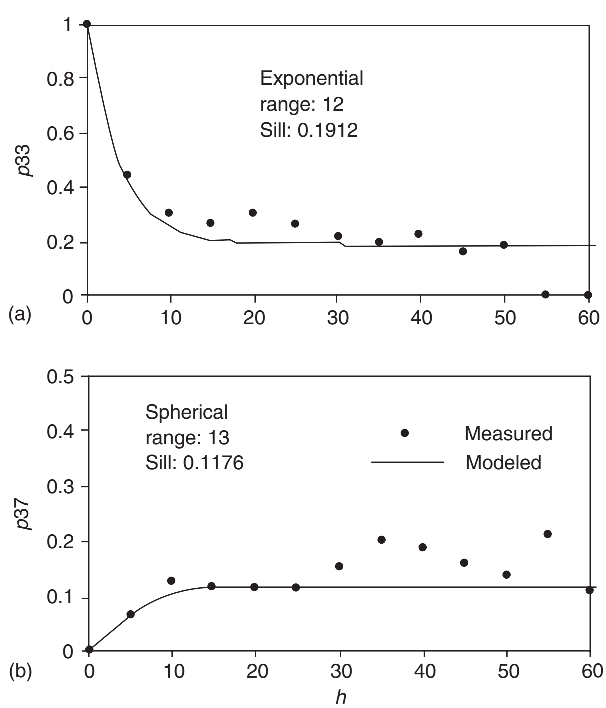

Figure 1 Transiogram modeling by mathematical models: (a) autotransiogram, (b) cross-transiogram. The scales along h axis are number of pixels.

The second use is to get continuous-lag transition probabilities from sample data through modeling. Usually sample data can only provide us transition probabilities with one or several specific lags (e.g., one-step or n-step transition probabilities), which is a severe limitation when transition probabilities at other lags are needed. The third use of transiograms is to characterize spatial (or temporal) heterogeneity of discrete variables. Correlation ranges and sills of transiograms and transiogram shapes are all reflections of spatial heterogeneity. In addition, class polygon size (mean length) may also be inferred from autotransiograms. Thus, transiogram may be used to infer polygon sizes of different land-use classes (e.g., residential, commercial, industry, park, and agriculture) or different urban planning district sizes with certain residential characteristics (low, medium, and high density). The fourth use of transiograms is to provide input transition probabilities to Markov chain simulation. Transiogram models can provide transition probabilities with any lag distance. The fifth use of transiograms is for data mining in large datasets. Through estimating transiograms from a large dataset of a categorical variable, one may find autocorrelation properties of single classes and complex relationships between different classes hid in the dataset. Thus, it may be used to determine if the magnitude or frequency level of some phenomena differ from one location to another. Therefore, it is expected that transiograms will be a useful spatial statistical tool in human geography to make generalizations concerning complex spatial patterns for studies of diseases, urban growth, land-use change, residential mobility, socio-spatial segregation, and vehicle traffic, etc.

Markov Chain Random Fields

The MCRF theory extends a single Markov chain for any dimensional modeling. The general solution of the conditional probability distribution of a MCRF Z at an unsampled location x was derived as

$$\Pr(z(\mathbf{x}) = k | z(x_1) = l_1, \ldots, z(\mathbf{x_m}) = l_m) = \frac{\prod_{i=2}^{m} p_{kl_i}(\mathbf{h_i}) \cdot p_{l_1 k}(\mathbf{h_1})}{\sum_{f=1}^{n} \left[\prod_{i=2}^{m} p_{fl_i}(\mathbf{h_i}) \cdot p_{l_1 f}(\mathbf{h_1})\right]} \quad [6]$$

where $p_{kl_i}(\mathbf{h}_i)$ represents a transition probability in the ith direction from state k to state l_i with a lag $\mathbf{h}_i$; $\mathbf{x}_1$ represents the neighbor from which the Markov chain moves to the current location x; m represents the number of nearest known neighbors (or locations); k, l_i, and f represent states in the state space $S = (1, \ldots, n)$; $\mathbf{h}_i$ is the distance from the current location to its nearest known neighbor $\mathbf{x}_i$. With increasing lag h, any $p_{kl}(\mathbf{h})$ forms a transiogram which represents spatial (auto or cross) correlation of classes. It can be seen that the conditional probability distribution equation of the MCRF is actually composed of transiograms. With changes of the number m and the directions of nearest known neighbors, the above general solution actually includes a set of different MCRF models for one-dimensional and multidimensional modeling. It is clear that MCRF models are nonlinear.

In practical use, however, the above general solution cannot be simply implemented directly, because it is necessary to consider a limited number of nearest known neighbors and the conditional independence of nearest known neighbors in cardinal directions for optimal simulations. Normally, considering four orthogonal cardinal directions is sufficient for two-dimensional modeling. So if only the nearest data locations in four cardinal directions are considered, the MCRF model in eqn. [6] is simplified as

$$\Pr(z(\mathbf{x}) = k | z(\mathbf{x}_1) = l, z(\mathbf{x}_2) = m, z(\mathbf{x}_3) = q, z(\mathbf{x}_4) = o) = \frac{p_{ko}(\mathbf{h}_4) \cdot p_{kq}(\mathbf{h}_3) \cdot p_{km}(\mathbf{h}_2) \cdot p_{lk}(\mathbf{h}_1)}{\sum_{j=1}^{n} \left[p_{fo}(\mathbf{h}_4) \cdot p_{fq}(\mathbf{h}_3) \cdot p_{fm}(\mathbf{h}_2) \cdot p_{lf}(\mathbf{h}_1) \right]} \quad [7]$$

where 1, 2, 3, and 4 represent the four cardinal directions considered. In directions 2, 3, and 4, transitions are from the current unknown location $\mathbf{x}$ to its nearest known neighbors, but in direction 1 (i.e., the coming direction of the Markov chain), the transition is from the nearest known neighbor $\mathbf{x}_1$ to the current location $\mathbf{x}$.

In simulation, a cardinal direction is replaced by a search sector to cover the whole search area (usually a circle) (Figure 2). Thus, four search sectors that equally split a search circle may be used to search for the nearest known neighbors, one from each search sector, for estimating the conditional probability distribution of a random variable at an unsampled location. In addition, because the nearest known neighbors found within a search radius may not always reach four, the needed MCRF models may be further simplified from eqn. [7]. A Markov chain sequential simulation algorithm can be used to conduct simulation.

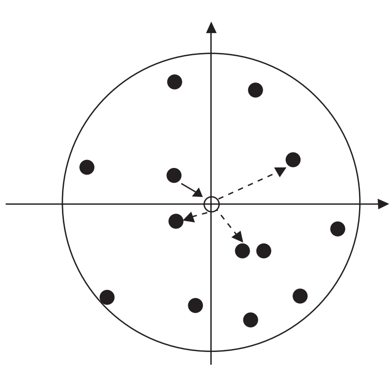

Figure 2 The searching sectors and neighborhood used in the random-path Markov chain sequential simulation algorithm.

Figure 3 shows simulated results of seven land-cover classes using a random path sequential simulation algorithm. The simulation was conditioned on a random sample set of 130 point data in a 35 km^2 area (a 295 by 295 lattice). Apparently, the patterns in simulated realizations are polygonal. Polygonal patterns are in accordance with the custom of area-class mapping and are also convenient for human understanding and data processing using GIS tools. Different realizations are imitative of each other but with apparent differences in details, which demonstrates the different possible configurations of the land-cover classes in reality. The optimal map based on maximum occurrence probabilities represents the best estimate based on samples. When samples are very sparse or simulation is unconditional, sills of transiogram models will have strong influence on the proportions of classes in simulated results. But with the number of conditioning samples increasing, the proportions of classes in the conditioning sample set will play a major role in determining simulated class proportions.

Although single realizations can represent spatial uncertainty by their differences in spatial distribution of classes, a more accurate and vivid way to quantify and demonstrate spatial uncertainty is using occurrence probability maps, which can be estimated from a large number of realizations or directly calculated. As shown in Figure 4, the maximum occurrence probability map represents the purity (or quality) of the optimal map, and occurrence probability maps of single classes represent uncertainty of spatial distribution of each class. In the maximum occurrence probability map, the white–gray stripes actually indicate the approximate location of class boundaries, which are called ‘transition zones’.

MCRF models have been used to model categorical geographical variables and may also be extended to model continuous geographical variables. It is expected that MCRF models will be useful in varied branches of human geography, particularly those involving the spread of disease (epidemiology), the practice of commerce planning, and exploration of traffic patterns in an urban core. However, how to use them and exactly where to use them are issues for human geographers to explore. For example, if one wants to map the spatial distribution of apartment rents or land prices in a city from a sample dataset, one may consider using two-dimensional MCRF models.

Issues

One-dimensional Markov chains have been used for decades in human geography and were proved useful in describing and modeling geographical phenomena and processes, such as patterns of human migration, reproductive behavior, and intergenerational occupational

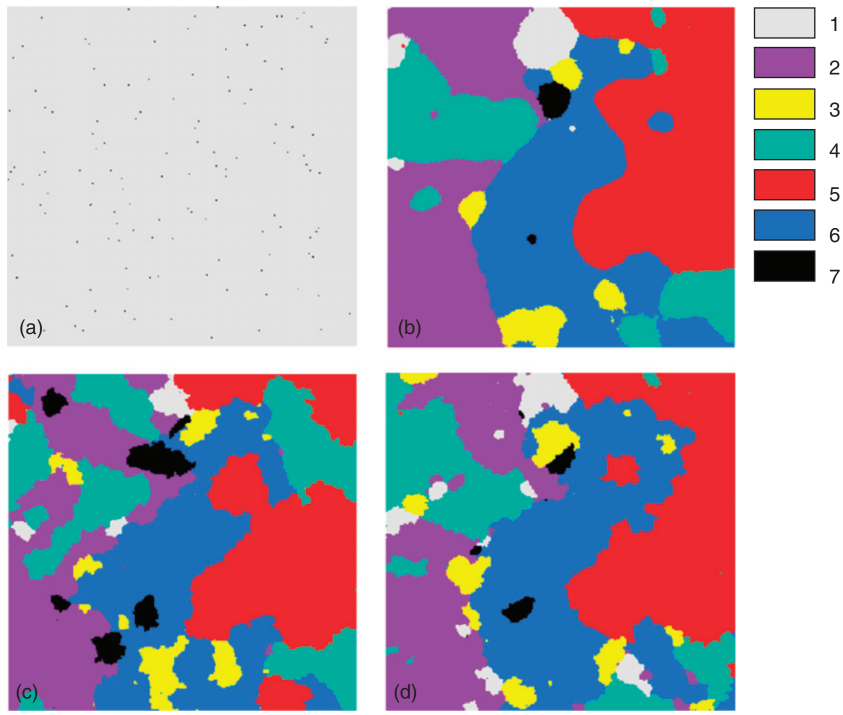

Figure 3 Simulated results using a random-path Markov chain geostatistical algorithm, conditioned on a random sample set of 130 points: (a) samples; (b) optimal map; (c) and (d) realizations.

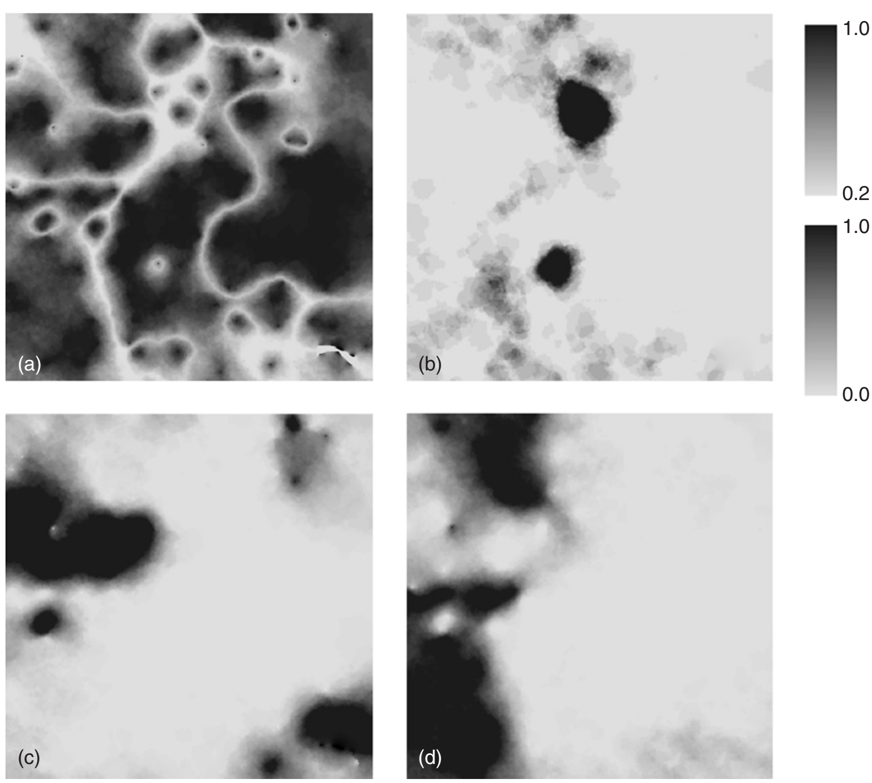

Figure 4 Occurrence probability maps generated using a random-path Markov chain geostatistical algorithm, conditioned on a random sample set of 130 points (correspond to Figure 3): (a) maximum occurrence probability map; (b), (c), and (d) occurrence probability maps of classes 7, 4, and 2, respectively.

mobility. However, it is still difficult to build high-order Markov chain models to deal with the complex property of social phenomena or processes.

Transiogram provides a way for estimating continuous transition probabilities from sparse samples and expert knowledge and for spatial heterogeneity characterization of categories. Considering the previous applications of TPM in human geography and the advantages of transiogram over TPM, it is expected that transiogram will also be a useful tool in human geography for measuring spatiotemporal changes. However, transiogram is just a two-point spatial measure. For measuring complex geographical phenomena and processes, it is obvious that multipoint statistics are needed.

There is a need for methods of analyzing large-scale spatiotemporal dependence in geographical studies because strong spatiotemporal dependence or autocorrelation exists in areally distributed variables. Because MCRF theory extended a single Markov chain for multidimensional simulation, it should be useful in dealing with multidimensional human geographical phenomena and processes, such as describing and predicting changes in land use, population distribution, residential structure, transport networks, and industrial structure and pattern. While many geographical phenomena and processes involve both spatial and temporal dependences, integrating the time dimension and space dimensions into the same MCRF model will be desirable.

See also: First Law of Geography; Kriging and Variogram Models; Monte Carlo Simulation; Simulation; Spatial Interpolation; Uncertainty.

Further Reading

Brown, L. A. (1970). On the use of Markov chains in movement research. *Economic Geography* 46, 393–403.

Clark, W. A. V. (1965). Markov chain analysis in geography: An application to the movement of rental housing areas. *Annals of the Association of American Geographers* 55, 351–359.

Collins, L., Drewett, R. and Ferguson, R. (1974). Markov models in geography. *The Statistician* 23, 179–209.

Li, W. (2007). Transiograms for characterizing spatial variability of soil classes. *Soil Science Society of America Journal* 71(3), 881–893.

Li, W. and Zhang, C. (2007). A random-path Markov chain algorithm for simulating categorical soil variables from random point samples. *Soil Science Society of America Journal* 71, 656–668.

Richardson, H. W. (1973). A Markov chain model of interregional savings and capital growth. *Journal of Regional Science* 13, 17–27.

Tang, J., Wang, L. and Yao, Z. (2007). Spatio-temporal urban landscape change analysis using the Markov chain model and a modified genetic algorithm. *International Journal of Remote Sensing* 28, 3255–3271.

Relevant Websites

http://en.wikipedia.org/wiki/Markov_chain
Classical Markov chains.