TeMA Journal of Land Use, Mobility and Environment

TeMA 1 (2015) 49-62
print ISSN 1970-9889, e- ISSN 1970-9870
DOI: xxxxxxxxxxxxxxxxxxxxx

review paper received 30 April 2015, accepted 2 July 2015

www.tema.unina.it

How to cite item in APA format:
Iacono, M., Levinson, D., El-Geneidy, A., Wasfi, R.. A Markov chain model of land use change in the Twin Cities, 1958-2005. *Tema. Journal of Land Use, Mobility and Environment*, 8 (3), XXX-XXX. doi: http://dx.doi.org/10.6092/1970-9870/2431

A MARKOV CHAIN MODEL OF LAND USE CHANGE

MICHAEL IACONO[A], DAVID LEVINSON[B], AHMED EL-GENEIDY[C], RANIA WASFI[D]

Department of Civil, Environmental, and Geo-Engineering, University of Minnesota
[a]e-mail: iaco0009@umn.edu
[b]e-mail: dlevinson@umn.edu

School of Urban Planning, McGill University
[c]e-mail: ahmed.elgeneidy@mcgill.ca

Department of Geography, McGill University
[d]e-mail: rania.wasfi@mail.mcgill.ca

ABSTRACT

The set of models available to predict land use change in urban regions has become increasingly complex in recent years. Despite their complexity, the predictive power of these models remains relatively weak. This paper presents an example of an alternative modeling framework based on the concept of a Markov chain. The model assumes that land use at any given time, which is viewed as a discrete state, can be considered a function of only its previous state. The probability of transition between each pair of states is recorded as an element of a transition probability matrix. Assuming that this matrix is stationary over time, it can be used to predict future land use distributions from current data. To illustrate this process, a Markov chain model is estimated for the Minneapolis-St. Paul, MN, USA (Twin Cities) metropolitan region. Using a unique set of historical land use data covering several years between 1958 and 2005, the model is tested using historical data to predict recent conditions, and is then used to forecast the future distribution of land use decades into the future. We also use the cell-level data set to estimate the fraction of regional land use devoted to transportation facilities, including major highways, airports, and railways. The paper concludes with some comments on the strengths and weaknesses of Markov chains as a land use modeling framework, and suggests some possible extensions of the model.

KEYWORDS:
Land use, Twin Cities, Statistical models, Markov chain, state dependence

1 INTRODUCTION

Modelling the dynamics of land use change in urban regions is an inherently difficult task. Despite improvements to the theoretical and empirical frameworks within which the problem of land use change has been cast, few researchers have been able to produce operational models with the ability to predict land use change accurately. Those who have experienced modest successes have largely done so at the expense of tractability and ease of interpretation. Meanwhile, there has been an emerging consensus that models attempting to predict land use change ought to incorporate probabilistic elements in order to make them more realistic and to represent the significant uncertainty that surrounds land development decisions. This paper describes the application of one type of probabilistic land use change model based on the notion of a Markov process. Within this process, the study area (in this case the Minneapolis-St. Paul, MN metropolitan area) is divided into a regular lattice of cells, each of which may take on one of 10 discrete land use states at any given time. At the heart of the Markov process formulation is the notion that the state of a cell at any time is a function only of its previous state. Transitions between states are governed by a matrix of transition probabilities, which are estimated based on actual land use data. Where the assumptions of the Markov process hold, the transitions of cells between states through time can be modelled and predicted as Markov chains. Markov chain models have a relatively simple and intuitive logic that makes them attractive alternatives to more complex formulations of stochastic land use models, at least for sketch planning purposes. Of interest is their ability to fore- cast over medium to long-term time horizons. In this study we use land use data for the Minneapolis-St. Paul (Twin Cities) region covering various years between 1958 and 2005 to calibrate a Markov chain model of land use change. The data represent a fine scale of spatial resolution, with the dimensions of each cell measuring 75 meters by 75 meters. This data set is applied to both "backcast" changes from the past to the present and to predict the distribution of land use decades into the future. The paper is organized as follows. The next section describes the properties of Markov chains and cites several of their applications to questions of urban land use. The third section formally introduces the structure of the model and the assumptions required for its application. The fourth section describes the cell- level data set constructed for this study, and uses it to develop an estimate of the amount of land use in the region devoted to transportation. the fifth section describes the results of the application of the Markov chain model to the regional land use data, generating historical predictions based on earlier periods of data and using more recent data to forecast several periods into the future. The sixth, and concluding, section comments on the strengths and limitations of the model while also suggesting some directions in which it might be generalized in order to increase its usefulness as a planning tool.

2 MARKOV CHAINS AND LAND USE MODELING

2.1 PROPERTIES OF MARKOV CHAINS

Markov chain models are essentially projection models that describe the probabilistic movements an individual in a system comprised of discrete states. When applied to land use and many other applications, Markov chains often specify both time and a finite set of states as discrete values. Transitions between the states of the system are recorded in the form of a transition matrix that records the probability of moving from one state to another. The definition of a system as a finite Markov Chain requires a certain set of properties to hold (Stokey and Zeckhauser, 1978). These include:

- a finite number of well-defined states that mutually exclusive and collectively exhaustive (meaning that the rows of the probability matrix must sum to one);
- the probabilities of the transition matrix must be the same for any two periods;

- probabilities have no memory, that is, the state tomorrow depends only on the state today (the Markov condition);
- time periods must be uniform in length or duration.

In practice, one or more of these conditions may not be met. This is especially true in the case of land use applications, where the uneven temporal availability of data often requires relaxation of the last assumption. Moreover, the assumption regarding constant transition probabilities (or stationarity of the system) is often rejected when tested as a statistical hypothesis, yet is still included in forecasting applications. Turner (Turner, 1987) argues that, in fact, land use change is not a strictly Markovian process, though it does have some such elements. For example, the transition of a land use cell between states may be influenced by state of neighboring cells as well, sometimes referred to as the "spatial neighborhood effect". Additionally, transition rates are often not constant through time, especially over longer periods. Thus, an important question may concern the optimal length of transition periods in Markov chains. Unfortunately, the transition probabilities estimated in most empirical applications are a function of data availability and take the length of transition periods as given.

2.2 PREVIOUS APPLICATIONS

Markov chains as a modeling tool evolved out of social and economic science research dating to the late 1950s. Empirical applications of Markov chains in urban and regional analysis began appearing in the 1960s. One such early application was Clark's use of Markov chains to model the movement of rental housing in U.S. cities (Clark, 1965). Using census tract data on mean contract rents, Clark described the movement of census tracts between 10 different rent classes in four different cities (Detroit, Pittsburg, Indianapolis and St. Louis) over the period from 1940 to 1960. Another application by Lever sought to describe the decentralization of manufacturing in the Clydeside region of Glasgow, Scotland, UK (Lever, 1972). Using postal directory data on 419 manufacturing firms for the years 1959, 1964 and 1969, Lever modeled the movement of manufacturing firms between four zones of the city as both a closed and an open system, with the latter formulation allowing for firm birth, death and inter-regional migration. Applications of Markov chains to urban land use dynamics began to appear in the 1970s as an alternative to the use of large-scale urban simulations models for land use forecasting. Bourne cited the ability to incorporate elements of inertia in land use succession processes as a key advantage of Markov chain models (Bourne et al., 2000). In particular, the matrix of transition probabilities could be seen as embodying important aspects of urban land use such as the durability of housing and other building stock. This was critical, since stock adjustment processes were largely absent from previous models of land use change. Bourne illustrated these principles by estimating transition matrices with data on central city land use from the municipality of Toronto over the period from 1952 to 1962. Key findings of this study indicated that land use in developed parts of urban areas tends to stay in the same state (land use class) despite the occurrence of rebuilding or structural modification. Changes in land use that did occur tended toward more intensive uses (e.g. residential to commercial), with scattered, vacant parcels among the most likely candidates for conversion. While Bourne's study relied on parcel-level data with recorded changes to the building stock, Bell exemplified the use of remotely-sensed data and the cell-based representation of land use that is common in most contemporary studies of land use change (Bell, 1974). Bell studied land use change on San Juan Island, WA from 1949 to 1971 by breaking the study area into 100 meter-by-100 meter (1 hectare) grid cells, using the remotely-sensed land use imagery. This data was used to test for independence of current and preceding land uses for the given years. Results indicated that land uses for the later year were not independent of the preceding land use, lending support for the Markov chain formulation. Additional empirical findings on tests of stationarity of the transition matrix and a continuous time formulation of the Markov chain model, where transition probabilities are replaced by rates of change, are reported in Bell and Hinojosa (Bell and Hinojosa, 1977). More recent studies using Markov chains for land use prediction have sought to broaden the scope of application of these models and probe new kinds of questions. Turner compared the results of a Markov chain model with two other types of spatial simulation models to forecast long-term changes in landscape cover in the Piedmont region of northern Georgia (Turner, 1987). Muller and Middleton provide an application to the Niagara region of Ontario, Canada, where land use data from five different points in time between 1935 and 1981 are used to estimate a three-state Markov chain to predict the consequences of urban growth (Muller and Middleton, 1994). McMillen and McDonald demonstrated the coupling of Markov chains with regression models (McMillen and McDonald, 1991). In order to estimate the influences of land values on zoning changes they estimated a price function to predict land values, which then serve as explanatory variables for the transition probabilities of a three-by-three matrix of land use zoning change. Weng integrated the use of geographic information systems capabilities and remote sensing with a Markov chain model to predict the possible land use consequences of rapid urbanization and industrialization in the Zhujiang Delta of China (Weng, 2002). Finally, Levinson and Chen provide a Markov chain model of land use change in the Twin Cities region using historical data (Levinson and Chen, 2005). The states of the model include both a land use class and an indicator of the presence and type of highway within each cell. The model is used to demonstrate the mutually interconnected evolution of transportation networks and land use patterns.

3 THE MODEL

The basic premise of the Markov chain model is that land use at some point in the future $(t+1)$ can be determined as a function of current land use (t), or mathematically,

$$X_{t+1} = f(X_t) \tag{1}$$

where $X_{(t)} + 1$ represents the land use at time $t+1$ and X_t represents land use at time t. The structure of the Markov chain model as applied to land use change involves a vector $\boldsymbol{n}_t$ with dimension $m \times 1$ (where m represents the number of states, in this case land use classes) describing the distribution of land use among current states and an $m \times m$ matrix of transition probabilities $(\boldsymbol{P})$ that governs the probability of transition between each pair of land uses, i and j. The model can then be written as a difference equation in matrix form Baker (Baker, 1989)

$$\boldsymbol{n}_{t+1} = \boldsymbol{P}\boldsymbol{n}_t \tag{2}$$

where $\boldsymbol{n}_{t+1}$ is another $m \times 1$ column vector describing the distribution of land use at time $t+1$. Since the transitions are probabilities, it follows that:

$$\textstyle\sum_{j=1}^{m} p_{ij} = 1 \quad i = 1,2, \dots m \tag{3}$$

meaning simply that the rows of the transition matrix must sum to 1. Maximum likelihood estimates of the transition probabilities can be obtained as (Anderson and Goodman, 1957):

$$\hat{p}_{ij} = n_{ij} / \textstyle\sum_{j=1}^{n} n_{ij} \tag{4}$$

where p_{ij} is the probability of transition between i and j and n_{ij} denotes the number of transitions from i to j. These values can all be obtained empirically. To test the validity of the Markov chain model, a useful first step is to test the null hypothesis that land use at one point in time, $t+1$, is statistically independent of land use at the preceding time period, t. This test can be conducted using standard contingency table techniques for cross-classified categorical data. The expected values for each cell indicating the number of transitions between i and j can be compared with the actual number of transitions to compute the test statistic, Pearson's chi-square, which is distributed χ^2 with $(M-1)^2$ degrees of freedom, where M indicates the number of land use classes (in this case 10). Under the hypothesis of independence, the expected number of transitions in each cell of the transition matrix $\hat{m}_{ij}$ can be calculated by:

$$\hat{m}_{ij} = n_{i+}n_{+j} \tag{5}$$

where n_{i+} denotes the marginal total of transitions for the ith row of the transition matrix and n_{+j} denotes the marginal total for the jth column of the transition matrix. Using these expected values, the test statistic (K^2) then takes the form:

$$K^2 = \sum_{i=1}^{I} \sum_{j}^{J} \frac{(n_{ij} + \hat{m}_{ij})^2}{\hat{m}_{ij}} \tag{6}$$

The test statistic is typically given the notation K^2 instead of X^2 to differentiate it from its distribution, which is chi-square. The null hypothesis of independence is almost universally rejected, indicating some level of dependency between successive land use states. Another important property of Markov chains, as identified in an earlier section, is the property of stationarity, particularly as it applies to the transition probability matrix. This property is critical for applications in which a Markov chain model is to be used for forecasting. The transition probability matrix $(\boldsymbol{P})$ is assumed to remain constant in successive periods, meaning that at any future period $t+k$, the matrix of cell transitions can be obtained by multiplying the vector of current land uses, $\boldsymbol{n}_t$ nt by the transition probability matrix $\boldsymbol{P}$, raised to the kth power $(\boldsymbol{P}^k)$. In most forecasting applications, the transition probability matrix is assumed to remain constant through successive time periods, and is seldom tested empirically. This study follows the work of Bourne (1971), who compared transition matrices for successive periods using simple correlations between cells of the matrix. By expressing the elements of one matrix $(\boldsymbol{P}_{t+1,t+2})$ as a function of another $(\boldsymbol{P}_{t,t+1})$, one can provide a rough check for stationarity by determining whether the correlation between matrix elements is significantly different from a value of one. In order to use the Markov chain model for prediction, an additional stochastic element is added. Since the transition probabilities represent estimates of the likelihood of conversion from one land use state at time t to one of 10 other states at time $t+1$, a mechanism is added to introduce randomness to the model and its predictions of future states. Since each row of the transition probability matrix sums to one, predictions of future land use states are obtained by drawing a pseudorandom number between zero and one, rounded to four digits. If the number falls within the probability space allocated to a particular land use state according to the transition matrix, then that state is chosen for conversion. This process is repeated for each land use cell in the data set. Predicted land uses can then be compared to actual observed land uses to summarize the accuracy of the model's predictions..

3 DATA

The land use data employed in this study build from a previous set of land use data used by Levinson and Chen (Levinson and Chen, 2005) in an earlier study of the Twin Cities. The expanded data set comprises a time series with observations for the years 1958, 1968, 1978, 1984, 1990, 1997, 2000 and 2005. Land use data for years prior to 1984 were manually digitized from paper copies of land use maps stored at the John

R. Borchert Map Library at the University of Minnesota. Data for selected years from 1984 to 2005 were obtained from the Metropolitan Council, the Twin Cities' regional planning agency and designated metropolitan planning organization (MPO), which maintains a parcel-level land use inventory for the region that is updated every few years. The parcel-level land use data was converted to a raster format and rectified to reduce geometric distortion. Some error remains due to the manual digitization process and the lower level of accuracy associated with earlier mapmaking processes. Differences in classification schemes for land use across years were addressed by adopting a common set of 10 generalized land use classes. These land use classes, along with their adopted abbreviations, include: Airports (AIRPOR);Commercial (COMM); Highway (HWY); Industrial (INDUST); Parks (PARKS); Public (PUBLIC); Railroads (RAILWA); Residential (RES); Vacant (VAC); Water (WATER).

The data set covers a large portion of the core seven counties of the Twin Cities region. Some portions of the region could not be covered due to a need to limit the analysis to the part of the region for which common land use data sets could be acquired for each year. The portions left out of the study area are comprised mostly of low-density residential and non-urban uses, which would likely be classified as vacant under the present scheme. The resulting study area covers approximately 3,426 square kilometers (1,322 square miles). The study area is partitioned into a grid of 75-meter by 75-meter cells, a spatial resolution much finer than the 188-meter square cells used in Levinson and Chen's study, leading to a roughly tenfold increase in the number of land use cells in the study area. This produces a data set containing over 610,000 cells. Each cell is assigned a land use class according to its predominant land use. Figure 1 shows the land use patterns in the region in 1958 and 2005, respectively, while Figure 2 presents a summary of trends among the land use classes from 1958 to 2005.

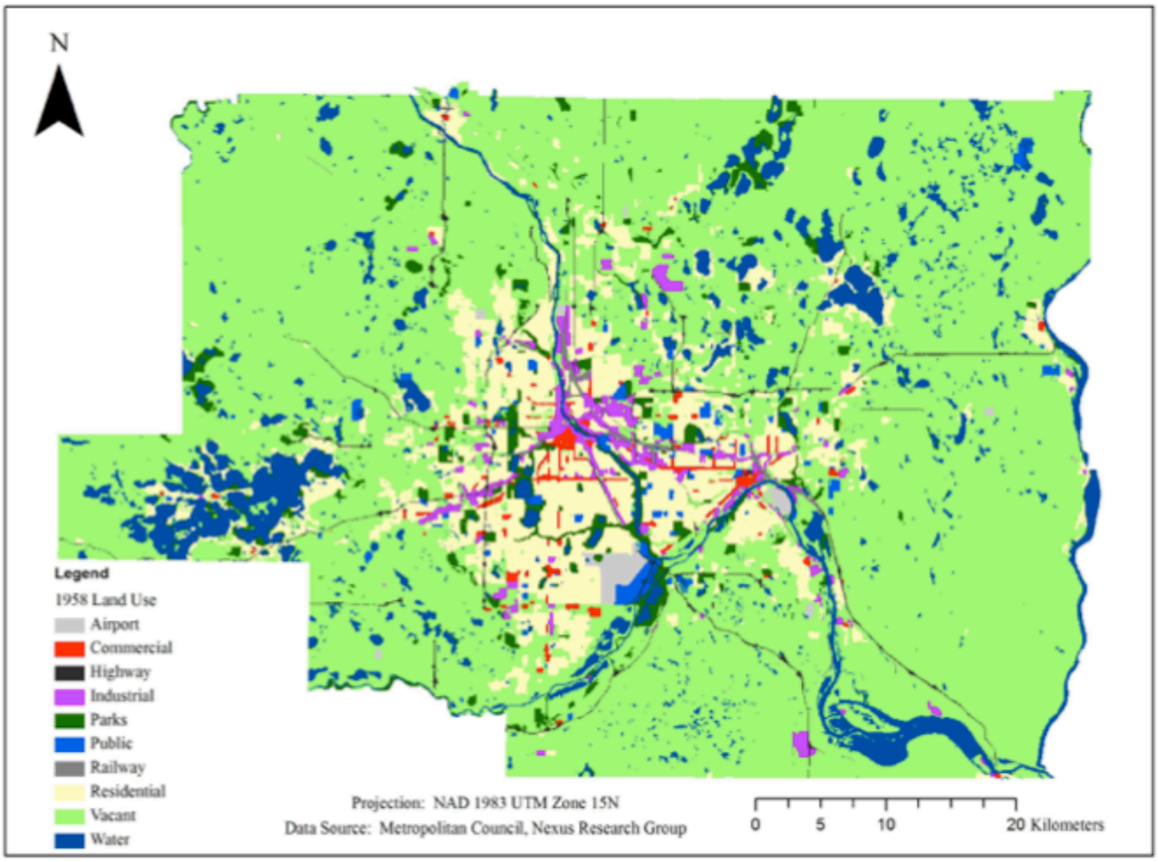

Fig. 1a Land use patterns in the Twin Cities region,1958

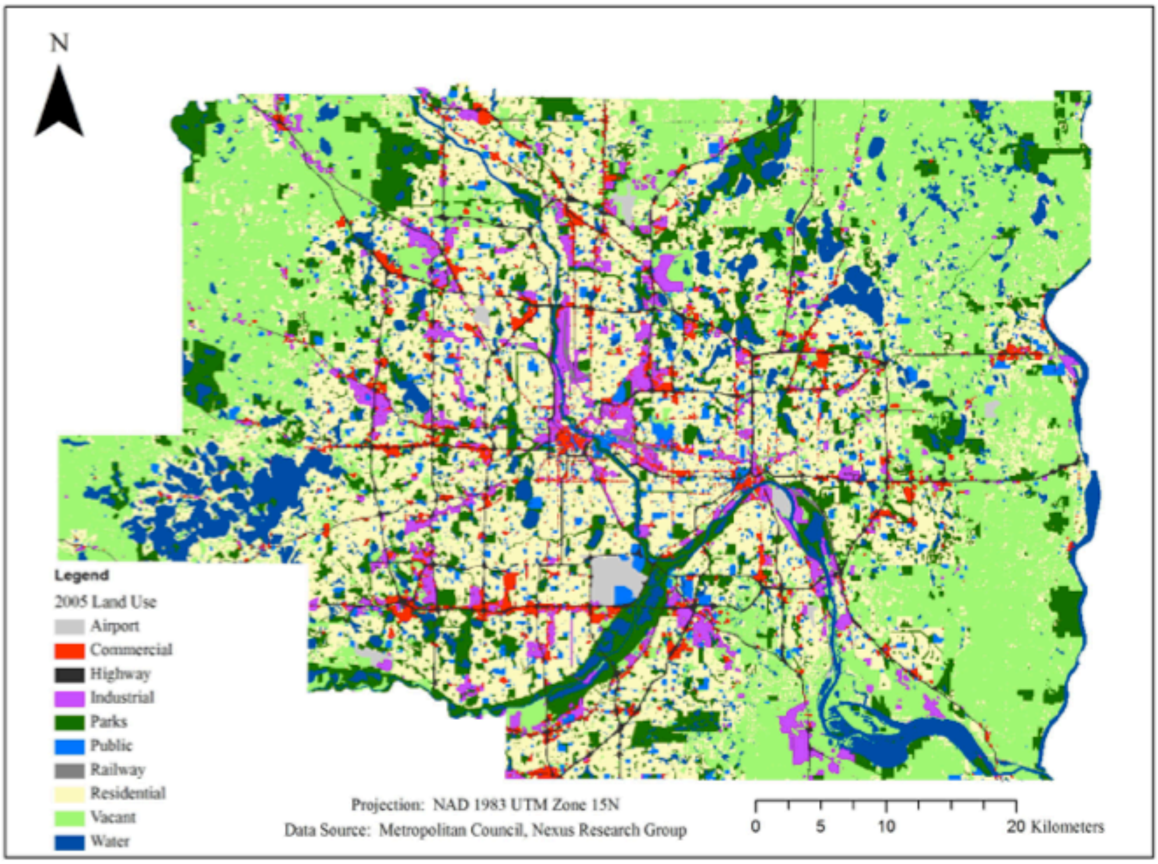

Fig. 1b Land use patterns in the Twin Cities region 2005

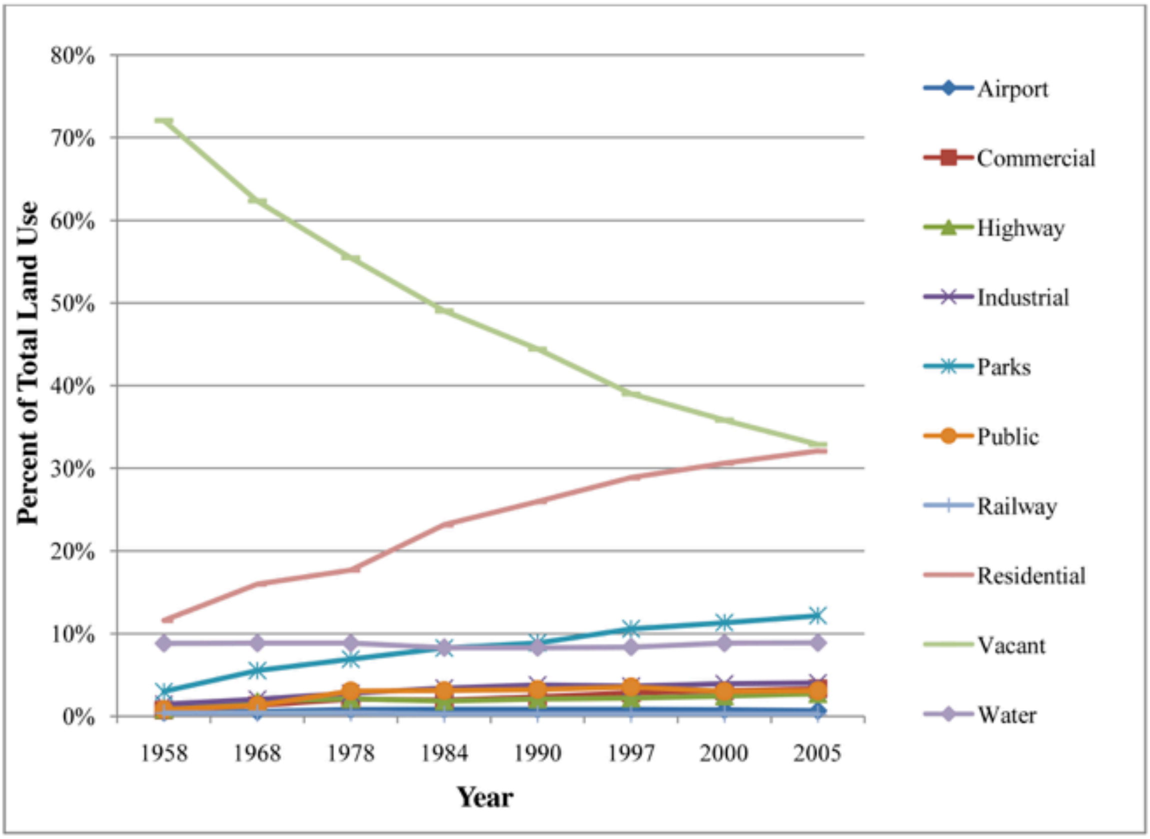

Fig. 2 Land use patterns in the Twin Cities region 2005

Virtually all land use classes have increased over this period, with the greatest increase in land use registered by the residential category. This growth has largely come at the expense of vacant land, as the region has been able to accommodate growth over the years via outward expansion. The data set contains three classes of land use related to transportation infrastructure: airports, railroads, and major highways. We can use the intermittent observations of land use to develop rough estimates of the amount of land that is consumed by transportation facilities and how it has changed over time as the region has developed. In the earliest year for which data are available (1958), transportation land uses covered 9,907 of the cells in the data set, the equivalent of about 1.6 percent of the total area in our sample or 55.7 km2 (21.5 mi2). Highways accounted for about one-half of all transportation-related land use. By 2005, these same three land uses covered a total of 22,187 cells, or 3.6 percent of the total area. Much of this growth came in the form of new highways, with highway land use increasing more than threefold. By comparison, the population of the 7-county core of the region, from which the land use data were drawn, increased by about 87 percent, from 1.5 million to 2.8 million. We interpret this estimate of transportation-related land use as a lower-bound estimate and, most likely, an underestimate. The land use data have no category for local roads which tend to be a denser network than regional highways, and treat parking as part of the respective land uses they serve. Other recent published estimates of parking coverage suggest that in urban settings, parking may account for 4 to 6 percent of total land use, while suburban settings tend to have lower amounts of coverage (mostly below 2 percent) (Davis et al., 2010a, Davis et al., 2010b). Were these two components to be added in to the total of transportation-related land use, the total coverage for the region would probably be somewhere in the range of 5 to 10 percent of all urban land use.

	AIRPOR	**COMM**	**HWY**	**INDUST**	**PARKS**	**PUBLIC**	**RAILW**	**RES**	**VAC**	**WATER**	**Totals**
AIRPOR	2874	12	2	148	9	90		59	160		3354
COMM	9	3709	59	588	238	508	1	2257	635	2	8006
HWY		15	10989	14	3	13		27	57		11118
INDUST	4	825	138	7641	343	496	21	1114	1886	4	12472
PARKS	383	364	137	378	18440	1191	4	3096	9622	76	33691
PUBLIC	62	226	5	95	354	4984	1	1688	644	3	8062
RAILW		3		9	3	4	2162	10	18		2209
RES	138	4597	261	1865	5579	5646	14	72155	7089	74	97418
VAC	1375	2976	1304	6198	17038	5676	35	27410	318548	174	380734
WATER		1	2	10	103	3		65	104	53636	53924
Totals	4845	12728	12897	16946	42110	18611	2238	107881	338763	53969	610988

Tab.1 Observed cell frequencies, 1968-1978

The K^2 statistic can be compared to a χ^2 distribution with $(10-2)^2 = 81$ degrees of freedom. With a critical region of $\alpha = 0.05$, values of the test statistic less than approximately 100 would indicate that land uses in 1978 were independent of those in 1968.

	AIRPOR	**COMM**	**HWY**	**INDUST**	**PARKS**	**PUBLIC**	**RAILW**	**RES**	**VAC**	**WATER**	**Totals**
AIRPOR	27	70	71	93	231	102	12	592	1860	296	3354
COMM	63	167	169	222	552	244	29	1414	4439	707	8006
HWY	88	232	235	308	766	339	41	1963	6164	982	11118
INDUST	99	260	263	346	860	380	46	2202	6915	1102	12472

PARKS	267	702	711	934	2322	1026	123	5949	18680	2976	33691
PUBLIC	64	168	170	224	556	246	30	1423	4470	712	8062
RAILW	18	46	47	61	152	67	8	390	1225	195	2209
RES	773	2029	2056	2702	6714	2967	357	17201	54014	8605	97418
VAC	3019	7931	8037	10560	26241	11597	1395	67225	211098	33631	380734
WATER	428	1123	1138	1496	3717	1643	198	9521	29898	4763	53924
Totals	4845	12728	12897	16946	42110	18611	2238	107881	338763	53969	610988

Tab.2 Expected cell frequencies, 1968-1978

With a computed K^2 of roughly $2.75\text{x } 10^6$, this is clearly not the case. Again, it should be noted that in the case of Markov chain models of land use, the hypothesis of independence is nearly always rejected. Historical dependence in land use is a strong force, as is indicated by the primacy of the diagonal elements of the observed transition matrix. Another way to examine the validity of the Markov chain framework is to test the stability or stationarity of the transition matrix. As described in an earlier section, one way to do so is to observe the correlation between the elements of matrices describing the transition probabilities. By regressing the matrix elements of a subsequent time period on a base period, it is possible to determine whether (and how far) the correlations between the two matrices deviate. The matrix of transition probabilities for the period from 1958 to 1968 will serve as a base period, since this is the earliest transition period for which data is available. Table 3 shows the results of three successive transition probability matrices being regressed on the original 1958 to 1968 matrix. The X and Y variables denote the response and predictor variables in the regression. The fit of the equation is summarized with the adjusted R^2 value.

				95% C.I.	
Y	X	Adj. R^2	β	Lower	Upper
1968-78	1958-68	0.977	0.98	0.95	1.01
1978-90	1958-68	0.943	0.948	0.902	0.995
1990-2000	1958-68	0.962	1.029	0.988	1.07
1968-78	1958-68	0.977	0.98	0.95	1.01
1978-90	1958-68	0.943	0.948	0.902	0.995

Tab.3 Summary of transition probability regressions

The value of the slope coefficient (β) is indicated, along with the lower and upper bounds of a 95% confidence interval for the mean value. In two of the three cases the 95% confidence interval includes the value of one, and in the third case the upper bound falls just short of one. While these results do not provide entirely conclusive evidence on whether the transition matrix is stationary, they offer some confidence that dramatic changes in transition probabilities are not occurring over time. Moreover, even a lack of stationarity does not need to preclude the use of Markov models. As Baker (Baker, 1989) has noted, stationarity can be assumed as a heuristic device for scenario generation using Markov chains. It is possible to evaluate how well the Markov chain model predicts land use change by using the historical time series to produce "backcasts" of land use for previous points in time. For example, the 1958 to 1968 transition probability matrix can be used as a base to predict forward in roughly 10-year increments to the years 1978, 1990 and 2000. Due to the different sources of data and data-generating processes noted for the years before and after 1984, we can provide "control" forecasts for the newer data using the 1984 to 1990 transition probability matrix as a base year matrix. These forecasts are provided for the years 1997 and 2005. Again, the land use conversion process in the model is governed by a random number generation procedure that draws values that correspond to the transition probabilities in the matrix for each initial land use state. Forecasts covering more than 10 years use the predicted land use distribution from 10 years prior as inputs to the forecast (e.g. forecast land use for 1990 is used as an input, along with the 1958-1968 probability matrix, for a forecast to the year 2000). This links the forecasts forward through successive time steps and preserves the Markovian principle that future states are only influenced by the present state. Summaries of the accuracy of the forecasts are provided in Table 4.

Base year matrix	Forecast Year	% Correct
1958-1968	1978	70
1958-1968	1990	55.2
1958-1968	2000	47.8
1984-1990	1997	84.4
1984-1990	2005	78.5

Tab.4 Forecast accuracy using historical time series data

As the results indicate, the accuracy of forecasts made using the 1958 to 1968 matrix of transition probabilities declines sharply over time. While all long-term forecasts can be expected to decline in accuracy the further they are asked to predict, there is a notable decline between the forecast years 1978 and 1990. This period coincides with the use of different sources of land use data which may not be entirely consistent and which may introduce additional inaccuracy to the forecast. The monotonic decline in accuracy also indicates that errors in forecasts from previous periods are fed forward into subsequent predictions. On the other hand, the forecasts made using a more recent transition matrix (1984 to 1990) as an input show a higher degree of accuracy and a more moderate decline over the second time step. This may be a result of more consistent data as well as a shorter transition period (6 to 8 years). Lastly, we are interested in using the Markov chain model to predict land use patterns several periods into the future. The most recent land use data are available for the years 1997, 2000 and 2005, indicating that the 1997 to 2005 period most closely matches the 10-year transition periods used throughout this study. Thus, a 1997 to 2005 transition probability matrix can be constructed and used for forecasting in 8-year increments. This matrix is reproduced below.

	AIRPOR	COMM	HWY	INDUST	PARKS	PUBLIC	RAILW	RES	VAC	WATER	Totals
AIRPOR	0.7388	0.001	0.0068	0.001	0.0325	0.0131	0	0.0055	0.1984	0.0029	1
COMM	0.0001	0.8187	0.0201	0.056	0.0045	0.0227	0.0002	0.0413	0.035	0.0015	1
HWY	0.0004	0.0107	0.9542	0.0054	0.0058	0.0031	0.0002	0.0094	0.0105	0.0001	1
INDUST	0.0004	0.071	0.0099	0.8371	0.0082	0.0086	0.001	0.0106	0.0517	0.0014	1
PARKS	0.0022	0.0036	0.0031	0.0025	0.9128	0.0062	0.0001	0.0116	0.0364	0.0214	1
PUBLIC	0.0001	0.0193	0.01	0.0384	0.0569	0.7364	0.0004	0.0223	0.1091	0.0071	1
RAILW	0	0.0065	0.0142	0.0201	0.011	0.0032	0.9139	0.0168	0.013	0.0013	1
RES	0	0.0024	0.0024	0.0009	0.0041	0.0023	0.0001	0.9634	0.023	0.0013	1
VAC	0.0004	0.0141	0.0099	0.0156	0.0513	0.0057	0.0002	0.0988	0.792	0.012	1
WATER	0.0001	0.001	0.0003	0.0014	0.0136	0.0002	0	0.0055	0.0096	0.9684	1

Tab.5 Transition probability matrix for 1997 to 2005

The 1997 to 2005 matrix is used to forecast forward through three time steps, yielding land use forecasts for the years 2013, 2021 and 2029. These forecasts are shown below in Table 6, along with the land use distribution in 2005, the base year.

	2005	2013	2021	2029	Change (2005-29)	Change (%)
AIRPOR	4047	3273	2674	2266	-1781	-44.0%
COMM	20296	22565	24114	25152	4856	23.9%
HWY	16635	19546	22041	24272	7637	45.9%
INDUST	24503	25961	27040	27620	3117	12.7%
PARKS	74251	81395	86454	89758	15507	20.9%
PUBLIC	18820	16793	15257	14013	-4807	-25.5%
RAILW	1505	1476	1454	1427	-78	-5.2%
RES	195934	211257	223401	233143	37209	19.0%
VAC	200837	171864	149304	131764	-69073	-34.4%
WATER	54160	56912	59249	61573	7413	13.7%

Tab.6 Land use forecasts for 2005 through 2029

Table 6 shows the land use distribution in each forecast year, along with the absolute and percentage changes through each time step. The land use forecasts for each period appear to be sensitive to abrupt, discontinuous changes that occur during the 1997 to 2005 period and are reflected in the transition matrix. The most notable effect is the prediction of a major decline in airport land. While there appears to have been a small decline from 1997 to 2005, this trend is projected out in each of the forecast periods, leading to a predicted decline of 44 percent from 2005 to 2029. This is probably not likely in a growing metropolitan area that anticipates continued growth in air travel in the coming decades. The same can be said of the trend in land used for highways, which is projected by the model to grow by roughly 46 percent. It would be useful to attempt to decompose this predicted growth by class of highway. Interstate and state trunk highway networks are already in place and are not likely to experience sharp increases in the near future, yet county highway networks, which tend to be more robust, may see substantial growth in newly-developing parts of the region. The model also predicts a major increase in residential land use, mostly at the expense of vacant land. This largely reflects the effects of the real estate boom of the late 1990s and early 2000s in the Twin Cities. Due to this reliance on past trends, the model will probably overpredict the demand for residential land use in the 2005 to 2013 period. Once new data become available, this observation can be tested.

4 CONCLUSION

This paper has demonstrated the application of a Markov chain model for fore-casting land use change in the Minneapolis-St. Paul metropolitan region. The Markov chain model has been shown to adequately describe the process of land use change, at least for short to medium-term time horizons. The extremely fine resolution of the land use data produced for this analysis allows for more detailed descriptions of land use transitions over time. The greater availability of data in recent years also allows for models that incorporate shorter transition periods, potentially leading to more accurate forecasts. Still, there are some aspects of the Markov chain model that deserve critical attention, and some directions of extension that could improve the model's output. These will be discussed in turn. One the most desirable qualities of the Markov chain model is its simplicity. It is able to describe the complex and long-term process of land use conversion in terms of simple transition probabilities, making it a potentially useful sketch planning tool.

However, this simplicity is also one of its greatest weaknesses. Since Markov chains are essentially projection models, they are not policy-sensitive and cannot easily incorporate the range of policy variables that might be of interest in predicting the impacts of various land use policies. The characterization of Markov chains as projection models also means that there is very little theory to guide their development. Except in cases where they are coupled with other types of models (e.g. McMillen and McDonald's zoning model), they may not encompass some of the important economic and regulatory forces shaping land use patterns in urban areas. These forces are often masked by the application of the transition probabilities. However, it is possible to introduce some of these factors directly into the model. Some applications have specified the transition probabilities themselves as functions of other variables (Brown et al., 2000; McMillen et al., 1991), thus making it possible to empirically estimate their determinants. One can imagine this being a possible path for introducing the influence of transportation networks on land use change within the MC framework. The use of transition matrices from a single period can also lead to forecasts that project short-term and perhaps discontinuous trends. An example of this was the projection of a major decline in airport land in the Twin Cities through 2029, despite countervailing trends in the underlying forces that drive the demand for airline services. A related matter is the application of transition matrices to residential land. Since housing markets are cyclical and are prone to boom-and-bust cycles, predictions based on a period of strong growth (or decline) may tend to overshoot (or undershoot) actual land use trends. Lastly, the Markov chain model, as applied in this study, does not account for neighbor effects. That is, land use in a particular location may be influenced not only by its previous land use, but also by the land uses of its neighbors. This principle has been incorporated into other types of cellular models of land use, such as cellular automata, which model land use as a function of the states of cells in a defined neighborhood. Modifying Markov chain models to incorporate this influence represents a potentially important improvement in model design. Indeed, there have been a handful of recent experimental efforts to design models with characteristics of both of these types of frameworks (de Almeida et al., 2003; Liu and Andersson, 2004; Pontius and Malanson, 2005). The basic Markov chain framework can also be extended in several directions to introduce greater detail and accuracy to processes of land use change. In addition to introducing neighbor effects, land use cells can be merged with data on the presence of transportation network links (Levinson and Chen, 2005) to describe the interaction between transportation networks and the demand for location among competing land uses. A further division of land use into classes based on intensity of use would also improve the model's detail. Residential uses in particular could be classified according to density or building height, along the lines of current zoning classifications. A similar classification scheme could be applied to commercial uses. Finally, more robust measures are needed to account for these additional influences in determining modeled outcomes. The evaluation measures employed in this study were fairly simple, and more elaborate frameworks are needed to model and forecast the interaction of land use with other dynamic processes at work within urban areas.

REFERENCES

Anderson, T. and L. Goodman (1957). Statistical inference about Markov chains. *Annals of Mathematical Statistics*, 28(1), 89–110. Available at http://www.jstor.org/stable/2237025.

Baker, W. (1989). A review of models of landscape change. *Landscape Ecology,* 2(2), 111–133. doi: http://dx.doi.org/10.1007/BF00137155.

Bell, E. (1974). Markov analysis of land use change: an application of stochastic processes to remotely sensed data. *Socio-Economic Planning Sciences,* 8(6), 311–316. doi: http://dx.doi.org/10.1016/0038-0121(74)90034-2.

Bell, E. and R. Hinojosa (1977). Markov analysis of land use change: continuous time and stationary processes. *Socio-Economic Planning Sciences,* 11(1), 13–17. doi: http://dx.doi.org/10.1016/0038-0121(77)90041-6.

Bourne, L. (1971). Physical adjustment processes and land use succession: a conceptual review and central city example. *Economic Geography* , 47(1), 1–15. doi: http://dx.doi.org/10.2307/143220.

Brown, D. G., Pijanowski, B. C., & Duh, J. D. (2000). Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *Journal of Environmental Management,* 59(4), 247–263. doi: http://dx.doi.org/10.1006/jema.2000.0369.

Clark, W. (1965). Markov chain analysis in geography: an application to the movement of rental housing areas. *Annals of the Association of American Geographers,* 55(2), 351–359. doi: http://dx.doi.org/10.1111/j.1467-8306.1965.tb00523.x.

Davis, A. Y., B. C. Pijanowski, K. Robinson, and B. Engel (2010). The environmental and economic costs of sprawling parking lots in the United States. *Land Use Policy,* 27(2), 255–261.
doi: http://dx.doi.org/10.1016/j.landusepol.2009.03.002.

Davis, A. Y., Pijanowski, B. C., Robinson, K. D., & Kidwell, P. B. (2010). Estimating parking lot footprints in the Upper Great Lakes Region of the USA. *Landscape and Urban Planning,* 96(2), 68–77.
doi: http://dx.doi.org/10.1016/j.landurbplan.2010.02.004.

de Almeida, C. M., Batty, M., Monteiro, A. M. V., Camara, G., Soares-Filho, B.S., Cerqueira, G.C., and Pennachin, C. L. (2003). Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. *Computers, Environment and Urban Systems,* 27(5), 481–509. doi: http://dx.doi.org/10.1016/S0198-9715(02)00042-X.

Lever, W. (1972). The intra-urban movement of manufacturing: A Markov approach. *Transactions of the Institute of British Geographers,* 56, 21–38. doi: http://dx.doi.org/10.2307/621540.

Levinson, D. and Chen, W. (2005). Paving new ground: a Markov chain model of the change in transportation networks and land use. In D. Levinson and K. Krizek (Eds.), A*ccess to Destinations* (pp. 243–266). Elsevier Science. ISBN: 9780080446783.

Liu, X. and Andersson, C. (2004). Assessing the impact of temporal dynamics on land-use change modeling. *Computers, Environment and Urban Systems,* 28(1), 107–124. doi: http://dx.doi.org/10.1016/S0198-9715(02)00045-5.

McMillen, D. and McDonald, J. (1991). A Markov chain model of zoning change. *Journal of Urban Economics*, 30(2), 257–270. doi: http://dx.doi.org/10.1016/0094-1190(91)90040-E.

Muller, R. and Middleton, J. (1994). A Markov model of land-use change dynamics in the Niagara region, Ontario, Canada. *Landscape Ecology,* 9(2), 151–157. doi: http://dx.doi.org/10.1007/BF00124382.

Pontius, G. R. and Malanson J. (2005). Comparison of the structure and accuracy of two land change models. *International Journal of Geographical Information Science,* 19(2), 243–265.
doi: http://dx.doi.org/10.1080/13658810410001713434.

Stokey, E. and Zeckhauser, R. (1978). A primer for policy analysis. W.W. Norton. ISBN-13: 978-0393090987.

Turner, M. (1987). Spatial simulation of landscape changes in Georgia: a comparison of three transition models. *Landscape Ecology,* 1(1), 29–36. doi: http://dx.doi.org/10.1007/BF02275263.

Weng, Q. (2002). Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. *Journal of Environmental Management* 64(3), 273–284.
doi: http://dx.doi.org/10.1006/jema.2001.0509.

AUTHOR'S PROFILES

Michael Iacono

Research Fellow in the Department of Civil, Environmental, and Geo-Engineering at the University of Minnesota (USA).

David David

Professor of Civil, Environmental, and Geo-Engineering at the University of Minnesota (USA).

Ahmed El-Geneidy

Associated Professor in the School of Urban Planning at McGill University.

Rania Wasfi

PhD Candidate in the Department of Geography at McGill University.