Faculty of Science and Engineering

Department of Spatial Sciences

Development of Tour Packages through Spatio-Temporal Modelling of Tourist Movements

Mohd Faisal Bin Abdul Khanan

This thesis is presented for the degree of
Doctor of Philosophy
of
Curtin University

October 2014

Declaration

To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature…………………………………………………………..
Date…………………………………………………………..

ABSTRACT

An ideal tour package should consist of itinerary elements such as the sequence of tourist attractions with the scheduled arrival times and visiting durations at each attraction, specific to each tourist according to their preferences and characteristics. Although seen as a determining factor for tourists' satisfaction by tour operators, a tour package is described as one of the fundamental, taken-for-granted symbols of the tourism industry. Conventionally, tour operators choose just some of a number of travel products, for instance, several places to visit or a transportation mode, and put them into a tour package without sufficiently understanding both the potential and actual needs of the tourists themselves. Furthermore, some tour operators have conflicting motives in producing tour packages such as adding many shopping activities with the intention of maximising their profits.

This research involves outlining several methodologies for designing tour packages. The methodologies are for personalised scheduling, activity planning and market segmentation, that have been developed by integrating tourists' spatio-temporal movements and their characteristics. The design of a tour package is directly linked to tourist market needs and demands using the Personalised Scheduling Method. Semi-Markov and scheduling processes are employed to produce an itinerary, personalised to a specific combination of characteristics for tourists. Other elements of a tour package such as route, activities and budget can be added to the itinerary to complete the tour package design process. The Activity Planning Method uses a Probability Density Function to estimate tourists' arrival times and visiting durations for planning the timing of activities at each attraction. The Market Segmentation Method applies the Expectation-Maximisation Algorithm to produce target markets of tour package products which are itineraries, attractions and activities based on tourists' characteristics.

A case study of Phillip Island, Victoria, Australia was undertaken to implement and evaluate each method developed. A survey was conducted to gather data of tourists' spatio-temporal movements and their characteristics. The Personalised Scheduling Method was implemented using the spatio-temporal movement dataset of female and

middle aged tourists, as the Phillip Island Nature Park is one of the target tourist markets. A number of itineraries were derived and suggested for female and middle age tourists. The Activity Planning Method revealed that a few activities on Churchill Island, one of the attractions located on Phillip Island, were often scheduled prior to tourists' arrivals or long after they arrived. This highlighted the importance of activity planning and marketing effort. Several tourist target markets were identified for various itineraries, attractions and activities, which are useful in understanding tourist needs and demands and identifying the competitive advantage of tour packages.

In conclusion, the tour package design methodology developed in this study is able to provide tourists with personalised itineraries and support other stakeholders such as attraction administrators and tour operators in managing the resources and facilities within attractions together with transport planning between tourism areas.

ACKNOWLEDGEMENTS

The journey of my PhD is probably the most challenging and life-changing activity of my 30 years of life. The joyous and distressing moments of my doctoral journey have been shared with many people. It has been a great privilege to spend several years since my Masters, and later on, PhD in the Department of Spatial Sciences at Curtin University. Its members will forever remain dear to me.

First and foremost, I owe a debt of gratitude to my supervisor, Dr Jianhong (Cecilia) Xia who patiently and consistently provided the vision, encouragement, constructive challenges and invaluable advice throughout my PhD journey and towards completing my thesis. I also appreciate my co-supervisor, Professor Bert Veenendaal and associate supervisor, Professor Panlop Zeephongsekul for their time and contribution in strengthening my methodology and improving my thesis.

I would also like to thank my sponsors, Ministry of Education, Malaysia and Universiti Teknologi Malaysia for funding my study. Not to forget, the staff of the Phillip Island Nature Park for their advice and assistance on the questionnaire design, sampling and feedback on this PhD.

Special thanks are also extended to Lori Patterson, Meredith Mulcahy, Caroline Rockliff and Pam Kapica for their administrative and academic assistance. I would also like to acknowledge Saud Aboshiqah, Xin Liu, Saad Al-Kahtani, Qian Sun, Ashty Saleem and Charity Mundeva for providing help and encouragement in times of need.

Finally, I wish to thank my parents, Abdul Khanan Kasin and Asmah Ab Rahman for their love and prayers that sustained me throughout my doctoral journey as well as my wife, Ainull Najhwar Abdul Razak and my kids, Ahmad Harraz Mohd Faisal and Ahmad Zaki Mohd Faisal for their unconditional support and patiently enduring this journey together. Your smiles and tears will eternally remain in my memory.

RELATED PUBLICATIONS

Journal papers:

Abdul Khanan, M. F., J. C. Xia, P. Zeephongsekul, and B. Veenendaal. "The Development of a Personalised Scheduling Method Based on the Spatio-Temporal Movement of Tourists for Designing Tour Packages." Submitted to *Tourism Geographies*, 2014.

Abdul Khanan, M. F., J. C. Xia, and B. Veenendaal. "Tour Package Activities Planning from the Perspective of the Spatio-Temporal Movement of Tourists." In Preparation for *Tourism Management*, 2014.

Abdul Khanan, M. F., J. C. Xia, and B. Veenendaal. "Clustering Tourists' Characteristics: Market Segments of Tour Packages and its Related Products." In Preparation for *Annals of Tourism Research*, 2014.

Book Chapter:

Abdul Khanan, M. F., J. Xia, and R. Inbakaran. 2012. "Individual Differences in the Tourist Wayfinding Decision-Making Process: A Case Study of Phillip Island, Victoria, Australia." In *Advances in Geo-Spatial Information Science*, Eds W. Shi, M. F. Goodchild, B. Lees and Y. Leung, 245-256. Leiden: CRC Press.

Conference Proceedings:

Abdul Khanan, M. F., and J. C. Xia. 2010. "Individual Differences in the Tourist Wayfinding Decision Making Process" *Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science*, 26-28 May 2010, *Hong Kong*: International Society of Photogrammetry and Remote Sensing.

Abdul Khanan, M. F., J. C. Xia, and B. Veenendaal. 2012. "Establishing a Framework for a Tour Package Design through the Modelling of Spatio-Temporal Movement of Tourist" *The 15th International Symposium of Spatial Data handling*, 22-24 August 2012, *Bonn, Germany*: International Geographical Union.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

PDF   :   Probability Density Function

EM   :   Expectation-Maximisation

PSM   :   Personalised Scheduling Method

APM   :   Activity Planning Method

MSM   :   Market Segmentation Method

ICT   :   Information and Communication Technology

IGA   :   Interactive Genetic Algorithm

GA   :   Genetic Algorithm

Starchild   :   Simulation of Travel/Activity Responses to Complex Household Interactive Logistic Decisions

CST   :   Constraint Satisfaction Technique

CATC   :   Computer Assisted Travel Counselling

CRS   :   Computerised Reservation System

DIS   :   Destination Information System

CBR   :   Case-Based Reasoning

AuSTO   :   Australian Sustainable Tourism Ontology

RBSim   :   Recreation Behaviour Simulator

GIS   :   Geographic Information System

GPS   :   Global Positioning System

AIC   :   Akaike Information Criterion

ER   :   Entity-Relationship

VBA   :   Visual Basic for Application

Weka   :   Waikato Environment for Knowledge Analysis

MLE   :   Maximum Likelihood Estimation

EPSEM   :   Equal Probability of Selection Method Sampling

ATCA   :   Activity Timing Comparison Analysis

AFA   :   Activity Frequency Comparison Analysis

CHAPTER 1: INTRODUCTION AND OUTLINE

1.1   Introduction

Transportation and tourist movement are considered as significant factors for international tourism (Page, 2005). When and how tourists move within the destinations are valuable information for tourism marketing, especially for developing tour packages by tour operators or the arrangement of activities by observing tourist movement by attraction administrators.

A tour package is a primary component of the tourism industry, seen by tour operators as a determining factor for tourist satisfaction (Geva and Goldman, 1991). It was described as "one of the fundamental, taken-for-granted symbols of the tourism industry" (Weaver and Lawton, 2010).

Some studies define the components of a tour package as consisting of accommodation, transportation and attraction, but ignore the vital aspect of tourist movement (Morrison, 1996; Perlitz and Elliott, 2000; Syratt and Archer, 2003). At the same time, much research that covers the aspect of spatio-temporal movement of the tourist within a tour package design do not include an actual tourist movement dataset such as a list of attractions visited, the arrival time and visiting duration

within attractions (Niknafs et al., 2003; Dunstall et al., 2004; Tam and Pun-Cheng, 2012).

A tour package is defined in this research as having a list of attractions with their arrival times and visiting durations by tourists. The list of attractions is logically sequenced and involves the arrangement of visitation from one attraction to another. It shows which attraction is to be visited earlier and later or first and last (Xia and Arrowsmith, 2005).

In this thesis, a methodology to design tour packages through the modelling of tourist spatio-temporal movement is developed. A scheduling method that is necessary for the integration of a movement sequence from one attraction to another with its itinerary element is proposed.

The research progresses to include the suggestion of activity starting time based on tourist arrival time within each attraction. Multiple starting times are suggested according to various tourist arrivals to suit a range of activity timing arrangements. At the end of this thesis, a method to associate tour package products such as itinerary, attraction and activity, according to different tourist characteristics, for instance age, gender and travel group, is introduced to acknowledge the importance of the marketing perspective of tour packages.

A case study was conducted at Phillip Island, Victoria, Australia to gather tourist spatio-temporal movement data that covers their movement around nine attractions. A mathematical model and scheduling process are used to develop a method of modelling tourist spatio-temporal movement and are evaluated using this spatio-temporal movement dataset.

Another mathematical model is applied to the arrival times and visiting durations of tourists in each attraction. Furthermore, the output of the itinerary is used to segment the itinerary market by using a tourist characteristic dataset through the implementation of a clustering algorithm.

1.2   Problem Formulation

The major challenge in managing tourism is the diversity of tourists competing for the same resources such as a tour package and the need to balance these multiple objectives while maintaining a positive tourism experience (O'Connor et al., 2005). According to Fennell (1996), the key to overcome this issue is by understanding tourist behaviour. One such behaviour is tourist movement (Cooper, 1981).

According to Xia et al. (2008), tourist spatio-temporal movement in terms of their way finding behaviour consists of two important components, namely physical movement and decision-making. This opinion is further strengthened by the definition of tourism itself which means the practice of travelling for recreation (Merriam-Webster, 2010). Travelling needs movement. Thus, it is inadequate to produce a tour package without considering the spatio-temporal movement of tourists.

The knowledge of tourist spatio-temporal movement patterns is essential to develop tour packages by understanding the location of popular sites and the timing of visits where it indicates how tourists combine tourist sites together and arrange their schedules (Xia et al., 2010). Furthermore, according to her, a dominant tourist spatio-temporal movement pattern could be used to assist attraction administrators in deciding how the daily program of activities should be arranged for attractions.

A tour package is significantly important where it comprises elements such as marketing, movement and activities (Jafari, 2000; Perlitz and Elliott, 2000). Although important, the proper methodology in designing tour packages is always taken for granted by tour operators. Dunstall et al. (2004) claimed that tour operators conventionally choose just some of the elements of the holiday, such as one or more travel products, or one or more places to visit and put them into their tour package. Furthermore, some tour operators have produced tour packages with an intention to maximise their profit such as adding too many shopping activities (Ap and Wong, 2001).

To date, very few tour packages have been created using proper quantitative techniques. Niknafs et al. (2003) implemented a Case-Based Reasoning (CBR) where tourist preferences such as interest and budget were the main subject in designing tour packages. Unfortunately, their techniques did not include a spatio-temporal movement aspect to better support their findings.

An improvement to the research by Niknafs et al. (2003) was the itinerary mapping technique to gather spatio-temporal movement data of tourists (Connell and Page, 2008). Alas, though spatial data were collected and analysed, they relied too much on sketch maps of tourist vehicular movement, resulting in a loss of fine detail as a result of using small scale maps (Lew and McKercher, 2006).

A greedy algorithm was implemented to find the shortest path from one attraction to another in tour packages (Tam and Pun-Cheng, 2012). Despite suggesting tour packages with proper timing elements such as arrival time and visiting duration at each attraction, the implementation of the shortest path concept sometimes does not suit touring scenarios. During tours, tourists could prefer longer routes to reach destinations due to the leisure motivation of passing by more places present between attractions during a tour.

The spatio-temporal movement in terms of visiting duration is also important in influencing the decision-making of tourists in choosing attractions (Sirakaya et al., 1996). A similar view was expressed by McKercher and Lew (2004) where tourist spatio-temporal movement was found to influence tour package modelling.

Finally, although the foundation of linking tourist spatio-temporal movement with tour packages design was established together with the link between tourist decision-making towards tour packages, little was done to look into the inter-relationship between spatio-temporal movement and tourist decision-making in the context of tourist preferences (McKercher and Lew, 2004).

Tour operators could fail to match tour packages, as a product of analysing spatio-temporal movement with tourist preferences via their characteristics as some of the attractions may only suit certain types of tourists (Kale et al., 1987). This inter-relationship between spatio-temporal movement and tourist decision-making is important in designing tour packages. Therefore, it must be taken seriously by stakeholders of tour packages.

1.3   Research Objectives

The primary aim of this research is to establish a method for developing tour packages by modelling tourist spatio-temporal movement. In order to achieve the primary aim, three objectives with related research questions are addressed.

The first objective is to develop a framework for developing tour packages that incorporates aspects of spatio-temporal tourist movement and tourist characteristics. Its related research questions are:

(i)   How do we best define a tour package and its components from the perspective of spatio-temporal movement?

(ii)   What are the issues in designing tour packages, especially in relation to spatio-temporal movement and tourist characteristics?

Such a framework is important when later, it will be used to devise a modelling technique for those movements as the most important component in designing tour packages. This is the second objective and it constitutes several research questions such as:

(iii)   What are the current methods to solve the tour package design issues?

(iv)   What methods can be used to spatio-temporally develop itineraries?

(v)   What methods can be used to estimate the arrival time and visiting duration of tourists for the purpose of suggesting starting times and visiting durations of various activities?

(vi)   What methods can be used to segment the market of tour package products according to diverse tourist characteristics?

(vii) What are the tourist characteristics for segmenting the market of tour package products?

Lastly, the third objective is putting into practice the modelling technique by implementing and evaluating it using a case study. Its research questions are:

(viii) Will there be various itineraries to be proposed as a result of diverse spatio-temporal movement patterns according to different tourist characteristics?

(ix) Will there be a different set of arrival times and visiting durations according to each attraction?

(x) Will there be a significant difference between the arrival time and visiting duration outcomes with the status quo of the timing of specific attraction activities?

(xi) Will there be a specific set of tourist characteristics associated with each tour package product?

This thesis does not attempt to model the spatio-temporal movement of tourists within each attraction as the scope only involves the macro-level movement of movement sequence instead of micro-level movement within an attraction. Nor does this thesis try to uncover the reasons behind each movement decision made by tourists. No cognitive decision-making models were involved in this thesis. A slight representation of cognitive perspective is demonstrated by the segmentation of tour package products according to different types of tourist characteristics.

1.4   Research Significance

The proposed research is significant in that it develops a method for developing tour packages by modelling tourist spatio-temporal movement. This research fills a gap by embedding tourist movement information in a tour package design process. This is done by including the output of scheduling actual tourist spatio-temporal movement and corresponding tourist characteristics as major elements of itineraries. Therefore, itineraries can be tailored for certain types of tourists according to their preferences and characteristics (Lopez, 2003).

On top of that, attraction administrators are able to improve the organisation of activities by enhancing the technique to estimate the arrival time of tourists (Asakura and Iryo, 2007). Concurrently, tour operators can offer diverse tour package products according to broader types of tourists by segmenting itineraries, attractions and activities markets (Xia et al., 2010). Ideally, by offering more types of tour package products, sales improve and profit increases.

This research is distinguishable from existing tour package research by emphasising scheduling of tourist movement, estimating the frequency of activities and starting time based on the arrival time of tourists and the development of customised itineraries. Other contributions from this research include:

(i)   Stakeholders, like attraction administrators, local authorities and government agencies are able to plan the future development of parks and surroundings such as improving facilities to assist tourist movement. Such facilities are like information kiosks, signboards, signposts and special roads for tourists to avoid roads used by locals.

(ii)   The tourist characteristic findings of this research are useful to plan the development of amenities to assist tourists with particular needs such as senior citizens and families with children. Such facilities may include toilets, special bus services, cafes and shaded areas.

(iii)   Providing a set of robust spatio-temporally enriched tourist movement information that includes their characteristics that are useful for obtaining the knowledge of tourist behaviour.

1.5   Thesis Outline

This section describes the thesis structure where it consists of eight chapters as depicted in Figure 1.1. The concepts, theories and methods of tour packages and its design process are briefly presented in these chapters. Each research question is also carefully addressed within these chapters.
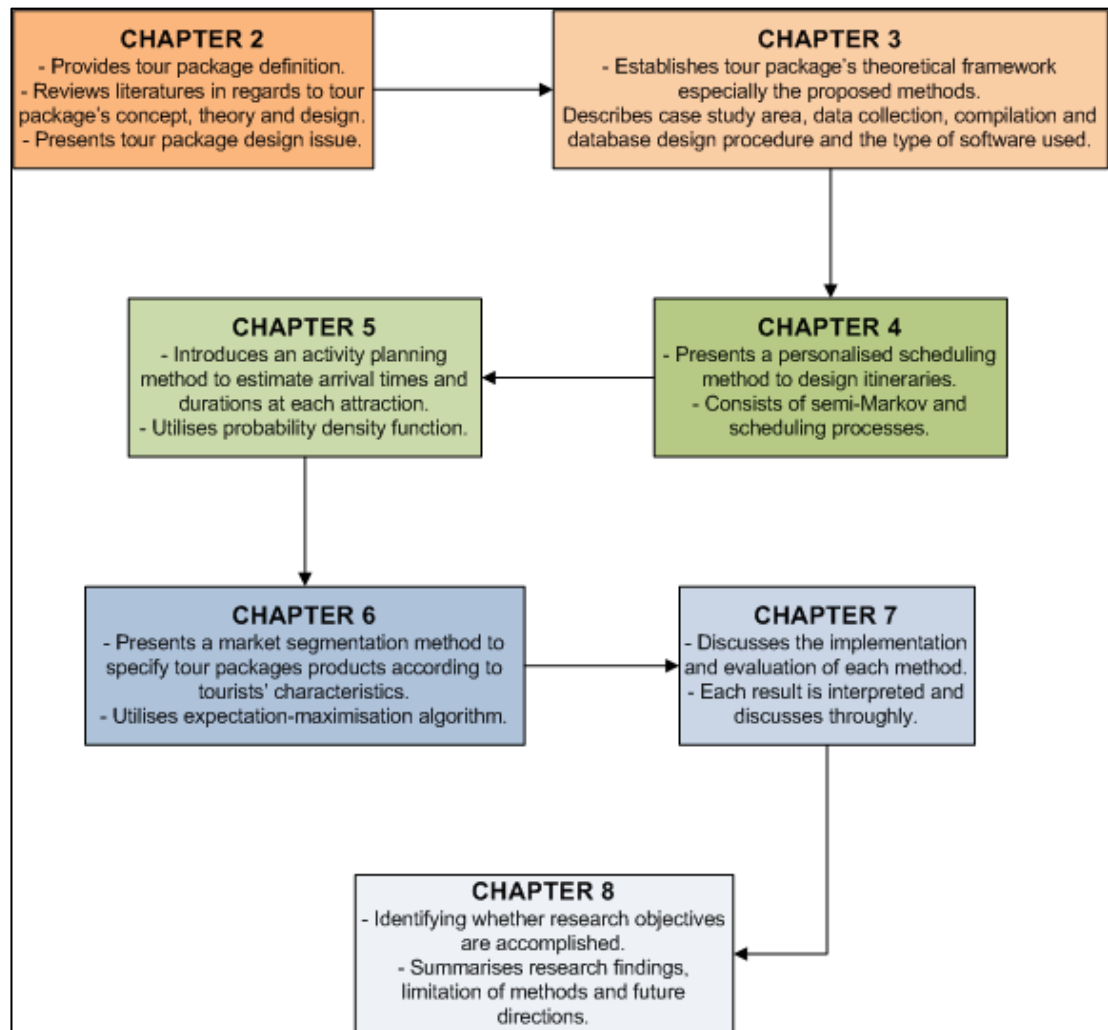
Figure 1.1   Thesis outline

Chapter 2 starts with reviewing the existing literature on various definitions of tour packages to provide an exact definition of a tour package for this research. Then, it reviews the appropriate literature in regards to the concepts and theories of tour packages and the existing methods of their design process.

Two types of methods are assessed in detail. They are methods directly and indirectly related to the design of tour packages. These indirect methods, although not directly related to tour packages, are able to contribute to the development of tour packages. This chapter concludes by presenting some issues of tour package design. The first and second objectives with research questions 1, 2 and 3 are addressed in this chapter.

Chapter 3 establishes the research methodology consisting of the theoretical framework of the research. This theoretical framework for designing tour packages is made up of five parts which are issues, stakeholders, stakeholder requirements, methodological requirements and solutions. The solutions highlight three methods to solve those issues. Again, the first and second objectives with research questions 2 and 3 are highlighted in this chapter.

A brief introduction of the case study area is given alongside with the data collection procedure. The data compilation, database design procedure and finally the type of software are presented to conclude this chapter.

The purpose of Chapters 4, 5 and 6 is to explicitly elaborate on the three methods mentioned above before providing the related results in Chapter 7. **Chapter 4** presents the Personalised Scheduling Method (PSM) to solve a tour planning issue for tourists. It applied the semi-Markov process, mentioned as the mathematical model in Section 1.1, to identify dominant movement sequences and arrival times for specific tourists' characteristics which will then be used to form a tour package itinerary. The second objective together with research question 4 is addressed in this chapter.

Chapter 5 introduces the Activity Planning Method (APM) to solve an activity planning issue for attraction administrators. The second objective together with research question 5 is highlighted in this chapter. A Probability Density Function (PDF), mentioned as the concept of probability in Section 1.1 is used to choose several arrival times and visiting durations where tourists arrive and spend at a particular attraction. This provides multiple options of arrival times and visiting durations for attraction administrators to consider.

Chapter 6 addresses the second objective together with research question 6 where the Market Segmentation Method (MSM) is suggested to solve the market segmentation issues faced by tour operators. Itineraries, attractions and activities are segmented according to respective tourists' characteristics using the Expectation-Maximisation (EM) Algorithm, mentioned as the clustering algorithm in Section 1.1.

Research question 7 is also highlighted where multiple tourists' characteristics are identified for the segmentation effort.

Chapter 7 discusses the implementation and evaluation of the entire methods presented in the previous three chapters through a case study. The first step involves the design of itineraries via the process of spatio-temporal movement modelling. The second step estimates the arrival time using PDF for activity planning. The final step involves the market segmentation process using the EM algorithm. Each result is interpreted and discussed. The third objective with all of its research questions are addressed in this chapter.

Chapter 8 concludes the thesis by giving a summary of the research findings. The limitations of the methods are discussed, and some of the future directions of this research are given. Most importantly, the research objectives and questions are reiterated to identify whether they are met or not throughout this thesis.

1.6   Chapter Conclusions

This chapter introduces the research by providing some background on the tour package design and tourists' spatio-temporal movement research. Later, this background is further detailed to formulate the problem of this research. Throughout the study, it establishes the objectives and the significance of this study. Finally, an outline of this thesis is given to provide a brief introduction on how this study is carried out.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter presents the findings of several existing research studies with regards to tour package design and identifies some issues that arise. As there are various definitions of a tour package, the chapter begins by listing some of those definitions. Next, a discussion on the history of tour package design is presented.

After that, a review of existing methodologies related to tour package design and implementation are outlined. The reviewed methodologies are categorised as two types of methods: (1) methods directly related to the design of tour packages and (2) methods indirectly related to the design of tour packages. Finally, several issues about tour package design are presented.

2.2 Defining a Tour Package

The definition of a tour package varies widely in the literature. Jafari (2000) defines a tour package as a form of travel organised by tour operators, and that the tour package represents a bundle of tourism products such as accommodation, transportation and tours. At least two of the products need to be marketed as one particular product to be considered a tour package. He also described the

interchangeable use of the term 'tour package' and 'itinerary' by tour operators. Syratt and Archer (2003) similarly define a tour package as a combination of tours, transport and accommodation, purchased in bulk by a tour operator to generate more profit and resold as part of a package.

The aspect of trip planning in a tour package was introduced by Morrison (1996) who stated that the tour package is a trip planned and paid for at a single price in advance, covering commercial transportation, accommodation, meals and sightseeing. Dellaert et al. (1995) added one more tourism product that forms part of a tour package which is attraction's activities.

The definitions of a tour package by Jafari (2000), Syratt and Archer (2003) and Morrison (1996) have a common feature of identifying the components of a tour package as accommodation, transportation and attractions. Notice that, the component of attractions is closely related to tours and sightseeing that is conducted by visiting attractions. An attraction is defined as a sight or place of interest in other literature.

Other tour package definitions focus more on the experience of tourists with references to the sequence and schedules for visiting attractions (Xia, 2007; Xia and Zeephongsekul, 2009; Tam and Pun-Cheng, 2012). Tam and Pun-Cheng (2012) refer to a tour package as comprising the following elements: the sequence of attractions, the journey duration, visiting durations, arrival times and transportation modes. This definition includes the spatial and temporal aspects of tourist movement in terms of the sequence and schedules for visiting attractions.

The range of spatio-temporal movement can differ according to several definitions. From a temporal perspective, a tour package can range from a day tour and can extend to weeks or even months (Dunstall et al., 2004; Tam and Pun-Cheng, 2012). From a spatial perspective, tour packages can include a single attraction to a list of attractions, a whole city or multiple cities and from a country to a group of countries (Dann, 1996; Heung and Chu, 2000).

Xia (2007) describes a tour package as a sequence of attractions, their schedule and target markets. This definition includes the marketing aspect of a tour package. The marketing aspect of a tour package includes efforts to relate tour packages to characteristics of tourists, better known as market segmentation (Thomson and Pearce, 1980).

Scheduling in a tour package includes determining suitable arrival times and visiting durations, that should be influenced by tourists' preferences and spatio-temporal movements (Xia and Zeephongsekul, 2009). This means observing the movement of tourists and including the spatio-temporal movement data in the analysis to design tour packages.

This study considers the definition of a tour package as a sequence of attractions, arrival times and visiting durations tailored to various tourist characteristics as the compulsory components for trip planning purposes. Tourists' characteristics include their demographic characteristics such as age and gender and non-demographic characteristics such as origin, preferred attraction and trip frequency. The definition is robust in a way that it combines the concepts of trip planning and tourist characteristics.

2. 3   A Brief History of Tour Package Design

The history of tour packages organisation began in 1841 when Thomas Cook organised the first trip for his members of the Teetotalers Club from Leicester to Loughborough, which at that time was a popular tourist destination (Enzensberger, 1996). He managed to convince Midland Counties Railway to slash the ticket prices when the ticket is purchased in bulk to cater for his group of tourists (Hunter, 2004).

When purchasing tour packages, most tourists are concerned with their budget (Budeanu, 2005). Therefore reducing the cost of tour packages has the chance to increase the possibility of tour packages to be bought. It is also noticeable that a tour package somewhat adheres to the preferences of a larger group of tourists where first, they prefer Loughborough as a destination (as in Thomas Cook's package) and

second, they prefer tour packages with lower prices (Enzensberger, 1996; Hunter, 2004).

Another development in tour packages is the introduction of cross continent tour packages to Europe and the USA in 1879, again by Thomas Cook (Syratt and Archer, 2003). Almost a century later, after World War 2, Vladimir Raitz organised the first British charter trip to Corsica using a chartered aircraft (Nilsson, 2011). Both instances, by Cook and Raitz highlighted that the tour packages have expanded geographically to include more countries indicating the importance of internationally observing tourist movement, to and from the destination.

It is useful to keep a record of the number of tourists who participated in certain tour packages that covered certain destinations. This information can be used in terms of marketing, to decide whether or not to retain those current tour packages or to design other tour packages with different destinations if the number of tourists who participated in the current tour packages is low. Although the example given is for international tour packages, this similar concept of observing tourist movement can be utilised to assess the marketability of tour packages at a lower level such as at the states or national parks level in terms of whether or not to include specific destinations.

The most important innovation that a tour package produces is the concept of attraction bundling into itineraries. Again, Thomas Cook took this initiative by assembling the element of travel and entertainment together with an efficient organisational framework that made it possible to provide the services at a price that people could afford (Brendon, 1991).

Furthermore, travel has become something pre-arranged with certain risks eliminated (Enzensberger, 1996). This can be easily understood whereby including other elements of a tour package such as transportation, food and accommodation, travel has become more convenient, and interruption during travel is less likely to happen.

Among the most significant improvements in tour packages after the 1990s is the introduction of Information and Communication Technology (ICT) specifically in terms of generating tour packages online. Tourists now are able to choose and customise their preferred tour packages according to their preferences in terms of attraction selection and their own timing (Pröll et al., 1999).

One example is manipulating a Case-Based Reasoning (CBR) recommendation system that relies on a relational database with interactive query management via a web-based platform (Bogdanovych et al., 2006). This saves time and tourists could avoid pressures of travel choices given by tour agents if the purchasing were made at the tour agency office.

2.3.1 Tour Package Research Development: Its Concept and Research

Although the invention of a tour package dates back to the 1840s, it was not until just over 40 years ago that tour packages started to receive the attention of tourism researchers. Stemming from the 1970s, research into tour packages of tourists can be studied from a different number of points of view.

Early research by Askari (1971) studied a number of different tour packages and compares their demand using a model. Using tourist and trip characteristics data such as income, tour price per day and the number of attractions visited per day, he was able to forecast the demand of tour packages in terms of the number of people purchasing any particular tour package. This was done by utilising an economic mathematical model that combined the aforementioned data. The specific trip characteristic data such as tour price per day and the number of attractions visited per day strongly reflect the aspect of tourist preference.

While Askari chose to economically model the demand of tour packages, Sheldon and Mak (1987) modelled the demand of tour packages in terms of tourists' demographic characteristics in choosing tour packages as a vacation mode. An empirical model that employs data such as tourists' characteristics (age and level of income) and trip characteristics (journey duration and the number of attractions visited) were tested using a logic analysis.

The model identified tourists who are elderly, intent on visiting several attractions, involved few people in the party, intent on making short visits and be first-time visitors to the attraction are most likely to execute their tour using tour packages. On the other hand, those who are less likely to purchase travel packages were rich tourists wanting longer visits, were repeat tourists and preferred to travel in larger parties (Sheldon and Mak, 1987).

Another study into how demographic and lifestyle of tourists affects the development of tour packages was conducted by Abbey (1979). As one of the pioneering researchers on marketing tour packages, he built a methodology consisting of ANOVA statistical analysis to test whether or not demography or lifestyle of tourists had influence on designing tour packages. In his study, lifestyle characteristics were described as predefined tourists' type according to their preferences such as historically inclined travellers, culture seekers, leisure seekers and nature seekers.

Tourists would make their own judgment of which type of tourists' lifestyle variables that were most suited to them. Tourists who executed their tour are given the choice of two tour packages, each one representing a demographic and lifestyle approach of tour packages and they were asked to indicate their preferences. At the end of the study, it was found that lifestyle characteristics are superior to demographic characteristics in influencing the design of tour packages.

Another perspective of the tour package research is related to the market segmentation of the tour package. A study by Thomson and Pearce (1980) identified some association between the characteristics of tourists such as their country of origin, age, gender, income and touring companions using a simple percentage analysis with different types of tour packages such as coach, camping or fly-drive tours. The research findings suggested that there were significant associations between certain tourist characteristics and their related tour packages as shown in Table 2.1.

Table 2.1   Tour package type with its associated

characteristics (Thomson and Pearce, 1980)

No.	Tour Package Type	Associated Characteristics
1	Coach tour	Couple tourists, older age and prefer comprehensive tour
2	Camping tour	Medium income earner, young tourists, predominantly female, single, on their first overseas trip, partner seeking and prefer cheap tours
3	Fly-drive tour	High income earner, couple tourists and seeking flexibility in their tour

A more fundamental study that defined the criteria of tour packages was done by Quiroga (1990) where these criteria were used to evaluate the efficiency of tour package elements. Quiroga listed four criteria which were: (1) a tour package that selects places to visit and its timing, (2) it controls tourist budget, (3) it acts as a mechanism to satisfy tourists, and (4) it combines adventure, novelty, escape and cultural experience, all with a margin of safety.

Notice that the third criterion involves satisfying tourists, thus tourists should be consulted for their preferences in designing tour packages. It is also interesting to know that, out of the mainstream characteristics such as the places, timing and costing of visits, Quiroga also mentioned safety as a criterion of tour packages. This can be observed with the inclusion of basic insurance for some tour packages nowadays that cover medical assistance and death, with the option to extend this basic cover to include wider coverage (Logitravel, n.d).

Enoch (1996) likewise listed four criteria of a tour package. They are: (1) it is a rational and effective way in achieving goals such as visiting the largest number of sites in a given duration, (2) it is usually less expensive than an individual trip to the same places, (3) a tourists' view is not considered when determining the contents of the package, and (4) it is not flexible in a way that once purchased, tourists have to strictly follow the plan and any changes are not permissible.

When Enoch's criteria are compared against Quiroga's, it is noticeable that there are two similarities where both of them stressed the importance of tour package in deciding places to visit and the arrangement of the timing. One more similarity would be on the significance of tour packages to lower the cost.

A noticeable difference between those two criteria is that Enoch mentioned the rigid characteristic of tour packages where tourists have no role in determining the content of tour packages. Both criteria can serve as a checklist when designing tour packages so as to achieve the purpose of tour packages and guarantee their robustness.

It can be seen that the subject of tour package research evolves from modelling the demand of tour packages in the 1970s to specifying the criteria of tour packages in the 1990s. Some findings from this research include the emergence of scheduling the visits to attractions for designing tour packages and the prominence of several types of tourist characteristics linked to some tour packages. The latest trend beginning in the mid-1980s is the research in designing tour packages quantitatively. This is elaborated more in the next section where various tour package design methodologies are critically reviewed.

2. 4 Methodologies of Tour Package Design

This review is divided into two parts: (1) the review of methods directly related to tour package design, and (2) the review of methods indirectly related to tour package design.

2.4.1 Methods Directly Related to Tour Packages Design

Several existing methods of tour package design focus more on the aspect of tour scheduling or the itinerary. One of these methods is the implementation of Electronic Tourism (E-Tourism) via CBR for designing tour packages in finding the best suggestion of travel schedule from one attraction to another (Niknafs et al., 2003). The tour package is the product of trip scheduling where a database of past trips (case-based) is built that greatly assists tourists with no existing knowledge of the tour's location.

Then, a tourist preferred trip (new case) is introduced where a similarity function is used to calculate the similarity between the tourist's preferred trip and previous trips. Hence, the choice of past trips on the similarity concept determines those for future trips. Although Niknafs et al. (2003) claim to suggest a travel schedule, the CBR method did not include any scheduling components such as visiting duration and arrival times at the attractions. Another drawback is that, tourists' constraints such as tourists timing availability and their budget is not considered in their method.

An enhancement of the research by Niknafs et al. (2003) was a study by Hsu et al. (2000) where they managed to suggest tour packages with attraction lists and visitation durations. An Interactive Genetic Algorithm (IGA) was used to flexibly and repeatedly change the initial tour package according to the evaluation of tourists based on the Genetic Algorithm's (GA) concept of crossover and mutation.

As an example in Figure 2.1, crossover happens when elements from two initial tour packages such as certain attractions swapped to form final tour packages with different attractions. This is in contrast to when in the initial tour packages mutation occurs resulting in a tour package changing its elements on its own as in the bottom right object in Figure 2.1 where the second element switches from 0 to 1.
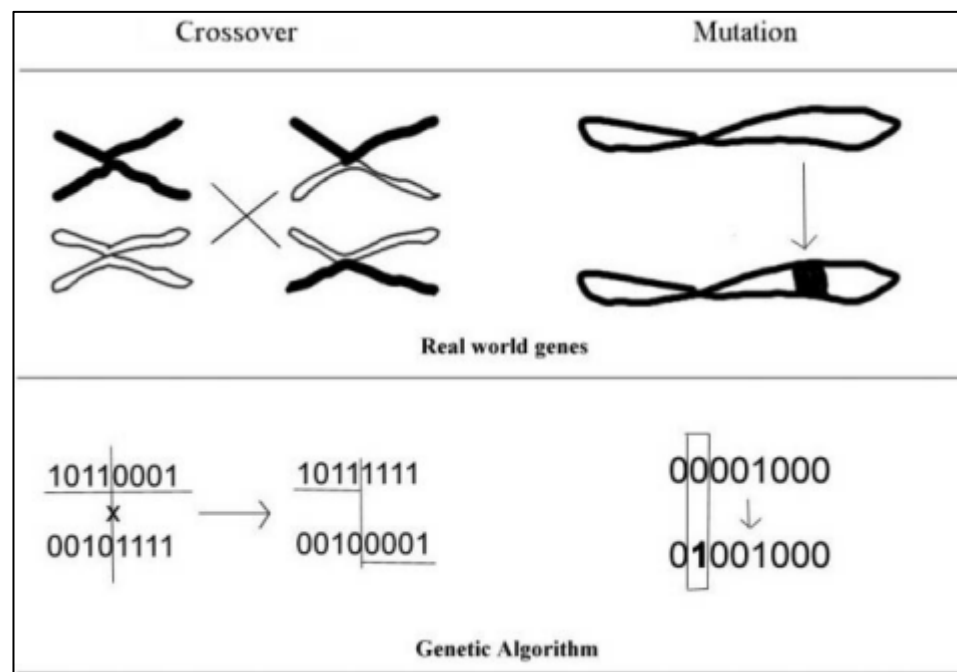
Figure 2.1   Distinguishing crossover against mutation (Kim and Cho, 2000)

An advantage of the system is that tourists are not required to define clear/crisp criteria (preferences) at the start. Instead they can learn to understand and change criteria through interaction with IGA from the moment the initial tour package was introduced. Despite the advantage, the system only suggests visiting duration as the temporal aspect of a tour package where this did not constitute a complete element of scheduling such as by introducing the arrival time for each attraction.

Some enhancement on CBR was made by Lopez (2003) by combining it with the Constraint Satisfaction Technique (CST) to produce a web-based system for generating scheduled tour packages. The method is holistic with, not only preferences considered, but also tourist constraints which are weighted when searching for similar past cases. Such constraints are like the opening hours of attractions and the maximum fee of entrance tickets that tourists are willing to pay.

The approach listed tour packages that include activities within an attraction, each with its timing element such as, arrival time, visiting duration and cost. The introduction of arrival time is an improvement when compared with the previous study by Hsu et al. (2000). Despite this enhancement, Lopez's method only suggested the arrival time and visiting duration randomly according to tourist preferences and attraction available time slot and does not implement the actual tourist movement dataset in terms of timing elements. This does not represent the actual behaviour of tourists, hence raising doubts on the authenticity of the suggestion of arrival time and visiting duration.

In their tour package design research, Tam and Pun-Cheng (2012) implemented a greedy algorithm to calculate the shortest paths from one attraction to another as an element of tour scheduling. Besides suggesting movement sequence from one attraction to another with its timing elements such as arrival time and visiting duration to spend at each attraction, they also suggested journey duration and transportation mode.

The suggestion of journey duration, as an improvement to the research by Lopez (2003) together with arrival time and visiting duration at each attraction constitute a complete temporal element for tour packages. While succeeding in suggesting an itinerary based on the spatio-temporal movements of tourists, a drawback of this model is that the shortest path does not always reflect the preferences of tourists.

The research by Hsu et al. (2000), Lopez (2003) and Niknafs et al. (2003) again highlighted the tour package design aspect of fulfilling personalised tourist preferences. The fulfilment of tourist preferences is another aspect in designing tour packages which is a fundamental concept of tour packages introduced by Thomas Cook (Hunter, 2004).

A concept of knowledge base that is a set of structured information was applied by Jakkilinki et al. (2007). In this research, an existing knowledge base known as Australian Sustainable Tourism Ontology (AuSTO) containing several classes as stated in Table 2.2 were implemented in designing itineraries for fulfilling tourist preferences. Each class has its own attributes.

Table 2.2   AuSTO classes and its attributes

(Jakkilinki et al., 2007)

No.	Class Name	Attributes
1	Involved Party	Represents a tourist, vendor or operator
2	Traveller Requirements	Stores the requirements of a traveller for a trip
3	Offering	Stores the information regarding vendor offerings
4	Resource	Stores information regarding tourism resources
5	Traveller Preference	Stores information regarding traveller preferences
6	Tourism Related Event	Stores information regarding tourism related events such as conference events
7	Traveller	Stores information regarding the traveller
8	Destination	Stores information regarding the destination

Implemented as a web-based system, AuSTO also connects between the entire classes that enable itineraries to be recommended containing elements from each class. An advantage of using AuSTO was that AuSTO includes a comprehensive set of classes most common to tour packages. It also defined the Involved Party class that shows the important role played by tour package stakeholders especially tourists.

Though complete, AuSTO's usage poses a risk where any inaccuracies or out-dated information is inherited into the system. Thus, using self-defined ontology based on a self-owned dataset could reduce the risk.

Loban (1997) defined the concept of 'matchability' to invent a computer-based decision support system known as Computer Assisted Travel Counselling (CATC) where it matches tourists' preferences against the attractions attribute. This answered an early call by Sirakaya et al. (1996) where they stressed the importance of designing personalised tours to match every possible type of tourist.

Initially, attraction attributes were compiled as a self defined knowledge base, an enhancement to the research by Jakkilinki et al. (2007). Mathematical models and multi-criteria decision making processes are then incorporated within the system. To evaluate the usability of the system, a comparison was made between CATC against other tour package design systems and concluded that his system was faster in exploring many alternatives and was able to handle more complex situations.

Though managing to design tour packages by considering tourist preferences with faster speed, the system did not explore the exact potential of personalising tours to fulfil tourist preferences. The scope of personalising tours only considers trip characteristics for instance journey duration, and the number of attractions visited and did not cover tourist characteristics such as tourist age and level of income (Askari, 1971).

Another focal aspect of tour package design is the inclusion of the timing aspect of activities. Activities are defined here as programs conducted at each attraction that tourists could participate in, such as horse riding, mountain biking and cultural performances (Becken and Simmons, 2002)

Some of the existing methods for planning the activity time involve the effort of scheduling the activity's starting time and visiting duration (Low et al., 1996; Dunstall et al., 2004). Both Low et al. (1996) and Dunstall et al. (2004) employed heuristic algorithms for scheduling activities into tourists' available time slots.

Tourists then use the scheduled time slots to choose their suitable arrival time for participating in the activities. This poses a risk where their arrival time might mismatch the actual starting time of activities that may be organised earlier by attraction administrators. They could miss the activities or arrive too early prior to the actual starting time of activities.

A system called Simulation of Travel/Activity Responses to Complex Household Interactive Logistic Decisions or STARCHILD implemented heuristic searching to produce multiple lists of scheduled activities (Recker et al., 1986a; Recker et al., 1986b). Later on, a utility function was used to evaluate the entire lists and select the one with the highest utility value.

The utility function is a mathematical function which ranks alternatives according to their benefit/loss according to different individuals (Oxford Dictionaries, 2013). Such examples of alternatives are admission fee, congestion level and the existence of any facilities in evaluating the attractiveness of an attraction (Apostolakis and Jaffry, 2005).

An advantage of this system lies in its ability to reserve an empty space for unplanned activities when quite often, there could be changes in the planning of activities. Recker et al. (1986b) also made an important mention of arranging activities in accordance with the arrival time of tourists. This is an improvement to the research by Low et al. (1996) and Dunstall et al. (2004) whereby establishing a connection between the starting time of activities and tourist arrival time, the risk where tourists could miss an activity is reduced.

The uniqueness of this system lies in its dependency on activities that later determined the attractions and was not rigid by instantly suggesting attractions without knowing the actual activities (Niknafs et al., 2003; Tam and Pun-Cheng, 2012). However, despite its ability to schedule activities, the timing aspect is discretised, i.e. not up to the level of hours and minutes (Recker, 1995). Moreover, activity duration was not included for completing the element of scheduling.

Similar to Recker et al. (1986a), Castillo et al. (2008) stressed more on the choice of activities in the design of a tour package but by implementing CBR. A CBR based system called SAMAP is used to select activities, attractions and appropriate transportation.

A planning module based on an artificial intelligence scheduling is then used to organise the activities and transportation to produce a sequenced tour package including an actual visiting duration for each attraction that is an improvement to the model of Recker et al. (1986a). Although timing element exists in terms of actual visiting duration to each attraction, Castillo et al. (2008) did not include another important temporal element which is the arrival time at each attraction.

Another tour package design aspect is the marketing perspective specifically the segmentation of the tour package market. The CBR was again performed by assimilating it with a decision-making dimension to create a model called DIETORECS for segmenting various tour packages according to different types of tourists (Fesenmaier et al., 2003). This extends the scope of tourist preference to include tourist characteristics such as tourist age, gender and decision styles.

Tourists have different choices for tourist attractions that are related to their characteristics, and they are classified by their decision styles: highly predefined users who are sure about their destination, accommodation oriented users, recommendation oriented users, geography oriented users, price oriented users and individual travellers who are uncertain about their destination.

DIETORECS then considers three types of tourists: (1) the most experienced tourists, the highly predefined users, who are given freedom to choose any tour package elements such as attraction, (2) the least experienced tourists, the individual travellers who are given a fully prepared tour package which they are recommended to adhere to, and (3) the mediocre tourists consisting of other types of tourists who are provided with the similar fully prepared tour package, but with a little freedom to change the elements according to their liking.

The entire suggestion was again based on the existing case of tour packages. Despite the fact that this method acknowledges the role of tourists when making the decision, the focus of tourist preferences based on the decision-making ignored the spatio-temporal element of tour packages where the only spatial element implemented by them is the tourist attraction and no temporal element such as arrival times and visiting duration within attractions were involved.

The review of each existing method related to tour package design revealed that there are a few design aspects for tour packages. The first design aspect is related to tour scheduling where the sequence of attractions, tourist arrival time and visiting duration are included. The second design aspect is in regards to the arrangement of activities, particularly their timing. Similar to tour scheduling, the starting time of activities based on tourist arrival time together with its duration is able to constitute a complete timing aspect of activities. The final design aspect is in terms of market segmentation of tour packages where various tour packages should specify their market according to different types of tourists.

2.4.2   Methods Indirectly Related to Tour Packages Design

Among the methods indirectly related to tour packages is the spatio-temporal movement analysis method that is found to have potential in the design aspect of tour scheduling. A prominent spatio-temporal movement analysis study that implemented an itinerary mapping analysis was used to analyse tourist movement in their itineraries (Connell and Page, 2008). In this research, tourists were required to provide their itineraries via movement patterns on maps where the entire movement patterns are then digitised into a Geographic Information System (GIS).

Maps were finally produced to show the multitude of itineraries via spatio-temporal movement patterns where these patterns are visually and statistically analysed. Although the system only analysed the historical pattern of itineraries and did not predict future itineraries, the practice of analysing historical itineraries can contribute to the development of future itineraries based on the assumption that tourist movement patterns can be replicated.

A mathematical modelling approach to analyse tourist spatio-temporal movement within the context of tour scheduling is Markov Chain and semi-Markov Chain modelling (Mednick, 1975; Lusseau, 2003; Xia and Zeephongsekul, 2009; Iosifescu et al., 2010; Xia et al., 2011; Cinlar, 2013). Mednick (1975) implemented the Markov Chain for modelling tourist movement in terms of making overnight stops within ten spots in Ontario province. Using the Markov Chain, he was able to produce a one-step transitional probability matrix that showed the probability of making overnight stops at one spot before going to another.

Despite the matrix outcome, he did not exploit the flexibility of the Markov Chain probabilities by using the concept of joint probability to show a complete movement sequence. Furthermore, he did not include any temporal element to his transitional probability.

An improvement to Mednick's Markov Chain method is the semi-Markov Chain where it differs from a Markov Chain in the way that it embeds a renewal process by including the temporal aspect of a tour package within its structure (Xia and Zeephongsekul, 2009; Xia et al., 2011). As an example, visiting duration at certain attractions was used as a parameter for predicting the probabilities of tourist movements within attractions.

Additionally, Xia et al. (2011) also fully exploited the flexibility of the Markov Chain probabilities. She combined multiple one-step transitional probabilities under the concept of joint probability to produce a complete movement sequence with its suggested visiting duration at any attraction.

A drawback of this research is that a complete movement scheduling as listed in Figure 2.2 is still impossible due to the absence of arrival time prediction. When combined with visiting duration prediction, a complete scheduling can be undertaken to produce a spatio-temporal itinerary.
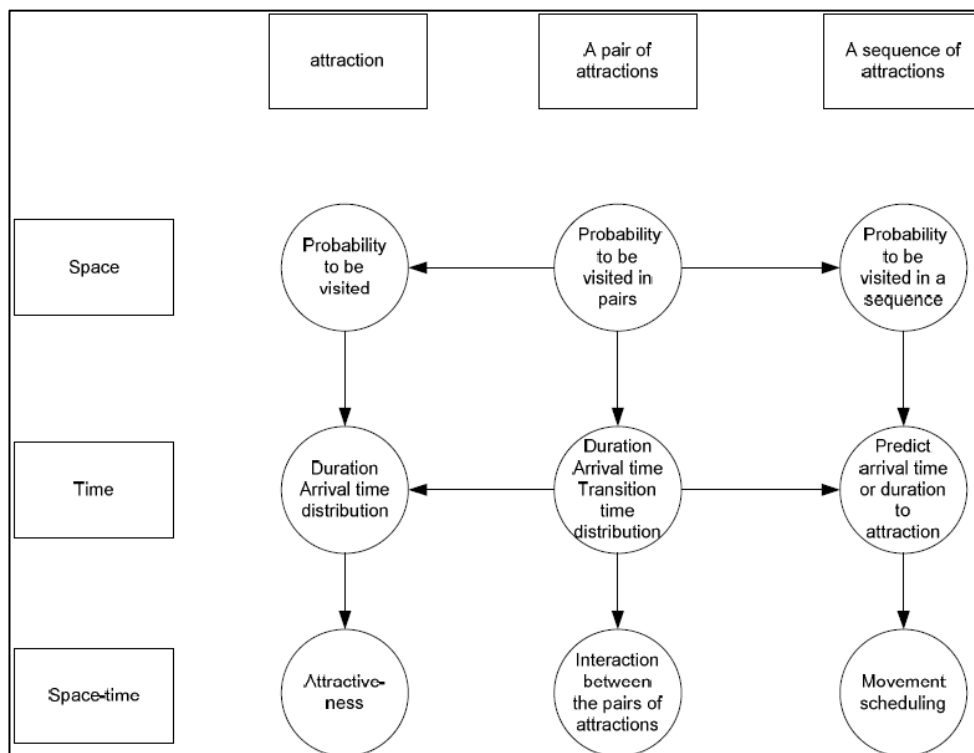
Figure 2.2   Movement scheduling combines arrival time and visiting duration (Xia et al., 2011)

There is also spatio-temporal and temporal movement research that covers the design aspect of tour scheduling and tourist preference but at the same time is able to contribute towards the arrangement of the timing of activities. A temporal movement study that focussed on the timing estimation utilised a Probability Density Function (PDF) in approximating the arrival time of buses to any destination (Knight, 1974).

As actual arrival time data are a continuous random variable, a PDF is used for the estimation where a set of arrival time data are fitted with different statistical distribution and the best distribution was chosen. The distribution is then used to find the probability density of arrival time for each time stamp.

An advantage of the study was that it suggested the usage of mode value, which is the highest frequency as a practical way in choosing the most suitable arrival time. Despite the fact that it is able to approximate the arrival time of coaches up to the level of minutes, it did not include the possibilities of having more than one mode, or more than one arrival time (multimodality phenomenon) for each arrival time

dataset. For tour packages, suggesting more than one arrival time is beneficial as tourists prefer flexibility in terms of arrival times at any particular attraction. Furthermore, duration as another component of activities timing scheduling that was absent from his research.

Mohammadian and Doherty (2006) used a PDF for modelling the duration of activities in explaining the phenomenon of activities scheduling. The PDF of four statistical distributions is implemented using an activities' duration dataset counted in minutes and the best distribution is selected based on t-statistics for finally explaining the observed duration. Using an activities duration dataset resulted in an improvement when compared with the research findings by Knight (1974). However, contrary to the research by Knight (1974), the fact that they did not include any dataset of arrival times, again lacked the complete element of the scheduling of activities. The phenomenon of multimodality was not included as well.

An agent-based system was implemented by O'Connor et al. (2005) to quantify tourists' sequential movement from one checkpoint to another within an attraction according to different tourist characteristics. An agent-based system is described as a system that is able to simulate tourist movement patterns via setting up predictive scenarios by defining the physiology and behaviour rules of tourists (Gustafson et al., 2003).

Tourist sequential movement is shown in Figure 2.3. The agent-based simulation software called Recreation Behaviour Simulator (RBSim) was used by O'Connor et al. (2005) to simulate the movement pattern of tourists with tourist movement data obtained from the Alge timing system.

Although some data were lost using the Alge timing system, the system was able to extrapolate the data, thus producing a list of sequential movements of tourists. They also made an important note of the different range of arrival times of tourists in each checkpoint where these arrival times should be considered when planning activities at each checkpoint.
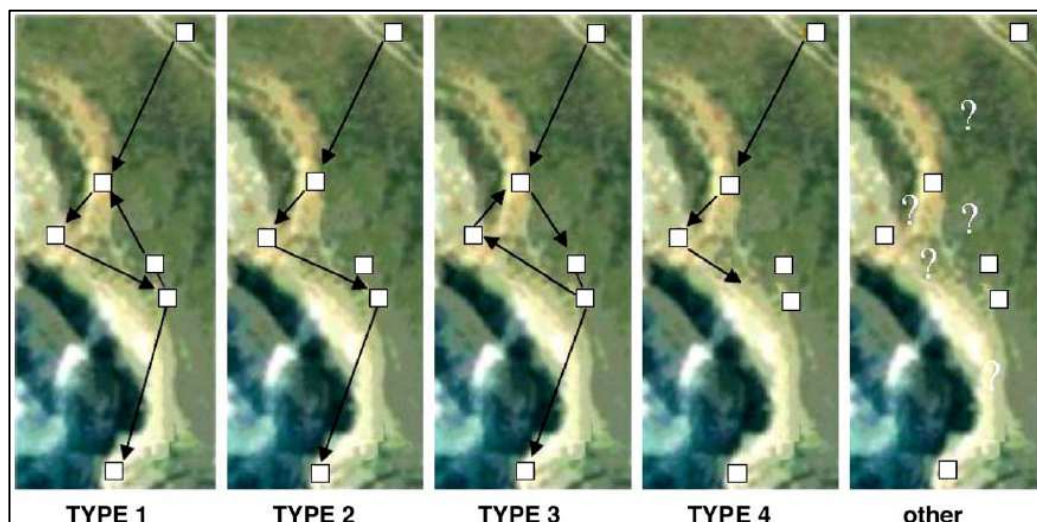
Figure 2.3   A list of tourists' sequential movements

(O'Connor et al., 2005)

One related shortcoming of the study is that they did not succeed in observing tourist preferences via tourist characteristics according to different movement patterns owing to the design of the study. Another shortcoming was that the lack of duration to complement the arrival times where this can further improve the arrangement of activities within each checkpoint.

Analogous to the tour package design related methods, the final tour package design aspect for indirectly related methods is the marketing aspect of a tour package. Research by Xia et al. (2008) that focussed on tourist way finding involved a spatio-temporal tourist movement model which covers visual analysis and simple statistics of frequency in describing tourist Global Positioning System (GPS) movement data. The movement data are related to three distinctive types of tourist characteristics in regard to the familiarity with the attraction: (1) familiar tourists, (2) partially familiar tourists, and (3) unfamiliar tourists.

At the end of the research, Xia et al. (2008) concluded that movement patterns were notably distinctive according to tourist familiarity where unfamiliar tourists had a tendency to use navigational aids such as signage. A drawback of this research was that only a simple statistical analysis was used to analyse a complicated set of tourist movement data according to different tourist characteristics.

The marketing aspect of tour package design for indirectly related methods is extended to include segmentation of tourist spatio-temporal movement patterns. Xia et al. (2010) used the Expectation-Maximisation (EM) Algorithm for clustering tourists according to tourist characteristics such as their age and gender for each type of tourist movement. This is an improvement to Xia et al. (2008) where EM algorithm is a more comprehensive clustering algorithm. A number of clusters were determined using the EM algorithm where each has its own associated characteristics.

As an example, Cluster 1 is declared as a young female in contrast to Cluster 2 which is declared as a young male. Each cluster has a particular pattern of tourists' movement. A significant advantage of the EM algorithm is that the likelihood (in identifying the cluster type) is guaranteed for each iteration (Budeanu, 2005).

This means, with each iteration of the EM algorithm, the likelihood of any particular tourist to be placed within any particular cluster increases. In the study, the authors admit that the insufficient amount of data made it impossible for the segmentation to be done for each movement pattern which was expected earlier.

Indirect methodologies of tour package design included in this review have applications in the field of spatio-temporal movement analysis, timing estimation, household activities scheduling and market segmentation. Although these methodologies are indirectly related, they have potential in contributing towards the design of tour packages.

Through comparing both methods related and indirectly related to tour packages, it is obvious that the tour package design aspects are similar in tour scheduling, activities timing arrangement and market segmentation. These aspects are important to robustly establish the framework of tour package design, and later on its specific methodologies.

2.5 Issues in Tour Package Design

There are three issues of tour package design identified from the existing literature: tour planning, activity planning and market planning issues. Some of the issues are backed by other supporting research.

Tour planning is an issue given that existing methods do not cover elements of scheduling by only suggesting attractions to be visited (Mednick, 1975; Niknafs et al., 2003). The itinerary mapping analysis did not even predict future itineraries (Connell and Page, 2008). The existing tour package scheduling research of IGA, SAMAP and semi-Markov Chain did not have a complete scheduling components by just including either arrival times or visiting durations but not both (Hsu et al., 2000; Castillo et al., 2008; Xia et al., 2011)

The unavailability of a complete tourists' spatio-temporal movement dataset can also result in a tour planning issue where the exact tourists' trends are not able to be identified (Hsu et al., 2000; Lopez, 2003; Jakkilinki et al., 2007). A few methods such as IGA and CBR combined with CST only implemented tourists' preferences data to design their tour (Hsu et al., 2000; Lopez, 2003) while one method using the AuSTO knowledge base by Jakkilinki et al. (2007) relied heavily on a secondary dataset over which they have no control.

The failure of CATC by Loban (1997) to holistically acknowledge tourist preferences by observing and associating tourist movement patterns with their characteristics such as age, gender and travel group is another reason for this tour planning issue to happen. This is risky as tour packages always adhere to the preferences of a larger group of tourists (Hunter, 2004).

Apart from suggesting potential attractions to be visited with its temporal components, a tour plan must also include possible routes for travelling within the attractions. Unfortunately, the respective tour package design research that utilised a greedy algorithm by Tam and Pun-Cheng (2012) used simplistic assumptions such as shortest path to derive routes. Finally, in regards to the probability concept, the existing research by Mednick (1975) did not fully exploit the Markov Chain ability

to quantitatively derive a complete transition probability from one attraction to another in forming a movement sequence for tour packages.

To conclude, tour planning can be summarised as having issues in regards to potential attractions to visit, visiting duration, arrival time to each attraction, possible routes for travelling to the attractions and the characteristics of tourists that are able to show their preferences in the movement pattern.

Activity planning is an issue where the arrangement of the timing of activities is not considered based on the integration of arrival time of tourists with current activities timing within attractions (Low et al., 1996; Dunstall et al., 2004). Instead, Low et al. (1996) and Dunstall et al. (2004) only considered tourists available time slots for the activities timing arrangement.

Starchild by Recker et al. (1986a) solved this issue by integrating tourists' arrival time with current activities timing. Despite this integration, the arrival time was only discretised. This is not ideal for stakeholders to plan activities in a timely and accurate manner.

Knight (1974) together with Mohammadian and Doherty (2006) suggested a solution to this issue where they managed to derive arrival times and durations in a continuous format up to the detailed level of minutes. Their research highlighted another significant activity planning issue where they did not consider a multi-modal scenario where more than one arrival time per attraction should be considered to give more flexibility for tourists in arranging activities.

Although this was solved by O'Connor et al. (2005) where multiple arrival times for various checkpoints were recorded, the unavailability of visiting durations to complement arrival times in their study resulted in another activity planning issue. This unavailability of visiting durations was also obvious in the Recker et al. (1986a) research. The research by Castillo et al. (2008) was notable in including visiting durations for activity planning but unlike the O'Connor et al. (2005) and Recker et al. (1986a) research studies, they did not include arrival times.

In conclusion, activity planning can be described as having issues in regards to the unavailability of the holistic element of activity time planning involving arrival times and visiting durations. The lack of detailed arrival times and visiting durations up to the level of minute together with the unavailability of multiple arrival times and durations of visits for an attraction are also of concern.

Tourists invest a considerable amount of money to purchase tour packages and in return they expect a high quality product (Phillips and Webster, 1983). However, some tour packages were not tailored according to tourist preferences based on appropriate characteristics of tourists (Loban, 1997). This is the basis of the market planning issue.

Some research studies that focus on the tourists' preference of tour packages, ignored the fundamental aspect of tour packages that should include timing elements such as arrival times and visiting durations (Fesenmaier et al., 2003). As a result, the link between tourist preferences and tour packages was only based on the element of attractions of these tour packages.

An improvement to this was a research by Xia et al. (2008) where she associated a tourist characteristic of destination familiarity with tourist spatio-temporal movement pattern of checkpoint sequence and arrival time for each checkpoint. In doing this, another market planning issue arises where she only conducted simple statistical analysis for linking these spatio-temporal movement patterns of tourists with their destination familiarity level.

Finally, in regards to the market segmentation concept, the existing research by Xia et al. (2010) was not able to fully exploit the EM algorithm for segmenting each tourist spatio-temporal movement pattern due to the inadequate amount of tourist characteristics data. Therefore, their earlier planning to observe the association of tourist characteristics for each movement pattern was aborted when only selected movement patterns were segmented (refer Section 2.4.2).

As a summary, market planning can be portrayed as having issues where firstly, the existing research did not link tourist preferences to a holistic tour package that consists of the tourist spatio-temporal movement pattern element (Fesenmaier et al., 2003). Other research only conducted a simple statistical analysis for market planning purpose while the final research did not have sufficient tourist characteristics data for undertaking segmentation for the entire movement pattern, thus producing a limited number of results (Xia et al., 2008; Xia et al., 2010).

2.6   Chapter Conclusions

This chapter evaluated existing research in regards to designing tour packages. Initially, several definitions of tour packages were listed and compared. Then, a brief history of tour packages was provided by emphasising their development.

Next, a review of possible methods of tour package design was presented and discussed. The review is divided according to two kinds of methods: (1) methods directly related to tour packages, and (2) methods indirectly related to tour packages. Finally, issues in tour package design were established based on the methods reviewed to guide this study in fulfilling objectives specified in Chapter 1.

Outstanding issues in regards to tour package design effort were identified according to several tour package design aspects, namely tour scheduling, fulfilling personalised tourist preferences, inclusion of an activity's timing and market segmentation. These outstanding issues were categorised into three kinds of issues which are tour planning, activity planning and market planning. These issues are considered in Chapter 3 for establishing the tour package design framework.

The inclusion of methods indirectly related to tour packages within this literature review revealed that methods that were not directly related to a tour package can be implemented to design a tour package. This is in relation to the spatio-temporal movement modelling or scheduling methods together with linking tourist characteristics for each spatio-temporal movement pattern.

CHAPTER 3: RESEARCH METHODOLOGY AND IMPLEMENTATION DESIGN

3.1 Introduction

This chapter discusses the research methodology and implementation design used throughout this research. It commences by identifying the requirements and stakeholders for designing tour packages based on the tour package design issues from Chapter 2.

Next, three methods that constitute the most important part of the research methodology are proposed to solve these issues. These methods, defined as solutions are briefly and separately described. Later on, the tour package design framework that contains essential components from tour package design issues and their solutions are presented.

The case study of this research is conducted in Phillip Island, Victoria, Australia. A brief introduction to Phillip Island is given where it justifies why the particular area is chosen for solving the tour package design issues. Finally, the data collection procedure, database design process and software types used are presented to conclude this chapter.

A preliminary tour package design framework is proposed in Figure 3.1 and includes five components which are issues, stakeholders, stakeholder requirements, methodological requirements and solutions. The flow of components starts from the inner layer of the figure which is the issues, advancing from one layer to another until the final layer of the solution.
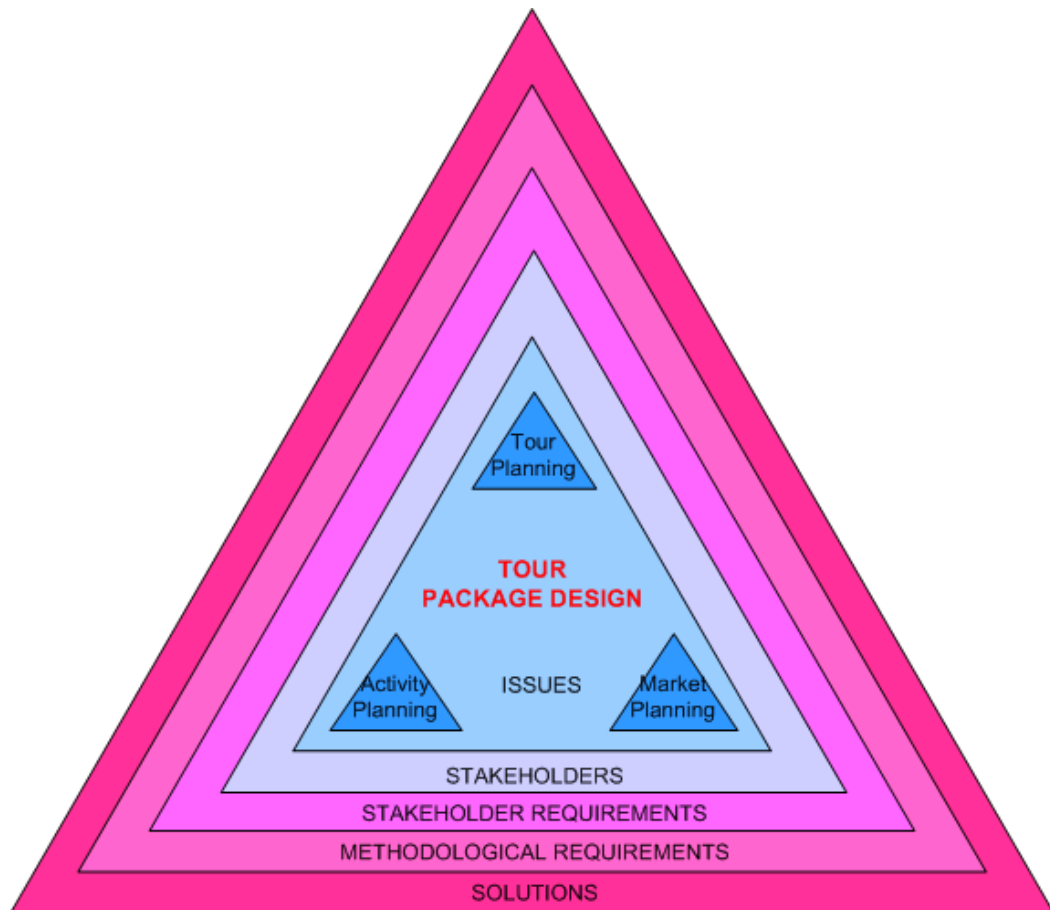
Figure 3.1   The preliminary tour package design framework

This flow of components is necessary for systematically developing the framework to achieve the objectives as mentioned in Section 1.3. The preliminary framework starts by presenting the tour package design issues faced by a tour package's stakeholders. Each stakeholder has their requirement in order for the issue to be solved. Another component to solve the issue is the methodological requirement. These two requirements are explained in Section 3.2. Finally, by incorporating these two requirements, three solutions for solving those issues are produced.

Currently, only one component which is the issue is populated, as described in Section 2.5. Each component will gradually be populated throughout this chapter until the framework is complete.

3.2 Tour Package Design Requirements

Two types of tour package design requirements are defined in this research: (1) stakeholder requirements, and (2) methodological requirements. Stakeholder requirements are the needs for tour packages to be designed for stakeholders while methodological requirements are the data needed to enable the methods as part of the solutions to be executed for tour package design.

There are three stakeholders of tour package design. The tour planning issue is unique to a tourist stakeholder. Tourists are the most important stakeholders of the tour package business since they are the main users of tour packages and customers of the other stakeholders. From a tourist perspective, planning a tour by means of scheduling is a basic need for a tour package. Tourists with different characteristics such as gender, age and travel group could prefer visiting different places at different times, all of which must be accommodated in a scheduling method.

For an attraction's administrator, a tour package can assist in determining activity or service time for tourists (Arrowsmith and Chhetri, 2003). The main purpose of a tour package for attraction administrators is to plan activities according to the arrival times and visiting durations of tourists at an attraction (Xia et al., 2009). To achieve this goal, a method to estimate the arrival times and visiting durations of tourists is needed.

For a tour operator, market planning is needed to understand their customers' characteristics and to provide them with a good tourist experience (Asakura and Iryo, 2007). A method to segment the market is required to identify target markets for a specific tour package product such as itineraries, attractions and activities.

There are two methodological requirements for tour package design: (1) tourist spatio-temporal movement, and (2) tourist characteristics. Attraction administrators and tour operators devise tour package products by observing tourist spatio-temporal movements during tours. Spatio-temporal movement of tourists is a reflection of tourists' travel preferences (Abdul Khanan et al., 2012).

Spatio-temporal tourist movement affects the development of tour package modelling (McKercher and Lew, 2004). One of the effects is how the knowledge of tourist movement patterns is essential to understanding tourist trip organisation such as attraction sequencing and timing of visits (Xia et al., 2010). Another effect is how tourists' movement decisions may assist attraction administrators in deciding how the daily program of activities should be arranged for an attraction in timely manners (Xia et al., 2010).

Another methodological requirement of tour package design is the inclusion of the characteristics of tourists. Characteristics such as age, gender and travel group (with whom the tourist travels with: alone, couple, kids, tour group) should be considered by acknowledging unique patterns of movements associated with each characteristic. Similar with spatio-temporal movement, tourists' characteristics can also have an effect on the development of a tour package.

Indirectly, characteristics can influence the spatio-temporal movement of tourists where each movement pattern can be unique based on distinctive preferences. This was shown by Arrowsmith et al. (2005) when they classified wayfinding routes based on gender, age and education differences, each with its own interests that have different movement patterns.

Xia et al. (2008) generated various maps that show different routes by tourists based on their level of familiarity of the surroundings. Abdul Khanan et al. (2012) identified the movement of tourists at a detailed micro level where each of the movements is associated with specific characteristics of tourists.

There are many characteristics that influence the decision-making of tourists during their tour, for instance socio-demographic profiles such as age and gender, and situational profiles such as travel groups and wayfinding strategies. Enoch (1996) emphasises the importance of socio-economic characteristics, where, for example, a particular group of tourists based on different social class may demonstrate distinctive preferences towards tour packages. Upper class tourists prefer a luxury tour package compared to other classes. A supporting finding by Wang et al. (2004) suggested that wives play a significant role in choosing the attraction, and they have a tendency in choosing an attraction with a shopping theme such as to buy souvenirs.

To summarise, tour package design requires the inclusion of two types of requirements, stakeholder requirements and methodological requirements. Three stakeholder requirements are presented where the different stakeholders have their own needs for tour package design efforts to be carried out.

Two methodological requirements are discussed where the first requirement of spatio-temporal movement describes the diverse movement patterns which are able to explain different preferences of tourists. For a proper scheduling process, the spatio-temporal movement must involve the arrival time and visiting duration to spend at each attraction. Journey duration if possible, should be included as well. The spatio-temporal movement dataset is specified according to various combinations of tourists' characteristics. As part of the method, further description of spatio-temporal movement with its categories is given in Section 4.2.

Tourist characteristics as the second methodological requirement such as gender, age and travel group are useful in explaining the differences of tourist preferences. As part of the method, further description of tourist characteristics with its types is given in Section 6.4.

3.3   Study Approach

The study approach lists the steps taken in addressing each tour package design issue to produce outcomes in achieving the objectives as stated in Chapter 1. It consists of six steps: solutions' proposal, case study area selection, description of methods, data

collection procedures, implementation and analysis/evaluation as shown in Figure 3.2.
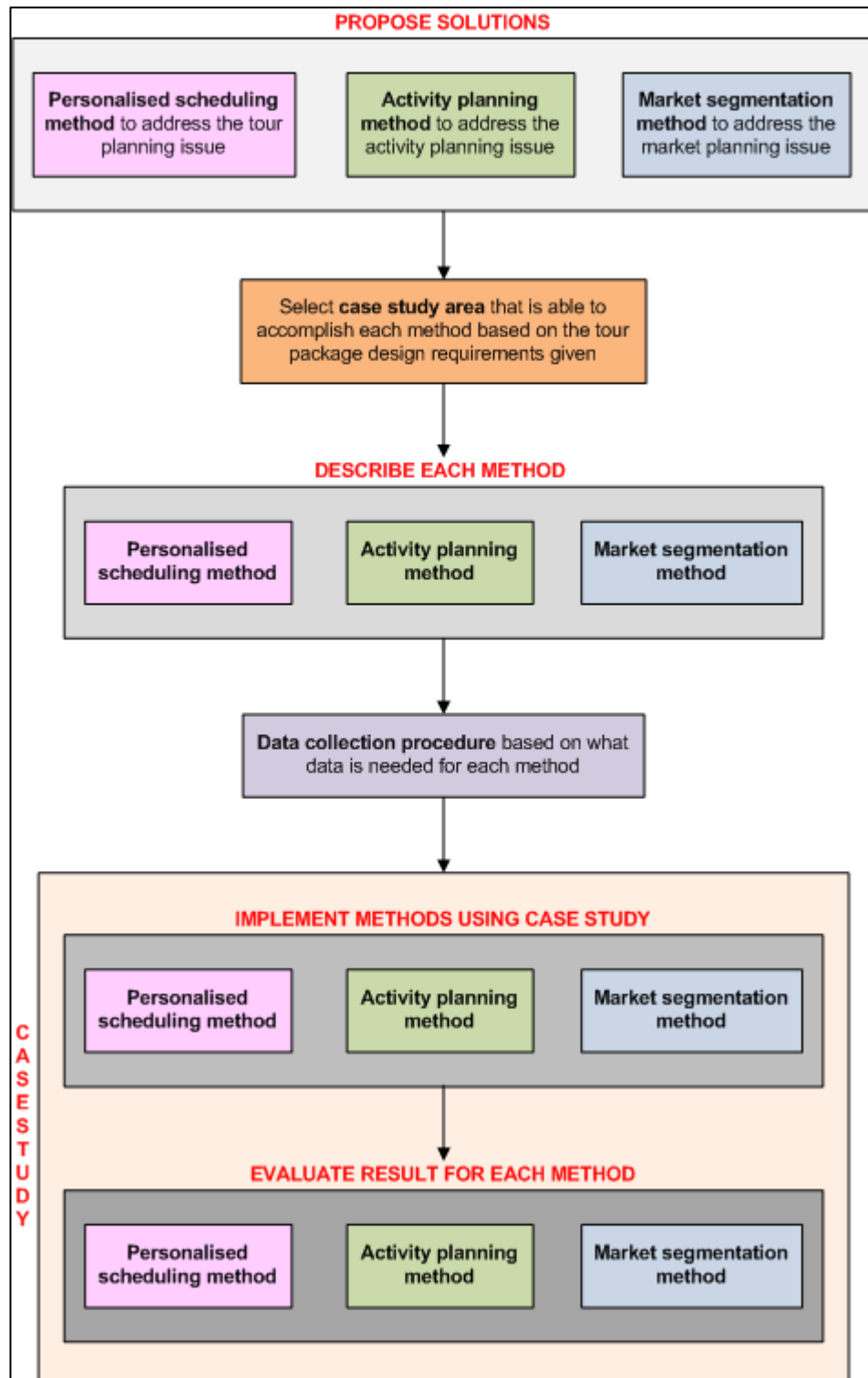
Figure 3.2   Study approach

Initially, several solutions for solving tour package design issues presented in Section 2.5 are proposed. These solutions are the final component of the tour package design framework. Three solutions are proposed to solve these issues, namely: (1) the Personalised Scheduling Method (PSM) to address the tour planning issue, (2) the Activity Planning Method (APM) to address the activity planning issue, and (3) the Market Segmentation Method (MSM) to address the market planning issue. These solutions are discussed in Section 3.4 and its particular sub-sections.

Next, the selection of a case study area is discussed where the case study area must be able to accomplish each method based on the tour package's design requirements given earlier. The case study provides an ideal ground for the realisation of each method in terms of each method's implementation evaluation. Section 3.6 discusses in detail the selection of a case study area, including its selection justification.

After selecting the case study area, each method is carefully described to ensure robustness of all methods. The methods are expanded in terms of detailing its methodological requirements or justification of data required, discussing the fundamental concept of its model and specifying analysis from the results obtained. Section 3.5 briefly details the steps in the description of this method while the next three chapters discuss the realisation of this step.

A significant part towards the end of this chapter (Section 3.6 and its particular sub-sections) is used to describe in detail the data collection procedure that was implemented. The data collection is crucial in gathering practical data to be implemented during the case study process. Details of data taken, sample size determination, sampling instrument, ethical consideration and sampling strategy are explained and justified.

Later on in this thesis, the case study is initiated in Chapter 7 where the first stage involves implementing the dataset within each solution. The tourists' spatio-temporal movement dataset linked to its respective tourist characteristics are applied within the PSM. The tourists' temporal movement dataset in terms of their arrival times and visiting durations based on specific attractions are applied within the APM. Tourists'

characteristics dataset is linked to its relevant tour package product such as itineraries, attractions and activities and is applied within the MSM.

The second stage of the case study is to evaluate the outputs from each method where finally, the efficiency for each case study is determined by considering whether or not the methods produced reasonable results. Both stages of the case study are briefly detailed in Section 3.5. The implementation and evaluation measures are carried out in Chapter 7.

3.4   Three Methods as the Proposed Solutions

This section briefly details the methods proposed as the solutions. The study approach that includes the steps of description, implementation and evaluation of a method are further detailed according to three separate workflows, unique for each method. Figure 3.3 presents this workflow. These steps are portrayed on the left side of the figure.

3.4.1   Personalised Scheduling Method (PSM)

The purpose of the PSM is to address the tour planning issue by developing an itinerary, according to tourists' preferences based on tourist characteristics such as gender, age and travel group. The produced itineraries include the sequence of attractions (i.e. where), arrival times and visiting durations (i.e. when) for a certain type of tourist (i.e. who).

These itineraries are customised for independent tourists. However, they can also be suggested for tour groups. Tourist spatio-temporal movement data in terms of their movement sequence from one attraction to another, the arrival time and visiting duration within each attraction are the focus of methodological requirements for this method.
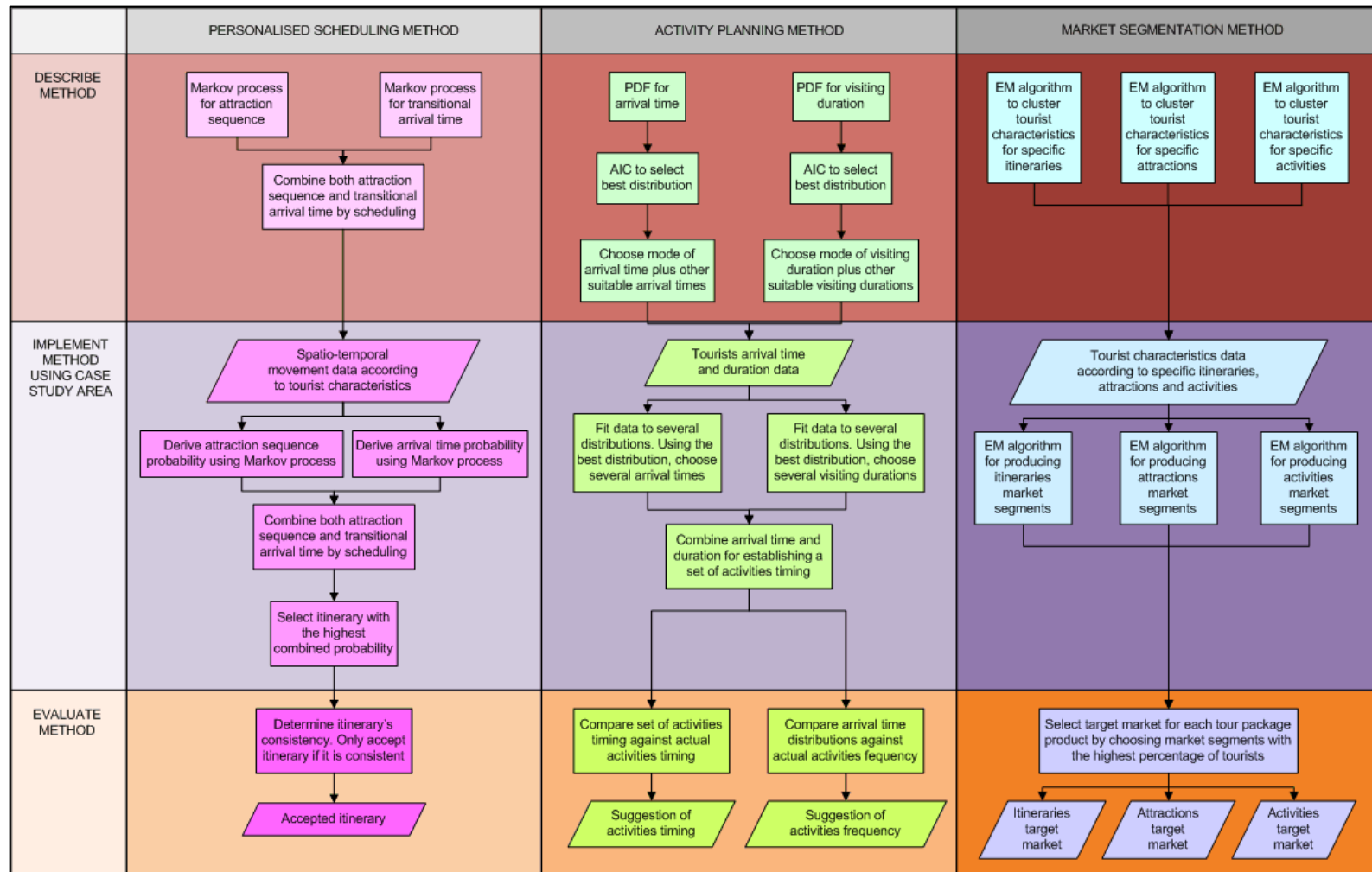
Figure 3.3   The workflow for each method

Tourist spatio-temporal movement can be a stochastic process. Therefore, a semi-Markov process is proposed for modelling the dominant spatio-temporal movement patterns. An advantage of a semi-Markov process is that the temporal element, such as the arrival time can be embedded within the model itself.

Additionally, a semi-Markov process can be exploited to combine multiple one-step transitional probability for producing itineraries that cover a complete movement sequence of tourists together with its suggested arrival time at any attraction (Xia et al., 2011).

The itinerary, derived from a combination of movement sequence and arrival time with the highest or relatively higher probability can later be chosen for designing tour packages. The description for this method in terms of the mathematical structure of the Markov process and scheduling as depicted in the top left corner of Figure 3.3 is discussed in Chapter 4.

The explanation of the implementation of this method in regards to producing movement sequence and arrival time probabilities and later on by combining both probabilities can be found in Chapter 4. This is shown in the middle left corner of Figure 3.3. The itinerary with the highest combined probability is chosen as the ultimate itinerary. The actual implementation using the Phillip Island case study can be found in Chapter 7.

As depicted in the bottom left corner of Figure 3.3., this itinerary is checked for its arrival time consistency. If any inconsistency is identified, the method is rerun until the consistency is reached. This procedure forms the evaluation step of this method. The detailed information for this evaluation step can be found in Chapter 4 while the actual evaluation using the Phillip Island case study can be found in Chapter 7.

3.4.2   Activity Planning Method (APM)

The purpose of the APM is to address the activity planning issue. This is by suggesting possible starting times and visiting durations for organising activities within each attraction based on the arrival times of tourists to each attraction

(Asakura and Iryo, 2007). Tourist temporal movement data in terms of arrival time and visiting duration data for each attraction are the focus of methodological requirements for this method.

The possible timing of activities are estimated using the PDF where it measures the probability of arrival times and visiting durations (Knight, 1974). The PDF is chosen due to its ability of estimating the arrival times and visiting durations as it can solve continuous random variables such as tourist arrival times. Furthermore, a PDF is able to suggest more than one arrival time or visiting duration by checking the mode of a probability distribution.

As depicted in the top middle corner of Figure 3.3, a PDF is used for fitting the arrival time and visiting duration datasets with distributions for instance Extreme Value, Nakagami, Weibull and Normal (Gaussian) distributions. Apart from unimodal PDFs, multimodal PDFs are also considered. An Akaike Information Criterion (AIC) is used to choose the best distribution where the distribution with the smallest figure of AIC is chosen. When suitable arrival time and visiting duration distribution curves are generated, the modes of arrival time and visiting duration that constitute the activities' timing element are identified for activity planning. The statistical outline of the PDF is further elaborated more in Chapter 5.

The explanation of the implementation of this method, shown as the central part of Figure 3.3 can be found in Chapter 5. Suitable arrival times and visiting durations obtained from the PDF and AIC procedure are combined to establish several sets of activities timing elements. The actual implementation using the Phillip Island case study can be found in Chapter 7.

As several activities' timing elements are estimated, comparisons can be made against the actual timing elements of activities within each attraction. This procedure forms the evaluation step of this method, shown in the bottom middle corner of Figure 3.3. If the comparison shows that they are different, attraction administrators have the option to retain the actual timing elements or they can adopt the estimated

timing elements as they are based on the actual spatio-temporal movement dataset of tourists. This can boost the participation in activities within their jurisdiction.

Finally, estimating more than one activity's timing elements create another aspect of activities organisation which is the activity frequency. Again, comparisons can be made in terms of the activity frequency if the suggestion of estimated activity, timing is adopted against the actual activity frequency.

Once more, attraction administrators have another option to implement the suggestion if they think it can improve the organisation of activities within their attraction. The detailed information for this evaluation step can be found in Chapter 5 while the actual evaluation using the Phillip Island case study can be found in Chapter 7.

3.4.3   Market Segmentation Method (MSM)

The purpose of the MSM is to address the market planning issue by clustering the market of tour packages' products, each into specific segments according to different types of tourists' characteristics. Tourist characteristics data such as age, gender and preference of attractions are the focus of methodological requirements for this method.

An Expectation-Maximisation (EM) Algorithm is used for clustering tourists' characteristics data according to each itinerary, attraction and activity. The EM algorithm is proposed to segment the market for tour package products where it has the ability to identify the likelihood of cluster parameters within a mixture model (see Figure 3.4) where the initial parameter is limited or unavailable (Witten and Frank, 2005).

In this research, the cluster mentioned is related to tourists' characteristics clusters. The model is called a mixture model as tourists characteristics as the subject of clustering can be assigned to more than one cluster. Furthermore, the EM algorithm is able to solve clustering issues when the dataset involved is nominal (Witten and Frank, 2005). Additionally, the EM algorithm used for clustering can produce more

than one cluster where each cluster has its associated tourists' characteristics linked to it. This means, a single itinerary can suits more than one type of tourists.
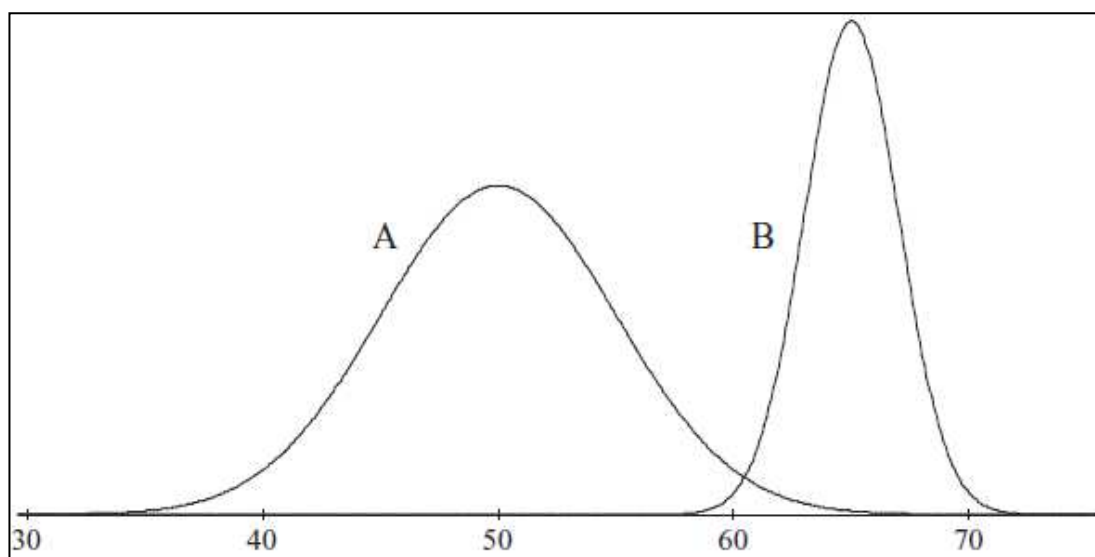
Figure 3.4   A mixture model highlighting two curves representing
clusters where each curve has its own parameters that
determine its shape (Witten and Frank, 2005)

The description of this method, shown at the top right corner of Figure 3.3 in terms of its mathematical structure of the EM algorithm is further detailed more in Chapter 6. The explanation of the implementation for this method, depicted as the middle right corner of Figure 3.3 using the Phillip Island case study can be found in Chapter 6. The output of this clustering procedure is several clusters or market segments where each cluster has some associations of tourists' characteristics. The actual implementation using the Phillip Island case study can be found in Chapter 7.

Finally, a target market is chosen for each tour package product, shown in the bottom right corner of Figure 3.3 as the evaluation step. A target market highlights a market segment, consisting of selected tourists that are selected for a particular purpose (such as to be promoted with specific tour packages) based on a certain criteria (Myers, 1996). Yet again, for commercial reasons, the target market from the market segment with the most number of tourists is selected where this is defined as the 'criteria' in the previous paragraph. The detailed information for this evaluation step

can be found in Chapter 6 while the actual evaluation using the Phillip Island case study can be found in Chapter 7.

3.5   The Complete Tour Package Design Framework

A complete tour package design framework is presented in Figure 3.5. The complete framework has similar shape and components to the preliminary framework, but now, each component is completely populated according to each issue. Each corner of the framework represents the specific issue with its related stakeholder, stakeholder requirements, methodological requirements and solution.

The first corner represents the tour planning issue where the stakeholder is the tourist. Tourists need to schedule their tour. To schedule the tour, tourist spatio-
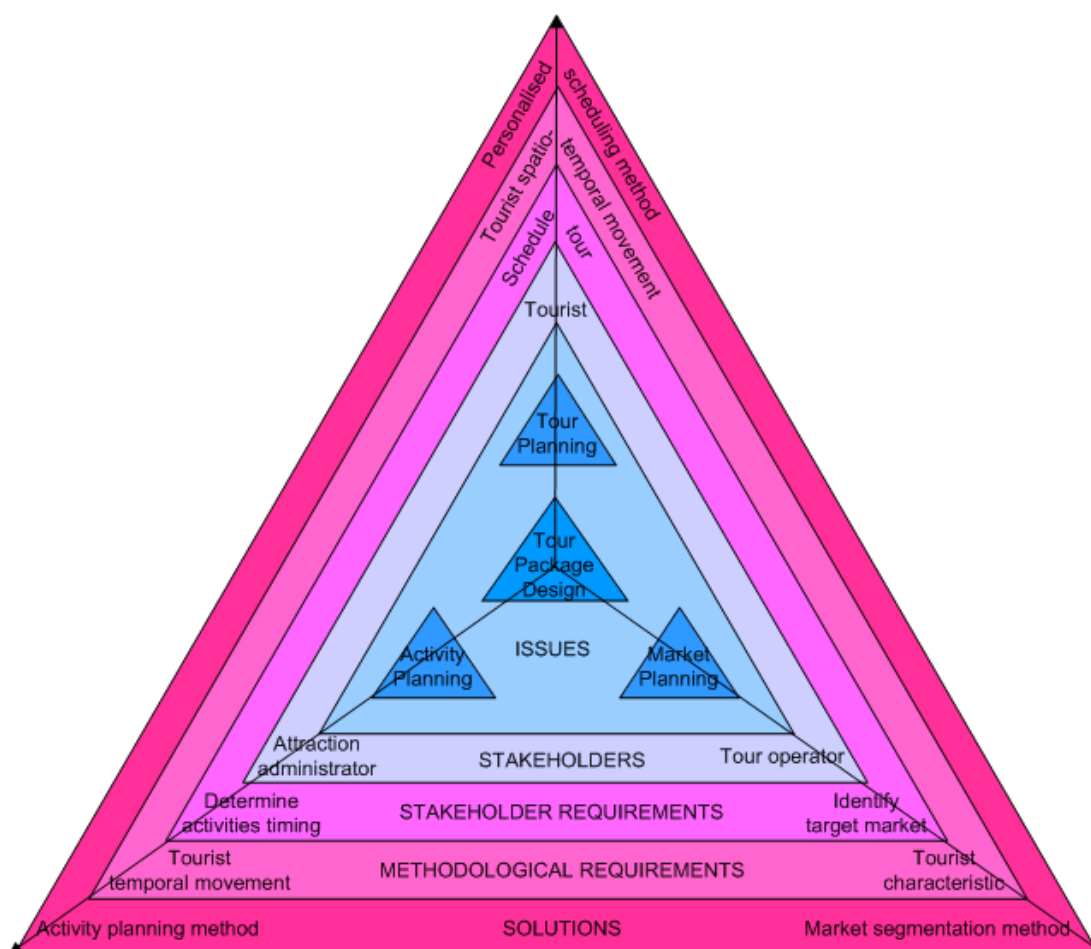
Figure 3.5   The complete tour package design framework

temporal movement data in terms of the movement sequence with arrival time and visiting duration for each attraction are needed. Finally, a PSM is proposed as a method to solve the tour planning issue by scheduling the tour.

The second corner represents the activity planning issue where the stakeholder is the attraction administrator. The attraction administrator needs to determine the timing of activities within their attraction. To determine the timing of activities, tourists' temporal movement data in terms of the arrival times and visiting durations are anticipated. Finally, an APM is proposed as a method to solve the activity planning issue by arranging the timing of activities.

The third corner represents the market planning issue where the stakeholder is the tour operator. Tour operators need to identify the target market of tour package products. To identify these target markets, tourist characteristics data such as age, gender and preference of attractions are anticipated. Finally, a MSM is proposed as a method to solve the market planning issue by segmenting the market of tour package products according to tourist characteristics.

3.6   Case Study Area

Phillip Island (see Figure 3.6) is located at Western Port, 120 kilometres south east of central Melbourne, Victoria, Australia (Phillip Island Nature Park, 2011). With an area of 10,000 hectares, stretching 26 kilometres long and 9 kilometres wide, it was awarded the status of Nature Park in 1996, the only one in Victoria. The centre of ecotourism in Victoria, it boasts around 346 species of birds and many other animals such as seals, koalas and the smallest species of penguin in the world (Artifishal Studios, 2012).

The permanent population is around 8,000 residents, but it received almost 2.7 million tourists (overnight and daytime) for the year ending June 2011 (Tourism Victoria, 2012b). Phillip Island has a temperate climate distinguished by its characteristic of having warm summers and cool winters. This proves beneficial to the tourism sector as during the summer, the number of residents there surges to 50,000 (Western Port Water, 2012).
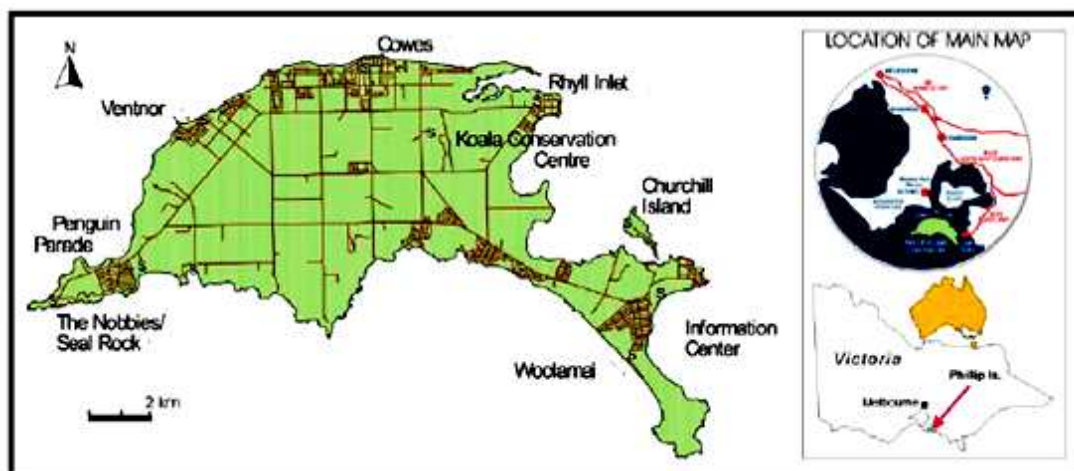
Figure 3.6   The map of Phillip Island with its main
attractions (Phillip Island Nature Park, 2005)

Tourism is important for the Phillip Island region where domestic Australian visitors spent a total of AUD 433 million during the economic year of 2012 (Tourism Victoria, 2012a). According to the latest available statistics, the economic year of 2010/2011 highlighted more than 250 tourism related businesses in the region where 84 % of the businesses were either micro or small businesses (Tourism Research Australia, 2012). Nature-based tourism dominated the tourism sector in Phillip Island with activities such as wildlife viewing and nature sightseeing (Tourism Victoria, 2008).

Few of these tourism related businesses are tourist attractions. Tourist attractions in Phillip Island can be divided into two: public and privately managed attractions. These tourist attractions can be further divided into two other aspects which are managed and unmanaged attractions.

A managed attraction is defined in this research context as an attraction that is managed by an official authority in a secluded area, most of them having a well-managed list of activities such as Phillip Island and Penguin Parade (Phillip Island Nature Park, n.d.). An unmanaged attraction has the opposite characteristics in a way that it is not managed by any official authority and is located usually in an open area. Normally, unmanaged attraction, such as Ventnor and Rhyll Inlet, do not have a proper list of activities.

Phillip Island is chosen as a case study area for several reasons. The first reason is that of the large number of tourists to Phillip Island. Deriving tourist movement probabilities according to their movement sequence, arrival times and visiting durations requires a large sample size to be collected, and Phillip Island seems to suit this purpose.

The second reason is because of the existence of several attractions within its area. Tourists usually commute between these attraction with some attractions having a well managed list of activities such as Churchill Island and Penguin Parade, complete with its timing (Phillip Island Nature Park, 2011). As one of the anticipated outcomes of the model is the movement sequence, having a case study area with several attractions where tourist usually commute between these attractions is suitable for the case study. For the APM, the existing activities with their actual starting times and recommended visiting durations can be compared with the estimated arrival times and visiting durations.

A further two reasons for the choice of the study area is related to the convenience of data collection. Firstly, as the entire attractions are located within the small island, it made the data collection process more convenient where the researcher could easily move from one attraction to another to monitor other data collectors. Secondly, the existence of Penguin Parade where 90% of tourists who visited Phillip Island usually concluded their tour also made Phillip Island a suitable case study area (Xia et al., 2010). The researcher with the data collection assistant spent most of their time here to increase the participation rate.

Table 3.1 lists the most typical attractions on Phillip Island. They represents the nature based tourism context that dominates the tourism sector in Phillip Island and their respective activities which is the subject of this case study (Tourism Victoria, 2008). They were chosen based on their popularity where 90% of tourists who visited Phillip Island will visit four of these attractions which are Penguin Parade, Information Centre, Koala Conservation Centre (Koala Centre) and Churchill Island (Xia et al., 2011). These four attractions are shown in Figure 3.7. Figure 3.8 shows

the movement patterns to four attractions, namely Penguin Parade, The Nobbies and Koala Conservation Centre that are listed by Xia (2007).

Table 3.1   List of attractions with their respective activities

No.	Attraction Name	Category	Activities
1	Penguin Parade	Managed	Interactive display at The Visitors Centre Penguin viewing Penguin movie
2	Koala Centre	Managed	Koala Boardwalk Woodland boardwalk Wildlife bushwalk Interactive learning at The Visitors Centre
3	The Nobbies	Managed	Special presentation at The Visitors Centre 'Just for Pups' kids session Boardwalk Seal viewing at The Seal Rock
4	Churchill Island Heritage Farm	Managed	Cow milking Wagon rides Blacksmith demo Working dogs Sheep shearing Whip cracking and boomerang throwing Baby animal visiting Visiting heritage buildings and gardens Animal viewing – Clydesdale horse, highland cattle and poultry
5	Cape Woolamai	Unmanaged	Coastal track walking (includes lighthouse and Pinnacle sceneries) Wildlife viewing
6	Rhyll Inlet	Unmanaged	Nature Walks (includes Oswin Roberts Koala Reserve, Conservation Hill lookout, Rhyll swamp bird sanctuary) Visiting historic sites
7	Cowes	Unmanaged	Visiting Cowes Cultural Centre Visiting museum Dining and shopping
8	Visitors Information Centre	Managed	Tour exhibition
9	Ventnor	Unmanaged	Coastal track walking (includes French and Flinders Island sceneries)

Figure 3.7   Phillip Island's top four attractions. From top right and clockwise: a boardwalk at The Nobbies, a cottage at Churchill Island, an entrance to a boardwalk at Koala Centre and Penguin Parade (Xia, 2007)

Both managed and unmanaged types of attractions are included for comparing the pattern of tourist arrival times and visiting durations in APM. The entire attractions are public attractions highlighting the absence of private attractions which are the limitation of this case study.

3.6.1   The Data Collection and its Procedure

Data collection is needed since no sufficient existing dataset related to the methodological requirements as discussed in Sections 3.2 to 3.5 exists. A related research by Xia (2007) managed to gather a tourist spatio-temporal movement dataset. However, she only gathered the movement sequence and data of arrival times of tourists without visiting durations where it is known that visiting duration is the   methodological   requirements   of   PSM   and   APM.
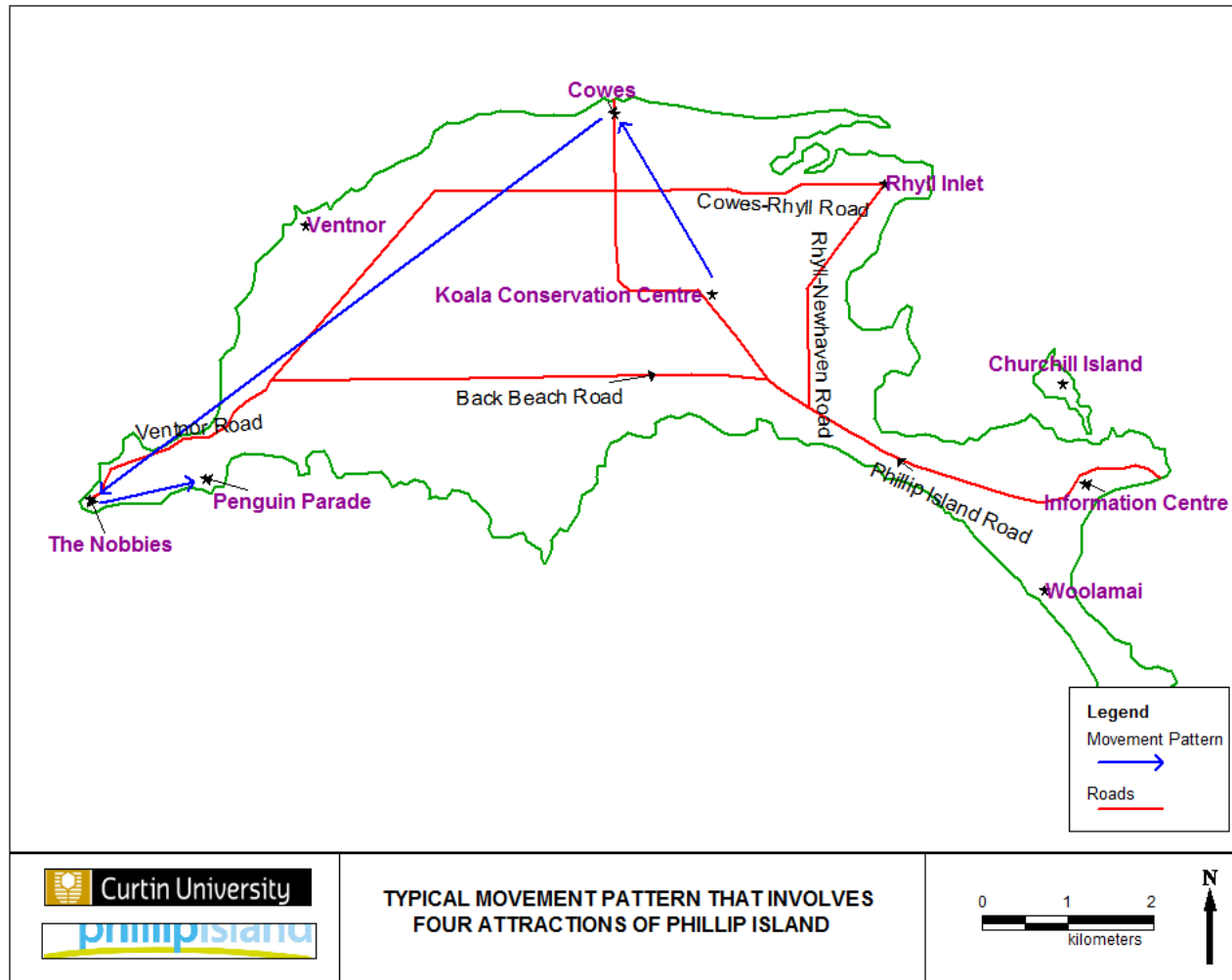
Figure 3.8   A typical movement pattern that involves four attractions of Phillip Island (Xia, 2007)

Furthermore, her sample size of 464 tourists is inadequate to produce detailed tour packages for various tourist characteristics (Witten and Frank, 2005). Another survey by Xia et al. (2010) which focussed on market segmentation faced similar issues of data inadequacy where smaller sample size caused similar issues of incapability to produce various tourist movement patterns as anticipated earlier.

The data collection procedure consists of several steps such as determining the details of data taken, calculating the size of the sample, choosing the sampling instrument and abiding by ethical considerations and lastly, planning the sampling strategy.

3.6.1.1 Details of Data Collected

The anticipated data are related to tourists' movements and their respective characteristics. The majority of the characteristics data are qualitative; however, some are quantitative for instance, visitation frequency, time spent on Phillip Island, age and annual household gross income.

The complete tourists' spatio-temporal movement data such as movement sequence, arrival time and visiting duration at each attraction are quantitative. Table 3.2 lists the details of data collected for each method. One important aspect is that different data are gathered for each of the three methods as discussed in Sections 3.2 to 3.5. However, some data are used for other methods.

Table 3.2   Details of data

No.	Method	Anticipated Data
1	PSM	Tourists' spatio-temporal movement data – sequence of attractions, arrival times and visiting durations at each attraction.
2	APM	Tourists' arrival times and visiting durations at each attraction
3	MSM	Tourists' characteristics data – age, gender, travel group, frequency of tour, purpose of tour, preferred type of attraction, overall duration of tour, origin, household gross income, level of education, mode of transportation.

Tourist spatio-temporal movement data meant for PSM are used in MSM to produce itineraries to be segmented. On the other hand, tourist characteristics data intended for MSM are used for PSM for personalising itineraries according to specific tourist characteristics. The dataset must enable the case study to accomplish the objective of each method.

3.6.1.2 Sample Size Determination

Sample size is an important issue in this study as a semi-Markov process that involves deriving movement probability according to tourists' arrival times at each attraction is unreliable if a small sample size is used. Furthermore, the statistical power of a Markov process would increase in parallel with the increment of the sample size (Tan and Yılmaz, 2002).

The calculation of sample size, ss is done based on the confidence level, z, population size, N, proportion, p and confidence interval, c using an online calculator for calculating sample size (Australian Bureau of Statistics, n.d.). It consists of two equations as shown below (Leiper, 1990). Equation (3.1) assumes an independent population therefore not incorporating population size. Equation (3.2) provides the corrected sample size, SS for a finite population, i.e. incorporating the population size figure.

$$ss = \frac{z^2 \times p \times (1-p)}{c^2} \tag{3.1}$$

$$SS = \frac{ss}{1 + \frac{ss-1}{N}} \tag{3.2}$$

The confidence level depicts the percentage of certainty that the sample statistic lies within the population parameter (Moore, 2001). The proportion marks the percentage of the population that possesses a specified parameter (Becken and Simmons, 2002). In this study, the proportion is 50% that tourists have specific movement patterns according to their characteristics.

As the exact proportion is unknown due to the uncertainty of whether tourists have specific movement patterns according to their characteristics, the 'worst case scenario' proportion value of 50% or 0.5 as this produces a conservative estimate of variance (Creative Research System, 2012). Lastly, the confidence interval indicates the desired level of accuracy of the parameter estimation in terms of how confident that the result is correct (Moore et al., 2012). This is expressed by the plus and minus figure of percentage away from the proportion figure to form the interval.

The figure of the population is 1,020,000 representing the overall number of domestic, intrastate, interstate and international daytime tourists to Phillip Island for the year ending June 2011, the most recent data during the time of the data collection process (Tourism Victoria, 2011). After the calculation of sample size using a 3 % confidence interval or 0.03 (to create the interval of 47% to 53% that tourists have specific movement patterns) and 95 % confidence level or 1.96, a figure of 1,066 tourists is obtained as the recommended sample size. The calculation of the sample size according to Equations (3.3) and (3.4) is provided below.

.

$$ss = \frac{1.96^2 \times 0.5 \times (1-0.5)}{0.03^2} = 1,067 \qquad (3.3)$$

$$SS = \frac{1,067.111}{1 + \dfrac{1,067.111 - 1}{1,020,000}} = 1,066 \qquad (3.4)$$

3.6.1.3 Sampling Instrument

A sampling instrument of a questionnaire was chosen as it is able to generate information not available from secondary sources about the characteristics of respondents, their perceptions and spatial behaviour (McLafferty, 2010). The questionnaire was divided into two main sections, each corresponding to the characteristics of tourists and their movement during their tour as discussed earlier.

The questionnaire consists of 14 questions (see appendix 1). Questions 1 to 13 that cover tourist characteristics aspect is placed in the first section while the only question that involves the tourist movement aspect is placed in the second section as the last question.

There are more than one tourist characteristic, for instance geographic, demographic, psychographic and behavioural characteristics. The anticipated tourists' characteristics data include age, gender, travel group, origin, preferred attraction and purpose of the tour. In terms of the MSM, this scope of tourist characteristics is explained further as segmentation variables in Chapter 6.

The only question in the second section provides tourists with a map where they are required to specify their movement details such as their movement sequence, arrival time and visiting duration at each attraction. They are asked to identify their starting time for their participation in each activity as well.

3.6.1.4 Sampling Strategy

The sampling technique is determined by two factors: sample size and study area characteristics. As the calculated sample size of 1,066 is large when compared to the preliminary plan of sampling timeframe of 8 days, a stratified random sampling by distributing the questionnaire in person at all attractions was devised.

For this study, a stratified random sample was obtained by separating the population of tourists into non-overlapping groups called strata, and then selecting a simple random sample for each stratum (Scheaffer, 2006). In this research, respondents were stratified according to the attraction where they were first handed the questionnaire.

Each stratum is guaranteed independent where there was not a single tourist captured twice at different attractions. Prior to participating, tourists were asked if they already responded previously. Several tourists who already took part in the survey, informed the researcher of their previous participation if they were asked to

participate again. The nature of having multiple attractions in this research is also another consideration when developing the stratified random sampling strategy.

At each attraction, respondents were randomly intercepted and were given the questionnaire later on (Xia, 2007). Locals of Phillip Island were excluded from the survey as they were assumed not to be tourists. On each day, the researcher moved from one attraction to another. In many cases, tourists' attendance at a certain attraction was assumed high during certain periods such as during lunch hour. During this period, the researcher had to attend an attraction that was known to have multiple selections of restaurants or cafes such as in Cowes.

Previous research has shown that tourists usually finish their tour in Penguin Parade (Xia et al., 2010; Xia et al., 2011). Therefore, a return box was placed there to assist tourists in returning their filled questionnaires. A return box was also placed at the Visitors Information Centre. Sampling was done during the summer break in February 2012 for eight days, from 9.00 am to 11.00 pm each day. This is the peak period when children have their school holidays and the number of tourists surges.

On each day when the sampling was concluded, data collectors skimmed through every returned questionnaire to ensure a sufficient range of data, i.e. itineraries and tourist characteristics were captured. A strategy was devised if there was an imbalance in the captured data. For instance, if there were insufficient data of older tourists, during the sampling session for the next day, they would gave more concentration for capturing data from older tourists.

3.7   Data Compilation and Database Design

A total of 1,536 sets of questionnaires were handed out to tourists at all attractions but priority was given to the four most popular attractions which are Visitors Centre, Churchill Island, Koala Centre and Penguin Parade. Attraction administrators estimate that more than 90% of tourists entering the national park visit these four attractions; it is therefore unlikely that significant movement patterns were missed by the survey (Xia et al., 2011).

Of the total number of questionnaires handed out, 1,310 questionnaires were returned with 1,104 questionnaires identified as valid responses in terms of tourists' characteristics and spatio-temporal movement data that were entered into a personalised tourist spatio-temporal movement database. The remaining 206 were incomplete and discarded. Table 3.3 details these figures according to each individual attraction where questionnaires were distributed.

Table 3.3   Returned and usable questionnaires according to each attraction

No.	Attraction	Returned Questionnaires	Valid Responses
1	Penguin Parade	509	413
2	Koala Centre	197	182
3	The Nobbies	273	230
4	Churchill Island Heritage Farm	160	144
5	Cape Woolamai	10	10
6	Rhyll Inlet	7	7
7	Cowes	81	51
8	Visitors Information Centre	73	67
9	Ventnor	0	0
	TOTAL	1,310	1,104

The personalised tourist spatio-temporal movement database comprises five database tables where one table concerns spatio-temporal movement while the remaining four are related to tourists' characteristics. Tourist characteristics data were entered into four separate database tables according to different types of tourist characteristics or segmentation variables, as described in the context of marketing: geographic, demographic, psychographic and behavioural. Yet again, these segmentation variables are explained further in Section 6.4 as part of MSM.

There are two miscellaneous yet important data not categorised as either spatio-temporal movement or characteristics data which are date/time and location where the questionnaires were distributed. Both are inserted into the spatio-temporal movement table. The entire database tables formed the personalised movement database where the term 'personalised' is defined as tourists' unique pattern of movement in regards to their specific characteristics.

The entire tables were defined as entities within a relational database as defined in Figure 3.9 depicting the Entity-Relationship (ER) diagram. Relations between these entities are established through a primary key that exists in all tables as columns or defined geospatially as attributes (Ramakrishnan and Gehrke, 2003).
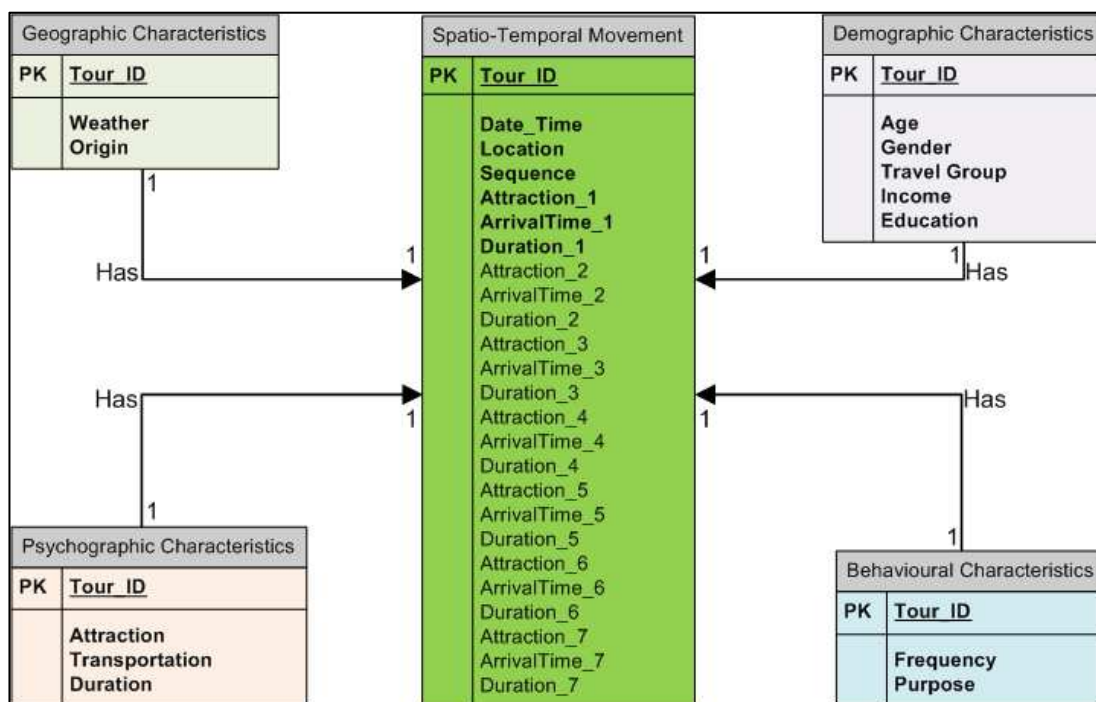
Figure 3.9   An ER diagram of tourist characteristics
and spatio-temporal movement database tables

Each entity has its specific instances and exists as rows or tuples within the table where these are the sets of records. In this case study, the primary key is defined as Tour_ID. A record in any of the four characteristics table with a particular Tour_ID must match a record with the same Tour_ID within the spatio-temporal movement table (Xia, 2007).

3.8   Software Implementation

Four types of software are used for this study. They are database and mapping, modelling, programming and statistical software. The tourists' spatio-temporal movement and characteristics dataset are stored as attribute database tables according to Figure 3.9 in the Esri ArcGIS 10.1 package. For the mapping, the geospatial data of Phillip Island, its attraction and road networks are stored as shapefiles in the same

package. Later on, the mapping tool making use of maps layout is exploited to design flow maps and tour package maps.

In terms of modelling, four software packages are used. The first one, Microsoft Excel 2012 is used to calculate the semi-Markov process related one-step transition probabilities and to construct the corresponding matrix. The second modelling software, MatLab R2010b is used to determine the PDF for calculating the one-step transitional arrival time probability with its corresponding transitional arrival time where later on, its corresponding matrices are built using Microsoft Excel 2012. SAS JMP 11.00 with its ability to incorporate the multimodality scenario of arrival times is used to estimate the PDF of arrival times and durations for APM.

The complete movement sequence and transitional arrival time probabilities are computed using built-in programming software within the Microsoft Excel environment called Visual Basic for Application (VBA). The source code for this purpose is attached in the Appendix. The last modelling software is Waikato Environment for Knowledge Analysis (Weka) 3.6.9 which is an open source software developed by The University of Waikato. It is used to conduct the EM algorithm within the MSM.

The final software is the statistical software for validating the outcomes of the PSM and the APM. The results of the one-step transition and one-step transitional arrival time probabilities are validated using a Chi-Square goodness of fit test within the IBM SPSS 20.0 statistical package. The results of the PDF for estimating arrival times and visiting durations in the APM are validated using a similar Chi-Square test.

3.9   Chapter Conclusions

This chapter introduces the research methodology that serves as an overall guide for this study. Initially, the requirements of tour package design are introduced where tourists' spatio-temporal movements and their characteristics are identified.

Subsequently, the general study approach is presented where the three methods are detailed namely the PSM, the APM and the MSM to solve three corresponding issues. Each method incorporates tourists' spatio-temporal movements and characteristics as the requirements of tour packages.

Later on, Phillip Island as the case study area is introduced, and the significance of its selection is discussed. The introduction of Phillip Island led to the discussion of the data collection procedure in detail. Phillip Island proved to be the ideal region to undertake the case study as it received a high number of tourists and contains several attractions.

Furthermore, approximately 95 % of the 1,104 tourists who provided a usable survey were found to be visiting either one or more of the attractions managed by the Phillip Island Nature Park authority such as Churchill Island, Koala Centre, The Nobbies and Penguin Parade. This is where the researcher spent the most time collecting the data.

Simultaneous with Xia's finding, most of the tourists (1004 tourists) visited Penguin Parade, and during the long queue to watch penguins they had an ample time to fill in the questionnaire (Xia, 2007). The concentration of tourists within these four attractions ensured the highest possible sample size obtained.

The location of the entire attractions that are restricted within the boundary of Phillip Island assisted in obtaining the highest number of sample size. The diversity of Phillip Island's tourists in terms of their characteristics and their spatio-temporal movement patterns created a multitude of results to be highlighted to all stakeholders.

Following the data collection, the data compilation and database design step was conducted. Finally, software implementation is described where four types of software that were used are revealed. The next chapter presents the first method of PSM to design tour packages.

CHAPTER 4: SCHEDULING TO PRODUCE TOUR PACKAGES

4.1 Introduction

In this chapter, the methodology of the Personalised Scheduling Method (PSM) is discussed. The methodology begins by detailing the categories of spatio-temporal movements and their components. Then, a PSM framework is proposed that contains elements such as input, process, output and tour package completion that can address the tour planning issue.

This chapter continues by presenting the concept of the semi-Markov Process that is utilised to solve the tourists' tour planning issue. Subsequently, the process of scheduling is discussed that paves the way for developing itineraries. Finally, some extra elements of the existing tour packages are listed, and that can be included within the itinerary output to complement the process of scheduling in order to fully produce tour packages.

4.2 Spatio-Temporal Movement Categories and their Components

For spatio-temporal tourist movement, the space component represents the location of the movement while the time component denotes the period at certain locations (Tussyadiah and Fesenmaier, 2007). Xia et al. (2011) detailed the temporal

movement so as to include the exact visiting durations and arrival times of tourists at certain tourist attractions. Tourists' spatio-temporal movement can be classified into scales of micro and macro-level movements (Xia and Arrowsmith, 2005).

The micro-level movement is based on a finer scale of location and time data whereas the macro-level movement involves a coarser scale. Table 4.1 compares the micro and macro-level movements in terms of tour package design.

Table 4.1   Comparison between the micro and macro-level movements in terms of itinerary design (Xia and Arrowsmith, 2005)

No.	Aspect	Micro-level movement	Macro-level movement
1	Definition	Movement within a tourist attraction	Movement between one attraction to another
2	Type of process	Continuous or discrete stochastic process	Discrete stochastic process
3	Granularity level	Finer	Coarser
4	Data level Spatial of detail	Individual coordinate (x,y) marking the location of tourist when moving	List of attractions
	Temporal	Up to minutes level of detail	Scheduling of movement between attractions

In terms of tourists' movement, a micro-level movement is a movement within a tourist attraction (Lau and McKercher, 2006). Consider a state space S, where in terms of tourist movement S represents a list of tourist attractions as the states. The tourists' movements at the micro-level can be presented by a continuous process $\{X_t, t \in T\}$ defined for a given possible geographic location arriving at t, where $T = \{t: 0 \leq t \leq +\infty\}$ (Stewart, 1994).

X_t is a set of spatial points (x, y) representing the sequence of locations of tourists undergoing movement for all time instances, t over a certain time interval, T (Murthy et al., 1990). Micro-level movement can be recorded spatially as a collection of coordinates in any coordinate system while temporally recorded to a finer detail of up to minutes to understand detailed movements (Batty, 2003).

Macro-level movement is a tourists' movement from one attraction to another, described as a discrete stochastic process $\{S = X_t, t = T\}$ where these attractions could be located some distance apart (up to thousands of kilometres) (Xia et al., 2011). X_t marks the attraction at time $T = \{0, 1, 2…\}$. For example, if there is a set of attractions visited by the tourists, then S will consist of those attractions with a similar set of discrete time instances for each attraction (Kemeny and Snell, 1960). Thus, a macro-level movement can be summarised as a process of sequencing tourists' movements from one attraction to another (Xia, 2007).

Designing a tour package involves working with both, micro and macro-level movement. A tour package involves suggesting multiple attractions to be visited within an accurate timeframe.

4.3   Other Elements for Completing Tour Packages

The inclusion of tourists' spatio-temporal movement components only highlighted the itinerary perspective of tour packages. Although these movement components are able to independently constitute a tour package, more elements are needed to expand a tour package so it is commercially feasible. Table 4.2 lists tour package elements from various contributions to the tourism literature.

Based on Table 4.2, a tour package can be summarised as generally having these elements: accommodation, transportation or routes, sightseeing, tour guides, meals, activities, budgets and optional tours. Much of the sightseeing elements which include attractions (defined as destinations and scenic spots in some of the literature) with its timing aspects such as arrival times and visiting durations are explained in Section 4.5.

Table 4.2   Tour package elements

No.	Author	Tour Package Elements
1	Askari (1971)	Escorted tours are all inclusive and include accommodation, transportation, sightseeing, baggage, customs and tour guides. Unescorted tour packages are more flexible where it includes all elements described in escorted tours except the tourist travels independently without joining the tour guide group.

2	Sheldon and Mak (1987)	Basic tour packages where only transportation, baggage handling and accommodation are included. Inclusive tour packages that includes all the basic elements plus some sightseeing and events at the attractions. All inclusive tours where meals and tour guides are added to the inclusive tours.
3	Dellaert et al. (1995)	Activities that will determine attractions and its timing.
4	Enoch (1996)	Transportation, accommodation, sightseeing (attractions and their visiting durations), budget.
5	Fesenmaier and JiannMin (2000)	Destinations, members of the travel group, timing (date of travel and length of trip), transportation mode, route, budget and activities.
6	Wang et al. (2000)	Pre-tour briefing, transportation (plane and coach), accommodation, meals, scenic-spot, shopping activities, optional tour, miscellaneous (tips, medical care and punctuality).
7	Wang et al. (2007)	Accommodation, transportation, shopping arrangement, optional tour, tourist guide.
8	ACS Distance Education (2011)	Transportation, accommodation, meals, the provision of a tour guide, transfers, cruises, rental cars, entrance fee to attractions, insurance and ticket fee for events.
9	Wong and McKercher (2012)	A basic package – transportation and accommodation. All inclusive packages – transport, accommodation, meals, sightseeing and the provision of a tour guide.

Three specifications are considered when deciding which elements are the most important elements to be included in tour packages as stated below. Section 4.6 provides an insight on the process to select the elements that are relevant to this research.

(i)   The elements must conform to the nature of tour packages defined in this research which is a day tour package covering a national park for tourists travelling independently;

(ii)   The availability of the elements within the aspect of geographical region and time;

(iii)   The accessibility to the data in regards to the element

4.4   The Personalised Scheduling Method (PSM) Framework

The PSM framework is the extension of the PSM's section of Figure 3.3. The framework consists of four elements: input, process, output and tour package completion is shown as Figure 4.1. The input element of Figure 4.1 represents the methodological requirement, i.e. data for the method, shown in Figure 3.3 within the top part of the PSM's implementation section. In terms of the data, focus is given to the transitional tourists' spatio-temporal movement dataset at the initial stage, the spatio-temporal movement dataset is specified according to particular combinations of tourist characteristics.

This enables the personalised spatio-temporal movement pattern to be observed as in Section 7.2 and its respective sub-sections, so the development of a personalised itinerary would be possible. The spatio-temporal movement dataset of tourists consists of their movement sequence from one location (attraction) to another with its particular arrival time and visiting duration spent at each attraction.

The process element of Figure 4.1 entails two processes which are the semi-Markov and scheduling processes that are highlighted in Figure 3.3, shown within the PSM's description and implementation sections. These processes are meant to transform the spatio-temporal movement dataset into an itinerary which is the output. Two kinds of probabilities are produced from the semi-Markov process which is the spatial movement transitions, i.e. movement sequence and transitional arrival time probabilities where later on, the scheduling process is used to combine both of these probabilities.

The cube and age with gender axes represents the output of the itinerary that is highlighted within the PSM's implementation and evaluation section of Figure 3.3. The cube details the itinerary consisting of spatial movement, arrival time and later on, visiting duration at each attraction and the axes define the itinerary as specific to particular tourists' characteristics. Finally the tour package completion element shows the addition of extra tour package elements such as routes, activities and budget into the produced itinerary for completing the process of designing personalised tour packages.
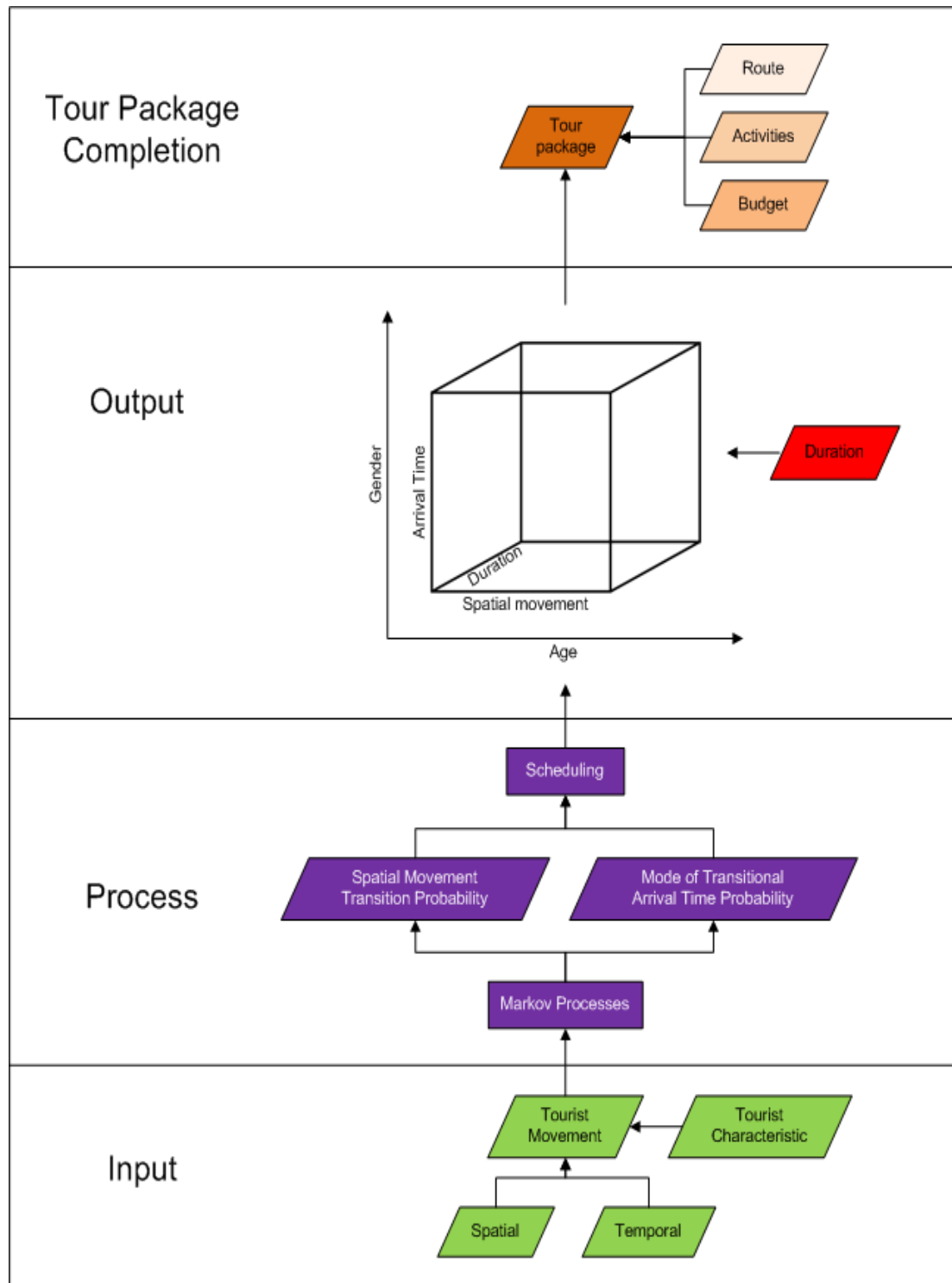
Figure 4.1: The Personalised Scheduling Method framework

4.5   Personalised Scheduling Methodology

Figure 4.2 emphasises the Process element of Figure 4.1 to design tour packages. Figure 4.2 starts by specifying the input element of Figure 4.1 to include movement sequence, arrival time and visiting duration. Then, the process element of Figure 4.1
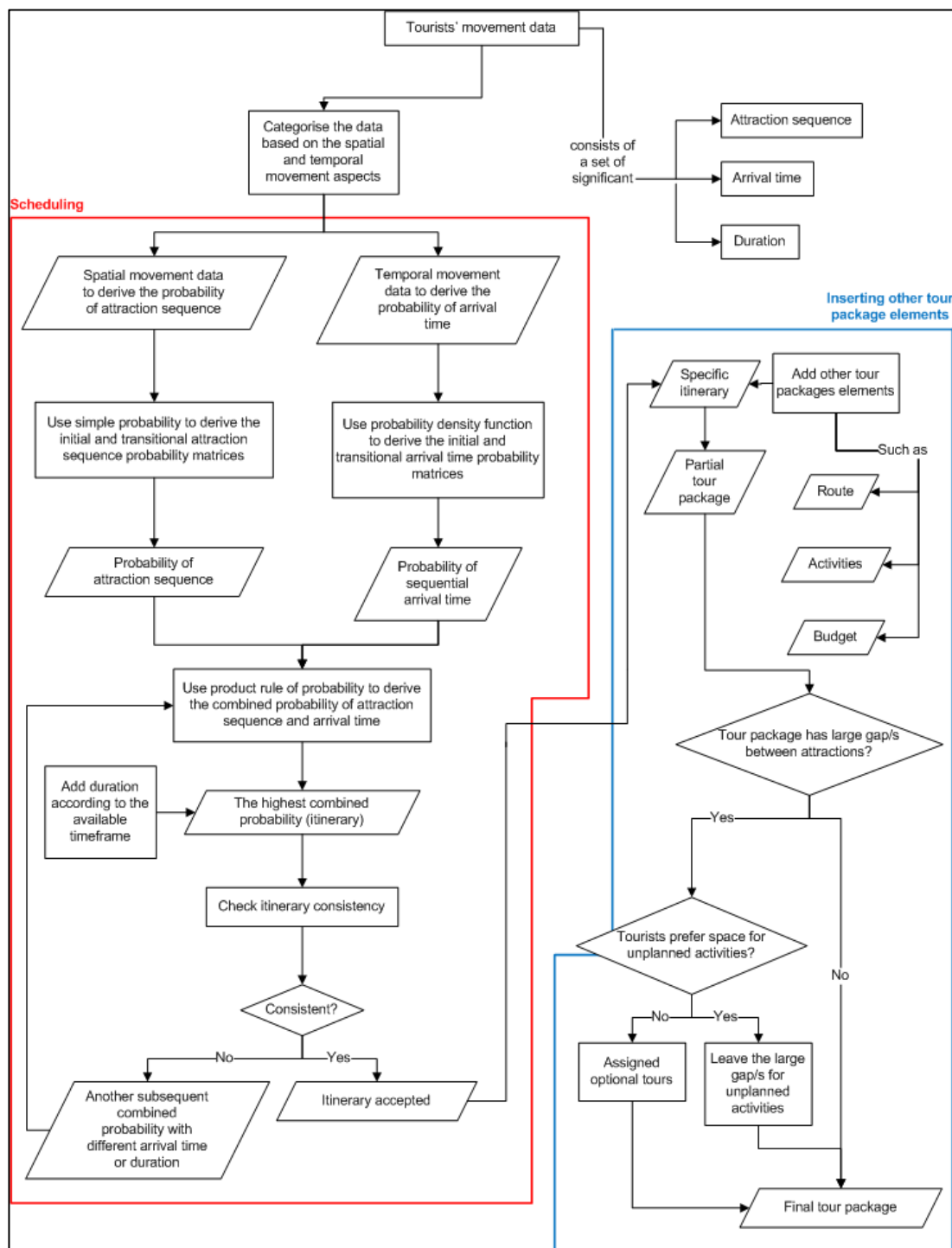
Figure 4.2   The methodology for designing tour packages

is described to consist of two steps where initially, itinerary elements are constructed, shown in the red section. Sections 4.5.1 and 4.5.2 explain this step that consists of two further steps which are the semi-Markov and scheduling processes.

A semi-Markov process is used to model tourist movement which is a stochastic spatial and temporal process. The product rule of probability is later used to merge the spatial movement transition probability with the mode of arrival time transition probability. The usage of mode is described further in Section 4.5.1. Note that, the bottom part of the red section of Figure 4.2 where the itinerary's consistency is checked represents the PSM's evaluation section of Figure 3.3. This evaluation step is detailed further in Section 4.5.2.

Later on, other elements are included in the itinerary elements for completing the tour package design effort, shown in the blue section. This is also shown as the tour package completion element in Figure 4.1 where it is further detailed in Section 4.6.

4.5.1   Semi-Markov Process

This chapter modelled tourist movements, a random spatial and temporal process, as a semi-Markov process. Let $\{X_t, t \in T\}$ be a family of random variables i.e. a stochastic process, which is indexed by time t belonging to the set $T = \{0,1,2,...,\}$ (discrete time) or $T = (0, \infty)$ (continuous time).   The random variable $X_t \in S$ where S is the state space consisting of the states (attractions) visited by a tourist during his or her excursion in a nature reserve, e.g. Phillip Island. In this exposition, it is assumed that t, as an element of time is continuous but $S = \{0,1,2,...,m\}$ is finite.

The process X_t can be modelled in various ways, but the most comprehensive approach which captures both the state visited and the time between consecutive visits is through the discrete Markov and semi–Markov process described below (Xia et al., 2011).

Definition 1 *Let X_o, X_1, X_2, . . . be the successive states visited by a tourist starting in state X_0 at times $0 = T_0 < T_1 < T_2$, . . . respectively. The stochastic process $\{X_t, t \in T\}$ is a semi–Markov process if it satisfies the following fundamental property:*

$$\Pr(X_{n+1} = j, T_n, T_{n+1} - T_n \le t) \mid X_0, X_1, \ldots, X_n; T_0, T_1, \ldots, T_n)$$
$$= \Pr(X_{n+1} = j, T_n, T_{n+1} - T_n \le t \mid X_n). \tag{4.1}$$

The interpretation of (4.1) is this: given that the current attraction that a tourist visited and the times when these visits occurred up to time T_n (arrival time at attraction *i*) are known then if the temporal movement is to be a semi–Markov process, the probability that the next visit at time T_{n+1} (arrival time at attraction *j*) is to state *j* and the time between the transition from the current visit to the next is within time *t*, depends only on the current state visited at time T_n, i.e. X_n. The conditional probability:

$$\Pr(X_{n+1} = j, T_n, T_{n+1} - T_n \le t \mid X_n = i) = Q_{ij}(t). \tag{4.2}$$

is called the semi–Markov kernel. The derivative of $Q_{ij}(t)$ with respect to *t* gives the Probability Density Function (PDF) $q_{ij}(t)$ which has the following probabilistic interpretation:

$$q_{ij}(t)dt \approx \Pr(X_{n+1} = j, T_n, T_{n+1} - T_n \in (t, t + dt) \mid X_n = i). \tag{4.3}$$

Various distributions and processes such as below are intrinsically related to the semi–Markov process, thereby making it a very powerful tool for modelling spatial and temporal movement of tourists.

(i) $$Q_{ij}(\infty) = \Pr(X_{n+1} = j \mid X_n = i) = P_{ij}. \tag{4.4}$$

Therefore, associated with the semi–Markov process, transition probabilities P_{ij} (or defined as spatial movement transition probability in the framework) , $i, j \in S$ is an underlying Markov process with a condition of arrival time of $T_{n+1} - T_n \leq t$.

(ii) $\quad F_{ij}(t) = \Pr(T_{n+1}^* - T_n^* \leq t \mid X_{n+1} = j, X_n = i)$ (4.5)

$F_{ij}(t)$ means the probability of tourists' stay at current attraction i with arrival time, t before moving to next attraction j.

(iii) $\quad \displaystyle\max_{-\infty < T < \infty} f_i(T) = f_i(T_i^*).$ (4.6)

where T_i^* is the mode of transitional arrival time, derived from using PDF at attraction i. For commercial reasons in terms of generating more profit for tour packages, $f_i(T_i^*)$ is evaluated at the mode which means the highest occurrence, hence the most popular arrival times that tourists choose. The most suitable mode of transitional arrival time is determined by deriving its probability, this time by implementing PDF.

(iv) Finally, this is the conditional probability with a transition from attraction i to j, the corresponding mode of arrival time (or defined as a mode of arrival time transition probability). Using the elementary probability argument, it follows that:

$$Q_{ij} = F_{ij}(t)P_{ij}.$$ (4.7)

provided that $P_{ij} > 0$. Note that if $P_{ij} = 0$, then the transition from i to j is impossible and so $T_{n+1} = T_n$ with probability 1.

(v) $\quad \Pr(T_1 - T_0 \leq t_1, T_2 - T_1 \leq t_2, ..., T_{n+1} - T_n \leq t_n \mid X_0 = x_0, X_1 = x_1, ...,$
$X_n = x_n) = F_{x_0 x_1}(t_1) F_{x_1 x_2}(t_2)...F_{x_{n-1} x_n}(t_n).$ (4.8)

This equation indicates that the times between transitions are conditionally independent given the knowledge of the states visited. Therefore, if the state space S consists of a single state, then the semi–Markov process is essentially a renewal process.

4.5.1.1 Model Validation

After obtaining two kinds of probabilities: spatial movement transition and mode of arrival time transition, the concern is, how well are these probabilities modelled using this methodology? That is to say, how compatible are the expected frequencies in regards to the probability for tourist spatio-temporal movement patterns (the calculated statistical value) with the observed frequencies, assumed to be representative of the entire tourist population.

Chi-Square Goodness of Fit Test, $\chi_{EF}^{2(70\%)}$ through Equation (4.9) is selected for validating the spatial movement transition probability model (Moore et al., 2012). For implementing this test, the entire sample is divided into training and

$$\chi_{EF}^{2(70\%)} = \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(O_{ij} - N_i P_{ij})^2}{N_i P_{ij}} \qquad (4.9)$$

testing datasets (Xia, 2007). The training dataset is used for feeding the model of the semi-Markov process where later on the expected frequencies can be obtained while the testing dataset is used to obtain the observed frequencies.

A set of observed frequencies, O_{ij} is compared with a set of expected frequencies, E_{ij} where ij corresponds to various spatial movement transition figures from one attraction to another using the Chi-Square Goodness of Fit test. The expected frequencies are obtained by multiplying any specific spatial movement transition probability, P_{ij} with N_i, the frequency figure of tourists who made the transitions starting from similar i attraction to all other attractions. The final output of

comparing these observed and expected frequencies produces the calculated statistical value which is the Chi-Square statistic, χ^2.

This Chi-Square statistic is then compared with the P-value, derived from the degrees of freedom, df and the confidence level, α. If the χ^2 value is outside the P-value, it can be concluded that the observed frequencies are compatible with expected frequencies, i.e. the model provides a good fit of the data (Moore et al., 2012).

An Akaike Information Criterion (AIC) is used to validate the PDF output for the mode of arrival time transitional probability model. Despite the PDF being used to computing the mode of arrival time, it is not the focus of PSM; readers are referred to Section 5.5 for the details of AIC.

4.5.2 Scheduling Tourist Movements

Consider the situation of modelling tourists' movement in Phillip Island consisting of m attractions. It is assumed here that each attraction is to be visited exactly once by a tourist and all attractions are covered. Hence there is a total of $m! = (m (m - 1)(m - 2)... 3.2.1)$ possible schedules that each tourist can take. The set of schedules is denoted by $P = \{s_1, s_2, \ldots, s_m!\}$.

The objective here is to select the optimal schedule(s) based on a maximisation of attraction utility. It is assumed that tourist movement follows a semi-Markov process defined in the Section 4.5.1 and then the following steps ensue:

(i) At time $T_o = 0$, the tourist starts her excursion at the state $X_o = 0$, i.e. from a location O outside of the nature reserve.

(ii) The approximate probability of the tourist following the schedule $S_i = t_1^{*(i)}, X_1^{(i)}; t_2^{*(i)} X_2^{(i)}; t_3^{*(i)} X_3^{(i)}; \ldots; t_m^{*(i)} X_m^{(i)}$, i.e. entering state $X_j^{(i)}$ at arrival time $t_j^{(i)}, j = 1, 2,\ldots, m,$ is given by

$$q_i = Q_{OX_1^{(i)}}(t_1^{*(i)}) Q_{X_1^{(i)}X_2^{(i)}}(t_2^{*(i)} - t_1^{*(i)}) Q_{X_2^{(i)}X_3^{(i)}}(t_3^{*(i)} - t_2^{*(i)}) ...$$
$$Q_{X_{m-1}^{(i)}X_m^{(i)}}(t_m^{*(i)} - t_{m-1}^{*(i)}) \tag{4.10}$$

Using (4.7), it can also be written as,

$$q_i = f_{OX_1^{(i)}}(t_1^{*(i)}) f_{X_1^{(i)}X_2^{(i)}}(t_2^{(*i)} - t_1^{*(i)}) f_{X_2^{(i)}X_3^{(i)}}(t_3^{*(i)} - t_2^{*(i)}) ... f$$
$$_{X_{m-1}^{(i)}X_m^{(i)}}(t_m^{*(i)} - t_{m-1}^{*(i)}) \times \prod_{j=1}^{m} P_{X_{j-1}^{(i)}X_j^{(i)}} \tag{4.11}$$

(iii)   The best scheduled movement or the most attractive combination of attractions is the maximisation of q_i, where using the product rule of probability, this combination of attractions includes the most frequently visited attractions and the most common, i.e. most popular arrival time at these attractions. The pattern with highest probability (F) is considered as a recommended tour package for tourists. Visiting duration is arbitrarily assigned based on the time frame between two arrival times (former and latter attractions).

$$F = \max(q_i) \tag{4.12}$$

(iv)   The final step is to check the consistency of the combination of attractions and arrival times. This signifies the PSM's evaluation section of Figure 3.3. For example, the most popular arrival time at attraction one estimated from the model could be later than the most popular arrival time at attraction two estimated from the model. If this happens, the next most popular arrival time is identified. This iterative process continues until the consistency emerges.

4.6   Selecting Other Elements for Tour Package Completion

The listed tour package elements in Section 4.3 are filtered according to the specifications also listed in the same section. The accommodation element is unrelated to this research as the itinerary is only designed for a one day tour package, thus accommodation is unnecessary.

The transportation element marks the public transportation for tourists to commute from one attraction to another. In terms of transportation element, it is notable that a scheduled bus service exists in the case study area (Australian Rail Maps, 2011). Despite having a sufficient service frequency of four to six per day, one disadvantage is that the bus route for Phillip Island only covers four out of nine attractions included in this research, thus making its selection impossible. A tour guide element is not needed since the tour package is prioritised for tourists who travel independently.

Routes from one attraction to another are important to ensure tourists arrive at their destination. The selection of route element is possible given that Phillip Island has a well defined road system that covers the entire attractions included in this research (Phillip Island Nature Park, 2011). The concept of shortest path in determining the route is implemented. Although argued by Abdul Khanan et al. (2012) that the shortest path route is not the reality in tourism, the fact that the case study is conducted in a national park guarantees that most of the routes are scenic, thus are enjoyable for tourists.

Tourists engage in activities when they arrive at an attraction. Such examples are an animal feeding session in a zoo, guided walks in a museum and having meals in town. An important aspect of activities arrangement is the timing aspect where activities should start after the arrival of tourists at any attraction. An activity or a list of activities must fit into the timeframe given. The inclusion of activities element is possible since each attraction in Phillip Island has specific sets of activities shown in Table 3.1.

The budget element of tour packages is included in this research due to its marketing importance when dealing with a large number of tourists to Phillip Island (O'Reilly, 2006). With only the money element involved (Goodwin, 1981), the addition of tourists' budget is only limited to the entrance fee of attractions or any activity fees within attractions that are not included in the entrance fee. Note that, the price of meals as a dining activity is not able to be included as tourists only decide on their meals once they arrive at the restaurant. Optional tours can be included if there is a large time gap between attractions. Any of the nine attractions can be included if they are not included in the existing itinerary.

Finally, four compulsory elements of tour packages are included which are route, activities, budget and optional tours. These four elements are inserted after the itinerary is derived. Attention should be given so that the insertion of these elements will not affect the itinerary's spatio-temporal movement structure. For instance: (1) the insertion of the journey duration might affects the arrival time at any attraction, and (2) the overall duration of activities within an attraction is longer than the visiting duration at the respective attraction.

4.7 Chapter Conclusions

This chapter described a method to schedule tourists' spatio-temporal movement towards designing tour packages. The spatio-temporal movement of tourists at the macro level that entails their movement sequence together with the arrival time and visiting duration spent at each attraction are utilised within the semi-Markov and scheduling processes to produce the itinerary of tour packages.

At the end of this chapter, extra elements of tour packages such as routes, activities, budget and optional tours are added for completing the design of tour packages to guarantee its commercial feasibility. The inclusion of these extra elements is ensured not to affect the existing spatio-temporal movement structure of itineraries.

CHAPTER 5: ACTIVITY PLANNING WITHIN ATTRACTIONS

5.1 Introduction

This chapter introduces the Activity Planning Method (APM) to plan the timing of tourist activities in various attractions. The activity timing is proposed based on tourist arrival times at each attraction. The chapter commences by detailing the factors influencing the starting times of activities and the factors influencing the tourist arrival times at attractions.

This chapter then progresses by introducing the framework of APM that can address the activity planning issue. It contains four elements which are input, process, output and analysis. The output element symbolises the suggestion of arrival times and visiting durations to plan the timing of activities. The analysis element is divided into two: an Activity Timing Comparison Analysis (ATCA) for comparing the suggestion of the timing of activities with the actual timing of activities and an Activity Frequency Comparison Analysis (AFCA) to compare the suggested frequency of activities with the actual frequency of activities.

This chapter continues by ascertaining the concept of Probability Density Function (PDF) as the process element of the APM framework. To conclude, some examples of the use of the PDF including multimodal PDF is given within the context of timing estimation of activities.

5.2 Factors Influencing the Activity Starting Times and the Arrival Times at Each Tourist Attraction

From a tour planning perspective, it is more suitable to conduct an activity when tourists arrive (O'Connor et al., 2005). This is to ensure more participation in activities by considering the peak hour of tourists' arrival. Tourists' arrival times and activity starting times can interact with each other. Any particular activity such as penguin viewing can motivate tourists to visit an attraction (for instance, Penguin Parade), thus increasing tourists' visitation rate (Leiper, 1990).

This can be seen in Figure 5.1 where the visitation rates at major attractions in Phillip Island are listed for the financial year 2012 – 2013. Notice that, the visitation rate at Penguin Parade is the highest due to the activity of penguin viewing (Tapper, 2006). Concurrently, a higher visitation rate leads to more activities conducted at any attraction.

Tourists' arrival times depend on the starting times of activities. In general, there are two factors that can influence the decision of the starting time of activity: the attraction type and the number of tourists. Apart from classifying tourist attractions as managed and unmanaged attractions as in Section 3.6, tourist attractions can also be divided into cultural attractions and natural attractions (Leisen, 2001). A cultural attraction is a man-made attraction such as a museum and art gallery while a natural attraction refers to geographical features that attract tourists to see it for example, a mountain or an animal landing site.

Theoretically, a cultural attraction can devise its activity such as a guided tour in a museum to start at anytime that they want. By doing this, attraction administrators can plan the starting times of multiple activities to avoid congestion at the main
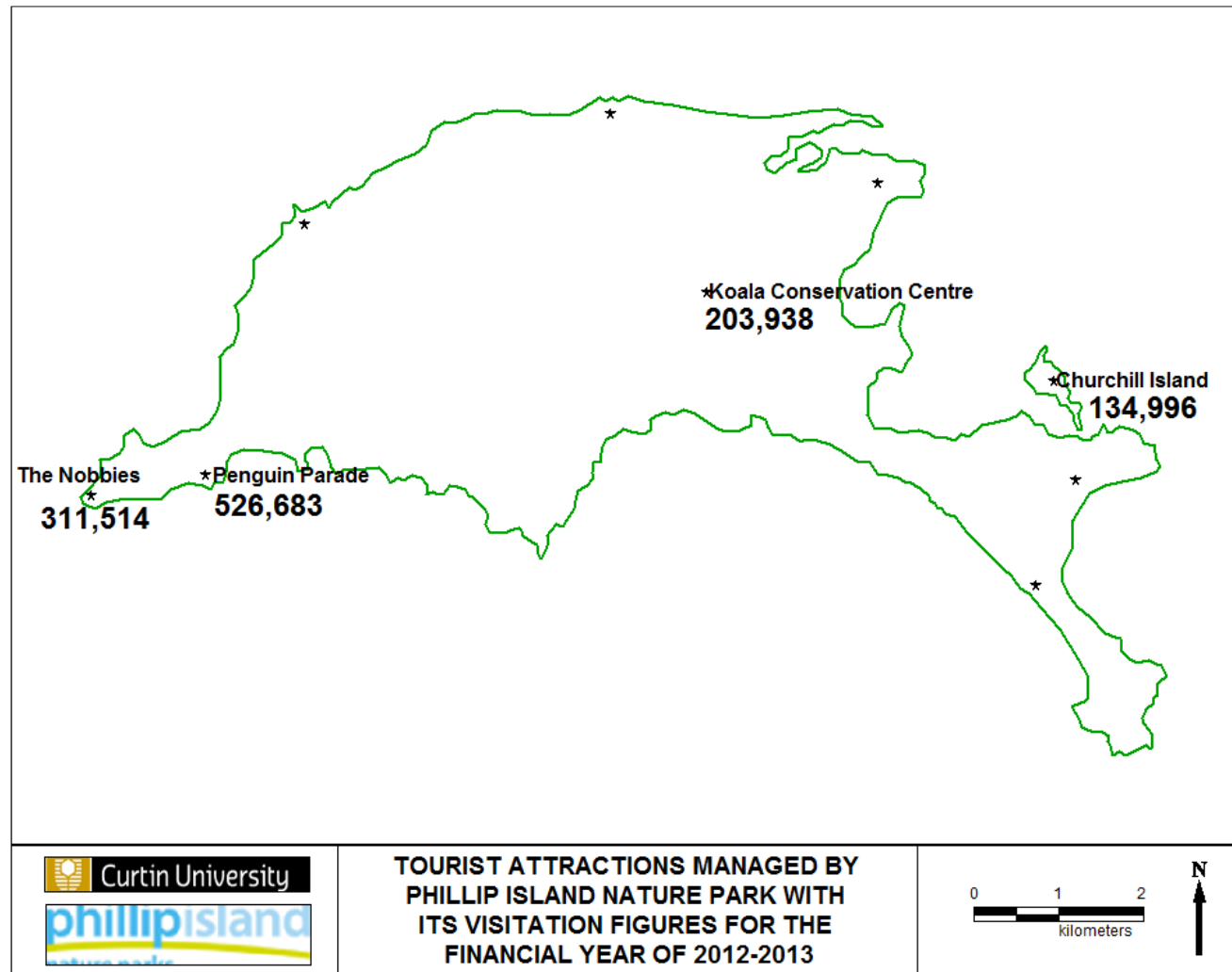
Figure 5.1   Tourist attractions of Phillip Island with its visitation figures (Phillip Island Nature Park, 2013)

entrance (Wanhill, 1980). Suggesting multiple starting times of activities covers the multimodality scenario that should be considered in APM.

Attraction administrators should also maximise the number of tourists participating in any activity. For this, they should consider the number of expected tourists when arranging the starting times of activities so that a reasonable number of tourists turn up for the activities. This can be done by estimating the possible arrival times of tourists when the mode of arrival times of tourists is anticipated. Similar to PSM, the mode of arrival time corresponds to the highest frequency of tourists arriving at any particular arrival time.

The starting times of activities can depend on the tourist arrival times. One factor which is the activity timing can affect the arrival time of tourists. Contrary to a cultural attraction, a natural attraction such as Penguin Parade in Phillip Island which includes a penguin landing activity is not able to devise the landing time of penguins as it is a natural process. As such, a park manager is left with no other option than to set the penguin landing activity at 9 pm during summer to correspond with the natural behaviour of penguins to arrive at that specific time period (Xia, 2007).

Similarly, tourists have to arrive prior to 9 pm to participate in the penguin landing activity, shown in Figure 5.2. This concentration of arrival time created another scenario of unimodality of PDF where logically, only one arrival time is suggested. However, for easing the congestion at attractions, several arrival times should also be considered.

As a conclusion, it was found that the starting times of activities depend on the type of tourist attraction. This raises the multimodality scenario of PDF. The starting times of activities also depends on the number of tourists arriving at any specific time which makes it necessary to identify the mode of arrival time of tourists. The timing of activities can influence tourists' arrival times at each attraction. This raises the unimodality scenario of PDF. The scenario of multimodality, mode and unimodality of arrival time should be incorporated when implementing the PDF.

Figure 5.2   Penguin landing and boardwalking activities at Penguin Parade and The Nobbies, both a natural attraction. During peak time, congestion happened at both attractions (Dunstall et al., 2004; Phillip Island Visitor Information Centre, 2013)

5.3   The Activity Planning Method (APM) Framework

The APM framework is the extension of the APM's section referred to in Figure 3.3. The framework consists of four elements: input, process, output and analysis as depicted in Figure 5.3. The input element represents the methodological requirement, i.e. data for the method, shown in Figure 3.3 within the APM's implementation section.

Contrary to the input element of Figure 4.1 in Section 4.4, focus is given only to tourist temporal movement data in regards to tourists' arrival time and visiting duration at each attraction. Tourists' arrival time and visiting duration at each attraction are combined to form the activity timing planning dataset.

The Process element incorporates the PDF to transform the activity timing planning dataset into usable information. This is highlighted in Figure 3.3, shown within the APM's description and implementation sections. The unimodality, multimodality and mode of arrival times of the PDF recommended in Section 5.2 are implemented. Outputs from the PDF process include modes and several other arrival times and visiting durations are shown within the APM's description section of Figure 3.3. This suggestion of arrival times and durations are combined to form a set of timing of activities and these are shown within the APM's implementation section of Figure 3.3.
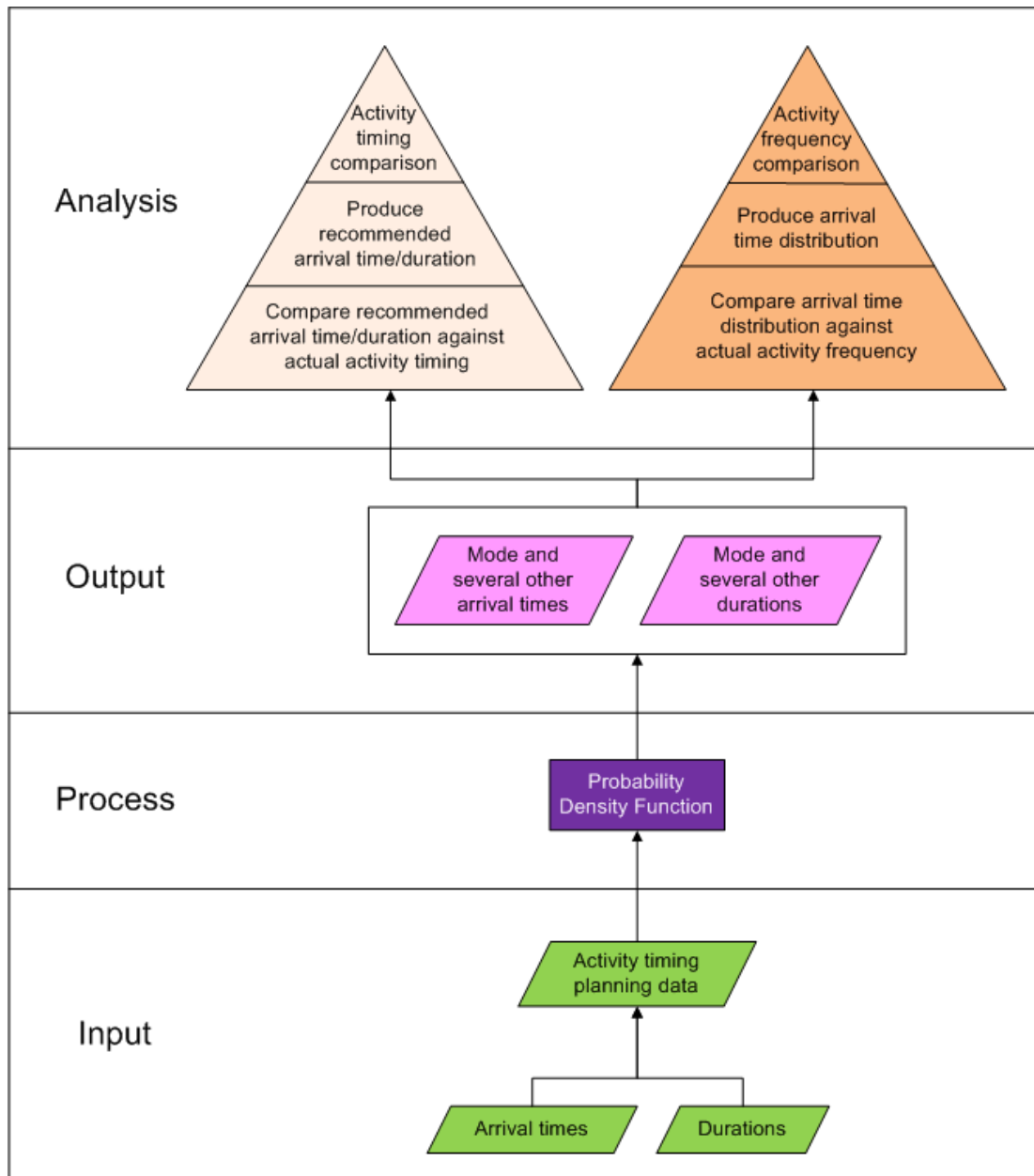
Figure 5.3   The Activity Planning Method framework

The most important element of the framework is the analysis element, shown as the evaluation section of Figure 3.3 where two types of analyses are each depicted as triple stage triangles. The top layer of the triangle shows the name of the analysis. The middle layer marks the product of the analysis, each derived from the output element of mode arrival times and visiting durations. The final layer is the bottom layer that portrays the steps taken to accomplish the analysis.

The most prominent analysis which is the ATCA has the same product as the first analysis. Nevertheless, it evaluates the estimated arrival times and visiting durations against the actual activity times and visiting durations already scheduled by the attraction administrators. If the comparison shows the estimated and actual arrival times and visiting durations are distinctive, attraction administrators have the option to utilise the estimated arrival times and visiting durations that are derived from actual tourists' activity timing dataset.

To conclude, the AFCA produces an arrival time distribution that highlights the frequency of estimated arrival times to be proposed to attraction administrators. This is evaluated against the frequency of current activities. Attraction administrators can use the estimated arrival time distribution to decide how frequent to organise any particular activity if the evaluation identifies dissimilarities between the estimation of the frequency of activities against the status quo. Later on, the estimated arrival time distribution is compared against the actual activity frequency within their attraction.

The AFCA is further extended by introducing arrival time and visiting duration distribution maps that show two maps, one each to describe the distribution of arrival times and the distribution of visiting durations for all attractions.

5.4   Probability Density Function

Continuous random variables such as arrival times and visiting durations are a set of possible values that are uncountable (Ott and Longnecker, 2010). As a continuous random variable X typically involves measurement where there are infinite values between any two intervals, the probability should be 0 that X equals any particular exact value x (Weiss et al., 2006):

$$P(X = x) = 0 \qquad \text{for all } x \in S \tag{5.1}$$

A continuous distribution can be characterised by its PDF of a random variable X over a set (usually an interval) to compute the probability that X takes a value in that set (DeGroot and Schervish, 2002). Let X be a continuous random variable; a non-

negative function f is said to be a PDF of X, for all real numbers $a < b$ (refer Figure 5.2 below):

$$P(a \leq X \leq b) = \int_a^b f(x)dx \qquad (5.2)$$

That is the probability, f that X (refer Figure 5.4) takes on a value in the interval (a,b) is the area in this interval and under the graph of the density function (Zhang, 2008).
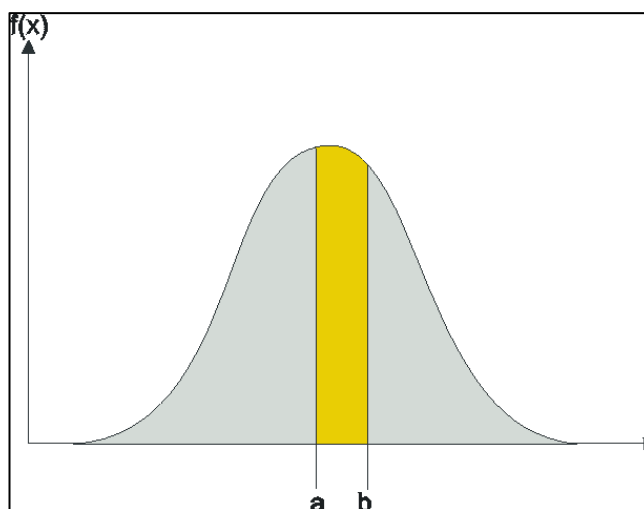
Figure 5.4   An example of a PDF (Reliasoft, 2002)

Notice that equation 5.1 is reflected in Figure 5.4 where the probability that the function is exactly equal to a specific value is zero, since a specific value has no area under the curve (Wikidot, 2011).

Every PDF must satisfy the following two conditions:

$$f(x) \geq 0, \text{ for all } x \qquad (5.3)$$

That is the PDF, $f(x)$ must be positive for all values of x, and:

$$\int_a^b f(x)dx = 1 \tag{5.4}$$

Where equation (5.4) describes that, since the PDF represents the entire sample space, the area under the PDF must equal to one.

5.4.1 Some Important Statistical Distributions

There are various existing distributions with different scale, shape and location parameters that ultimately determine the shape of the curve (refer Table 5.1).

Table 5.1 Important statistical distributions

No.	Distribution	Scale Parameter	Shape Parameter	Location Parameter
1	Log-Normal	μ	σ	γ (γ=0 for two parameter distribution
2	Gamma	θ	κ	None
3	Weibull	α	β	γ (γ=0 for two parameter distribution
4	Normal (Gaussian)	σ^2	None	μ

Apart from the parameters, the data form can also contribute to the different shape of the curve either it is monotonically increasing or decreasing, logistic and constant over time (Mohammadian and Doherty, 2006). In the case of the dataset of arrival times and visiting durations, each set of arrival time and visiting duration can be fitted into different distributions.

5.5 Estimation of Arrival Time and Visiting Duration Using Activity Timing Dataset

The process element of Figure 5.3 is summarised in Figure 5.5. A PDF is applied to fit several arrival time datasets shown as the Input element in Figure 5.3 using various statistical distributions such as LogNormal, Generalised Extreme Value, Weibull and Normal (Gaussian). A PDF of arrival time, T is a function, $f(t)$ such that for any two arrival times, a and b with $a \le b$,
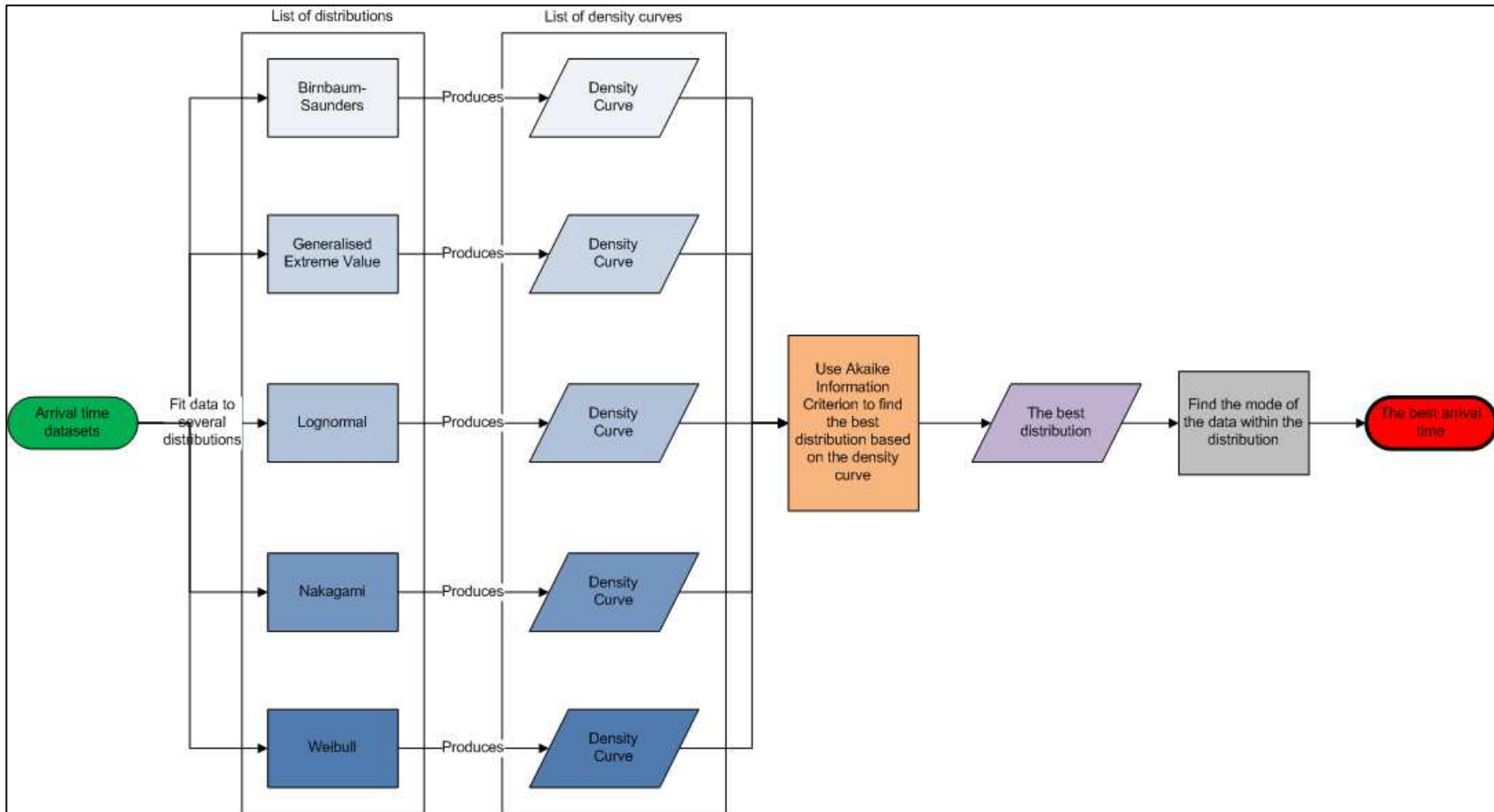
Figure 5.5   PDF process

$$P(a \leq T \leq b) = \int_{a}^{b} f(t)dt \qquad (5.5)$$

The explanation of (5.5) is the same as (5.2). One important phenomenon should be observed about the multimodality of the PDF. Unlike the unimodal PDF that only has one mode, the multimodal PDF happens when there are two or more modes when a distribution is fitted to the dataset (Brassard and Correia, 1977). This is due to the mixture of two or more different unimodal distributions (Pfingsthorn and Birk, 2013). Figure 5.6 displays the multimodal PDF. X-axis represents the figure of the variable while y-axis represents the probability density figure. It can be observed from Figure 5.6 that more than one peak or multiple modes exist, shown by the multiple arrows.
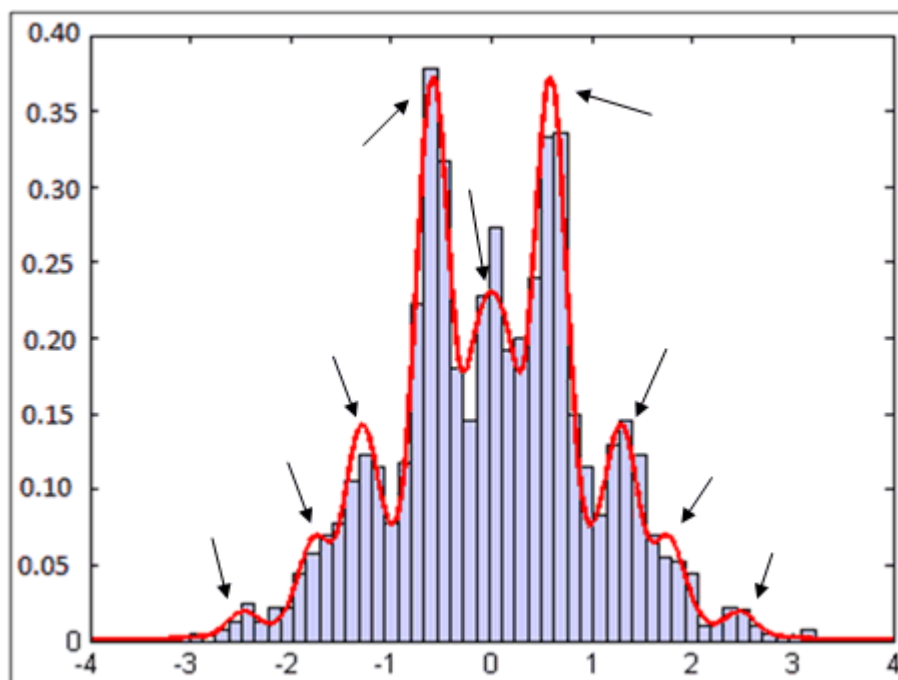
Figure 5.6   A Multimodal PDF (Mathworks, 2013)

This multimodal phenomenon can happen when the dataset contains a complex variation in terms of its arrival time and duration instances. In terms of this research, the occurrence of multimodality can bring benefits in the estimation of arrival times and visiting durations as it can suggest more than one arrival time and duration. The challenge now is to determine which distribution can best fit the arrival time datasets.

To achieve this, the Akaike Information Criterion (AIC) is implemented where it can validate the model, i.e. by measuring the relative goodness of fit of a statistical model (Awad, 1996).

The selection of AIC is not solely on its ability to ascertain the goodness of fit (shown as L in equation 5.6), but as it penalises the increasing number of parameters used, shown as k in Equation 5.6 (Bozdogan, 2000). This penalty discourages overfitting where reducing the number of parameters improves the value of AIC. At the end of the distribution selection process, the model with the least value of AIC is selected.

$$AIC \quad = \quad -2 \log L + 2k \tag{5.6}$$

Figure 5.7 depicts the comparison of two fitted distributions, i.e. probability density curves, one unimodal and the other one multimodal to the arrival times dataset. The x-axis represents the arrival time while the y-axis represents the probability density figure shown as the curved line. The green curve line represents a unimodal distribution while the red curve line represents a multimodal distribution. Similarly as in the previous chapter, some datasets do not provide a significant distribution in the fitting process. In this case, a simple probability is used to substitute the PDF.
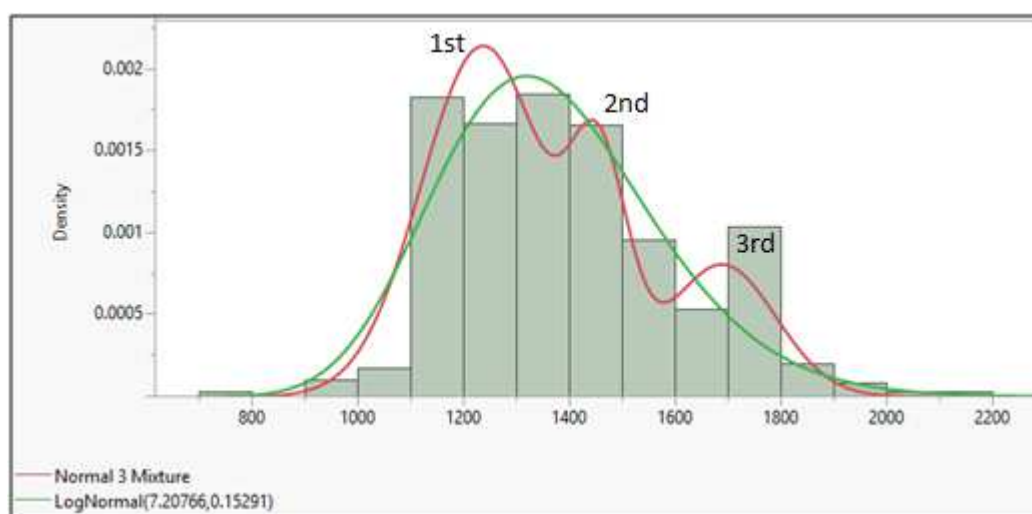
Figure 5.7   A Histogram showing two distributions

fitted to the arrival times dataset

The task now is to identify the best arrival times to suggest for attraction administrators. Activities can be conducted more than once in a day. Therefore, this method is able to propose more than one arrival time. This is done by carefully observing the multimodal distribution of Figure 5.7. The best arrival time corresponds to the curve area with the highest density, marked as 1st, therefore producing the highest probability. Notice that the mode, i.e. the highest frequency of arrival time is at 12.25 pm.

Other suitable arrival times correspond to the other peaks from Figure 5.7 that marks the other modes. These are marked as 2nd and 3rd peak in the red curve area and are observed as 2.20 pm and 5.00 pm respectively. These suggestions of mode and several other arrival times correspond to the Output element of Figure 5.3. The Chi-Square Goodness of Fit test is finally used to validate the model with the same procedure shown in Section 4.5.1.1. The whole procedure in this section is then applied to the visiting duration dataset.

The Output element that represents the suggestions of mode and several other arrival times and durations are brought forward in the Analysis element of Figure 5.3, i.e. the APM's evaluation section of Figure 3.3. The product of this analysis is presented in Sections 7.3.2.1 and 7.3.2.2.

5.6   Chapter Conclusions

This chapter presented a method to plan the timing of activities that includes arrival times and visiting durations. Initially, the factors influencing activity starting time and the arrival time at each attraction are identified. The factors are type of tourist attraction, the number of tourists arriving at any particular time and timing of activities. The identification of these factors highlighted the importance of including unimodality, mode and multimodality within the context of PDF.

The fundamental concept of PDF is presented where multiple probability density curves are obtained according to different distributions when an AIC is used to choose the best distribution. The smallest value of AIC depicts the best distribution where this distribution is used to identify the mode or arrival time and duration. This

suggestion of arrival times and durations is extended for the subsequent analyses of activity timing comparison and activity frequency.

CHAPTER 6: MARKET SEGMENTATION OF TOUR PACKAGE PRODUCTS

6.1 Introduction

This chapter presents the methodology of the Market Segmentation Method (MSM) to segment the market of tour package products such as itineraries, attractions and activities. It begins by introducing the concept of market segmentation and its relationships with tourist characteristics.

Next, a framework for MSM composed of four elements, namely input, process, output and target is introduced that can address the market planning issue. This chapter then progresses where segmentation variables are determined, which is followed by the description of the Expectation-Maximisation (EM) Algorithm adopted in this research. Finally, the approach of market targeting is presented.

6.2 The Connection between Tourist Characteristics with Tour Package Products

Tourist characteristics were believed to have an impact on the tour package selection (O'Connor et al., 2005; Xia et al., 2008; Abdul Khanan, 2009). This is based on tourist preferences towards tour package offerings. Tourist preferences are assumed to be based on their characteristics, where in this research context, tourist characteristics are utilised to observe tourist spatio-temporal movement pattern.

A good example of a tourists' characteristics study was research by Hsu et al. (2009) where factors that affect tourists' choice of destination were identified. In the research, several characteristics of tourists in terms of tourists' visitation purpose for visiting attractions such as to rest, to relax and to enjoy environmental serenity were listed. Several outputs of the tourist visitation purpose (tourist preferences) linked to its specific tourist attractions are highlighted. As an example, tourists who seek to socialise, experience night life and shopping prefer Taipei 101 which is the tallest building in Taiwan.

Research by Kale et al. (1987) examined the travel preferences of youth segments towards tour package offerings. The research focussed more on the demographic characteristics of tourists where one of their findings suggested younger tourists aged 18 to 35 should be offered a tour package that is filled with activities and almost non-stop action. They also recognised the tendency of younger tourists to choose scenic and natural wonder attractions in their tour package.

Another research topic of linking tourist characteristics with tour package products is market segmentation. According to Smith (1956), market segmentation involves "viewing a heterogeneous market as a number of smaller homogenous markets, in response to differing preferences, attributable to the desires of consumers for more precise satisfaction of their varying wants". In terms of tourism, the varying wants refer to tour package products while the heterogeneous market means individuals, groups or organisations with one or more similar characteristics that cause them to have similar product needs (Pride et al., 2007)

In their market segmentation research, Thomson and Pearce (1980) used a statistical analysis to identify associations between tourist demographic characteristics with types of preferred tour packages. One of their results is that coach tours are more associated with elderly couples, many of whom are retired. In contrast, a camping theme tour package is mostly used by youth who are under 25, most of them single female, on their first overseas trip seeking the company of other young people and on a relatively cheap package although most of them are medium income earners.

Recall from Section 4.3 that most of the definition of tour packages listed several mainstream elements of tour packages such as tourist's budget, overall duration, preferred activities and accommodation. A distinctive study by Xia et al. (2010) took a step further in market segmentation by segmenting not according to these mainstream elements but according to the dominant movement pattern of tourists based on a set of real tourist movements dataset. Using a procedure called an EM algorithm, target markets of selected tourist characteristics were identified for these dominant tourist movement patterns.

To conclude, it was identified that particular tourist demographic characteristics for instance, age and gender did have preferences towards the offerings of tour package products. The correct selection of tourist characteristics to be segmented within this research is anticipated for showing some association with tour package products.

6.3   The Market Segmentation Method (MSM) Framework

The MSM framework is the extension of the MSM's section of Figure 3.3. Presented as Figure 6.1, the framework also consists of four elements namely input, process, output and target market.

The input element represents the methodological requirement, i.e. data for the method shown in Figure 3.3 within the MSM's implementation section. Initially, tourist characteristics defined as segmentation variables in marketing are determined where these variables signify the subject to be segmented. Such examples of tourist characteristics are age, gender and preferred type of attraction.

A specific set of tourist characteristics dataset is the main subject in this research and is carefully selected from a broad selection of existing tourists characteristics. The chosen characteristics dataset is later divided according to different itineraries, attractions and activities. The itineraries were obtained using the earlier Personalised Scheduling Method (PSM).
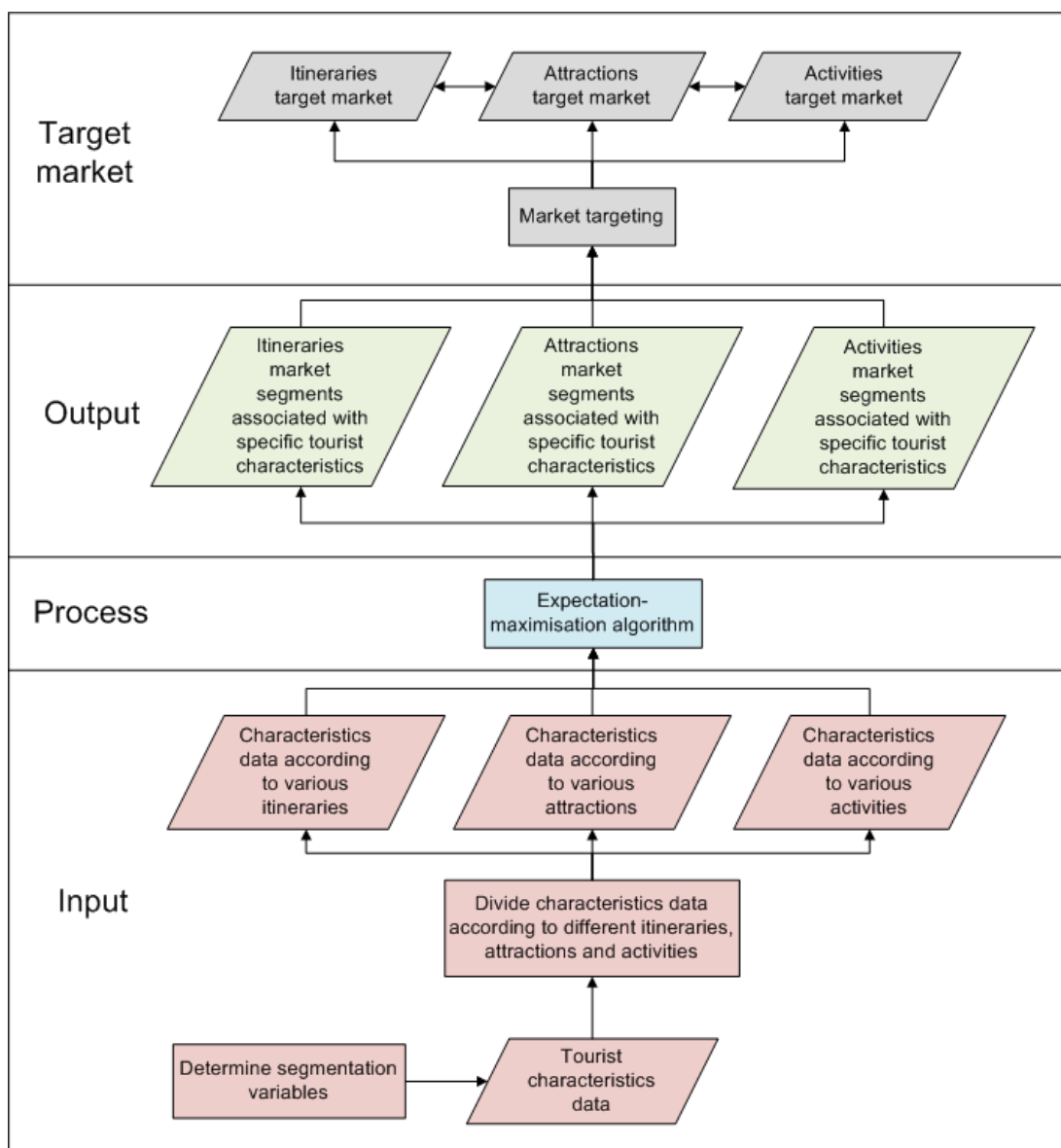
Figure 6.1   The Market Segmentation Method framework

The Process element marks the EM algorithm that is highlighted in Figure 3.3, shown within the PSM's description and implementation sections. The EM algorithm is used to segment or cluster the characteristics dataset into certain number of market segments for highlighting some significant association of characteristics that exist in specific tour package products.

The outputs of a particular number of market segments with associated characteristics are produced where later on, the selection of a target market is conducted to select the most significant market segment that can best represent any

specific tour package products. This is shown within the MSM's evaluation section of Figure 3.3.

Finally, the relationships between these target markets in regards to different tour package products are identified. Questions such as, "does a specific target market, which participates in any particular itinerary hence tour package, also visit any attraction included within that particular tour package or vice versa?" and "does a specific target market, which visits any particular attraction, also participate in activities included within that particular attraction or vice versa?" are highlighted.

These are represented by the double pointing arrows in the framework diagram. The entire elements of the market segmentation framework are detailed in the subsequent sections.

6.4   Determination of Segmentation Variables

A preliminary element of tour package products' market segmentation is the segmentation variables where in this chapter these variables are described as tourist characteristics (Bloom, 2004; Liu et al., 2010). In general, tourist characteristics are divided into four different variables which are geographic, demographic, psychographic and behavioural (Figure 6.2) (Pride et al., 2007).

Geographic variables are variables influenced by the distinction of geographical areas such as origin, population density and climate (Frank et al., 1972). Origin can be presented in various scales: country, state, county and township (Flognfeldt Jr, 1999). As described by Beane and Ennis (1987), demographic variables are the most prevalent variable as "consumers are placed on definite scales of measurement which are easily understood". Examples of these variables are age, gender and household income.

Psychographic variables involve values and lifestyle variables that are able to describe the cognitive structure of the individual where it is seen as a complement to the demographic characteristics for configuring tourism products (Nicolau and Más,

2005). The lifestyle variables cover three aspects which are activities, interests and opinions (Morrison, 1996).
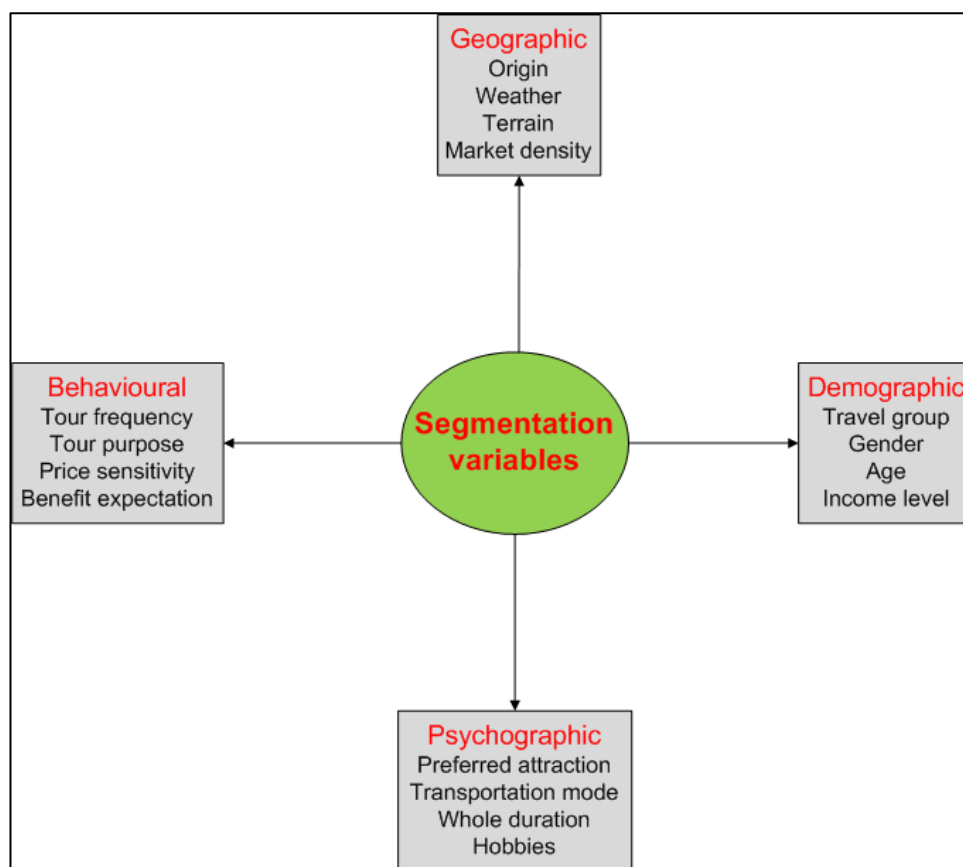
Figure 6.2   List of segmentation variables (Pride et al., 2007)

Behavioural variables are individual variables in relation to tourists attitude towards, use of or response to a product (Kotler and Keller, 2006). These include user frequency, user occasions and user status. As a plethora of variables exist, the question now is which variable should be chosen. The answer is to look back at the objectives of tourists' market segmentation (Xia et al., 2010).

For instance, a study by Kaynak and Yavas (1981) compared the similarity and differences between using traditional demographic and geographic variables with a trip purpose variable towards expenditure patterns, attitudinal orientations and satisfaction level of trips. As stated, the demographic variables in use are age, gender and income while origin is used to represent the geographic variable. The

classification of tourists as vacation/relaxation, business or visitor symbolises the trip purpose aspect of their research.

Finally, 12 segmentation variables that cover all four variable types are used for the market segmentation purpose. These variables are listed in Figure 3.8 as geographic, demographic, psychographic and behavioural characteristics which reflect the entire tourist characteristics that have been collected.

6.5   Clustering and the EM Algorithm

Clustering is a type of segmentation technique where the basic idea is to divide a number of cases (usually respondents) into homogenous segments according to pre-specified variables which are assumed to reflect the similarities of individuals within the segments and the dissimilarities between them (Dolnicar, 2002; Everitt, 2011).

The definition of clustering is parallel with the market segmentation definition presented in Section 6.2. The number of segments and characteristics of each segment are then determined by the dataset and clustering method used (Wedel and Kamakura, 2000).

It is important to note that the method presented in this chapter is an inverse method of the PSM as depicted on Figure 6.3 when it comes to the segmentation of tour packages. The PSM starts with specifying the tourist movement dataset according to certain tourist characteristics and in return, a tour package as the end product is obtained. The MSM is initiated by first having a tour package and, finally the specific characteristics for each respected tour package are obtained.

The EM algorithm as a particular type of clustering method is an iterative procedure for computing MLE in situations where, but for the absence of some additional data, MLE would be straightforward (McLachlan and Thriyambakam, 2008). The benefit of using the EM algorithm is that the likelihood (of identifying the cluster type) is guaranteed for each iteration (Trombe et al., 2011).
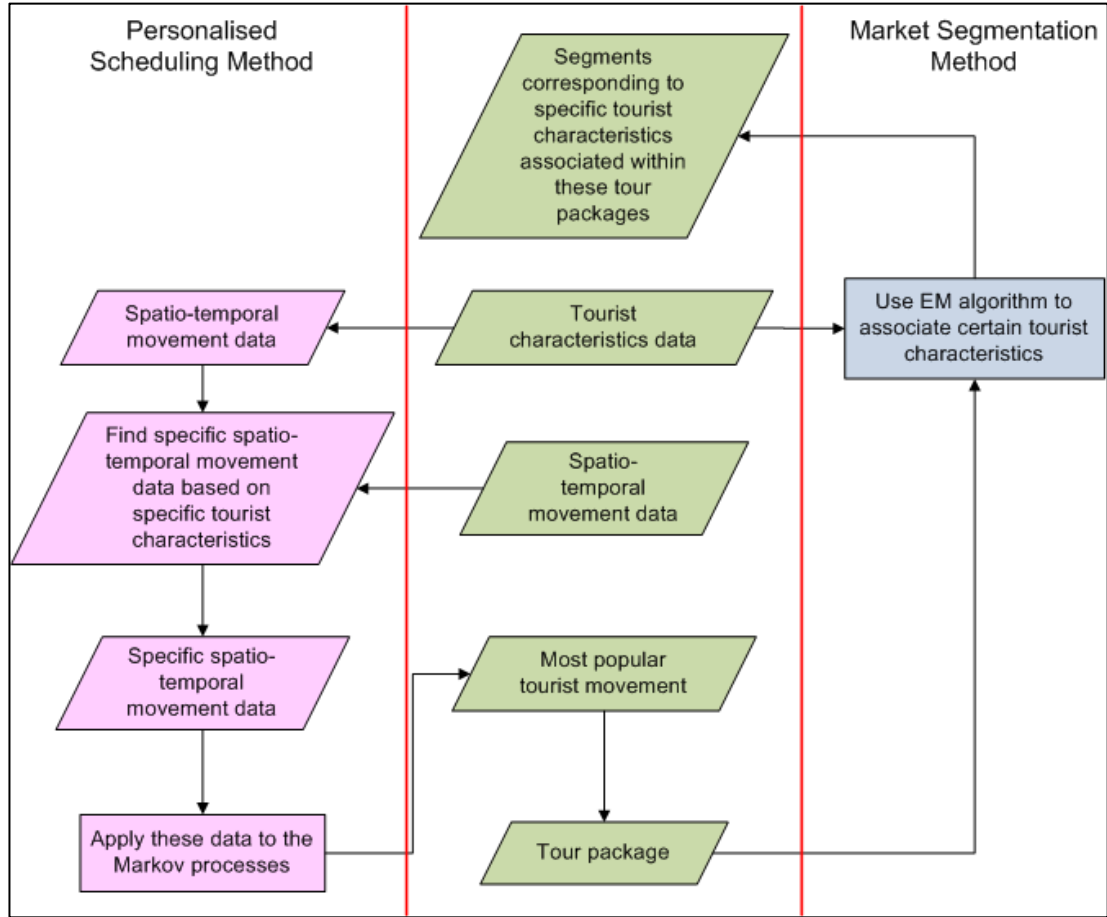
Figure 6.3   The MSM as an inverse method of the PSM

For a comprehensive introduction to EM algorithm theory, readers are referred to Witten and Frank (2005), McLachlan and Thriyambakam (2008), Gupta and Chen (2011), Xia et al. (2010) and Lin (2013). The rest of this section provides a brief summary of the EM algorithm theory extracted from those studies.

The fundamentals of the EM algorithm starts with considering a mixture model composed of K clusters which are given by:

$$p(y \mid \theta) = \sum_{j=1}^{K} \alpha_j \rho_j(y \mid \theta_j) \tag{6.1}$$

where $\rho_j(y \mid \theta_j)$ is the Probability Density Function (PDF) for the j-th cluster and the mixture weights, α represents the proportion of data belonging to the j-th

cluster such that $\sum_{j=1}^{K} \alpha_j = 1$. The parameters, θ_i of the cluster densities together with the mixture weights, α_i constitute the parameters, θ of the mixture model.

The EM algorithm for clustering procedures can be summarised as having three necessary steps as below. Equations (6.2) and (6.3) are derived from Xia et al. (2010):

(i)   Expectation (E) step: The initial association of specific parameters are established into specific clusters. Compute the probabilities w_{ij} that the i-th observation belongs to the j-th cluster using Bayes rule as in Equation (6.2). This is done at iteration t, given parameter estimates $\theta^{(t)}_i$ and $\alpha^{(t)}_i$.

$$w_{il}^{(t+1)} = \frac{\alpha_l^{(t)} \rho(y_i \mid \theta_l^{(t)})}{\sum_{j=1}^{K} \alpha_j^{(t)} \rho(y_i \mid \theta_j^{(t)})} \qquad (6.2)$$

(ii)   Maximisation (M) step: The calculation of the maximum likelihood of those specific parameters within its assigned clusters is carried out. For the purpose of updating the model parameters, the cluster membership probabilities are fixed and the likelihood is maximised. Equation (6.3) gives the estimation of maximum likelihood of the mixture weights.

$$\alpha_j^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} w_{ij}^{(t+1)} \qquad (6.3)$$

(iii)   The weights are then integrated into Equation (6.1) whereby parameters $\theta_j^{(t+1)}$ are ascertained by likelihood maximisation. These expectation, maximisation and model updating steps are reiterated until convergence parameter likelihood estimation is reached.

The EM algorithm is a soft clustering algorithm. In terms of clustering nominal dataset such as in this research, it assigns a frequency count with a fractional figure for any instance of tourists' characteristics within any cluster (Witten and Frank, 2005). This simply means that, any instance can be probabilistically split between several clusters.

The EM algorithm also produces a log-likelihood figure for each of its executions to measure how likely a particular cluster is (He and Zhang, n.d.). A larger figure of the log-likelihood means the better clustering result it is. In fact, this is the validation step of the EM algorithm. For validating the clustering result, a cross-validation method is used.

For this validation purpose, each market segmentation dataset containing various instances of tourist characteristics are replicated three times. This replication process was done using a random sampling method called Equal Probability of Selection Method Sampling (EPSEM) via the SPSS 20 statistics package. This replication produces three separate datasets with each dataset consisting of 90% of the original dataset with possibly a different combination of instances of tourist characteristics records (Crockett et al., 2011). When segmentation is implemented with these three separate datasets, the comparison of the log-likelihood figure is made where the clustering result with the smallest value of log-likelihood is chosen.

6.6   Market Targeting

The last step in the market segmentation process in this research is the process of market targeting. Market targeting is a detailed examination and evaluation of the range of possible market segments, usually only one market segment for a specific tour package and its products, and the degree to which the tour operator can satisfy some, or all of their expectations (Pride et al., 2007). It is done with a group of tourists called the target market, which consists of specific tourists that are selected for a particular purpose (such as to be promoted with specific tour packages) based on a certain criteria (Myers, 1996).

It is quite impossible for a tour operator to cater for the whole potential market segment of tourists. As an example, a tour operator who focuses on backpacking and camping tour packages must prioritise the youth market segment in order to achieve profitability. This is also known as concentrated market targeting (Phillip Island Nature Park, 2013). The aim of the market segmentation process in this chapter is to identify all four characteristics of tourists as presented in Section 6.4 and if they have any associations with specific tour packages, attractions visited and activities participated based on their spatio-temporal movement.

There could be a range of market segments identified for each tour package with its related product. However for this research, the segment with the heaviest users is selected as the target market for generating more profit (Evans and Stabler, 1995). For example, if three different market segments are identified for a tour package, the target market segment is the one with the highest percentage of members (Xia et al., 2010). Indeed, this market targeting process reflects the MSM's evaluation section of Figure 3.3.

6.7   Chapter Conclusions

This chapter demonstrates a method to segment the market of tour packages, attractions and activities. To start with, it details the connection between tourist characteristics with tour package products. Four types of tourist segmentation variables namely, geographical, demographic, psychographic and behavioural (such as origin, gender, tour purpose and length) are used to identify their potential link with any of those tour packages with its related products.

The essential concept of the EM algorithm is presented where this EM algorithm clusters tourist characteristics data for each tour package product. Various market segments are produced which require us to implement the final step of market targeting for choosing the target market.

CHAPTER 7: CASE STUDY, IMPLEMENTATION AND EVALUATION

7.1   Introduction

This case study, implementation and evaluation chapter contains the implementation and evaluation of each of the three methods presented in Chapters 4, 5 and 6. This chapter begins with the implementation of the Personalised Scheduling Method (PSM). The data in regards to the personalised tourist spatio-temporal movement is implemented within the context of the semi-Markov and scheduling processes to produce itineraries followed by tour packages.

Next, this chapter presents the implementation of the Activity Planning Method (APM). The data in regards to the tourist temporal movement is implemented within the context of the Probability Density Function (PDF) for estimating tourist arrival times and visiting durations. The output of the estimation of arrival times and durations are later used to plan the activities timing within attractions.

Finally, this chapter implements the Market Segmentation Method (MSM). The data in relation to the tourist characteristics is implemented towards the Expectation-Maximisation (EM) algorithm for producing various market segments by clustering the tourist characteristics dataset according to specific tour package products.

Subsequently, target markets are identified from several market segments where the marketing effort can be focussed upon.

By undertaking the implementation and evaluating the results, this case study validates the concepts and models presented in Chapters 4, 5 and 6. Results from each method are carefully explained and discussed to draw significant conclusions.

7.2 The Personalised Scheduling Method (PSM)

This section provides the implementation and evaluation of the PSM to develop personalised itineraries and tour packages. Initially, a cohort that corresponds to female and middle aged tourists with 391 records was chosen from the sample of 1104 tourists. From the survey, these characteristics are the most popular combination that involves two types of characteristics (see Table 7.1).

Table 7.1   Top five tourists' profiles based on combination of characteristics

No.	Tourists characteristics combination	Number of records
1	Female and middle aged	391
2	Male and young	285
3	Female with partner	175
4	Female with friends and relatives	172
5	Male with friends and relatives	143

The above female and middle aged tourists' dataset of 391 records was divided into two subsets of data: 70 % for training and 30 % for testing where this sub-section implemented the training dataset. This was done using a random sampling method called Equal Probability of Selection Method Sampling (EPSEM) using the SPSS 20 statistics package (Crockett et al., 2011). Therefore, the corresponding training dataset has records 274 while the testing dataset has 117.

7.2.1 The Estimation of Spatial Movement Transition Probability

The frequencies of the spatial movement transition of tourists are listed as a flow map in Figure 7.1. The map lists multiple transitions from one state to another (nine attractions considered on Phillip Island), alongside with its frequencies figures. It
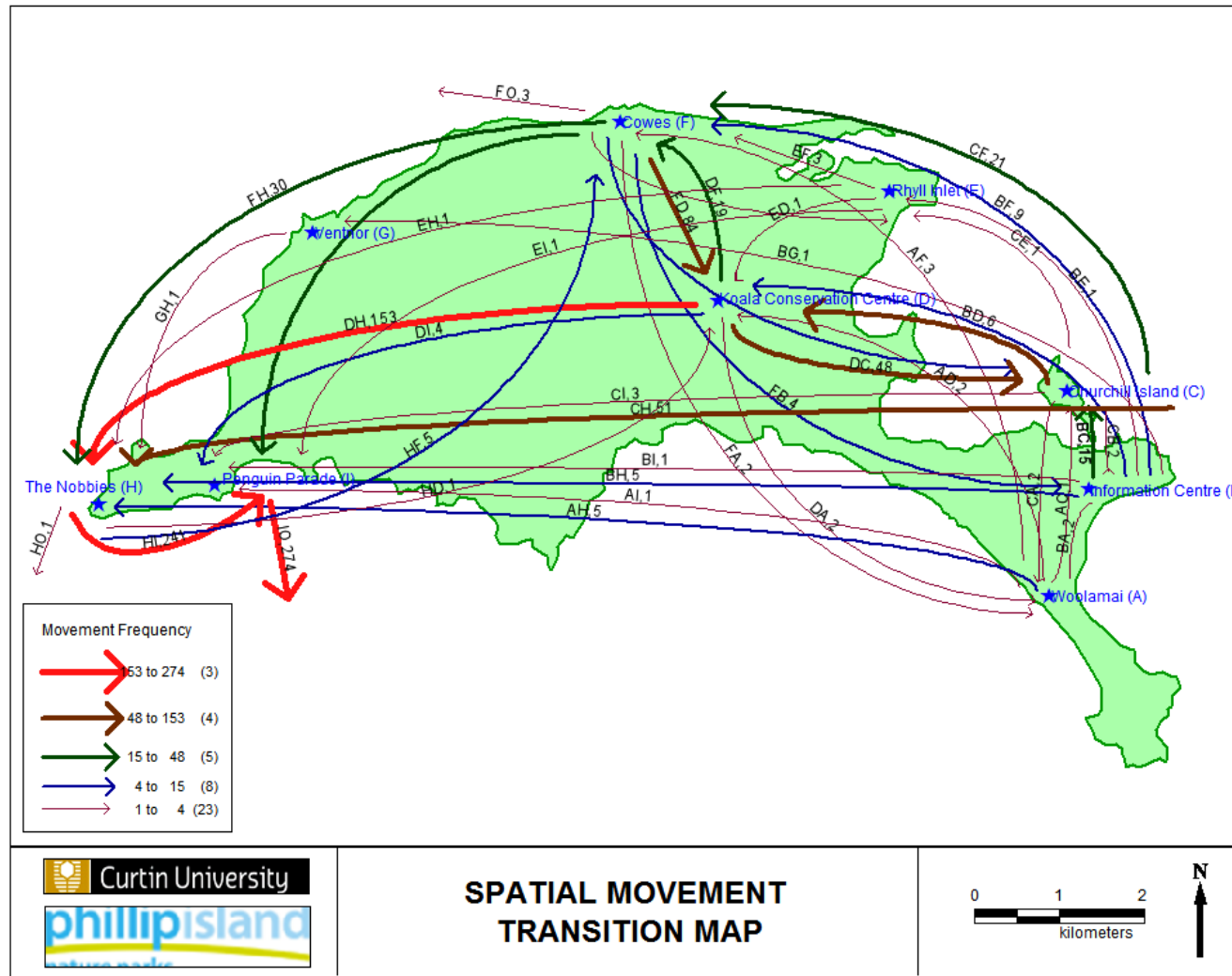
Figure 7.1   The spatial movement transition map

also includes an absorbing state, 'OUT', where tourists leave the states that mark the completion of their tour. These are shown as transitions FO, HO and IO where the arrows exit Phillip Island.

For instance, from The Nobbies (H) to Penguin Parade (I), the number of tourists who made that transition is 241. The other significant transitions with 50 or more tourists making the transition are Koala Centre (D) →The Nobbies (H) (153), Cowes (F) → Koala Centre (D) (84), Churchill Island (C) → Koala Centre (D), Churchill Island → The Nobbies (51) and Penguin Parade (I) → OUT (O) (274). Figure 7.1 also highlighted the aspect of the thickness of the flow line where the higher the frequency of the transition is, the thicker the flow line will be. Some transitions do not exist where this indicates no tourists making that transition, such as from D to E.

Based on the spatial movement transition map (Figure 7.1), a list of spatial movement transition probabilities $P_{ij}, i, j \in S$, was calculated using the semi-Markov process (See Equation 4.4). A spatial movement transition probability map (see Figure 7.2) with a same list of transitions as Figure 7.1 is presented, but this time, it lists the figures of transition probability instead of frequency.

For instance, from Rhyll Inlet (E) to Cowes (F), the transition probability is 0.5. This probability of 0.5 is derived using Equation (4.4) by dividing the number of tourists making the E to F transition which is 3 with the summation of the number of tourists making transitions starting at E, which is 6 (altogether adding E to A, E to B, E to C, E to D, E to F, E to G, E to H, E to I and E to Out). This same method of calculation is used for other transitions starting at different attractions.

The other significant transitions with 50 % or more tourists making the transition are Koala Centre (D) →The Nobbies (H) (0.677), Cowes (F) → Koala Centre (D) (0.535), The Nobbies (H) → Penguin Parade (I) (0.972) and Penguin Parade (I) → OUT (O) (1). These pairs of attractions are believed to be more popular among tourists than other pairs. Akin to Figure 7.1, some transitions do not exist where this indicates no tourists making that transition, such as from E to G. This is because the
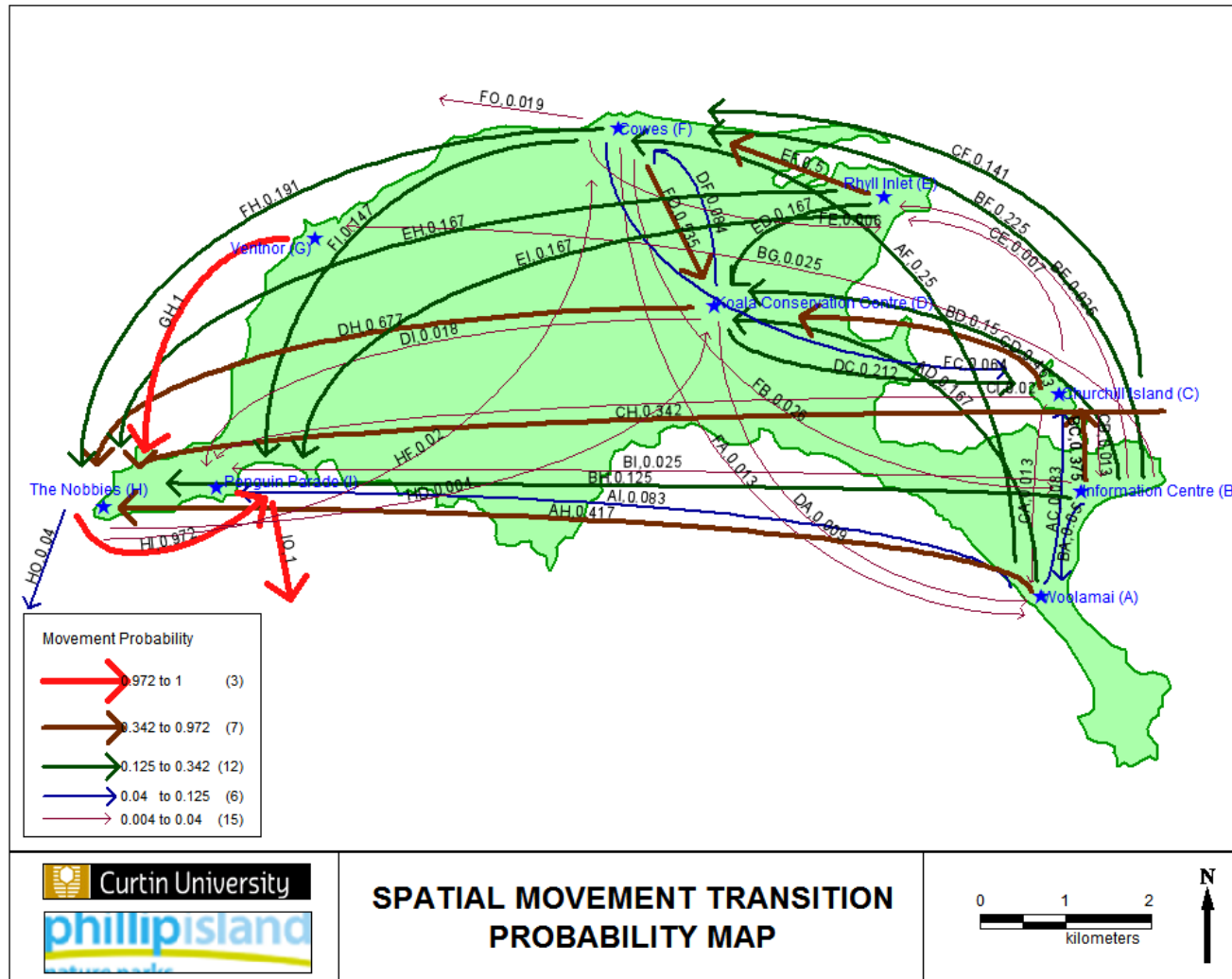
Figure 7.2   The spatial movement transition probability map

majority of tourists visiting E (Rhyll Inlet) and G (Ventnor) are local tourists who are excluded from the survey.

After getting the entire figures of probabilities from Figure 7.2, a stationary discrete semi-Markov process as described in equation (4.1) was used to model the actual movement of tourists in order to obtain the probability of their movement sequence from one attraction to another. As an example, for the actual movement pattern of 'FDHIO', the transition probabilities of FD, DH, HI and IO needs to be multiplied where later, a figure of 0.329 as the probability of the movement sequence is obtained. In fact, 'FDHIO' is the actual movement sequence with the highest figure of probability of the movement sequence, i.e. the most popular movement sequence for four attractions as shown in Table 7.2.

Table 7.2   Three most popular actual movement sequences for four attractions

No.	Movement sequence	Movement sequence probability
1	FDHIO	0.329
2	CDHIO	0.305
3	BCHIO	0.125

7.2.1.1 Validating the Spatial Movement Transition Probability Model

After obtaining the probability results of the spatial movement transition probability model, the key question now is how well these probabilities are modelled using this methodology? In this section, the results were validated using Chi-Square Goodness of Fit Test, using Equation (4.9).

Using the observed frequencies from the testing dataset of the transition from Rhyll Inlet (E) to Cowes (F), the spatial movement transition probability figure of the similar transition, P_{ij} and the frequency figure of tourists starting their transition from attraction i to all other attractions, N_i, the Chi-Square $\chi_{EF}^{2(70\%)}$ is calculated as:

$$\chi_{EF}^{2(70\%)} = \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(O_{ij} - N_i P_{ij})^2}{N_i P_{ij}}$$

$$= \frac{(6.425 - 26 \times 0.5)^2}{26 \times 0.5} = 3.325$$

To obtain the P-value, one must have a degree of freedom, df and the confidence level, α. The degree of freedom in this example is three since there are only four transitions that begin from attraction E. The figure of 0.05 is selected as the confidence level, thus producing a P-value of 7.815. Notice that the $\chi^{2(70\%)}_{EF}$ value is outside the P-value, thus the observed counts are compatible with the expected counts (based on the spatial movement transition probability), i.e. the model provides a good fit of the data (Apostolakis and Jaffry, 2005).

The validation results using a similar Equation (4.9) for each spatial movement transition is presented in as a matrix in Table 7.3. The matrix shows the 90-step transitions from one state to another (nine attractions considered on Phillip Island). Similar with Figures 7.1 and 7.2, it also includes an absorbing state, 'OUT', where tourists leave the states that mark the completion of their tour. The matrix reads from each attraction in the leftmost column to each attraction in the top row as one pair of transitions. For instance, from The Nobbies (H) to Penguin Parade (I), the expected counts (χ) and the observed counts (PV) 0.192 and 7.815 respectively.

The entirely observed counts (PV) are compatible with the entire expected counts (χ) for each movement transition, i.e. the model provides a good fit for each movement transition. This is not surprising as according to Xia (2007), a general Markov process such as Equation (4.4) produces a one-step transitional probability that does not involve the assumption of stationarity. This means the choice of future attractions would depend on which stage of the tour the decision has to be made. This assumption was ignored in her research that made some of the validation results unfit.

The GH and IO transitions are not included in the validation process as both do not produce fractional probabilities by having probabilities of 1, thus the corresponding cells are marked with a hyphen. Other transitions marked with the hyphen indicate the unavailability of tourists making that transition.

Table 7.3   Validation result for spatial movement transition probability

	Woolamai (A)	Visitors Centre (B)	Churchill Island (C)	Koala Centre (D)	Rhyll Inlet (E)	Cowes (F)	Ventnor (G)	Nobbies (H)	Penguin Parade (I)	OUT
Woolamai (A)		-	$\chi = 0.84$ $PV = 9.488$	$\chi = 1.39$ $PV = 9.488$	-	$\chi = 1.69$ $PV = 9.488$	-	$\chi = 1.701$ $PV = 9.488$	$\chi = 0.84$ $PV = 9.488$	-
Visitors Centre (B)	$\chi = 1.805$ $PV = 14.07$		$\chi = 5.859$ $PV = 14.07$	$\chi = 4.335$ $PV = 14.07$	$\chi = 0.951$ $PV = 14.07$	$\chi = 5.406$ $PV = 14.07$	$\chi = 0.951$ $PV = 14.07$	$\chi = 3.828$ $PV = 14.07$	$\chi = 0.951$ $PV = 14.07$	-
Churchill Island (C)	$\chi = 0.974$ $PV = 12.59$	$\chi = 0.974$ $PV = 12.59$		$\chi = 9.945$ $PV = 12.59$	$\chi = 0.494$ $PV = 12.59$	$\chi = 7.75$ $PV = 12.59$	-	$\chi = 11.03$ $PV = 12.59$	$\chi = 1.44$ $PV = 12.59$	-
Koala Centre (D)	$\chi = 1.96$ $PV = 9.488$	-	$\chi = 9.306$ $PV = 9.488$		-	$\chi = 4.981$ $PV = 9.488$	-	$\chi = 4.988$ $PV = 9.488$	$\chi = 3.86$ $PV = 9.488$	-
Rhyll Inlet (E)	-	-	-	$\chi = 3.053$ $PV = 7.815$		$\chi = 3.325$ $PV = 7.815$	-	$\chi = 3.053$ $PV = 7.815$	$\chi = 3.053$ $PV = 7.815$	-
Cowes (F)	$\chi = 1.949$ $PV = 14.07$	$\chi = 3.798$ $PV = 14.07$	$\chi = 7.991$ $PV = 14.07$	$\chi = 11.93$ $PV = 14.07$	$\chi = 0.987$ $PV = 14.07$		-	$\chi = 13.05$ $PV = 14.07$	$\chi = 11.15$ $PV = 14.07$	$\chi = 2.886$ $PV = 14.07$
Ventnor (G)	-	-	-	-	-	-		-	-	-
Nobbies (H)	-	-	-	$\chi = 0.992$ $PV = 7.815$		$\chi = 4.8$ $PV = 7.815$	-		$\chi = 0.192$ $PV = 7.815$	$\chi = 0.992$ $PV = 7.815$
Penguin Parade (I)	-	-	-	-	-	-	-	-		-

7.2.2   The Estimation of the Mode of Arrival Time with its Probability

The mode of arrival time is the highest occurrence time that tourists arrive at an attraction or a state before moving to the next state, for example, arriving at Cowes at 3.00 pm before moving to Rhyll Inlet. Using Equation (4.6), the mode of arrival time transitional probability was calculated that shows the probability of the arrival time mode at attraction i, given that the movement starts at i and then moves to attraction j.

For estimating the mode of arrival time probability at attractions, a series of possible PDFs shown as Equation (5.5) was implemented, with the Akaike Information Criterion (AIC) as depicted in Equation (5.6) adopted for selecting and validating the best distribution based on the goodness of fit and number of parameters. The arrival time dataset was categorised according to each transition. Each transitional arrival time dataset was then fitted within each distribution to choose the most suitable distribution based on the smallest figure of AIC.

In this case study, the Birnbaum-Saunders density function presented the best distribution fit for transitional arrival time at the Koala Centre (D), before moving to Woolamai (A) as it produces the lowest value of AIC as shown in Table 7.4. This smallest figure of AIC for Birnbaum-Saunders means it is better in capturing the entire dataset if compared to other common distributions such as Weibull and Gamma (Xia et al., 2011). Figure 7.3 depicts the distribution fit of arrival times for the DA transition.

Table 7.4   Three distributions with the smallest figure of AIC

No.	Distribution's Name	AIC Figure
1	Birnbaum-Saunders	52.722
2	Weibull	52.889
3	Gamma	53.687
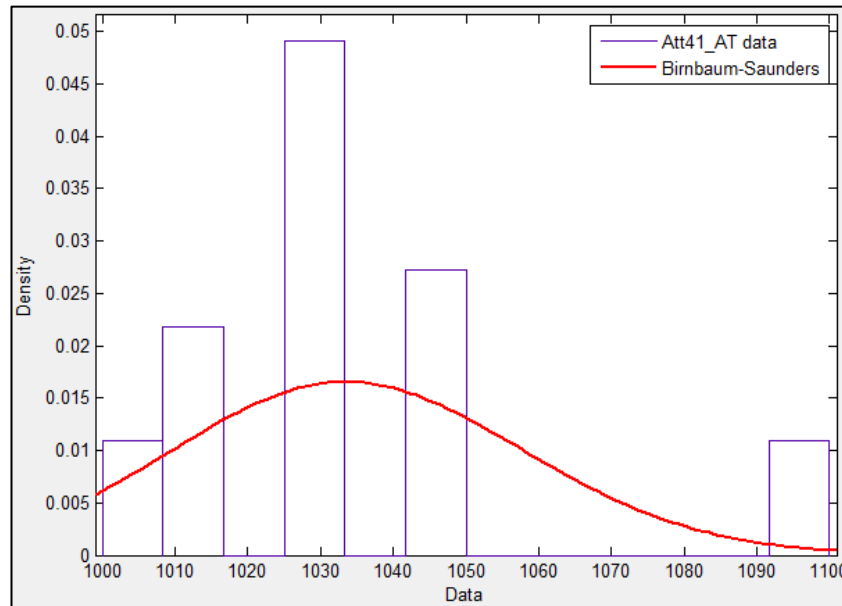
Figure 7.3   The distribution fit of arrival times for the DA transition

This signifies the Birnbaum-Saunders as the distribution that provided the best goodness of fit while implementing a reasonable number of parameters that do not overfit the dataset. This procedure is repeated for all other transitions where other transitions might result in a different distribution selected, since tourists' arrival times are different, resulting in a different fit of distribution functions.

Tables 7.5, 7.6 and Figure 7.4 each shows three different types of information: (1) the best fit distribution for each transition, (2) the AIC value in selecting and validating the best fit distribution, and (3) a flow map that reflects the mode of arrival time transitional probability. The way of reading Tables 7.5 and 7.6 is similar to Table 7.3 while the way of reading Figure 7.4 is similar to Figures 7.1 and 7.2. Based on Table 7.4, some significant arrival time transitions with more than 40 % tourists, i.e. more popular transitional arrival times are Churchill Island (C) → Rhyll Inlet (E) (0.426), Cowes (F) → Rhyll Inlet (E) (0.426) and The Nobbies (H) → Koala Centre (D) (0.453).

Other significant arrival time transitions based on the more popular spatial movement transition of Table 7.3 are Koala Centre (D) → The Nobbies (H) (0.003), Cowes (F) → Koala Centre (D) (0.002), The Nobbies (H) → Penguin Parade (I) (0.002) and Penguin Parade (I) → OUT (1). Notice that the first three probability figures are

Table 7.5   The list of distribution fits for each transition

	Woolamai (A)	Visitors Centre (B)	Churchill Island (C)	Koala Centre (D)	Rhyll Inlet (E)	Cowes (F)	Ventnor (G)	Nobbies (H)	Penguin Parade (I)	OUT
Woolamai (A)		-	Gamma	Log-logistic	-	Extreme value	-	?	Birnbaum-Saunders	-
Visitors Centre (B)	Inverse-gaussian		Generalised Pareto	Generalised extreme value	Birnbaum-Saunders	Nakagami	?	Generalised Pareto	Nakagami	-
Churchill Island (C)	?	Weibull		Generalised Pareto	Inverse-gaussian	Birnbaum-Saunders	-	Logistic	Weibull	-
Koala Centre (D)	Birnbaum-Saunders	-	Generalised Pareto			Generalised Pareto	-	Generalised extreme value	Extreme value	-
Rhyll Inlet (E)	-	-	-	Extreme value		Extreme value	-	Inverse-gaussian	Weibull	-
Cowes (F)	Extreme value	?	Inverse-gaussian	Logistic	Logistic		-	Generalised extreme value	Weibull	Generalised pareto
Ventnor (G)	-	-	-	-	-	-		Inverse-gaussian	-	-
Nobbies (H)	-	-	-	Weibull	-	Nakagami	-		T-location scale	Logistic
Penguin Parade (I)	-	-	-	-	-	-	-	-		Weibull

Table 7.6   The AIC value for the best fit distribution

	Woola-mai (A)	Visitors Centre (B)	Churchill Island (C)	Koala Centre (D)	Rhyll Inlet (E)	Cowes (F)	Vent-nor (G)	Nobbies (H)	Penguin Parade (I)	OUT
Woola-mai (A)		-	23.127	35.883	-	51.896	-	?	32.754	-
Visitors Centre (B)	163.838		73.511	107.846	44.79	56.874	?	87.985	64.837	-
Chur-chill Island (C)	?	20.221		651.764	78.151	357.268	-	401.321	301.453	-
Koala Centre (D)	52.722	-	874.581		-	719.336	-	634.52	595.857	-
Rhyll Inlet (E)	-	-	-	245.234		355.631	-	145.794	35.783	-
Cowes (F)	95.925	?	287.948	235.061	153.35		-	464.648	559.904	31.241
Ventnor (G)	-	-	-	-	-	-		12.314	-	-
Nobbies (H)	-	-	-	160.409	-	223.256	-		1142.335	267.848
Penguin Parade (I)	-	-	-	-	-	-	-	-		1567.233
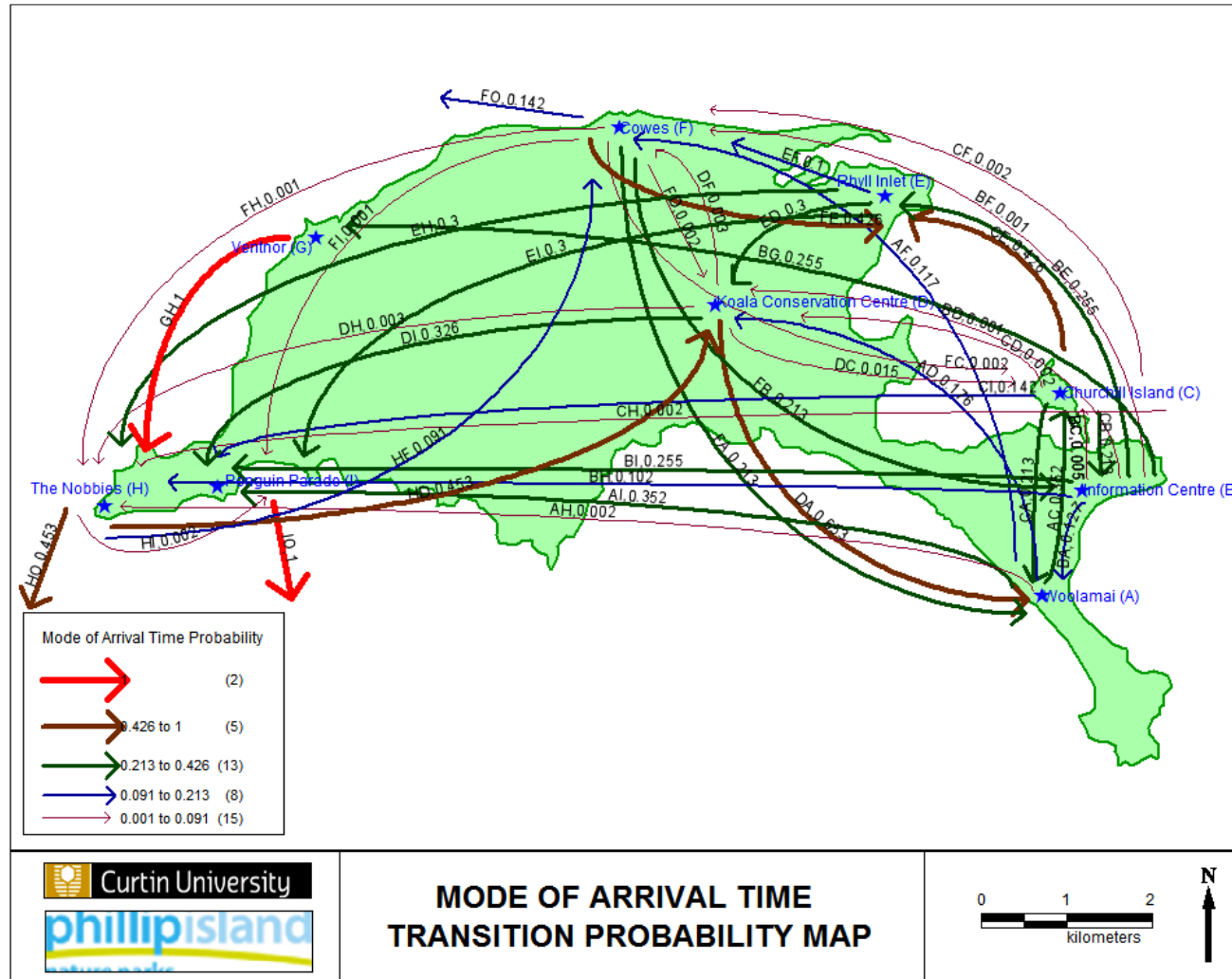
Figure 7.4   The mode of arrival time transition probability map

small, although many tourists make this transition based on Figure 7.2 which means tourists arrival times vary differently. The figure for the Penguin Parade (I) → OUT should also be small as tourist arrival times vary differently, however, because it is the only transition from Penguin Parade where the figure is normalised, the probability figure of 1 is obtained.

The sign, '?' in Tables 7.5 and 7.6 denotes that there are no distributions that can be fitted into the dataset because the size of the dataset is too small. In this situation, one simply calculates the simple probability, by dividing the frequency of the high occurrence, arrival times with the total size of the dataset.

The mode of transitional arrival time is presented as a flow map in Figure 7.5 that lists the entire mode of arrival time for all transitions. The way to read this flow map by taking transition CE as an example is the arrival time at attraction C before moving to attraction E is 3.00 pm. Another example of transition FD iscthe arrival time at attraction F prior to moving to attraction D is 12.20 pm.

Some important findings corresponding to a higher figure of mode of arrival time transition probability, i.e. more popular transitional arrival times (Figure 7.4) arise, such as the arrival times at The Nobbies (H) prior to shifting to Koala Centre (D) and the arrival times at Cowes (F) before shifting to Rhyll Inlet (E) are 10.00 am and 1.30 pm respectively.

Other significant transitional arrival times based on the popular spatial movement transitions probability of Figure 7.2 are the arrival time at The Nobbies (H) before moving to Penguin Parade (I) at 4.00 pm and the arrival time at Penguin Parade (I) before exiting the tour (O) at 7.19 pm. This is expected based on observation during the data collection period where many tourists arrive at 4.00 pm to 5.30 pm and 6.30 pm to 7.30 pm for The Nobbies and Penguin Parade respectively.

Similar to the spatial movement transition probability, a stationary discrete semi-Markov process as described in Equation (4.1) was used to model these arrival time patterns of tourists in order to obtain the probability of their transitional arrival time.

Figure 7.5   The mode of arrival time transition map

As an example, for an actual movement pattern of 'FCEIO', the transitional arrival time probabilities of FC CE, EI and IO are multiplied altogether where the final figure of 2.56×10^{-4} as the probability of movement sequence is obtained. Indeed, 'FCEIO' is the actual movement sequence with the highest figure of the transitional arrival time probability, i.e. the most popular transitional arrival times for four attractions, as shown in Table 7.7.

Table 7.7   Three actual movement sequences with the most
popular transitional arrival times for four attractions

No.	Movement sequence	Transitional arrival time probability
1	FCEIO	2.56×10^{-4}
2	DAFHO	1.53×10^{-4}
3	DAFIO	7.64×10^{-5}

7.2.3   Scheduling Tourists Movement

The probability of spatio-temporal movement patterns, hence the itinerary, was calculated using Equation (4.12). Basically, this probability is the product of the spatial movement transition probability map (Figure 7.2) and the mode of arrival time transition probability map (Figure 7.4).

Table 7.8 illustrates three examples of these calculation results for three actual movement patterns for four attractions with the highest value of the probabilities of spatio-temporal movement, i.e. the most popular itinerary when compared with other actual movement sequences: BEHIO, FCEIO and FDHIO. In this table, it can be seen that the product rule of probability is put in place whereby the spatial movement transition probability is multiplied against the mode of arrival time transition probabilities.

Although FDHIO and FCEIO have the highest spatial movement transition and mode of arrival time transition probabilities respectively, BEHIO appeared to have the highest probability of spatio-temporal movement pattern (itinerary) at 6.2×10^{-7}, shown as the first record in Table 7.8. Finally, the visiting duration at an attraction is

estimated arbitrarily from the difference between arrival time at the attraction and the arrival time at the next attraction.

Table 7.8   The ranking of the probability of spatio-temporal movement pattern for four attractions

No.	Movement sequence	Movement sequence probability	Transitional arrival time probability	Probability of spatio-temporal movement pattern (Itinerary)
1	BEHIO	4.05×10^{-3}	1.53×10^{-4}	6.2×10^{-7}
2	FCEIO	1.39×10^{-5}	2.56×10^{-4}	3.56×10^{-9}
3	FDHIO	3.29×10^{-1}	6.85×10^{-11}	2.3×10^{-11}

7.2.4   Ensuring the Consistency of an Itinerary and Completing Tour Packages Design

A consistent itinerary is described as an itinerary with the proper flow of timing that sequentially corresponds to the movement sequence. Table 7.8 does not consider the inconsistency cases of itineraries. An inconsistent itinerary represents a physically impossible movement sequence, i.e. an itinerary with temporal elements that do not correspond to the movement sequence.

Table 7.9 highlights a case of this inconsistency in the first record where the arrival time at The Nobbies as the latter attraction is earlier than Koala Centre which is the earlier attraction. If a violation of consistency occurred, the subsequent most frequent arrival time mode was chosen. The process was iterated until consistency is achieved as shown as the second record in Table 7.9. Indeed, this consistency checking step symbolises the PSM's evaluation section of Figure 3.3 that was further detailed at the bottom left section of Figure 4.2.

Table 7.9   The consistent and inconsistent itineraries

No.	Sequence	Status	Itinerary
1	BDHIO	Inconsistent	Visitor Centre at 9.30 am, 15 minutes. Koala Centre at 4.15 pm, 45 minutes. The Nobbies at 4.00 pm, 2 ½ hours. Penguin Parade at 7.19 pm, 3 hours.
2	BDHIO	Consistent	Visitor Centre at 9.30 am, 15 minutes. Koala Centre at 10.15 am, 2 hours. The Nobbies at 4.00 pm, 2 hours. Penguin Parade at 7.19 pm, 3 hours.

Table 7.10 shows a complete tour package which is the expansion of the itineraries shown as the second record in Table 7.9 after the inclusion of the elements of the tour package. The suggested routes are the shortest path routes ensuring least journey duration that are taken to move from one attraction to another. Although shortest, most of them are scenic routes that are able to satisfy tourists. The suggestion of routes is flexible, tourists can opt to choose other routes that suit their preferences.

Table 7.10   Example of a tour package

Sequence	Tour Package
BDHIO	Visitor Information Centre at 9.30am, 15 minutes. **Turn left to Phillip Island Road from the main entrance of the Visitor Information Centre** Koala Centre arrives at 10.15am, spend 2 hours. Activities – interactive learning at the Visitors Centre (10.15am), Koala Boardwalk (10.45am), Woodland Boardwalk (11.15am), Wildlife Boardwalk (11.45am). Cost – AUD 11 (adult), AUD 5.50 (child). **Turn right to Phillip Island Road from the main entrance of Koala Centre** OPTIONAL TOURS. Cowes arrives at 12.30pm, spend 3 hours. Activities – lunch (12.30pm), visit cultural centre (1.30pm), visit museum (2.30pm). Cost – Free (except lunch). **Turn right to Phillip Island Road from the main entrance of the museum and turn right at Ventnor Road.** The Nobbies arrives at 4pm, spend 2 hours. Just for Pups kids' session (4pm), special presentation at the Visitors Centre (4.30pm), boardwalk (5pm), seal viewing (5.30pm). Cost – Free. **Turn left to Ventnor Road from the main entrance of The Nobbies** Penguin Parade at 7.19pm, 3 hours. Activities – interactive display at the Visitors Centre (7.19pm), penguin movie (8pm), penguin viewing (8.45pm). Cost – AUD 22.10 (adult), AUD 11 (child).

The total budget of the tour package for an adult is AUD 33.10 and for a child is AUD 16.50 where this covers the cheapest entrance fees for Koala Centre and Penguin Parade. If tourists opt for further enjoyment at Penguin Parade, they can spend more to purchase specialised packages such as the guided ranger tour, Penguins Plus and the ultimate adventure tour (Phillip Island Nature Park, 2012).

In terms of activities, only one scheduled activity is included in the tour package which is the penguin viewing activity at 8.45 pm. All other activities are self guided activities, thus the timings are arbitrarily assigned to adhere to the arrival time and visiting duration at each attraction.

One optional tour is included where tourists are recommended to visit Cowes for activities such as lunch, visit the cultural centre and museum. This optional tour can be excluded if tourists request to do so. A map that shows the tour package of BDHIO is presented as Figure 7.6. It clearly shows the attraction locations and routes from one attraction to another together with other details described above.

The PSM procedure from Sections 7.2.1 until 7.2.4 is repeated for specific combination of tourist characteristics where these significant characteristics were identified from previous research (Enoch, 1996; Arrowsmith et al., 2005). The output of this procedure is presented in Table 7.11 that shows the recommended itineraries for specific types of tourists together with their figure of probability of spatio-temporal movement pattern. These itineraries are itineraries with the highest combined probability according to each combination of characteristics with at least three attractions.

Combinations of characteristics for instance 'male and young' (1st record of Table 7.11) and 'female and middle aged' (2nd record of Table 7.11) are purposely selected where these are comparable characteristics of male against female and young against middle aged, so as to facilitate the direct comparison between those respective itineraries.

Similar cases of direct comparison are repeated for a comparable combination of characteristics such as 'male, young and international' (3rd record of Table 7.11) against 'female, middle and local' (4th record of Table 7.11) or 'experienced and preferred sightseeing' (5th record of Table 7.11) against 'inexperienced and preferred wildlife' (6th record of Table 7.11).
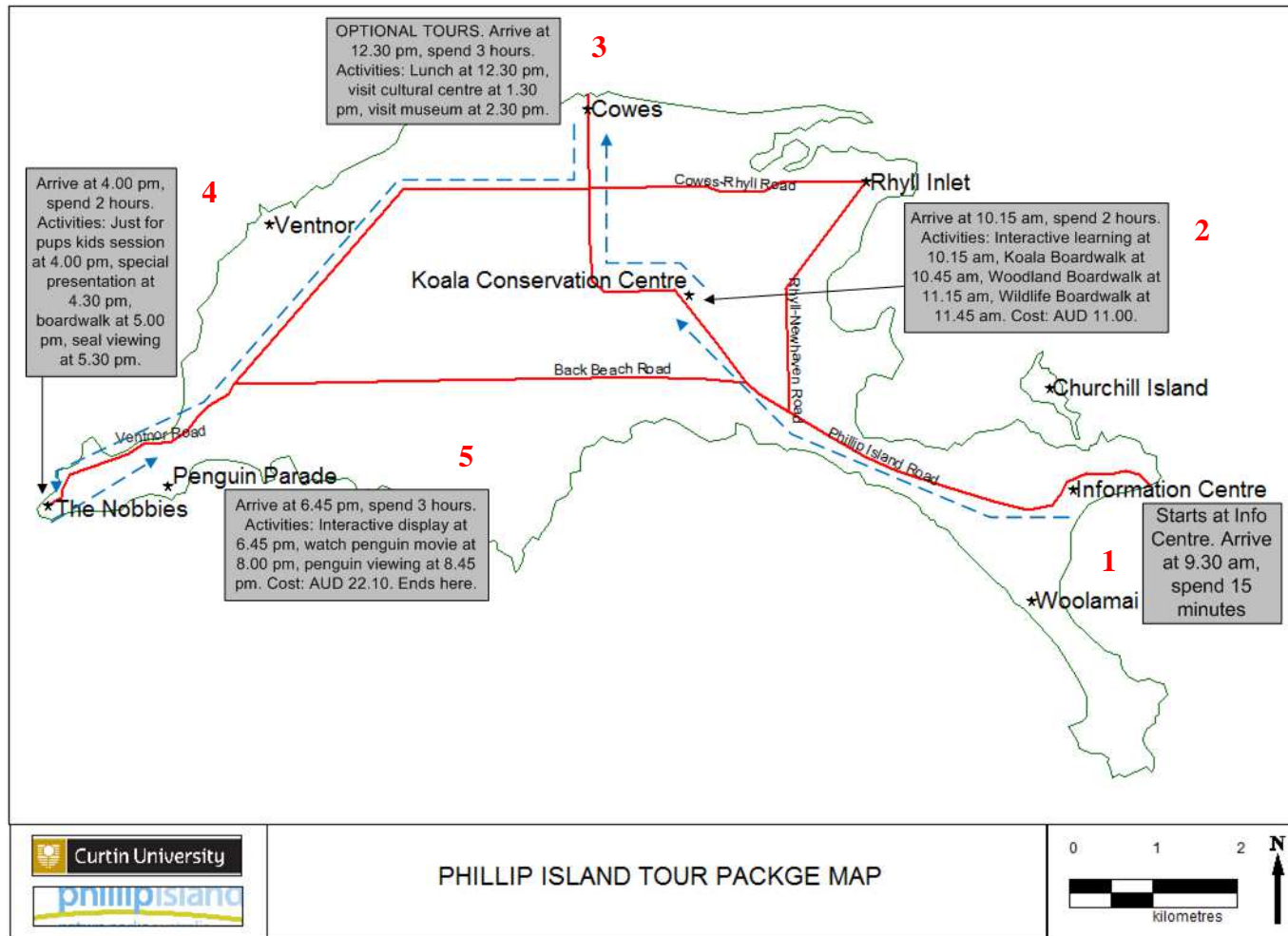
Figure 7.6   The tour package map of Phillip Island for the movement sequence 'BDHIO'

Table 7.11   Recommendation of itineraries for certain types of tourists

No.	Type of Tourist	Itinerary	Probability
1	Male and young	Koala Centre at 10.31 am, 2 hours. Churchill Island at 1.15 pm, 2 ½ hours. The Nobbies at 5.09 pm, 2 hours. Penguin Parade at 7.27 pm, 3 hours.	1.14×10^{-3}
2	Female and middle aged	Information Centre at 3.00 pm, 15 minutes. Rhyll Inlet at 3.30 pm, 15 minutes. The Nobbies at 4.00 pm, 2 ½ hours. Penguin Parade at 7.19 pm, 3 hours.	6.18×10^{-10}
3	Male, young and international	Churchill Island at 11.31 am, 2 hours. Koala Centre at 2.00 pm, 40 minutes. Cowes at 2.43 pm, 45 minutes. Rhyll Inlet at 3.45 pm, 1 hour. The Nobbies at 5.11 pm, 1 ½ hours. Penguin Parade at 7.10 pm, 3 hours.	3.71×10^{-9}
4	Female, middle age and local	Koala Centre at 4.00 pm, 1 hour. Cowes at 5.15 pm, 1 ½ hours. Penguin Parade at 7.15 pm, 3 hours.	8.27×10^{-4}
5	Experienced and preferred sightseeing	Koala Centre at 2.00 pm, 40 minutes. The Nobbies at 5.11 pm, 1 ½ hours. Penguin Parade at 7.10 pm, 3 hours.	9.8×10^{-4}
6	Inexperienced and preferred wildlife	Information Centre at 3.00 pm, 15 minutes. Koala Centre at 3.30 pm, 1 ½ hours. The Nobbies at 5.45 pm, 1 ½ hours. Penguin Parade at 7.37 pm, 2 ½ hours.	6.08×10^{-6}
7	Inexperienced, preferred wildlife and with partner	Information Centre at 9.30 am, 15 minutes. Churchill Island at 11.25 am, 1 hour. Koala Centre at 12.30 pm, 1 hour 30 minutes. Cowes at 3.45 pm, 1 hour. The Nobbies at 5.06 pm, 1 ½ hours. Penguin Parade at 6.45 pm, 2 ½ hours.	9.53×10^{-5}
8	On a tour group	Koala Centre at 3.47 pm, 1 hour. The Nobbies at 5.10 pm, 2 hours. Penguin Parade at 7.55 pm, 2 hours.	5.03×10^{-6}
9	Half day	Koala Centre at 3.12 pm, 1 ½ hours. The Nobbies at 5.15 pm, 2 hours. Penguin Parade at 7.34 pm, 2 ½ hours.	1.22×10^{-9}
10	Partner with children	Churchill Island at 12.34 pm, 2 ¼ hours. Koala Centre at 3.00 pm, 1 ½ hours. The Nobbies at 4.53 pm, 1 ½ hours. Penguin Parade at 6.37 pm, 3 hours.	6.83×10^{-10}

Another type of direct comparison that can be made is by comparing itineraries that have different numbers of associations of characteristics, but both with the same combination of characteristics, for example 'male and young' (1st record of Table 7.11) against 'male, young and international' (3rd record of Table 7.11) or 'inexperienced and preferred wildlife' (6th record of Table 7.11) against 'inexperienced, preferred wildlife and with partner' (7th record of Table 7.11).

7.2.5 Discussion and Evaluation of PSM

There is a clear distinction that highlights different preferences when comparison is made for 'male and young' (1st record of Table 7.11) against 'female and middle aged' (2nd record of Table 7.11) tourists' itineraries where male and young tourists visit the Koala Centre and Churchill Island while female and middle aged tourists visit the Information Centre and Rhyll Inlet. Both itineraries visited The Nobbies and Penguin Parade as anticipated.

The finding of the 'male and young' tourists itinerary highlights the dissimilarity against the findings by Kale et al. (1987) where they mentioned tourists aged between of 18 and 35 should be offered a tour package with non-stop activities. This might be due to the differences of the characteristics of dataset used by their research as compared to this research. However, a finding of the 'male, young and international' tourists' itinerary (3rd record of Table 7.11) confirmed their statement where obviously activities are non-stop with a tight schedule (Kale et al., 1987).

Next, the finding of 'male, young and international' tourists' itinerary confirmed a statement by Abdul Khanan et al. (2012) where the itinerary contains at least one attraction associated with a woodlands theme which is Rhyll Inlet that is popular with a mangrove boardwalk. This highlighted the affinity of male tourists towards the woodlands' theme attraction.

On the other hand, the finding of attractions for 'female, middle aged and local' tourists' itinerary (4th record of Table 7.11) includes Koala Centre and Penguin Parade where these attractions are categorised as the wildlife type of attraction. This

has again confirmed the findings of Abdul Khanan et al. (2012) that female tourists shown affinity towards a wildlife theme. Notice that, the itinerary for 'male, young and international' tourists includes a large number of attractions (six) as compared to 'female, middle aged and local' tourists' itinerary.

The phenomenon of attractions similarity was observed when comparing itineraries for both, experienced (5th record of Table 7.11) and inexperienced tourists (6th record of Table 7.11). A distinction between both types of tourists is the emergence of the Information Centre as one of the attractions visited by inexperienced tourists. In fact, the other 'inexperienced' tourists' itinerary (7th record of Table 7.11) contains the Information Centre as well.

This is understandable as the first time tourists seek information about tour offerings on Phillip Island. Similarly the 'male, young and international' tourists' itinerary (3rd record of Table 7.11), and the 'inexperienced, preferred wildlife and with partner' tourists' itinerary (7th record of Table 7.11) highlighted the large number of attractions where each itinerary contains six attractions.

This is unsurprising as the majority of the overseas tourists (131 of 151 tourists) have no experience of visiting Phillip Island and inexperienced tourists want to enjoy more by visiting as many attractions as they can. Both itineraries almost share the same attractions except the 'male, young and international' itinerary includes Rhyll Inlet while the 'inexperienced, preferred wildlife and with partner' itinerary contains the Information Centre.

Another interesting output is related to the tour group tourists' itinerary (8th record of Table 7.11) where it starts late at 3.47 pm as a result of most of the tour groups departing from Melbourne during the early afternoon. This is based on observations during data collection. Since they arrived at a later time, only three important attractions, all managed by the Phillip Island Nature Park authority were covered as shown in Table 7.11. They are Koala Centre, The Nobbies and Penguin Parade. This highlighted the prominence of these attractions.

It is also worthwhile to note the late arrival time of 7.55 pm at Penguin Parade for tour group tourists. Unlike other types of tourists, tour groups have their own dedicated lane, therefore not requiring them to join the long queue to enter Penguin Parade. Another anticipated finding was the higher number of tourists representing elderly tourists (23 tourists) that join the tour groups (53 tourists). This confirms the findings of Thomson and Pearce (1980) who mentioned that elderly tourists were keener than other age groups to join tour groups. On the other hand, a half day tourists' itinerary (9th record of Table 7.11) had a similar number of attractions as the tour group itinerary but with difference in terms of its schedule.

Finally, the partner and children tourists' itinerary again highlighted the prominence of attractions managed by the Phillip Island Nature Park authority, i.e. Churchill Island, Koala Centre, The Nobbies and Penguin Parade that are visited by them. In terms of scheduling, the arrival time of 12.34 pm at Churchill Island for the partner and children tourists' itinerary is significant where they arrive exactly prior to the wagon ride activity that is organised at 1.00 pm.

In conclusion, it can be seen that a variety of itineraries was produced for a combination of diverse tourists' characteristics. This shows that the initial step of specifying the spatio-temporal movement dataset according to selected tourists' characteristics is efficiently able to produce personalised tour packages.

As anticipated, the movement sequence and the transitional arrival times were able to be quantitatively assigned using the semi-Markov process. However, some of the transitional arrival times are not able to be quantified using the PDF; therefore a simple probability estimation was put in place.

The validation process for the spatial movement transition probability model using Chi-Square goodness of fit test and transitional arrival time probability model using the AIC showed that both models are able to produce reliable results. The evaluation method of checking the consistency of each itinerary produces itineraries that have reliable timing. Finally, the effort of inserting other tour package elements such as

routes, activities and budget has transformed these itineraries into a complete tour package as defined in Section 4.6.

7.3   The Activity Planning Method (APM)

This section provides the implementation and evaluation of the APM to plan the timing of activities at each attraction. Churchill Island is the focus of the APM's implementation where it represents the attraction with the most scheduled activities (Table 7.12). This suits the Activity Timing Comparison Analysis (ATCA) and the Activity Frequency Comparison Analysis (AFCA) well in terms of comparisons that are made against actual timing and frequency of activities.

Table 7.12   Figures of scheduled and unscheduled activities for each attraction in Phillip Island

No.	Attraction	Scheduled Activities	Unscheduled Activities
1	Penguin Parade	2	1
2	Koala Centre	0	4
3	The Nobbies	1	3
4	Churchill Island	7	3
5	Woolamai	0	2
6	Rhyll Inlet	0	2
7	Cowes	0	3
8	Visitors Centre	0	1
9	Ventnor	0	1

7.3.1   Deriving Modes and Several Other Arrival Times and Durations

An arrival time dataset of Churchill Island was fitted to each distribution to identify which distribution produces the smallest figure of AIC. Table 7.13 shows the distribution with the minimal figure of AIC; hence the best fit is the Normal 5 Mixture distribution, representing the best goodness of fit while implementing a reasonable number of parameters that do not overfit the arrival time dataset of Churchill Island.

The Normal 5 Mixture multimodal distribution fit of arrival times for Churchill Island is shown as Figure 7.7. The mode of arrival time, i.e. the most popular arrival time for Churchill Island is at 10.30 am. It is forecast that the most number of tourists

arrive at this time. Other significant arrival times are 1.15 pm, 11.10 am, 12.20 pm and 2.30 pm.

Table 7.13   Five distributions with the smallest AIC
for the Churchill Island Arrival Time Dataset

Churchill Island	Normal 5 Mixture	Log-Normal	Gamma	Normal	Weibull
AIC	8007.3	8494.37	8497.7	8510.97	8556.59
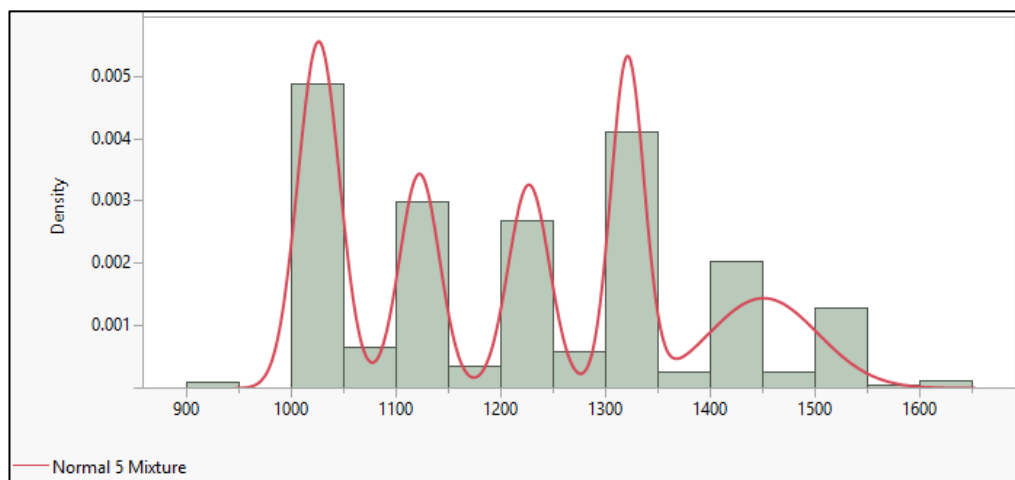
Figure 7.7   Significant arrival times for Churchill Island

On the other hand, the mode of visiting duration, i.e. the most popular visiting duration for Churchill Island is 2 hours as shown in Table 7.14. This figure represents 29.27 % of the cohort of 656 tourists visiting Churchill Island, derived by using a simple probability as no significant distribution is found using the Churchill Island visiting durations' dataset. Other noticeable visiting durations are 1 hour 30 minutes, 1 hour, 1 hour 45 minutes and 1 hour 15 minutes. These visiting durations were rounded to the nearest 15 minutes.

Table 7.14   Significant visiting durations for Churchill Island

No.	Visiting Duration	Count of Tourists	Percentage
1	2 hours	192	29.27
2	1 hour 30 minutes	147	22.41
3	1 hour	101	15.4
4	1 hour 45 minutes	68	10.37
5	1 hour 15 minutes	60	9.15

The above procedure was replicated using the arrival time dataset of Koala Centre, representing an attraction with no scheduled activities. The Normal 5 Mixture distribution was again implemented towards the arrival time dataset due to its minimal AIC figure when compared with other distributions shown in Table 7.15.

Table 7.15   Five distributions with the smallest AIC

for Koala Centre Arrival Time Dataset

Churchill Island	Normal 5 Mixture	Log-logistic	Extreme Value	Weibull	Gamma
AIC	8574.31	8652.9	8655.56	8698.21	8720.26

Again, this represents the best goodness of fit, where at the same time implementing a reasonable number of parameters that do not overfit the arrival time dataset of Koala Centre.

For comparison, Figure 7.8 and Table 7.16 depict the multimodal distribution fit of arrival times and significant visiting durations for the Koala Centre. A simple probability is used to derive the significant visiting durations, including the mode of visiting durations due to the absence of significant distribution when dealing with the Koala Centre visiting durations' dataset.
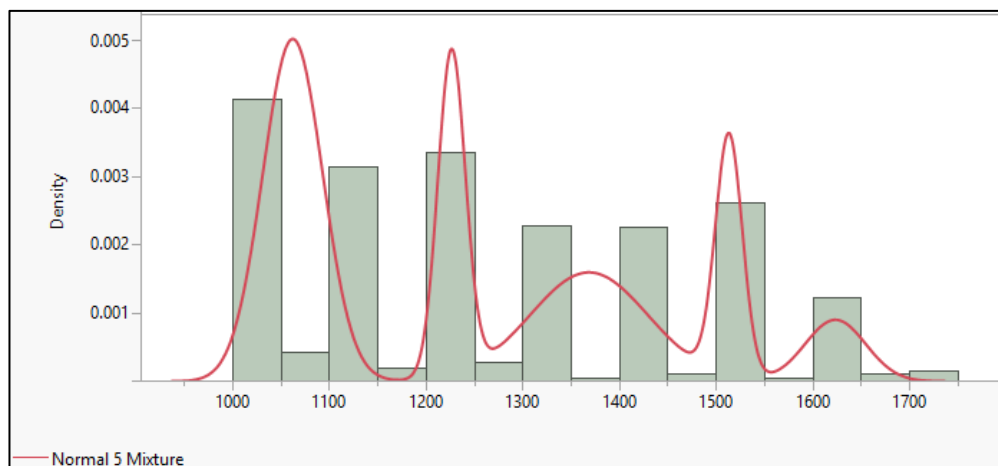
Figure 7.8   Significant arrival times for Koala Centre

Table 7.16   Significant visiting durations for Koala Centre

No.	Visiting Duration	Count of Tourists	Percentage
1	1 hour 30 minutes	263	33.38
2	1 hour	256	32.49
3	1 hour 15 minutes	130	16.5
4	2 hours	91	11.55
5	45 minutes	17	2.16

The mode of arrival time, i.e. the most popular arrival time from Figure 7.8 for The Koala Centre is at 10.40 am. Other significant arrival times are 12.15 pm, 3.10 pm, 1.35 pm and 4.10 pm. Conversely, the mode of visiting duration, i.e. the most popular visiting duration from Table 7.16 is 1 hour 30 minutes, representing 33.38 % of the cohort of 788 tourists visiting The Koala Centre.

The APM procedure in Section 7.3.1 is repeated for the remaining attractions. The output of this procedure is presented in Table 7.17 that shows the most popular arrival times and visiting durations for the entire attraction in Phillip Island.

Table 7.17   The most popular arrival times and
visiting durations for each attraction

No.	Attraction	Arrival time	Visiting duration
1	Woolamai	4.20 pm	1 hour
2	Information Centre	9.25 am	15 minutes
3	Churchill Island	10.30 am	2 hours
4	Koala Centre	10.40 am	1 hour 30 minutes
5	Rhyll Inlet	4.25 pm	1 hour
6	Cowes	12.20 pm	1 hour
7	Ventnor	5.30 pm	30 minutes
8	The Nobbies	5.10 pm	1 hour 30 minutes
9	Penguin parade	7.00 pm	3 hours

7.3.2 Activity Timing Comparison Analysis (ATCA)

Table 7.18 shows Churchill Island's actual activities' starting times for scheduled activities together with the actual visiting durations for both, scheduled and unscheduled activities arranged by Phillip Island Nature Park authority. They were also confirmed by the researcher during the data collection process.

Table 7.18   Churchill Island's actual activities

starting times and visiting durations

No.	Activity	Starting time	Visiting duration
1	Cow milking	11.00 am, 2.10 pm	3 minutes
2	Wagon ride	1.00 pm	5 minutes
3	Sheep shearing	2.25 pm	10 minutes
4	Whip cracking	2.40 pm	5 minutes
5	Working dogs	3.05 pm	10 minutes
6	Blacksmith demo	3.20 pm	10 minutes
7	Boomerang throwing	3.35 pm	5 minutes
8	Visiting baby animal	Unscheduled	15 minutes
9	Visiting heritage buildings and gardens	Unscheduled	30 minutes
10	Animal viewing	Unscheduled	20 minutes
	Total recommended visiting durations		**1 hour 53 minutes**

When comparison is made between the findings of arrival times of Churchill Island (Figure 7.7) and these actual activities' starting times (Table 7.18), it can be observed that three of the arrival times, which are 1.15 pm, 11.10 am and 2.30 pm from Figure 7.7 almost correspond with the activities timing of wagon ride, cow milking and sheep shearing.

On the other hand, the total actual visiting duration of 1 hour 53 minutes from Table 7.18 almost matches the most popular visiting duration of 2 hours (Table 7.14). Furthermore, the total actual visiting duration also sits between the two highest occurrences of significant visiting durations which are 2 hours and 1 hour 30 minutes from Table 7.14.

Finally, attraction administrators can use the result of the most popular arrival times as the starting time of any activities such as at 10.30 am (see Figure 7.7). From there they can schedule the entire activities non-stop for 2 hours (see Table 7.14 for the mode of visiting duration). This is able to ensure optimum participation of tourists as the most popular arrival times and visiting durations are in used.

7.3.3   Activity Frequency Comparison Analysis (AFCA)

Notice that in Table 7.18 all scheduled activities are conducted once daily except cow milking, which is held twice at 11.00 am and 2.10 pm. To accommodate various tourist arrival times shown in Figure 7.7 (that also highlights the distribution of arrival times), it is suggested that these arrival times are adopted and with it, comes the increment of frequency.

If attraction administrators want to adopt the mode of arrival time at 10.30 am alongside with the second and third highest occurrences of arrival times (Figure 7.7) as the starting time of non-stop activities for 2 hours (Table 7.14) as stated in the previous sub-section, the frequency of activities is automatically increased to three for each of those activities.

This has the possibility to increase the number of visitors who previously are not able to visit because the starting times of the activities do not match their timing availability for visitation. During school holidays, attraction administrators can also adopt all five arrival times (Figure 7.7) so as to increase the frequency of activities frequency to five times as a response to the higher number of tourists during this period.

Figure 7.7 shows the pattern of distribution of the most popular arrival times across the entire attractions in Phillip Island. Notice that from Figure 7.7, the majority of arrival times for The Nobbies and Penguin Parade are as anticipated where most tourists arrive later during the evening. Furthermore, the arrival time unimodal distribution fit for the Penguin Parade shows the prominence of the penguin viewing activity where the arrival time uniformity produces the unimodality scenario. This output is different from Churchill Island, The Koala Centre and Cowes which produced multimodality distribution. In these attractions, tourists' arrival times vary wherein the specific case of Churchill Island, tourists might choose their different arrival time to coincide with their preferred activities.

As depicted in Table 7.17, the majority of tourists arrived at Penguin Parade at 7.00 pm to coincide with the penguin viewing activity actually scheduled at 8.45 pm. Usually they have to arrive significantly earlier to be among the first in the long queue. As The Nobbies is near to Penguin Parade, most tourists chose to arrive at 5.10 pm where after finishing their visit at The Nobbies, they can immediately go to Penguin Parade.

7.3.4   Discussion and Evaluation of APM

Based on Table 7.17, the mode of arrival time at the Information Centre is at 9.25 am. This highlighted that most tourists who came to the Information Centre, visited it as the first attraction of the day to retrieve information for planning their tour. There are also attractions where the arrival time varies during the day, according to Figure 7.9. One example is Cowes where, although most tourists arrive at 12.20 pm for lunch, based on observations, there are tourists who came later during the evening to have their dinner. Cowes is the township of Phillip Island, where businesses and restaurants are mostly located.

In terms of visiting duration, the prominence of Penguin Parade where tourists stay the longest arises. As noted earlier, most tourists prefer to come early at 7.01 pm (refer Table 7.17) to join the queue, although the penguin viewing activity only started at 8.45 pm. The majority of tourists who visited the Penguin Parade stayed for three hours, based on Table 7.17.

According to Table 7.17, most tourists who visited the Information Centre only spent 15 minutes to retrieve information for planning their tour. It is also important to note that the entire attractions managed by Phillip Island Nature Park authority that includes the Churchill Island, Koala Centre, The Nobbies and Penguin Parade have their mode of visiting duration of at least 1 ½ hours, based on Table 7.17. This is due to the nature of activities within this attraction that are well defined and properly managed.
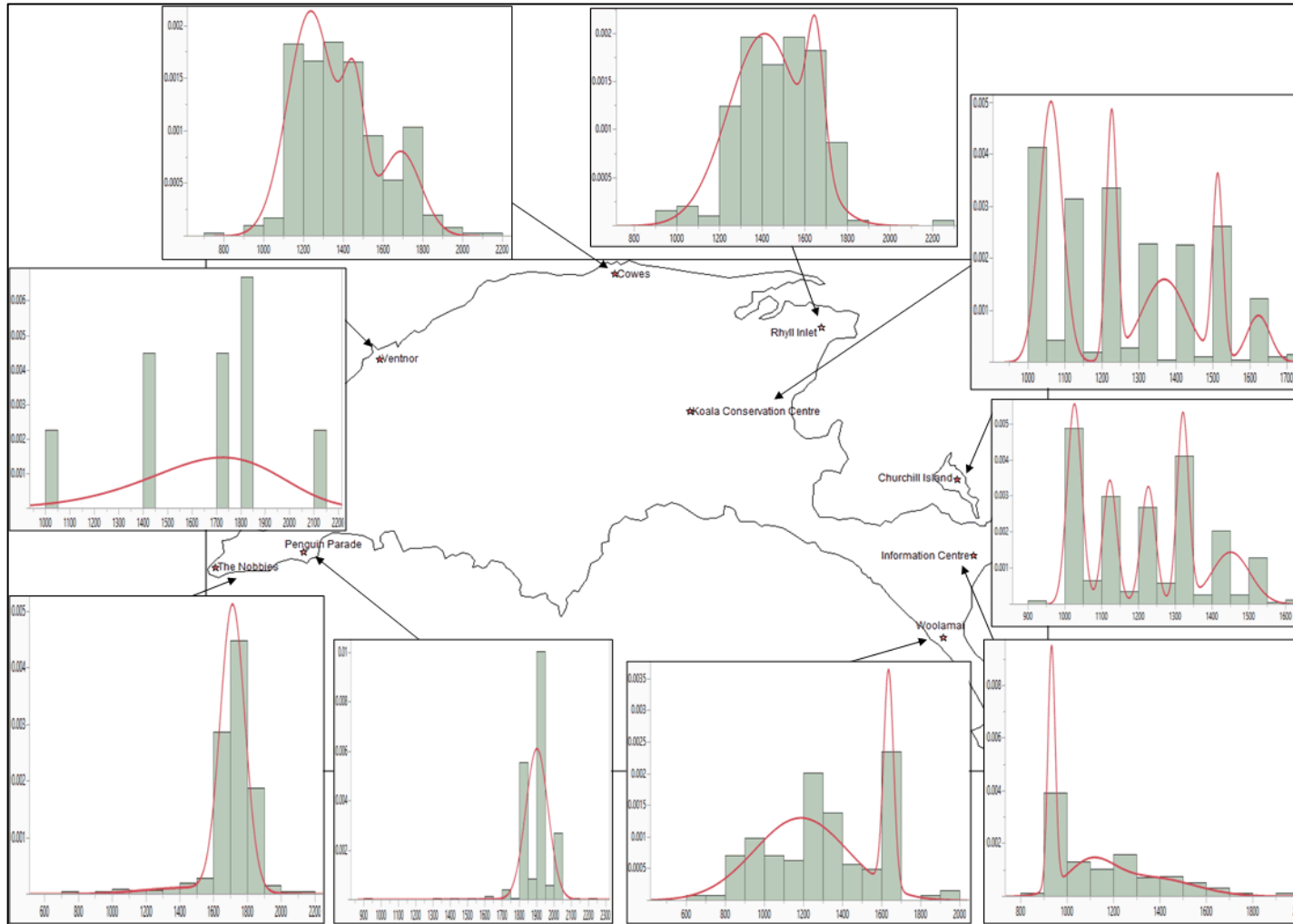
Figure 7.9   Distribution of arrival times across the entire attractions

To conclude, various suggestions of arrival times and durations were estimated for each tourist attraction. As anticipated, the most popular arrival time and duration were estimated for each attraction. Some attractions, such as Churchill Island and Koala Centre highlighted several suggestions of arrival times (multimodality scenario) while attractions such as The Nobbies and Penguin Parade, highlighted a unimodality scenario where the majority of tourists arrived at a single arrival time. These can be observed from Figure 7.9.

The PDF has been proven to be sufficient in suggesting not just the most popular arrival times, but also other significant arrival times, as shown in Figures 7.7 and 7.8 in Section 7.3.1. This provides flexibility for attraction administrators in setting up several starting times for similar activities. The emergence of multimodal PDF when compared to unimodal PDF in fitting the arrival times dataset was highlighted where eight out of nine PDFs are multimodal PDFs. However, similar with PSM, the entire visiting durations were not able to be estimated using the PDF, instead by using simple probability.

The selection and validation process using the AIC, as shown in Tables 7.13 and 7.15 in Section 7.3.1 proved that the model provides reliable results. The evaluation methods of ATCA and AFCA (Sections 7.3.2 and 7.3.3) show that the suggestion of several arrival times and durations presented advantage against the current timing arrangement for current activities. If these multiple arrival times are adopted by attraction administrators as shown in Section 7.3.3, this provides flexibility for tourists to arrive at their preferred time.

Tourists who are not able to visit attractions because they miss the activities which are only conducted once, can now choose their preferred arrival times as a result of activities being conducted more frequently. This is able to increase the visitation rate.

7.4   The Market Segmentation Method (MSM)

A 'CDHIO' itinerary is chosen for the market segmentation of itineraries. This represents the most popular itinerary, derived using the PSM, when a complete dataset of tourists' spatio-temporal movement unspecific to any tourists'

characteristics were used. In terms of the market segmentation of attraction and activities, Churchill Island and wagon ride are chosen. Churchill Island is chosen as it is one of the attractions included in the 'CDHIO' itinerary with the most comprehensive set of activities, as shown in Table 7.12. Furthermore, the dataset in terms of activity participation for other attractions are insufficient.

The selection of the wagon ride for the market segmentation of activities is due to its popularity when compared with other scheduled activities conducted on Churchill Island as shown in Table 7.19. Notice that the entire selection of tour package products for market segmentation that involves solely Churchill Island is purposely made to look into the relationship that exists between them, shown as double pointing arrows in Figure 6.1. This is done in Section 7.4.5.

The selection of the 'CDHIO' itinerary, Churchill Island and wagon ride do not limit the MSM to be applied for other itinerary, attraction and activities. For clarification, this MSM can be implemented as long as the other tourism products contain a set of tourists' characteristics with its respective classifications. Therefore, the 'CDHIO' itinerary, Churchill Island and wagon ride are sufficient to represent all possible tourism products.

Table 7.19   Activities at Churchill Island with
its respective numbers of participants

No.	Activity	Number of participants
1	Blacksmith demonstration	76
2	Cow milking	81
3	Sheep shearing	65
4	Wagon ride	83
5	Whip cracking	80
6	Working dogs	77

7.4.1   The Tour Package Market Segmentation

A total of 65 tourists travelling using the 'CDHIO' itinerary are automatically classified by the Weka modelling software into four clusters (Witten and Frank, 2005). These clusters are based on one geographic characteristic (origin), four demographic characteristics (gender, age, household income and level of education),

two psychographic characteristics (reason of tour and favourite type of attraction) and four behavioural characteristics (experience, travel group, overall duration and transportation mode). These four clusters constitute the columns while the characteristics with its classifications form the rows in Table 7.20.

Table 7.20   Market segmentation of itinerary 'CDHIO'

Characteristics	Classification	Cluster			
		1 25%	2 26%	3 32%	4 17%
Experience	No experience	3.6493	17.7008	21.6289	4.021
	Experienced	14.8532	1.0064	1.2403	8.9001
	Total	18.5025	18.7072	22.8691	12.9211
Reason of tour	Holiday	12.5134	16.7008	19.8022	7.9836
	Education	2.1147	1.0013	1.0589	1.8251
	Business	1.0017	1.9999	1.0095	1.9889
	Accompany friend/relative	4.8728	1.0052	2.9986	3.1234
	Total	20.5025	20.7072	24.8691	14.9211
Favourite type of attraction	Wildlife	7.094	15.7497	11.1632	5.993
	Adventure	1.8399	1.0074	3.1424	2.0103
	Sightseeing	7.4562	1.9525	8.7717	5.8197
	Heritage	3.1492	1.9975	1.791	1.0622
	Culture	1.9633	1	1.0008	1.0359
	Total	21.5025	21.7072	25.8691	15.9211
Travel group	Friend/relative	5.0207	1.971	10.8892	5.1191
	Partner with children	12.6437	1.0628	9.5099	7.7836
	Alone	1.0037	1.0014	1.9818	1.0131
	Partner only	1.8316	1.7587	2.4055	1.0042
	Tour group	1.0027	15.9132	1.0828	1.0012
	Total	21.5025	21.7072	25.8691	15.9211
Overall duration	2-3 day	9.0851	1.0397	7.953	6.9222
	4-5 day	1.0056	1.005	2.027	1.9623
	Full day	7.4709	3.9947	13.6051	4.9293
	Half day	1.9914	14.6631	1.2618	1.0837
	More than a week	1.9496	1.0045	1.0223	1.0236
	Total	21.5025	21.7072	25.8691	15.9211
Gender	Female	12.1385	10.9248	6.712	5.2248
	Male	6.364	7.7824	16.1572	7.6964
	Total	18.5025	18.7072	22.8691	12.9211
Age	Middle aged	2.5702	7.0084	13.8703	11.5511
	Young	14.945	7.7795	8.9889	1.2866
	Old	1.9873	4.9193	1.01	1.0835
	Total	19.5025	19.7072	23.8691	13.9211
Origin	Overseas	4.3804	15.7827	21.67	1.1668
	Australia	14.1221	2.9245	1.1991	11.7543
	Total	18.5025	18.7072	22.8691	12.9211
Household	Low	3.0019	12.7122	9.1983	1.0876

income	High	15.4005	4.9775	7.7222	2.8998
	Medium	1.1002	2.0174	6.9486	9.9338
	Total	19.5025	19.7072	23.8691	13.9211
Level of education	Tertiary-postgraduate	10.0766	8.9829	6.9351	6.0054
	Tertiary-undergraduate	7.4211	7.733	11.9371	5.9088
	Secondary-completed	2.0048	2.9912	4.9969	2.007
	Total	19.5025	19.7072	23.8691	13.9211
Transportation mode	Campervan/motor home	3.1383	1.0567	6.7713	1.0338
	4wd	5.9404	1.0158	3.9745	1.0694
	Passenger (Car/van)	10.4212	2.7215	13.0406	11.8167
	Bus	1.0027	15.9132	1.0828	1.0012
	Total	20.5025	20.7072	24.8691	14.9211

For automatically determining the number of clusters, Weka uses a tenfold cross-validation where "beginning with one cluster, Weka continues to add clusters until the estimated log-likelihood decreases" (Beane and Ennis, 1987; Witten and Frank, 2005; He and Zhang, n.d.). Table 7.20 shows the clustering results. Cluster 3 represents the majority of tourists with 32 % of overall tourists while clusters 1, 2 and 4 represent 25 %, 26 % and 17 % each, respectively. This clustering result is validated using the cross-validation procedure as presented in Section 6.5. Notice that, the result with the largest figure of log-likelihood, i.e. the best clustering result out of three separate datasets is Dataset 1 as shown in Table 7.21.

Table 7.21   Comparison of log-likelihood figures across three
separate datasets for tour package segmentation

No.	Dataset	Log-likelihood
1	Dataset 1	-8.1795
2	Dataset 2	-8.5643
3	Dataset 3	-9.0791

The shaded cells depict the significant characteristic types in regards to each cluster. Here, only characteristic types that are truly significant are chosen (i.e: have large differences of count against other characteristic types) or at least double/nearly double when compared with the subsequent highest count are chosen.

Insignificant characteristics of certain clusters are shown as clusters without shades on some of the clusters' characteristics' types. Notice from Table 7.20, 'experience', 'reason of tour', 'origin' and 'transportation mode' are significant characteristics for clustering the tourists. This is reflected in the table where each cluster has shaded cells for these characteristics.

Refer again to Table 7.20, in terms of 'experience' characteristic, each cluster 1 and 4 had 15 (14.8532) and 9 (8.9001) tourists assigned to those clusters. With regards to 'origin' characteristic, clusters 2 and 3 pointed out the domination of overseas tourists while clusters 1 and 4 are closely related to their domestic counterpart. Cluster 2 highlights the prominence of bus as a transportation mode with 16 out of 21 tourists using them as a means of transport around the island whereas in the other three clusters, passenger vehicles such as car or van dominate. In terms of the characteristic of 'reason of tours', all clusters are associated with holiday as the reason. Table 7.22 is a simplified version of Table 7.20 that highlights the types of characteristics associated with each cluster.

Table 7.22   Types of characteristics of tourists
for each cluster of 'CDHIO' itinerary

Cluster	Percentage	Classification of Characteristics
1	25	Experienced, female, young, domestic and high income earner tourists, came for holiday, travel with a partner and children and travel in a passenger vehicle.
2	26	Non-experienced, overseas and low income earner tourists, came for a holiday, prefer 'wildlife' attractions and travel with a half day tour group in a bus.
3	32	Non-experienced, male, overseas and tertiary undergraduate tourists, came for a holiday, a full day visit and travel with a passenger vehicle.
4	17	Experienced, middle aged, domestic, medium income earner tourists, came for holiday and travel with a passenger vehicle.

7.4.2   The Attractions Market Segmentation

A total of 457 tourists who visited Churchill Island are automatically classified by the Weka modelling software, using the same procedure described in Section 7.4.1 into six clusters (Witten and Frank, 2005). This is shown in Table 7.22 containing

similar tourists' characteristics used in the itinerary's market segmentation. The structure of Table 7.23 is similar to Table 7.20 in regards to its rows and columns. Notice that, the increasing number of tourists also increases the number of clusters. Table 7.23 depicts the output of clustering. Cluster 6 represents the majority of

Table 7.23   Market segmentation of Churchill Island attraction

Characteristics	Classification	Cluster					
		1	2	3	4	5	6
		20%	5%	27%	5%	11%	32%
Experience	No experience	21.76	25.405	107.8	1.622	5.2098	138.2
	Experienced	73.374	1.0762	15.318	21.58	48.903	8.7506
	Total	95.133	26.482	123.12	23.201	54.113	146.95
Reason of tour	Holiday	45.08	23.41	103.69	21.68	41.273	132.87
	Education	2.0129	1.0139	5.3268	1.006	5.0821	5.5583
	Accompany friend/ relative	46.645	1.0064	15.026	1.1144	8.3956	8.8125
	Business	1.0037	3.0001	1.0051	1.0006	1.2854	1.7052
	Others	3.3913	1.0516	1.0772	1.4	1.0769	1.0029
	Total	98.133	29.482	126.12	26.201	57.113	149.95
Favourite type of attraction	Wildlife	33.113	20.286	81.112	2.4676	29.929	82.093
	Sightseeing	39.59	4.1685	24.978	11.461	16.369	43.434
	Heritage	17.367	3.0238	5.4302	4.1393	6.3411	13.699
	Adventure	2.7993	1.0021	12.326	7.062	1.4683	8.3423
	Culture	5.2641	1.0016	2.2743	1.0716	3.0056	2.3828
	Total	98.133	29.482	126.12	26.201	57.113	149.95
Travel group	Friend/ relative	80.087	1.1448	51.447	2.2857	5.6521	27.383
	Partner with children	2.0885	1.036	17.33	1.8611	45.491	51.193
	Alone	1.2734	1.0694	13.127	1.2169	1.0887	2.2247
	Partner only	12.029	1.0885	42.873	17.618	3.5555	66.836
	Tour group	2.633	25.143	1.3399	3.2172	1.1414	1.5257
	Total	99.133	30.482	127.12	27.201	58.113	150.95
Overall duration	2-3 day	29.499	1.0201	38.168	12.564	18.259	68.49
	4-5 day	7.1515	1.0684	7.3345	2.074	1.6322	13.74
	Full day	49.372	7.2062	62.805	6.2442	29.651	53.722
	Half day	4.0701	19.167	14.852	2.9291	1.9468	3.0359
	More than a week	3.5812	1.016	2.4053	1.8985	1.124	7.975
	6-7 day	5.4604	1.004	1.5563	1.4911	5.4995	3.9887
	Total	99.133	30.482	127.12	27.201	150.95	58.113
Gender	Female	51.542	13.31	45.518	6.0726	30.443	61.115
	Male	43.591	13.172	77.603	17.129	23.67	85.837
	Total	95.133	26.482	123.12	23.201	54.113	146.95
Age	Middle aged	37.156	10.127	14.959	3.8388	27.986	72.934
	Young	42.386	11.285	103.96	5.117	24.776	54.475
	Old	16.59	6.0707	5.1999	15.245	2.3515	20.542

	Total	96.133	27.482	124.12	24.201	55.113	147.95
Origin	Overseas	5.4759	23.349	89.754	2.3828	5.3945	106.64
	Australia	89.657	3.1325	33.367	20.818	48.718	40.307
	Total	95.133	26.482	123.12	23.201	54.113	146.95
Household income	Low	34.523	17.204	105.59	5.7025	4.2132	7.7689
	High	17.364	6.049	2.1812	9.5229	31.761	85.122
	Medium	44.246	4.2288	16.351	8.9757	19.138	55.06
	Total	96.133	27.482	124.12	24.201	55.113	147.95
Level of education	Tertiary-postgraduate	16.084	10.136	18.905	5.3409	30.783	59.751
	Tertiary-under gra-duate	44.525	11.291	56.548	6.0267	20.453	66.156
	Secondary-completed	26.148	6.0481	43.905	12.194	2.783	21.922
	Secondary-partial	10.374	1.0013	4.8292	1.64	2.0925	1.0627
	Primary	1.0011	1.0046	1.9331	1.0001	1.0018	1.0593
	Total	98.133	29.482	126.12	26.201	57.113	149.95
Transpor-tation mode	Campervan/ motor home	1.7003	1.0835	11.78	3.2415	6.9681	43.226
	4wd	20.876	1.0439	24.836	5.3331	14.295	24.616
	Passenger (Car/van)	70.307	1.2381	81.582	13.283	33.672	79.92
	Bus	2.5881	25.047	1.0075	3.2157	1.1331	1.0085
	Walking	1.7001	1.0229	6.0673	1.1148	1.044	1.0509
	Motorcycle	1.9626	1.0462	1.8479	1.0134	1.0005	1.1294
	Total	99.133	30.482	127.12	27.201	58.113	150.95

tourists with 32 % of overall tourists while clusters 1, 2, 3, 4 and 5 represent 20 %, 5 %, 27 %, 5 % and 11 % each. This clustering result is again validated using the cross-validation procedure as presented in Section 6.5. Notice that, the result with the largest figure of log-likelihood, i.e. the best clustering result out of three separate datasets is Dataset 1 as shown in Table 7.24.

Table 7.24   Comparison of log-likelihood figures across three separate datasets for attraction segmentation

No.	Dataset	Log-likelihood
1	Dataset 1	-10.06545
2	Dataset 2	-10.12312
3	Dataset 3	-20.45637

The criteria for choosing the significant characteristics' types are similar as Section 7.4.1. From Table 7.24, it can be observed that 'experience', 'overall duration', 'origin' and 'transportation mode' are important characteristics for clustering the

tourists within the attraction context. The entire clusters do have associations with one of the characteristic types of these characteristics.

In regards to 'experience' characteristic, clusters 2, 3 and 6 are associated with non-experienced tourists, each with 25, 108 and 128 tourists falls under this characteristic's type while the other clusters have associations with experienced tourists with 73, 22 and 49 tourists for each of the clusters 1, 4 and 5. Another significant characteristic that emerges is the 'overall duration' where most of the clusters are linked to full day tourists except for clusters 2 and 4 where each of the clusters involves half day and 2-3 day tourists.

Corresponding with the 'experience' characteristics, the similar clusters of 2, 3 and 6 are identified as overseas tourists' clusters while clusters 1, 4 and 5 are related to domestic tourists. This is a logical finding as overseas tourists have lower probability in prior visits to Churchill Island, and thus Phillip Island during the earlier years when compared to domestic tourists. In terms of transportation mode, almost all clusters have their tourists travelling in a passenger vehicle except cluster 2 where these tourists are touring on a bus. Cluster 2 is also defined as a tour group cluster that had a half day tour. Table 7.25 is a simplified version of Table 7.23 that highlights the types of characteristics associated with each cluster.

Table 7.25   Tourists' characteristics' types for each
cluster of Churchill Island's attractions

Cluster	Percentage	Classification of Characteristics
1	20	Experience, domestic and tertiary undergraduate tourists, travel with friends/relative for full day in a passenger vehicle.
2	5	Non-experience, over sea and low income earner tourists, came for a holiday in a tour group for half day in a bus, prefer 'wildlife' attractions.
3	27	Non-experience, male, young, over sea and low income earner tourists, came for a holiday and prefers 'wildlife' attractions, a full day visit and travels with a passenger vehicle.
4	5	Experience, male, old, domestic and secondary-completed tourists, came for holiday with only a partner, spend 2-3 days and travels with a passenger vehicle.
5	11	Experience and domestic tourists who travel with a partner and children for a full day in a passenger vehicle, came for a

		holiday and prefers wildlife attractions.
6	32	Experience, over sea and high income earner tourists, came for a holiday, prefers wildlife attractions and travel on a passenger vehicle for a full day.

7.4.3   The Activities Market Segmentation

A total of 83 tourists who participated in the wagon ride activity are automatically classified by the Weka modelling software, using the same procedure described in Section 7.4.1 into two clusters (Witten and Frank, 2005). This is shown in Table 7.26 containing similar tourists' characteristics used in the itinerary's market segmentation.

Table 7.26   Market segmentation of wagon ride activity

Characteristics	Classification	Cluster	
		1 64%	2 36%
Experience	Experienced	24.8308	13.1692
	No experience	29.8991	19.1009
	Total	54.7299	32.2701
Reason of tour	Accompany friend/relative	3.1974	7.8026
	Holiday	50.5255	20.4745
	Education	1.9987	4.0013
	Business	1.0083	1.9917
	Total	56.7299	34.2701
Favourite type of attraction	Wildlife	19.4819	10.5181
	Heritage	11.6641	16.3359
	Sightseeing	21.5707	3.4293
	Culture	3.9856	1.0144
	Adventure	1.0276	3.9724
	Total	57.7299	35.2701
Travel group	Partner with children	45.7234	6.2766
	Friend/relative	1.5098	23.4902
	Partner only	8.4967	3.5033
	Total	55.7299	33.2701
Overall duration	Full day	28.2616	6.7384
	2-3 day	18.5958	13.4042
	4-5 day	3.004	1.996
	Half day	4.8552	8.1448
	6-7 day	1.0195	4.9805
	More than a week	2.9937	1.0063
	Total	58.7299	36.2701
Gender	Female	23.5489	12.4511
	Male	31.1809	19.8191

	Total	54.7299	32.2701
Age	Middle aged	31.575	9.425
	Young	22.1334	19.8666
	Old	2.0215	3.9785
	Total	55.7299	33.2701
Origin	Australia	33.9266	13.0734
	Overseas	20.8033	19.1967
	Total	54.7299	32.2701
Household income	Medium	23.7249	6.2751
	High	29.4524	4.5476
	Low	2.5526	22.4474
	Total	55.7299	33.2701
Level of education	Tertiary-postgraduate	22.6321	5.3679
	Tertiary-undergraduate	25.4631	17.5369
	Secondary-partial	1.3218	5.6782
	Secondary-completed	7.3129	5.6871
	Total	56.7299	34.2701
Transportation mode	Passenger (Car/van)	31.4439	25.5561
	4wd	18.3299	3.6701
	Campervan/motor home	5.9549	1.0451
	Walking	1.0007	2.9993
	Bus	1.0005	1.9995
	Total	57.7299	35.2701

The structure of Table 7.26 is similar to Table 7.20 in regards to its rows and columns. Notice that, the decreasing number of tourists also reduces the number of clusters. Table 7.26 depicts the output of clustering. Cluster 1 represents 64 % of overall tourists while Cluster 2 represents 36 % tourists. This clustering result is validated using the cross-validation procedure as presented in Section 6.5. Notice that, the result with the largest figure of log-likelihood, i.e. the best clustering result out of three separate datasets is Dataset 1 as shown in Table 7.27.

Table 7.27   Comparison of log-likelihood figures across three
separate datasets for activity segmentation

No.	Dataset	Log-likelihood
1	Dataset 1	-9.71178
2	Dataset 2	-9.82316
3	Dataset 3	-10.45321

Once more, the criteria for choosing the significant characteristics' classifications are as same as the previous two sub-sections. Table 7.26 shows that the' reason of tour', 'travel group' and 'overall duration' emerged as significant characteristics when it comes to clustering within the activity context. In terms of 'reason of tour' characteristics, both clusters clearly show that tourists' reasons for touring is for holiday where the first and second clusters have each, 51 and 20 tourists touring for the respective reasons.

The market segmentation for activity highlighted the importance of 'travel group' characteristics. In terms of travel groups, most of the tourists (46 out of 56) in the first cluster are tourists who travel with a partner and their children, whereas 23 out of 33 tourists in the second cluster travel with either friends or relatives.

A distinction of the characteristic type in terms of 'overall duration' characteristics once more emerges. In the first cluster, the majority of 28 from 59 tourists are touring for a full day while, in the second cluster, the majority of 13 from 36 tourists had a 2 to 3 day tour. Table 7.28 is a simplified version of Table 7.26 that highlights the types of characteristics associated with each cluster.

Table 7.28   Tourists' characteristics' type for each cluster of wagon ride activity

Cluster	Percentage	Classification of Characteristics
1	64	Tourists who come for a holiday with a partner and children having a full day tour.
2	36	Young, low income earner and tertiary undergraduate tourist, who travels for holidays with their friends and/or relatives for a 2-3 day tour, travels using passenger vehicle.

7.4.4   Target Market Selection

Although there are several ways to determine the target market as described in Section 6.6, the selection of cluster with the most user/highest percentage is proposed. This can increase profits based on higher sales (Myers, 1996). For this reason, clusters 3, 6 and 1 from Tables 7.20, 7.23 and 7.26 are each selected for the purpose of tour packages, attractions and activities market segmentation and are shown in Table 7.29.

Table 7.29   Tourists' characteristics types for the entire tour packages' products

Type	Classification of Characteristics
Itinerary CDHIO	Non-experienced, male, overseas and tertiary undergraduate tourists, came for holiday, a full day visit and travel in a passenger vehicle.
Churchill Island visitation	Experienced, overseas and high income earner tourists, came for a holiday, prefers wildlife attractions and travel on a passenger vehicle for a full day.
Wagon ride activity	Tourists who come for a holiday with a partner and children having a full day tour.

7.4.5   Discussion and Evaluation of MSM

From Table 7.29, it can be clearly seen that two characteristic types emerge in the entire tour package products, which are full day tourists and tourists who came for the purpose of the holiday. Therefore, in terms of a full day tourist, tour operators in collaboration with the Churchill Island administrator should pay attention in terms of facilitating them during their tour as they appear to dominate the entire tour package products in relation to Churchill Island.

This can be done by preparing more places to rest within Churchill Island assuming that full day tourists, as they tour for the whole day might get tired during their tour. Larger cafes and a kids' playground should be able to provide a better experience for them, as well.

Other significant characteristics which appear in both the CDHIO itinerary and Churchill Island visitation are overseas tourists and tourists who travel in a passenger vehicle. In regards to tourists who travel in a passenger vehicle, tour operators can advise the administrator of Churchill Island to construct more parking bays to facilitate finding a parking bay. Tour operators should also devise an aboriginal culture presentation during the participation of tourists in their tour package to accommodate overseas tourists.

Tourists with characteristic types that appeared in the Churchill Island visitation, but did not appear on the CDHIO itinerary might mean that they visit other attractions in the itinerary. Concurrently, tourists with characteristic types that appeared in the

Churchill Island visitation, but did not appear in the wagon ride activity, might again mean that they participated in other activities within Churchill Island.

In conclusion, several useful results shown in Tables 7.26, 7.29 and 7.32 as outcomes from three different market segmentations are produced. Although popular tourists' characteristics types such as tourists' age and gender do not come with any significant result as anticipated, the emergence of other characteristics such as 'overall duration' and 'reason of tour', prompted more serious research to be conducted on these characteristics. This can be observed in Table 7.33.

An implementation of the EM algorithm for post-hoc clustering highlighted the advantage of using this algorithm in terms of its efficiency and ease of use without needing us to have prior knowledge of the expected cluster (Everitt, 2011). The validation procedure of cross-validation using log-likelihood, described in Section 6.5 provides convincing results as shown in Tables 7.22, 7.25 and 7.28. The evaluation method of determining target market based on the segment with the most users shown in Section 7.4.4 provides tour operators with marketable tour package products. This can increase the chance of any itinerary, as an example to be purchased.

7.5   Chapter Conclusions

This case study presents the realisation of all three methodologies where each methodology was implemented, validated and evaluated to ensure the accuracy of results produced. The PSM managed to model the spatio-temporal movement of tourists using the semi-Markov and scheduling processes. Multiple itineraries were produced for various characteristics with the most popular one being the 'BEHIO' itinerary for female and middle aged tourists. Tour packages were derived later in Section 7.2.4 by adding other elements of tour package such as route and budget for completing the PSM.

The implementation of the APM in estimating various arrival times and visiting durations within each attraction is notable whereby these outcomes were used to plan the timing of activities for certain attractions. A distinctive set of popular arrival

times and visiting durations were identified for different attractions. The multimodality scenario where several significant arrival times emerged for Churchill Island and The Koala Centre proved that the PDF is flexible in suggesting these arrival times. The ATCA and AFCA where the output of estimating arrival times were compared with the current arrangement of timing highlighted some differences where the output of arrival time suggestion if adopted is possible to increase the visitation rate.

Lastly, the MSM managed to link several target markets of tourists' characteristics' classifications with certain tour packages, attractions and activities. A unique set of tourist characteristics was identified within each target market where later in Section 7.4.5, the relationship between these target markets were ascertained. The selection of target market based on the segments with most users can provide assurance to a tour operator where if these target markets are adopted, the chance of any itinerary to be purchased can increase.

Ultimately, this case study provided the best opportunity to test and verify the theories and concepts developed in each method presented in Chapters 4, 5 and 6 by implementing the case study data. In general, the results provided a good indication of each method in solving related issues highlighted in Chapter 3 based on the discussion and evaluation. The validation procedures for each method also ensured that the results would be as accurate as possible.

CHAPTER 8: CONCLUSIONS AND RECOMMENDATIONS

8.1 Introduction

This concluding chapter covers three important aspects. Primarily, this chapter identifies whether or not the findings of this research meet the objectives and research questions. Some limitation of the theories, methods and the implementation of this research are also addressed. Finally, this chapter, hence the thesis concludes by discussing the future direction of research.

8.2 Did The Research Achieve its Objectives?

The primary aim of this research was to establish a method for developing tour packages intended for tourists by modelling their spatio-temporal movement. This was accomplished through three objectives as stated in Section 1.3. In achieving these objectives, several research questions were put forward.

The first research question dealt with defining a tour package in terms of tourists' spatio-temporal movement. Several existing definitions of tour packages and comparison were made in Chapter 2 in terms of its components. The exact definition was given in Section 2.2 as having a compulsory component of a sequence of attractions, arrival times and visiting durations tailored to various tourist

characteristics for trip planning purposes. The definition was then extended in Section 4.6 to include several more components which are routes, activities and budget.

The second research question dealt with the current issues in designing tour packages in relation to tourists' spatio-temporal movement and their characteristics. As attained in Section 2.5, three issues of tour package design were identified for each tour package stakeholder in this research: (1) tour planning for tourists, (2) activity planning for attraction administrators, and (3) market planning for tour operators. Based on these issues, a tour package design framework was finally established in Section 3.5 by suggesting solutions to each issue that incorporates stakeholder and methodological requirements

The third research question was concerned with the current methods for solving the tour package design issues. Various existing methods that are possible to solve tour package design issues were identified in Section 2.4. These methods were specified into two types of method: (1) methods directly related to tour package design, and (2) methods indirectly related to tour package design. The methods indirectly related to tour package design were analysed based on their potential to design tour packages especially by incorporating tourists' spatio-temporal movement and characteristics.

The fourth research question was concerned with the method to develop spatio-temporal develop itineraries. The Personalised Scheduling Method (PSM) that consists of semi-Markov and scheduling processes were chosen in Section 3.4 and Chapter 4 to develop itineraries. In addition to its ability to distinguish the probability of tourist's transition from one attraction to another, the semi-Markov process is also able to identify the transition with the highest probability, i.e. the most popular movement sequence. Combined with the Probability Density Function (PDF), the semi-Markov process improves the design of itinerary components.

The fifth research question dealt with the method for estimating tourists' arrival times and visiting durations for activity planning at each attraction. The Activity Planning Method (APM) that made use of the PDF was chosen in Section 3.4 and

Chapter 5 to accomplish this task. Besides identifying the mode of arrival times, i.e. most popular arrival times, the PDF is also able to handle the multimodality scenario where other modes of arrival times can also be identified. This provides more flexibility for attraction administrators to arrange their timing of activities.

The sixth research question dealt with the method to produce the target market of tour package products according to tourists' characteristics. The Market Segmentation Method (MSM) that includes the Expectation-Maximisation (EM) algorithm was developed in Section 3.4 and Chapter 6 to derive these target markets. The ability of the EM algorithm to cluster without initially having to know the clusters' parameters, as described in Section 3.4.3 provides the ease of use of this algorithm (Witten and Frank, 2005). Furthermore, the EM algorithm can handle nominal variables. These two advantages have made the process of market segmentation simple and accurate.

The seventh research question was concerned with the possible tourist characteristics for segmenting the market of tour packages' products. The potential tourist characteristics for market segmentation were identified during the step of "determining the segmentation variables" in MSM in Section 6.4. There are four types of characteristics: (1) geographic, (2) demographic, (3) psychographic, and (4) behavioural. The usage of these characteristics ensured a robust output of market segmentation as depicted in Section 7.4.

The eighth research question was concerned whether there will be various itineraries from a diverse set of spatio-temporal movement patterns of tourists. The outcome of the PSM in Section 7.2.5 that produced various itineraries for female and middle aged and other type of tourists confirmed this. Each outcome was guaranteed to be the most popular itinerary for each type of tourist.

The ninth research question dealt with whether there will be different sets of activities timing, i.e. arrival times and visiting durations for each attraction. The outcome of the APM in Section 7.3.1 with distinctive sets of activities timing confirmed this. Again, each outcome was guaranteed to be the most popular

activities timing for each attraction. Certain attractions highlighted a multimodality scenario where more than one significant arrival time were produced. The Activity Timing Comparison and the Activity Frequency Comparison Analyses (ATCA and AFCA) of APM also responded to the tenth research question which concerned whether there are differences between the activities' timing outcomes with the current arrangement of activities timing. Several significant differences of activities' timing were identified when the comparison was made.

Finally, the eleventh research question was concerned with whether there will be a specific set of tourist characteristics associated with each tour package product. A number of significant associations between tourists' characteristics classifications towards these tour package products were able to be ascertained in Section 7.4.4. Some relationships of tourists' characteristics between different target markets were also able to be established.

In essence, the aim of establishing a method to design tour packages was clearly achieved by answering the set of research questions. This method can be utilised by tour operators in packaging the compulsory elements of tour packages so that the specific tour package can later on serve as a useful guide during a tour. Additionally, the other two methods for the planning of activities and market segmentation can be used by both attraction administrators and tour operators to complement the process of designing tour packages to ensure the robustness of these tour packages.

Ultimately, the consideration of tourists' spatio-temporal movements and their characteristics through various modelling methods is the major research contribution of this study. Hence, these methods manage to address the research gap of the limited availability of incorporating the exact spatio-temporal movement aspect within tour packages. Apart from contributing directly to the tourism management sector, these methods can contribute to political and transportation implications via its scheduling findings. This can be achieved if the tourism regulatory agency enforces tour agencies to include scheduling information when designing tour packages. Public transportation agencies can utilise this scheduling information to device a public

transportation schedule within any specific tourism area that contains several attractions such as in Phillip Island.

8.3   Limitations of Research

Several limitations exist in this research and are discussed in this section. Most of the limitations are divided into two categories: the first one is related to modelling methods and the other one in terms of case study implementation.

The semi-Markov process was the fundamental element for the modelling part of this research to design tour packages. It was mentioned by Xia as a novel yet efficient way for the modelling of spatio-temporal movement of tourists (Xia, 2007). Unlike Xia's research, this research brought the semi-Markov process to the next level by implementing it for estimating the probability of transitional arrival times with the assistance of PDF.

Yet it comes with a limitation where owing to the usage of the second or subsequent most frequent arrival times inconsistencies could have occurred for any itinerary (refer Section 7.2.5). As understood, any itinerary was chosen as a result of that particular itinerary having the highest ranking when compared with other itineraries. When this itinerary is inconsistent it forced the usage of the second or subsequent most frequent arrival time, the combined probabilities might also decrease, dropping the ranking when the comparison is again done.

This could result in another consistent itinerary which at the beginning had a lower ranking to emerge as the one with the highest ranking. Consideration should be made either to retain or drop the earlier highest ranking itinerary if it is inconsistent, instead of replacing the arrival time to simplify the process.

Though having an overall sample size of 1,104 tourists, the preliminary process of specifying the tourists' spatio-temporal movement dataset in relation to their characteristics can be made to only include a maximum of two characteristics. This is due to the insignificant sample size of below 200 when including more than two characteristics (see Table 7.1, only a maximum three characteristics is combined).

Therefore, the PSM with its efficiency can only design tour packages for up to two characteristics. To overcome this, a higher sample size is needed.

Another limitation in regards to the case study implementation is again related to the small sample size in regards to the participation of activities in certain attractions that affects the market segmentation process (see Section 7.4). The only significant sample size is related to the Churchill Island attraction where it was found that Churchill Island had a set of proper scheduled activities when compared to other attractions where most of the activities were unscheduled.

Therefore the only activity segmentation done was for Churchill Island. Another reason for the unavailability of the activity dataset might be due to tourists thinking that specifying the starting time of each activity is a bit complicated.

8.4   Future Research Directions

Designing tour packages by means of modelling tourists' spatio-temporal movement combined with their characteristics through the application of the semi-Markov and scheduling processes are considered innovative. For that reason, there are several future research directions to further expand this research area.

It is known that PDF is able to effectively model a set of random variables (Stirzaker, 2003). However, as shown in the previous page, this is not always the case when it comes to uniform datasets, especially in terms of visiting duration. Therefore, to overcome this, it is proposed that the implementation of other measures, specifically the hazard function to describe the distribution of visiting duration dataset. Moreover, the hazard function is frequently known for duration models (Arellano, 2008).

Two limitations are identified in regards to the sample size. Primarily it involves inadequacy of the spatio-temporal movement dataset that hinders us to specify more than three characteristics for the modelling process while the other issue is the smaller sample size of tourists who completed the activities section of the questionnaire.

For the first limitation, it is suggested that the sampling process should be longer and by recruiting more than one researcher to collect those data. To solve the second limitation, the simplification for the process of obtaining activities data should be undertaken for instance by requiring tourists to mark only the activities that they participated in.

The implementation of PSM in this research only involved a day tour package whereas in reality tour packages are diverse in terms of their overall duration scope such as a three day and a whole week tour packages. Therefore, it is proposed that the PSM is used to cater for a multi day tour package that consists of more tourism products such as accommodation, meals and transportation (Syratt and Archer, 2003).

For destinations with a well organised public transportation system, it is recommended that the inclusion of an actual and real-time schedule of public transportation for the easement of tourists travelling from one place to another is undertaken (Tam and Pun-Cheng, 2012). One such example is by implementing Google transit data (Tapper, 2006).

From the case study, it can be noticed that some attractions notably Penguin Parade have a higher visitation rate due to the popularity of the smallest known penguin species. This higher visitation rate resulted in the issue of long queues where tourists have to arrive very early in order to be at the front of the queue. Frequently, this causes discomfort to certain types of tourists such as tourists who travel with kids, especially smaller babies.

In order to improve this situation, it is proposed that the Phillip Island Nature Park authority adopt the method of queuing theory where using this theory, a model to predict the queue lengths and waiting times is constructed so that the output can be used to better manage the admission process (Sundarapandian, 2009).

Finally, as Phillip Island is essentially a closed system, it is suggested that another case study area where the settings are more challenging with multiple entrances are used. Some challenges involving the data collection method and the execution of the modelling method are expected. For an open system such as Swan Valley in Western Australia, tourists come from multiple entrances. Therefore the data collection method has to be organised more efficiently to capture an adequate sample size.

An open system can also include several tourist attractions. Thus, the same modelling method has to be sufficient to incorporate more attractions and more movement patterns. The results from this new case study area could be extended to other areas in order to verify the efficiency of the methods presented in this study towards modelling tourists' spatio-temporal movement from one attraction to another.

REFERENCES

Abbey, J. R. (1979). Does life-style profiling work? *Journal of Travel Research, 18*(1), 8-14.

Abdul Khanan, M. F. (2009). *Individual differences in the wayfinding decision-making process.* (Master's thesis), Curtin University of Technology.

Abdul Khanan, M. F., Xia, J. C., & Inbakaran, R. (2012). Individual differences in the tourist wayfinding decision-making process: a case study of Phillip Island, Victoria, Australia. In S. Wenzhong, M. Goodchild , B. Lees & Y. Leung (Eds.), *Advances in Geo-spatial Information Science* (Vol. 10, pp. 245-256). Boca Raton: CRC Press.

ACS Distance Education. (2011). How to organise package tours. Retrieved April 14, 2013, from http://www.acs.edu.au/info/hospitality/tourism/package-tours.aspx

Apostolakis, A., & Jaffry, S. (2005). A choice modeling application for Greek heritage attractions. *Journal of Travel Research, 43*(3), 309-318.

Arellano, M. (2008). Duration models. Retrieved July 22, 2013, from http://www.cemfi.es/~ARELLANO/DURATION-MODELS.PDF

Arrowsmith, C., & Chhetri, P. (2003). Port Campbell National Park patterns of use: a report for the development of visitor typology as input to a generic model of visitor movements and patterns of use. Melbourne: Department of Geospatial Sciences, RMIT University.

Arrowsmith, C., Chhetri, P., & Zanon, D. (2005). Monitoring visitor patterns of use in natural tourist destinations. In C. Ryan, S. Page & M. Aicken (Eds.), *Taking Tourism to the Limits: Issues, Concepts and Managerial Perspectives* (pp. 33-52). The Netherlands: Elsevier.

Artifishal Studios. (2012). Phillip Island official visitor guide. Retrieved June 25, 2012, from http://www.visitphillipisland.com/wp-content/uploads/2012/11/Phillip-Island-Official-Visitors-Guide-2013-Final-Low-Res.pdf

Asakura, Y., & Iryo, T. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice, 41*(7), 684-690.

Askari, H. (1971). Demand for package tours. *Journal of Transport Economics and Policy*, 40-51.

Australian Bureau of Statistics. (n.d.). Sample size calculator. Retrieved October 17, 2011, from http://www.nss.gov.au/nss/home.nsf/pages/Sample+size+calculator

Australian Rail Maps. (2011). Cowes to Wonthaggi. Retrieved April 14, 2013, from http://www.railmaps.com.au/routedetails.php?RouteSelect=951

Awad, A. M. (1996). Properties of the Akaike information criterion. *Microelectronics Reliability, 36*(4), 457-464.

Batty, M. (2003). Agent-based pedestrian modelling. In P. Longley & M. Batty (Eds.), *Advanced Spatial Analysis: The CASA Book of GIS* (pp. 81-106). Redlands, California: ESRI Press.

Beane, T. P., & Ennis, D. M. (1987). Market segmentation: a review. *European Journal of Marketing, 21*(5), 20-42.

Becken, S., & Simmons, D. G. (2002). Understanding energy consumption patterns of tourist attractions and activities in New Zealand. *Tourism Management, 23*(4), 343-354. doi: http://dx.doi.org/10.1016/S0261-5177(01)00091-7

Bloom, J. Z. (2004). Tourist market segmentation with linear and non-linear techniques. *Tourism Management, 25*(6), 723-733.

Bogdanovych, A., Berger, H., Simoff, S., & Sierra, C. (2006). *Travel agents vs online booking: tackling the shortcomings of nowadays online tourism portals.* Paper presented at the ENTER E-Tourism Conference, Lausanne, Switzerland.

Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology, 44*(1), 62-91.

Brassard, J. R., & Correia, M. J. (1977). A computer program for fitting multimodal probability density functions. *Computer Programs in Biomedicine, 7*(1), 1-20.

Brendon, P. (1991). *Thomas Cook: 150 years of popular tourism*. London: Secker (Martin) & Warburg Ltd.

Budeanu, A. (2005). Impacts and responsibilities for sustainable tourism: a tour operator's perspective. *Journal of Cleaner Production, 13*(2), 89-97.

Castillo, L., Armengol, E., Onaindía, E., Sebastiá, L., González-Boticario, J., Rodríguez, A., . . . Borrajo, D. (2008). SAMAP: an user-oriented adaptive system for planning tourist visits. *Expert Systems with Applications, 34*(2), 1318-1332.

Cinlar, E. (2013). *Introduction to stochastic processes*. New York: Courier Dover Publications.

Connell, J., & Page, S. J. (2008). Exploring the spatial patterns of car-based tourist travel in Loch Lomond and Trossachs National Park, Scotland. *Tourism Management, 29*(3), 561-580.

Cooper, C. P. (1981). Spatial and temporal patterns behaviour of tourist. *Regional Studies, 15*(5), 359-371.

Creative Research System. (2012). Sample size calculator.   Retrieved November 24, 2011, from http://www.surveysystem.com/sscalc.htm#one

Crockett, A., Rafferty, A., Higgins, V., & Marshall, A. (2011). Weighting the social surveys ESDS government.   Retrieved August 27, 2012, from http: //www. esds. ac. uk/govemment/docs/Weighting. pdf

Dann, G. (1996). The people of tourist brochures. In T. Selwyn (Ed.), *The Tourist Image, Myths and Myth Making in Tourism* (pp. 61-81). Chichester: John Wiley and Sons.

DeGroot, M. H., & Schervish, M. J. (2002). *Probability and statistics*. Boston: Addison-Wesley.

Dellaert, B., Borgers, A., & Timmermans, H. (1995). A day in the city: Using conjoint choice experiments to model urban tourists' choice of activity packages. *Tourism Management, 16*(5), 347-353.

Dolnicar, S. (2002). A review of data-driven market segmentation in tourism. *Journal of Travel and Tourism Marketing, 12*(1), 1-22.

Dunstall, S., Horn, M. E. T., Kilby, P., Krishnamoorthy, M., Owens, B., Sier, D., & Thiebaux, S. . (2004). An automated itinerary planning system for holiday travel. *Information Technology and Tourism 6*, 195-210.

Enoch, Y. (1996). Contents of tour packages: a cross-cultural comparison. *Annals of Tourism Research, 23*(3), 599-616.

Enzensberger, H. M. (1996). A theory of tourism. *New German Critique*(68), 117-135.

Evans, N. G., & Stabler, M. J. (1995). A future for the package tour operator in the 21st century? *Tourism Economics, 1*(3), 245-263.

Everitt, B. (2011). *Cluster Analysis* Hoboken: Wiley.

Fennell, D. A. (1996). A tourist space-time budget in the Shetland Islands. *Annals of Tourism Research, 23*(4), 811-829.

Fesenmaier, D. R., & JiannMin, J. (2000). Assessing structure in the pleasure trip planning process. *Tourism Analysis, 5*(1), 13-27.

Fesenmaier, D. R., Ricci, F., Schaumlechner, E., Wöber, K., & Zanella, C. (2003). DIETORECS: Travel advisory for multiple decision styles. *Information and communication technologies in tourism, 2003*, 232-241.

Flognfeldt Jr, T. (1999). Traveler geographic origin and market segmentation: The multi trips destination case. *Journal of Travel and Tourism Marketing, 8*(1), 111-124.

Frank, R.E., Massey, W.F., & Wind, Y. (1972). *Market segmentation*. Madison: Prentice-Hall.

Geva, A., & Goldman, A. (1991). Satisfaction measurement in guided tours. *Annals of tourism Research, 18*(2), 177-185.

Goodwin, P. B. (1981). The usefulness of travel budgets. *Transportation Research Part A: General, 15*(1), 97-106.

Gupta, M. R., & Chen, Y. (2011). *Theory and use of the EM Algorithm*. Hanover, Maryland: Now Publishers Inc.

Gustafson, E., Nestler, J., Gross, L., Reynolds, K., Yaussy, D., Maxwell, T., & Dale, V. (2003). Evolving approaches and technologies to enhance the role of ecological modeling in decision-making. In H. H. Dale (Ed.), *Ecological Modeling for Resource Management* (pp. 135-164). New York: Springer.

He, B., & Zhang, Y. (n.d.). Clustering documents in large text corpora. Retrieved February 14, 2012, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.119.8839&rep=rep1&type=pdf

Heung, V. C. S., & Chu, R. (2000). Important factors affecting Hong Kong consumers' choice of a travel agency for all-inclusive package tours. *Journal of Travel Research, 39*(1), 52-59.

Hsu, F. C., Chen, J. S., & Chen, P. (2000, May, 2000). *Interactive genetic algorithms for a travel itinerary planning problem.* Paper presented at the The 2nd Asia Pacific Conference on GAs and Applications,, Hong Kong.

Hsu, T. K., Tsai, Y. F., & Wu, H. H. (2009). The preference analysis for tourist choice of destination: A case study of Taiwan. *Tourism Management, 30*(2), 288-297.

Hunter, F. R. (2004). Tourism and empire: the Thomas Cook & Son Enterprise on the Nile, 1868–1914. *Middle Eastern Studies, 40*(5), 28-54.

Iosifescu, M., Limnios, N., & Oprisan, G. (2010). *Introduction to stochastic models*. Hoboken, NJ: Wiley.

Jafari, J. (2000). *Encyclopedia of tourism*. London: Routledge.

Jakkilinki, R., Georgievski, M., & Sharda, N. (2007). *Connecting destinations with an ontology-based e-tourism planner*. Paper presented at the ENTER E-Tourism Conference, Ljubljana.

Kale, S. H., Mcintyre, R. P., & Weir, K. M. (1987). Marketing overseas tour packages to the youth segment: an empirical analysis. *Journal of Travel Research, 25*(4), 20-24.

Kaynak, E., & Yavas, U. (1981). Segmenting the tourism market by purpose of trip. *International Journal of Tourism Management, 2*(2), 105-112.

Kemeny, J. G., & Snell, J. L. (1960). *Finite Markov chains*. Princeton, NJ: Van Nostrand.

Kim, H. S., & Cho, S. B. (2000). Application of interactive genetic algorithm to fashion design. *Engineering Applications of Artificial Intelligence, 13*(6), 635-644.

Knight, T. E. (1974). An approach to the evaluation of changes in travel unreliability: a "safety margin" hypothesis. *Transportation, 3*(4), 393-408.

Kotler, P.J., & Keller, K.L. (2006). *Marketing Management*. Upper Saddle River, NJ: Prentice Hall Higher Education.

Lau, G., & McKercher, B. (2006). Understanding tourist movement patterns in a destination: a GIS approach. *Tourism and Hospitality Research, 7*(1), 39-49.

Leiper, N. (1990). Tourist attraction systems. *Annals of Tourism Research, 17*(3), 367-384.

Leisen, B. (2001). Image segmentation: the case of a tourism destination. *Journal of Services Marketing, 15*(1), 49-66.

Lew, A. A., & McKercher, B. (2006). Modelling tourist movement: a local destination analysis. *Annals of Tourism Research 33*(2), 403-423.

Lin, D. (2013). An Introduction to Expectation-Maximization. Retrieved April 26, 2013, from https://courses.csail.mit.edu/6.867/wiki/images/b/b5/Em_tutorial.pdf

Liu, Y., Ram, S., Lusch, R. F., & Brusco, M. (2010). Multicriterion market segmentation: a new model, implementation, and evaluation. *Marketing Science, 29*(5), 880-894.

Loban, S. R. (1997). A framework for computer-assisted travel counseling. *Annals of Tourism Research, 24*(4), 813-834.

Logitravel. (n.d). Does my package holiday include a travel insurance? Retrieved June 22, 2013, from http://www.logitravel.co.uk/preguntas-frecuentes/holidays-does-my-package-holiday-include-a-travel-insurance-1_159.html

Lopez, B. (2003). *Holiday scheduling for city visitors.* Paper presented at the ENTER E-Tourism Conference, Helsinki.

Low, B. T., Cheng, C. H., Wong, K. F., & Motwani, J. (1996). An expert advisory system for the tourism industry. *Expert Systems with Applications, 11*(1), 65-77.

Lusseau, D. (2003). Effects of tour boats on the behavior of bottlenose dolphins: using Markov chains to model anthropogenic impacts. *Conservation Biology, 17*(6), 1785-1793.

Mathworks. (2013). Slice Sample. Retrieved July 25, 2013, from http://www.mathworks.com.au/help/stats/slicesample.html

McKercher, B., & Lew, A. A. (2004). Tourist flows and the spatial distribution of tourists. In A. A. Lew, C. M. Hall & A. M. Williams (Eds.), *A Companion to Tourism* (pp. 36-48). Malden, Maryland: Blackwell Publishing.

McLachlan, G. J., & Thriyambakam, K. (2008). *The EM algorithm and extensions.* Hoboken, N.J.: Wiley-Interscience.

McLafferty, S. L. (2010). Conducting questionnaire wurveys. In N. Clifford, S. French & G. Valentine (Eds.), *Key Methods in Geography*. London: Sage Publications.

Mednick, H. (1975). A Markov chain model of travel patterns of US visitors to Ontario. *Journal of Leisure Research, 14*(26), 246-255.

Merriam-Webster. (2010). Tourism. Retrieved April 1, 2010, from http://www.merriam-webster.com/dictionary/tourism

Mohammadian, A., & Doherty, S. T. (2006). Modeling activity scheduling time horizon: duration of time between planning and execution of pre-planned activities. *Transportation Research Part A: Policy and Practice, 40*(6), 475-490.

Moore, D. S. (2001). *Statistics : concepts and controversies*. New York: W. H. Freeman.

Moore, D.S., McCabe, G.P., & Craig, B.A. (2012). *Introduction to the Practice of Statistics*. New York: W.H. Freeman & Company.

Morrison, A. M. (1996). *Hospitality and travel marketing*. Albany, NY: Delmar Publishers.

Murthy, D.N.P., Page, N.W., & Rodin, E.Y. (1990). *Mathematical modelling: a tool for problem solving in engineering, physical, biological, and social sciences*. Oxford: Pergamon Press.

Myers, J. H. (1996). *Segmentation and positioning for strategic marketing decisions*. Chicago: American Marketing Association.

Nicolau, J. L., & Más, F. J. (2005). Stochastic modeling: a three-stage tourist choice process. *Annals of Tourism Research, 32*(1), 49-69.

Niknafs, A. A., Shiri, M. E., & Javidi, M. M. (2003, September 1-5, 2003). *A case-based reasoning approach in e-tourism: tour itinerary planning*. Paper presented at the The 14th International Workshop on Database and Expert System Applications, Prague.

Nilsson, J. H. (2011). Service in the air: a century of hospitality aboard passenger airlines. Lund: Lund University.

O'Connor, A., Zerger, A., & Itami, B. (2005). Geo-temporal tracking and analysis of tourist movement. *Mathematics and Computers in Simulation, 69*(1), 135-150.

O'Reilly, C. C. (2006). From drifter to gap year tourist: Mainstreaming backpacker travel. *Annals of Tourism Research, 33*(4), 998-1017.

Ott, L.R., & Longnecker, M. (2010). *An introduction to statistical methods and data analysis*. Belmont, California: Brooks/Cole Cengage Learning.

Oxford Dictionaries. (2013). Utility function.   Retrieved December 16, 2013, from http://www.oxforddictionaries.com/definition/english/utility-function

Page, S. (2005). *Transport and tourism*. Harlow: Prentice Hall.

Perlitz, L., & Elliott, S. G. (2000). *The products*. French Forest, NSW: Prentice Hall.

Pfingsthorn, M., & Birk, A. (2013). Simultaneous localization and mapping (SLAM) with multimodal probability distributions. *The International Journal of Robotics Research, 32*(143). doi: 10.1177/0278364912461540

Phillip Island Nature Park. (2005). How to get to Phillip Island.   Retrieved April, 21, 2010, from http://www.penguins.org.au

Phillip Island Nature Park. (2011). Strategic plan 2012 - 2017. from http://www.penguins.org.au/assets/About/PDF-Local-Community/Strategic-Plan-2012-17sm.pdf

Phillip Island Nature Park. (2012). Buy tickets.   Retrieved January 22, 2012, from https://www.penguins.org.au/buy-tickets/

Phillip Island Nature Park. (2013). Annual Report 2012 - 2013. Cowes, Victoria: Phillip Island Nature Park.

Phillip Island Nature Park. (n.d.). Attractions.   Retrieved December 23, 2011, from http://www.penguins.org.au/attractions/

Phillip Island Visitor Information Centre. (2013). The Nobbies.   Retrieved August 22, 2014, from http://www.visitphillipisland.com/listing/the-nobbies/

Phillips, R. G., & Webster, S. (1983). *Group travel operating procedures*. New York: CBI.

Pride, William M., Elliott, G., Rundle-Thiele, S., Waller, D., Paladino, A., & Ferrell, O. (2007). *Marketing : core concepts and applications*. Milton, QLD: John Wiley.

Pröll, B., Retschitzegger, W., & Wagner, R. R. (1999). *Holiday packages on the web.* Paper presented at the ENTER E-Tourism Conference, Innsbruck.

Quiroga, I. (1990). Characteristics of package tours in Europe. *Annals of Tourism Research, 17*(2), 185-207.

Ramakrishnan, R., & Gehrke, J. (2003). *Database Management Systems*. Boston: McGraw-Hill Education.

Recker, W. W. (1995). The household activity pattern problem: general formulation and solution. *Transportation Research Part B: Methodological, 29*(1), 61-77.

Recker, W. W., McNally, M. G., & Root, G. S. (1986a). A model of complex travel behavior: part 1 - theoretical development. *Transportation Research Part A: General, 20*(4), 307-318.

Recker, W. W., McNally, M. G., & Root, G. S. (1986b). A model of complex travel behavior: part 2 - an operational model. *Transportation Research Part A: General, 20*(4), 319-330.

Reliasoft. (2002). Defining distributions. Retrieved August 31, 2013, from https://www.weibull.com/hotwire/issue12/relbasics12.htm

Scheaffer, R. L. (2006). *Elementary survey sampling*. Belmont, CA: Thomson Brooks/Cole.

Sheldon, P. J., & Mak, J. (1987). The demand for package tours: a mode choice model. *Journal of Travel Research, 25*(3), 13-17.

Sirakaya, E., McLellan, R. W., & Uysal, M. (1996). Modeling vacation destination decisions: a behavioral approach. *Journal of Travel and Tourism Marketing, 5*(1-2), 57-75.

Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *The Journal of Marketing, 21*(1), 3-8.

Stewart, W.J. (1994). *Introduction to the Numerical Solution of Markov Chains*. Princeton, NJ: Princeton University Press.

Stirzaker, D. (2003). *Elementary Probability*. Cambridge: Cambridge University Press.

Sundarapandian, V. (2009). *Probability Statistics And Queuing Theory*. Bonn, Germany: Prentice-Hall Of India Pvt. Limited.

Syratt, G., & Archer, J. (2003). *Manual of travel agency practice*. Oxford: Buttweworth-Heinemann.

Tam, H., & Pun-Cheng, L. S. C. (2012). Evaluation of online itinerary planner and investigation of possible enhancement features. In A. G. O. Yeh, S. Wenzhong, Y. Leung & C. Zhou (Eds.), *Advances in Spatial Data Handling and GIS* (pp. 195-210). Berlin: Springer.

Tan, B., & Yılmaz, K. (2002). Markov chain test for time dependence and homogeneity: an analytical and empirical evaluation. *European Journal of Operational Research, 137*(3), 524-543.

Tapper, R. (2006). *Wildlife watching and tourism: a study on the benefits and risks of a fast growing tourism activity and its impacts on species*: UNEP/CMS Convention on Migratory Species and TUI.

Thomson, C. M., & Pearce, D. G. (1980). Market segmentation of New Zealand package tours. *Journal of Travel Research, 19*(2), 3-6.

Tourism Research Australia. (2012). Regional tourism orofile for Phillip Island 2011/12. Retrieved November 11, 2011, from http://www.ret.gov.au/tourism/documents/tra/regional%20tourism%20profiles/2012/regionaltourismprofiles_2011-12.xls

Tourism Victoria. (2008). Tourism Victoria annual report. Retrieved July 23, 2011, from http://www.tourism.vic.gov.au/IMAGES/STORIES/DOCUMENTS/STRATEGIESANDPLANS/TV-ANNUAL-REPORT-07-08.PDF

Tourism Victoria. (2011). Facts and figures. Retrieved November 24, 2011, from http://www.tourism.vic.gov.au/facts-and-figures/

Tourism Victoria. (2012a). Domestic tourism expenditure in Victoria. Retrieved April 27, 2012, from http://www.tourism.vic.gov.au/IMAGES/STORIES/DOMESTIC%20TOURISM%20EXPENDITURE%20IN%20VICTORIA%20YEAR%20ENDING%20DECEMBER%202008-2012.PDF

Tourism Victoria. (2012b). Facts and figures. Retrieved May 27, 2012, from http://www.tourism.vic.gov.au/facts-and-figures/

Trombe, P. J., Pedersen, M. W., & Madsen, H. (2011). Chapter 4 - estimation by the EM algorithm. Retrieved March 30, 2012, from http://www2.imm.dtu.dk/~mawp/02433/ch4_slides.pdf

Tussyadiah, I. P., & Fesenmaier, D. R. (2007). *Interpreting tourist experiences from first-person stories: a foundation for mobile guides.* Paper presented at the The 15th European Conference on Information Systems, St Gallen.

Wang, K. C., Hsieh, A. T., Chou, S. H., & Lin, Y. S. (2007). GPTCCC: An instrument for measuring group package tour service. *Tourism Management, 28*(2), 361-376.

Wang, K. C., Hsieh, A. T., & Huan, T. C. (2000). Critical service features in group package tour: An exploratory research. *Tourism Management, 21*(2), 177-189.

Wang, K. C., Hsieh, A. T., Yeh, Y. C., & Tsai, C. W. (2004). Who is the decision-maker: the parents or the child in group package tours? *Tourism Management, 25*(2), 183-194.

Wanhill, S. R. C. (1980). Charging for congestion at tourist attractions. *International Journal of Tourism Management, 1*(3), 168-174.

Weaver, D. B., & Lawton, L. (2010). *Tourism management*. Milton, QLD: John Wiley & Sons.

Wedel, M., & Kamakura, W. A. (2000). *Market segmentation : conceptual and methodological foundations*. Boston: Kluwer Academic.

Weiss, N. A., Hardy, M., & Holmes, P. T. (2006). *A course in probability*. Boston: Pearson/Addison Wesley.

Western Port Water. (2012). Phillip Island Recycle Water Scheme. Retrieved April 19, 2012, from http://www.westernportwater.com.au/WEBFILES/SERVICES/RECYCLED%20WATER/REGIONAL%20HEALTH%20AND%20ENVIRONMENTAL%20PLAN%202012.PDF

Wikidot. (2011). Chapter 7: normal distribution. Retrieved August 31, 2013

Witten, I.H., & Frank, E. (2005). *Data mining: practical machine learning tools and techniques*. San Francisco: Morgan-Kauffmann Publishers.

Wong, C. U. I., & McKercher, B. (2012). Day tour itineraries: searching for the balance between commercial needs and experiential desires. *Tourism Management, 33*(6), 1360-1372.

Xia, J. C. (2007). *Modelling the spatial-temporal movement of tourists.* (PhD thesis), RMIT University, Melbourne.

Xia, J. C., & Arrowsmith, C. (2005, December 12-15, 2005). *Managing scale issues in spatio-temporal movement of tourists modelling.* Paper presented at the The International Congress on Modelling and Simulation, Melbourne.

Xia, J. C., Arrowsmith, C., Jackson, M., & Cartwright, W. (2008). The wayfinding process relationships between decision-making and landmark utility. *Tourism Management, 29*(3), 445-457.

Xia, J. C., Evans, F. H., Spilsbury, K., Ciesielski, V., Arrowsmith, C., & Wright, G. (2010). Market segments based on the dominant movement patterns of tourists. *Tourism Management, 31*(4), 464-469.

Xia, J. C., & Zeephongsekul, P. (2009, July, 2009). *Space and time analysis of tourist movements using semi-Markov processes.* Paper presented at the The 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Cairns.

Xia, J. C., Zeephongsekul, P., & Arrowsmith, C. (2009). Modelling spatio-temporal movement of tourists using finite Markov chains. *Mathematics and Computers in Simulation, 79*(5), 1544-1553.

Xia, J. C., Zeephongsekul, P., & Packer, D. (2011). Spatial and temporal modelling of tourist movements using semi-Markov processes. *Tourism Management, 32*(4), 844-851.

Zhang, L. (2008). Probability density functions.   Retrieved March 21, 2012, from http://www.math.utah.edu/~lzhang/teaching/3070summer2008/DailyUpdates/jun26/sec4_1.pdf

APPENDIX 1: RESEARCH'S QUESTIONNAIRE

January 2012

**Department of
Spatial Sciences**

Excellence in:
- Surveying
- GIS
- Mapping
- Mine and Engineering Surveying

GPO Box U1987
Perth Western Australia 6845

Telephone + 61 8 9266 7566
Facsimile + 61 8 9266 2703
Email   spatial@curtin.edu.au

Project title: The Development of a Tour Package Method through the Modelling of Spatio-Temporal Movement of Tourists

Dear Participant,

My name is Mohd Faisal Bin Abdul Khanan and I am a Doctor of Philosophy's candidate at Curtin University's Department of Spatial Sciences. My Supervisor is Dr Cecilia Xia. We are currently undertaking a joint project as above with the Phillip Island Nature Park that aims to develop a tour package method by understanding tourist movement. This project has been approved by the Curtin University's Human Research Ethics Committee, approval no. SMEC-53-10.

Your participation in this research will involve filling in a questionnaire designed to ascertain basic movement patterns and socio-demographic data. The questionnaire will take no longer than 10 minutes of your time. The questionnaire is designed to find out more about how you tour Phillip Island. The collected data will make it possible to quantify and classify current park visitors based upon your socio-demographic background and movement patterns. Tourist movement modelling will be developed based on this survey. This will allow park managers to offer more customised services such as tour packages and develop more effective and efficient strategies for managing natural and recreational resource in parks. Participation in this research is voluntary and you may withdraw at any time. We welcome you to examine the study materials before deciding whether you wish to participate.

The data collected will be analysed and aggregated for publishing in professional journals. Because the data will be generalised individual responses will not be identified and your anonymity will be protected. All individual responses will be destroyed at the end of the research.

If you have any queries regarding this project, please contact Mohd Faisal Bin Abdul Khanan on (08) 9266, 4255, e-mail m.abdulkhanan@postgrad.curtin.edu.au or Dr Cecilia Xia (BSc, MSc, PhD (RMIT)) on (08) 9266 7563, e-mail c.xia@curtin.edu.au.

Yours faithfully

Mohd Faisal Bin Abdul Khanan and Dr Cecilia Xia

Form No. _____

Date: _____   Time: _____

Location where questionnaire is handed: _____

Weather (please tick one box only):

Sunny ☐   Fine ☐   Cloudy ☐   Light shower ☐   Rain ☐

Q1. How many times have you visited Phillip Island, including today's trip?

☐ Once
☐ Twice
☐ Three times
☐ More than three times

Q2. Why are you visiting Phillip Island today? Please tick all that apply.

☐ Holiday, leisure, recreation
☐ Educational purpose
☐ Business purpose
☐ Accompany friends, relatives
☐ Others _____

Q3. What type of attractions that attract you most in your Phillip Island trip today? Please tick one box only.

☐ Sightseeing(i.e.: Woolamai)
☐ Wildlife (i.e.: Penguin Parade)
☐ Heritage (i.e.: Churchill Island)
☐ Culture (i.e.: Cowes Cultural Centre)
☐ Adventure (i.e.: surfing at Rhyll Inlet)

Q4. In terms of the type of group you are travelling, are you travelling with: (Please tick one box only).

☐ Travel alone
☐ Partner only
☐ Partner with children
☐ Friends or relatives
☐ Tour group
☐ Special group (business/ research)

Q5. How long will you spend for your trip in Phillip Island?

☐ Half day
☐ Whole day
☐ 2-3 days
☐ 4-5 days
☐ 6-7 days
☐ More than a week

Q6. Are you a:

☐ Male
☐ Female

Q7. Which age group are you in?

	18-24
	25-34
	35-44
	45-54
	55-64
	Above 64

Q8. What is your place of residence? Please fill in one box only.

Australia: State

Overseas: Country

Q9. What is your annual gross household income?

	Below AUD 50,000
	AUD 50,001-AUD 70,000
	AUD 70,001-AUD 90,000
	AUD 90,001-AUD 110,000
	AUD 110,001-AUD 130,000
	Above AUD 130,000

Q10. Which best describes the highest level of education you have reached? Please tick one box only.

	Primary
	Secondary-partial
	Secondary-completed
	Tertiary-undergraduate
	Tertiary-postgraduate

Q10. Which best describes the highest level of education you have reached? Please tick one box only.

	Primary
	Secondary-partial
	Secondary-completed
	Tertiary-undergraduate
	Tertiary-postgraduate

Q11. What type of travel mode do you use during your tour in Phillip Island today?

	Walking
	Motorcycle
	Passenger vehicle (i.e car or van
	4WD
	Campervan/motor home
	Tour bus/coach

Q12. Could you write down the approximate time you arrive and duration spent (in hours) at each attraction for today's trip only? Please also write down the approximate time you started participating in each activities as stated below.

Cowes

Arrival time	
Duration spent	
	Starting time
Visit Cowes Cultural Centre	
Visit museum	
Shopping	
Dining	

Rhyll Inlet

Arrival time	
Duration spent	
	Starting time
Nature walks (Oswin Roberts, Conservation Hill, Rhyll swamp bird Sanctuary)	
Visit historic sites	

Ventnor

Arrival time	
Duration spent	

The Nobbies

Arrival time	
Duration spent	
	Starting time
Presentations at the Visitors Centre	
Just for pups kids session	
Boardwalk	
Seal viewing at Seal Rock	

Churchill Island Heritage Farm

Arrival time	
Duration spent	
	Starting time
Cow milking	
Wagon rides	
Blacksmith demo	
Working dogs	
Sheep shearing	
Whip cracking and boomerang throwing	
Visit baby animal	
Visit heritage buildings and gardens	
Animal viewing	

Penguin Parade

Arrival time	
Duration spent	
	Starting time
Interactive Display at the Visitors Centre	
Penguin viewing	
Ranger guided tour	

Koala Conservation Centre

Arrival time	
Duration spent	
	Starting time
Koala Boardwalk	
Woodland Boardwalk	
Wildlife Bush Walk	
Interactive Learning at the Visitors Centre	

Cape Woolamai

Arrival time	
Duration spent	
	Starting time
Coastal track walking (lighthouse and pinnacles)	
Wildlife viewing	

Visitors Information Centre

Arrival time	
Duration spent	

Phillip Island Rd
Cowes - Rhyll Rd
Lock Rd
Rhyll - Newhaven Rd
Ventnor Rd
Ventnor Beach Rd
BASS COAST
Back Beach Rd

APPENDIX 2: VBA PSEUDOCODE'S TO DERIVE PROBABILITIES OF MOVEMENT SEQUENCES AND TRANSITIONAL ARRIVAL TIMES

```vba
'Codes to derive probabilities of movement sequences
Sub MoveSeq()

'Declare the entire variables
Dim biprobability As Variant
Dim Monoprobability As Double
Dim Result As Variant
Dim AP As Variant
Dim Dest As Variant
Dim Movement As Variant
Dim P(0 To 100) As Variant
Dim D As Variant
Dim A As String
Dim BP(0 To 100, 0 To 100) As Variant

Dim i As Long
Dim j As Long
Dim k As Long
Dim T As Long
Dim DN1 As Integer
Dim DN2 As Integer

 'Establish array values from some of the variables above
  biprobalility = Range("B6:K14").Value
  Monoprobalility = Range("B21:K21").Value
  Movement = Range("O2:O13").Value
  AP = Range("B18:K18").Value
  Dest = Range("B22:K22").Value '
  Result = Range("P2:P13").Value
  Range("U3").Value = biprobalility(3, 2)
```

```
'Go through each element of the movement sequence array
 For i = LBound(Movement, 1) To UBound(Movement, 1)


    T = Len(Movement(i, 1))
    Result(i, 1) = 1
     A = Mid(Movement(i, 1), 1, 1)
     Range("V3").Value = Monoprobalility(1, 3)


    'Find the individual spatial movement transition probability for each transition
     For k = 1 To 10
       If A = AP(1, k) Then
         P(1) = Monoprobalility(1, k)
         DN1 = Dest(1, k)
       End If
     Next k
     Range("V3").Value = P(1)
     Result(i, 1) = Result(i, 1) * P(1)
     Range("V3").Value = Result(i, 1)


  'Multiply the individual spatial movement transition probability for a specific
   movement sequence to form the final movement sequence probability
    'For T = Len(Movement(i, 1)) To 1 Step -1
       For j = 2 To T
        For k = 1 To 10
          If Mid(Movement(i, 1), j, 1) = AP(1, k) Then
          DN2 = Dest(1, k)
           P(j) = biprobalility(DN1, DN2)
           End If
         Next k
         Result(i, 1) = Result(i, 1) * P(j)
         DN1 = DN2
         Next
  Next i
```

```vba
    Range("P2:P13").Value = Result


End Sub


'Codes to derive probabilities of transitional arrival time
Sub ArrTime()


'Declare the entire variables
Dim biprobability1 As Variant
Dim Monoprobability1 As Double
Dim Result1 As Variant
Dim AP1 As Variant
Dim Dest1 As Variant
Dim Movement1 As Variant
Dim P1(0 To 100) As Variant
Dim D1 As Variant
Dim A1 As String
Dim BP1(0 To 100, 0 To 100) As Variant


Dim i1 As Long
Dim j1 As Long
Dim k1 As Long
Dim T1 As Long
Dim DN11 As Integer
Dim DN21 As Integer


 'Establish array values from some of the variables above
  biprobalility1 = Range("B27:K35").Value
  Monoprobalility1 = Range("B40:K40").Value
  Movement1 = Range("O2:O13").Value
  AP1 = Range("B39:K39").Value
  Dest1 = Range("B41:K41").Value '
  Result1 = Range("Q2:Q13").Value
```

```
Range("U7").Value = biprobalility1(3, 2)


'Go through each element of the movement sequence array
  For i1 = LBound(Movement1, 1) To UBound(Movement1, 1)


    T1 = Len(Movement1(i1, 1))
    Result1(i1, 1) = 1
     A1 = Mid(Movement1(i1, 1), 1, 1)
     Range("V7").Value = Monoprobalility1(1, 3)


    'Find the individual transitional arrival time probability for each transition
     For k1 = 1 To 10
       If A1 = AP1(1, k1) Then
         P1(1) = Monoprobalility1(1, k1)
         DN11 = Dest1(1, k1)
       End If
     Next k1
     Range("V7").Value = P1(1)
     Result1(i1, 1) = Result1(i1, 1) * P1(1)
     Range("V7").Value = Result1(i1, 1)


    'Multiply the individual transitional arrival time probability for a specific
     movement sequence to form the final transitional arrival time probability
      'For T1 = Len(Movement1(i1, 1)) To 1 Step -1
        For j1 = 2 To T1
          For k1 = 1 To 10
            If Mid(Movement1(i1, 1), j1, 1) = AP1(1, k1) Then
            DN21 = Dest1(1, k1)
             P1(j1) = biprobalility1(DN11, DN21)
            End If
          Next k1
          Result1(i1, 1) = Result1(i1, 1) * P1(j1)
          DN11 = DN21
```

```
        Next j1
    Next i1
     Range("Q2:Q13").Value = Result1


End Sub
```