This paper presents a novel method for modeling the spatio-temporal movements of tourists at the macro level using Markov Chains methodology. Markov Chains are used extensively in modeling random phenomena which results in a sequence of events linked together under the assumption of first-order dependence. In this paper, we utilize Markov Chains to analyze the outcome and trend of events associated with spatio-temporal movement patterns. A case study was conducted on Phillip Island, which is situated in the state of Victoria, Australia, to test whether a stationary discrete absorbing Markov Chain could be effectively used to model the spatio-temporal movements of tourists. The results obtained showed that this methodology can indeed be effectively used to provide information on tourist movement patterns. One
significant outcome of this research is that it will assist park managers in developing better packages for tourists, and will also assist in tracking tourists’ movements using simulation based on the model used.

Introduction

The spatio-temporal movement of tourists is a complex process. Nevertheless, it is vital for a park manager to have some knowledge of these movements as well as locations of popular sites in the park. Agent modeling software such as RBSim 3 (Itami, 2003) can be used to model these movements under different management scenarios. Modeling can also be undertaken based on the desired level of spatial and temporal accuracy and purpose of trips. For example, spatio-temporal movements can be modeled as continuous processes with high resolution or as discrete processes with low resolution (Hornsby, 2001). Resolution can be measured in both spatial term (recording movements to the nearest centimeter, meter, or kilometer) or in temporal term (recording movements to the nearest second, minute, hour, day, or year). If the objective of the exercise is to understand detailed movements, then modeling will need to be at the micro-level (Batty, 1997; Jacquez et al., 2005; Timpf, 1998; Wang and Manning, 1999). Otherwise, it should be modeled at the macro-level (Forer and Simmons, 1998; Xia and Arrowsmith, 2005).

A variety of techniques are available to track tourist movements. The traditional tracking techniques are based on observations and interviews which require the researcher to follow an individual tourist and record his or her movements (Dumont et al., 2004). Participants can also be asked to trace or retrace their spatial
movements on a cartographic map using self-administered questionnaires (Fennell, 1996). Recently, many new tracking techniques such as Global Positioning System (GPS) (Arrowsmith et al., 2005), Timing Systems (O’Connor et al., 2005), Camera-based Systems (Haritaoglu et al., 1998), Personal Digital Assistant (PDA) Tracking (Hadley et al., 2003; Loiterton and Bishop, 2005), Mobile Phone Tracking (VeriLocation, 2004) have been utilized to record movement information of tourists resulting in a fairly high degree of resolution.

Several statistical techniques, such as Log-Linear Models, Logistic Regression and Markov Chains, have been applied to analyze movement data of large population. For example, the effects of habitat fragmentation on the movements of cotton rats could be identified by the Log-Linear methodology (Diffendorfer et al., 1995) and simple Markov Chains techniques have been used to elucidate the migration pattern of people (Constant and Zimmerman, 2003; Tobler, 1997). Data mining techniques, such as Expectation Maximization (EM) clustering, have also been used to identify the patterns of tourist movements for different tourist typologies (Wang and Manning, 1999; Xia and Arrowsmith, 2005).

In this paper, we present a method for modeling the spatio-temporal movement of tourists at the macro level using Markov Chains analysis. The theory of Markov Chains is a powerful technique which has been used widely in various disciplines for analyzing series of events which are linked together by first-order dependence. We applied this technique to estimate the probabilities of tourists visiting a sequence of attractions in a nature park on Phillip Island, which is situated in the state of Victoria, Australia. These probabilities are estimated using data which was collected
over a period of time from tourists visiting the park. One significant outcome arising out of this work was that park managers can use the results to develop suitable tour packages and arrange activities at appropriate times for tourists at each attraction. Further, the data can also be used to produce agent-based simulation models such as RBSim3 (Itami, 2003).

This paper is organized as follows. In the next section, we give a brief review of the theory of Markov Chains without going into any great detail but sufficient for what is required in this paper. Subsequent sections will describe the spatio-temporal movement of tourists on Phillip Island, estimate the probabilities of various movement patterns using absorbing Markov Chains and then validate the model using a standard test. Finally, some conclusions of our studies are drawn in the last section.

**Markov Chains**

A *stochastic process* is defined as a family of random variables \( \{X_t, t \in T\} \) defined on a given probability space and indexed by \( t \) belonging to a parameter set \( T \). The set \( T \) can be regarded as the time sequence of the process and is either discrete or continuous. In the discrete case \( T = \{1, 2, \ldots\} \) and in the continuous case \( T = [0, \infty) \). The range of \( X_t \) produces a *state space* \( S \) which is again either discrete or continuous. When \( S \) is finite or countable, the process is referred to as a *discrete stochastic process*.

A discrete stochastic process is referred to as a *Markov Chain* (MC) if the future evolution of \( X_t \) is dependent only on its present state. This is defined formally as
A stochastic process $X_t$, $t \in T$, taking values in a discrete set $S$, which we may take as the set of positive integers for convenience, is a Markov process if, for any sequence $t_0 < t_1 < t_2 < \ldots < t_n < t$, positive integer $n$ and values $i_0, i_1, \ldots, i_n$ and $j \in S$, the following identity involving conditional probabilities holds:

$$
\Pr(X_t = j | X_{t_n} = i_n, X_{t_{n-1}} = i_{n-1}, \ldots, X_{t_1} = i_1, X_{t_0} = i_0) = \Pr(X_t = j | X_{t_n} = i_n).
$$

(1)

Markov and related processes are comprehensively covered in the literature and have been applied extensively in diverse disciplines (Iosifescu, 1980; Kemeny and Snell, 1976; Stewart, 1994).

When $T = \{1, 2, \ldots\}$, the set of conditional probabilities

$$
\Pr(X_{n+1} = j | X_n = i) = p_{ij}(n)
$$

for $i, j \in S$ and $n = 1, 2, \ldots$ is called the set of one-step transition probabilities of the MC. One-step transition probabilities satisfy

(i) $p_{ij}(n) \geq 0$ for all pairs $i, j \in S$.

(ii) $\sum_{j \in S} p_{ij}(n) = 1$.

If the distribution of the initial state, i.e. $X_1$, is known, say

$$
\Pr(X_1 = i) = v_i
$$

$i \in S$, then, applying the product rule of probability and identity (1), the joint probability of the event

$$
\{X_1 = i_1, X_2 = i_2, X_3 = i_3, \ldots, X_n = i_n\}
$$
is given by

\[ \Pr(X_1 = i_1, X_2 = i_2, X_3 = i_3, \ldots, X_n = i_n) = v_{i_1} p_{i_1i_2}(1) p_{i_2i_3}(2) p_{i_3i_4}(3) \ldots p_{i_{n-1}i_n}(n-1). \]  

(3)

A MC is stationary if its one-step transition probabilities \( p_{ij}(n) \) are independent of \( n \). In other words, the probability of moving from one state to another is invariant of the epoch at which the transition takes place. In this case, we write \( p_{ij}(n) = p_{ij} \) for all \( i, j \) and the joint probability (3) can be simplified to

\[ \Pr(X_1 = i_1, X_2 = i_2, X_3 = i_3, \ldots, X_n = i_n) = v_{i_1} p_{i_1i_2} p_{i_2i_3} p_{i_3i_4} \ldots p_{i_{n-1}i_n}. \]  

(4)

One-step transition probabilities of a stationary MC can be arranged in a matrix, denoted by \( P \), called the one-step transition probability matrix, where

\[
P = \begin{pmatrix}
p_{11} & p_{12} & p_{13} & p_{14} & \cdots \\
p_{21} & p_{22} & p_{23} & p_{24} & \cdots \\
p_{31} & p_{32} & p_{33} & p_{34} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]

When the state space \( S \) is finite with cardinality \( m \), then \( P \) is a finite square matrix with dimension \( m \).

The \( n \)-step transition probabilities \( \Pr(X_{n-1} = j | X_0 = i) \) of a stationary MC will be denoted by \( p_{ij}^{(n)} \), \( i, j \in S \). Since

\[
\Pr(X_{n-1} = j | X_0 = i) = \sum_{i_n-2, \ldots, i_2, i_1} \Pr(X_{n-1} = j, X_{n-2} = i_{n-2}, \ldots, X_2 = i_2, X_1 = i_1 | X_0 = i)
\]

it is easy to show, on applying (4), that \( p_{ij}^{(n)} \) is the \((i, j)\) element of the matrix

\[
P^n = P \times P \times \ldots \times P
\]  

(5)
which is obtained by multiplying the matrix $P$ by itself $n$ times.

**Modelling Spatio-Temporal Movements of Tourists**

Spatio-temporal movements of tourists can be modeled at both the micro and macro level. At the micro level, they are represented by a continuous stochastic process $\{X_t, t \in T\}$, where $T = [0, \infty)$, taking values in a state space $S$. If $S$ consists of various spatial points representing the locations of the persons undergoing movements, then it is a continuous state space. On the other hand, for a nature tourist destination such as a park, there will be a number of attractions such as various lookouts and geomorphologic features situated in the park. We may consider the state space $S$ as comprising of these attractions and other auxiliary locations, therefore, $S$ is discrete.

At the macro level, movements of tourists are represented by a discrete stochastic process $\{X_t, t \in T\}$, where $T = \{1, 2, \ldots\}$. In this paper, we focus on the movement process at the macro level and also assume that $S$ is discrete. We let $S = \{A_1, A_2, \ldots, A_k\}$ where $A_i, i = 2, 3, \ldots, k$ are the tourist locations and $A_1$ represents the state “OUT”, i.e. the region exterior to the general area where the tourist sites are located. We will model the stochastic process representing the movement of tourists within $S$ by a stationary discrete MC where each movement occurs after a unit time step. Thus the movement of a tourist over a sequence of times from $t = 1$ to $t = m$ will be represented by the sequence

$$M(m) = (A_{i_1}, A_{i_2}, A_{i_3}, \ldots, A_{i_m})$$

where the last state reached is always $A_1$, i.e. “OUT”. This is because the trip
would always end up by a tourist leaving the park. In addition, this state will be an \textit{absorbing state} since the process terminates once it reaches that state. Thus the Markov chain we are considering is a finite \textit{absorbing Markov chain}. Due to the absorbing state, the other attractions are \textit{transient states}, i.e. starting from any of these attractions, and assuming that the process is allowed to go on for some time, it would eventually reach the absorbing state at an exponentially fast rate (Kemeny and Snell, 1976, p. 43).

By the Markov property (refer to (3))

\[
Pr(M(m)) = Pr(A_{i_1}) Pr(A_{i_2}|A_{i_1}) Pr(A_{i_3}|A_{i_2}) \ldots Pr(A_1|A_{i_{m-1}})
\] (6)

where we note that the last state reached is "OUT". The one-step transition probability matrix of the process is given by \( P = (p_{ij}) \) where \( i \) and \( j \) represent states \( A_i, i = 1, 2, \ldots, k \). Note that \( p_{11} = 1 \) since \( A_1 \) is an absorbing state.

\textbf{Survey and analysis}

In order to test whether a stationary discrete MC could effectively be used to model the spatio-temporal movements of tourists, a survey was conducted on Phillip Island, Victoria, Australia, to gather vital information. The survey implemented collected data on items such as tourists’ profiles, arrival times to each attraction, duration of times they spent there and the sequence of routes they used to reach the attraction. The self-administered questionnaire technique was used to track tourists’ daily movements on the island at the macro-level. Eight hundred questionnaires were distributed from 6th to 8th of March 2004 and from 17th to 20th of January 2005 to tourists at various entry points to the park. In total, five hundred questionnaires
were returned with 464 entered into the database for analysis. The remaining 34 questionnaires were incomplete and therefore discarded.

**Study area**

Phillip Island, located at the mouth of Westernport Bay, is 140 kilometers south-east of Melbourne, the capital city of the State of Victoria, Australia. There are a large number of natural flora and fauna species on the island, and these include penguins, koalas, seals, shearwaters, mangroves, wetlands, sandy beaches and rugged rocky cliff faces. The tourist attractions on Phillip Island are Churchill Island, Cowes, the Koala Conservation Center, Nobbies/Seal Rock, Penguin Parade, Rhyll Inlet, Ventnor and Cape Woolamai (indicated on the map in Figure 1), where visitors can experience wildlife in its natural environment. We refer the reader to Lawson (2002) for a detailed description of the park and its major attractions. For further

![Figure 1: Map of Phillip Island](image)
analysis, the following attractions and auxiliary locations are coded as follows and these constitute the state space $S$ of our stationary MC:

$$
\begin{align*}
A & : \text{Information Center} \\
B & : \text{Cape Woolamai} \\
C & : \text{Churchill Island} \\
D & : \text{Koala Conservation Centre} \\
E & : \text{Rhyll Inlet} \\
F & : \text{Cowes} \\
G & : \text{Penguin Parade} \\
H & : \text{The Nobbies/Seal Rock} \\
J & : \text{Ventnor} \\
OUT & : \text{Outside the Park}
\end{align*}
$$

**Markov Chains analysis**

In our case study, we considered the movement of tourists on Phillip Island as moving around nine attractions listed above. A stationary finite absorbing MC will be used to model the movements of these tourists between each attraction from the moment they enter the park until the moment they completed their visits. The states of the chain are the nine attractions with an additional absorbing state which we labeled “OUT”, which signals the completion of their daily trips. The conditional probability of moving from one attraction to another is the transition probability of the Markov Chain, e.g. the transition probability of going from attraction $A_i$ to
attraction $A_j$, $i \neq j$, between time $n$ to $n + 1$ is given by

$$Pr(A_j(n + 1)|A_i(n)) = \frac{Pr(A_j(n + 1) \cap A_i(n))}{Pr(A_i(n))} \quad (7)$$

where the symbol $A_i(n)$ represents the event of a visit to state $A_i$ at $t = n$ and $Pr(A_j(n + 1) \cap A_i(n))$ the probability of visiting both attraction $A_i$ and $A_j$, with attraction $A_i$ being visited first then followed by $A_j$. Since $A_i(n)$ can be partitioned into mutually exclusive events, i.e.

$$A_i(n) = \bigcup_{j=1}^{k} (A_i(n) \cap A_j(n + 1)),$$

from (7) we further obtain

$$Pr(A_j(n + 1)|A_i(n)) = \frac{Pr(A_j(n + 1) \cap A_i(n))}{\sum_{j=1}^{k} Pr(A_i(n) \cap A_j(n + 1))}. \quad (8)$$

By the assumption of stationarity, the transition probabilities (7), and hence (8), will be independent of $n$.

A graph tree (see Figure 2) is derived which summarizes all sequences of movements by tourists and give frequency of visits at each attraction as well as frequency of movements from one attraction to the next (Arrowsmith et al., 2005). We note that the maximum number of attractions visited by each visitor on a daily trip was seven. Based on this graph tree, the probability of any sequence of movements from the time a tourist enters the park until the time she leaves it can be calculated by using equation (3) once we are able to estimate the transition probability matrix and the initial probability vector. These probabilities will be estimated by the Maximum Likelihood Estimation method using information contained in the graph tree.
Figure 2: Graph Tree
The initial probability of visiting an attraction $A_i$, $Pr(A_i(1))$, is estimated by counting the number of visits to the attraction as a first destination divided by the total number of visits to all first attractions. For example, in the case of attraction $D$,

$$Pr(D(1)) = \frac{95}{451} = 0.21.$$ 

The initial probabilities of all nine attractions are tabulated in Table 1 below.

Table 1: Initial Probabilities

<table>
<thead>
<tr>
<th></th>
<th>$Pr(A)$</th>
<th>$Pr(B)$</th>
<th>$Pr(C)$</th>
<th>$Pr(D)$</th>
<th>$Pr(E)$</th>
<th>$Pr(F)$</th>
<th>$Pr(G)$</th>
<th>$Pr(H)$</th>
<th>$Pr(J)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Probability</td>
<td>0.01</td>
<td>0.06</td>
<td>0.07</td>
<td>0.21</td>
<td>0.03</td>
<td>0.33</td>
<td>0.18</td>
<td>0.10</td>
<td>0.01</td>
</tr>
</tbody>
</table>

The estimation of (8) is also straightforward using information contained in the graph tree. For example, to estimate the transition probability of going from $D$ to $C$, we apply the following steps:

1. Count the number of movements that satisfy the profile $D(n) \cap C(n+1)$ for $n = 1, 2, \ldots m - 1$ where $m$ is the maximum number of possible visits. For example, $N(D(1) \cap C(2))$ is the number of movements where tourists began their trips at $D$ and then move to $C$.

2. Sum these frequencies, i.e. $\sum_{n=1}^{m-1} N(D(n) \cap C(n+1))$.

3. Repeat Step 1 and 2 for all states in $S$ other than $D$ and sum all these frequencies to obtain the total number of one-step movement patterns starting in state $D$. 

13
(4) Divide the number obtained in (2) by that of (3).

The estimated one-step transition probabilities are tabulated in Table 2.

Table 2: Transition Probability Matrix

<table>
<thead>
<tr>
<th></th>
<th>OUT</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>OUT</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.200</td>
<td>0.000</td>
<td>0.200</td>
<td>0.200</td>
<td>0.400</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>B</td>
<td>0.043</td>
<td>0.000</td>
<td>0.000</td>
<td>0.217</td>
<td>0.043</td>
<td>0.284</td>
<td>0.088</td>
<td>0.239</td>
<td>0.043</td>
<td>0.000</td>
</tr>
<tr>
<td>C</td>
<td>0.103</td>
<td>0.000</td>
<td>0.102</td>
<td>0.333</td>
<td>0.026</td>
<td>0.308</td>
<td>0.102</td>
<td>0.026</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D</td>
<td>0.052</td>
<td>0.000</td>
<td>0.022</td>
<td>0.030</td>
<td>0.000</td>
<td>0.300</td>
<td>0.481</td>
<td>0.193</td>
<td>0.185</td>
<td>0.007</td>
</tr>
<tr>
<td>E</td>
<td>0.000</td>
<td>0.000</td>
<td>0.077</td>
<td>0.115</td>
<td>0.000</td>
<td>0.577</td>
<td>0.193</td>
<td>0.038</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>F</td>
<td>0.014</td>
<td>0.000</td>
<td>0.017</td>
<td>0.031</td>
<td>0.010</td>
<td>0.000</td>
<td>0.568</td>
<td>0.339</td>
<td>0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>G</td>
<td>0.948</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>0.002</td>
<td>0.022</td>
<td>0.000</td>
<td>0.026</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>H</td>
<td>0.014</td>
<td>0.000</td>
<td>0.005</td>
<td>0.019</td>
<td>0.010</td>
<td>0.191</td>
<td>0.746</td>
<td>0.000</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>J</td>
<td>0.000</td>
<td>0.000</td>
<td>0.133</td>
<td>0.000</td>
<td>0.133</td>
<td>0.200</td>
<td>0.400</td>
<td>0.134</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The probabilities of movement patterns can be computed using equation (3) once the one-step transition probabilities and the distribution of the initial state have been estimated. For example, the probability of travel pattern $D \rightarrow F \rightarrow G$, i.e. the Koala Conservation Centre (D) is the first destination visited, followed by Cowes (F) with Penguin Parade (G) being the last destination visited before exiting.
the park is calculated using Table 2 and Table 1 above to equal

\[
\Pr(DFG\ OUT) = \Pr(D) \times \Pr(F|D) \times \Pr(G|F) \times \Pr(OUT|G)
\]

\[
= 0.21 \times 0.481 \times 0.568 \times 0.948 = 0.054.
\]

Each movement pattern will also depend on the number of attractions visited. Therefore, in order to identify the significant patterns from the data collected, we first categorize these patterns according to the number of attractions visited. The most significant movement patterns will then be identified by the values of the probabilities of these patterns conditioned on the number of attractions. To obtain these probabilities, we divide the probability of each pattern consisting of \(n\) visits, \(n = 1, 2, \ldots, 7\) by the sum of the probabilities of all patterns consisting of \(n\) attractions. For example, for \(n = 1\), the probabilities of the events \(G, D\) and \(C\) are estimated as 0.168, 0.01 and 0.008 respectively, giving a total probability of 0.186 for a one-attraction pattern. The conditional probability of visiting a single attraction \(G, D\) or \(C\) is therefore equal to 0.168/0.186 = 0.903, 0.01/0.186 = 0.054 or 0.008/0.186 = 0.043 respectively, thus making \(G\) the most popular destination for a one-attraction visit to the park. The same process was applied to \(n = 2, \ldots, 7\) visits and, for the sake of illustration, we tabulate the results for the case \(n = 4\) in Table 3 below and note that \(DFHG\) was by far the most popular \(n = 4\) sequence of attractions. Note that all possible movement patterns end in state \(G\), i.e. Penguin Parade. This is because the highlight of this attraction, the parade of penguins, takes place after sunset and it is therefore the final destination of most tourists to the island.
The value of the above approach in predicting tourists’ movements using data from a graph tree should be fairly apparent. Although a graph tree offers much useful information, one of which is the frequency of visits to each attraction, it does not assist organizers such as a park manager, say, for marketing purposes, to select the most attractive sequence of attractions. However, as we have just illustrated, a model utilizing Markov Chain techniques will do precisely that. In addition, the estimated Markov Chain model could be used to simulate movements process, by incorporating it within a multi-agent modeling system, such as RBSim3 (Itami, 2003), thereby assisting park managers to simulate sequence of visits using expert knowledge.

**Model validation**

In this section, we address the question of how well Markov Chain technique can
be used to model tourists’ movements on Phillip Island. Our objective is to assess how close the observed frequencies are to the expected frequencies when (4) is used to calculate the probabilities of different movement patterns. This objective will be achieved using the Chi-Squared Goodness of Fit Test (Devore, 2004).

We adopted a data mining approach by dividing the data set, consisting of 464 records, into two parts: *training data* and *test data*. The selection of training data was implemented by a random sampling method called EPSEM (Equal Probability Selection Method) (Kish, 1965). This data was used to fit the Markov Chain model, i.e. to estimate the transition probability matrix $P$ and initial probability vector. For a fixed number of visits $n$ and corresponding number of movement patterns $N(n)$, the expected frequencies of movement patterns $E_i$ obtained thereby were tested against corresponding observed frequencies $O_i, i = 1, 2, \ldots N(n)$, from the test data using a Chi-Squared Goodness of Fit Test. We selected 70% of the data (325 records) as training data, leaving the remainder as test data and then repeated the process by selecting 60% of the data (278 records) as training data. The objective here is to gauge the effect sampling bias has on the results obtained. Table 4 and Table 5 summarize the results obtained from these two cases, respectively. We note that the Chi-squared Test statistic $\chi^2$ is obtained using the equation

$$
\chi^2 = \sum_{i=1}^{N(n)} \frac{(O_i - E_i)^2}{E_i}.
$$

The p-values given in the tables indicate the degree of significance in the results. Customarily, a p-value of 0.05 or less offers evidence against the model, i.e. the Markov Chain model provides a poor fit to the data. As is evident from the tables,
Table 4: Chi-Squared Test Statistic when 70% of data is used for training

<table>
<thead>
<tr>
<th></th>
<th>n=1</th>
<th>n=2</th>
<th>n=3</th>
<th>n=4</th>
<th>n=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi^2 )</td>
<td>0.13</td>
<td>3.27</td>
<td>3.39</td>
<td>15.19</td>
<td>4.39</td>
</tr>
<tr>
<td>p-value</td>
<td>0.72</td>
<td>0.20</td>
<td>0.07</td>
<td>0.004</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 5: Chi-Squared Test Statistic when 60% of data is used for training

<table>
<thead>
<tr>
<th></th>
<th>n=1</th>
<th>n=2</th>
<th>n=3</th>
<th>n=4</th>
<th>n=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi^2 )</td>
<td>0.37</td>
<td>1.60</td>
<td>6.92</td>
<td>2.32</td>
<td>4.12</td>
</tr>
<tr>
<td>p-value</td>
<td>0.83</td>
<td>0.45</td>
<td>0.14</td>
<td>0.31</td>
<td>0.04</td>
</tr>
</tbody>
</table>

in the majority of cases, the Markov Chain model did provide a good fit to the data. However, p-values generally decline with increasing \( n \), leading one to suspect that either the stationary assumption breaks down or higher -order Markov Chains are required for longer trips.

Conclusion

This paper presents a novel method for modeling the spatio-temporal movements of tourists at the macro level using Markov Chain methodology. Markov Chains provide a powerful tool for analyzing trends and outcomes of a series of events that are linked together by first-order dependence. The technique was specifically used in this paper to calculate the probabilities of movement patterns by tourists.
in a natural park. A case study was undertaken at Phillip Island, from which an extensive survey was conducted to collect data for the purpose of testing whether a stationary, absorbing, discrete Markov Chain could be used to model the spatio-temporal movements of tourists visiting the park.

A fraction of the data was used to fit the model and the remainder to test the model. The model was then validated using a series of Chi-Squared Goodness of Fit Tests, the results of which suggest that the Markov Model is appropriate as long as the length of a visit is not too long. When it is, the fit is not as good and we conjecture that in this case, the stationary and first-order dependence assumptions underlying the model break down. However, we believe that the level of generality in this work is sufficient for many practical situations that could occur in a natural environment. In future work, we will relax some of these assumptions and compare the results of these new studies with the results obtained in this paper. We also plan to model the continuous movement process at the micro level using continuous time Markov process.

Finally, a significant outcome of this study is that it will assist park managers in identifying significant movement patterns of tourists in their parks. This will assist them in designing better tourist packages as well as providing more attractive combinations of attractions. In addition, the probabilities obtained from employing the approach of this paper could be used as input into agent-based simulation system such as RBSim3 (Itami, 2003) and then used to evaluate tourist behavior under different management scenarios in order to isolate potential problems and obtain optimum outcomes.
References


